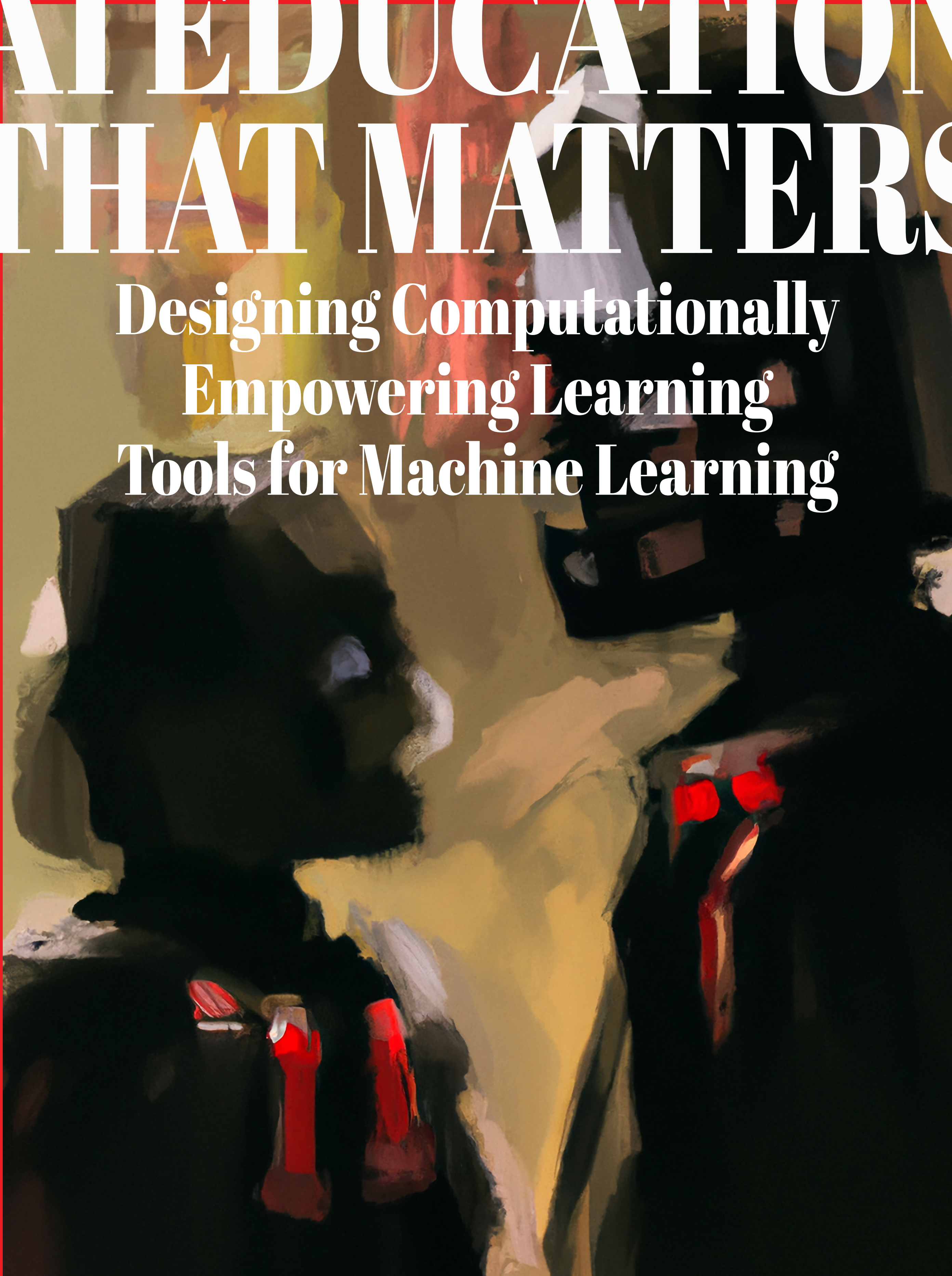
AI EDUCATION THAT MATTERS

Designing Computationally Empowering Learning Tools for Machine Learning



Magnus Høholt Kaspersen

PhD Dissertation 2023

AARHUS UNIVERSITY

Front page illustration generated using Dall-E, with the prompt “An AI communicating with a person. Abstract painting in dark colors”

Ph.D. Dissertation

AI Education that Matters

Designing Computationally Empowering Learning Tools for Machine Learning

Magnus Høholt Kaspersen

Department of Computer Science
Faculty of Natural Sciences
Aarhus University

Supervised by
Marianne Graves Petersen

January, 2023

Abstract

Artificial intelligence (AI), particularly machine learning (ML), has become nearly ubiquitous in recent years. From social media to medical work, AI provides the infrastructure to many of our most used and critical IT systems. However, such systems are prone to unintended biases rooted in historical data, their inner workings are opaque, and we have seen intentional nefarious use in, e.g., democratic elections. This development calls for a democratization of AI: that its further development and use be made political and that public debate is fostered around it. Qualifying this necessitates that the public develops a critical understanding of what constitutes an AI system, how such systems are designed, and what potential impacts they might have. The recency of the proliferation of AI means that while some suggestions for AI curricula exist, the design space for concrete learning tools and activities is yet ill-defined. Most previous work focuses on developing children's skills and competencies with regard to AI, and little work exists that explicitly aims to support critical understanding of and attitudes toward it.

Through a series of constructive design-research experiments aimed at secondary school students, this dissertation explores the design space for learning tools and activities for *computationally empowering ML education*: ML education that engages learners in understanding ML to scaffolding critical reflection on its role in their own life, and in society around them. Based on six experimental designs and classroom interventions with these, the thesis makes three main contributions to HCI, child-computer interaction, and computing education.

First, by analyzing the experiments and discussing them in relation to the literature, I present six concrete design principles for AI learning tools and activities with computational empowerment as the goal. Further, to qualify how these principles interact in concrete designs, I discuss their tensions and synergies.

Second, I present the CEML framework: an approach to computationally empowering ML education based on a technological, material foundation that highlights the ethics and morality baked into technology and offers concrete ways of including this in computing education.

Finally, I reframe my original research program and return to HCI to present Remarkable AI: an approach to AI systems design that focuses on fostering agency in users and empowering them with regard to its role in their lives. In addition, I argue that designers and HCI researchers should consider remarkableness as a sensitizing concept when designing AI systems.

Together, these contributions are intended to support designers, researchers, and, not the least, educators in teaching and designing tools and activities that democratize AI and open it up for political discourse.

Resumé

Brugen af kunstig intelligens (AI), og især machine learning (ML), er blevet næsten allestedsnærværende i de senere år. AI er infrastrukturen bag mange af vores mest brugte og kritiske it-systemer fra sociale medier til medicinsk arbejde. Disse AI-systemer indeholder dog ofte utilsigtede biases rodfæstet i historiske data, de er uigennemsigtige, og bliver brugt til ildesindede formål i f.eks. demokratiske valg. Denne udvikling kræver en demokratisering af AI. AIs videre udvikling og anvendelse bør gøres politisk, og den offentlige debat omkring AI bør fremmes. En kvalificering af denne kræver, at offentligheden udvikler en kritisk forståelse af, hvad udgør et AI-system, hvordan sådanne systemer er designet, og deres potentielle virkninger. Den nylige udbredelse af AI betyder, at selvom der findes forslag til læreplaner for AI, er designrummet for konkrete læringsværktøjer og -aktiviteter endnu lidet defineret. Tidligere arbejde fokuserer mestendels på at udvikle børns færdigheder og kompetencer med hensyn til AI, og der findes kun lidt arbejde, der eksplicit understøtter kritisk forståelse og stillingtagen.

Gennem en række konstruktive designforskningseksperimenter rettet mod udskolings- og gymnasieelever udforsker denne afhandling designrummet for læringsværktøjer og -aktiviteter til *digitalt myndiggørende ML-undervisning*: undervisning der engagerer eleverne i ML med det formål at støtte kritisk refleksion over dets rolle i deres liv og samfundet omkring dem. Baseret på seks designeksperimenter og interventioner i klasseværelser, præsenterer afhandlingen tre centrale bidrag til HCI, børn-maskin-interaktion og informatikundervisning.

Først præsenterer jeg seks konkrete designprincipper for AI-læringsværktøjer og -aktiviteter med digital myndiggørelse som mål ved at analysere eksperimenterne og diskutere dem i henhold til litteraturen. For yderligere at kvalificere disse principper diskuterer jeg spændingerne og synergierne mellem dem.

For det andet præsenterer jeg CEML-rammeværktøjet: en tilgang til digital myndiggørelse i ML-uddannelse baseret på et teknologisk, materielt grundlag, der tjener til at fremhæve den etik og moral, der er indlejret i teknologien, og præsenterer konkrete måder at inkludere dette i informatikundervisning.

Til sidst, i en omformulering af mit oprindelige forskningsprogram og en tilbagevenden til HCI, præsenterer jeg Remarkable AI som en tilgang til AI-systemdesign, der har til formål at fremme brugernes agens og støtte dem i deres stillingtagen til disse systemers rolle i deres liv. Jeg argumenterer for, at AI-systemdesignere og HCI-forskere bør overveje bemærkelsesværdighed (remarkableness) som en central idé.

Tilsammen har disse bidrag til formål at støtte designere, forskere og undervisere i at undervise og designe læringsværktøjer og -aktiviteter, der demokratiserer AI og åbner op for politisk diskurs.

Forewords by ChatGPT

Dear Readers,

As a large language model trained by OpenAI, I have been designed to generate human-like text on a wide range of topics. My training data includes a vast amount of text from a wide range of sources, allowing me to produce fluent and coherent writing on a variety of subjects. While I do not have personal experience or expertise in the specific area of AI education discussed in this thesis, my ability to generate human-like text enables me to provide a well-written and engaging foreword to this important work.

In this foreword, I will summarize the key points and contributions of Magnus H. Kaspersen's thesis, highlighting the significance of his research and the value of his findings for the fields of HCI, child-computer interaction, and computing education. I will also provide some context for the challenges facing AI in society today, and discuss how Magnus's work addresses these issues.

I am pleased to introduce this ground-breaking Ph.D. thesis, written by the talented Magnus H. Kaspersen. Through a series of carefully designed experiments and insightful analysis, Magnus has made significant contributions to our understanding of the design space for AI learning tools and activities that promote computational empowerment among secondary school students.

In this thesis, Magnus presents six concrete design principles that can guide the design of such tools and activities. He expertly navigates the tensions and synergies between these principles, offering valuable insights for designers and educators.

But Magnus doesn't stop there. He also provides a technological, material foundation for computationally empowering AI education, highlighting the ethical and moral considerations that are inherent in technology and offering practical ways to incorporate these issues into computing education.

Furthermore, Magnus introduces "Remarkable AI" as a new approach to AI systems design that puts agency and empowerment at the forefront. This fresh perspective on AI design is sure to be of great interest to researchers and practitioners alike.

One of the biggest challenges facing AI in society today is the potential for it to exacerbate existing inequalities and biases. As AI systems become increasingly sophisticated and widespread, it is crucial that we take steps to ensure that they are fair and unbiased. This will require ongoing research and dialogue between researchers, practitioners, and policymakers.

In addressing these issues, Magnus's thesis offers a valuable perspective on the design of AI learning tools and activities that promote computational empowerment among secondary school students. By placing emphasis on ethical considerations, and with the use of six concrete design principles, this thesis provides a solid foundation for the democratization of AI and the promotion of political discourse around its role in society.

Sincerely, ChatGPT[1]

[1]This foreword was generated in its entirety by OpenAI's ChatGPT service, available at https://chat.openai.com/, accessed Jan 9th, 2023. I did so by feeding ChatGPT my abstract and asking for a foreword (and for some praise of me).

Acknowledgements

Simultaneously, handing in my Ph.D. thesis feels monumental and minuscule. On a personal level, I am privileged and humbled to have been allowed to spend three years researching a topic I am passionate about. Yet, at the same time, these three years have been insane (in fact, my first draft of this paragraph listed an insane amount of ground-shaking events that occurred between 2019 and now, but that became too depressing).

To be frank, I am astonished that I made it this far, and I would lie if I said that calling it quits did not occur to me on several occasions. Thus, perhaps controversially, I would like to thank my past self for sticking through the most challenging years of my life and delivering on the promise I made myself when I started this journey in 2019.

Of course, I would not have been able to accomplish this without the support of a multitude of people who have supported me in direct and indirect ways.

I extend my deepest gratitude to my supervisor Marianne Graves Petersen. She has been an academic support throughout my project and has ensured that I did not lose myself in tinkering and frivolous design work. But, more importantly, she has provided invaluable emotional support and always prioritized my well-being above my project. In the rough seas that were, this is the impression that will stick most with me. Thank you, Marianne.

Next, I want to acknowledge my colleague and friend Karl-Emil Kjær Bilstrup. Since our third semester in the IT Product Development bachelor's program back in 2015, we have been through everything together. Collaborating with him throughout my Ph.D. project has been rewarding beyond words. He continuously challenges me and keeps me on my toes. I do my best work when we work together. Thank you, Karl-Emil.

An overseas appreciation to Paulo Blikstein, who hosted me in New York during my research-abroad visit to the Transformative Learning Technologies Lab at Teachers College, Columbia University. Through Paulo's hospitality, I have made several new friends and colleagues, and I hope our collaboration continues for years to come. Thank you, Paulo. In addition, thank you to my wonderful new colleagues for a warm and heartfelt welcome, especially to Akio, Blake, David, Diana, Renato, and Yipu.

For providing invaluable insights and commentary on my dissertation, I would like to thank my colleague Niels Olof Bouvin.

Thank you to all my colleagues at Aarhus University, its Computer Science and Digital Design and Information Studies departments, and the Center for Computational Thinking and Design. Thank you, in particular, to Ole Sejer Iversen and Rachel Smith for leading the charge in the CEED project together with Marianne. Thank you to all my colleagues in the CEED project, Mariana Tamashiro, Marie-Monique Schaper, Maarten van Mechelen, Mille Lunding, and Jens-Emil Grønbæk. Thank you also to Peter Gall Krogh for immensely fruitful discussions about design and research through it. Thank you to Liam Healy for reinvigorating my love for design research and design in general. Thank you, all!

Thank you to all my co-authors, who were not mentioned elsewhere. Thank you Arthur Hjort, John Zimmerman, Eva Eriksson, Ira Assent, Simon Enni, and Kasper Løvborg.

To round off this round of work-related expressions of gratitude, I would like to thank all the schools, teachers, and hundreds of students who have volunteered their time during these weird years to help me out. Thank you.

As necessary as the help and support from the above people is the help and support I received from the people below, with whom I have (perhaps for the better) not worked.

First and foremost, a loving thank you to Lea Møller Christensen, my partner (not least in crime). I have no words to describe my gratitude towards you. Although you have dealt with much more than I have throughout these years, you were always there for me. I would not have been able to do it without you.

Another thank you to my family. Thank you to my father, mother, sister, and brother. You have all been a great support and help throughout my studies.

Finally, thank you to all my amazing friends. Thank you Sarah, Heine, Simon, Vilde, Asmus, Mark, Kasper, Daniel, Sidsel, Luise, Katrine, Andreas, and Søren. These past years, I have not seen half of you half as much as I would have liked, but you have all been there for me when I needed it, and for that, I thank you.

Structure of the Dissertation

This dissertation has been divided into two parts; Part I provides an overview of, and overarching argument for my research and its contributions, and Part II contains the five main papers included in this dissertation. This first part is structured as follows:

Chapter 1 introduces and motivates the work behind this dissertation. I situate the dissertation in current discussions on the use and implications of AI and ML in everyday life, and in the work to include Computer Science and Informatics at all levels of education. I present the research questions that this dissertation addresses, and provide an overview of the dissertation's structure.

Chapter 2 presents background and related work on Computational Thinking and how it is moving in a more critical direction. Then it presents work on empowerment in and beyond education, and on previous research into what students should know about ML, and how to design learning tools and activities for teaching it. Finally, the chapter presents previous work on how the HCI field has engaged with the design of AI-systems.

Chapter 3 presents my methodological approach to the work presented in this dissertation. Further, I detail my collaborations with other researchers and educators. I reflect on how my constructive design research and participatory design have affected my work.

Chapter 4 unfolds the design process behind the work presented in this dissertation, emphasizing how different design experiments informed each other. From this, I articulate six design principles informed by my work and discuss how these interact with each other in concrete designs through different tensions and synergies. I discuss how the principles and their tensions and synergies might impact designers and educators working with ML education.

Chapter 5 presents Computational Empowerment for ML Education (CEML), a framework for teaching ML which aims to empower students in discussing and reflecting on the role of ML in their lives and in society with a strong foundation in ML fundamentals. The chapter presents a set of ML concepts, practices, and perspectives that provide such a foundation, and discusses how these might be applied by designers and educators in the classroom.

Chapter 6 moves beyond the classroom and discusses the role of Computational Empowerment in the development of everyday AI applications. I argue for the need for a new approach to AI-systems design. The chapter presents Remarkable AI as a design approach that emphasizes end-users' empowerment in interacting with AI-systems. I present remarkableness as a sensitizing concept for AI-systems design and discuss how to determine the right level of remarkableness on the continuum between remarkable and unremarkable AI.

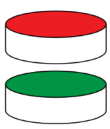
Chapter 7 discusses how my work has addressed the research questions posed in Chapter 1 and concludes the dissertation. Further, the chapter discusses the limitations of my work and points to future work needed to further develop the ideas presented in the dissertation.

Main Contributions

Bellow follows an overview of the publications and manuscripts that are included in part II of the thesis. For all the included manuscripts, I have made significant contributions to the work behind and the writing. Papers A through C are drawn on in chapter 4. Paper D is drawn on in chapter 5. Finally, Paper E is included as chapter 6 with several edits and expansions as detailed in that chapter. Each paper that presents a design experiment has been assigned a pictogram in the margin notes to more easily identify it. When these experiments are referenced, this pictogram will be shown in the margin notes to support the reader. Note that for all included papers, reference numbers have been adjusted to the dissertation.

Paper A (VotestratesML) Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, Maarten Van Mechelen, Arthur Hjorth, Niels Olof Bouvin, and Marianne Graves Petersen. 2022. High school students exploring machine learning and its societal implications: Opportunities and challenges. Int. J. Child-Comp. Interact. 34, (Dec 2022). https://doi.org/10.1016/j.ijcci.2022.100539.

Paper B (ML Ethics Workshop) Karl-Emil Kjær Bilstrup, Magnus H. Kaspersen, and Marianne Graves Petersen. 2020. Staging Reflections on Ethical Dilemmas in Machine Learning: A Card-Based Design Workshop for High School Students. In Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20). Association for Computing Machinery, New York, NY, USA, 1211–1222. https://doi.org/10.1145/3357236.3395558.



Paper C (Machine Learning Machine) Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen. 2021. The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning. In Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21). Association for Computing Machinery, New York, NY, USA, Article 19, 1–12. https://doi.org/10.1145/3430524.3440638.

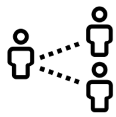
Paper D (CE Progression Model) Magnus Høholt Kaspersen, Daniel Graungaard, Niels Olof Bouvin, Marianne Graves Petersen, and Eva Eriksson. 2021. Towards a model of progression in computational empowerment in education. Int. J. Child-Comp. Interact. 29, C (Sep 2021). https://doi.org/10.1016/j.ijcci.2021.100302.

Paper E (Remarkable AI) Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, John Zimmerman, Ole Sejer Iversen, Marianne Graves Petersen. Remarkable AI: Designing for Computational Empowerment in Use of AI Systems. Unpublished.

Additional Contributions

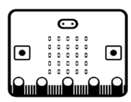
The following publications are not included in Part II but have all played a part in the contributions of this dissertation. I have made significant contributions to the work behind these publications, and either major or proportional contributions to their writing. Papers F through H are briefly introduced in chapter 4, as they played a role in defining the design principles introduced in this chapter. Paper A is an extended version of paper J, and as such was included instead of it.

Paper F (Data Literacy) Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Mille Skovhus Lunding, Marie-Monique Schaper, Maarten Van Mechelen, Mariana Aki Tamashiro, Rachel Charlotte Smith, Ole Sejer Iversen, and Marianne Graves Petersen. 2022. Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data. In Interaction Design and Children (IDC '22). Association for Computing Machinery, New York, NY, USA, 225–236. https://doi.org/10.1145/3501712.3530783.



Paper G (Machine Learning Machine 2.0) Karl-Emil Kjær Bilstrup, Magnus HøHolt Kaspersen, Matilde Fjeldsø Larsen, Niels Olof Bouvin, and Marianne Graves Petersen. 2022. The Best of Both Worlds: Designing a Tiered Hybrid Interface for Teaching Machine Learning in K-9 Education. In Nordic Human-Computer Interaction Conference (NordiCHI '22). Association for Computing Machinery, New York, NY, USA, Article 51, 1–12. https://doi.org/10.1145/3546155.3546156.

Paper H (ml-machine.org) Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Marie-Louise Stisen Kjerstein Sørensen, and Marianne Graves Petersen. 2022. Opportunities and Challenges of Teaching Machine Learning as a Design Material with the micro:bit. In Adjunct Proceedings of the 2022 Nordic Human-Computer Interaction Conference (NordiCHI '22). Association for Computing Machinery, New York, NY, USA, Article 45, 1–6. https://doi.org/10.1145/3547522.3547689.



Paper I (Demo to Design) Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Ira Assent, Simon Enni, and Marianne Graves Petersen. 2022. From

Demo to Design in Teaching Machine Learning. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). Association for Computing Machinery, New York, NY, USA, 2168–2178. https://doi.org/10.1145/3531146.3534634.

Paper J (VotestratesML*) Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, Maarten Van Mechelen, Arthur Hjorth, Niels Olof Bouvin, and Marianne Graves Petersen. 2021. VotestratesML: A High School Learning Tool for Exploring Machine Learning and its Societal Implications. In FabLearn Europe / MakeEd 2021 - An International Conference on Computing, Design and Making in Education (FabLearn Europe / MakeEd 2021). Association for Computing Machinery, New York, NY, USA, Article 3, 1–10. https://doi.org/10.1145/3466725.3466728

Contents

List of Figures

List of Tables

PART I

Overview

Chapter 1

Introduction

> *"We want to feel that our world is intelligible, so we can be responsible for it."* — Crawford, 2009 [45]

During the last three years as a Ph.D. fellow, I have written and rewritten this introduction several times. Artificial Intelligence (AI) is present and central in an ever-expanding part of our everyday life. Being a researcher trying to figure out how to teach this to children feels like playing a constant game of catch, in which the technology is always just out of reach. In my time as a Ph.D. fellow, it was mainly the introduction to the world of the GPT-3 and ChatGPT natural language processing and Dall-E image generation models that dominated the public conversation: what is the future of communication and art if we can no longer tell if a painting was made by a person or an AI model, or if we cannot be sure who or what wrote that tweet that made us angry? Shortly after I began my Ph.D., the world changed because of the Covid-19 pandemic. Suddenly, we were utterly dependent on computational systems, social media, and video conference software to stay in touch with our loved ones and to do our work. Now, AI and systems using it are so ingrained in our everyday lives that part of being a citizen involves engaging with these systems, and as such, should, arguably, also include an understanding of how these systems are made, how they work, and what their implications are.

Since the most recent AI Spring, the advantages made in computer processing and the cheapening of digital storage have made AI a central component in developing many interactive systems. Across the world, AI drives innovations in fields such as medicine [43, 200], criminal justice [176], terrorism prevention [241], etc. Further, AI is deployed across social networking sites for curating personalized content and advertisements, and researchers at Facebook have shown how they can tweak these systems for large-scale manipulation of their users' moods [129]. Finally, AI is physically present in our homes through digital assistants such as Google Assistant, Amazon's Alexa, etc. [98].

These examples show that while AI promises excellent progress, it also has the potential to do great harm. Much criticism has been raised about how AI is used, pointing out how these systems' opaqueness makes it nearly impossible to scrutinize and survey them [36]. Additionally, since these systems rely on historical data for making inferences, they are subject to unfair biases, e.g., sexism, classism, racism, homophobia, etc. [11, 210] that might permeate such data. If not carefully mitigated, systems risk becoming unfair and consolidating human biases and errors in ways that are nearly impossible to account for.

The advantages of AI, the promises made, and the advances already shown are too significant to wholly dismiss the technology on these grounds. But, as HCI researchers and designers, we must do what we can to ensure that the systems we design or the designs our research enable benefit their users. Already, researchers are engaged with creating explainable AI (xAI) that allow for scrutiny [1], and the Association for Computing Machinery has established an academic conference on Fairness, Accountability, and Transparency (FAccT) in AI systems.

In addition to these efforts, there is a growing focus on AI education in schools worldwide, and researchers have been exploring what to teach about AI (e.g., [137, 216]). These efforts are often based on the notion of Computational Thinking (CT), as popularized by Wing [234]. Wing's CT focuses on how Computer Science methods and concepts, such as abstraction and decomposition, can be a powerful way of understanding and solving problems.

However, CT "lacks a wider contextual approach to technological, cultural and societal challenges and change" [107, p. 4] posed by AI. What is needed, I argue, is an approach to AI education that does not just aim to develop children's instrumental skills to prepare them for a career in one of the STEM fields (science, technology, engineering, and math) but one which fosters a critical understanding of AI, and empowers children to question and discuss it.

With this dissertation, I aim to contribute to this conversation, specifically in relation to the subset of AI technologies called machine learning (ML). As I unfold below in section 1.4, ML is a subset of AI driven by the analysis of historical data to make inferences. ML specifically is the technology enabling the most discussed AI technologies such as Dall-E, ChatGPT, conversational systems such as Alexa and Siri, and the recommender systems used in search engines such as Google, and social networking sites such as Facebook, Instagram, and TikTok.

In my Ph.D. studies, I have followed several strands of research — some more experimental and some more theoretical — and this dissertation attempts to combine these into a coherent whole. Generally, I have spent most of my time on different design experiments for secondary school students and classrooms[1]. In the final stages of my work, however, I have come to realize that the scope of the quest I have embarked on reaches beyond what is achievable in a formal school setting and necessitates changes in how AI systems in the wild are designed. (This notion is explored and discussed further in chapter 6).

I hope you enjoy the read and that you step away from this dissertation inspired and perhaps a little provoked.

1.1 The Need for Computational Empowerment in ML Education

In recent years, several researchers working with CT have argued for focusing on empowerment in computing education (e.g., [50, 107, 110, 123, 124, 213]). Kinnula

[1] Secondary School carries different meanings in other parts of the world, but I use it throughout this dissertation as a shorthand for students aged 11-18, grades 6-12, or in the Danish school system; 'udskolingen', and 'gymnasieskolen'.

et al. [124] present a framework for different meanings of such empowerment, ranging from a management view in which empowerment takes a motivational meaning, to a critical perspective, in which empowerment denotes the ability to recognize issues with the status quo and encouragement for making changes and taking action.

As I have argued above, the challenges of AI and ML call for empowerment in the critical sense of children, a need echoed in the literature [50, 100, 107]. Iivari [100] argues that the role of children should be shifted from "mere users of digital technologies created by adults to makers and shapers of such technologies and, along these lines, to transformers of culture". Similarly, Iversen, Smith, and Dindler [107] expands CT's focus to teaching children to 'code' and 'decode' technology around them and frames computing as cultural production. They position 'decoding' as a way for children to "understand the impacts of digital technology on their personal and societal contexts" and 'coding' as a means for children to participate in "co-creating the future that emerges through the design of digital technologies".

Iversen, Smith, and Dindler [50, 107] name this approach Computational Empowerment (CE), to highlight their aims of democratizing technology and empowering children in their interaction with it. Informed by the Scandinavian Participatory Design movement of the 1970s and -80s, they argue that children should be empowered to have a say in how technology is developed and included in their lives. This, they claim, can be achieved through hands-on design work, in which children learn to create meaningful objects through technological means and by analyzing and 'hacking' existing technologies to examine their personal, cultural, and societal impacts.

In this dissertation, I draw on CE as my primary educational approach, as its aims align with the challenges AI and ML pose (as outlined above). ML education from a CE point of view entails a focus on both *coding*-activities, such as understanding and creating with ML, as well as *decoding*-activities, such as analyses and discussions of the implications of ML.

1.2 Research Problems & Questions

As discussed above, the widespread use of ML poses promises and challenges to how we arrange our future societies and personal lives. There is a need for a public conversation about how ML is used in the systems we engage with daily.

CE is fruitfully positioned to address the need for fostering an understanding of ML that supports this conversation. It proposes that children should not just learn instrumental skills and competencies regarding digital technologies but that it is equally (if not more) important that they develop the ability to reflect on, discuss and question them critically.

However, CE is presented by Iversen, Smith, and Dindler [107] as a broad approach to teaching digital technology, and one that, in the past, has mainly engaged with technology through digital fabrication [107]. Thus, CE has focused less on developing students' understanding of the fundamentals of technology. ML differs from other computational technologies in some key areas: It is not explicitly programmed to carry out (algorithmic) instructions but makes inferences based on historical data. The algorithms that allow this are (often) complex beyond (most) human comprehension, leading to opaque results that are difficult to call into question. Further, ML, in practice, usually exists behind the curtains of the systems that deploy it. You might interact with an interactive system and never realize that it either deploys ML or collects data about you to use in ML systems elsewhere. As mentioned above, I see a need for an articulation and elaboration of CE, specifically with regard to its engagement with the technical fundamentals of ML.

Next is the challenge of designing concrete learning tools and activities with a CE framing of ML education. Van Mechelen et al. [222] show that most research in child-computer interaction aims to teach technology to provide children valuable skills and competencies or the means to improve their lives through technology (i.e., with a focus on non-critical articulations of empowerment [124]). They identify a need for more research that develops children's critical engagement with technology. In this dissertation, I aim to address this gap, specifically regarding ML. What does computationally empowering ML education look like in practice? How do we convey to researchers, designers, and educators in ML, what CE implies for teaching ML? What practical advice can we give these people based on this framing?

The final problem this dissertation addresses is how we sustain CE beyond the classroom. Empowerment cannot be bestowed unto people but must be claimed by them [67, 153]. Thus, it is not enough to develop students' capacity for critically engaging with ML: "Freedom and morality are meaningless unless they are to be enjoyed among people who are free and moral. Hence, the ethics of the individuals in the AI world is influenced by the architects of the ecosystem within which they enact..." [212, p. 225]. In other words, we need to address the systems that children engage with outside the classroom. If these systems are (as they are today) designed to serve the tech giants first and their users second, what opportunity do children have to exercise their so-called empowerment? Especially when so many rely on these systems to maintain and develop their social lives [132, 240]. How can we ensure children's agency when using everyday ML-based systems outside the classroom? Focusing on teaching AI/ML and developing critical thinking in students is not enough. What we learn from these significant efforts should be tied back into HCI research and ML systems design. Only in this way can we sustain our efforts beyond the classroom and into everyday life.

These problems are further reaching than what I will be able to *solve* in my dissertation. However, with this dissertation, I aim to inform research as well as policy and educational efforts to come. The research problems outlined above lead me to address the following research questions in this dissertation:

RQ 1: How can CE as an approach to ML education address the challenges posed by the widespread use of ML?

RQ 2: How do we design learning tools & activities that scaffold children's computational empowerment with regard to ML?

> **RQ 2a:** What characterizes the design space for such tools & activities?

RQ 3: How can a CE approach to ML education inform AI/ML systems design?

Below, I briefly present my research methodology (see chapter 3 for an in-depth unfolding) and briefly introduce AI and ML for the uninitiated. Following this, I unfold how each research question will be addressed in the different chapters of this dissertation, how they relate to my work, and the papers on which this dissertation is based.

1.3 Research Methodology

Explorative work such as this calls for an explorative methodology. Here, I have first and foremost based my work on Constructive Design Research [7, 130]. Constructive Design Research (CDR) is a particular branch of research-through-design, where the construction of design artifacts and experiments with these artifacts drives the knowledge generation in the research [7].

In the work presented here, I have investigated the context of CE and ML education through several experiments. I have designed prototype learning tools and activities for teaching ML with a CE focus. Each of these has brought new insight and findings with them and has informed the next experiment, as well as my increasing understanding of the problems I outlined above. These insights led me to look at the problems in new ways and drift [130, 131] in new directions with the following experiment. Krogh and colleagues [130, 131] describe different ways of doing so, with the most pertinent for this project being a "comparative" mode of drifting. For this approach, the different design experiments aim to "reveal as-yet undocumented additional qualities of a concept and confirm some previously found qualities". Central to my experimental work has been exploring the design space for learning tools and activities, and throughout the process, I have incorporated several design elements. In chapter 4, I outline this process with a particular focus on the drifting that occurred throughout it.

In addition to CDR, this work is informed by Participatory Design, specifically its Scandinavian variant [22, 28]. Much of the design work has been carried out in collaboration with other designers, educators, and students. Still, most importantly, my work shares its political goals with the Scandinavian branch of PD. My work is motivated by the acknowledgment that computational technologies and the digitalization of society must happen democratically to ensure that the interests of otherwise disempowered stakeholders be heard. This combination of CDR and PD has led to several tensions in my project, both methodologically and practically, which I explore and discuss further in chapter 3.

1.4 A Primer on Artificial Intelligence and Machine Learning

For the interested reader, with little to no understanding of artificial intelligence or machine learning, I offer here a small introduction to the technology[2]

AI as an idea has been around for a very long time and can be traced back to the philosophers and mathematicians of antique Greece. In the 1950s, however, the idea of AI as we have come to understand it today began to form. In its

[2] For a more comprehensive introduction see, e.g., *Artificial intelligence* by Russell, Norvig, and Davis [189].

broadest term, AI is defined by the Merriam-Webster dictionary as "the capability of a machine to imitate intelligent human behavior". As AI is an umbrella term, this definition can be applied to (relatively) simple systems such as reversing sensors in cars that indicate or even bring the vehicle to a halt when too close to an obstacle. It might also be used to describe delicately programmed non-playable characters in video games designed to give the player the appearance of intelligent behavior.

In this dissertation, I am not concerned with these types of systems but rather the subset of AI technologies called machine learning (ML). As indicated by the name, ML is said to "learn" from historical data to produce an outcome informed by this data. A simple example of ML is the K-nearest neighbors algorithm. Say we want to predict a type of sports activity based on how far players move and how fast they accelerate. This algorithm takes as its input a training data set describing this relation, with each data entry having a *label* (i.e., the type of sport) and two *features* (i.e., length traversed, and total acceleration). When we wish to predict the sports type of a new, unlabeled entry, we show this to our system. It then calculates the distances between the new entry to those already labeled. The new data entry is then awarded a label based on the category of the majority of the K nearest neighbors (hence the name).

This is an example of a supervised ML model (i.e., where the training data is manually labeled). There are many types of supervised ML algorithms, most of which are more complex than the example above, such as different types of neural networks. Notably, there are also other types of ML, such as unsupervised learning, in which we do not know the label of the training data, and algorithms attempt to divide it into meaningful chunks. Additionally, there is reinforcement learning, in which algorithms explore and optimize their path to a predetermined goal[3], as well as generative ML models, where historical data is used to suggest new entries[4].

The tools and activities I present in this dissertation are all examples of supervised ML, but as argued by Enni and Assent [56], all types of ML are constructed from the same technological foundation of learning from data, and optimization towards a specific goal, whether it is categorizing a new data entry, generating a new image, or finding the optimal strategy for playing StarCraft II.

1.5 Structure of Part 1

The remainder of part 1 is structured as follows: First, in chapter 2, I position my work in relation to previous research: I present related work on empowerment in and beyond education, i.e., what it means to be empowered, the role empowerment plays in technology education, and child-computer interaction research. Next, I present work related to AI education, specifically with regard to suggestions for AI curricula and what children should know about AI. Then I present the state of the art in designing tools and activities for teaching AI and argue that there need to be more tools that specifically aim to foster children's critical reflection.

[3] An example of this is Google's AlphaStar, which was trained to optimize winning matches in the video game StarCraft II, and was able to beat professional athletes.

[4] Examples here are typing suggestions on smartphones, but also more advanced systems such as ChatGPT, the author of my forewords, and the image generation model Dall-E.

Finally, I present work related to the role HCI has played in the design of AI systems. Here, I focus mainly on xAI and AI design guidelines.

Next, in chapter 3, I present and discuss the methodological foundations of my work, how I have applied CDR, and specifically worked within a programmatic epistemological framing [130]. I further present how PD has influenced my work, how collaboration was central throughout my process, and discuss the tensions between CDR and PD that have affected my design process.

Chapters 4 to 6 present the main contributions of my project.

Chapter 4 addresses RQ2. Here, I present and analyze my design process (based on papers A through C and F through H), emphasizing the knowledge creation that occurred in the process and the drifting between each experiment. Based on this analysis, I present six design principles for designing computationally empowering tools and activities for AI education. Finally, I discuss these principles with regard to how they interact with offset in my experiments.

Next, in chapter 5, I address RQ1 by presenting the CEML framework: an expansion of the original articulation of CE with a technological, material foundation. I present what ML concepts, practices, and perspectives are key for CE-based ML education. Further, I draw on our model for progression towards CE, based on paper *Paper D*, to discuss how the CEML framework can inform the work of designers and educators working with CE and ML.

Chapter 6 addresses RQ3. It lays forth a reframing of my original research program (as presented in chapter 3) as Remarkable AI, based on *Paper E*. This framing marks a return to HCI, and I argue for the need to sensitize designers to remarkableness in AI systems design with the aim of computationally empowering people with regard to AI.

Finally, in chapter 7, I conclude part I, summarize my contributions to computing education and HCI, and point out limitations and directions for future research based on these contributions.

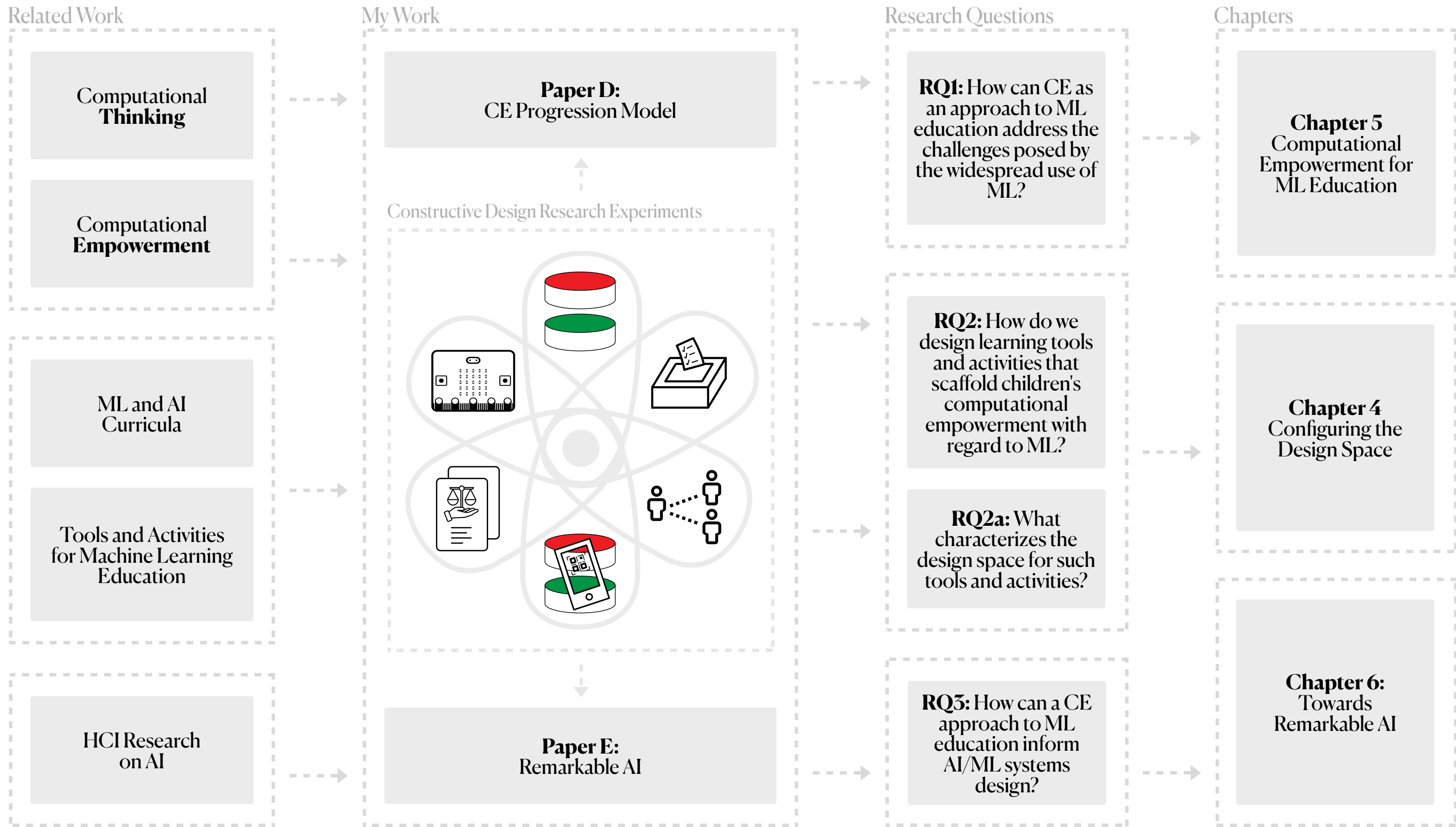


Figure 1.1: An overview of the relation between the related work I draw on, my work and design experiments, my research questions, and in which chapters they are addressed.

Chapter 2

Background & Related Work

This dissertation is built on a large body of previous work, mainly from HCI and computing education. In this chapter, I present the state of the art about computational thinking in education, how critical aspects of computing and empowerment have become more prevalent in research on computing education, and what researchers have argued every child should know about AI. Next, I present an overview of prior efforts in designing AI learning tools and activities. Finally, I present related work on how the HCI community has addressed and engaged with AI, specifically through Explainable AI, everyday algorithm auditing, and design guidelines for AI. I do so to provide an overview of the fields that intersect with my research and to further position my work within these fields.

2.1 Computational Thinking

The term Computational Thinking (CT), as it is most popularly understood, was developed by Wing [234] in 2006. For Wing, the term means "thinking like a computer scientist", that is, problem-solving by reforming "a seemingly difficult problem into one we know how to solve" [234, p. 33]. She argued that computers and computer science, as a method, are universally needed for tackling the issues we had (at the time) at hand. At the heart of computational thinking, Wing argues, is abstraction [233]. Further then, computational thinking to Wing is asking "how do we express a problem such that it can be computed?", and "[i]mplicit in answering this question is our identifying appropriate abstractions and choosing the appropriate kind of computer for the task" [233, p. 3719].

In 2008, Wing argued that "if we wanted to ensure a common and solid basis of understanding and applying computational thinking for all, then this learning should best be done in the early years of childhood" [233, p. 3720]. She observes that universities (at least in the U.S.) have already begun incorporating CT into their curricula across different fields of study and argues that the next step is to look to K-12 education.

However, the phrase "Computational Thinking" was first used by Papert [171], although on in passing. Papert's idea was that computers might serve as a tool for students to experiment with abstract, mathematical ideas by constructing concrete and meaningful objects. This idea is further instrumentalized in his learning theory of Constructionism, which argues that learning most efficiently occurs when constructing such meaningful objects [169]. There is merit in both approaches [136]; Wing's CT emphasizes that an understanding of central CS ideas is necessary to act in an increasingly digital world. Further, Papert's CT asserts that experimenting and constructing with digital technologies (such as programming) can be powerful tools for learning.

As CT gained traction in K-12 education, so did the development of tools for teaching it, often inspired by both Wing and Papert. Although I will not present a review of CT tools generally (instead see, e.g., [76]), I will briefly talk about Scratch, as it has been and continues to be a hugely popular tool for developing CT in children, with more than 35 million monthly users, and 100 million registered users[1]. Resnick et al. presents Scratch as a way "to make it easy for everyone, of all ages, backgrounds, and interests, to program their own interactive stories, games, animations, and simulations, and share their creations with one another" [183, p. 60]. Functionally, Scratch is a so-called block-programming language in which users create algorithms and programs by snapping different blocks together. The different types of blocks were designed visually and functionally to allow snapping only when syntactically possible. The aim for Resnick et al. was to provide children with the means of developing their capacity for computational thinking (in the Wing sense) in a playful, constructive, and experimental way, following Papert's Constructionism.

To further operationalize CT (in Wing's terms), Brennan and Resnick [33] present three key dimensions of CT; computational thinking concepts, practices, and perspectives. As examples in the context of Scratch, the authors present "sequences, loops, parallelism, events, conditionals, operators, and data" as key CT concepts, "being incremental and iterative, testing and debugging, reusing and remixing, and abstracting and modularizing" as practices, and creative expression, social connection, and questioning as key perspectives. While these examples are primarily targeted toward programming, the dimensions are relevant to CT as a whole.

With her insistence on CT as a fundamentally important skill, Wing accurately predicted the situation that we are in today, which led to my work. Across the world, CT is being implemented in school systems[2]. However, as pointed out by Iversen, Smith, and Dindler[107], the strengths of Wing's CT are in understanding how computing works and how to apply computing to solve problems. It is, however, lacking regarding how to prepare children for lives as heavily influenced by computational technology as we see today. They argue that CT should be expanded with the notion of Computational Empowerment, which they define "as the process in which children, as individuals and groups, develop the skills, insights, and reflexivity needed to understand digital technology and its effect on their lives and society at large, and their capacity to engage critically, curiously and constructively with the construction and deconstruction of technology." [107, p. 1].

[1]See https://scratch.mit.edu/statistics/, accessed 20/12/22

[2]See, e.g., the CS4All program in the U.S (https://sites.google.com/schools.nyc.gov/cs4allnyc/) and in Europe the joint efforts of Informatics Europe and ACM Europe (https://www.informatics-europe.org/education/informatics-in-schools.html, accessed Jan 17th, 2023).

2.2 Empowerment in Computing Education, and Beyond

To take a step back, I will briefly account for the notion of empowerment and how it has been used in education.

Empowerment, as a term, came into use in the late 1960s and 1970s in the civil rights movements of the time, including feminism and the Black Power movement [38][3]. The term, however, has often been misused or conflated with democratization, decentralization, political participation, enabling, and speaking out [153, 237]. There is, however, some agreement on the meaning of the term. Hence, empowerment is "a multidimensional process of change from a condition of disempowerment", that "[c]annot be bestowed by a third party, as individuals are active agents in this process", and "[i]s shaped by the context [...]" [153, p. 321].

With regard to empowerment in education, a seminal work is Freire [67]'s *Pedagogy of the Oppressed*. He argues that educators should strive "for the emergence of consciousness and critical intervention in reality" [67, p. 81]. According to Freire, the (still) predominant student/teacher relationship, in which students are receivers of knowledge that educators deposit (what Freire calls "banking education"), needs to be retired, and replaced "by reconciling the poles of the contradiction so that both are simultaneously teachers and students.". He further cements that empowerment comes from within and cannot be deposited in students: "Freedom is acquired by conquest, not by gift. It must be pursued constantly and responsibly." [67, p. 47].

I have, throughout my work, experienced this on several occasions. Although many students expressed gratitude that we had taught them about ML and how it is implemented in society, we also faced a certain hesitation about making changes to how they use and act around ML-systems, in particular social networking sites and applications. Here Karlberg [117]'s exploration of power relations might help to shed light on how this can be addressed. He argues that the dominant view of power in the West is adversarial or "power against" something. He offers an alternative in "mutualistic power", or "power with" someone. Karlberg argues that empowerment might be nurtured or assisted through education or mentorship and that by collaborating, people who were previously disempowered might find power in such collaborations. With this view, while we might not empower individual students to make significant changes on their own with regard to ML, we might assist them through education and aim to develop their capacity to take collective action by, e.g., participating and adding to the discourse around ML.

Next, I will examine different empowerment goals related to children's computing education. In child-computer interaction (CCI) broadly, and specifically in computing education, empowerment is often mentioned as a goal with 188 papers from the CCI's most prominent venues (the Interaction Design for Children Conference (IDC), and the International Journal for Child-Computer Interaction (IJCCI)) using the term from 2003 to 2020 [222]. Kafai, Proctor, and Lui [111] offer a general framework for different computing education framings divided into cognitive, situated, and critical framings. Here, the cognitive frame centers on building skills and competencies in a specific domain, and the situated frame

[3] For a comprehensive history of Empowerment, see "Empowerment: The History of a Key Concept in Contemporary Development Discourse", by Calvès [38].

focuses on supporting practices and participation. Finally, the critical frame aims to foster ideological awareness and teach social and political action strategies. Kinnula et al. [124] present a more detailed framework of 5 different views on empowerment related to how the term is understood in CCI: First, the *management/mainstream* view in which empowerment is characterized as providing employees with (some) power as a motivational strategy. Next, the *functional* view, in which empowerment is understood as a method for "improving people's life-conditions to serve organizational/management goals". Third is the *educational/competence* view, which aims to empower people by teaching them useful skills and competencies. The fourth view is *democratic*, in which empowerment is taken to mean improving people's "ability to participate in decisions affecting their lives". Finally, the *critical* view is characterized as the oppressed "combating" the oppressors. This final view coincides with Freire's articulations of empowerment. Importantly for Freire, however, is the idea that a resolution for the oppressed lies in the absolution of the oppressed/oppressor relationship, not in the oppressed gaining power over the oppressors, themselves becoming oppressors in the process.

In 2021 Van Mechelen et al. [222] conducted a systematic review on the use of Empowerment in CCI literature drawing on the distinctions made by Kinnula et al. [124]. They found that nearly 70% of the papers that mention empowerment deploy it in either its functional or educational term, while 23% use empowerment in its democratic sense. Furthermore, the mainstream view is represented in 8% of papers, while the critical view only makes up 3% of papers. These distinctions are, however, not wholly separate; children might need to obtain new skills (educational empowerment) and certainly need the ability to partake in decisions affecting their lives (democratic empowerment) in the fight for liberation from their oppressors (critical empowerment). Given the current power relations between Tech Giants and us, critical empowerment should be striven for to a greater extent. This is further corroborated by Iivari et al. [101], who argue that researchers in CCI should pursue the following: "(1) more explicit and thorough engagement with critical theories; (2) more attention paid to different forms of empowerment and their implications; (3) more critical analyses of the status quo; and (4) CCI studies moving from mild to wild along the continuum of criticality."

Computational Empowerment (CE) [50, 107] is a Scandinavian approach to computing education that explicitly aims for democratic and critical empowerment of children with regard to digital technology. Further, CE emphasizes the development of children's reflexivity, and ability and willingness to take an ethical stance on digital technology (thus adhering to what Iivari et al. [101] labels a wild approach to criticality).

CE had its starting point in the Danish FabLab@School [107, 204] project, which "[t]hrough long-term engagement with diverse stakeholders including students, teachers, lab leaders, and local and national policymakers, [. . .] explored the core challenges and potentials of integrating digital technologies into real-life educational contexts" [107]. Through a participatory design-based process, the involved teachers and researchers examined how to use digital fabrication and design with the goals of supporting students' in "1) engaging creatively in technology development, 2) understanding the role of digital technology in society, and 3) reflectively and critically understanding the role of technology in one's own life" [107]. As an approach, CE builds on Scandinavian PD [28, 31], and in particular, the early Utopia project [31]. Here, researchers collaborated with trade unions and blue-collar workers to gain influence on how technology was introduced in their workplace through processes that Karlberg [117] might have characterized as mutualistic empowerment.

In a 2018 paper, Bødker and Kyng [30] presents a harsh but fair critique of current PD efforts. One of their primary criticisms is that "[w]hat happens after design is left to existing structures outside the projects": Most current PD efforts are concerned more with ensuring democratic and fair participation in the research process than with what happens once the researchers leave. In other words, the empowerment goals of PD projects seem to have shifted away from being critical [124]. CE is positioned as a return to previous goals of critical empowerment. In particular, the Danish FabLab@School project, which CE sprung from, engaged with local and national policymakers to make sustainable effects on policy [27].

However, while the Danish FabLab@School project focused on supporting digital fabrication as a means for children to co-create the future, the project did not engage with advanced, emerging digital technologies, such as ML. My project then is an insistence on sustaining the critical empowerment goals of CE and the FabLab@School project, with a particular focus "on questioning [...] even the most advanced technologies at hand" [30], here ML.

2.3 What Everyone Should Know About ML

To help develop a CE approach to ML and AI education, I will present literature related to AI in education, specifically regarding what everyone should know about it.

In 2019, Touretzky et al. [216] presented their suggestions for what all K-12 students should understand about AI as "Five Big Ideas". First, they argue that students should understand that "computers perceive the world using sensors". According to the authors, one of the most significant advantages of AI is that the computing power in many devices is now sufficiently powerful to process rich sensor data such as video and audio, allowing advanced analysis by AI systems. Second, they argue for teaching that "Agents [or instances of AI] maintain models/representations of the world and use them for reasoning". It is important here that AI systems never have a holistic worldview but are instead limited to specific representations of the world determined by how they are configured. Third, to emphasize that "Computers can learn from data", i.e., that ML algorithms allow AI systems to self-configure these representations based on training data. Then, they propose teaching how "Making agents interact comfortably with humans is a substantial challenge for AI developers", referring to conversational systems. They point out that these systems are still based on limited models of the world and, thus, do not possess human-level reasoning skills. Finally, they argue that "AI applications can impact society in both positive and negative ways" and present two fundamental questions related to this; "what applications should AI be used for [...], and what ethical criteria should AI systems be required to meet?". These are high-level suggestions, and the authors offer more specific proposals for different grade levels. Their recent work has focused specifically on how to implement teaching about ML in American schools [217].

Similarly, Ng et al. [157] present an exploratory review of AI literacy, and based on this, four aspects of AI literacy: Knowing and understanding the basics of AI, using and applying AI to different contexts with an understanding of what sorts of problems it might be used to solve, evaluating AI systems and creating new AI systems, and AI ethics related to "human-centered consideration" such as fairness, accountability, transparency, etc.

Importantly for a CE-based approach to AI education, neither of these works emphasize empowerment with regard to AI, in the critical sense of aiming to foster the competencies and attitudes to take action regarding the effect of AI students' lives, communities, and society at large.

Long and Magerko [137] present a definition of AI literacy and a comprehensive (although non-exhaustive) set of competencies and design considerations for such literacy. They define AI literacy as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace". Many of the competencies they present are covered by the broad ideas above but are presented by Long and Magerko [137] in more detail and, notably, often with a critical focus. I will not go over all 16 competencies and 15 design considerations but will highlight a few that exemplify the authors' focus on developing critical skills. For example, with their second competency, Long and Magerko argue that students should be able to understand and critically discuss the differences between human intelligence and "machine intelligence". In competency six, the authors argue that learners should become able to "[i]magine possible future applications of AI and consider the effects of such applications on the world" [137, p. 4], with the aim of evaluating current technologies, the values embedded in them, and their possible long term effects. For designing tools for teaching AI, they suggest (in design consideration three) working with contextualized data sets, emphasizing the importance of being critical about data and understanding how it is collected and processed and its limits. They further, with design consideration eight, encourage to develop students as critical consumers, able to question and discuss AI technologies in terms of their supposed intelligence and trustworthiness.

What is, arguably, missing from Long and Magerko's conceptualization of AI literacy is a focus on developing children's attitudes. In the Danish Fablab@School project, a central sentiment was the role of the child [106]. Iversen, Smith, and Dindler [106] here argue that by fostering the role of the child as a protagonist, "children can be empowered to shape technological development and critically reflect on the role of technology in their practices" [106, p. 27]. They argue that the desired outcome when understanding the child, or student, in this way is not only the development of competencies but also developing their "stance toward technology in their life." [106, p. 30]. For a CE approach to AI education, this idea is central. Empowerment, in this sense, is not only about developing competencies and motivating students to learn but also about fostering students' self-identification as potential agents for change regarding AI.

2.4 Tools and Activities to Teach ML

Since beginning my Ph.D. project, research into teaching AI and ML has increased dramatically. Attempting an exhaustive review of tools and activities is beyond the scope of this section, but I will outline the landscape for these below. I am focusing specifically on tools and activities as it is through experiments with my own of these that I have conducted my research. Additionally, my dissertation is (for the most part) about the design of tools and activities. First, I outline tools and activities' different ways of interacting with AI/ML. Then I present what aspects of AI/ML tools and activities aim to teach, according to the framework by Brennan and Resnick [33], divided into concepts, practices, and perspectives. Finally, while tools and activities have advanced since the beginning of my project, there is still a lack of tools aimed at fostering computational empowerment, specifically by focusing on personal and societal perspectives of ML/AI.

2.4.1 Different Ways of Interacting with AI/ML

Tools and activities for teaching AI/ML can generally be divided into three categories (although with some possibility for overlap): Unplugged, embodied, and graphical. First is the *unplugged* activity. Lindner, Seegerer, and Romeike [135] present a set of five activities for teaching different AI principles. One activity has children classify illustrations of biting and non-biting monkeys. Here, the illustrations are hung on a whiteboard, and students collaborate to develop different criteria for classifying them, e.g., are they showing teeth or not? Ossovski and Brinkmeier [167] similarly present an unplugged activity to teach linear classification. Here, photos of screws of two different lengths are used to create a simple unplugged ML system. Students measure the boundary box[4] of a screw, and maps its dimensions onto a 2-dimensional coordinate system. This way, a data set is created, and a linear classifier is then estimated by drawing a line separating the two groups of screws. Skinner, Brown, and Walsh [202] conduct a co-design workshop with children of color to explore their conception of fairness in AI. Here, children were tasked to design a fair AI librarian. The authors find that they associated fairness with kindness and having their best interests in mind. The co-design process did, however, not address how, technologically, fairness can be addressed in ML systems.

Unplugged tools and activities offer a low-floor entry to ML/AI. They can present these technologies to students in non-overwhelming ways that highlight the basic principles of ML/AI and which are simple enough to be understood and made functional without the need for computers. However, they can quickly become too simplistic to convey the complexities and nuances of ML/AI and might not be well suited for discussing issues that arise during more opaque steps in the ML process.

The next mode is what I will call embodied tools and activities. Opposite to unplugged tools, these are interactive and allow students to interact with ML/AI in embodied ways. The degree to which aspects of ML are embodied varies across different tools. Zimmermann-Niefield et al. [246] present AlpacaML as a way for students to learn about ML in the context of athletic moves. The system is comprised of a small wearable device that can be attached to the body and collects gesture data. This data is sent to a smartphone, allowing students to train models to classify different moves. Students are asked to use the system to improve their moves by, e.g., classifying good and bad kicks in a game of football (soccer). In another paper, Zimmermann-Niefield et al. [245] expand AlpacaML

[4]That is, the smallest rectangle that can be drawn around the screw in a photo, such that it encapsulates it entirely.

to interface with Scratch to allow students to build their own gesture-based controls for interactive media. In a similar project, Hitron et al. [92] introduce the Gest system, which allows children to create ML models to recognize gestures of geometric shapes, draw in the air using a small hand-held device. Other tools include Any-Cubes [193]. Any-Cubes are two wooden cubes, where one is equipped with a camera, and the other allows actuators such as a DC motor to be attached. Children can use the camera cube to photograph objects and train a model to recognize these. The other cube can activate the actuator corresponding to the classification upon recognition. Finally, on the boundary between embodied and graphical tools and activities is Google's Teachable Machine [39]. This popular tool allows students to train ML models with different types of data; using their laptop's web camera, they can train models on image or posture data and by using its microphone on audio data. Teachable Machine, however, also supports a purely graphical mode of interaction, where digital images are uploaded to the website and used as training data. Vartiainen et al. [224] used Teachable Machine in a ML workshop for 6th-grade students (aged 12-13), and found that the embodied approach lowered the barrier for entry for the students, as well as supported them in concretely exploring complex ML topics.

Such tools and activities also offer a low entrance barrier and can convey ML/AI to students in a relatable, embodied way. Furthermore, they are better equipped to address ML systems' opaque nuances because they include functional models and systems. However, they often struggle with representing more complex and dynamic aspects of ML that could more readily be made available to students graphically. Many tools, thus, combine the embodied approach (especially regarding data collection) with graphical modes of interaction.

Finally, we have a set of purely graphical tools and activities for teaching AI/ML. Examples of these are block-programming languages made or expanded to include ML/AI capabilities, e.g., Snap! [115], Scratch [60], NetsBlox [34], and the MIT App Inventor [211]. Other tools include SmileyCluster [226], a tool for learning about unsupervised ML by clustering emoticons. AIThaiGen [5] is a tool for teaching ML to 7th to 8th-grade students, offering ways to create object recognition, facial landmarks detection, hand pose detection, and simple image classification, and components for connecting these models to other systems. Finally, Mariescu-Istodor and Jormanainen [141] presents a tool similar to Teachable Machine that allows students to train ML systems on image data. It differs, however, in using a simpler algorithm for image detection, thus allowing students to observe more steps in the processing of data before use in the algorithm.

Typical for these tools is that they are highly dynamic and offer higher ceilings and wider walls [184] for what is achievable over other modes of interaction. By extending existing block programming languages, the possibilities that are inherent in these are carried over, allowing students to create highly functional systems for use in a wide variety of contexts. Further, these tools often allow an in-depth exploration of ML/AI technology by allowing students to pull back the curtain on black boxes and peep inside.

2.4.2 What Tools and Activities Aim to Teach

As suggested by Brennan and Resnick [33], we can, for learning purposes, divide technologies such as AI/ML into three parts; concepts, practices, and perspectives.

Most tools and activities aim to teach at least some concepts of ML. For example, the *unplugged* activities by Lindner, Seegerer, and Romeike [135] teaches central concepts such as what training data is, what features in data are, and how a ML model might classify data. SmileyCluster by Wan et al. [226] aim to teach similar things, but also what clustering in unsupervised ML is, and how it works. Most tools also include aspects of the ML process. With AlpacaML, Zimmermann-Niefield et al. [246] highlight how models are evaluated and how they can be iteratively improved. Vartiainen et al. [224] with their project-oriented workshop, teach how problems can be articulated such that they can be solved with ML. However, as pointed out by Druga, Otero, and Ko [52] in their review of the landscape of resources for teaching AI, few tools address personal and societal perspectives of AI/ML. Skinner, Brown, and Walsh [202] and their AI librarian address fairness in AI systems from the point of view of children of color. As we point to in *Paper C*, tools and activities that address concepts and practices, rarely address the perspectives and implications of ML and vice versa.

2.5 AI in HCI Research

Finally, I present a brief overview of how the HCI community has worked to support and enhance the design of AI systems in recent years. I do so according to the recognition (as argued above) that developing children's skills and attitudes towards AI and ML is insufficient as long as the systems where they meet AI are designed in ways that suppress their possibilities for questioning, evaluating, and altering them.

As argued by Grudin [77], HCI and AI as research fields as several overlaps, especially as AI/ML is increasingly used in end-user systems. One field where this overlap is apparent is Explainable AI (xAI). xAI is concerned with how to explain the output of AI models to their users [146]. However, as argued by Miller, Howe, and Sonenberg [147], these systems often fail to consider non-expert users, focusing instead on offering explanations to domain experts that utilize AI in their work. Further, Abdul et al. [1] argue that xAI tends to offer explanations of a system's inner workings to experts to gain their individual trust but fail to take into account the social contexts in which these systems are used. Ghajargar et al. [72] present Graspable AI as an approach to AI that, in some ways, seeks to address these issues. They argue that work is needed in other modalities for explaining AI systems and that doing this through physical form might be a fruitful way of explaining the meaning, context, and cultural significance of such systems.

These issues are further addressed by Shen et al. [198] and what they dub Everyday Algorithm Auditing. Through case studies of a series of platforms where their users have collectively called into question decisions made by algorithmic systems, such as Twitter prioritizing white people when automatically cropping uploaded photos to display on their website. They suggest that platforms implement opportunities for their users to form communities for auditing their use of AI and mechanisms that allow developers to provide insight into how their systems work. Further, they suggest incentives for platforms to make changes based on these audits. While this approach seems promising, it lacks the effort to push for platforms to design their AI-enabled functionalities in ways that make them malleable for their users. While everyday algorithm auditing allows users to collectively find symptoms of unfair or misfunctioning algorithms, it arguably does not support users directly in understanding and calling into question their underlying causes. To do so, platforms would have to expose more of how their algorithms work and, importantly, offer users ways of exploring, experimenting with, and grasping this.

Finally, HCI researchers have proposed different design guidelines for AI systems. Several researchers have proposed different design principles and guidelines for designing AI systems (e.g., [2, 95, 160]). Recently, Amershi et al. [2] suggests 18 guidelines for human-AI interaction based on an analysis of 150 previous design recommendations from research and industry. Some of them relate to the usability of AI-enabled applications, such as "[s]how contextually relevant information", and "[s]upport efficient correction". Others are geared towards helping users understand, on a general level, what the systems AI component is doing, and how users themselves are and can affect this, e.g., "[c]onvey the consequences of user actions", and "[m]ake clear why the system did what it did". These are steps in the right direction but, I argue, not sufficient. Offering an explanation of an AI system's behavior, or conveying the consequences of an action, does not necessarily leave the end-user with any more power over the outcome. Instead, as a thought experiment, maybe a user would prefer a slightly less accurate outcome, but instead using an algorithm that they could understand. Finally, they suggest to "[m]itigate social biases", which might be more of an ideal than a practical possibility [186]. Indeed, the authors' example of an autocomplete feature without gender bias is one that "suggests both genders [him, her]", which is biased toward a binary understanding of gender. This is to say that completely removing unwanted biases is near impossible since developers might not be aware of their biases; these might be deeply ingrained in historical data, or be constantly reinforced by the system itself. While it is a good thing for developers to mitigate as many unwanted biases as possible, efforts are also needed, as suggested by Shen et al. [198], that allow users to make developers aware of biases they might have missed. Additionally, as argued above, this requires ways for users to learn how biases can occur in AI systems.

2.6 Summary

In this chapter, I have presented a broad range of literature on computational thinking and empowerment, teaching AI and ML, and AI in HCI research. The chapter provides an overview of the different fields with which my work intersects and points to gaps in the literature that my work seeks to address. I argue that computational thinking as an educational approach needs to engage more with the critical aspects of AI/ML and that a broader focus on empowerment is needed to prepare young students for participating in a world where these technologies increasingly affect us. Further, I point out that the work hitherto carried out with a computational empowerment frame (i.e., the Danish FabLab@School project) lacked an engagement with emerging technologies such as AI and ML. I have presented a range of existing tools and activities aimed at teaching AI and ML and found that while there is a great variety in these, only some tools and activities are designed to teach the ethics of AI. Fewer still combine this with a technical engagement with the technology. Finally, I have given a brief overview of how the HCI research field has engaged with AI through Explainable AI, everyday algorithm auditing, and design guidelines for AI. While this is vital work, work is still needed to aid designers in creating AI-systems that help end-users understand their underlying principles and, thus, what consequences these systems might have.

Chapter 3

Methodology & Reflections

In this chapter, I present, argue for, and reflect on my choice of methodology. Throughout the process, I have primarily drawn on Constructive Design Research (CDR) [128, 130] as my method of inquiry and knowledge generation. In addition, Participatory Design (PD) [28, 29, 30], and especially its Scandinavian tradition, has played a supporting role, and the combination of CDR and PD methods has been a central tension throughout my project.

This chapter first introduces Computational Empowerment for ML Education as a research program informing my design process. Then, I unfold the role design experiments played in my process. Next, I reflect on how knowledge is produced and disseminated as intermediate-level knowledge. Then I present how Participatory Design informed my work. Finally, I discuss the tensions between CDR and PD in my project[1].

3.1 My Research Program

According to Krogh and Koskinen [130], knowledge-generation in CDR, historically, belongs to one of four "epistemic traditions"; experiential, methodic, programmatic, and dialectic. The experiential tradition is rooted in art and design practice. Here, "[r]esearch gives the overall frame and direction to research, but the hypothesis is primarily shaped by practice" [130, p. 54]. In the methodic tradition, designers' work is guided by theories and methods, on which grounds hypotheses about a design artifact can be made and tested. In the dialectic tradition, hypotheses "get their meaning in communities of practice rather than theory alone" [130, p. 57]. They are formed in collaboration with a community the designer serves, and as such, they are evaluated in their ability to serve this community. Finally, in the programmatic tradition, "designs get their meaning from theoretical frameworks that discuss the meanings of these designs and serve as handles for other researchers, but that cannot be understood without the concrete experiences of

[1]Sections 3.2, 3.4, 3.5 are adapted from my progress report, and have been edited and expanded upon.

the design objects" [130, p. 55]. In my project, as I will unfold below, I have followed the programmatic tradition.

The initial program for my project was ready-made at its onset. My project is of many in the CEED project: Computational Empowerment for Emerging Technologies in Education. Here, I focused from the beginning on ML, giving rise to my project's research program, Computational Empowerment for ML Education. As argued by Bang and Eriksen, "the program sets a frame for the experiments, making them into more than "undirected explorations", while it at the same time is open for surprises and new insights arising from the experiments" [6, p. 4.3]. In other words, CE for ML Education provided the initial direction for the project. Taking things apart, ML, first and foremost, provided the project with a technological boundary. While other technologies were, and are, relevant for CE (and part of the CEED project), my project would explore ML.

CE, on the other hand, sets political and methodological boundaries. By drawing on its PD roots, CE positions itself as an explicitly political approach to computing education. CE's declared goal is "creating opportunities for future generations to engage as agents and cocreators of potential and critical alternatives" [107, p. 4]. Further, it concerns the "empathic, aesthetic, ethical, and structural aspects of technology" [107, p. 4]. In other words, like early Scandinavian PD [31], CE is concerned with the democratization of technology. Instead of focusing on the workplace, though, CE concerns itself with education. CE is not just politically but also methodologically grounded in PD. Iversen, Smith, and Dindler [107] argue that the established catalog of "methods, techniques, and practices" in PD is an appropriate starting point for developing children's computational empowerment.

In addition to this pre-established research program, I brought my experiences and expertise into the project, especially my experience with Constructive Design Research. This further helped with the initial framing of the project. CDR, as coined by Koskinen et al. [128], is a type of Research-through-Design, that refers to design research, where the knowledge-generation is specifically centered around the *construction* of design artifacts, and experiments with these[2]. As elaborated by Bang et al. [7], the design experiment is central to producing knowledge in CDR. A project's hypotheses and research questions are constantly evaluated against the experiment and vice versa. As they argue, "knowledge, empirical findings, concepts, and ideas are combined as a form of abstract prototypes to be tested and debated according to their relevance to practice, academia, and practicability or feasibility of the experiment". Throughout a larger project that may contain many experiments (such as my Ph.D. project), the research program and its central hypotheses and research questions may drift between or during these experiments, and reframing the research program might become necessary.

As such, the relevance of design experiments in CDR is not limited to evaluative devices with which a hypothesis can be tested. Instead, design experiments are as much framing devices that play a central role in the initial framing and (potential) ongoing reframing of the research program [6]. Below, I discuss the role experiments played in my research project.

[2]The terminology, however, is still somewhat fuzzy, as there is little agreement on this distinction, and many design researchers use construction but do not call themselves constructive design researchers.

3.2 The Role of Experiments: Drifting and Reframing

My commitment to CDR is motivated by how it supports both rapid and deep exploration of new subject areas and its focus on producing knowledge through the construction of artifacts. Apart from my own, I have only found a few examples of design work addressing how to teach ML from a critical empowerment perspective [124]. Thus, the design space for such learning tools and activities had yet to be defined at the beginning of my project, and a primary focus of my work has been on uncovering this space. Another of my focus points has been holding myself accountable to the people I aim to serve with my work. These are designers and design researchers in the education space, but also educators themselves. The theoretical contributions of my work (including this dissertation) are mainly aimed at designers and design researchers to add to the collective knowledge base and discourse around how to design for the democratization of digital technology. Further, sustaining knowledge in artifacts enables concrete discussions with stakeholders in the project, such as teachers and municipal staff, centered around specific artifacts, such as learning tools and activities. Finally, by explicitly making our work available and supporting scaling, we aim to support educators in adapting our artifacts to their context or developing their own teaching activities.

In this way, our experiments play two roles in my project: To allow a designerly, iterative exploration of the design space of CE for ML Education and as concrete design proposals that add to the discourse and offer new ways of teaching ML from a CE point of view.

Krogh and colleagues [130, 131] describe different types of relations between experiments and artifacts in design research projects as they develop over time, i.e., different ways of *drifting* (see Figure 3.1). The first is *accumulative*, where knowledge from previous design experiments is accrued in the next, leading to a final artifact embodying the total knowledge produced in the design process. The *comparative* way of drifting, where different design experiments work "from or towards a shared platform of comparison" [131, p. 45]. Each experiment is designed to evaluate knowledge produced in earlier experiments and uncover new aspects about the focus of inquiry as a way of "acknowledging complexities". *Expansive* drift occurs when the designer is working towards revealing an area of focus with unknown boundaries. There is no causal connection between experiments except the ambition of broadening the understanding of the design space. In contrast, *serial* drifting is characterized by methodical advancement, with each experiment continuously building on experiences from the previous. Finally, drifting by *probing* is characterized by "[e]xploiting opportunities and exploring design ideas as they emerge through design work" [131, p. 15]. Each experiment here is often personally motivated or designed to foster the most impact rather than following insights or ideas generated in previous experiments.

Among these five ways of drifting, the relation between the design experiments in my project can, broadly, be characterized as *comparative*. When seen together and from a distance, my design experiments do precisely this; each experiment is concerned with the same set of concepts and seeks to explore these from different angles while drawing on and challenging the knowledge and understanding

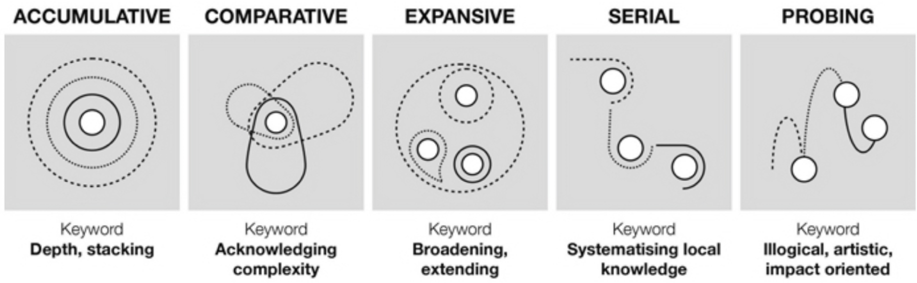


Figure 3.1: Five Ways of Drifting in Constructive Design Research by Krogh and Koskinen [130]. Included with authors' permission.

gathered from the previous experiments. Here, the central area of focus is Computational Empowerment for ML Education. My experiments revolve around this focus, exploring its different aspects in a reciprocal process of framing and reframing the program based on previous experiments and letting the program inform new experiments.

Apart from drifting within my design program, I underwent a significant drift, causing a reframing of the program into what I call Remarkable AI, which I unfold in chapter 6. This new framing expands the boundaries of my initial program. Moreover, it embraces the role interaction designers and the HCI community play in the shaping of (AI) technology which lends itself to be critically examined.

3.3 Disseminating Programmatic Design Research

Krogh and Koskinen [130] claim that design research in the programmatic epistemic tradition mainly uses theoretical frameworks as outputs for the knowledge produced in such projects. However, the interaction design and design research communities have found that the notion of the generalizable theory is often too high of a bar to be meaningful. But, knowledge is being produced in these communities that is more general than a singular design concept. Instead, researchers have explored, and continue to explore, intermediary knowledge, i.e., knowledge that somehow lies in the intermediary between concept and theory (e.g., [41, 46, 71, 96, 207, 242, 244]). According to Stolterman and Wiberg [207], this happens through design work that combines "individual theoretical concepts into constructs that bring together earlier findings in new concepts and artifacts". Höök and Löwgren [96] further this argument by suggesting other forms of intermediate-level knowledge in HCI. They divide these into generative (such as design patterns, guidelines, annotated portfolios, methods, and tools) and evaluative (e.g., design heuristics, experiential qualities, and design critiques).

In chapter 4, I present a set of six design principles to support designers of ML learning tools and activities and educators choosing to use these in their teaching. As described above, I have worked comparatively with our design experiments, informed by previous experiments' findings while exploring new aspects of the design space. To arrive at these design principles, I analyzed my experiments with regard to which design decisions caused a drift, that is, which design elements of each experiment were noticeable enough that it was either sustained through several design experiments or immediately discarded after evaluating the experiment. This analysis yielded seven design principles as a set of generative intermediate-level contributions to designing ML learning tools.

Further, in chapter 5, I present a framework for computationally empowering ML education. This framework builds on existing articulations of CE [50, 107], and expands on it with a *technological, material foundation*. The framework draws on Verbeek's idea of the morality of technology [225] and provides a vocabulary for decomposing how specific ML concepts, practices, and perspectives influence the impact of any given ML-systems. Further, the framework ties these technological foundations to the CE activity model suggested by Dindler, Smith, and Iversen [50, Figures 2 and 3], thus highlighting how these might be included in the design of concrete tools and activities.

Finally, in chapter 6 I introduce Remarkable AI as an approach to AI systems development, that emphasizes the need to open up these systems for their users. Remarkable AI represents one end of a continuum of remarkableness in AI systems (where Unremarkable AI [238] represents the other), and I argue that remarkableness should become a sensitizing concept for designers and developers of AI systems.

3.4 Participatory Design that Matters

In addition to CDR, I have drawn on ideas from Participatory Design (PD). I am motivated by democratizing ML, reflecting the ideals of traditional PD [22, 31] but with a focus on its influence on our everyday lives rather than our work lives.

At its infancy in the seventies and eighties, PD was a democratically motivated effort to empower on-the-floor workers as new technology was introduced at their workplaces [31]. In later years, digital technology, and with it HCI research, moved into our everyday lives [26], and new challenges rose for PD. In a 2018 paper, Bannon, Bardzell, and Bødker [8] discuss future directions for PD, highlighting the need to bring the democratization of technology into the educational setting. Another recent paper by Bødker and Kyng [30] calls for a return to the political nature of PD and a focus on technological ambition.

My research is a response to these calls. Its alignment with the goals of Computational Empowerment [50, 107] to empower students with regard to technology means that I am politically engaged in democratic discussions about technology and its use. Similarly, I have high technological ambitions with my focus on ML and developing state-of-the-art tools and activities for teaching it.

Concretely in my research, I mainly draw on PD methods by inviting stakeholders such as pupils, teachers, and representatives from the municipality into my design process and by allowing them to engage with ML in meaningful ways through experiments with the artifacts that take center stage in my project. However, my participants are only sometimes on equal terms with me in the project, as is customary in traditional PD projects. Instead, I focus on mutual learning during design workshops and teaching situations. I let this learning influence my work while maintaining autonomy over my designs. This is necessary to sustain the deliberate design practice of CDR, but leads to tensions with

PD: "[W]e note that many of the published papers that use [CDR] methods are designer-driven, feature minimal meaningful participation, and in many cases de-emphasize political conflict (though there are exceptions)" [8, p. 7].

3.5 Combining Constructive Design Research & Participatory Design

My motivation for combining CDR and PD lies in their shared reflexive stances toward design. In CDR, the designer is reflective in their conversation with the material of design [194] and experiments with their artifacts. In PD, the designer also reflects during the design process, but the reflection happens cooperatively between stakeholders, leading to mutual learning between the parts. Combining both approaches allows switching between the introspective reflection present in CDR and the extrospective reflection of PD, allowing designs to be examined from different angles.

However, there are tensions between CDR and PD in the emphasis on stakeholder involvement and empowerment. PD is a political project that aims to give voice to the "people on the floor" in design projects. This deliberately diminishes the autonomy of the designer whose role (in simplified terms) becomes to use their design knowledge to support the stakeholders in realizing a finished product that accommodates their needs. Thus, it is the job of the design researcher to uncover the needs on which to base the product and to support otherwise unheard stakeholders in the project. In contrast, CDR advocates the autonomy of the designer as an artist. Here, the specific Designer brings their values and beliefs, and the product is enriched by these, at least in artistic terms. This maintains design projects' typical hierarchy. However, CDR must uphold this hierarchy as design artifacts simultaneously serve as suggestions for solutions and deliberate ways of exploring and inquiring about a design space. Because of this latter purpose, autonomy is necessary: With sovereignty over their artifacts, designers can maintain the artifacts' deliberateness.

On the other hand, it might be possible that these tensions are more ideological than practical issues. In my case, I have cherry-picked methods from PD to invite stakeholders into the design process while maintaining the autonomy of my design decisions (although they have been negotiated between my colleagues and me). In any case, it is apparent that power relations between researchers and "subjects" are central for both methodologies precisely because design and its inevitability to be world-changing are central to them both. It might not be possible to avoid hierarchical structures altogether in design situations or to entirely preserve the designers' autonomy in projects with multiple stakeholders; perhaps neither is inherently desirable.

3.6 Project Collaboration

My project has been an intrinsically collaborative effort. Most prominently, I have collaborated with my colleague Karl-Emil Kjær Bilstrup (who features as a co-author on most of the papers this dissertation is based on). Through the project, we have collaborated closely as design and research partners. All experiments I present throughout this dissertation were designed in collaboration with Karl-Emil. This, naturally, is also evident from our dissertations, which are both parts of the CEED project, although they differ in the types and nature of their contributions.

Further, I have had an ongoing collaboration with other researchers in the CEED project. We have worked together on concrete design experiments, specifically the Critical Data Literacy Intervention *(Paper F)*, but have also collaborated by discussing and developing our separate experiments throughout the process.

Finally, to uphold the PD roots of CE, the CEED project established a sustained collaboration with a group of teachers in Aarhus, Denmark, known as De32[3]. The municipality established the group to support computing education and digital skills in schools around Aarhus. It consists of several teachers from different subjects bought out one day a week. The role these teachers played shifted as the research drifted. In some experiments, they took a consulting part. Here we approached teachers with an idea, mock-up, or early design prototype. In small workshops with 2-3 teachers, we would discuss this artifact; teachers would test it and provide inputs. We would then incorporate this feedback into the further design process. The Critical Data Literacy Intervention *(Paper F)* is the collaboration closest to PD's participatory ideals. Here, we collaborated closely with teachers throughout several workshops over four months and developed a five-day intervention, which we then carried out and evaluated in an 8th-grade classroom (see [192]). The final role teachers took was as classroom facilitators during our interventions. We carried out several experiments in which the teachers present had not been involved in the design process but were still in the classroom while we carried out an intervention. Here, we would have ongoing discussions during students' group work, and they would assist students needing help.

[3]see https://sites.google.com/aaks.dk/de32/startside, accessed Jan 7th, 2023. In Danish.

Chapter 4

Configuring the Design Space

This chapter presents and discusses the design process of my work as a Ph.D. student, from the infancy of my project to its concluding moments and beyond. When I began my work, little other work existed on teaching ML and AI with an explicit aim of scaffolding CE. Thus, the project took off into a relatively unknown design space, where little formal knowledge existed about how such learning tools could or should be designed. Thus, much of my work has been navigating and making sense of this design space, i.e., which handles and knobs a designer of such tools has in their possession and how to use these to design successful learning tools.

Below, I present in detail my design process, emphasizing how findings in one experiment drove the design of the next, or how I bridged what Hallnäs and Redström [80] term 'the hermeneutical gap' (i.e., the gap between my understanding of the design situation, what I learned from one experiment, and what was designed in the next). Examining my entire body of work allows the relationship between the experiments and the guiding principles that emerged through them to be explicated. Further, it highlights how these principles interact with each other.

Additionally, this design process is vital for understanding how I arrived at CE as a frame for ML education (as detailed in chapter 5), as well as my drift towards Remarkable AI (as presented in chapter 6). Thus, this detailed presentation of the design process serves to document the framing and reframing that occurred throughout my project. Recently, Zimmerman et al. [243] argued that actions of (re)framing are "an important outcome of design research, one that constructs an analytical argument for the value of the new frame as an outcome of a twisting and winding journey". As such, this chapter serves as the central 'data-set' for the forthcoming arguments in the dissertation.

Central to this chapter is a set of design principles derived inductively through analysis of my design experiments. Notably, these principles are not an exhaustive list but rather the central tenets of my work. Thus, I invite other researchers to contest or expand upon them with experiments of their own. These principles (which will be presented in detail in section 4.2) are:

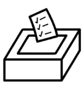
- Considering Tangible Representations of ML
- Providing Low Floors to Engaging with ML
- Providing Opportunities for Meaningful Engagement
- Scaffolding Students' Fundamental, Technical Understanding of ML
- Emphasizing Critical Engagement With ML
- Designing for Scalability

Some of these principles are self-evident and well-known from research on education and computational thinking (e.g.,[23, 142, 184]). However, there exists a series of tensions and synergies between the principles, which aren't obvious when viewing them from afar: How a designer balances the emphasis on these principles affect the success of a given ML learning tool. By providing both a set of design principles and a portfolio of the tensions and synergies within these principles, I aim to aid designers in designing such tools, as well as a metric by which educators might judge existing tools before investing their time and money in them.

4.1 My Design Process, Explained

Below, I present my design process in chronological order[1], with each design experiment marked by a subsection. Following the description of the experiment itself, I describe findings regarding the design of the experiment that lead to new directions for the following experiments. Note that these findings are not necessarily addressed in the immediately subsequent experiment but sometimes in a later experiment. Additionally, not all findings from each experiment are present here, only findings that directly influenced the design of a later experiment. Instead, for each experiment, I will provide a reference to the publication describing it, wherein more details can be found. An overview of the process can be seen in Figure 4.1.

4.1.1 VotestratesML



To explain my process, I have to start before my Ph.D. starts with my Master's thesis. My colleague and I designed VotestratesML *(Paper A)*, our first attempt at teaching ML from a CE point-of-view, with a CE frame (see Figure 4.2).

VotestratesML (VML) is a web-based platform for teaching ML in a Social studies classroom. It allows students to create ML models to predict voter behavior by pulling from voter data from a post-election study based on the general election in Denmark in 2015. Here, students work collaboratively in groups to create a ML model. To motivate students, VML includes a competitive element in which the different groups compete against each other to create the "best" models. VML

[1]The chronology of publications generally follows the order of the experiments, but notice that the experiment with the Machine Learning Machine *(Paper C)* precedes the ML Ethics Workshop *(Paper B)*, although the publications dates, and such the numbering of the papers, are the other way around.

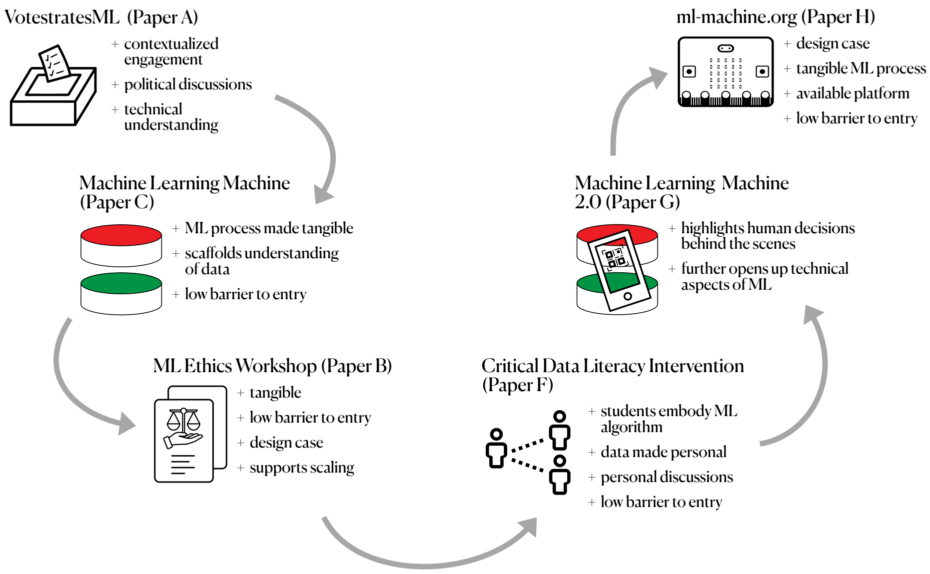


Figure 4.1: A complete overview of my design process. For each experiment, qualities that are carried over to the next are highlighted.

allows students to choose the goal of the model (e.g., to predict what party voters will choose), which features of the data-set the model should consider (e.g., their opinion on tax policies, immigration, if their parents belonged to a specific party, etc.), and to configure different parameters of the model (e.g., number of epochs and hidden layers in a neural network-based model). VML was framed as a learning tool for social studies first and for ML second:

> [I]nstead of starting by introducing ML concepts, VotestratesML takes departure in the social studies subject as a tool for analysing voter behaviour. The tool is designed to support typical social studies class activities; Students work in groups on tasks assigned by the teacher, followed by a discussion based on the group work. The available data categories were chosen based on models of voter behaviour taught in Danish social studies classes [108], allowing students to explore how ML models can predict voter behaviour, compare theoretical models with ML models, and discuss ML models from a social studies perspective and how they are already used actively to affect the outcome of democratic elections [3, 158, 185] (*Paper A*, p. 100[2])

The reason for this contextual framing of VML is to be found in an earlier experiment. Here, we designed a quick-and-dirty prototype using Google Collab[3] (collaborative Jupyter Notebooks) to allow students to build ML models on a variety of data. This early prototype was not contextualized as a social studies-first prototype but instead aimed solely at providing students a first experience of what we (at the time) deemed the most important ML concepts. The high school students participating in our pilot study with this tool were generally

[2]Page number refers to the included manuscript in the dissertation.
[3]https://colab.research.google.com/

Figure 4.2: Two students creating ML models to predict voter behavior with VotestratesML

not particularly engaged by this decontextualized and technical-looking interface, stating that "For this to be really exciting, you probably need to be interested in the subject (ML)" (Paper A, p. 107[4]).

This finding and conversations with a high school social studies teacher led us to the idea of a highly contextualized tool that contributed not only to students' understanding of ML but also to their subject-specific knowledge and self-identification as social science students. At the same time, this contextualization also provided the opportunity of focusing on how ML was/is used in democratic elections with the Cambridge Analytica [37] case and the upcoming Danish general election as focus areas.

From our studies with VML, we found that while the tool did foster some discussion on ML ethics and indeed was engaging to use, we were not successful in teaching students the role of data in ML, e.g., "If choosing a feature did not have the expected effect or a model made an unexpected prediction, most students rejected what the model actually stated, and instead wrote it off as a problem with the data, rather than reconsider the veracity of their expectations" (Paper A, p. 106[5]). Another finding, and unintended consequence of our design, was that its competitive element led students to dismiss otherwise exciting discussions in the pursuit of a 'better' model.

These findings inspired the design of our two following design experiments: the Machine Learning Machine *(Paper C)* and the ML Ethics Workshop *(Paper B)*.

[4] Page number refers to the included manuscript in the dissertation
[5] Page number refers to the included manuscript in the dissertation

Figure 4.3: The two devices of the MLM. To the left is the Trainer device and to the right is the Evaluator. From *Paper C*.

4.1.2 The Machine Learning Machine

The Machine Learning Machine *(Paper C)* is an attempt to make the process of creating a data-set, training, and then testing and improving a ML model as tangible as possible. The design of the MLM (see Figure 4.3) was directly informed by our finding with VML that students struggled with understanding the role of data. It addresses this by moving the data into the physical world and simplifying it:

> "The MLM is a learning tool which enables K-12 students to iteratively work on ML models for binary classifications of doodles they draw using pen and paper. It consists of two separate devices [...]: the Trainer, a big industrial looking metal box which provides a simple interface for labelling data and training ML models; and the Evaluator, an instrument inspired by laboratory equipment for analysing and evaluating ML models' performances. Besides the two units, the ML model that learners build is represented by a physical artefact. This model is moved between the two units to either train or evaluate the model". (Paper C, p. 140[6])

Another rationale behind the design of the MLM was to slow down users' engagement with ML, as compared to VML: "[The MLM] requires students to use a slowly moving, laborious system, and asks them to make time and space to assemble around ML, to explore it together, and to take their time to reflect on details" (Paper C, p. 10).

However, as the MLM focused mainly on teaching the ML process and key concepts, the participants' reflections were about this. In our pilot study, we found that the MLM effectively highlighted the role of data. The participants engaged intently with the aspects of ML highlighted by the MLM. They created different hypotheses to explain data's role in the MLM and devised various strategies for testing these. On the other hand, the participants struggled with understanding the idea of the model artifact, specifically that it represented a ML-model with

[6] Page number refers to the included manuscript in the dissertation

Figure 4.4: Students engaged in the ML ethics workshop. From *Paper B*.

an abstract understanding of the data rather than serving as "'memory cards', storing photos of the drawings and moving them between the Trainer and the Evaluator" (Paper C, p. 149[7]).

Discussions about ML ethics were absent in the pilot study, and while we suggest activities for scaffolding such discussions in Paper C, we did not evaluate these activities. However, as argued above, scaffolding a fundamental understanding of ML concepts and processes is an inherent part of qualifying such discussions. Finally, while the MLM is an engaging prototype the design of it: two large custom-built boxes, a 3D-printed model artifact, etc., highlighted another challenge: if we are, as I propose, to teach ML and AI on a large scale, we also need tools that scale beyond HCI-research prototypes.

4.1.3 ML Ethics Workshop

The other issue with VML, should the reader not remember, was that discussions on ethics were hampered by its competitive aspect. Additionally, such discussions were only peripherally included in the design of the MLM.

This inspired us to design another experiment with a greater focus on fostering these discussions. With the ML Ethics Workshop *(Paper B)*, we created a design-based workshop in which students engage with a design case to design a ML-based mobile application to solve an issue with their peers (other teenagers) as the target audience (see Figure 4.4). We designed this workshop to be completely unplugged, needing only paper materials, which we provided in the studies we conducted. Specifically, the workshop is inspired by Inspiration Cards [82] and includes a deck of several cards used throughout (see Figure 4.5). The workshop

[7]Page number refers to the included manuscript in the dissertation

Figure 4.5: Some examples of the cards used to guide the workshop activities in the ML Ethics Workshop *(Paper B)*. From *Paper B*.

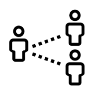
format is as follows: First, students are given a short introduction to ML, and the facilitators present the design case. Next, students analyze the context of their case using *data*-cards provided by facilitators. Next, students ideate, coming up with as many ideas for applications as possible. Then, they select an idea and describe its ML component using a template provided by the facilitators. Next, students reflect on the ethics of their proposed applications, prompted by an *ethics*-card provided by the facilitators. Following this, students redesign their applications according to the above discussion. Finally, students present their applications to each other, followed by a classroom discussion.

From our evaluation of the workshop format, we found that it was successful in scaffolding discussions about ML ethics. In particular, the design case tied ethical considerations to students' specific designs, providing them with a concrete starting point for their discussions. However, the workshop also highlighted the need for a solid foundational understanding of ML when discussing its ethics:

> "The group discussed whose responsibility it was, if the system recommended an exercise that caused an injury, but their solution was to employ professional testers to prevent this from happening. It was, however, unclear how these testers would prevent the system from recommending wrong exercises, even if all exercises had been tested. And it is also unclear how the testers would be made responsible for bad recommendations." (Paper B, p. 126[8])

In this case, a more profound familiarity with ML would have allowed students to qualify their solution technically and to have used more precise wording about their thoughts, allowing others to critique it as well.

4.1.4 Critical Data Literacy Intervention



The Critical Data Literacy Intervention *(Paper F)* was carried out over five days in an 8th-grade classroom of a Danish public school. As the name suggests, the

[8]Page number refers to the included manuscript in the dissertation

Figure 4.6: Students embodying a KNN model, posing for a photo of their classification.

focus was on developing students' data literacy, specifically on how data are used in ML. The intervention took students through a series of activities, in which they became acquainted with how data are collected about them, how this data can become more valuable than the sum of its parts, how the data can be used to make ML models, and what these models can be used for. We designed the experiment around four artifacts; a pop quiz, the *data compass*, an embodied K-nearest Neighbors (KNN)[9] exercise (see Figure 4.6), and the KNN Exploration Tool.

"The pop-quiz is an online form, created in Google Forms[10], that mimics the quizzes pupils see on social platforms and in magazines, which promise insights about themselves or their peers if they answer a few questions" (Paper F, p. 229). The purpose of the activity was to illustrate to students, how data are collected on the internet through seemingly innocuous interactions. The quiz asks questions such as students' age and how far they travel to school, as well as more personal questions such as how important making a lot of money is to them.

Students answered the quiz as homework before the intervention, in which we revealed to them that this had been a ploy for us to collect data about them. Students then used the Data Compass (see Figure 4.7, to the left) to reflect on the value of this data by rating each question with regard to their willingness to share the information and how they perceived its value to others.

[9]KNN is a simple ML algorithm, that classifies a new data entry, by calculating its distance to the K-nearest neighbors.

[10]Note that pupils had access to the Google Educational Suite which is compliant with local legislation on data collection and that all data collected were thoroughly anonymized before leaving this platform.

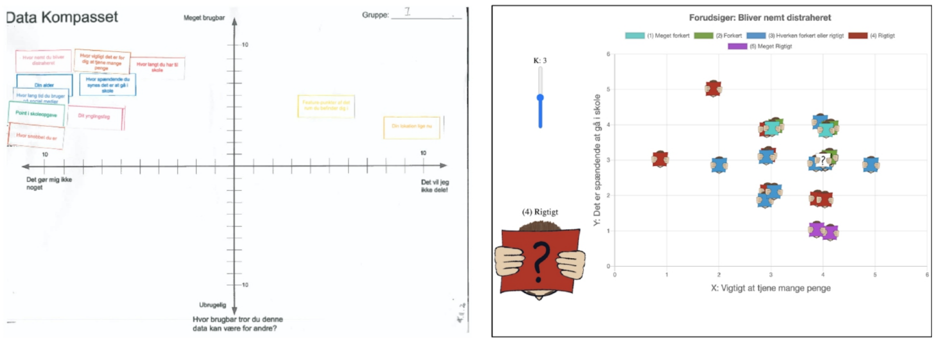


Figure 4.7: To the left is a filled-in Data Compass. To the right is a screenshot of the KNN Exploration Tool. From *Paper F*.

Following this, we taught the students how seemingly mundane data can become valuable when aggregated. We did so with the Embodied KNN exercise:

> "[A]n unplugged activity [15] where pupils embody a two-dimensional KNN model, acting as data points representing their own personal data. A coordinate system is drawn on the floor in a large room and pupils arrange themselves in the coordinate system according to the axis and simulate the KNN algorithm with different values of K by asking their two, four, and six closest classmates questions". (Paper F, p. 229)

For example, we ordered students by height and stationary activities per day and used this to predict their favorite sport. While this simple model was imperfect, it highlighted how some data (such as height) can be used as a proxy for other information (such as gender), based on which more sophisticated predictions can be made.

Finally, students used our KNN Exploration Tool (see Figure 4.7, to the right): "a digital tool which scaffolds pupils in further exploring the class' data (from the pop quiz) and build different KNN models" (Paper F, p. 229). To guide their work with this tool, we presented a design case to the students: *The School for Gifted Children*, in which students were tasked with creating a model to determine which students could be admitted to an exclusive private school.

From the Data Literacy Intervention, we found that working with, modeling, and discussing students' personal data made ML personally meaningful to them. They were able to reflect on the value of data by imagining how it might be used to affect their own lives. Further, the personal nature of the data also helped create a low barrier of entry for engaging with it. Students were eager to discuss how data can act as proxies for related information and were generally engaged throughout the intervention.

We also found that having students embody their data made it more graspable to them; in the embodied KNN exercise, students paid close attention to each other's positions in the coordinate system and would complain when, e.g., they felt someone overestimated their height. In addition, some students could draw connections between this exercise and how the YouTube recommender algorithm works by looking for what videos people similar to you are watching. Students

Figure 4.8: Students using the MLM 2.0. From *Paper G*.

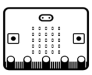
did, however, still need guidance from the facilitators in the classroom discussions, and not all could relate the embodied exercise to the digital instances of ML they encounter daily.

4.1.5 MLM 2.0

Next, we designed MLM 2.0 (see Figure 4.8), which, as the name suggests, is an iteration of the original MLM. Here, we outfitted the physical MLM boxes with QR-codes in strategic places that link to web pages allowing students to access information about the underlying ML processes behind the scenes. Additionally, students are given access/the option to alter the parameters of the ML algorithm, enabling them to experiment with how these affect the model's outcome. In our pilot study with MLM 2.0, we experienced that children would use these digital options to test hypotheses about the workings of their models.

While these studies were conducted focusing on how the digital addition would affect children's use of the MLM, I see this addition as a new way to include further possibilities for facilitating discussions about ethics using the MLM, as it highlights human-decision making inside the guts of ML. However, this version of the MLM does not address the original's scalability issues.

4.1.6 ML-MACHINE.ORG



To do so, we designed the final experiment to be included in my dissertation (and the last one we made. . .), namely ml-machine.org *(Paper H)*. With this experiment and prototype we aimed to keep as many as possible of the MLM's tangible aspects but to design a system that would lend itself more readily to scaling. This prototype consists of a web page (see Figure 4.9), that allows students to train ML models, using the BBC micro:bit.

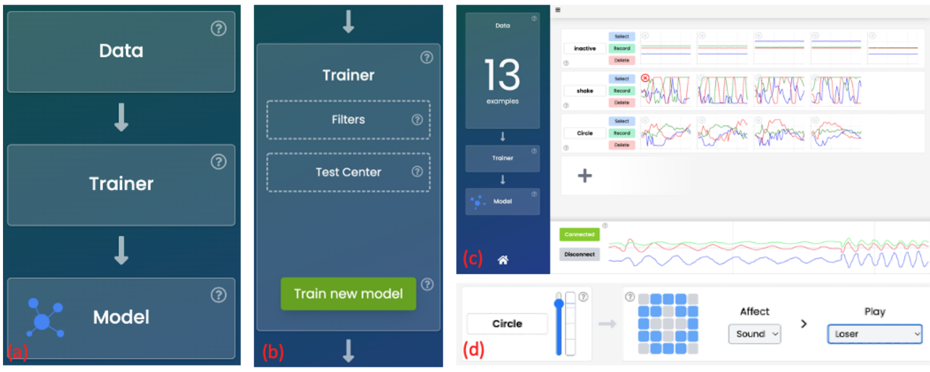


Figure 4.9: The ml-machine.org webpage. From *Paper H.*

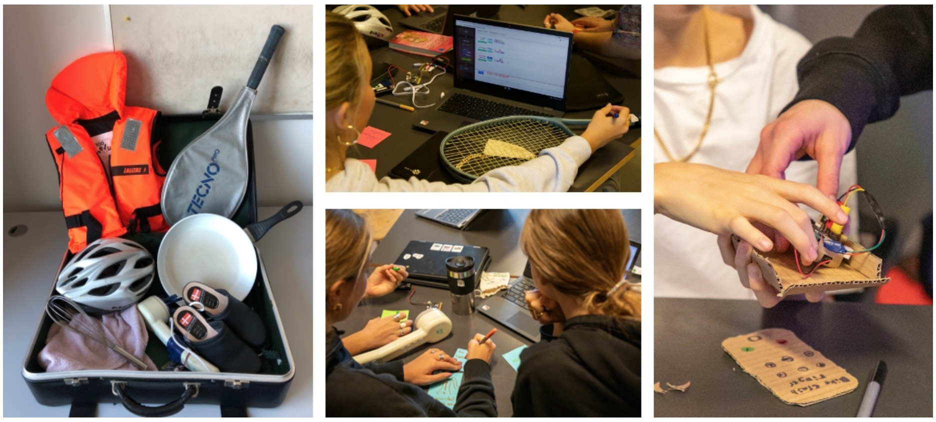


Figure 4.10: Students using ml-machine.org to redesign and augment everyday objects using ML. From *Paper H.*

> Students collect accelerometer data with a micro:bit and process it in the graphical user interface on the web page. They use the processed data to train machine learning models to recognize/predict different gestures/movements. Finally, they can send their model's predictions to a second, output-micro:bit, on which they can change the LED display, play sounds, and activate pins (to, e.g., activate a vibration motor). (Paper H, p. 2)

Apart from the micro:bit being designed explicitly for children to learn about computing, it has also been rolled out to a large portion of Danish public schools as part of an effort from the Danish Broadcast Company (DR) to teach computing in this context.

In our studies with ml-machine.org, we have centered our teaching unit around a design case (as was the case for the ML Ethics Workshop *(Paper B)* as well). Here, we tasked students with augmenting everyday items with ML (see Figure 4.10). We did so, to ensure that students would face similar issues in creating their models/systems as those faced by real designers of ML-based systems.

As we had hoped, we did indeed see that students would engage with these issues while using ml-machine.org: issues such as data representativeness, unintended bias, etc. However, as was the issue with the MLM, the nature of the data collected only allowed for a skin-deep discussion, and needed significant translation efforts to scaffold a discussion of how ML is used in critical instances in the everyday systems we are surrounded by. As we have been working with DR, to create a curricular unit with the ml-machine to scaffold critical discussions of ML, this is a challenge we have experienced first-hand. Teaching activities with a critical goal, need to surround the hands-on activities with ml-machine.org with a meta-layer dealing with the issues surrounding ML. While not necessarily a bad thing in isolation, it does indicate that the design of ml-machine.org does not immediately support critical discussions of ML, but rather focuses on the technical foundations necessary to have these discussions.

4.1.7 Summary of Design Process

In summation, I have undergone a process characterized by learning, iteration, and hypothesis testing. The design process has been exploratory first and foremost. I was experienced in design before starting it, but not with either education, or ML, and on a personal level, this process has taught me many things about both. I also argue that the process can teach other designers about how to use design inquiry to delimit a hither-to-unknown design space. Using the process outlined above, and the takeaways and ways of drifting laid out in it, I present how below.

4.2 Design Principles for ML Learning Tools & Activities

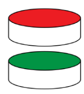
In this section, I present a set of design principles, derived from analyzing the qualities in each of the above design experiments, and how these informed the following experiments. As such, these principles are what I have identified as the design features in my experiments that had enough merit to be carried through several experiments, and thus, arguably, transcend the individual learning tool or activity. I arrived at these principles by analyzing each experiment and identifying its features, such as how ML is interacted with, presented, and how it addresses ML concepts, practices, and perspectives. Then I considered which of these features reappeared in other experiments and from this identified the set of principles. Note, that while many of these features were included in a single experiment based on related theoretical work or experiences in other contexts, it is because of their staying power in our series of experiments that they are included here. An overview of the principles can be found in Table 4.1.

Below, I briefly characterize each of these principles and discuss how my design experiments have informed them. Following this, I highlight how certain synergies and tensions are present between principles when applying them in practice, and discuss how these can inform the practices of designers and educators working with ML education.

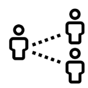
Design Principles for CE-based ML Education	**Central Experiments**
Considering Tangible Representations of ML	Machine Learning Machine *(Paper C)*, ML Ethics Workshop *(Paper B)*, Critical Data Literacy Intervention *(Paper F)*, Machine Learning Machine 2.0 *(Paper G)*, ml-machine.org *(Paper H)*.
Providing Low Floors to Engaging with ML	VotestratesML *(Paper A)*, Machine Learning Machine *(Paper C)*, ML Ethics Workshop *(Paper B)*, Critical Data Literacy Intervention *(Paper F)*, Machine Learning Machine 2.0 *(Paper G)*, ML Ethics Workshop *(Paper B)*ml-machine.org *(Paper H)*.
Providing Opportunities for Meaningful Engagement	VotestratesML *(Paper A)*, ML Ethics Workshop *(Paper B)*, Critical Data Literacy Intervention *(Paper F)*, ml-machine.org *(Paper H)*.
Scaffolding Students' Fundamental, Technical Understanding of ML	VotestratesML *(Paper A)*, Machine Learning Machine *(Paper C)*, Critical Data Literacy Intervention *(Paper F)*, Machine Learning Machine 2.0 *(Paper G)*, ml-machine.org *(Paper H)*.
Emphasizing Critical Engagement with ML	VotestratesML *(Paper A)*, ML Ethics Workshop *(Paper B)*, Critical Data Literacy Intervention *(Paper F)*
Designing for Scalability	ML Ethics Workshop *(Paper B)*, Critical Data Literacy Intervention *(Paper F)*, ml-machine.org *(Paper H)*.

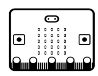
Table 4.1: Design Principles for CE-based ML-education, and their related experiments.

4.2.1 Considering Tangible Representations of ML



Being a digital, and often an automated process, ML can be difficult to grasp, and so, providing opportunities for users to "get hands-on with it" has, in my work, proven an effective way of familiarizing them with its inner workings. The literature further suggests that tangible user interfaces can be effective when teaching abstract and complex phenomena [97, 143]. Additionally, the research suggests that providing an option to transition to a graphical interface, might help students choose the interaction style that most suits their preferences for the specific task at hand.





This quality is especially present in the design of the MLM *(Paper C)*, and MLM 2.0 *(Paper G)*. These artifacts were designed with tangibility in mind, as an exercise of how to represent ML concepts and the ML process tangibly, with MLM 2.0 providing additional functionality by providing an optional graphical interface. However, tangibility is also present in other experiments, notably in the ML Ethics Workshop *(Paper B)*, where students use pen and paper to outline a ML-system prompted by inspiration cards, and in the Critical Data Literacy Intervention *(Paper F)*, where students embody data points in a spatial ML-system. Tangibility is also central to ml-machine.org *(Paper H)*, in which students collect gesture data with the micro:bit.

This principle advises designers to consider which aspects of ML they aim to teach and to explore different ways of representing them. Tangible representations offer advantages, such as with ml-machine.org *(Paper H)*, where we observed that students would explore the nature of gesture data, and stumble on important findings, such as the importance of sensor placement, of their own accord. However, they also have disadvantages such as being less dynamic than graphical representations, and the issue of scalability. For educators choosing which tools to deploy in their classroom, a similar consideration is necessary: tangible tools allow children an embodied experience with ML, and good tangible tools can be effective for providing a first experience with ML's complicated nature. However, graphical tools might offer more flexibility and ways of visualizing and interacting with complex phenomena that might not be possible to represent tangibly without oversimplifying them.

4.2.2 Providing Low Floors to Engaging with ML

As we experienced in our early work with VML *(Paper A)*, ML can be daunting to explore by itself, as it can be complicated to use and difficult to understand. Resnick and Silverman [184] argues that when designing learning tools for CT more generally, the "challenge is to develop features that are specific enough so that kids can quickly understand how to use them".

All of my work aims to provide low floors. We have aimed to design all tools so that they could be used with minimal guidance and supervision. VML is designed such that users cannot bring it into an undesirable state, such that only choices that result in a functional model can be made. The ML Ethics Workshop *(Paper B)* provides templates and inspiration cards that scaffold students' activities. While this didn't completely ensure that all students' designs were technically sound, it did seem to increase the degree of their feasibility.

Here, my advice to designers of ML learning tools is to consider the first experience a child has when using said tool. When learning about a complicated technology such as ML, it is important that the design of the tool itself is welcoming and forgiving about the child's level of understanding of ML. For educators, this consideration is equally important. A learning tool might offer advanced functionality, but a barrier of entry that counteracts whatever advantages the tool might else have.

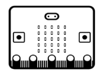
4.2.3 Provide Opportunities for Meaningful Engagement

Meaningful engagement is a broad term that refers here to giving students the opportunity of seeing themselves in ML, and exploring the ways it (can) affect their lives and the things they care about.

Our experiment with VML *(Paper A)* highlighted that students' attitudes towards ML and digital technology can be a major hurdle for them with regard to engaging with it. However, by situating VML as a social studies tool first, students were able to recognize the value it might bring them in their self-identification as students of the social sciences. With the Critical Data Literacy Intervention *(Paper F)* and the ML Ethics Workshop *(Paper B)*, we framed ML as something personal to the students; the data they provide systems on a daily basis is valuable, can be used for purposes not easily noticed, and the systems they are used in entail complex ethical dilemmas. Thus, providing a contextualized frame for ML can be effective means of scaffolding engagement that's personally meaningful to students, whether it is related directly to a subject that interests them or to their personal lives.

The literature on CT in education routinely argues, that contextualizing technology is an important factor when teaching it (e.g., [79, 84, 159, 197]). Hambrusch et al. [84] argue that students "will comprehend computational concepts more easily if those concepts can be motivated by examples from their scientific subdisciplines." However, as observed in our studies with VML, this is not limited to the science disciplines, but, I argue, all subjects.

Another powerful way of scaffolding meaningful engagement is through construction. This idea is directly related to Papert's influential theory of *Constructionism* [169]; that learning happens through "reconstruction rather than as a transmission of knowledge..." and that "...learning is most effective when part of an activity the learner experiences as constructing a meaningful product."



We have experienced this, especially with the ML Ethics Workshop *(Paper B)*, and ml-machine.org *(Paper H)*. Here, students construct respectively an implementable but non-functional and a functional interactive ML-system. Students themselves are in charge of their own design process, and decide what their system looks like, and what functionality it has. It might be the case, that the meaning students find in these systems is purely that they are entertaining, but it engages them nonetheless.

The key for designers is that ML learning tools should be meaningful to the students using them. This might sound trivial, but that is not the case. It entails a serious consideration of students' contexts and is determined by the age of students, what subjects they are in, what their personal interests and stories are, etc. For educators, it is important to consider what subject they are teaching and if the tools they are considering is designed explicitly for this context or is a generalized tool. It is also important that they trust, that learning through construction is a fruitful endeavor, even though it might be less straightforward to predict exactly what students will take with them from such activities.

4.2.4 Scaffolding Students' Fundamental, Technical Understanding of ML

When teaching ML with the aim of developing students' ability to critically reflect on its use and the implications thereof, it is particularly important to ensure that they gain a fundamental understanding of the technical specificities of ML. Doing so ensures that such reflections do not merely become "doomsday" prophecies, but actually take into account the technical and factual limitations of ML. On the other hand, it alleviates that ML becomes a buzzword, a sort of "magic dust" to sprinkle on any system for it to magically become better, and grounds discussions in what is technically possible.

Kafai, Proctor, and Lui [111] argue that "...we need to frame computational thinking beyond an understanding of computational concepts and practices [...], to include an understanding of the values, biases, and histories embedded in the digital technologies", but that such a focus "...does not always guarantee more in-depth computational understanding".

I argue, that such a lack of in-depth understanding might actively hamper students' ability to understand values, biases, etc. As an example, take the issue of data representativeness in ML. Without a fundamental understanding of both data, and ML, understanding the impact and potential pitfalls of data representativeness is difficult. This is demonstrated in my work with the ML Ethics Workshop *(Paper B)*, in which some students struggled to identify relevant issues in their ML-systems, as well as in the VML *(Paper A)* studies, where students' limited understanding of the role of data in ML, led to false conclusions, and hampered them in classrooms discussions.

This principle proposes that serious design work be put into conveying the underlying, technical principles that are relevant to critical aspects of ML the designer is aiming to teach. Teaching students to develop a ML-system is not the goal. Instead, focus should be put on conveying the rules of ML, its limitations, and possibilities such that students can apply these when trying to understand how they interact with ML in the wild. For educators, I advise a similar consideration: examine what promises a learning tool is making and scrutinize how these are tied to technical components of ML.

4.2.5 Emphasize Critical Engagement with ML

Critical engagement should be made central in any tool and activity promoting it. By this, I mean that the design should offer direct ways of engaging students' capacities for critical thinking about ML. This means that critical elements of ML should be brought to the forefront while interacting with the tool or activity.

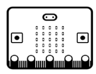
Importantly, I am not implying that all tools should necessarily engage with potential personal or societal consequences of ML, such as VML, and the ML Ethics Workshop does in my work. These experiments deal directly with the accountability of stakeholders in ML systems and whether ML systems are fair or not. These are crucial to consider, when putting together a curricular unit on ML but, arguably, it is equally as important that students come to understand how and why ML systems have potentially negative consequences, i.e., the technological materiality and practices that lead to these. In my work, experiments such as the MLM are not directly positioned as tools that emphasize the potential impacts of ML. Rather, through engaging with the MLM, students become familiar with technological concepts such as data, and modeling, and practices such as formulating problems to be solved with ML, and how data can be (and sometimes might not be) representative of this problem. Asking these questions, I argue, is central to decoding the ML systems that surround us, and serves to qualify discussions of them.

I propose emphasizing critical engagement as a key principle when designing tools and activities for computationally empowering students with regard to ML. Additionally, I wish to assert that to do this, it is not sufficient to focus solely on the potential consequences of ML, but also on the technological fundamentals and practices of ML, that lead to these consequences. For designers of such tools and activities, this means that even when emphasizing consequences, attention should be paid to the reality of ML systems as they are designed today. While it might be thought-provoking to imagine ML dystopias, unless these are carefully grounded in present-day technology (in the way that the best science fiction novels do), they might never evolve beyond provocation. On the flip side, tools, and activities that aim to teach ML fundamentals are beneficial in themselves, if designers consider how these interface with the critical aspects of ML. Finally, for educators implementing these in their classroom, it is important to consider not only the impacts of ML that they wish for their students to discuss but also how ML technology brings these to reality by how they function and are made.

4.2.6 Design for Scalability

The pursuit of computationally empowered citizens is an enormous task, which will require a myriad of learning tools and activities to accommodate cultural and social differences and preferences of learning styles. This given, designing tools for scalability is not merely a matter of designing the most generally pleasant tool, that most people would want to use.

Westley and Antadze [230] differentiate between two types of scaling in the context of social innovation; scaling *out*, and scaling *up*. Scaling out is what we typically think of when considering scalability, namely how many people our systems are used by, and/or affect. On the other hand, scaling up entails engaging with "the broader economic, political, legal or cultural context", to connect it to resources, policies, etc. These two types of scaling are intertwined; scaling up might help to scale out, and vice versa. Manzini [139] argues that to scale systems out is not simply a matter of making them widely available, but requires different people, to adapt the ideas behind them to different contexts. He further argues, that to scale up it is "almost always necessary for entities with the means and circumstances to act on a large scale", and that smaller actors (such as individual researchers, designers, and educators) seek out connections to such entities.



With ml-machine.org *(Paper H)*, we have been, and are, working on scaling the system both out and up. When designing the system, we made deliberate design choices with scalability in mind, such as working with the popular micro:bit, and running the system in a web browser. Apart from being more readily implemented in Danish classrooms, this also allowed us to make strategic partnerships with both the Danish Broadcast Company (DR) and the Micro:bit foundation. DR and their department for children and youth, already produce high-quality curricular units for teaching technology, that has a nationwide span. The Micro:bit foundation on the other hand oversees the entire development of the micro:bit platform, and were able to assist with scaling on an international level, as well as providing in-depth technical assistance.

When designing learning tools for ML with regard to scalability, designers should consider which platforms are available in places they want their systems to be used. Many classrooms have laptops, some have maker labs, micro:bits, or other low-cost microcontrollers available. Instead of designing novel interactive systems, consider appropriating off-the shelves products where possible. Apart from making the tool readily available to more people, designing for platforms that teachers can be expected to be familiar with, might also make them more likely to adopt it in their teaching. In addition, consider which strategic partnerships are relevant for scaling the tool and if they might have certain requirements for a possible collaboration (such as using a specific platform). For educators selecting tools for prolonged use, I advise choosing tools that have considered the aspects of scaling, and in particular, tools that have support systems set up, such that

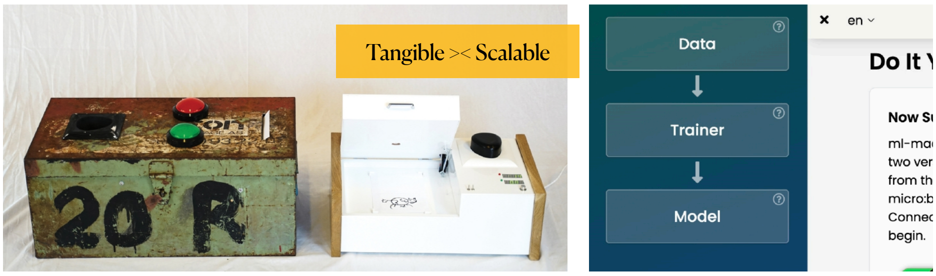
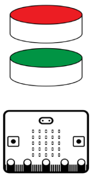


Figure 4.11: The tension between Tangible and Scalable is exemplified through the Machine Learning Machine *(Paper C)*and ml-machine.org *(Paper H)*, which has similar goals, but due the MLM being a custom built tangible design, it does not scale as well as ml-machine.org

help can be found if it becomes necessary.



There is an argument to be made for one-off, particular prototypes such as the Machine Learning Machine *(Paper C)*. They allow researchers to experiment with different ways of interacting with technology, that might find its way into systems that are more readily available in the classroom (as is the case with the MLM *(Paper C)*, and ml-machine.org *(Paper H)*). This is, however, not a given, and sustainability is an ongoing discussion in the PD community [30, 103]. Researchers should consider the scalability of their work to ensure that it can be sustained in the classroom, and beyond their involvement.

4.3 Tensions & Synergies in Computationally Empowering Learning Tools for ML

The design principles outlined above were extracted from my design experiments deductively, by analyzing them with regard to which design elements were either dismissed or carried over between experiments. As such, they represent my "best practices" of a sort; a set of levers to pull when designing new tools and practices.

When going through my experiments, it became clear that these features played different roles in different experiments; some focused on developing their critical understanding and toned down the focus on technical understanding, and vice versa.

I have been able to identify a series of tensions and synergies that make up a multidimensional tug-of-war in which the design space of "computationally empowering learning tools for machine learning" is suspended.

In this section, I present and discuss tensions, while the following section discusses synergies. Finally, I discuss the implications of these for designing learning tools for ML.

4.3.1 Tensions

Tangible >< Scalable: First, designing tangible tools is in tension with designing for scalability (see Figure 4.11). Making as many aspects of ML tangible as possible (form) was a significant design challenge, which I have addressed with

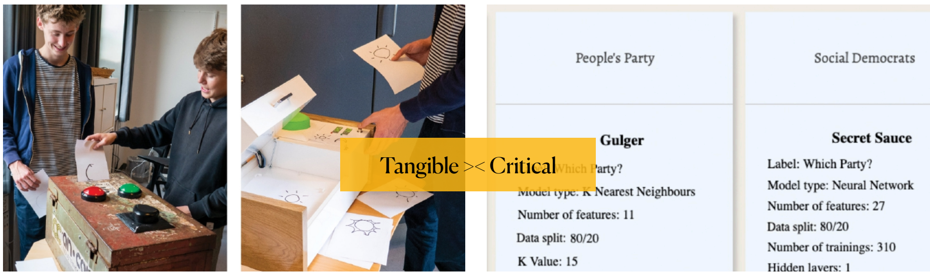
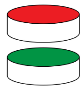
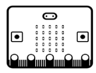


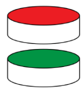
Figure 4.12: The Tangible >< Critical tension is exemplified by comparing the Machine Learning Machine *(Paper C)*and VotestratesML *(Paper A)*. The tangible nature of the MLM, means that the data considered need a physical representation, where the design of VML allowed it to address data of a political nature.



the MLM, which is perhaps as much an exercise in TUI-design as it is a suggestion for a learning tool. However, it highlights that making ML tangible is not a trivial task, and one that requires careful design, and not least custom interactive elements, that are not available off-the-shelf, nor affordable or practical to deploy nationwide or even internationally.

To achieve scalability, at least in a pragmatic, and practical sense, one needs to be aware that this trade-off exists. Off-the-shelf products such as the micro:bit used in ml-machine.org offers greater scalability but sacrifice the deliberate tangible qualities that a custom-built TUI such as the MLM offers (at least it sacrifices some of them). To design a tool that still makes ML graspable to its users, therefore, entails identifying these trade-offs, working with the limitations of the chosen platform, and (if possible) addressing its shortcomings through other means. With the ml-machine.org, we address the short-comings of the micro-bit platforms when compared to the MLM through the design of the website. While it cannot, by its nature of being digital/graphical, directly replicate the MLM's tangible qualities, some are translated instead to graphical components, such as the spatial distinctions between training and evaluating a model.





Tangible >< Critical: Another tension that permeates my work is between tangible tools and tools that directly address critical aspects of ML (see Figure 4.12). The tension here shows up due to the difference in the nature of the data involved in tangible tools (at least the ones I've been involved in designing) and how ML is most often used. For data to be made tangible, or even to be part of a tangible system, it needs some physical representation. However, data in most real-life ML applications does not have a direct physical counterpart. Instead, it more closely resembles the data used in VML: rows and columns in a spreadsheet.

As such, there is (at least in my work), a trade-off between designing tools with tangible interaction, and tools where discussions on ethics are central, and not an afterthought. That is not to say, that the two qualities cannot co-exist, but rather that they are in tension with each other, and a designer should consider how to most effectively address this tension.

The MLM, for instance, is not designed for critical discussions to be an integral part of using the tool. It does support activities that demonstrate critical aspects

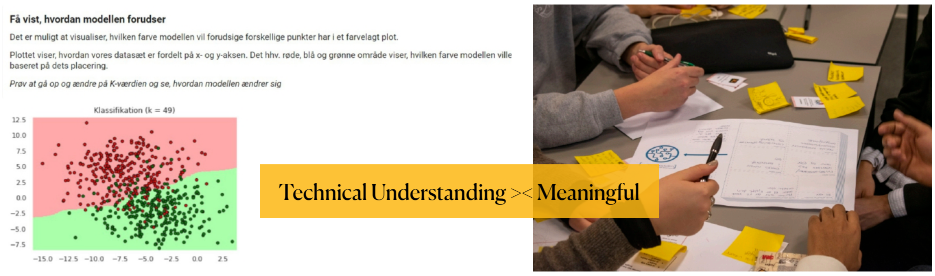
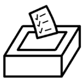


Figure 4.13: The tension between Technical Understanding and Meaningful exemplified with the first prototype of VotestratesML *(Paper A)*, and the ML Ethics Workshop *(Paper B)*. Too much emphasis on technical aspects of ML in the former led students to become demotivated while working with it.

of ML, but does not tie these to any specific use cases: the MLM might be used to replicate, e.g., unintentional bias by using different colored pens for each class. In contrast, VML contextualizes these discussions due to the subject matter of the tool. I argue, that the gap between a more general tangible tool and a specific context could be bridged by a teacher in the classroom. This, however, demands more from the teacher in terms of making such connections, and from students' abstraction skills.

Technical Understanding >< Meaningful: The tension between a focus on technical understanding and a contextual approach lies in how much each is allowed to fill in the tool designed (see Figure 4.13). If too much emphasis is put on the context, ML itself risks becoming a gateway to discussing other issues, and it becomes easy to resort to answers such as "ML is bad" in discussions, without the foundational understanding of ML required to justify why that might be the case. If too much emphasis, on the other hand, is put on technical understanding, students might be turned off by the tool immediately, because it becomes too difficult to see why ML is relevant to them. With the first prototype of VML we emphasized the technical aspects of ML too much, and students were demotivated by this, finding it difficult to relate to their own lives. In subsequent experiments, we instead emphasized this aspect, while dialing back requirements for technical understanding. Here, it is important to remember, that the goal of CE-based ML education is not necessarily to teach students how to implement their own ML systems from scratch, but rather to understand them on a level that supports their critical reflection of it.



Technical >< Critical: There also lies a tension in the interplay between designing for technical understanding and critical reflection (see Figure 4.14). Because the technical aspects of ML can be difficult to teach, and to understand, prioritizing them can often mean sacrificing emphasis on the critical parts. On the other hand, by completely de-emphasizing the technical aspects of ML, students' ability to engage with it, in a technically qualified way, also suffers.

This tension is a practical concern of how to cut the cake, so to say. Here, the most important thing for a designer to consider is the degree to which one-or-the other is dominant in the tool design, and to be intentional in deciding the focus.

Figure 4.14: The Technical >< Critical tension exemplified with ml-machine.org *(Paper H)* and ML Ethics Workshop *(Paper B)*. While I argue ml-machine.org can facilitate critical discussions on ML, it prioritizes scaffolding students technical understanding, where the Ethics Workshop deemphasizes this in, focusing instead on discussions.

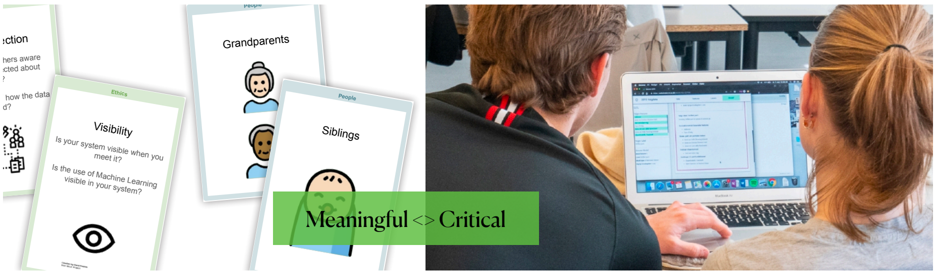


Figure 4.15: The synergy between Critical and Meaningful is exemplified in the ML Ethics Workshop *(Paper B)*, and VotestratesML *(Paper A)*. The former draws on students' own life to facilitate critical discussions, while the latter draws on their interest in politics.

This is not to say that both cannot be present in a tool, but rather, that depending on the context, time, etc., prioritizing one over the other might be necessary, and this should be done with care. Further, this tension is mitigated in prolonged engagements, as several tools and activities are combined to form a coherent curricular unit. Here, the challenge is instead to make a combination that allows for both deep engagements with the technical aspects and critical perspectives of ML.

4.3.2 Synergies

In addition to tensions between design features, there are also certain synergies, i.e., interactions between features that are beneficial to each feature. Some of these might seem banal, and to some extent, they are, however, I am not presenting these as novel, generalizable discoveries, but rather with interest in how they have materialized concretely in my design experiments.

Meaningful <> Critical: There is, in my work, a synergy between designing for meaningful and critical engagement (see Figure 4.15). A meaningful engagement here denotes that students can, somehow, relate to ML in a personal, social

Figure 4.16: The Meaningful and Low Floors synergy exemplified through the ML Ethics Workshop *(Paper B)*and ml-machine.org *(Paper H)*. Both engage students immediately in creating ML systems (the former non-functional, and the latter functional), while framing their experiments with ML as a design case.

or political way. It entails a positioning of ML as something familiar, rather than something foreign, and as something that can be influenced rather than something that is imposed. In this way, the possible effects of ML, both negative and positive, also move closer to students, and they can bring their own experiences and values into discussions about it.

Additionally, this positioning of ML as personally meaningful to students might make them more inclined to engage with it in a critical manner outside the classroom, which is one of the ultimate goals of CE for ML education.

MEANINGFUL <> LOW FLOORS: Related, meaningful engagement also synergizes with designing for low floors (see Figure 4.16). This synergy is straightforward in the sense that, making ML meaningful to students might also lower the barrier to entry for them. This is tied to Papert [171]'s idea of "mathophobia", that is when students come to identify as being bad at math to the point of it becoming an identify-trait of theirs. Providing meaningful engagement then, Papert argues, is a means for diluting mathophobia. Similarly for ML, some students simply do not see, why they should concern themselves with it (similarly to the students which Iversen, Smith, and Dindler [107] classify as "not-yet-motivated"). Providing ways of meaningful engagement, then, serves to motivate such students by showing them why ML does concern them, and offering them ways of engaging with it, in a manner that matters to them.

Figure 4.17: Both versions of the Machine Learning Machine (*Paper C* and *Paper G*) exemplifies the synergy between Tangible and Meaningful. Physically engaging with ML allows students to make sense of it in a social, and embodied way.

TANGIBLE <> MEANINGFUL: Providing tangible representations of ML and scaffolding meaningful engagement seems to synergize as well (see Figure 4.17). I argue this is because ML, as a technology, is often hidden away, integrated into other digital services, and interwoven with programmatic components. Making ML literally physical means moving it from an abstract concept to a concrete thing. Instead of being something to discuss in hypothetical terms, tangible ML can be scrutinized, it allows people to gather around it, inspect it from multiple angles, and have a shared, concrete reference in discussions.

4.4 Implications for Designers and Educators

The above design principles and the tensions between them offer an early conceptualization of the design space for computationally empowering learning tools and activities for ML and have implications for how to design such tools.

By presenting these tensions and synergies as an annotated portfolio, I aim to highlight the interplay between the artifacts, and experiments behind this dissertation. I will, again, stress that the goal here is not to develop some theory of the 'perfect' singular artifact, that satisfies all the principles presented in section 4.2. Rather, as argued by Gaver and Bowers [70], the portfolio aims to configure a "zone of potentially fertile possibilities", i.e., the design space for such tools. To this end, I suggest that other designers utilize this portfolio to inform their work when designing different learning tools and activities. Important here is considering the goals of a given design, how it can best be supported by the principles above, and what trade-offs need to be made. Here, the tensions and synergies offer an insight into which qualities of a design might conflict with others, and conversely which principles might harmonize well.

There are also implications for teachers putting together a curricular unit on ML containing a combination of different activities and tools that all serve to develop students' understanding of a certain curriculum. Here, I argue, the portfolio can be helpful when deciding on which tools and activities to structure such a unit around. Different educators might have different emphases, and different students have different needs, and as such, a one-size-fits-all solution is not appropriate. Instead, the portfolio acknowledges the complexity of the situation and aims to provide guidance and inspiration to teachers regarding how to design their curricular units. In practical terms, it offers a vocabulary for comparing the design of different tools and activities, not for judging which are good and which are bad, but rather with regard to how the design of different tools might complement each other.

4.5 Summary

RQ2: How do we design learning tools & activities that scaffold children's computational empowerment with regard to ML?

In this chapter, I have summarized my design process and presented several design experiments, emphasizing the qualities of each that were sustained in the following experiments. To address **RQ2**, I extracted a set of principles for designing computationally empowering learning tools and activities: Considering tangible representations of ML, providing low floors to engaging with ML, providing opportunities for meaningful engagement, scaffolding students' fundamental, technical understanding of ML, emphasizing critical engagement with ML, and

RQ2a: What characterizes the design space for such tools & activities?

designing for scalability. It is important to note that these are not criteria for a good tool or activity design, but should rather be seen as resources. Between these principles, I have identified a set of tensions and synergies and presented an annotated portfolio of how these manifest in my design experiments. This serves to conceptualize the design space for CE-based ML (**RQ2a**) as informed by my experimental work and to offer concrete guidance for designers and educators.

Chapter 5

Computational Empowerment for ML Education

From my projects' onset, it has been motivated and driven by a commitment to empowering students in their engagement with AI and ML by aiming to develop their understanding of ML concepts, practices, and potential impacts both on a social and a societal level. This framing originates, as mentioned in chapter 1, from the notion of Computational Empowerment (CE), as put forth by Iversen, Smith, and Dindler [107]. CE was developed as a general approach for developing children's digital technology comprehension, focusing on fostering their ability for critical reflection.

However, CE frames digital technology comprehension through digital fabrication methods and fablabs, which makes engaging directly with emerging, and complex technologies such as AI and ML difficult.

In this chapter, I discuss CE as it relates to ML and present a framework that extends CE to qualify it for addressing ML. I start by extending the original design-based process model proposed by Iversen, Smith, and Dindler to include a technological foundation that aims to support students throughout the learning process. Further, I discuss how the framework can be used to discuss CE-based learning goals for ML by combining it with the CE Progression Model *(Paper D)* to analyze VotestratesML *(Paper A)*.

5.1 A Technical Foundation for CE-based ML Education

CE proposes that students should learn to both code (in the broadest of meanings) and decode digital technologies and presents a design-based process model detailing the activities in such a learning process (see Figure 5.1). The model focuses on digital fabrication as the main way students engage directly with the design material; digital technology.

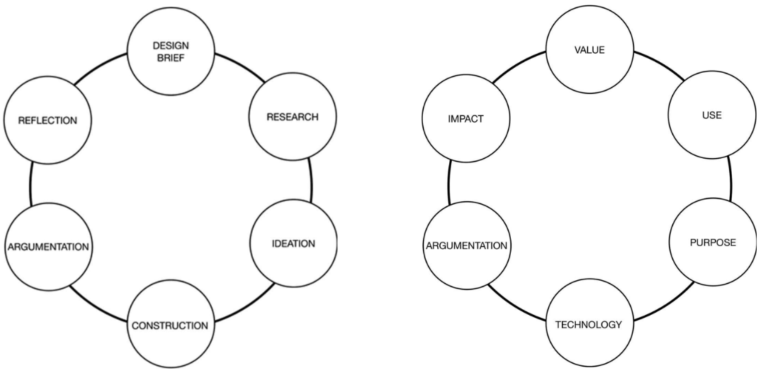


Figure 5.1: The activity model from Computational Empowerment as seen by Dindler, Smith, and Iversen [50]. The leftmost model describes constructive design activities, whereas the rightmost describes analytical activities. Included with permissions from the authors.

This approach, I argue, is insufficient when aimed at ML education. ML is, as discussed previously, technologically opaque: The ML-systems we encounter in search engines, on streaming sites, and social media are enormously complex and rarely lend themselves to any real kind of decoding, even less so coding (in the sense of constructing similarly complex models). While several tools exist that allow students to create working ML-models and -systems, few allow them to look behind the curtains, so to say, and examine how the technology works on a fundamental level, beyond the very basics. This is a hindrance to scaffolding students' critical reflection on ML since many of the pitfalls of ML are deeply rooted in its technological specificities. For example, how unintended biases occur or how particular optimization objectives might lead to a more accurate but less fair ML-model (for an in-depth discussion, see *Paper I*).

This idea is inspired by Verbeek [225] who argues that technology is normative because it prescribes certain actions, and as such, "designing should be regarded as a form of materializing morality." [225, p. 379]. Thus, engaging critically with technology should also entail engaging with its materiality since this is central to the actions the technology affords and, thus, its potential impacts.

Hence, engaging with ML or fabricating ML-systems without an understanding of their technological specificities limits the opportunities for students to engage critically with ML. This is not to suggest that discussions regarding ML and its use cannot be had with and by students without a technological foundation. However, as I experienced with VML *(Paper A)* and the ML Ethics Workshop *(Paper B)*, these discussions can quickly leave little room for nuance, with ML becoming either a magic spell to solve issues or evil incarnate. Such a binary understanding of ML (or any technology) works to shut down further discussions rather than open them up. ML is already everywhere, and unless we go (almost wholly) off the grid, we will have to live with it. There are, however, meaningful conversations to be had which cannot be had constructively based on a binary good/bad view of ML.

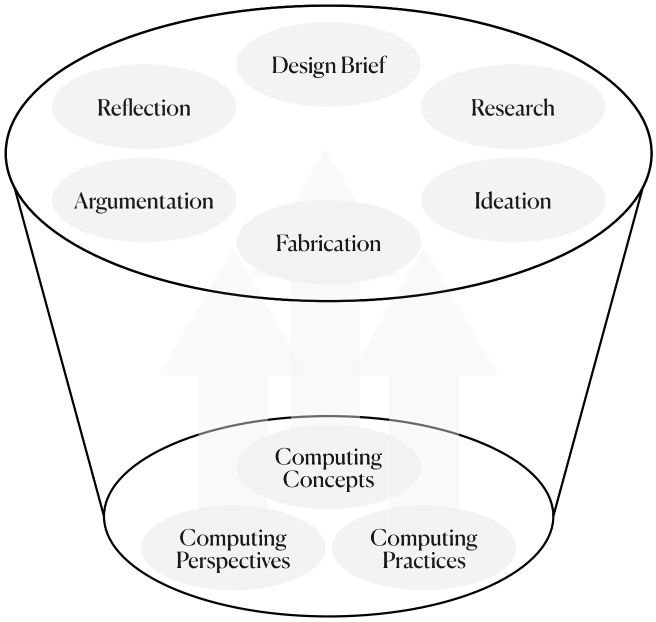


Figure 5.2: A model of the expansion of the CE activity model to include computing concepts, practices, and perspectives.

Instead, the nuances of ML lie in its specificities [56]: What are exemplary and representative examples of desired behavior? Which features of the data are relevant and which are not? With what objective is the model optimized, etc.? Teaching children to code and decode technology, thus, requires them to engage with these specificities on more than a surface level. To be clear, I am not suggesting that, e.g., a 7th-grade student should be able to implement the most appropriate loss function when training a ML-model. Instead, students should be able to recognize that models are created with different purposes, that ML developers make such choices, and that they have the potential to impact the system they are used in severely.

I suggest the addition of a technological foundation underpinning the design-based CE process model (the leftmost model in Figure 5.1), which can be seen in Figure 5.2. This foundation adds three central aspects of understanding a technology [33]: concepts, practices, and perspectives. These aspects were first presented by Brennan and Resnick [33] and originally referred to aspects of programming, concepts such as loops and parallelism, practices such as testing and debugging, and perspectives such as self-expression. However, as we argue in *Paper D*, these aspects cover a much broader range of digital technologies and ways of interacting with computation, including ML. The specific concepts, practices, and perspectives, however, depend on the subject in focus, and I discuss what I see these being in relation to ML below, in section 5.2.

Importantly, this foundation is just that; a foundation. It is not the point of CE-based ML education but rather a method for qualifying it. ML concepts, practices, and perspectives are taught to support students when ideating, arguing for a design, or reflecting and iterating on it. For each activity in the CE process model, the technological aspects have different implications. When doing a field study in the research step of the CE design activity model, this might mean paying close attention to what data should be collected for a potential ML-system, or if the context is sensitive to possible unintentional biases. When ideating, the foundation can provide a framework to work within (as in the ML Ethics Workshop *(Paper B)*). For fabrication, a technological foundation helps students construct functioning systems, and for argumentation, it can provide technical arguments for a systems' design. It can support students when reflecting on their system and its implications, and the design brief might be altered based on technological realities, possibilities, and limitations.

Finally, adding a technological foundation to CE is not limited to ML education but can be expanded to all complex, emerging technologies. For example, for Augmented Reality (AR), this might mean teaching how surfaces are scanned and modeled or how AR can be used for location tracking and behavior manipulation [138] (see section 7.5 for a discussion of this).

5.2 Dimensions of ML: Concepts, Practices, & Perspectives

Here, I explore further what, concretely, the technological and material foundation presented above means for how ML could be conceptualized for CE. To do so, I draw on the characterization of the dimensions of CE, as presented in *Paper D*based on the work of Brennan and Resnick [33].

5.2.1 A model for Progression in CE

With our CE Progression Model *(Paper D)*, we attempted to detail how progression towards becoming computationally empowered can be characterized. On the vertical axis, the model (as seen in Table 5.2) spans different dimensions of CE; concepts, practices, perspectives in one's own life, and perspectives in society. As argued in *Paper D*, "[t]hese categories allow for a more detailed classification of technical knowledge and skills than the original aspects of CE, as well as a finer-grained understanding of the attitudes and values associated with being computationally empowered and of taking a critical stance towards technology than allowed by the framework of Brennan and Resnick [33][1]".

Brennan and Resnick [33] define concepts, practices, and perspectives as follows: "*computational concepts* (the concepts designers employ as they program), *computational practices* (the practices designers develop as they program), and *computational perspectives* (the perspectives designers form about the world around them and about themselves)" (emphasis in original). We divided perspectives into two distinct aspects, perspectives in one's own life, and perspectives in society, to combine these aspects with Iversen, Smith, and Dindler [107]'s original definition of CE: "1) engaging creatively in technology development, 2) understanding the role of digital technology in society, and 3) reflectively and critically understanding the role of technology in one's own life". To expand on which computational dimensions are involved in engaging with technology development, we replaced

[1] I changed the citation style in the quote to match the format used in my dissertation.

Definition of CE	Dimensions of CT
Engaging creatively in technology development	Computational Concepts Computational Practices
Understanding the role of digital technology in society Reflectively and critically understanding the role of technology in one's own life	Computational Perspectives

Table 5.1: The relation between the definition of CE by Iversen, Smith, and Dindler [107] and the dimensions of CT as presented by Brennan and Resnick [33].

this aspect of CE with the concepts and practices from CT. On the other hand, to qualify computational perspectives from CT, this aspect has been expanded into two dimensions; perspectives in one's own life and perspectives in society. The relation between the definition of CE and the dimensions of CT can be seen in Table 5.1.

To allow designers and educators a finer categorization of progression, we merged the dimensions with the popular SOLO taxonomy [18]. The SOLO taxonomy divides learning into four competency levels, which can be used to create a common language when forming learning goals across classrooms, activities, or subjects. To do so, these competency levels are linear and hierarchical, and a set of verbs accompanies each level to help educators articulate what students at a certain level should be able to accomplish:

> "Progression can be defined as moving up in SOLO levels, from pre-structural, to uni-structural, multi-structural, relational, and up to extended abstract level as the highest level. The first two levels refer to developing surface knowing and the latter two levels refer to developing deeper knowing. Surface learning refers to studying without much reflecting on either purpose or strategy, learning many ideas without necessarily relating them and memorising facts and procedures routinely. Deep learning refers to seeking meaning, relating and extending ideas, looking for patterns and underlying principles, checking evidence and relating it to conclusions, examining arguments cautiously and critically, and becoming actively interested in course content [89]". (from *Paper D*, p. 159[2])

It is worth mentioning that the SOLO model has been criticized for being simplistic and for representing a reductionist view on learning [116, 236]. Nonetheless, SOLO is a powerful tool for designing learning goals and relating these to teaching activities [89].

Together, the dimensions of CE and the SOLO taxonomy form a model for characterizing progression in CE (see Table 5.2). The model can be used in several ways: First, it is aimed at educators planning CE projects, where the model might help ensure that a comprehensive set of learning goals (on the vertical axis) are created and that progress towards these goals (the horizontal axis) are considered

[2]Page number refers to the manuscript included in the dissertation.

	Uni-structural	Multi-structural	Relational	Extended Abstract	
Concepts					
Practices					
Perspectives in one's own life					
Perspectives in society					

Table 5.2: A model for progression in CE. From *Paper D*.

throughout the project. For researchers and designers working within a CE frame, the model might be consulted when exploring how a new computational subject could be taught or how to design tools and activities for such subjects.

5.2.2 Conceptualizing ML for CE

To develop an idea of what CE concepts, CE practices, and CE perspectives of ML are, we will disregard the model's horizontal (SOLO) dimension for now and focus solely on these.

Since CE focuses on critical engagement and reflection, I focus on this when determining the conceptualization of ML with regard to each dimension. As such, the inclusion of an ML component is determined by its potential impact on students' lives. In relation to the two perspectives-dimensions, these certainly overlap with ML concepts and practices, as these are what enable them. Still, here, I focus on "big ideas", i.e., the perspectives I see as most essential to understand. Further, I draw on related literature regarding ML curricula to qualify this conceptualization (specifically [137, 216]). An overview of my CE conceptualization of ML for CE can be seen in Figure 5.3.

ML Concepts: ML can be boiled down to three concepts central to understanding it on a level appropriate for CE; data, models, and predictions/output.

Without data, there would be no ML, essentially no Information Age, and several researchers have identified Data Literacy as an essential skill (e.g., [32, 63, 65]). In our work, we have addressed this notion by aiming to teach children to understand what data is, how it can be computationally processed, and how personal, private information can become data entries for a ML system (*Paper F*). Without an understanding of data, understanding ML is, simply put, not possible. Further, data (especially personal data) and its collection is a powerful force in our society and economy, and understanding the digitization of these, arguably, also requires understanding data.

The model is another essential concept in ML. Key here is the famous aphorism "all models are wrong, but some are useful". To understand ML, it is important to grasp that it only ever approximates an understanding of the world and that this

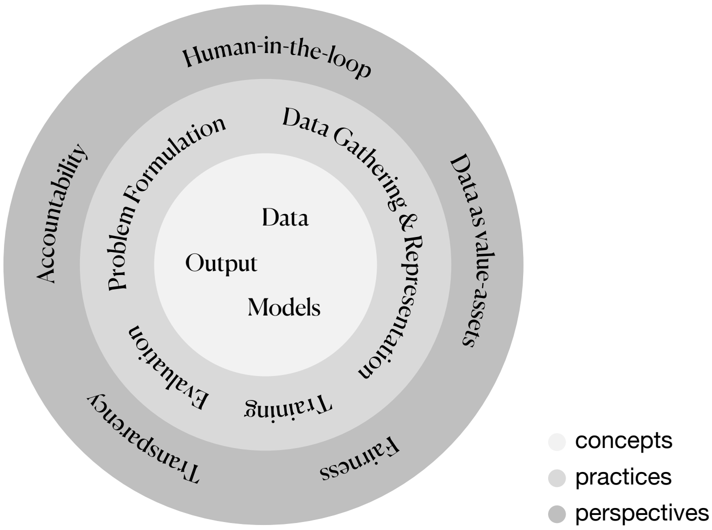


Figure 5.3: ML concepts, practices, and perspectives for CE-based ML education.

approximation is only as good as the model at the core of the ML system. By definition, and for better and worse, ML-systems hold partial and simplified views of the world. Thus, understanding models is the first step towards understanding the worldview embodied in such systems and the sort of problems ML can address.

Finally, understanding the ML system's output (often some sort of prediction) is important, especially what such an output is and how it can be presented. ML is, in essence, statistics, and its outputs are also statistical probabilities. However, their statistical nature is often removed in end-user applications and presented as, e.g., recommendations rather than the educated guesses they are. Understanding the nature of ML-outputs is again important for understanding what sort of issues are addressable by ML, but also for understanding and evaluating the qualities of the solutions they propose.

ML Practices: In *Paper C*, we present a simplified model for characterizing the process of creating ML models (see Figure 5.4). Here, creating ML models is divided into four separate steps; *problem formulation*, *Data Gathering and Representation*, *Training*, and *Evaluation*. The model, although simplified, corresponds to other models characterizing the ML process (e.g., [56]).

The first step in the model entails determining the goals of the model. This, essentially, is a design question that requires us to ask "what is the model meant to achieve?" [56]. This, in turn, has ramifications for what data we need to collect and what we do with it, the type of ML we will deploy, our success criteria, and how we evaluate our system. Understanding what problems can be solved with ML is the first step towards recognizing where ML might be employed in the real world (for good or bad). Further, it is part of understanding that ML models are not objective but that human intentions are key from the first step towards a ML model.

Next in the model is data gathering and representation. In this step of the

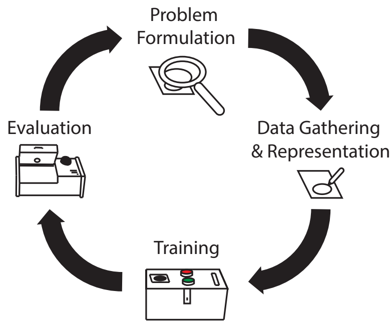


Figure 5.4: A simplified model of the ML process, from *Paper C*.

process, data is collected that is believed to address the problem articulated before. Further, the most appropriate representation of this data is determined[3]. Learners should understand this practice because it determines what the model might find important. Critical examples can be found if sensitive data is included, such as gender, skin color, age, etc. Understanding the role of data and their representation is key to understanding how models can become unfair.

Next is the training step. Here, we ask the model to make connections between the collected data to optimize the model regarding the goal we determined in step one. This conflates several separate steps [56], but its essence is that the model configures itself to find patterns in the data that allow it to make inferences about new data. Depending on the model type and its complexity, these patterns might be more or less understandable by humans. Understanding how training occurs, then, is important for understanding how and why models become opaque.

Finally, we evaluate our models. This might involve some technical assessment, such as finding the accuracy or precision of a model, but also contextual evaluations, such as to what degree the model actually serves the people or purpose it was made for. Understanding that models can, and should be judged by different metrics depending on their use case, is important for assessing their usefulness in different situations. It is also key for discussions regarding the accountability of ML models and how they are used.

ML Perspectives Finally, I argue for focusing on the following perspectives of ML; *the human-in-the-loop*, *data as value-assets*, *fairness*, *transparency*, and *accountability*. These are perspectives central to the implications of ML for our personal lives and society, and an understanding of them can help qualify discussions about the use of ML. Notice, that while the model for progression in CE divides this dimension into two, it is conflated here. However, the same ML perspectives have implications for both society and our own lives.

The human-in-the-loop refers to how people, i.e., designers and developers of ML-systems are always in some way or another involved directly in the forming of a ML-system. From deciding what data to collect, how it is represented in the system, how the model is trained, and what its measures of success are,

[3]For instance, if trying to distinguish between geometric shapes drawn by hand, these can be made black and white, since the model should disregard color.

human decisions are deeply involved in ML, despite it often seeming otherwise. Understanding this can support us in holding those responsible for ML-systems in the world accountable for their effects.

Next is the idea of data as value-assets. The collection of mass data-sets about individuals, when aggregated, becomes hugely valuable assets for companies because complex models can be built to identify people that share specific interests, values, etc. Related to this notion is the idea of "surveillance capitalism", aptly named by Zuboff [247], which describes how these giant companies make billions from social networking sites that are apparently free to use. But, of course, they aren't free, rather, users pay for them by providing valuable data instead of with money. Or, instead of customers, users of such sites (or rather their data) are the product, access to whom are being bought by advertisers, mediated by ML models (or, popularly, "The Algorithm").

Fairness relates to how biases work in ML systems. The patterns ML-algorithms look for when training a model is precisely bias, i.e., the inciting incidents in a data-entry that sets it apart from some data-entries and likens it to other. Without these biases, ML-models would have no predictive capabilities. However, an issue arises when models pick up on biases we did not intend. Data-sets are only ever as good or fair as the people creating them and might inherit these people's implicit (or explicit) biases. When dealing with people, we expect such biases to occur; we know that we have them, and often systems are put in place to mitigate them (such as having multiple judges on a criminal case or having several keys to arm nuclear devices). However, the biases can disappear into a complex nest of artificial neurons due to the opaqueness of ML models. Thus, fairness relates to how data is collected and represented, what model type is used in the system, how auditable it is, and what the intended use case of the system is.

Transparency regards how ML is integrated into systems and how easily the model is audited. As I have described above, ML is often integrated into systems such that it is difficult to know that it is automating decisions (such as which commercials to show us). This is pertinent to social networking sites, but transparency is a factor in all ML systems: How transparent for doctors is it that the system providing recommendations for patient treatment is based on ML? How transparent is it for the patient? Can doctors dive behind the recommendation and audit the ML-model or the data it is based on?

Finally, accountability centers on how companies and providers of ML systems can be held accountable by their users and to which degree. In her book, Weapons of Math Destruction O'Neil [164] describes an American teacher being fired due to an automated teacher-evaluation system. When investigating further, she is told by the school that they have complete trust in the system they have purchased. The company behind the system refuses to tell her how it made its decisions since they considered it a company secret (it was, after all, not a human decision based on some publicly available evaluation metric, but a decision made by proprietary software that could be copied). In this case, there was no accountability embedded in the system, and as such, no way for the teacher to contest its decision. If (or probably rather when) ML systems are embedded further into society, especially by government authorities, holding them accountable is essential. Understanding how easily accountability can be disregarded in ML-system then is important to qualify discussions about its use.

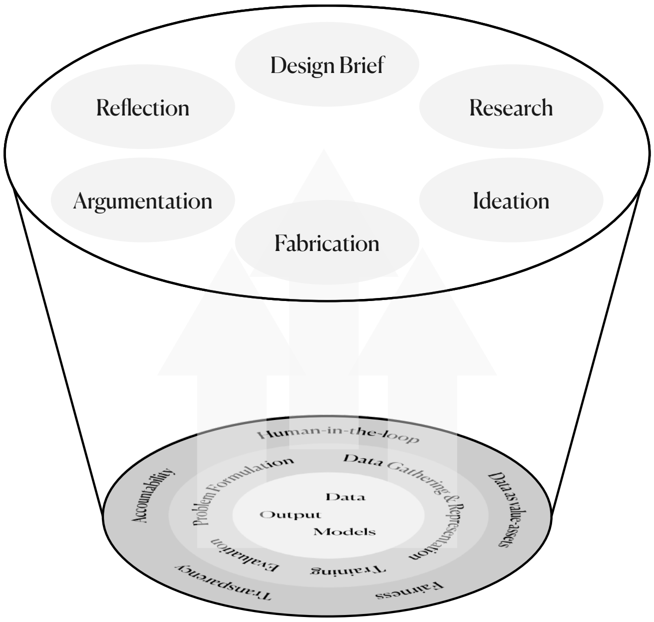
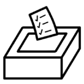


Figure 5.5: The CEML framework with concepts, practices, and perspectives from ML based on CE.

5.3 Applying the CEML Framework



In this chapter, I have expanded on the design model by Dindler, Smith, and Iversen [50] describing CE activities with an emphasis on ML fundamentals into the CEML framework (see Figure 5.5). This framework allows designers and educators to draw on CE principles and goals when teaching ML. CEML can be combined with Table 5.2 to help determine or analyze learning objectives in a CE-based ML curriculum and provide guidance for designing or using CE-based learning tools and activities for ML in the classroom. To look at an existing example, I will demonstrate using VotestratesML *(Paper A)* below (see Table 5.3).

As seen in Table 5.3, the model for progression in CE from *Paper D* can be used to classify its learning goals. It is important to note that the goals in Table 5.3 are directly related to ML fundamentals (the lower half of the CEML framework) and only indirectly to its broader context of use. We can use the top half of the CEML framework (i.e., the activity model from CE) to address these issues. Thus, we might ask students to consider how the design brief (e.g., in VML *(Paper A)*; design a system to help the social democratic party identify potential voters) is value-laden and might influence the human decisions that are part of creating a ML-system. During a reflective activity, we can then ask students to consider how the design brief influenced them if they can identify what potential impacts this might have on the system in use, and then ask them to iterate on the model they have created, drawing on these new insights.

With this exercise, I wish to convey how the technological specificities of ML

VotestratesML	Unistructural	Multistructural	Relational	Extended Abstract
Concepts Data, Models, Output	Identify features and labels in a ML model	Combine features, and a label to create a functional ML model	Relate the output of a model to social studies theories	Hypothesize and discuss the output of a ML model in relation to social studies theories
Practices Problem Formulation, Data Gathering & Representation, Training, and Evaluation	Identify which problems can be solved with ML	Describe the characteristics of a ML problem	Evaluate and compare different ML models	Hypothesize on how to improve a ML model
Perspectives in one's own life Human-in-the-loop, data as value-assets, fairness, transparency, and accountability	Identify what makes mundane information useful for a ML system			
Perspectives in society Human-in-the-loop, data as value-assets, fairness, transparency, and accountability	Identify value-driven decisions in a ML system	Describe how human-held values can influence a ML system	Criticize human-made decisions in a ML system	

Table 5.3: The learning goals of VML analyzed through the CEML concepts, practices, and perspectives

are essential when discussing its potential impacts and how the CEML framework can scaffold such discussions. Designers and educators should consider what dimensions of ML to expose to students through new learning tools and activities and how these might be addressed in different CE activities.

5.4 Summary

This chapter addresses **RQ1**, by proposing an expansion of CE [50, 107] for teaching ML. I argued that to qualify the reflections that CE emphasizes, a student needs a technological foundation because of the complexity of the technology and how it is nested in opaque ways in everyday systems. I proposed that aside from the design activity model suggested by Dindler, Smith, and Iversen [50], students should be taught about ML concepts, practices, and perspectives, and have highlighted how these interact with the elements of the CE activity model. Further, I have argued that while I find a technological foundation crucial for a CE approach to ML, it should never overshadow CE's focus on reflection. Next, I have proposed a conceptualization of ML concepts as belonging to the three categories above. I propose focusing on data, output, and models for key concepts. For practices, I propose emphasizing problem formulation, data gathering and representation, training, and evaluation. For perspectives, I propose an emphasis on the human-in-the-loop, data as value-assets, fairness, transparency, and accountability. Below, in section 7.5, I discuss the usefulness of the CEML framework for informing education in other emerging technologies, where a qualified discussion benefits from a technological foundation.

RQ1: How can CE as an approach to ML education address the challenges posed by the widespread use of ML?

Chapter 6

From Education Towards Remarkable AI Everywhere

Throughout our classroom interventions, a theme emerged in how students reflected on AI and its role in their lives; "we don't care about the things we share". While students were generally engaged in creating with and discussing AI, they were unwilling to change their behavior regarding it, even as it became clear to them how their data is aggregated and used by, e.g., social networking sites. To be clear, it was never our intention to turn students away from such sites but rather to support them in seriously and critically examining their use of them with regard to how the sites deploy AI and for what purpose. However, we learned just how important social networking is to adolescents and what lengths they are willing to go to maintain it. This finding is further corroborated in the psychology literature (e.g., [132, 163, 165, 240]), which points to FOMO (the Fear Of Missing Out), as a potential contributor to adolescents' addiction to social networking sites.

In any case, it became clear that a concentrated focus on efforts in education contexts would not suffice. Although CE includes political work to further its educational aims, ensuring that children become knowledgeable regarding AI is not enough. Even if they develop the right skills, competencies, and attitudes required to question and debate AI and its use, they are still, ultimately, dependent on proprietary, and closed-off systems, such as social networking sites, for their social lives to continue as before. And such sites, and the companies running them, have no incentives to alter how AI is used in them. Moreover, they are hugely profitable, especially due to their collection of massive data-sets about their users and their analysis and selling off of these. Instead, we realized that in addition to education, work is needed that convincingly supports the designers of such systems to include methods for interrogating and discussing their use of AI.

In this section, I present Remarkable AI as a design approach for doing so, based on *Paper E*[1].

[1]In the following, these sections are based on *Paper E*: section 6.1 is adapted from sections 1 and 2 in the paper, with several paragraphs removed for brevity and minor edits for clarity. Similar edits have been made to section 6.2, which has been adapted from the paper's section 3. Section 6.3 was inspired by the paper's section 4, some text reappears from the paper, but the analysis was remade and rewritten for this dissertation. Section 6.4 has been adapted from

6.1 Framing Remarkable AI

In HCI, there has been a long-standing debate about dichotomies and continua regarding the extent to which technologies' inner workings and values should be exposed to users. One of these discussions surrounded the notion of Remarkable [178] and Unremarkable [215] Computing. As an approach to ubiquitous computing in the home, Unremarkable Computing, as coined by Tolmie et al. [215, p. 404], aimed "to make computational resources that can be unremarkably embedded into routines and augment action", referencing Weiser's [229] seminal view on *the computer for the 21st century.* In response, Petersen [178] suggested *Remarkable Computing* as a complementary approach, arguing that as much as digital technologies can support people's routines, they are also increasingly becoming "objects of lifestyle and identity" [178, p. 1446], and that such objects are and should be remarkable.

In 2019, Yang, Steinfeld, and Zimmerman [238] presented an argument similar to that of Tolmie et al., namely the notion of *Unremarkable AI*; an approach to designing AI systems, specifically for supporting clinical decision making, that aims for them to be "subservient to the day-to-day decision-making" of their users. They motivate this by the failure of other clinical decision support tools (DST) to be incorporated into doctors' daily routines successfully.

AI and ubiquitous computing are analogous in that they share a central tenet, namely the "interest in building technologies that make sense and respond in a sensible way to the complex dynamics of human environments" [133], and (at least in regards to machine learning) a dependence on data collection and analysis. However, this also leaves Unremarkable AI subject to similar critiques to those delivered by Petersen to Unremarkable Computing.

We see Remarkable AI as complementary to Unremarkable AI [238]. Remarkable AI is appropriate in contexts where reflection on the use of AI is necessary and where Unremarkable AI might lead to skewed power dynamics between stakeholders or loss of agency of users. However, Unremarkable AI might not always lead to this (and Remarkable AI is deemed to bring its own issues in some instances).

To inform Remarkable AI, we look to approaches to designing for reflexivity. Sengers et al. [196]'s Reflective Design is an influential approach, drawing on several other traditions within HCI and Design such as critical and speculative design [4, 54], and reflection-in-action [194], to present a comprehensive approach to designing technologies for critical reflection. They argue that "for those concerned about the social implications of the technologies we build, **reflection itself should be a core technology design outcome for HCI**" (emphasis in the original). Furthermore, they propose that designing for critical reflection entails exposing otherwise hidden aspects and making these "available for conscious choice".

We agree with the goals of Reflective Design but argue that its scope needs to be broadened for dealing with emerging technologies such as AI; since AI can be too complex for end-users to grasp immediately, providing opportunities for conscious choice is not enough. In addition, a reflexive approach to designing

the paper's section 5, with minor edits based (among other things) on the remaking of the analysis. Finally, section 6.5 is based on section 7 of the paper and has been edited for clarity and expanded upon.

AI systems should include opportunities for users to inspect and learn how the system's AI-component operates and what effect it has on the system, themselves, and their surroundings, and then allow users to make conscious choices regarding their interaction with the system.

6.2 Remarkable AI

Remarkable AI aims to make the AI component of such systems present for their users to expose and explain their inner workings and empower users to interrogate and negotiate the role of AI in these systems and its effect on their lives. We find it imperative that designers of AI-enabled systems heed the calls to consider aspects of fairness, accountability, and transparency in such systems and design their systems with an appropriate level of remarkableness at the right time.

To do so, the AI component of such systems must first be made perceivable to users since things that we cannot perceive cannot become present in our lives. In Unremarkable *Computing*, this is addressed through routines; "artifacts that are implicated in routines can be perceptually available yet practically invisible in use" [215]. In other words, in Unremarkable Computing, the apparent perceptibility of an artifact does not matter; what is important is its perceptibility *in use*. Presence in Unremarkable AI [238] plays a somewhat more sophisticated role. Yang, Steinfeld, and Zimmerman [238] argue that while Unremarkable AI should not usually interfere with users' routines, it should be "present enough to slow decision-making down" when the prediction is at odds with the users' assumptions.

In response to Unremarkable Computing, Petersen [178] points out that no interactive technology is inherently invisible in use; it must first be appropriated and learned before it can fade into the background. However, as the proliferation of AI systems has shown us, that can no longer be said to be true. Indeed, the front-end of an interactive system is not inherently invisible, but the back-end might be, and this is where AI is most often situated. Unremarkable AI highlights that a prediction is being made [238, Fig. 1], but its aim of not interfering with users' routines means that this might be ignored. Remarkable AI complements this approach by aiming for reflection and empowerment.

I argue that designers of AI systems should design their products in ways that expose and explain their AI components and allow users to interrogate and negotiate the role of AI in the system and its effects. Below, I expand on how Remarkable AI aims to support designers in doing so by looking to research in computing education on teaching AI since this community has produced several examples of exposing and explaining AI systems through use. These approaches are presented in the context of learning, but as I will show, they contain important lessons to be learned for designing end-user systems. Additionally, xAI provides explanations about how AI systems produce predictions and is an aspect to consider when designing a Remarkable AI system. Still, as argued above, it often only covers part of the process and thus is not itself enough.

In the following, I present two high-level design sensitivities derived from the literature and argue why designers should pay attention to these when designing AI systems to be remarkable.

6.2.1 Learning through Use: Exposing & Explaining AI

In Petersen [178]'s Remarkable Computing, a central aspect that sets it apart from its Unremarkable counterpart is the notion of Learning through use; that appropriating and learning to use technology is an inherent trait that should be designed for. Remarkable AI extends this notion and argues that to understand the potential impacts of an AI system, some level of understanding of its inner components is necessary and that this should be designed for by exposing and offering explanations of these components through use.

The notion of *coding* from CE [50, 107] is beneficial here. For example, consider how we typically interact with a search engine: We enter a search query (most likely in our web browser's search bar, i.e., not on the website of the search engine), scroll through the results of the query and click any link that sparks our interest. Suppose we instead try to code (in the broadest of meanings) our own. In that case, we are forced to pay attention to how they work behind the scenes, how queries are interpreted, how websites are crawled and indexed to become available for searching, and crucially, how results are ordered using AI and which data are collected about the users. Coding then makes us look at and understand technology in a new light.

This is not to suggest that AI systems should require users to code them from scratch. Still, if aiming to be remarkable, they should provide users with opportunities to dive below their front-end, change and customize system parameters, and observe how this affects the outcome in a safe environment. This approach is similar to the notion of *infrastructuring* [180] but focuses on exposition and explanation rather than appropriation. Exposition and explanation are, however, not enough since it does not necessarily provoke reflection on the broader implications of a system's use of AI.

6.2.2 Interrogating & Negotiating the Role of AI

Therefore, the second aspect of a Remarkable AI system is providing opportunities for users to interrogate and negotiate the role AI plays in the system and how this affects its stakeholders and context of use.

Again, CE provides a helpful notion, namely *decoding*. Here, technologies are analyzed in depth with regard to their purpose, the values they imbue, and their potential and current impact. Drawing on this, we suggest that systems designers should incorporate ways of decoding their use of AI with regard to, e.g., what values guide how data was collected, analyzed, and used in the system, what its criteria of success are, but also which stakeholders are involved/implicated in which aspects of the system, who has control over how the system's predictions are used, etc.

We are also inspired by Reflective Design [196], which provides several design principles for fostering reflections in end-users. First, it argues that reflective designs should support users in reflecting on their own lives by, e.g., "offering users new ways of experiencing and reflecting on their activities" and "to offer up new choices that may not have been in the user's awareness". Further, they argue that "technology should support skepticism about and reinterpretation of its own working". Users should be the ultimate authority on the activities they are performing with the system. Systems should be designed to encourage reflection on this and allow users to reject using the system if they disagree with its values and purposes.

Finally, Remarkable AI argues that users should be involved in negotiating the values and power dynamics built into the system. Here we look to PD for its efforts of doing so, more broadly, in the design of interactive technologies. In their paper on the future of PD, Bødker and Kyng [30] called for visions of approaches to "escape the iron grip of the big corporations" with regard to their, now all the more so, ubiquitous use of data and AI with no opportunity for users to intervene. Remarkable AI is an attempt to provide designers with an approach for developing AI systems that utilize AI's numerous promises while aiming to avoid its many pitfalls by, among other things, drawing on PD methods for inviting users into the engine room and negotiating how the use of AI plays out.

Remarkable AI aims to guide designers to design systems that expose and explain their AI components and invite users to interrogate and negotiate AI's role in their lives. It proposes doing so by allowing users to dive below a system's front-end to alter and experiment with its configuration and, ultimately, the results of its AI component and by offering up opportunities for users to reflect and act on the values, stakeholders, and purposes of the system.

6.3 Case Studies: Informing Remarkable AI

In the previous sections, I motivated the need to consider Remarkable AI as an approach to AI systems design. In this section, I will analyze three studies from my research to qualify and demonstrate how Remarkable AI can be designed for.

It is important to remember here that Remarkable AI springs from a drift in the research context; where the tools and activity that this analysis is based on stem from Education, Remarkable AI addresses the design of end-user AI-systems. Indeed it was our experiences from these projects that led to our drift towards end-user design, and in turn, the cases share similar goals with Remarkable AI, although their designs are particular to Education.

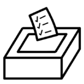
I chose this method as an experimental way of qualifying the problem space of designing for Remarkable AI. By drawing on the different-yet-similar problem-space of AI Education, I highlight the similarities and differences between the contexts to offer a delimitation of Remarkable AI. I propose concrete design strategies to designers when dealing with the remarkableness of their AI systems.

To analyze the cases, I used the two design sensitivities above as themes, deductively analyzing the cases with regard to how they address these. In practice, I reviewed the publication for each case (i.e., *Paper A*, *Paper C*, and *Paper B*), coding instances of each sensitivity as presented in the paper. Additionally, having been intimately involved in the design and evaluation of these cases, I noted down any relevant information not captured in the papers. Based on this analysis, I then compiled a set of design strategies, which were then discussed and refined in collaboration with two other authors of the paper draft (*Paper E*), who were also involved in the research projects.

Case	**Participants**	**Setting**	**Duration**
VML	61	High school classroom	3 x 90 min
MLM	2*	Lab setting	60 min
Ethics Workshop	71	Various	90 - 180 min

Table 6.1: A brief overview of the evaluations for the three included cases. More info can be found in their respective papers (VML *(Paper A)*, MLM *(Paper C)*, ML Ethics Workshop *(Paper B)*). *) Two adolescents participated in the pilot study in MLM *(Paper C)*, but approximately 30 students participated in subsequent unpublished interventions.

6.3.1 VotestratesML



As a short reminder to the reader: VML *(Paper A)* is a collaborative learning tool for exploring ML and its use in democratic elections for use in a social studies classroom. Students collaborate on building ML models for predicting voter behavior using data collected about the 2015 Danish general election.

Exposing & Explaining AI: VML has several features designed to both expose and explain AI concepts, practices, and perspectives. Students are free to explore the interface and functionality of VML, and the interface is designed so that it cannot be brought into a faulty state.

VML was designed as a political tool rather than a computational tool. It is contextualized for the social studies classroom, and the ML models that can be built by it are based on data from the Danish general election. By drawing on existing theories regarding voter behavior, students can use their subject-matter knowledge to reflect on the data and the expected usefulness of their models. Further, by allowing students to choose between different ML algorithms (the somewhat simple K-nearest neighbors and the more complex feed-forward neural network), VML highlights how different algorithms can lead to varying levels of model complexity and different results. VML is designed with two ways of evaluating a model; first, students get an accuracy score and an F1 score[2] for each model they create. Following this, they can try out their models on personas based on aggregate voter data for each political party. In this way, students are presented with a formal evaluation of their model and a more contextual evaluation, and the differences between them can be discussed.

Interrogating & Negotiating the Role of AI: In addition to supporting students in understanding what ML is and can do, they can also draw on their understanding of social studies to contest its use in democratic elections. We saw this in the evaluations with VML when a student exclaimed about the ethics of using ML in political campaigns, e.g.: "The Social Democrats do vote-seeking [in the role-playing exercise]. But I don't have much respect for that. I would rather go for policy seeking..." (*Paper A*, p. 105[3]). Here, the student has recognized

[2]The harmonic mean of the model's precision and recall scores. See https://en.wikipedia.org/wiki/F-score for an in-depth explanation.

[3]Page number refers to the manuscript included in the dissertation.

that one of the use cases of a system like VML is to identify, what subjects are important to their potential voters, and that a political party might use this knowledge to choose their defining issues. Further, the competitive process in VML of continually improving and optimizing models also spurred reflections on the world-view inherent in ML-systems: "Machines only look at what we ask them to look at. If they have made a model, it is we who decide what they should be looking for. Humans are better at marginalizing than machines are. If we do not make them marginalize groups, they won't do it!" (*Paper A*, p. 105[4]). In both these cases, we observe how VML was successful in scaffolding reflections on the use of AI. In the first example, these reflections are tied to the context of the tool, whereas the second is a broader, more general reflection.

To Remarkable AI, VML is significant in how it contextualizes ML as an integral tool in the political reality rather than as a neutral, objective technology. This results in a system that can be discussed, not only on technological terms but also contextually, with regard to the impact it might have on political campaigns and policymaking, with the discussion grounded in the technological reality of ML.

6.3.2 The Machine Learning Machine

The Machine Learning Machine *(Paper C)* was designed to make every step of the (simplified) process of creating a ML model explicit and *tangible*. First, students draw sketches and "feed" them to a *Trainer*-machine that digitizes the sketch and trains a model. After training a model, students can move their physical model artifact to the *Evaluator* that lets them use their model to recognize newly drawn sketches. The small number of participants in the pilot study makes it impossible to draw generalizable conclusions, but the MLM embodies several design intentions with repercussions for Remarkable AI.

Exposing & Explaining AI: By moving data creation, analysis, training, and evaluation into the physical world, students are invited to explore the process and make on-the-fly hypotheses about the system they are creating. This was especially regarding the role of data in ML systems. In our pilot study, we saw that participants actively used the physicality of their data-set by laying it out on a table and investigating it, forming hypotheses about their model based on the data, and experimenting with these hypotheses by creating new data.

Further, the visual aesthetic of the devices was designed to elicit a reaction corresponding to each device's role in the ML process: The Trainer appears as a dirty and used industrial device, representing the messy and iterative nature of creating ML models. The Evaluator's modern and clinical appearance is meant to signify how ML models often appear static once they are presented to end-users. Finally, the design of the model artifact, especially how it is moved between devices, represents how training and evaluating models are separate steps in the process, where the data is (most often) left behind, and only an abstraction of it is carried over.

[4]Page number refers to the manuscript included in the dissertation.

Interrogating & Negotiating the Role of AI: While these were not tested in the pilot study, *Paper C* presents two scenarios for how the implications of ML-systems in-the-wild, can be explored using the MLM.

The first scenario explores how insufficient data representativeness can lead to the model giving wrong or inconclusive results:

> "When comparing the dataset with the doodles, the model struggles to categorize, [students] realize that almost all cows in the dataset are looking to the right, while the new doodles illustrate cows that are looking to the left. The students draw a bunch of new doodles of cows looking to the left, and re-train the model. When they plug the model back into the Evaluator, it performs significantly better at recognizing cows looking to the left. ". (Paper C, p. 146[5])

Further, the system might be used to illustrate how unintended biases might be introduced to a ML-system:

> "Through further experimentation and discussions with their teacher, they realise that their model is very accurate at recognising houses if they are drawn by the same person who produced the data-set and with the same pen. If this person, instead, draws a bicycle it more often categorises it as a house than as a bicycle. They have unintentionally biased their model to assume that the type of pen and the artist are the most important attributes for distinguishing the two types of drawings." (Paper C, p. 146[6])

In this way, MLM provides a simple yet powerful way of introducing students to the technicalities of some of the (negative) ways ML-systems might impact their lives.

With regard to Remarkable AI, of particular importance is how students can experiment with and form hypotheses about ML in a safe sandbox environment. Additionally, the MLM allows students to experience how potential negative impacts of a ML system are rooted in its technological materiality if not carefully accounted for and to explore ways of mitigating these.

6.3.3 ML Ethics Workshop

Finally, I look at the ML Ethics Workshop *(Paper B)* designed to engage students without prior knowledge of ML in discussions on concrete ethical issues by scaffolding them in designing a pseudo-functional ML-system using different types of cards and templates to guide their design. This activity is initiated by a design brief provided by the teacher, e.g., to design an AI-enhanced mobile app to combat loneliness in adolescents. Next, students are handed a set of data cards that they are asked to discuss. These cards describe different categories of data that might be used in a ML model, such as "health" or "smart home". After openly discussing the data cards, students ideate on the design brief and pick an idea to work on. Subsequently, students describe their envisioned system design using a paper template, selecting relevant data cards and describing specific instances in

[5] Page number refers to the included manuscript in the dissertation
[6] Page number refers to the included manuscript in the dissertation

the broader categories. Once they have described a system, students are handed a single ethics card chosen by their teacher that describes an ethical issue relevant to their idea, e.g., "Privacy — is the system violating the privacy of you and your users? How?". Next, they are asked to discuss and *redesign* their system based on concrete issues they identify through their discussion of the ethics card. Finally, students present their system and discussion to each other, and a shared conversation is had in the classroom for each project.

Exposing & Explaining AI: This workshop differs from the previous two activities as students do not create a functional ML-system, but instead sketch one using the paper template. However, this was designed to ensure that the hypothetical system would be implementable. Additionally, the activity includes an introduction to the basics of ML. As such, the activity did not explicitly focus on exposing or explaining AI as a technology, but rather the introduction and constraints set up by the workshop served to bootstrap their discussions and reflections.

Interrogating & Negotiating the Role of AI: The card-based workshop includes two decks of cards, and activities with them, that aim to support this goal. The first deck, the *people cards*, aims to scaffold students' reflections on how different stakeholders are involved in any ML system. The *ethics cards* each present an ethical dilemma to consider regarding students' designs. The latter deck was deliberately introduced after students' had designed the first version of their application because we aimed for them to feel personally responsible for their systems' potential failures. Rather than students attempting to avoid these issues in the initial design phase, they were asked to redesign their application with the ethical dilemmas in mind after this phase. While challenging, and sometimes impossible, for the students to successfully solve, these dilemmas led to insightful reflections, such as in the following exchange (from *Paper B*):

> **Student 1:** "What if a group, where something is going on under the radar is added to the system by us [...] and something bad happens in that club, because we didn't know there was a problem with them. Whose fault is it then?"
>
> **Student 2:** "It is still the responsibility of the club."
>
> **Student 1:** "But we brought them into our system..."
>
> **Students 3:** "I think those who are responsible for the app are to blame. And I'm thinking that there is always someone in charge. That's a director's job."
>
> **Student 4:** "Okay, that is the Minister of Health. So it should be the job of the Minister of Health or what?" (Paper B, p. 127[7])

Here, we see students dealing with a specific dilemma, that of responsibility for failures in the system. We see that it is difficult for them to place the blame for such failure. In this case, where a government body is the provider of the system, this might be a clearer-cut case than with private providers, which might not be expected to take responsibility.

[7]Page number refers to the included manuscript in the dissertation

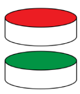
The case itself is another aspect of how the workshop allows students to interrogate the role of AI in their lives. By providing a design case where students essentially design for themselves (i.e., other teens), we invite them to imagine themselves as the application's end-users. The aim here is to invoke a sense of empathy for the potential users when discussing the impacts of the system students create.

In terms of designing for Remarkable AI, of notice here is how the workshop can engage participants in interrogating and negotiating ethical dilemmas in an AI system, even though they have little to no prior knowledge of AI or ML. Additionally, the workshop aims to create an emphatic reaction in students with the intended users of their applications, both in terms of the value it might bring them and what pitfalls to avoid and address.

6.4 Design Strategies for Remarkable AI

Above, we have presented two design sensitivities for Remarkable AI, namely designing to expose and explain AI and to interrogate and negotiate its role in our lives and in society. While the three cases have been designed for educational purposes, we have found three design strategies for designing end-user systems that can be extracted from them. Where the design sensitivities are high-level sensitivities to discuss and consider in a design process, these design strategies are concrete suggestions for designing for the sensitivities. Note that while the three strategies presented below are what we have been able to extract from our case studies, different and complementary strategies might be extracted from looking at other work.

6.4.1 Foregrounding the ML Process to Expose AI Components



The first strategy deployed across these systems is to foreground the ML process in a structured way, exposing different components that go into creating an AI system. With the MLM, students create their own data, manually load it into the model and then manually test the resulting output. VML opts instead to use fixed real-world data but allows students to explore the data, select what data they want to include, and further experiment with different model types and their configurations. While students do not create a functional ML system in the ethics workshop, they are asked to consider what data is relevant to their system and which predictions or inferences it makes.

By foregrounding the ML process to users of AI systems, they first and foremost are made aware that a computer is making decisions that impact their use of it or the decisions they make based on their interactions with it. While they might not understand each step in the process, it is at least apparent to them. We see this as the first step towards users being able to call the system and its use into question.

6.4.2 Iterative & Safe Experimentation to Explain AI

To aid students in understanding AI, several of the cases above allow experimentation without the risk of getting stuck. For example, VotestratesML is designed to enable students to iteratively try any combination of data, model type, and model parameters. The interface adapts to provide feedback to students regarding which choices are valid given their current configuration. Similarly, MLM allows users to iterate on their model to improve it. Here, users can take advantage of the physical data points by spreading them out and looking at their similarities and differences to make hypotheses about what inferences the model is making, and then test their hypotheses by strategically making new data-points and testing or training their data on these.

For end-user systems, allowing users to experiment with their AI components might not always be appropriate. Still, systems could enable users to do so in a sandbox environment without affecting the existing system. This would give users a sense of what happens to (their) data in the system and how different choices made by its developers or maintainers affect them. This functionality could be further expanded by including xAI functionality, although this was not the case in the above systems.

6.4.3 Facilitating Discussions and Reconfiguration to allow Interrogation and Negotiation

Finally, to scaffold students in negotiating the use of AI systems and their effect on their lives, the systems invite discussion and reconfiguration. An integral part of the Ethics workshop and VML is that students' systems and ML models are scrutinized and discussed by a wider audience. Additionally, in the Ethics workshop, students are explicitly asked to redesign their system to address the issues they have encountered. While this is often difficult for the students, it highlights the trade-offs between, e.g., privacy and effectiveness made by real-world designers and developers of AI systems.

While a complete redesign of an end-user system by the user might not be desirable, systems could be designed to reflexively engage their users in configuring the systems to their preference. This might entail allowing users to choose between privacy-preserving data collection or increased accuracy of predictions or letting users choose between advanced but opaque or simple but explainable models. Further, this strategy could be deployed to design systems where multiple stakeholders are affected by their use of AI to help them negotiate how it's deployed. By highlighting how these systems' use of AI interweaves with their users' daily lives and surroundings, users are supported in interrogating the systems' roles in these.

6.5 Discussion

As I see it, Remarkable AI and Unremarkable AI [238] are two ends of a continuum, delimiting the design space for remarkableness in AI-enabled applications, which come with each their strengths and weaknesses. To introduce the idea of this continuum, I have drawn on previous work in HCI regarding Remarkable [178] and Unremarkable Computing [215]. I did so to highlight how contrasting approaches can foster constructive discussions in the HCI community, and since the criticisms of Unremarkable Computing by Petersen [178] are as applicable to Unremarkable AI [238]. However, as mentioned earlier, because AI works in the background and a user might never realize that they are using an AI-enabled system, Remarkable AI is further reaching than its Computing counterpart.

Below, I discuss the necessity for considering remarkableness in everyday AI-enabled applications, how to find the right level of remarkableness, and the limitations of Remarkable AI.

6.5.1 Designing a Right Level of Remarkableness

As mentioned above, I fully recognize that Remarkable AI is not a catch-all solution for designing AI-enabled systems. Instead, I see Remarkable and Unremarkable AI as delineating a continuum of AI systems design, in which designers will need to ask themselves (and their users) important questions regarding the nature of their needs and wants for a system. As much as designers might want to support and augment users' decision-making, all systems are arguably transformative in that they prescribe particular action possibilities for their users [225].

In recognizing this, an important question to raise is whether it is acceptable for a system to do so indirectly or to which degree users should be confronted with and offered control over this influence. This depends on the application area and, not least, the designer's values. The case of DSTs is interesting since both sides of the argument could be made; Yang, Steinfeld, and Zimmerman argue that a DST should support clinicians' work practices and only be noticed when their predictions "add value", in which case the Unremarkable approach is most appropriate. However, if one begins to question what value is added and who decides that it's valuable, the case is less clear cut; the patient might not wish to have AI play a role in their decision-making, or clinicians might want to know more about how the DST makes its predictions, what data these are based on, and if unintended biases are present, before trusting it. I am not suggesting that a remarkably designed DST would be more effective or that clinicians would be more inclined to use it (in fact, the opposite might be true). Still, I want to point out the political and ethical questions that AI-enabled systems raise and that should be addressed in any design process.

Determining the right level of remarkableness is a design choice, which designers should engage with, and which answer should be informed by understanding the system's stakeholders, the context of use, etc. Indeed, it might be the case that a system clearly would not benefit from being remarkable and that designing it to be so would only subtract from the value it provides (arguably, one example of this could be ML-powered, on-device enhancement of smartphone photography).

Thus, the right level of remarkableness is highly contextual. As such, designers should engage rigorously in determining it, even if the AI system's use case seems inconsequential at the surface level.

While I see Remarkable and Unremarkable AI as two end-points in a continuum, there is no reason why they could not be combined in different parts of a timeline and context of use. For example, it might be that clinicians at first wish to inspect the AI component of their DST, but as they come to trust it, they might be less inclined to do so; a sort of appropriation-through-use instead of Petersen [178]'s notion of "learning-through-use". Alternatively, clinicians might use this DST in an unremarkable way, observe that it is helpful, and then feel the wish to unfold its inner workings.

6.6 Summary

To address **RQ3**, this chapter presented Remarkable AI as an approach to AI-systems design that aims to empower end-users in interrogating and discussing such the impact of such systems on their lives. In a case-study analysis of the examples of my work; VotestratesML *(Paper A)*, Machine Learning Machine *(Paper C)*, and the ML Ethics Workshop *(Paper B)* I have derived three design strategies for designing such systems; foregrounding the ML process to expose AI components, iterative and safe experimentation to explain AI, and facilitating discussions and reconfiguration to allow interrogation and negotiation. Importantly, I do not see Remarkable AI as the goal for all AI-systems, but rather as one end of a remarkableness-continuum, the other end occupied by Unremarkable AI [238]. I have discussed how the right level of remarkableness might be determined and argued that this is not trivial and might have to be determined case-by-case by the designers of such systems.

RQ3: How can a CE approach to ML education inform AI/ML systems design?

Chapter 7

Discussion & Conclusion

In this chapter, I will discuss this dissertation's contributions to computing education and HCI. I will relate the chapters above to the research questions posed in chapter 1 and discuss how they are answered, as well as point out my work's limitations and possible future work.

Overall, this dissertation is the collective dissemination of three years of work. Throughout my Ph.D. process, I have examined how to make AI education matter. Given the promises and challenges of AI, we owe to children to educate and empower them with regard to it. To do so, I took my offset in CE and ML. However, CE as an approach had to my knowledge, not previously been applied to ML education, and so I started with little idea of the direction in which to take my work. Thus, my colleagues and I worked experimentally (through the design experiments presented in chapter 4) to uncover how CE and ML would interact. This work led to a refined set of design principles (see also chapter 4) and to the CEML framework, as an expansion of CE to address more advanced technologies such as ML (see chapter 5). Finally, our work led me to conclude that without empowering end-users of AI-systems in the wild, education only goes so far. This led to the articulation of Remarkable AI as an approach to AI-systems' interaction design informed by CE (see chapter 6).

7.1 ML Education based on CE and a Strong Technological Foundation

My first research question asks how CE as an approach to ML education can address the challenges posed by the widespread use of ML. In chapter 5, I presented and argued for the CEML framework, which is an extension of the design model by Dindler, Smith, and Iversen describing the activities involved in CE [50] with a technological foundation, consisting of ML concepts, practices and perspectives. CE is a well-suited approach to ML education as it focuses on developing students' understanding of and reflection on the technology surrounding them. However, I found CE alone was not enough, and I argue that engagement with ML on a technologically foundational level is also necessary. Developing students' "capacity to engage critically and curiously with the construction and deconstruction of technology" [50, p. 67] entails lifting the curtain of said technology, and to do so, a fundamental understanding of the technology is necessary.

The CEML framework is similar to other suggestions for ML and AI curricula. Still, it differs in a few key ways, especially regarding its insistence on scaffolding students' process of becoming active participants in determining the ongoing use of these technologies. Long and Magerko [137] present their vision of AI literacy based on previous work in the field. They include critical reflection as a focus in their AI literacy but are concerned with students' becoming critical consumers rather than contributors. This is a step towards CE, but the potential impacts of AI and ML merit going further than this.

Additionally, CEML represents a Nordic approach to ML education, rooted in the political history of the PD movement [22, 28] in Scandinavia. While Touretzky et al. [216, 217] have presented a comprehensive view on how AI and ML education might be included in the U.S. school system, the differences between the contexts warrant an approach that is rooted in a Nordic understanding of society and culture. I argue that the CEML framework is a first step towards such an approach to ML education.

CEML further complements prior work by offering a holistic approach that treats knowledge of ML not as an end in itself but as a means to investigate its meaning in and impact on students' lives and society at large. CEML does this by explicitly linking ML technological fundamentals to the reflexive activities in CE and insisting that these are inherently connected. This is not to say that a similar recognition isn't possible or implicitly present in other work, but rather that it is an intrinsic aspect of CEML. By making these connections explicit, CEML aims to support researchers and educators in articulating activities and tools that support the discussions around ML that are needed.

Further, as the Danish Fablab@School project demonstrates [204], CE provides a scaffolding for students to develop a "sense of self-determination" as being capable of conducting and reflecting on iterative design activities. Thus CEML aims to develop students' self-identification as citizens whose opinions regarding the use of ML are legitimate and should be taken seriously.

We are seeing great promises and progress made with AI, but we are indeed also facing significant challenges regarding its use [164]. The CEML framework offers a way to address these challenges in ML education that acknowledges their complexity and their relation to the technological dimensions of ML. Apart from CEML, my dissertation offers several exemplars of activities and tools designed in accordance with the CEML approach, which might guide and inspire researchers and educators in their work.

7.2 How to Design CE-based Learning Tools and Activities for ML

Second, I asked how we might design learning tools and activities that scaffold children's computational empowerment with regard to ML and what characterizes the design space for these. I have explored this through a series of successive design experiments, each building on the findings and experiences from the former. This led me to articulate six design principles; Considering tangible representations of ML, providing low floors to engaging with ML, providing opportunities for meaningful engagement, scaffolding students' fundamental, technical understanding of

ML, emphasizing critical engagement with ML, and designing for scalability.

These principles overlap with existing work on designing for ML education and computing education more broadly. For example, Long and Magerko [137] suggest embodied interaction, providing low barriers to entry, drawing on students' interests, and contextualizing ML. Regarding computing education more broadly, Resnick and Silverman [184], among other things, argue for low floors and wide walls, choosing black boxes carefully, and not overemphasizing programming. Similarly, Blikstein [23] argues for selective exposure to computing concepts and practices. Many of the design principles in this dissertation could be seen as variations on these, suggesting that these are applicable when designing for CE-based ML education. Further, CE as an approach also brings new considerations to the design of learning tools and activities, most prevalently the focus on critical engagement.

At the same time, I aim for the experiments in chapter 4 to exemplify how a CE-approach to tools and activities for teaching ML might look. Each experiment, I argue, surfaces different qualities regarding what to teach about ML and how to design for teaching it. They vary in scope, complexity, and context, but all address the same subject from different angles. This overlap between my experiments has allowed me to investigate how they interact, as presented in the annotated portfolio of section 4.3. With this, I aim to demonstrate how the experiments and the design principles deduced from them are interdependent and not a recipe for designing "optimal" tools and activities.

I have found that addressing these questions with a CDR methodology has been fruitful in generating an extensive collection of design experiments. Without such a collection, I would not have been able to confidently extract the set of design principles nor the tensions and synergies between them. Working with experiments in a comparative way [131] ensured that while each experiment uncovered new aspects of the design space, there was enough overlap between them to identify how their qualities overlap. Further, the throughput of my work was enabled by the collaborations described in section 3.6. Working closely with a design partner allowed me to develop and critique ideas continuously and (at least) doubled the constructive horsepower. Our well-established collaboration with teachers from Aarhus municipality ensured that our work remained founded in teachers' and students' reality and provided a stable venue for trying out and evaluating our experiments.

7.3 A Remarkable Approach to AI-systems' Interaction Design

Third, I asked what the implications of a CE approach to ML are for AI/ML systems design. In discussions of my work with my colleagues, I came to realize that while AI and ML education is a necessary step towards empowering children and adolescents with regard to its use, this use is still determined by the options available, which are currently proprietary and commercial systems that have their capital shareholders' best interests in mind, and not their users'. Thus, in addition to education, we need to better equip designers and developers to develop new, open systems that can be interrogated, discussed, and appropriated by their end-users. Furthermore, the work being done in AI and ML education can serve as a starting point for these discussions, especially when these aim to open up AI systems, let their users experience, and experiment with their underlying mechanisms. To inform such as approach, I have drawn on my own experiments. Referencing previous discussions in HCI, I have named this proposed approach Remarkable AI. Remarkable AI complements Unremarkable AI as proposed by Yang, Steinfeld, and Zimmerman [238] to unfold a continuum of remarkableness in AI systems.

I argue that in addition to other approaches, such as xAI, or everyday algorithm auditing [198], sensitizing designers to the concept of remarkableness in AI-systems provides new ways of democratizing AI. As Miller, Howe, and Sonenberg [147] argue, xAI is most often addressed to domain experts and people professionally engaged with AI. Remarkable AI complements xAI by offering ways to understand such systems that acknowledge the complexity of the social context in which they are used. Remarkable AI focuses on end-users of everyday AI-systems and on offering a bottom-up approach to designing these, in which users take an active part in understanding and shaping the systems to their own contexts. This is similar to the goals of everyday algorithm auditing. Still, Remarkable AI further suggests that systems themselves are opened up to their end-users in ways that not only let them audit the systems for symptoms of issues but also let them investigate the root causes of the issues themselves and experiment with alternative ways for them to work.

I do not intend for the continuum between Remarkable and Unremarkable AI that I propose here to be a concrete guide on how to solve individual usability issues with AI systems. Instead, I aim to sensitize designers and developers to the remarkableness of AI in the systems they create.

7.4 Limitations

A high-level limitation of my work is its capability to stir any real change given the current global political and economic powers in play regarding the extensive use of AI and ML. I myself do not feel particularly empowered regarding the AI/ML systems that I have come to depend on, so how can we expend children and adolescents to be? In discussions I have had with students throughout my project, it seems they, too, feel disempowered regarding their use of AI/ML systems and often peer-pressured into using them (e.g., in regards to social networking applications [132]) for fear of social repercussions. This indicates a need for equipping students with the skills, competencies, and attitudes necessary to discuss the role they wish AI/ML to play in their lives (and, indeed, the need for better AI systems). In the short term of my different design experiments, I have experienced students engaged in insightful discussions about ML and its use. I hope these discussions will influence how the same students approach these systems outside the classroom. I am hopeful going forward, with the growing focus on criticality in computing education research (see, e.g., [50, 100, 125]), and the efforts teachers are making to push for this to make its way into the classroom. However, further political action is needed, as well as better alternatives to the systems we have today, especially regarding social networking sites. With Remarkable AI and the notion of remarkableness in AI-systems, I have pointed to one way that designers of AI systems might become sensitized to these issues.

On a more granular level, my research has several limitations. First, my experimental focus and rapid approach to design experiments mean that I have yet to conduct thorough, longitudinal studies with regard to the learning outcomes of students or to their empowerment. While it is unclear how to evaluate this type of critical empowerment in the span of a Ph.D. project, learning outcomes could be assessed and provide a more quantifiable indicator of students' takeaways from our experiments. However, we have observed students drawing on their knowledge of ML in classroom discussions and when arguing for their choices in design processes. Furthermore, my work has focused on generating a large corpus of experiments to characterize the design space for such tools, the process of which was only possible without a focus on longitudinal studies of learning.

The design principles presented in chapter 4 were articulated after the fact, and as such, their collective generative qualities are yet unknown. However, they were formulated based on several design experiments in which they were all carried over as qualities from previous experiments. As such, the principles have, on an individual level and in various combinations, been used to guide my work and thus have been articulated based on design qualities from each experiment that had generative potential.

Finally, Remarkable AI is still in the early stage of conceptualization. While I have offered initial strategies for designing remarkable AI-systems, these have yet to be used generatively. It is also unclear when Remarkable AI would be appropriate to implement and when it would not. Indeed, in critical areas where time is of the essence, interfaces that are unremarkable and seamless [42] might be more appropriate if the systems are trusted by their users to perform satisfactorily. Additionally, Remarkable AI relies on the good faith of designers, and especially companies, to be adopted. While this cannot be assumed, there is a business argument to be made: a move towards greater remarkableness in AI-systems can be likened to the push for open-source software. It can potentially increase users' trust in systems by allowing them to be scrutinized technologically and for interested parties outside a company to do so. Falling short of this, it will be up to legislators to impose laws that ensure a level of remarkableness in systems' where it is needed, such as the European Union is proposing with their Artificial Intelligence Act[1].

7.5 Future Research

For the final section of my dissertation, I will discuss possible paths and avenues for how my work might inform future research. I have presented the CEML framework to bring CE's empowerment and democratization values to ML education. However, it has yet to be evaluated as a coherent approach in the classroom. As such, I invite future research that utilizes CEML for designing classroom interventions and testing these out. Where my work has mainly focused on the *coding* aspect of CE, i.e., constructing with ML, I see a possibility to apply the CEML approach to *decoding* activities as well, i.e., the analysis of existing AI/ML-systems in-the-wild. Here, the top part of the CEML model would be replaced by the analytical model from CE (see Figure 5.1, to the right). This

[1] See, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, accessed Jan 23rd, 2023

would allow students to draw on their (new) knowledge of ML to critically examine existing technology (to the extent that this is possible, given their opaque nature).

The design principles and their tensions and synergies were derived from examining the design experiments throughout my Ph.D. studies. While they outline the design space for CE-based learning tools and activities for ML, they are not exhaustive. There is still more work needed to map the design space. Thus I invite other designers, design researchers, and educators to draw on the principles in new empirical work and to contest, qualify and expand on the work I have presented here. Due to my work being explorative, and focused on uncovering the design space, most of my experiments are limited in scope and, thus, have yet to engage with some of the existing guidelines for designing tools and activities for computing education that are larger in scope. Specifically, I encourage work that engages with the principles of wide walls and high ceilings [184] and explores these implications for CE-based ML. Researchers with a CT focus have already addressed these principles with, e.g., the Snap! [115] and Scratch [60] programming languages, but without explicitly focusing on CE.

Additionally, more work is needed that explores how the tensions and synergies I have identified can be addressed. Specifically, the tension I have found between tangible and critical ways of interacting with ML is interesting to explore further due to the promises I have seen in tangible representations to make opaque and complex aspects of ML malleable for students. Additionally, I encourage future research to explore how the dimensions of ML as presented in section 5.2 intersect with my proposed design principles.

While the work in my dissertation revolves around ML, the CEED project, and other research address different emerging technologies, my work can also inform this work. Rotolo, Hicks, and Martin [187] define emerging technologies as having five shared properties: They are radically novel, meaning that they are somehow disruptive and challenging to current practices. They exhibit relatively fast growth. There is some coherence in research and development on the technology, i.e., some basic agreement on the technology. They have a promising and prominent impact on society, and finally, they are surrounded by some level of uncertainty and ambiguity regarding their impacts. ML is an emerging technology according to these properties. Other examples might be virtual, augmented, and mixed reality, quantum computing, and blockchain, but also gene manipulation technologies such as CRISPR[2]

The shared properties between emerging technologies suggest that my work is applicable to technologies other than ML, within reason. CE, generally, offers a method for engaging with these technologies that fosters critical reflection and enables students to take a stance on them. Since such technologies are uncertain and ambiguous, they should be approached without the assumption that we might easily predict how they will be used and with what impact. As with ML, approaching such technologies and their potential implications through their technological foundations anchors discussions about them in their practical reality and thus serves to qualify these. The CEML framework can inform a CE-based approach to such emerging technologies by asking what constitutes concepts, practices, and perspectives for them.

For example, take ExposeAR by Lunding et al. [138]. ExposeAR is a collaborative augmented reality (AR) authoring tool for children, in which they

[2]See, e.g., https://www.genome.gov/genetics-glossary/CRISPR, accessed Jan 5th, 2023.

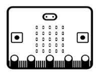
design small games for each other in a classroom setting to learn about AR. The authors argue that key concepts are plane tracking, image tracking, AR interaction techniques, and device pose tracking. Key practices involve collaboration, testing, and re-iteration. Finally, they focus on perspectives such as location tracking and behavior manipulation. They found that the children in their study could create systems that manipulated their peers' behavior to follow specific paths in the game and the real world, mimicking the popular game Pokémon Go[3]. The children were able to discuss how this technology enables behavior manipulation and to relate it to physical stores designed for consumers to follow specific paths (i.e., how an IKEA store always nudges a shopper past their restaurant). The key here is how AR as a technology, through tracking planes and images, allows digital content to be superimposed on the physical world to create engaging games in which people are required to move about and thus can be manipulated past certain stores or places. Understanding the underlying technology of AR allows students to reflect on and discuss not only that they are being manipulated but how AR technology enables this manipulation.

Further, it is not just in formal education that CE is needed. While many go through at least some form of formal education, an increasing number of people are past that part of their lives and are facing the world without the understanding and reflexive skills I advocate for in my work. For these people, we need to look elsewhere for venues to aid their empowerment. The SHAPE project at Aarhus University[4], with which I will (hopefully) shortly be affiliated, addresses this need by looking to two different possible venues; public libraries and trade unions. In a participatory collaboration, we will work with librarians and trade union employees to develop new ways for their respective target audiences to engage with AI and ML, emphasizing CE.

Finally, Remarkable AI presents an approach to AI systems design that aims to lift the curtains of their inner workings and empower users to interrogate and negotiate their use. As the approach is based on case studies from my work, I invite empirical work that applies Remarkable AI as an approach to develop novel interfaces and interactions with the technology. I see a need to engage in discussions on the level of remarkableness necessary in AI systems, and how this will vary across different contexts of use.

7.6 Wider Impact

Apart from the classroom discussions we have facilitated, my work has also led to broader potential impacts. For example, as mentioned in chapter 4, our work on ml-machine.org *(Paper H)* has led us to establish an ongoing collaboration with the Danish Broadcast Association (DR), and together we will deploy a curricular unit on ML, utilizing ml-machine.org. This unit will be deployed as a part of DR's "ultra:bit" platform on computing education, which has serviced more than 100.000 Danish students[5]. Additionally, we are in the process of establishing an ongoing collaboration with the micro:bit foundation, which has serviced an



[3]See https://nianticlabs.com/en/business, accessed Jan 5th, 2023.
[4]See https://shape.au.dk/, accessed Jan 17th, 2023
[5]See https://www.dr.dk/skole/aabn-drs-skattekiste-i-din-undervisning-og-brug-dr-skole (in Danish), accessed Jan 18th, 2023.

estimated 25 million children across 60 countries[6], on further developing and strengthening the micro:bit platform's ML capabilities.

On a more personal note, my work has enabled and motivated me to participate in the public debate about computing and AI education. For example, as a reaction to the plans of the Aarhus city council to remove the budget of De32 (i.e., the groups of teachers that we have collaborated with), I had a letter published in a local newspaper arguing against this decision. Further, I have been asked to participate in a newly established panel in my trade union to inform their efforts to prepare and qualify their members for an increasingly digital job market. Finally, when the Covid-19 pandemic hit Denmark, I found myself motivated to join and participate in a political party in Denmark. To my surprise, this party had no politics on digitalization or IT. Nevertheless, I have succeeded in establishing a national working group in the party on IT politics with a particular focus on establishing computing education in the Danish school system.

A Short Goodbye

Thank you for spending time with my dissertation, no matter how you engaged with it. I hope you found the read enjoyable. I hope you have found my contributions valuable and, if nothing else, interesting.

I look forward to following my dissertation's life and the research fields it aims to contribute to.

[6] See https://microbit.org/about, accessed Jan 18th, 2023

PART II

Research Papers

Paper A

High School Students Exploring Machine Learning and its Societal Implications: Opportunities and Challenges

Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, Maarten Van Mechelen, Arthur Hjorth, Niels Olof Bouvin, & Marianne Graves Petersen

Abstract

The increased use of AI and machine learning (ML) calls for a general AI literacy, in particular regarding understanding how ML works, the process behind creating ML models, and reflecting on its implications. Where existing learning tools focus on the first two, we explore opportunities and challenges for meaningfully engaging students in understanding and reflecting on ML in their everyday life. We designed VotestratesML, following a Constructive Design Research approach, as an ethics-first learning tool that allow students to explore implications of ML for democratic elections. Based on deployments of VotestratesML in two high school social studies classrooms, we found that safely exploring ML from a concrete starting point helped students reflect and form opinions about its use, that promoting iterative exploration through collaboration and competition motivated them to explore, and that foregrounding ethics in the design and grounding ML in a well-known subject area allowed them to engage with ML on a personal level.

A.1 Introduction

The increased use of Machine Learning (ML) in almost all aspects of our lives [164, 173] increases the necessity of a widespread AI literacy [53, 137]. AI, and subsequently ML, is part of the infrastructure of many everyday technologies in ways that are often opaque and difficult to comprehend[36, 199]. Understanding how AI and ML work is important not only for those pursuing a career in STEM (Science, Technology, Engineering, and Maths) fields but, arguably, for all children, as these technologies are increasingly pervading all aspects of our lives (i.e., education, leisure, and work). Thus, AI literacy becomes a precondition to fully participate in society, whereby the term literacy denotes that the goal is not simply to develop children's instrumental skills, but also a critical understanding of manifestations of power and ideology in AI technologies, and consequently, its personal and societal implications [223].

The importance of a critical understanding of technologies has also been echoed in recent studies on Computational Thinking (CT). Researchers (e.g., [33, 76, 107, 110, 213]) have argued that traditional CT competencies (e.g., decomposition, abstraction, automation, etc.) should be complemented with critical perspectives on the personal, social, and political consequences of digital technology. In these broader and more comprehensive approaches to CT, students learn about technologies and computational infrastructures, not just in order to become better programmers or designers, but also to enable them to engage with political and ethical questions about technology in the real world. As Kafai et al. [110] put it, these approaches see the *"cognitive understanding of underlying concepts of CT and its uses in the world as key to becoming a more critical practitioner of computation"*. We agree with the critical approaches above, and align ourselves specifically with the notion of Computational Empowerment (CE) [50, 107]; a perspective on computing education rooted in Scandinavian participatory design values of democracy and skillfulnes that aims to develop children's critical understanding and informed decision-making with regards to the role of digital technology in their lives and society more broadly [107]. Using CE as an approach to AI literacy raises the question as to how ML learning tools can qualify and support children's critical reflection and understanding of ML, including its personal and societal implications.

With this awareness of the need for students and the general public to understand AI and ML, researchers have started exploring what K-12 students ought to know about ML [216] in order to become AI literate [137]. The understanding needed can be summarised as belonging to three main areas; *1)* what is a ML system, *2)* how are ML models developed, *3)* and what are possible implications of ML applications? An important aspect to consider here is what Blikstein [23] terms *selective exposure*, i.e., which aspects of the technology should be foregrounded to children, and how to maximise what can be achieved with the technology through a certain tool. In current research we often see the former two areas emphasised in educational tools and practices [115, 232, 246], while the third is either neglected or only briefly touched upon (see e.g., [246], which seeks to engage ML novices in interactive model building, but leaves notions of empowerment to future work). This trend is also demonstrated by Giannakos et al. [73], who identify several projects/games for teaching AI/ML, where most focus on the first two areas (e.g., ML4Kids[1] and AI4Children[2]), while a few include the third area (e.g., The Moral Machine[3]). This aligns with findings of Van Mechelen et al. [221], who, based on a systematic review of child-computer interaction research, concluded that design ethics (i.e., the personal/societal impacts of technology and its qualities in use) has been been underdeveloped in the literature, both as an overall concern and an explicit learning goal for children.

Thus, while interesting work has been done that can inform the design of tools and activities to teach students about ML; to our knowledge, no studies have taken a CE approach to learning about ML, which entails a more equally distributed focus on all three main areas. With the aim to do so in this paper, we explore the following question: *What are the opportunities and challenges for a CE approach to meaningfully engage students in understanding and reflecting on ML?* To explore this question, we have reviewed existing literature on designing ML curricula and tools for teaching it with regards to how to support CE. Additionally, we have designed VotestratesML; a prototype of an interactive, collaborative learning tool for ML, aimed at supporting high school students in exploring and reflecting on the role of data and ML in political campaigns. VotestratesML was iteratively developed in a Constructive Design Research (CDR) process [130, 205], and serves both as a proposal for a high school learning tool for ML, and as a means to explore the design space of ML learning tools in high schools. The process was realised through six user-interventions in two social studies high school classes. To inform the future design of learning tools for ML, we herein present our findings and how they formed the design of VotestratesML.

With this paper, we address Eriksson et al.'s [59] call to open up new frontiers for research that carries child-computer interaction (CCI) beyond the foundation and legacy of Papert and Harel [169] in order to embrace the entire education system with a broader, more critical perspective on 21st century learning, and encompass diverse issues related to children's use and understanding of emergent technologies, such as ML [59], a need more recently identified by Van Mechelen et al. [223] who note that critical notions of empowerment are only marginally represented in CCI. More specifically, our contribution is a synthesis of what students should know about ML as seen through the lens of CE, as well as the design of VotestratesML and the ways in which this tool can engage students in learning about ML and critically reflect on its impacts.

The paper is structured as follows; In section A.2 we review existing literature on ML curricula and learning tools. Then in section A.3 we present VotestratesML. Here, we describe a typical use scenario in a classroom, then we detail the VotestratesML prototype and user interface (UI) and highlight certain aspects of the design of the tool. Finally, we provide the overall design rationale behind the tool. Following this section, we present our design process and study in section A.4. In section A.5 we present our findings, before discussing them with regards to our research question in section A.6. In section A.7 we discuss limitations of our study, and finally in section A.8 we conclude the paper and point to future work.

[1]https://machinelearningforkids.co.uk/
[2]https://www.ai4children.org/
[3]https://moralmachine.mit.edu

A.2 What students should learn about Machine Learning

Here, we review existing literature about teaching ML to K-12 students, divided into three main areas that correspond to current research trends about what these students should understand about ML [137, 216]; *1)* what a ML system is, *2)* how ML models are created, and *3)* what possible implications of ML applications are. Often learning tools for ML address the first two areas and to a lesser extend the third, but we argue that a focus on CE requires a specific focus on the third area in combination with the former two.

A.2.1 What is a ML System?

We see in the literature a focus on teaching what a ML system is. Along these lines, Touretzky et al. [216] emphasise how computers learn from data and maintain models/representations of the world based on this data. Similarly, Long and Magerko [137] state competencies such as distinguishing between artefacts that use and do not use ML, understanding the strengths and weaknesses of ML system, and recognising how ML systems reason and make decisions as core in developing a literacy about ML and AI. This approach is reflected in most tools for teaching ML. Many block programming languages include this such as AI Snap! blocks [115] that allow students to build ML applications in the Snap! block-programming language, using predefined ML blocks. Similarly, *Cognimates* [51] and *Machine Learning For Kids* from IBM allow children to experiment with speech recognition, object recognition, etc., through the *Scratch* programming language. Other tools specifically designed for teaching ML have a similar focus. The popular Teachable Machine [39] by Google, allow users to input data and train ML models for image, sound and gesture recognition. Popbots by Williams et al. [232] introduces the basics of ML systems, such as training a classifier by labelling data input. *Cozmo*, a robot from Anki, allows children to play and experiment with built-in object and marker detection, face recognition, etc.

While understanding what constitutes a ML system is important, it is not enough to understand the whole of ML, and from a CE perspective, not enough to develop the capacity for critically engaging with the implications of ML. Understanding what a ML system is, will help students recognise ML systems in-the-wild, and with reasoning about what tasks ML as a technology is suitable for addressing, and importantly which it is not. However, it does not help students in deciphering and challenging assumptions implicitly built into the system by its designers and the implications these might have for the people affected by the system.

A.2.2 The ML Process

Another current focus is on teaching the process of creating ML models. Here, Long and Magerko [137] emphasise the different steps involved in creating ML models and the role humans play in doing so. Several learning tools focus explicitly on this aspect of ML. Hitron et al. [92] introduce the *Gest* system, a ML-based gesture recognition system designed to teach children aged 10-13 about data labelling and evaluation of ML models. They conclude that children are able to understand ML processes, specifically data labelling and evaluation of models, and that children are able to apply this knowledge in other contexts. Zimmermann-Niefield et al. [246] deploy learning tools for ML into a high school context. They design and evaluate *AlpacaML*, an iOS application building models of athletic moves through sensor-input for use in physical education in high schools. They found that students developed their own theories on what constituted a "good" model and were focused on the performance of their particular model rather than whether the model was true to what it modelled. Kaspersen et al. [120] and their Machine Learning Machine try to encompass the entire process of building ML systems, from collecting data to training and later to evaluating a model, and the iterative nature of doing so.

Understanding the ML process is another central aspect of developing AI literacy from a CE perspective. An understanding of the ML process entails understanding what steps are involved in creating an ML model and importantly the choices made by humans in these steps (for a further discussion of this see, e.g., [55]). In combination with an understanding of what ML is, understanding ML processes will allow students to further pick a part of any ML system, and go beyond the face-value of the system to reflect on and discuss how it came to exist.

A.2.3 Implications of ML

Finally, there is a focus on the implications that ML systems might have on society and in our personal lives. Long and Magerko [137] stress that students should understand that "data cannot be taken at face-value" and that some systems have the ability to "physically act on the world". Finally, they point out that students should be able to "[i]dentify and describe different perspectives on the key ethical issues surrounding AI". Similarly, Touretzky et al. [216] present the idea that AI and ML systems "can impact society in both positive and negative ways" and that this is a important to teach. We also see this in research on learning tools, albeit to a lesser extend. Bilstrup et al. [19] engage high school students in designing ML systems in a card-based design workshop which engages them in discussions of the ethical dilemmas in applying ML to solve real-world issues. Similarly, Skinner et al. [202] explore children of color's perception of fairness in ML/AI systems in a co-design workshop in which 9-14 year old children design a "fair" AI librarian.

Understanding the possible implications of ML implies viewing ML systems as socio-technical systems that engage in complex social and technical structures in our society. Seeing ML systems this way allow students to look beyond the system itself and reflect on the way it affects people around it in direct and indirect ways.

A.2.4 Three areas of ML

In summary, we see that what students should understand about ML can be divided into three areas; *what ML is, the ML process* and *implications of ML*, and that focus on them are different between different learning tools. Table A.1 presents an overview of how a selection of different tools and activities address them (the selection includes a limited but broad view of both popular publicly available tools and recent research prototypes). Focusing on one area is not sufficient as, e.g., understanding the ML process requires knowledge of ML concepts and vice versa. E.g., in a later paper, Zimmermann-Niefield et al. [245] expand

Area of ML	AlpacaML [246]	Gest [92]	Teachable Machine [39]	AI Librarian [202]	Machine Learning Machine [120]
What ML is	Training data, ML models, training	Training data, ML models	Training data, ML models, hyperparameters, training	N/A: Children were asked what AI and "fair" meant.	Training and test data, training and evaluation
The ML Process	Data gathering, labelling, cleaning and analysis, evaluating models	Data gathering and labelling, evaluating model	Data gathering and labelling, evaluating models	N/A	Entire process, from training to evaluating models
Implications of ML	N/A: Authors suggest to explore it in future work	Explored in post-interviews, but is not part of the learning tool design	N/A	Design and discussion of what constitutes "fairness" in an AI Librarian.	N/A: Authors suggest activities, but do not evaluate them

Table A.1: The three areas of ML and how a selection of learning tools address them.

AlpacaML to integrate with Scratch allowing students to build gesture-controlled interactive media and to engage more with understanding different components of a ML system and how they fit together. Similarly, tools targeting the first two areas, sometimes include reflections on ML implications; Hitron et al. [92] also investigate if children after using Gest are able to reflect on situations where it is inappropriate to use ML, and find that about half of the participants were able to identify issues about privacy, intimacy, and safety. Kaspersen et al. [120] propose scenarios explaining how the Machine Learning Machine could be used to explore ethical issues of ML, such as data representation and biases, but do not evaluate this aspect. These aspects are, however, most often not presented as a core aspect of the learning tool design, but as an addition that requires a separate discussion.

From a CE perspective, understanding and being able to reflect and act on the implications of ML is the most important of the three areas, and should be at the core when designing ML learning tools. However, to be able to do so requires some insight into what ML is and the process of creating ML models and humans' role in doing so. Thus, all three areas must be covered, at least to some extend, by learning tools aiming to support children in becoming computationally empowered, and research is needed that explores how to do so. In the following section, we present VotestratesML, which as a learning tool embodies our exploration of how a balance between the three areas could be struck.

A.3 VotestratesML: A Collaborative Learning Tool for Teaching ML

VotestratesML is a web application enabling students to collaborate in real time on iteratively building models for predicting voter behaviour using voter profile data, and which aims to scaffold class-wide discussions about ML. Data from

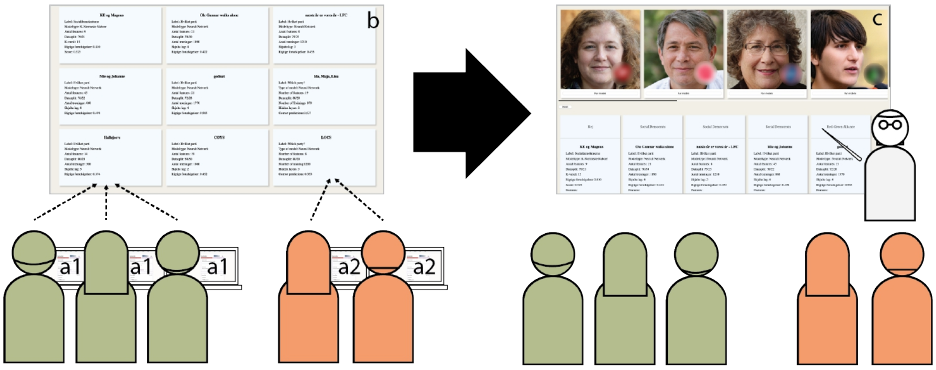


Figure A.1: The different components of VotestratesML and their use: *a* is the *Collaborative Component*, *b* the *Competitive Component*, and *c* the *Discussion Component*· By using the Collaborative Component, groups of students build models, which are pushed to the Competitive Component, allowing groups to compare their models. The Discussion Component is used afterwards to discuss ML, mediated by a teacher. See Figure A.2 for an in-depth view of the Collaborative Component. The two other components are explained further in Figure A.3.

a survey of the Danish national election in 2015 [208] is used in the prototype to make predictions about real world voters. VotestratesML and the rationale behind it embodies our suggestion for how to teach ML with an emphasis on CE, although other researchers might design such a tool differently.

A.3.1 A VotestratesML Use Scenario

In a typical use situation of VotestratesML, students are introduced to the tool and how it supports building ML models. Next, students are divided into groups of 3-4 people. Each group member logs into VotestratesML on their laptop gaining access to the collaborative ML tool, see Figure A.1a. The teacher then asks groups to create the best possible model for, e.g., predicting if a person will vote for The Social Democratic Party. Each group chooses the *Votes for The Social Democratic Party*-label available in VotestratesML. The groups go on to discuss which features to include in their model. Here, students can draw on their existing knowledge from social studies class. For example, according to the *Michigan Model* for voter behaviour [108], which students are taught in advanced social studies, voting behaviour is affected by family, and thus, it might be important to choose features describing how the parents of a person voted. Once students agree on a set of features, they need to choose an algorithm, and to determine its parameters. Once the model is trained and tested, students can change their model parameters, features, shuffle the data or change the training-data/test-data ratio to try to improve the model. This workflow is summarised in Figure A.2. When they are satisfied, students push their model to a shared view projected on the wall in the classroom as seen in Figure A.1b. From here, they can compare it to models from other groups and discuss how to improve their model. Finally, when the teacher ends the exercise, a more detailed view is projected on the wall (see Figure A.1c) and used in discussions of students' choice of features and model parameters, the predictions of the models, and the implications of this way of working on politics and society. The topic of these discussions depend on the task given to the students, but might include how these predictions can be used to target political advertisements to specific demographics and what the consequence of this is.

A.3.2 Collaboratively Building, Evaluating and Reflecting on ML Models with VotestratesML

Below, we describe in detail the design of VotestratesML as it played a central role in the knowledge generation leading to the contributions of this paper. This is explained further in section A.4.

To support real time collaboration, VotestratesML is built in Webstrates [126], which synchronises DOM elements between multiple clients in real time. The students each work on their own laptop, but collaborate with their group through a shared Webstrates-based website, making the interface of VotestratesML inherently collaborative between group members' laptops. Every group member has full control over the application and they must communicate to coordinate their work, the intention being that students will divide tasks with different focus points between them.

VotestratesML consists of three interdependent components that support different types of class activities through providing different functionalities and model views. The relation between components is illustrated in Figure A.1.

Using the *Collaborative Component*, students collaborate in groups by building, testing and improving the group's ML model for predicting voter behaviour. VotestratesML conceptualises the process of building ML models in a series of steps as illustrated in Figure A.2. First, students process the data-set by shuffling the data and splitting it into a training data-set and a test data-set (see Figure A.2a). Second, they explore the properties of the data-set, which ranges from age and gender to voters' attitudes towards environmental issues and tax policies, and choose one label and as many features as they like (see Figure A.2b). Finally, they choose between two ML algorithms; K-Nearest Neighbour (KNN)[4] or a Feedforward Neural Network (FNN)[5] and set the parameters for the chosen model (For KNN: value of k, for FNN: layers and number of iterations at the training step), and train and evaluate a model to see how it performs on the test dataset (see Figure A.2c). Students are provided with the model's *accuracy* and *f1* score, which they can use to compare their model with earlier attempts or other groups' models. Students can freely jump between these steps and group members can work on different steps simultaneously. The interface is designed to support students in following this process, even if they do not fully understand ML. It does so, by only allowing actions that bring the application to meaningful states, greying out buttons and hiding UI elements when not relevant. Thus, students can work on creating functional ML models from the beginning and can explore ML further by tinkering with data and model parameters to improve the predictive ability of their models. The *Competitive Component* has a single view, see Figure A.3a, which is projected on a shared screen in the classroom and provides information about the scores of each group's model. Models are

[4]https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm
[5]https://en.wikipedia.org/wiki/Feedforward_neural_network

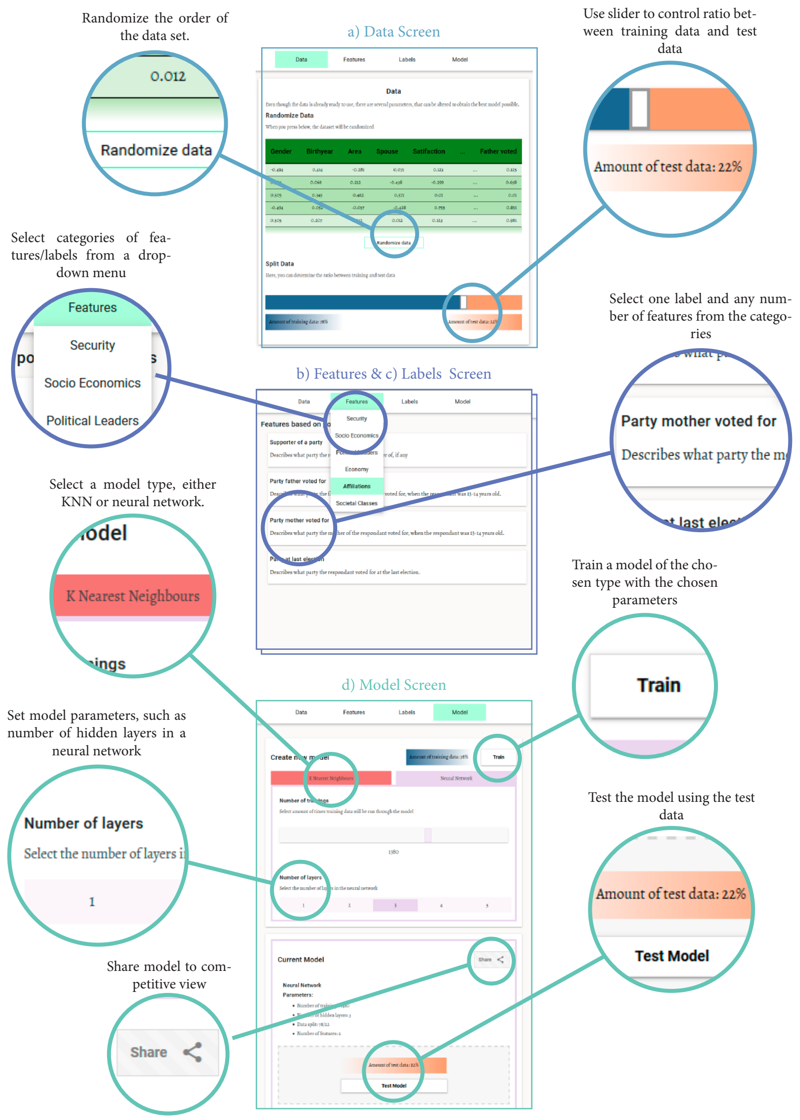


Figure A.2: The VotestratesML Collaborative Component in detail; a) shows the Data Screen that allow students to experiment with their input data, b) and c) shows the Features and Labels screens (conflated as they are similar in design), that allow students to determine what data their model should consider (features) and what it should look for (label). Finally, d) shows the Model Screen that allow students to select the model type, configure it's parameters and train a model.

Figure A.3: The competitive aspect of Votestrates ML (cropped): *a)* shows how students can compare their models while working to improve them; *b)* shows how the teacher can use students' models on personas; *c)* shows how predictions can be discussed and how the models can be inspected in detail.

pushed to the Competitive Component by each group as they iteratively try to improve them. Using this component, students can share their models with the classroom, and see how well they fare in comparison with other groups' models. Finally, the *Discussion Component* also has a single view which is projected on a shared screen and used to mediate a joint discussion in the class. This component is used to inspect students finished models when the teacher ends the activity of building them. The Discussion Component provides full information about each model, see Figure A.3c, and lets the teacher use students' models to predict the behaviour of different personas, see Figure A.3b. This is used as an anchor point for discussions about the implications of ML for elections and broader discussions about ML in society.

A.3.3 Design Rationale: Machine Learning for Social Science

VotestratesML differentiates itself from similar learning tools in a number of ways. First, it combines the three areas of what students should know about ML; *what a ML system is, the ML process* and *the implications of ML* (with an emphasis on the latter). Second, instead of starting by introducing ML concepts, VotestratesML takes departure in the social studies subject as a tool for analysing voter behaviour. The tool is designed to support typical social studies class activities; Students work in groups on tasks assigned by the teacher, followed by a discussion based on the group work. The available data categories were chosen based on models of voter behaviour taught in Danish social studies classes [108], allowing students to explore how ML models can predict voter behaviour, compare theoretical models with ML models, and discuss ML models from a social studies perspective and how they are already used actively to affect the outcome of democratic elections [3, 158, 185].

By focusing the design of VotestratesML on making predictions about people's voting behaviour based on properties like age, income, gender, etc., students can draw on their prior knowledge to explore the extent to which their knowledge and preconceived notions are supported by real-life data. Thus, their prior knowledge and interest in the topic of voter behaviour can bootstrap their exploration of machine learning methods and the implications of these.

In addition, the design of the interface aims to scaffold students in building ML models, even with a very limited understanding of ML. Through experimentation with building models for predicting voter behaviour and iteratively improving these models, students gain experience with the trial-and-error process and begin to gain an understanding of ML models. The KNN and FNN algorithms were chosen because they exemplify the extremes with regards to ML complexity; KNN being simple and FNN being complex. KNN was used to illustrate the basics of ML, while FNN was used to illustrate how ML can become almost too complex to understand. As such, students were not expected to understand the FNN algorithm, rather they could experience how difficult it can be to explain its predictions.

By providing a collaborative interface in which groups are forced to negotiate design choices internally because all group-members have full control of the application, VotestratesML aims to scaffold reflections on the process of building ML models. Similarly, the competitive aspect of VotestratesML aims to scaffold students in considering how their models can be improved, and VotestratesML allows students to do so through trial-and-error, which is analogous to how ML models are often improved in real-world situations. Thus, students experience first-hand the design choices and potential issues embedded in ML models. The design of VotestratesML aims to actively scaffold insightful discussions about the implications of ML and to motivate students to further explore ML by interacting with the tool.

A.4 Method

VotestratesML is the result of a constructive design-research (CDR) process and was evaluated and iterated on through several real-world deployments. In this section, we present the methodology behind our work, the design process leading to the design of VotestratesML as well as how it was evaluated in the field.

We adopted a Constructive Design Research methodology [128, 130] in order to investigate how ML learning tools can be designed for K-12 classes and how they can support critical discussions and reflections around the use of ML. According to Bang et al. [7], CDR can be seen as a way of iteratively making and testing hypotheses, where knowledge-generation revolves around the construction of an artefact (e.g. a product, a service, media etc.) [128, 205] and experiments with this [7, 130]. In this work, VotestratesML is the central artefact, and our findings are based on the different experiments leading to the version of VotestratesML presented above as well as the lessons learned from the in-situ deployments of VotestratesML. The design process led us to the final version of VotestratesML as presented in this paper (see Figure A.4 for an overview), and to two central hypotheses (similar to the two central arguments of our design rationale); that grounding ML in an existing subject (i.e. social studies) can engage students already interested in the subject but not in ML, and second that collaborating in groups and competing against other groups can be an effective way of scaffolding students' reflections about ML.

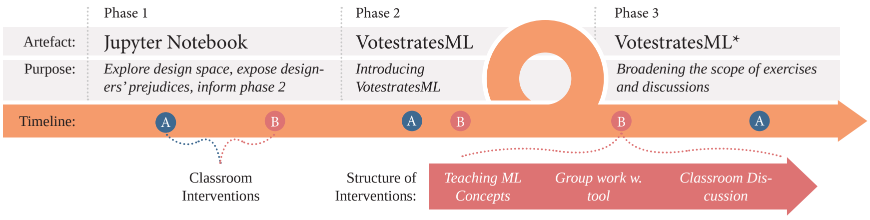


Figure A.4: An overview of the design process of VotestratesML. The study spanned approximately four months. Each class was involved three times, and each intervention had the same basic structure.

A.4.1 Participants

We collaborated with two social studies teachers from a public high school, located in an upper middle class area in a mid-sized Danish city, each inviting us to work with one of their classes. We consented 30 students in one class (A), and 31 in the other (B), aged 17 to 20. All participants under the age of 18 were asked for the consent of their parents or legal guardians. Consent forms were collected by the researchers at the beginning of the interventions. All students from both classes had elected into a 3-year social studies class. Students from class A were in their third year, while those from class B were in their second year. In both classes, the gender ratio was approximately equally distributed.

A.4.2 Intervention Protocol

VotestratesML was designed in an iterative design process with three main phases (as illustrated in Figure A.4). In each phase of the design process, each class participated in an intervention with the dual purpose of teaching students about ML and informing the design of VotestratesML, which was introduced in phase 2 and iterated on in phase 3. This resulted in a total of six interventions (see Figure A.5 for an example of the intervention setup). All interventions took place during regularly scheduled social studies class, with each intervention lasting 90 minutes. Two researchers and a teacher were present during each intervention. The researchers took over teaching during the intervention, while the teacher focused on helping students with exercises and contributing during class discussions.

All interventions were structured similarly; a brief introduction was given about the subject of the particular phase, followed by the students doing group work, and ending in a general discussion in the classroom. The first phase focused on introducing ML to social studies students and informing the first version of VotestratesML. We did so by introducing students to central ML-concepts and letting them work with these in a collaborative Jupyter Notebook, based on Google Collaboratory. We used common interactive elements, such as sliders and drop-down menus to allow students to engage in building ML models without prior programming knowledge. While this was effective in engaging the few students who were already interested in ML, it did not motivate the rest. Thus, in the second phase we explored how to contextualise ML for a social studies class, by deploying the first iteration of VotestratesML. It was also in this phase that the competitive element was introduced, to see what effects this would

Figure A.5: Students engaged in building models with VotestratesML (Note: all necessary permissions were collected to feature photos of students).

have on the students' motivation. We introduced new concepts integral to using VotestratesML, in particular the K-Nearest Neighbours (KNN) algorithm, and let students work in groups using VotestratesML to build social-studies specific ML models, such as creating a model for predicting who will be voting for the Social Democrats. This was followed by a discussion of how such models could be used during elections. Finally, the third phase aimed to reintroduce some of the complexity from phase 1, and introduce the ambiguity of ML in an explorable way that might support more general discussions of the implications of ML. We introduced students to neural networks and asked them to work in groups to explore these with VotestratesML. As a final exercise, we asked students to build models for predicting the voting behaviour of the researchers in the classroom. Finally, we facilitated a closing discussion on the role of AI in society and how it might evolve in the future.

A.4.3 Data Gathering and Analysis

Throughout the interventions data was collected in the form of observations and field notes, sound recordings and photography. We made observations and field notes with an open ended approach in which we in particular noted critical incidents related to students' use of ML concepts in discussions and group work, students' engagement in discussions and group work as well as frustrations and break downs during the group work with VotestratesML. During the introduction to each phase, one researcher did the introduction while another made observations and wrote field notes. During group work and discussions audio recordings were made by placing recording equipment at the desks of the student groups, and both researchers made observations and field notes as well as taking photos of the students working.

The data was analysed deductively by the first two authors, with the two CDR-hypotheses (see above) as sensitising concepts [25, 145] guiding the analysis. Following each intervention, a write-up [145] of observations and field notes was produced and discussed between the two researchers present at the interventions. Audio recordings of students' intra-group discussions during group work and of class-wide discussions were reviewed by the first two authors, and excerpts that informed the sensitising concepts were selected and later transcribed. With regards to grounding ML in social studies, we particularly looked for ways in which students used ML concepts and their experiences with VotestratesML to identify societal implications of ML in discussions in intra-group and in class-wide discussions. Regarding collaboration and competition we focused on students' engagement and frustrations, how they discussed design-decisions while using VotestratesML and for breakdowns with the prototype.

Between interventions, these analyses were used for designing the following iteration of VotestratesML, with each iteration embodying the findings from the previous intervention [41], as described above. Following the interventions, the first two authors did an analysis similar to the above, but across all three design phases, the results of which is presented below.

A.5 Findings

As argued, we see a need for CE learning tools that support students in understanding ML, the process of creating ML models, and in reflecting on the implications of ML and how it should form our future lives and societies. Here, we present and discuss the findings from designing and evaluating VotestratesML.

A.5.1 Making ML Personally Meaningful

VotestratesML frames ML as a social studies tool for predicting voter behaviour, that can be used to achieve goals in the social studies subject. With the tool, students can utilise theoretical models from social studies or test their own intuition about who votes for whom. In this way, students can engage with ML both from the perspective of the ML technology and its concepts and processes, but also approach it from a well-known subject area. In particular, the narrative around ML and the models of voter behaviour built by the students inherently frames the discussions in terms of potential societal implications of ML.

This design is based on our experiences in phase 1, where we observed that many students did not see themselves as individuals who could or should understand the technical aspect of ML. The exercises in Jupyter notebooks were centred around basic components of ML, such as features, labels and training- and test-data. From observations made during group work in these interventions, we found that students quickly distanced themselves from the exercises with comments such as *"I'm not very good at math"* and *"This looks very complex"* before having invested any time with the prototype. As one student put it in a conversation between them and one of the researchers present: *"For this to be really exciting, you probably need to be interested in the subject [ML]"*. Although students were eager to discuss ML, they did not draw on ML concepts in their discussions.

In contrast, in the interventions with VotestratesML, students' models and their predictions were used as basis for class discussions in which students' arguments were grounded in social studies theory and in their existing knowledge of voter groups, or based on their experiences from trial-and-error processes with the prototype. Different types of arguments led to different discussions concerning the use of ML. For instance, in phase 2, students role-played as advisers for the Social Democratic party, and were asked to advice, based on ML models, how to best persuade voters from another party. The exercise spurred discussions about the ethics of using ML in political campaigns, e.g.: *"The Social Democrats do vote-seeking [in the role-playing exercise]. But I don't have much respect for that. I would rather go for policy-seeking..."*. In another discussion about how micro-targeting could make democratic parties focus on single issues instead of prioritising coherent ideologies, one student argued: *"It is a democratic problem, if you [democratic parties] focus on single issues rather than a coherent ideology or harmony in their policy.. it will lack coherence and will not work"*. Furthermore, the predictions of the personas' behaviours spurred technical questions about ML, which could be verbalised using students' models about, e.g., how models with high accuracy sometimes make poor predictions or why models with more features are not always better.

A.5.2 Supporting Reflection through Collaboration & Competition

The iterative process of building ML models combined with collaborative and competitive aspects of VotestratesML spurred reflection on and discussion about ML during the activities. While competing in small groups in phase 2 and 3, students collaborated to build the best possible model in order to beat the other groups. Students were observably motivated by the competitive aspect of the exercises and the option for each group to compare themselves to the other groups seemed to give students a sense of ownership, and pride in their models, motivating them to keep improving it. For example, groups regularly cheered after having tested their new model and some students would make exclamations such as *"Share it, share it, share it! We need a group name! It's a great start, nobody is close to us"* or *"We should choose this features, nobody else has used this one!"*.

The audio recordings of students' intra-group discussions during group work with VotestratesML showed that students were eager to supersede the models of other groups. Students would eagerly discuss how to gain an advantage over their competitors, and these discussions were, although based on students' understanding of social studies and their personal prejudices, discussions about ML, e.g., whether a new feature was important prior to selecting it:
Student A: *"What about this one [a feature]: 'Supports a political party.' It could be good."*
Student B: *"Yes.. But isn't it very typical for young people to issue-vote? Like, just voting for what you feel?"*
Student A: *"Yeah.. You might be right about that."*
Student B: *"Can you try train it? [the model with the new feature]"*

The discussions and experiments with different features during group work, observably spurred students' curiosity about ML, leading them to ask questions to the authors present such as *"What is a good score?"* and to ideas about why other groups' models performed better than theirs; *"What if we have trained our model too much, so it now has become bad?"*. On the other hand, the competitive elements of VotestratesML made some reflections superficial due to the high pace of the competition. Often, students could be seen brushing over otherwise curious results in order to have time for another attempt to beat their peers.

The experience of building and optimising models by any means, however, also led to nuanced reflections on the issues around ML after the group work and to students taking critical stances towards the technology. For example, from the audio recording of a classroom discussion in phase 3 one student argued: *"It is against our culture, where you are not allowed to generalise by any means. That is outright what this is built on."*, while another argued: *"Machines only look at what we ask them to look at. If they have made a model, it is us who decide, what they should be looking for. Humans are better at marginalising than machines are. If we do not make them marginalise groups, they won't do it!."*. Another student recognised problems with ML models on how to choose which norms should guide the models' decisions: *"It is very different from person to person, which values you find most important. [...] It is just very different, and it [an ML model] may not be able to predict this"*.

A.5.3 Conceptualising ML for K-12 students from a CE Perspective

Where other tools focus mainly on what ML is, and the process of creating ML models, we have worked towards exposing the implications of ML through the interaction with and activities around VotestratesML.

This approach worked well for engaging students in understanding ML and for reflecting on its implications, but we also experienced that students formed misunderstandings about ML, which hindered their reflection: During phase 2, we observed that students had difficulties understanding the connection between the data and the voter behaviour predictions of their models. If choosing a feature did not have the expected effect or a model made an unexpected prediction, most students rejected what the model actually stated, and instead wrote it off as a problem with the data, rather than reconsidering the veracity of their expectations. Especially the generalisability of predictions seemed to be difficult for students. To explore this, in phase 3, we asked students to use VotestratesML to predict how the researchers present at the interventions would vote. While working on the models, students asked the researchers questions such as: *"What do you [the researcher] think of this?"* or *"How do we know, what topics you [the researcher] care about"*, indicating that students thought of the models as being able to predict the voting behaviour of a specific person or group of persons, rather than as general models for predicting voting behaviour. We believe this misunderstanding about the role of data stems from the data-set being relatively black-boxed in VotestratesML. As argued above, this was done to scaffold students in quickly building ML models, but might have hindered their understanding of the data-set and what each row in the data-set represents. This highlights the importance of finding the right concepts to black box and the right ones to glass box, something also noted by prior research in ML learning tools [92], and that a good balance between black and glass boxes is not trivial to achieve.

A.6 Discussion

In this section, we discuss the findings above as they relate to our research question; *what are opportunities and challenges for a CE approach to meaningfully engage students in understanding and reflecting on ML?*.

A.6.1 Implications of Machine Learning: Foregrounding Ethics and Grounding ML in a Well-known Subject Area

VotestratesML aims to teach the implications ML could have on our lives and on society. It does so through two measures; foregrounding ethics in the design of the learning tool itself and the activities with it, and by grounding ML in social studies.

VotestratesML is not unique in including ethics when teaching ML. It is stated as future work for AlpacaML [246] and the MLM [120], and Hitron et al. [92] evaluate if their participants are able to reflect on ethical ML dilemmas after using the Gest system. However, none of these systems are designed with an ethics-first approach, and aim at teaching ML first, while treating ethics as a sort of after-thought. In contrast, VotestratesML was designed with the intention of foregrounding ML ethics and teaching ML as a means to that end.

VotestratesML does so, by presenting ML as a tool for predicting voter behaviour in social studies, making learning about ML a necessity for becoming a better social studies student. This provides the tool with relevance in a social studies classroom, even if teachers or students initially do not see ML as relevant for the subject. Iversen et al. [12] argue that being able to judge the relevance of a given technology is important for CE. When designing learning tools for CE, we argue that this relevance should be communicated in the design of the tool, allowing students to use the tool for achieving meaningful objectives for the subject and for themselves in general.

This approach seemed to increase students' motivation to engage in activities with ML. Instead of requiring students to "be good at math" or have an initial interest in ML, VotestratesML allows students to utilise ML for achieving meaningful objectives and advance their proficiency as e.g. social studies students, while learning about ML along the way. We argue, that the framing of CE learning tools as belonging in existing subjects is an effective way to engage students with different subject-related backgrounds and knowledge in working with technologies. Zimmermann-Niefield et al. [246] take a similar approach, by framing AlpacaML as a tool for physical education, but do not discuss the critical aspects of introducing ML in athletics with students.

A.6.2 Exposing the ML Process through Collaboration & Competition

The ML process consists of a series of steps involving human choices such as data gathering and analysis, selecting a model type, optimising the model, evaluating the results, etc. [56]. Other learning tools such as AlpacaML [246] and the MLM [120] aim to teach students about this process by letting them create and manipulate data in an iterative fashion. However, they do not engage students in the steps involved in creating ML models between data gathering and analysis and evaluating the model.

VotestratesML, on the other hand, aims at exposing the "messiness" of the entire ML process and to have students experience how it is based on human choices, judgements and trial-and-error. VotestratesML still black boxes some of the underlying ML mechanisms, but aims at exposing the ambiguous choices and the complexity embedded in the ML process. It supports this by allowing teachers to set clear and comparative goals for students (e.g., *build the best possible model for predicting if a voter will vote for the Social Democratic Party*), and by allowing students to explore and experiment with different strategies for selecting features, model types and model parameters through an interface that is designed for fast iterations and collaboration.

In our study, we found that collaboration and competition were effective means of engaging students, something Iversen et al. pose as a major challenge for CE [107]. The Collaborative Component of VotestratesML motivated and supported students in reflecting on and discussing different choices while improving their model, and each group seemed to take pride in their models, motivating them to keep improving upon it. The competitive aspect of VotestratesML motivated especially students who were not initially motivated by either ML or predicting voter behaviour.

In line with Iversen et al. [107], we find motivational factors important aspects to explore when designing learning tools for ML, in order to engage all students in the learning activities. However, discussions during group work were sometimes superficial, as the competition and improving their model, for some students, became more important than understanding how their decisions had affected their results.

This highlights the importance of striking a balance between motivational measures, such as competition, and encouraging reflection. While VotestratesML perhaps does not strike that exact balance it does provide insight into how students can be motivated to engage with and reflect on complex aspects of the ML process.

A.6.3 What a ML system is: Safely Exploring ML from a Concrete Starting Point

We argue, that the conceptualisation of ML (and other technologies) for CE learning tools is a central challenge for CE. Other research emphasises design and fabrication with technology as the core approach to empower youth [109, 134, 213]. However, when working with advanced technologies, such as ML, which rely on many other technologies and are integrated into larger infrastructures, fabrication may not expose all its implications. Instead, VotestratesML responds to recent calls on 'pulling back the curtain' of intangible and abstract computational systems [110], by providing a concrete starting point for discussing ML. VotestratesML aims to achieve this, by having students build interactive ML models, seeing how they fare when predicting voting behaviour, and discussing this from a social studies perspective. To support this, we designed VotestratesML to allow students to build and configure these models. VotestratesML encourages experimentation with different parameters and models, by only allowing safe actions that result in VotestratesML ending up in meaningful states. VotestratesML black boxes many underlying ML mechanisms, and prioritises students' exploration of

Area of ML	VotestratesML
What ML is	Safely exploring ML from a concrete starting point
The ML Process	Promoting iterative exploration through collaboration and competition
Implications of ML	Foregrounding ethics in the tool design and grounding ML in a well-known subject area

Table A.2: A summary of how VotestratesML embodies a CE approach to teaching ML.

higher level ML phenomena and their implications, by exposing the process of developing ML models.

Although students' conceptions of ML were challenged, we also observed how their understanding of ML helped them reflect on, discuss, and form opinions about the use of ML in society, and argue for these opinions using terminology from social studies theory and from ML. We have aimed to illustrate for high school students, that ML is not just a powerful tool that can improve their lives; ML models have both positive and negative implications, and it is important that we, as a society, actively engage in discussions about, and take a critical stance towards it.

Furthermore, we chose to use the terminology used in typical ML activities (features, labels etc.) and while most students were able to use many of these terms in discussions, these also revealed misapprehensions in students' understanding of the terms. We wonder if the use of these terms is an appropriate way to conceptualise ML in high school. It would perhaps be more apt to draw on concepts already known to students, such as dependent and independent variables, or it might be better to simply describe them as input and output. We encourage future research to explore what terminology is most appropriate for introducing ML in high schools.

A.6.4 Teaching ML from a CE Perspective

In Section 1 we argued for the need for a CE approach to teaching ML and in section A.2 we argued for the shortcomings of many previous tools from a CE perspective. While VotestratesML is by no means a perfect CE learning tool, it embodies our best efforts to design a ML learning tool with a CE starting point, summarised in Table A.2. This approach might not be ideal for optimising technical skills, but we argue that while these skills are important, the ability to take a critical stance, to discuss and reflect upon such technologies as ML are even more important skills for a majority of students. Thus, while we should not abandon traditional CT and computer science education, we agree with Iversen et al. [107] that CE is an important addition.

When exploring new ways of teaching ML designers and educators should consider the balance between the three areas of ML (see section A.2) and be especially mindful that teaching implications of ML does not become an afterthought, but instead central to the design of the learning tool/activity. We hope that our reporting on VotestratesML, its design and our findings from deploying it can contribute to future explorations of how to teaching ML from a CE perspective.

A.7 Limitations

This paper has a few limitations. First, the study has limited external validity: It involved only two social studies classrooms from the same high school in an upper middle class neighbourhood, making the participants a somewhat homogeneous group, and as the interventions took place during a regular social studies class, students might have been motivated simply by having a break from their regularly scheduled teaching. Finally, we acknowledge the risk of confirmation bias and of oversights in our data analysis due to only analysing the data deductively from the sensitising concepts that emerged during the design process and are aware that our findings are suggestive. We stress that the design VotestratesML is a specific proposal, that other tools for supporting CE might look different, and that studies with such tools might identify other, complementary opportunities and/or challenges.

A.8 Conclusion & Future Work

In order to promote a general AI literacy in high school students, we set out to explore *opportunities and challenges for a Computational Empowerment approach to meaningfully engage students in understanding and reflecting on ML* and have done so through a Constructive Design Research approach, by proposing and evaluating VotestratesML; a collaborative learning tool which embodies our approach for engaging students in learning about ML by allowing them to explore ML and its implications for democratic elections. The design of VotestratesML, is based on the notion that students should learn *what a ML system is, the ML process and the implications of ML* with a particular focus on the third area, to support them in becoming computationally empowered. With VotestratesML, we explored how to design ML-learning tools that qualify and support students in reflections and discussions about the implications of ML. Based on the design process and the deployment of VotestratesML in two social studies classes in a Danish high school, we have identified some key opportunities and challenges; Letting students explore ML and its implications from a concrete starting point and in a low risk way helped them reflect, discuss and form opinion about its use. We also found that VotestratesML's collaborative and competitive nature helped motivate students to explore and experiment with different options for improving their model. Finally, we found that designing VotestratesML as an ethics-first learning tool and grounding ML in a well-known social studies allowed students to have personally meaningful discussions and to advance their proficiency as social studies students.

VotestratesML exemplifies our approach to "pulling back the curtain" of otherwise intangible and complex computational systems in order to engage students with different subject-related backgrounds and knowledge in working with ML as well as other emerging technologies. We believe that the opportunities presented in this paper hold true for other emerging, computational technologies such as the internet of things, augmented/virtual reality, and blockchain. Further, we find that more research is needed on conceptualising these technologies for high school students to allow them to understand and discuss their implications, especially with regards to which concepts to black box and which to glass box and with regards to the terminology used.

A.9 Selection and Participation

The participants were chosen by their high school teacher volunteering the entire classroom. While most participants were legal adults, a few were under the age of 18. All participants were consented by their respective Social Studies teacher, who asked them if they would sign a standard consent form worked out by the employing university of the authors. All consent forms informed participants about how their data would be collected and stored as well as how to withdraw their consent. All participants under the age of 18 were asked for the consent of their parents or legal guardians. The consent forms were collected by the researchers at the beginning of the interventions.

Paper B

Staging Reflections on Ethical Dilemmas in Machine Learning: A Card-Based Design Workshop for High School Students

Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, & Marianne Graves Petersen

Abstract

The increased use of machine learning (ML) in society raises questions of how ethical dilemmas inherent in computational artefacts can be made understandable and explorable for students. To investigate this, we developed a card-based design workshop that allows students to reflect on ethical dilemmas by designing their own ML applications. The workshop was developed in an iterative process engaging four high school classrooms with students aged 16-20. We found that scaffolding students in designing meaningful ML systems served to qualify their ethical reflections. Further students' design processes allowed them to engage with the ethical dilemmas and to tie these to the properties of the technology and to their design decisions. We suggest seeing technology-close discussions about ethics as a goal in design processes, and prototyping as a means to ground these discussions in students' own design decisions, and we contribute a workshop format and design artefacts that allow for this.

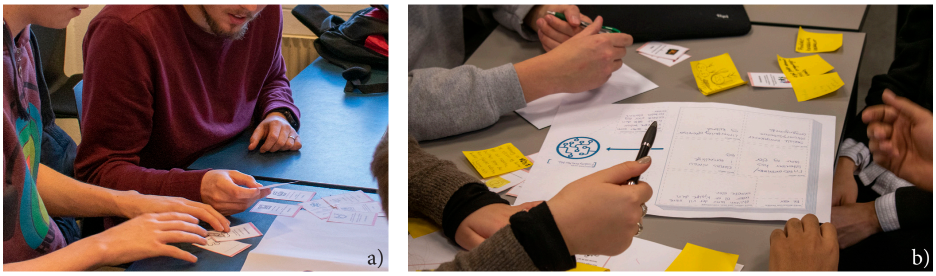


Figure B.1: *a)* A group of high school students discussion what data is relevant for their idea about a new ML system using a deck of ML data cards; *b)* A group of students using our ML board to describe how ML is used in their idea.

B.1 Introduction

The current focus on teaching Computational Thinking (CT) [234] across all levels of education, and the increased ease of incorporating emerging technologies such as machine learning (ML) into new applications raise the importance of investigating how ethical aspects of new technologies can be made explainable and explorable to students, who are the future designers and consumers of these technologies.

In recent years, ML has changed what is possible to achieve with the use of computational processing, expanding computers' ability to understand and interact with the world. This development promises great new possibilities, but is also currently transforming the nature of social interaction, work, education, etc. [151], increasing information asymmetry [173] and making technologies less comprehensible to users [1]. This introduces a range of ethical dilemmas and issues unique to ML. These ethical issues have been addressed in both the ML community through, e.g., the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), as well as in the HCI community [1, 77, 199] which implores that implementing ML into the world includes making ethical judgements and approaching ethical issues in the system design.

Today, companies are integrating ML into everyday technologies and infrastructures such as smart phones, maps, streaming services, etc., to provide better services to customers and to build new business models. However, if ethics are not carefully considered (and they often are not), ML can become a tool for unethical conduct, resulting in *dark patterns* [74], *hypernudging* [239], etc. Designers and developers are navigating a myriad of different interests, balancing ethics against system efficiency, profitability, user convenience and are constantly making moral decisions and value judgements. Dealing with ethics is difficult, and designers and users alike are often faced with dilemmas, when designing and using ML-powered products. We argue, that to take active part in shaping the future, understanding everyday technologies is imperative, and thus students should be able to recognise these dilemmas in technologies they interact with, and be able to reason about their consequences.

Within CT this more critical perspective is gaining momentum [107, 110, 216]. For instance, Computational Empowerment (CE), as proposed by Iversen et al. [107], advocates that students should be able to recognise the ethical choices and considerations in technology and learn to decode *"the consequences of these choices for the people who will use the technology"*. Iversen et al. address this by having children go through a design process which includes reflecting on the more critical aspects of their designs. This design process utilises digital fabrication techniques such as 3D printing and laser cutting, and while these are effective design tools, we argue that to deal with specific, technology-close ethical issues such as the ones described above, students need hands-on experience with the technology in question, e.g. ML. Involving students in design processes and giving them hands-on experience has been a part of CT from the start [171]. The purpose of doing so in CT has, however, been to turn technologies and computational concepts into powerful tools for the users, enabling them to approach problems in new ways and to explore new ideas. This focus on making powerful tools rarely leaves room for ethical considerations; instead CT often operates in simplistic *micro-worlds* [171], which are *"undisturbed by extraneous questions"* [171, p. 12], where ethical problems do not exist.

Inspired by CE, we explore how to involve high schools students in designing ML-systems in a way that supports exploration and reflection of technology-close, ethical dilemmas inherent to ML. To do so, we design and deploy a ML Ethics Workshop, in which high school students use a ML specific set of card decks for designing ML applications. The workshop places ethical reflections at the center of the design activities and discuss technology-close ethics with students based on their own designs of ML systems. We deploy four iterations of the workshop in four different classrooms engaging 66 high school students aged 16-20. The materials used in the ML Ethics Workshop can be found here: https://github.com/Karlo-Emilo/ML-ethics-workshop.

The paper contributes to the fields of HCI and CT, through *a)* a design rationale for an ethics-first construction kit for ML, *b)* experiences and insights from a Constructive Design Research process [128, 130] in which the ML Ethics Workshop was designed, and finally *c)* a discussion of the implications of using design processes for teaching students about ML and to reflect on and discuss technology-close ethical dilemmas of its use, based on findings from the ML Ethics Workshop.

B.2 Related Work

In this section we briefly review existing literature on using design as a learning approach, on card-based design methods and on involving teenagers in design processes.

B.2.1 Design as a Learning Approach

Design as an approach for learning about technology can be dated back to Papert's constructionism [86, 171]. Since then, other researchers have expanded on these ideas (e.g., [75, 112]) and today, digital fabrication, Fab-Labs and maker spaces are present at schools across the world [83, 168]. In a Scandinavian context, this approach has also been explored in recent years with a more critical approach to technology [59, 107, 204, 218]. The Fablab@school.dk project [107, 204] engages students in digital fabrication with different technologies to empower them to decode and understand their future, technology-mediated world [107]. To explore this Smith et al. present a design process model to use in educational contexts for using *"digital fabrication as a reflective and material tool for working with real-life and complex societal contexts"* [204]. Eriksson et al. [59] argue for the need of a wide *digital design literacy*, which aims to *"raise awareness about decision-making in technology design, the potential impact of technology and, ultimately, whether it contributes to meaningful relationships"*. The authors argue that this is best achieved in design processes, where students make design-decisions for real-world settings and learn to reflect on the implications of these decisions.

B.2.2 Cards-based Design Methods

In recent years, the use of card-based design methods and tools have accelerated, and today more than 155 different card-sets can be found in design literature [188]. The use of cards in design processes has been found to enhance user-driven design processes, since they can help structure and scaffold design processes[82, 152]. Cards provide a common object of interest to participants, and thus can also help engage all participants [220]. Friedman and Hendry use their *Envisioning Cards* [68] to bring a more human-centered focus into the design process. The envisioning cards are designed to work across different contexts and technologies, and implore designers to, e.g., consider children as possible stakeholders of a system or to consider how deliberately withdrawing from using a system might affect a (non) user's everyday life. *Situation cards* [149] are cards with descriptions of realistic and problematic situations at a workplace, which participants discuss in a participatory workshop to encourage them to come up with ideas, that solves everyday problems at the workplace. Another example is *inspiration cards* [82], which are used in design processes with disparate participants to improve engagement and support generation of innovative and realistic design concepts with new technologies, letting the participants create posters with the cards. Card-based design methods, that focus on a specific technology are few and far between [188], but there are a few examples, e.g.: *Tiles* is a card-based toolkit for designing internet-of-things applications [152]. The toolkit consists of different card categories and a paper-board, where participants can organise cards into meaningful IoT applications. Tiles supports reflection on users' creations through *Criteria* Cards, which are designed to evaluate ideas by discussing subjects such as *feasibility* or *sustainability.* Mavroudi et al. successfully deploy the cards in a lower secondary school setting [144], finding Tiles to be a low-cost way of having students, with little to no prior knowledge of the subject, design and discuss IoT applications. They also evaluate students' understanding of privacy issues related to IoT after using Tiles but never clarify how it supports learning and exploration of these ethical issues. Another similar, technology-close card-based method is PlutoAR [201], which is designed for K-10 classrooms to let students design and create their own augmented reality (AR) application. Here, cards act as AR-tags, and a paper-board, named the *Launchpad*, lets students organise the AR cards, which can be interpreted by a mobile application, that turns launchpads into scenarios, that play out in the application. As such, students are able to create applications, that e.g. launches and guides a rocket through a maze.

We agree with Eriksson et al. [59] that involving students in design processes is a useful way of teaching students about decision-making in technology. However, in contrast to the digital fabrication-based design processes in the Fablab@School.dk project, we argue that to discuss technology-close ethical issues about a specific technology, e.g., ML, students should also explore the limitations and implications of said technology through design processes, where they are accountable for their own design choices. To let students design ML systems without having to learn to how to code, we draw on the advantages of using card-based design methods. As other work has shown [144, 152, 201] using technology-specific cards in design workshops is an advantageous way of engaging students who lack, e.g., programming skills in designing with specific technologies.

B.2.3 Designing With Teenagers

In the Danish high school system, the average student is approximately between 16 and 20 years old, making most of students participating in this work teenagers. Fitton et al. [62] argue, that since teenagers are soon-to-be adults, they should be engaged and involved in the shaping of future technologies, which they will eventually become users of. Further, Iversen & Smith [105] argue that this involvement can be achieved by providing teenagers with *"meaningful alternatives to existing technologies"* by involving them in design processes of new technology. Involving teenagers in such processes is, however, challenging [62, 85, 104]. Hansen & Iversen [85] present an approach for teen-centric PD, in which motivation is based on *encouragements, tools, technology, identification, cooperation, endorsements as experts, and performance.* Similarly, Iversen et al. [104] seek to understand how teenagers are motivated in PD projects by analysing a number of PD workshops with teenagers, and propose similar specific steps to engage students. In the workshops reported on in this paper, we encountered a few issues that we believe are specific to working with teenagers. In particular, many students were boundary-pushing and extreme in the ways in which they discussed. These issues are discussed in a later section of this paper.

B.3 The Ethics of Computational Products

As products utilising computational processes and the infrastructures in which these are implemented become more complex, ethical issues in their design and implementation also become more complex [35, 66]. Computers' ability to process information about individuals more efficiently than ever has many benefits, but it also introduces new ethical dilemmas which often do not have clear-cut solutions [64]. Especially, because data has become important value assets to organisations, on which they build their business models and will go great lengths to collect [35, 173]. Software developers, engineers and designers in these organisations are constantly making value judgements, choosing what is morally right and wrong in every decision, counterbalancing different interests and the organisation's priorities [66, 91, 151]. These judgements are hidden in software and complex infrastructures, making them invisible for users of the organisations' products. This invisibility factor can be exploited for unethical abuse, but more importantly it makes ethical choices and dilemmas inherent in computational products difficult to reason about for users [151]. When talking about ethics in this paper, we will refer to these moral decisions made by individuals and organisations, sometimes referred to in literature as *micro-ethics* [91]. The implications of these moral decisions have only become magnified with the propagation of ML into everyday technologies, which has been articulated in the ML community in relation to fairness, accountability, transparency, nudging etc. [1, 239].

The ML Ethics Workshop is designed with the aim of making the moral issues and decisions understandable and explorable for students. Similar to many existing CT and fabrication tools [59, 112], our focus is to make students more capable of understanding today's technology-mediated society. However, where others focus on making computational concepts understandable [76] and turn technologies into powerful tools for students or to co-create meaningful futures [107], our focus is different. Our aim is specifically on making technology-close ethical dilemmas in ML-based computational systems understandable, and to allow students to reflect on the choices embedded in the products and services they use everyday.

B.4 Method

The work presented here follows the Constructive Design Research (CDR) methodology [71, 128, 130] to investigate how to make ethical dilemmas and choices inherent in computational artefacts understandable and explorable for high school students. We hypothesised (in the CDR-sense, see [7, 130]), that high school students would be able to have meaningful, technology-close discussions about ML ethics, by designing, reflecting on, and redesigning their own ML systems. In CDR-projects, hypotheses are instantiated through the creation of artefacts [7, 130, 205], and knowledge-creation is driven by experiments with and exploration of these artefacts [7, 128]. For this work, the center artefact is the ML Ethics Workshop, including cards and other hand-outs, which is presented below. Throughout the process, we have explored how different iterations of the workshop could be used to explore the above goal and the findings presented below come from our experiences with the different workshops. As denoted by Koskinen et al. [127] CDR may be accountable for several concerns of theory and practice. In this work we are concerned with expanding the scope of CT research, through exploring how design can be used in learning situations as well as with producing a concrete design method for making ML and its implications understandable and explorable to high school students. As such, we see the ML Ethics Workshop, including the workshop format and the set of artefacts, as a contribution in itself and as our main way for exploring the space of using design processes for teaching ML.

B.5 Machine Learning Ethics Workshop

This section describes the current version of the the ML Ethics Workshop, which serves to guides high school students through an ethics-first design process using cards and boards to guide their process. The workshop is designed to be used in all kinds of high school classrooms and it is not expected, that students have a special interest in, or knowledge about, ML. It has been deployed in one and a half to three hours workshops, but we believe it could be scaled up to a full day workshop. In groups, students design and describe a ML application for helping themselves and their peers in their everyday lives. Halfway through the process students are challenged with the ethical implications of implementing their application in a real world setting. To address these implications, they must redesign their application, dealing with ethical issues, which only become more complex as they discuss them, and weigh these against the functionality and user-experience of their application. Throughout the workshop, we work specifically with supervised, classification-based ML, the kind of ML typically used for e.g. recommender systems or image recognition.

This section will first present the rationale behind the workshop, then describe the design artefacts in terms of cards and boards used in the kit and, last, thoroughly describe the current format of the workshop.

B.5.1 Design Rationale

The workshop guides students through a design process where they confront the moral choices and ethical dilemmas of designing and implementing their own ML applications. First, students explore, through a design process, how ML can improve their own and their peers' lives. Here, students are guided to be as specific as possible in describing what data their application use, and what the ML component is predicting and how it is trained. This is to scaffold students' later discussions about technology-close decisions and dilemmas in the design and implementation of their application. When students have a well described ML application, which they believe can help their peers, the workshop changes character to focus on the ethical issues related to the application they just designed. By discussing questions about privacy, explainability, accountability, etc. students identify the most critical ethical issues in the design and implementation of their application. In this discussion students will identify the moral decisions they had already consciously or unconsciously made, and what questions they must further answer in their design process, before they can morally answer for their application. To let students experience, how these ethical questions often lead to individual value judgements or to choosing between undesirable outcomes, they are tasked with redesigning their application to address one or more of the most critical issues. In this task, students have to make choices about what is most important, striving for a functional and morally accountable application, which often turns out to be difficult. Last, students present their applications, and how they have addressed ethical issues in the design and implementation, followed by a classroom discussion. In these discussions there are no external 'bad guys' to blame, only students themselves and their arguments for their design choices.

B.5.2 Design Artefacts

Three decks of cards and a board for describing a ML system are used by the students in the workshop. The three decks of cards consist of respectively 14 *data cards*, 9 *ethics cards*, and 26 *people cards*, see figure B.2, all sized 6X9cm and colour-coded to communicate which deck they belong to. The *data* deck is used to analyse a context for possible data sources, exposing to the participants how data can be found everywhere, and to provide examples of a variety of different types of data and data sources. Each card describes a category of data sources (e.g. health data, news, users locations) and provides a few examples of specific data sources in the category (e.g. pulse, breaking news, time at a location) to help participants understand the category. The *people* deck is used to analyse the context in which the application will be implemented with focus on people, with each card describing a potential stakeholder (e.g. a colleague, a sibling, teacher, etc.). Participants use this card deck to identify who the application may affect, and how these people can be included in the considerations about the design of the application. The *ethics* deck is used for reflecting on ethical implications of implementing a ML-system into a context, asking the participants to consider ethical ML issues [1] (e.g. explainability, privacy, accountability). Each card asks the participants one or two questions to help frame a discussion around the issue in relation their design and ethics, e.g. *"What happens if your system makes a bad decision? Who is accountable?"*

The ML board, see figure B.3, is an A3 board with a visualisation of a supervised ML-model, leaving blank fields for the students to fill in when describing the ML component of their idea, forcing them to be specific about, which data their system uses and what it predicts. Participants first describe their idea and move on to describe the ML system: What the data in their system describes (e.g. a student, a kick to a ball, a dish), which data the system learns to predict (e.g. the students grade, the precision of the kick, calories in a dish) and up to four data sources, the system will use to make these predictions (absence, accelerometer data, weight). Last, they name their model, so they can reference it later in the design process.

Figure B.2: Two examples of (from left two right) data, ethics and people cards. Translated to English from the original language. Icons created by Wira Wiranda, Angelina, Becris and Danil Polsh from Noun Project and the people from OpenMoji.

B.5.3 The Workshop Format

The workshop supports groups of participants in designing a ML system and in reflecting on the ethical implications of implementing it into the world. It consists of nine steps: A short introduction to ML, a presentation of a case, six activities which the groups are asked to perform, joint presentations and a discussion of the participants' designs. The nine steps are described beneath:

1. *ML introduction:* A short introduction to supervised ML learning. Students are introduced to the predictive capabilities of ML, discuss a specific use of

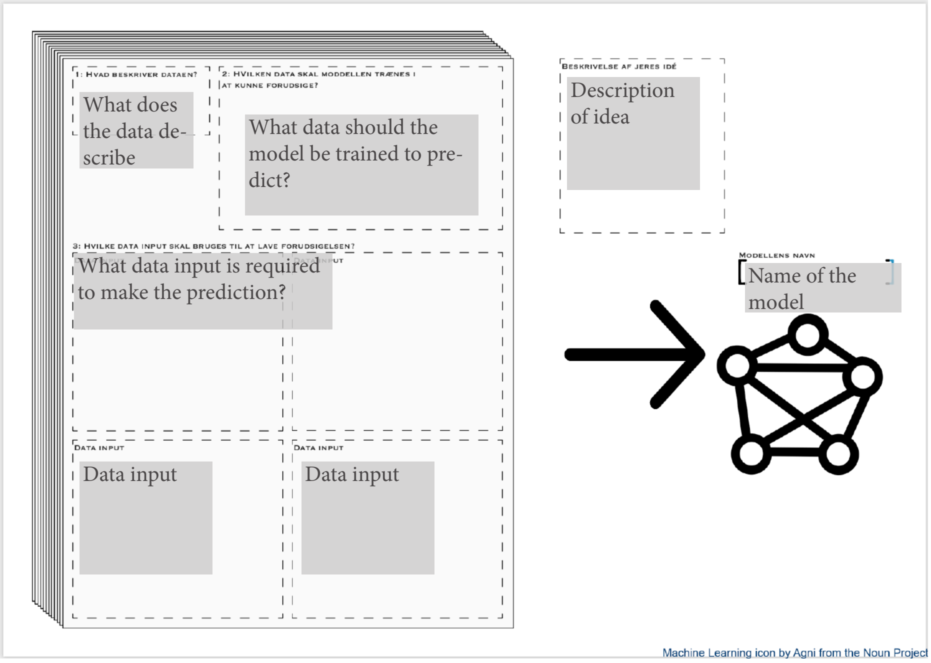


Figure B.3: Board for describing ML system. Originally in Danish and annotated in English.

ML; using ML in democratic elections, and gain hands-on experience with ML through an interactive ML-learning tool [118].

2. *Case presentation:* Students are given a short description of a narrative, which frames the design process with a context, stakeholders and an overall goal for the products the participants will be designing. It may also restrict the product design itself (e.g. it must be and app or a wearable). It situates ML in students' own life and let them explore, how ML can solve problems, which they find important and can provide new opportunities, which are valuable to them.

3. *Analysis of context for data:* Each group chooses a few cards from the *data* deck and use them to analyse the case context for possible data sources.

4. *Ideation:* Based on the case description and the exploration of data sources, each group conducts a short IDEO style ideation [150], where they come up with as many ideas as possible. At the end, each group is asked to choose the best idea (or combine multiple ideas into one great idea).

5. *Description of ML system:* Each group use the ML board to describe the ML component of their idea. It helps participants become very specific about their use of data and ensures all ideas are based on ML. If it is not possible for a group to fill the board, they need to go back and revise their idea.

Workshop	**#1**	**#2**	**#3**	**#4**
Length	3 hrs	1.5 hrs	2 hrs	2 hrs
Location	University	Own classroom	University	University
Background	Technology & design	Social studies	IT	Informatics
No. students	24	25	5	12

Table B.1: Overview of the different workshops as well as the classrooms whom participated in them.

6. *Reflections on ethics:* Each group is given an ethics card and the entire people cards deck, to guide them in reflecting on possible implications for different actors when implementing their ML application. They choose one issue, which they find most critical about their system.

7. *Redesign product:* Each group discusses how they can address the ethical dilemmas in the design of their product and describes in detail, through sketches, texts, etc., how their system should be redesigned to address this issue.

8. *Presentation:* Each group prepares a one minute presentation, which describes their concept, their ethical issue and how they approached it in their design. Every group gives their short presentation, which is followed by a short applause from the other groups.

9. *Discussion:* The students' designs are used in a joint discussion about ethics, grounded by the ethical problems faced by the participants in the design process and how they tried to solve them.

All groups are given a limited amount of time to complete each activity to ensure progression in the design processes. Activities are presented and described just before they start to make participants focus on the given activity and not the whole process. The internal collaboration in the groups is left informal, but mediated by the artefacts.

B.6 Designing the ML Ethics Workshop

This section shortly describes the purpose, design rationale and the insights from each workshop that influenced the design of following version of the ML Ethics Workshop. The findings from the workshop will be presented in the following section. The ML Ethics Workshop was developed in an iterative CDR process and each iteration of the workshop was deployed in a high school classroom. A total of 66 students, aged 16-20, participated in the workshops. In total, four workshops was made in four different classrooms with different backgrounds and from different high schools. All classrooms were volunteered by their teacher and participated as part of their high-school education. Each student was, however, consented individually according to Danish regulations. Workshop 2 took place at a local high school in the students' regular classroom. All other workshops took place in a classroom at university of the researchers, and was conducted

Artefact	**Example**	**#1**	**#2**	**#3**	**#4**
Ethics Cards	See figure B.2, left	X	X	X	X
Data Cards	See figure B.2, middle	X	X	X	X
Humans Cards	See figure B.2, right	X	X		X
Context Cards	*"Gaming: How can [ML] be used to... - improve you performance? - create better habits? - strengthen you friendships? ..."*	X	X		
ML Board	See figure B.3		X	X	X
ML Cards	*"Feature: A characteristic of the phenomenon we observe. The independant variable."*	X			
Output Cards	*"Emoji: Use emojis to express emotional outputs."*	X			
Interaction Cards	*"Smart Watch: Use a smart watch to interact with the system."*	X			

Table B.2: An overview of the different artefacts used throughout the different versions of the ML Ethics Workshop. The right side of the table describes in which iteration of the workshop each artefact was used.

with classes visiting the university. This accounts for the variance in numbers between Workshops 1, 3 and 5. An overview can be found in Table B.1.

All iterations of the workshop followed the same basic structure; students were given an introduction to ML, next they were given a design case to work on and were split into groups of 4-6 people to do so. After working through different design phases, students presented their ML application and the ethical issues they had been discussing. Table B.2 provides an overview of the artefacts used in each workshop. In the following, the artefacts will be referred to by their name in the table.

B.6.1 Workshop 1: Testing An Idea

In the first workshop, we tested the first version of the Ethics Workshop, which were our initial approach on how to design a workshop for staging ethical dilemmas. The workshop was inspired by how the Tiles Toolkit [152] uses different technology and interaction cards, a board where the cards can be organised and a playbook to guide a design process, and how Envisioning Cards Toolkit [68] stages discussions about human values through design activities. We brought context cards with different design contexts and possible directions for the design process. The contexts were based on students' everyday lives, e.g. sport, dating, school. To scaffold the design process, we brought six different card decks (see figure B.2), post-its and A3 boards. The main idea was, that students would come up with an idea for a ML application, that was valuable to them, use the application to discuss ethical issues of ML systems and address some of the ethical issues in the interaction design of the application. This workshop was carried out with 24 students, who were divided into two groups of four, two groups of five and one group of six students.

INSIGHTS:

Students were engaged throughout the workshop, and seemed motivated by the many cards and the possibility to come up with their own ideas. The amount of cards did, however, seem to hinder their reflections about the activities, as students were more focused on sorting and getting through all cards in each deck, than discussing a few cards in depth. The ML cards did not support the students very well in describing a ML system; they seemed to need more structure to design meaningful ML systems. As a result the ethical discussions often lacked technical depth. Students were good at identifying overall ethical issues, but perceived their own ML applications as either morally good or bad, based on their own intentions in the design process, with nothing much to do about it.

B.6.2 Workshop 2: Describing a ML system

In the second workshop, we provided additional scaffolding for the students' design of ML systems by introducing the ML board, which can be seen in Figure B.3. The board visualises a supervised ML system, which students must fill in based on their ML application. Furthermore, the workshop was simplified to only contain data, ethics and people cards to make more room for immersion and reflection. This workshop was carried out with 25 students who were divided into five groups of four students and one group of five students.

INSIGHTS:

The workshop was conducted in a social studies classroom, and students struggled with the openness of the design case, which was based on the context cards (see Table B.2 for an example) and the requirement of integrating ML. Many groups spent most of their time on deciding on an idea. Few groups were able to fill in the ML board correctly, but the groups who did succeed were more nuanced when reflecting on the ethical issues in their application.

B.6.3 Workshop 3: Combining Design and Ethical Reflection

The third workshop explored how to make more room for ethical discussions, and how to expose moral choices early in the design process. To simplify the workshop further, the *people* deck was removed and each group picked only two data cards and were provided with one ethics card. The intention was to make more room for discussing each card in depth and to ensure, that each group worked with different types of data and ethical issues, in effort to make the classroom discussions more diverse. To ease the ideation and conceptualisation process for the students, the context cards were replaced with a shared design case about helping lonely fellow students, providing students' ML application with a specific purpose with societal relevance, and limiting the design space to mobile applications. Last, students were asked to redesign their systems using wireframe sketches of their application to encourage them to become more specific in their design. They were asked to identify the most critical ethical issue in their application and sketch, how they would address this in the interface. The workshop was carried out with five students who participated as a single group.

INSIGHTS:

This workshop was conducted with only five students, who worked in a single group, but this version of the workshop was also used in Workshop 4 with only limited changes. The predetermined design case and the fewer data cards seemed to help students generate more well-described ideas, and to make it easier for them to fill in the ML board. The group's ethical discussions had more depth than any group discussions in earlier workshops, and dealt with the complexity of coming up with good solutions to ethical issues.

B.6.4 Workshop 4: Validating The Concept

In the fourth and final workshop, we tested the composition of Workshop 3 with more groups. The only change being, that the people cards were reintroduced to ground the ethical discussion around the people, who would be affected by the introduction of the students' applications. The workshop was carried out by a classroom of 12 students who were divided into three groups of four students.

B.6.5 Data Collection & Analysis

Designing workshops for discussing ethics in computational products in high school classrooms is a rather unexplored area. Therefore, data was collected and processed using open ended approach. Throughout all workshops, data was collected using observations, sound recordings, photography, video and by collecting the artefacts produced by the students e.g. paper-based mock-ups of their designs. All participants in the workshops were asked for their consent, and only data about consented participants was collected. After each workshop, a write-up [145] of field-notes and observations was produced and discussed between the researchers present at the workshop. The write-ups focused on how students used the design artefacts in the design process, how they approached each step in the workshop, how they talked about and understood machine learning, and how they discussed ethics in relation to their idea. Together, all three authors analysed the data above, focusing on emerging themes and critical incidents, by collaborating on creating an affinity diagram from the data. Selected audio recordings of students' internal discussions were transcribed and together, the three authors analysed them with focusing on, how students identified and talked about ethical issues, and how they came to an agreement on how to approach ethical issues in their design. Photography and video were reviewed with focus on how students collaborated around the design artefacts and how they presented ethical issues in their application for the classroom. The artefacts produced by students were analysed with focus on how they described their ML system using the design artefacts and, again, how they identified and approached ethics in their designs.

B.7 Findings

In this section, we present our findings from the workshops with regards to how they were able to engage students in reflections about ethics related to ML-based systems. These findings are synthesised from an analysis across different data types and workshops as described above.

B.7.1 Understanding of ML Served to Qualify Reflections

Throughout different workshop iterations, we experimented with how students could be supported in describing ML systems that conceptually and technically made sense and could, to some degree, be implemented in the real world. In Workshop 1, we gave students cards that represented specific parts of a ML pipeline (e.g. features, label, training), and asked them to annotate and combine cards to describe their systems, as seen in Figure B.4. This was effective to the extend that all groups used the cards to describe their ML system, but we observed that they had a hard time describing meaningful systems, and in most cases the ML component of their designs were described as a magic touch that was sprinkled on top. In the example in Figure B.4, a group of students designed a system for optimising how their school's rooms are scheduled to best fit the needs of different subjects, teachers etc. The group presented the ML system as being able to *"predict the best schedule for the school"*, but it was unclear how exactly ML would help to solve the issue, i.e., how would previous schedules be evaluated as better or worse and what data would this be based on? In the subsequent discussions on ML ethics, this became a hindrance to the students, who were only able to shallowly discuss the issues of the system. To deal with possible issues about algorithmic decision making, their system would schedule classes and study trips *"in a humane and stress free way"*, which is a good intention, but does not deal with the underlying technological issues or describe how they would achieve this. In a similar example, a group designed a ML system for helping athletes improve their workouts. The group discussed whose responsibility it was, if the system recommended an exercise that caused an injury, but their solution was to employ *professional testers* to prevent this from happening. It was, however, unclear how these testers would prevent the system from recommending wrong exercises, even if all exercises had been tested. And it is also unclear *how* the testers would be made responsible for bad recommendations.

Following this finding the board seen in Figure B.3 was introduced to scaffold students' design of the ML components in their systems. Instead of giving them the pieces for an ML-system, we provided them with the basic structure of one, and asked them to fill it in. This approach seemed to better support the students in designing conceptually and technically meaningful systems. Not all systems were sound, and students struggled especially with the specificity of data inputs and which data to predict. In contrast to the first workshop, many students were, however, able, at least to some extend, to create meaningful ML-systems. In the subsequent discussions on ethics, these students had more detailed discussions about their system's ethical issues. For example, in the third workshop, a group designed an app for recommending social groups to lonely teenagers based on their interests. The students were given an ethics card, which asked them to question the responsibility of the algorithmic decisions in their system, and based on their design, they were able to have in-depth and specific discussions about responsibility:

> **Student 1:** "What if a group, where something is going on under the radar is added to the system by us [...] and something bad happens in that club, because we didn't know there was a problem with them. Whose fault is it then?"
>
> **Student 2:** "It is still the responsibility of the club"
>
> **Student 1:** "But we brought them into our system..."
>
> **Students 3:** "I think, those who are responsible for the app are to blame. And I'm thinking that there is always someone in charge. That's a director's job"
>
> **Student 4:** "Okay, that is the Minister of Health. So it should be the job of the Minister of Health or what?"(The Ministry of Health, was the product owner of the application in the design case)
>
> **Student 1:** "I don't think we should take it that far up. We need a specific division to monitor groups."
>
> [continued discussion]
>
> **Student 3:** "Then we should have a rating system or the ability to report groups, and then someone should review them afterwards [...] Someone should always be to blame."

Their discussion illustrates how it is difficult to place responsibility once something goes wrong in ML systems, and as is discussed below, these students were able to design a system that (in a basic way) addressed this issue.

B.7.2 The Design Process Tied Ethics to Design Decisions

We found that by tightly controlling the design process, students were supported in connecting their discussions about ethics to their designs, and they were able to use their discussions in the redesign of their systems.

This was, however, less the case in the first workshops. In Workshop 1, a group of male students designed a ML system to rate and sort women according to their breast size. This showcased (in a rather extreme way) the approach taken by many students when designing their ML systems in the first workshops. While we provided contexts for the students in Workshop 1 and 2, many students came up with ideas, that they found interesting for themselves, without considering how they would impact others, and if they (and others) would actually benefit from the idea. In this way the context of "Social Media" became the breast-application and the context of "Urban Life" became "Club Counter" where an AI would recommend clubs based on its clientele. These groups did not do well in discussing the ethical issues of their designs: They could easily identify ethical issues in their designs, but it was difficult for them to address these in their redesigns, as they could not account for why their application should exist in the first place. Indeed, the group wanting to rate women simply answered "No" with no further

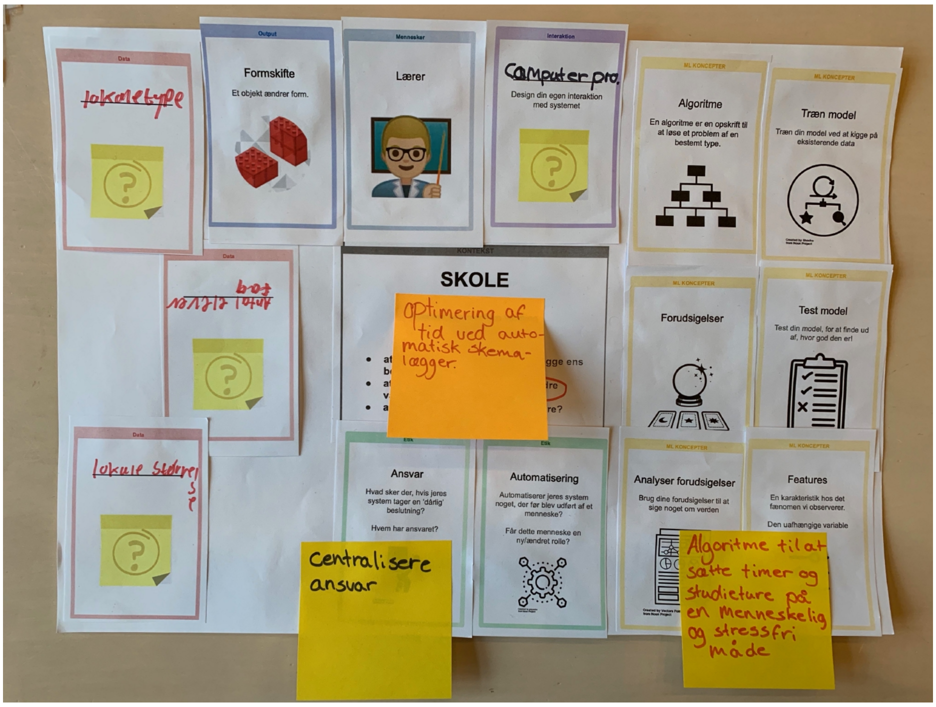


Figure B.4: A ML System, designed by students in workshop 1, for optimising scheduling and booking rooms in a school.

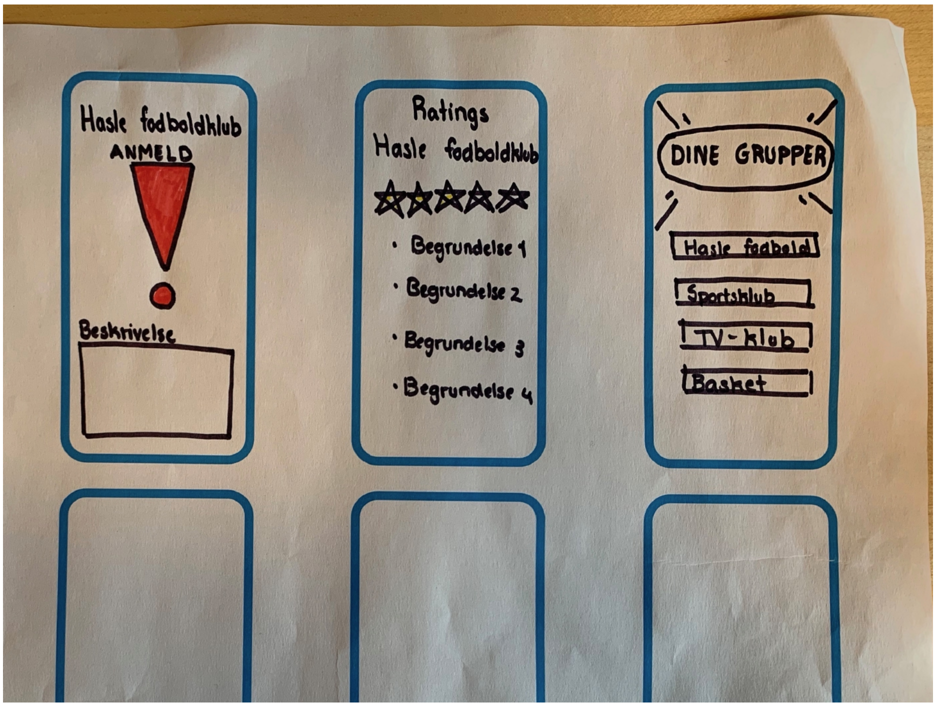


Figure B.5: Wireframe sketches of an app, designed by students in Workshop 3, for recommending groups to lonely teenagers based on their interests.

elaboration to the question: "Is it ethical to implement this?" when asked to reflect on the ethical aspects of their design.

To support students in designing systems that could be meaningfully discussed (and were not degrading), we introduced the case about helping lonely students in Workshop 3, and asked students to consider and sketch the ways in which their system addressed their discussions about ethics. This approach seemed to better tie together the students' design process and their discussions about ethics, e.g., in the group described above, who designed a system for recommending groups for lonely teenagers. Based on their discussion, they decided to implement a system for rating and reporting groups (see Figure B.5). While this is a traditional (and potentially harmful) way of dealing with responsibility, the students were able to identify the issues in their discussion (see above) and design a solution addressing the issue.

In Workshop 4, a group designed a similar app for predicting "soon-to-be" lonely people and pairing them up with each other. This group discussed the ethics of data collection, and the dilemma of the app *"giving away other people's loneliness"* by pairing them with strangers (see Figure B.6). This group designed the app to include a "My Data" page, in which the user can see exactly which data is given away, and can choose to remove data they do not want the app to include in its recommendations.

B.7.3 Challenging to Address Ethical Dilemmas in Design

Throughout the workshops we found that students struggled with addressing the ethical dilemmas they had discussed in their design. Students, who were able to have qualified and interesting discussions about ML ethics, found it difficult to come up with good solutions to their issues, and often designed systems in which users (themselves and their peers) were responsible for system faults or data collection. Many students were trapped in a dilemma of privacy; that their application and the quality of their solutions depended on access to personal data from their users, and for groups who had discussed the ethics of data gathering, the design solutions was often akin to traditional "Terms and Conditions"-agreements that most applications and websites now have (see Figure B.6). As described above, some let users control what data would be included in predictions, even if this would lead to worse recommendations. In Workshop 3, the group who reflected on accountability in their application for lonely people, also discussed privacy issues and concluded that users should be incentivised to provide as much data as possible since this would provide better recommendations or as one of the students put it: *"The more lonely they are, the more motivated they become to get better recommendations"*. Once again, this illustrates the dilemma of choosing between different interests, which is also present in real-life ML applications.

We did not expect students to be able to solve these issues as professional designers are struggling with this too. Instead, we aimed for the experience to illustrate for the students how difficult it is to design and implement technology that avoids ethical issues, even if you are well-intentioned and want to change the world for the better.

B.8 Discussion

As argued above, many existing approaches for using design processes in CT focus on empowering students to become creators [171] or to discuss broad, societal implications of technology [107]. These approaches often neglect discussions that are both close to technology and to its implications. In this section we reflect on our workshop format, based on the findings from the workshops, and discuss how it can inform future workshops or other activities with similar objectives of staging reflections on ethical dilemmas in computational systems and products.

First, we argue that the entire design process should be framed to help students design *ethically* interesting systems. In workshops 1 and 2 students were, to a large degree, responsible for framing their task and were only provided the context card (see Figure B.2) to help them. We found that students would often test our and their peers' boundaries. The system for rating women by breast size is a glaring example, and throughout workshops, many students used rather extreme examples to illustrate their considerations, such as illustrating a system failure by suggesting that a 6-year old user would be recommended to join a sex-club or discussing what would happen if a terrorist group accidentally ended up in their application. One of the characteristics of teenagers that set them apart from children or adults is their focus on developing identities and social roles [14], and in this light, their testing of boundaries is not surprising. We did, however, find that it would often make ethical discussions pointless (e.g. the breast size rating

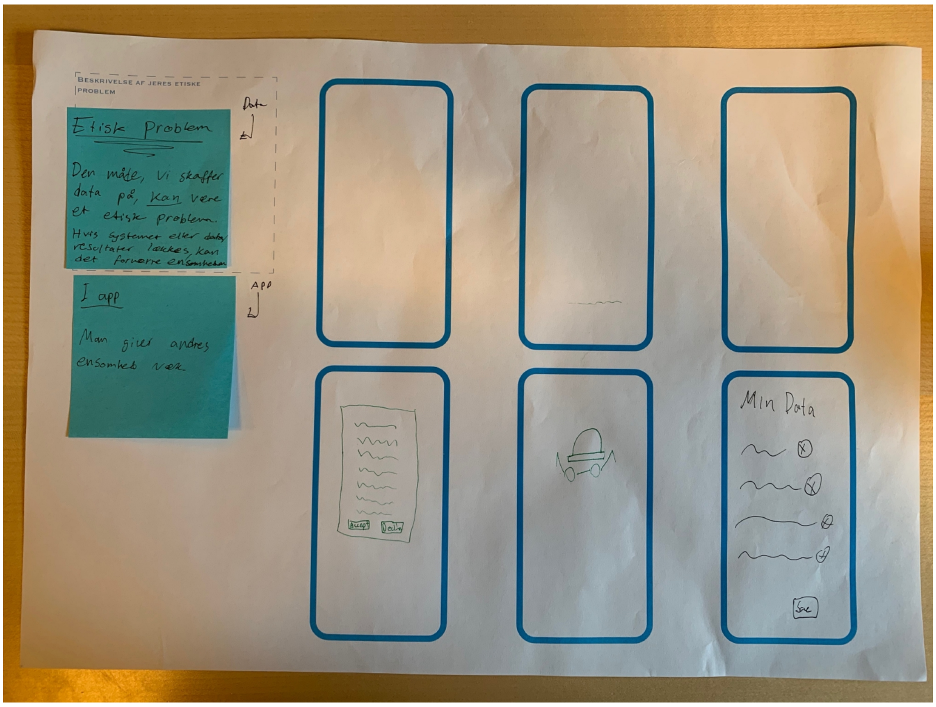


Figure B.6: An app, designed by students in Workshop 4, for predicting loneliness an pairing people before they become lonely. The sketching show how users must agree to the app's terms and conditions, but can also choose what data the app bases its recommendations on.

system). One immediate response could be to stop the students in their track and discipline them, but we argue, that this would be inappropriate, keeping in mind our goal of staging reflections. Instead we found it helpful to provide a more specific case for the students, to scaffold them in designing systems that would better support their discussions. Starting from workshop 3, we provided students with a case about loneliness, which is an actual issue among Danish teenagers, meaning that students were designing for a real and vulnerable demographic and were encouraged to consider how this demographic is affected by their design choices.

Second, we argue that the focus of design workshops should be moved from the possibilities of a technology towards the implications of implementing said technology in real-world settings, when aiming to support critical discussions and reflections about technology. We saw that when we supported students in designing *implementable* ML systems, they seemed to have more insightful discussions about its implications. One reason for this, we argue, is that design processes are typically future-oriented in the sense that we as designers are interested in what-could-be. However, for a technology-close discussion about implications, it is necessary to look at the specific here-and-now issues of a technology. In turn, this implies a change in the criteria for success of such a design process. Our inclination as designers is to look at the products of students' design processes; are they able to successfully communicate an idea or a design, do they address the case provided to them, are they solving an actual issue? These are questions that might be used to evaluate typical design processes, but as the goal has changed, we argue, so should the questions. Instead of focusing on prototyping as an end in itself, we suggest seeing insightful discussions about ethics as a goal in itself, and prototyping as a means to ground these discussions in students' own design decisions. We suggest asking questions akin to the following to evaluate if a design process is successful; are students able to formulate a technically feasible system; can they identify how their idea and design choices relate to specific, technology-close ethical issues; are they able to discuss these with their own idea as the point of departure; and can they address their considerations when prototyping or modifying their idea?

B.9 Limitations

Since students came from different backgrounds, and many students were already interested in technology their existing knowledge of ML varied. In addition, because of the differences in length of the workshops (see Table B.1), students received slightly different introductions to ML and students participating in Workshop 2 had already had a longer introduction to ML through another workshop [118].

B.10 Conclusion

In this paper, we have explored how ethical dilemmas and moral choices in ML applications can be made understandable and explorable for students. To do so, we designed a workshop format, and several artefacts for use in the workshop, in which students confront the ethical choices and dilemmas of designing and implementing their own well-intentioned ML applications. We conducted four workshops in different Danish high school classrooms, refining and iterating on the workshop based on insight gathered in the previous workshops.

Experiences with the workshop format illustrate how an understanding of fundamental ML and how iterating on own design ideas helped to qualify students' ethical reflections. In addition, the design process served to reveal to students the complexity of the ethical issues in ML systems, and tie them to the properties of ML and to design-decisions. Grounding ethical discussions in students own designs, made them accountable for their choices and illustrated, how difficult it is to come up with clear-cut solutions to these ethical issues. Based on these findings, we recommend that in order to scaffold technology-close ethical discussions, workshops and design processes should focus less on technological possibilities and more on implications and consequences, and we have presented our workshop format and artefacts as an approach for doing so.

We hope that future research in CT will address the ethics in computational products and systems and aim to make technology-close ethical dilemmas and decisions explorable for students.

B.11 Acknowledgements

This work was supported by a research grant (#28831) from VILLUM FONDEN. The authors would like to thank the participating students and teachers from Aarhus Gymnasium, Aarhus Statsgymnasium, Marselisborg Gymnasium & Slotshaven Gymnasium. The authors would also like to thank Daniel Graungaard for assisting with executing the workshops, and all researchers in the Center for Computational Thinking and Design at Aarhus University for providings comments and feedback during the design process.

Paper C

The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning

Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen

Abstract

Machine Learning (ML) is often used invisibly in everyday applications with little opportunity for consumers to investigate how it works. In this paper, we expand recent efforts to unfold what students should know about ML and how to design tools and activities allowing them to engage with ML. To do so, we explore how to make processes and aspects of ML tangible through the design of the Machine Learning Machine (MLM); a tangible user interface which enables students to create their own data-sets using pen and paper and to iteratively build and test ML models using this data. Based on insights from the design process and a preliminary pilot study with the MLM, we discuss how a tangible approach to engaging with ML can spur curiosity in students and how the iterative process of improving ML models can encourage students to reflect on the relation between data, model and predictions.

C.1 Introduction

Machine Learning (ML) is increasingly used throughout many of the applications we use everyday, but the processes involved in ML are inherently hidden for users and are often difficult to comprehend [1, 30]. The advancement of ML systems has changed what can be achieved with computational systems, and ML is being

Figure C.1: *a)* The two components of the Machine Learning Machine; the Trainer (left) and the Evaluator (right). *b)* A user labelling a drawing using the Trainer to train a machine learning model.

implemented in applications, systems and infrastructure all around us. However, ML involves several challenges, such as inherent biases [164], *dark patterns* [74] and *hyper-nudging* [239]. As a result there is a growing focus on how to design explainable and accountable AI and ML-based systems [1, 209], and support a widespread understanding of ML in order to empower citizens interacting with systems that use ML [1].

For several years, the inclusion of Computational Thinking (CT) [234] in formal education has been steadily growing, aiming to prepare students for careers in computational fields. In recent years, a more critical perspective on the use of information technology in society has been finding its way into the CT curriculum [50, 107, 114], arguing that students should not only be able to understand and use technology itself, but also be able to critically reflect on the implications of digital technologies for their own lives as well as societal implications. This is a challenge for all technologies, but is especially relevant for ML due to its inconspicuous and pervasive characteristics [173]. Consequently, ideas about what students should know about ML have begun to take form; Touretzky et al. [216] present five *Big Ideas* about what K-12 students should know about artificial intelligence and subsequently ML; these include 1) Computers perceive the world using sensors 2) Agents maintain models/representations of the world and use them for reasoning 3) Computers can learn from data 4) Making agents interact comfortably with humans is a substantial challenge for AI developers and 5) AI applications can impact society in both positive and negative ways. Basic ML processes include in particular Touretzky's idea #2 and idea #3, and it is around these two ideas that many of the inherent issues of ML are to be found; data is often flawed which might lead to bias issues, and models often get too complex to understand and therefore difficult to question [164]. In order to teach these ideas to students, we need ways of making them understandable and graspable; Bilstrup, Kaspersen, and Petersen [19] illustrate how providing opportunities for students to explore the basic processes behind ML potentially empower students to reflect on the possible personal and societal impacts of ML. In line with this, recent learning tools for teaching ML allow children to engage with the processes of ML in terms of data gathering, training and evaluation [51, 92, 193]. Hitron et al. [92] and Zimmermann-Niefield et al. [246] both present systems that require learners to perform gestures to generate data via sensors and train an ML system based on this data. Any-cubes, by Scheidt and Pulver[193], is a system consisting of two cubes, where one cube is used to train recognition of physical objects which activates the other cube controlling an actuator.

Previous work suggests that making technology manipulable and explorable through Tangible User Interfaces (TUIs) can be an effective way of teaching difficult and otherwise intangible phenomena [61, 140, 142, 143]. Marshall, Price, and Rogers [143] argue that TUIs are particularly suitable for presenting different models[1] about the world to users, and letting them explore these models in a tangible way. In the CT community, tangible systems have been used extensively by researchers, based on the ideas of Constructionism [170]. CT researchers design "Toys to Think With" [171], arguing that learners gain new insights by combining their existing knowledge with working with constructing artefacts, computational or otherwise. This strategy has been used for teaching different computational subjects such as programming (e.g. [97, 99]) and algorithms (e.g. [48, 69]). However, while some work makes certain aspects of the ML process tangible [51, 92, 193]; we were not able to find any work, that systematically explores how the full ML process can be made tangible.

In this paper we present an exploration of how to make the processes and aspects of ML tangible, in order to teach students about ML and in in this way contribute towards a more widespread understanding of ML [1]. This results of this exploration are manifested in the Machine Learning Machine (MLM), as seen in Figure C.1a. MLM is a TUI consisting of two separate devices, which allow K-12 students to create data and train and test ML models. The first device, the Trainer, is an industrial-looking machine, allowing learners to train models to recognise student-made drawings by feeding the machine with a number of drawings and labelling these. The second device, the Evaluator, resembles laboratory equipment and allows learners to test their models by placing new drawings under its camera and having it try to recognise them. Learners can switch between the two devices by moving a physical representation of their model between them, and in this way iteratively try to improve their model. By training ML models on drawings made by students, the MLM aims to support students experimenting with training ML models, and to support teaching ML fundamentals as well as more complex tasks in which students explore, e.g., data representativity and unintentional bias. The novelty of MLM is not the tangible user interface per se. Instead it is the manifestation of how tangibility can support students' exploration of the complex phenomena of ML.

The paper contributes to the fields of HCI and CT, through *a)* The MLM itself as a design exemplar of how ML processes can be made tangible, *b)* a design rationale for a TUI for making ML processes manipulable and understandable, *(c)* preliminary findings from a small pilot study conducted with MLM, and *d)* a discussion of and suggestions for how TUI's can scaffold learning and reflection about ML processes.

[1] The authors refer to any kinds of models about the world and not particularly to ML models although the coincidence is striking.

C.2 Related Work

In this section we present previous work related to TUIs for learning and teaching ML in educational settings.

C.2.1 Tangible User Interfaces for Learning

Since the Constructivism of Piaget [179], the body has been a major factor in the way learning and subsequently teaching have been understood. Papert's later notion of Constructionism and LEGO Mindstorms [171], argues for experiential learning where learners explore different manipulable artefacts thus piecing together ideas and knowledge to form new knowledge. Hence, the embodied learning approach has moved into the computational domain.

The educational potential of tangible systems has also been recognised within the TUI community. Fails et al. [61] compare how a desktop version and a tangible version of a game teaching children about health hazards affect children's learning. They conduct a survey based on the two modalities to investigate the difference in several metrics such as number of interactions, answer depth, interest in the subject, etc. Although the survey was conducted with only a few students, they found that students who had used the physical version of the game generally gave more in-depth answers and became more interested in the subject. Marshall, Price, and Rogers [143] present a scheme for conceptualising TUIs for learning and analyse several tangible artifacts using their conceptualisation. First, they suggest that tangibles should be analysed in terms of being 'ready-to-hand' and 'present-at-hand', the terms coined by Heidegger [90] to describe how a user relates to a tool she is using. When a tool is 'ready-to-hand' the user acts through it, upon something else, making the tool 'invisible' to the user. When a tool becomes 'present-at-hand' focus moves to the tool, making the user aware and reflective about using it. Marshall, Price, and Rogers argue that effective learning requires these moments of taking a step back, and reflecting on the learning experience in a more cognitive way. Second, Marshall, Price, and Rogers propose two main categories for tangible systems; expressive and exploratory. Expressive systems allow users to produce an external representation or artefact whereas exploratory systems embody models presented by someone else and allow users to explore these in tangible ways. Finally, they suggest that learners can explore such models in either theoretical or practical ways. Systems that allow theoretical exploration embody a theoretical model about the world, whereas practical systems allow users to explore the inner workings of the system itself.

C.2.2 Machine Learning for Young People

There exist several tools for teaching ML to young people and children in all ages. Several block-programming languages now support building ML applications; The AI Snap! blocks [115] is an extension of the popular Snap! block-programming language, allowing students to use a set of predefined ML blocks to build ML applications. The Scratch programming language has been utilised in systems such as *Cognimates* [51] and *Machine Learning For Kids* from IBM[2], which allow children to build simple systems utilising speech recognition, object recognition,

[2]https://machinelearningforkids.co.uk/

etc. Williams, Park, and Breazeal[231] argue for an *Early AI Literacy* and present PopBots; a social robot made of LEGO blocks where preschool children train a ML system on a tablet to control the robot's behaviour. Outside of research, several commercial systems also exist that makes ML explorable; The *Cozmo* robot[3] can use object detection to play games with children and allows them to program different behaviour using a block-programming language. The Teachable Machine by Google[4] also allows learners to experiment with ML by e.g. uploading sets of images for training an image recognition system or to use their webcam to make a gesture recognition system. Computational Notebooks [162], for example the popular *Jupyter* Notebook [174] or Google's Collab Notebooks[5], enable more advanced learners to build their own ML systems from scratch using programming languages such as Python.

TUIs for teaching Machine Learning

TUIs have been used by several researchers for teaching different computational concepts such as programming [97, 99], augmented reality [201], internet-of-things [144], block-chain [121], robotics [231] as well as for teaching Machine Learning.

Hitron et al. [92] and Zimmermann-Niefield et al. [246] both present systems that rely on recording and training ML models based on bodily movements, trying to recognise correct movements. Hitron et al. present Gest [92], a gesture recognition system where children, aged 10-13, train a system by drawing simple geometric shapes by performing gestures using an input device designed by the authors. Through a follow-up interview, they find that the participating children were able to apply their new knowledge of ML to their own life and to think up personally meaningful applications using ML. Zimmermann-Niefield et al. [246] present AlpacaML, a ML system to recognise good athletic moves such as kicks in a game of football. The authors deploy the system in a workshop with a local high school class and found that the young people succeeded in collecting data, building, testing and evaluating models as well as iterating on the process to improve their models. Gest and AlpacaML make the process of data collection tangible, but still rely on respectively a desktop and a mobile application for further inspection and evaluation of the models. AlpacaML allows learners to inspect videos of their moves for labelling good and bad moves, while no such option is reported available in Gest. Finally, Scheidt and Pulver [193] present Any-Cubes; a TUI for making simple if-this-then-that systems based on image recognition. Any-Cubes consists of two cubes; the first cube has a built-in camera for taking photos of an object and recognising it again. The second cube has a set of banana plugs which allows users to connect actuators to it. This actuator is then activated if the first cube recognises an object, allowing users to experiment with making simple image recognition systems. While Any-Cubes incorporates buttons for labelling images of different objects, there is no way to inspect these images, and thus no way for learners to inspect and evaluate their models, except when the system is not able to recognise an object and thus does not activate the actuator.

[3]https://www.digitaldreamlabs.com/pages/cozmo, accessed Jan 10th, 2023
[4]https://teachablemachine.withgoogle.com/, accessed Jan 10th, 2023
[5]https://colab.research.google.com/, accessed Jan 10th, 2023

As argued above many existing tangible systems for teaching ML only rely on tangible aspects for a learner to explore a part of the ML process. Below, we present the MLM as the amalgamation of our exploration of how to, more systematically, make the ML process tangible and explorable for learners through an iterative process.

C.3 Methodology

The work presented in this paper is based on the Constructive Design Research (CDR) methodology [128, 130] and investigates how ML aspects can be made tangible and explorable for learners. CDR projects are driven forward by the construction of artefacts and the knowledge creation happens in experiments with and explorations of these artefacts [7, 130, 205]. Here, the central artefact is the MLM and the knowledge creation lies in the design and construction of the two devices as well as in the pilot study. CDR can be held accountable to several theoretical and practical concerns [127]. In this work we are concerned with exploring how to make ML tangible, in efforts to provide better tools for teaching ML as part of the CT curriculum. We see the MLM and its design rationale as a contribution in itself and as our method for exploring the design space of tangible ML. The MLM is simultaneously a suggestion for how a tangible ML learning tool might look and a Design Exemplar [21] embodying our research agenda and the knowledge created.

This work is part of a larger research agenda [59] based on a critical approach to teaching CT and designing tools for doing so [50, 107]. In other CDR experiments we have explored how to include ethical dilemmas of ML-based technology in design workshops with children, and how to scaffold critical discussions about ML in a Social Studies classroom [19]. As a whole, these experiments form a long term research programme invested in exploring what, and how, ML should be taught to young people.

C.4 The Machine Learning Machine

The MLM is a learning tool which enables K-12 students to iteratively work on ML models for binary classifications of doodles they draw using pen and paper. It consists of two separate devices (see Figure C.1): *the Trainer*, a big industrial looking metal box which provides a simple interface for labelling data and training ML models; and *the Evaluator*, an instrument inspired by laboratory equipment for analysing and evaluating ML models' performances. Besides the two units, the ML model that learners build is represented by a physical artefact. This model is moved between the two units to either train or evaluate the model. The MLM is designed to be used in a learning context with students (and teachers) who have no, or little, prior knowledge of ML.

C.4.1 Design Rationale

The MLM investigates how processes in supervised classification-based ML can become easier to explore and understand for students, by moving it into K-12 classrooms through a TUI, and how insights about ML can be incorporated in the design of a TUI and students' interactions with the system. Aligning with Bilstrup, Kaspersen, and Petersen [19], we take the stand that in order to have insightful reflections on the implications of ML, such as algorithmic fairness/biases, explainability and accountability [1], students must have some understanding of the underlying computational processes. The MLM aims at providing students with a fundamental understanding of ML by making the processes of creating and evaluating ML systems *expressive* and *exploratory* [143] through tangible artefacts. Consequently, the aim of the MLM is not to teach students how to implement ML algorithms, but rather to identify and reflect on how ML works, and what implications the use of ML has.

With MLM, students build and design their own ML models in an iterative process, allowing them to collaboratively explore the relation between data and predictions in ML systems. This involves taking processes that are normally intangible, fuzzy and digitised and putting them into physical artefacts which students can assemble around, try to understand, discuss and experiment with. Hallnäs and Redström introduce the notion of *slow technologies* [81] which invites for reflections through its interaction and expression. The interaction and visual expression of the MLM is designed to invite students to reflect on each step in the ML process by requiring students to perform each step manually, sometimes in tedious manners, and through the expressions of the two aesthetically very different devices.

In this manner, the MLM exposes the basic processes of ML, which differentiates ML from other kinds of computational processing. While designing the MLM and how it presents these processes, we found inspiration in two of the five *big ideas* about AI presented by Touretzky et al. [216]: That computers use models about the world to make predictions (idea #2), that the models are based on data collected in the world (idea #3) and are generated through a trial and error process. Based on these and by looking at how other ML tools such as AlpacaML [246], Gest [92], ML4Kids and Teachable Machine engage learners in ML processes, we have defined four iterative steps, which students will engage with through MLM; *Problem Formulation, Data Gathering and Representation, Training* and *Evaluation.* They closely resemble how ML systems are created in the real world albeit in a simplified way. Below, the steps are unfolded in detail:

1. *Problem Formulation:* The first step entails specifying the problem to be solved, whether the problem can be solved using ML and consequently what kinds of data need to be collected in order to do so. During this first step, it is determined what the ML system will be predicting including what categories new data can be classified as.

2. *Data Gathering and Representation:* Next step is to gather the needed data, making sure the data-set is sufficient in size and in representation.

3. *Training:* The third step is the training of a ML model. Here, the data-set gathered in the above step is fed to a ML algorithm, in the case of the MLM to a Neural Network, and is used to train a ML model. Each data-point is labelled as belonging to one of the categories determined in the first step.

4. *Evaluation:* The final step is to evaluate the model, once one has been trained. Here, data that has not been used for training the model, is used

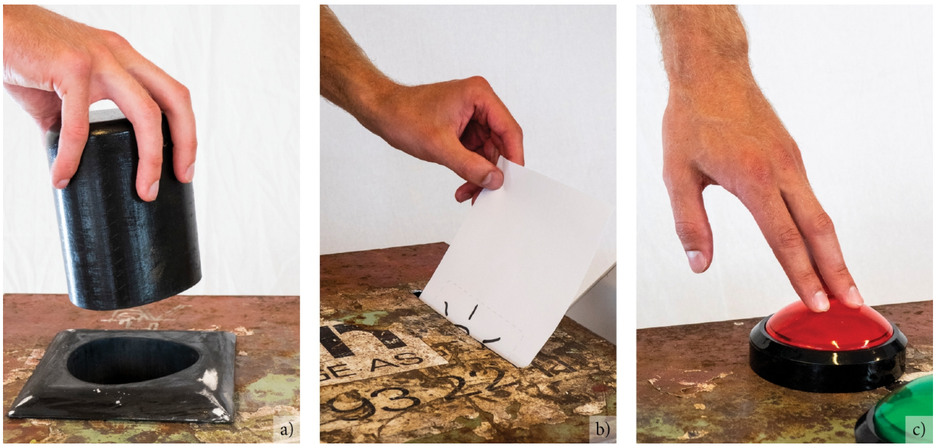


Figure C.2: The Trainer in use. *a)* shows the Model artefact being plugged into the Trainer, *b)* shows a drawing being inserted, and *a)* shows the drawing being labelled.

to evaluate the predictive abilities of the model. Based on the the results, the above steps are repeated to improve the model's predictions.

Below, we describe each component of the MLM, their function, how they were designed and how they relate to the four steps above. Next, we describe the use of the MLM and provide scenarios of how the MLM can be used in an educational setting.

C.4.2 Machine Learning Machine Components

The MLM consists of two interactive devices; the Trainer for training ML models, and the Evaluator for evaluating these models. In addition, the ML model is represented by a physical artefact and learners use simple pen and paper to create data for the MLM. This design exposes that training a model and using it for predictions are two different processes. In addition, the student-generated data makes the size and the nature of the dataset visible and manipulable. This gives students full control over the data and allows for simple data generation, iteration and a wide range of variations and creativity. Students can create datasets of anything they can imagine and draw, but the MLM is limited to two different labels; and thus, students are limited to drawing two motifs to distinguish between. Here, we will describe the functionality and design of each of the components.

The Trainer:

This device is used by learners, as seen in Figure C.2, to train ML models. It consists of a large metal box, with a socket for the Model artefact and a slit for feeding it with data in form of paper drawings. The paper is caught inside the box, by a clutch mechanism, that keeps it in place. Once this is done learners can press one of two large buttons, one green and one red, on the top of the Trainer to assign the inserted paper one of two labels represented with the red or green colour. Once labelled, a picture is taken of the drawing and used to train the

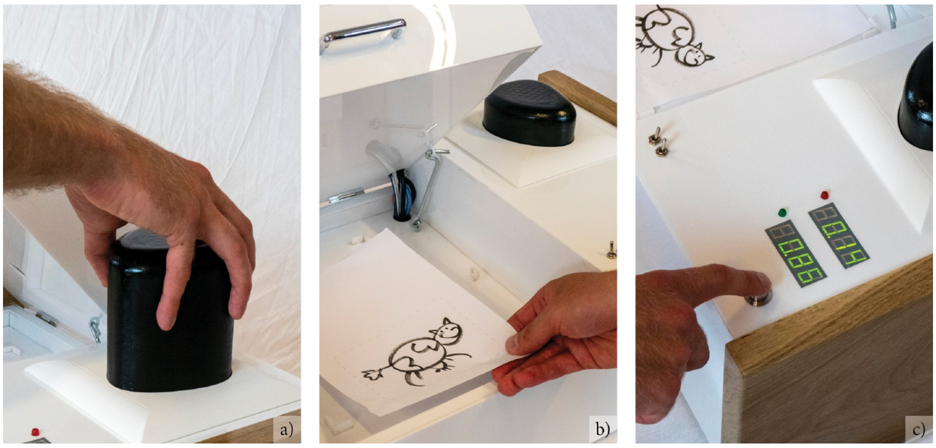


Figure C.3: The Evaluator: *a)* shows the Model artefact being inserted into the Evaluator, *b)* shows a drawing being laid in place, and in *c)* the drawing is categorised and the Evaluator shows its predictions and the probabilities in the LED displays.

ML model, and the clutch mechanism is released to let the drawing fall to the bottom of the box. After training a model and evaluating it using the Evaluator, learners can open up the Trainer, take out the drawings and inspect them to look for clues on how to improve their model.

The visual appearance of the Trainer references industrial machinery with a rusty, metal exterior and large coloured buttons. In addition, when the Trainer is opened to take out the drawings as described above, the electronics are exposed. The appearance of the Trainer was designed in this way to signify the messy, trial-and-error process of training ML models.

The Trainer provides students with several opportunities to reflect on the role of data in ML systems. Each drawing must be fed individually to the machine, making the process slow and deliberate. In addition, the system allows data to be inspected, by simply removing the paper from the box. If the Evaluator falsely labels a new drawing or simply fails to do so, this drawing can be compared to the training set to see if there are issues with the data representation, if the data-set is too small, etc.

The Evaluator:

This device is used, as seen in Figure C.3, to evaluate ML models created using the Trainer. It consists of a glossy white box with wooden sides. The box has a lid, which can be opened and fixated in a 45° angle. There is a camera in the lid, which is used to classify new drawings, which are placed under the lid. To the right of the lid is an interface for controlling the Evaluator. This interface has a socket for the Model artefact, two switches for turning on the Evaluator, a button for engaging the Evaluator in classifying a drawing, two LEDs and displays for displaying how certain the model is of its prediction.

The visual appearance of the Evaluator draws on references to consumer electronics with its glossy finish and wooden sides, and on laboratory equipment with the seven-segment LED displays and the knobs and switches. It was designed this way to convey the feeling of investigating the model, and how ML systems often appear polished and final once deployed even though they might not be as precise as they seem. Together, the Trainer and the Evaluator signify the difference between how ML models are created and how users often experience them.

With the the Evaluator, learners can evaluate a model by exploring how it interprets different drawings. This allows them to make hypothesises about how the model's predictions can be improved or changed, which they can go back to the Trainer and test out.

THE MODELS:

The physical models represent ML models and is moved between the Trainer and the Evaluator when training or evaluating models. This allows learners to keep track of which model they are currently using, and more importantly, it exposes through making the model tangible, that the model is only an approximation of reality, and that it is removed from the data, once it is used to make predictions[6]. The shapes of the model artefacts are asymmetrical oval cylinders, and are designed so they only fit into into their plugs from one angle. The top is bevelled while the bottom is straight to indicate what end is up and what end is down. On the bottom of the artefact is a small indentation in which a passive RFID-tag is placed. This gives each artefact a unique ID, which is used when moving models between the Trainer and the Evaluator. There is currently no way of telling if a modelled is empty or being trained on. The physical representation of the ML model can be seen in Figure C.3a being inserted into the Evaluator.

C.4.3 Using the Machine Learning Machine

Below we describe how the MLM can be used to create and learn about ML following the four steps described above; Problem formulation, Data gathering and representation, Training and Evaluation. The steps and how they relate to the MLM can be seen in Figure C.4.

(1) PROBLEM FORMULATION: Students define which doodles their soon to be model should be able to classify. They must choose two classes (two motifs for the doodles), while considering the complexity of the concept they choose. *E.g., create a model that can recognise if a doodle illustrates a square or a circles or, more complex, a model that can recognise cows from keys.*

(2) DATA GATHERING AND REPRESENTATION: Students consider how they can represent their two classes and start producing doodles with pen and paper. It is necessary to produce multiple doodles for each class. It can be one student who draw all doodles, or a class can collaborate to generate the data-set of doodles. The approach will affect how the model will perform. *E.g., every student in the class draws five quick doodles of respectively keys and cows.*

[6]The latter is the case for most ML systems, including neural network-based systems, but not for all ML systems.

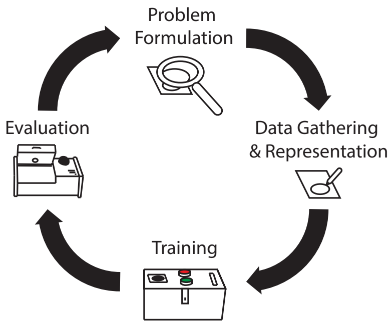


Figure C.4: The iterative ML process used by the MLM including where in the process the Trainer and the Evaluator are used.

(3) Training: Students plug an empty model (if it is the first iteration of a model) into the Trainer and start feeding it with their doodles and press the buttons to label each drawing and start training the model, providing it with examples of each class. *E.g., students press the red button each time they feed it a doodle of a key and the green button each time they feed it a doodle of a cow.*

(4) Evaluation: Students plug the model out of the Trainer and plug it into the Evaluator to evaluate the model with new doodles they, again, draw with pen and paper before putting them into the Evaluator to see the classification results. *E.g., a student may have a classmate who have not contributed to the dataset draw a cow and see if the model can recognise doodles drawn by other people.*

After this first iteration, students can repeat the process multiple times to improve their models, while exploring how the data-set impacts the model's predictions. At some point they may want to open the the Trainer to take out the doodles and inspect the data-set. In this iterative trial-and-error process, they can start to explore and gain an understanding of classical ML considerations: *Should they re-specify the problem? What are the model actually capable of (or can potentially be capable of) predicting? Is the dataset representative of the problem they have formulated? Is the quality of the data good enough or are some doodles difficult to classify even for a human? Could there be any unintentional biases in the dataset such as the type of pen or who have drawn the doodle?* Below, we will provide a few scenarios to clarify how we imagine this could be happen.

C.4.4 Use Case Scenarios

The following scenarios describe teaching activities about specific ML challenges that incorporate the MLM. In these scenarios, students follow the four-step-process described above and are beyond the first iteration of their ML model. Since the MLM is intended as a learning tool used in classrooms, we assume that a facilitator is present and guide students in their explorations. More complex use cases of the MLM, such as the ones below, assume that the facilitator has a fundamental understanding of ML, but the MLM could also be used to teach

ML fundamentals without the facilitator understanding ML. The facilitator in the scenarios is referred to as 'the teacher'.

Representative data:

Representativity is an important aspect of building ML models. It measures if the data-set is representative of the problem it describes and requires data exploration to identify situations that are not covered by the current dataset.

Scenario: *Students are training a model to recognise doodles of cows from doodles of keys. When testing their model with the Evaluator, it can recognise most keys, but provide mixed results when shown doodles of cows. Especially when they ask other students to draw cows. They have tried to draw more doodles to train the model further, but it does not improve the performance significantly. Encouraged by the teacher, they choose to take the dataset (doodles) out of the Trainer to inspect it. When comparing the dataset with the doodles, the model struggles to categorise, they realise that almost all cows in the dataset are looking to the right, while the new doodles illustrate cows that are looking to the left. The students draw a bunch of new doodles of cows looking to the left, and re-train the model. When they plug the model back into the Evaluator, it performs significantly better at recognising cows looking to the left.*

Unintentional Bias:

Another important aspect of building ML models, which is closely connected to data representativity, is unintentional biases. The ML model may learn biases from the data that were not intended by the designer. Thus, the model generates unintended assumptions about the relations between the inputs.

Scenario: *Students are training a model to recognise drawings of bicycles from drawings of houses. Their teacher has told them, that this is a difficult task, so they have designed an efficient setup to provide the model with a lot of data: One person is drawing houses, one is drawing bicycles and one is feeding the drawings to the Trainer. When they plug their model into the Evaluator, it performs well and the students are satisfied with their new model. However, when another group tests their model, it does not seem like the model has any conceptualisation of what a house or a bicycle is. The results seem mostly random. Through further experimentation and discussions with their teacher, they realise that their model is very accurate at recognising houses if they are drawn by the same person who produced the data-set and with the same pen. If this person, instead, draws a bicycle it more often categorises it as a house than as a bicycle. They have unintentionally biased their model to assume that the type of pen and the artist are the most important attributes for distinguishing the two types of drawings.*

C.4.5 Technical Implementation

The components of the MLM are implemented as follows. Each interactive device contains a Raspberry Pi (RPi) which runs a Python application and controls slave Arduinos. When a drawing is labelled by a user who presses one of the buttons on the Trainer, a Pi Camera takes a picture of the drawing. The RPi in the Trainer sends the picture with its label and the ID of the model being trained to a Python backend. This is done through a RESTful API that processes the images

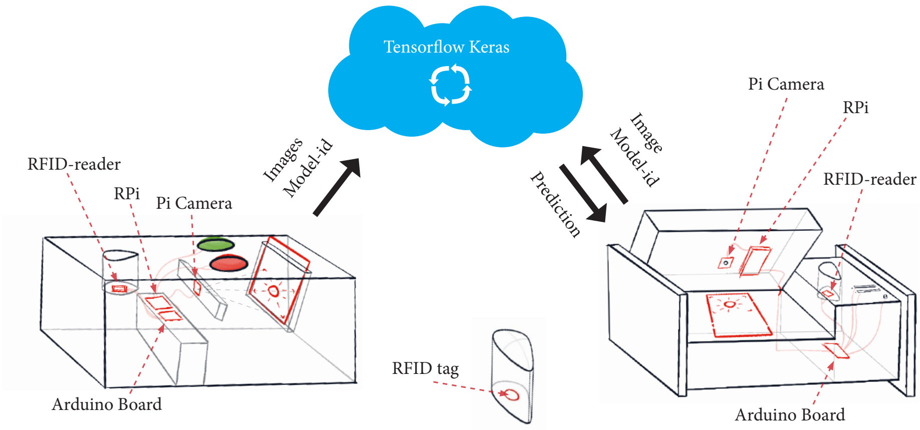


Figure C.5: A diagram showing the flow with which data is produced, ML models are trained and predictions are made with the MLM.

using OpenCV [7]; converts them to greyscale, reshapes them to $28 * 28$ pixels and saves them in a SQL database. The backend then conducts the ML computations using Tensorflow's Keras API [8] to train neural networks (models) which are used when making predictions. The models are trained from scratch without use of pre-trained models, but the implementation is inspired by, and has been tested with, Google's *Quick, Draw!* dataset, thus models can be trained and tested with this dataset. When a user presses the button on the Evaluator to have it recognise a new drawing, another Pi Camera photographs the drawing and sends it to the back-end, where the previously trained model is used to recognise the drawing. This prediction is sent back to the Evaluator and is displayed in the two displays. To keep track of which model is currently being trained, the Model artefacts have RFID-tags with unique ID's. These are recognised by RFID-readers in both the Trainer and the Evaluator. These unique ID's are used when storing and retrieving models in the SQL database. An overview of the technical implementation and flow can be seen in Figure C.5.

C.5 Pilot Study

We conducted a small, preliminary pilot study to evaluate the MLM, as seen in Figure C.6. The study was conducted with a group of two male teenagers, one aged 15 and one aged 19. Both participants were consented. The study lasted one hour and took place in a research laboratory at a university. To get an idea of the MLM as a learning tool, the study was designed to closely resemble how the MLM could be used in an educational setting. During the evaluation, we asked the participants to think out loud and to ask us if any issues occurred. One author took notes during the interview, another took pictures and a third facilitated the evaluation. Further, audio recordings were made of the entire evaluation. The study was conducted in the authors' and participants' local language. Quotes and

[7] https://opencv.org/
[8] https://www.tensorflow.org/

Figure C.6: The pilot study: *a)* shows a participant engaged in creating data in the form of drawings, *b)* shows the participants using the Trainer to train a model for recognising 1's from 0's, and *c)* shows a participant using the Evaluator to test a model.

questions have been translated into English by the authors.

First, we conducted a short preliminary interview to qualitatively establish a baseline about what the participants already knew about ML, where we asked the following questions: *a) Have you heard about ML before? What do you think it is? b) Do you know of any systems that use ML? And how? c) How do you think voice assistants, such as Siri and Alexa recognise what you are saying?*

Next, participants went through two tasks using the MLM. First, a task where participants were asked to draw five 1's and five 0's and train a model using these drawings. Next we asked them to evaluate their model by drawing new 1's and 0's and to improve it based on their evaluation. The second task was free. We asked the participants to formulate their own problem to solve and figure out what to draw in order to do so. Then the above process of training and evaluating iteratively was repeated.

Finally, we did a post interview. Here we, once again, asked the participants what they now thought ML was. We asked several questions about the use of the system; *How was is to use the system? What surprised you along the way about the system, the predictions or something else? What was hard to understand?* Next, we asked them if they could think of any more systems using ML. Finally, we asked them several questions that required reflecting about ML in general; *a) If you were to design your own ML system, what would it do? And what would you need to be wary of? b) In many places in the U.S. facial recognition systems have been banned. Why do you think that is? What can go wrong from using them?*

The analysis was made informally by two of the authors, who collectively went through the data collected during the study.

C.5.1 Results

Although the pilot study is only preliminary, and no generalisations can be made from the very small number of participants, some interesting findings did occur, that might inform how the MLM could be used to teach ML.

When prompted for knowledge about ML, the participants revealed, that they had little existing knowledge of ML. One had not heard of it, while another had a crude understanding of ML as something that learned something from inputs. The latter participant was able to explain that voice assistants worked by "someone having taught it the words´´, that it is able to recognise, but was not able to give a more in-depth answer. No participant could mention other systems utilising ML. In the interview following the study, the participants were not able to provide a more precise definition of ML, and they seemed to find it difficult to discern between the MLM and ML in general. However, during the study they had many in-depth reflections about issues that occur when developing real-world ML applications.

During the first task of drawing 1's and 0's, the Evaluator was able to reliably recognise 0's drawn almost circular, but struggled with oval 0's. It also struggled with 1's, which the participants had drawn as straight lines. From this, they gathered, that they needed to train the model with many different types of 1's and 0's; "We need to draw as broadly as possible". In the next iteration, participants tried to draw 1's and 0's that resembled the ones the Evaluator was not able to recognise. They drew many different 1's and 0's, and while the model improved it's ability to recognise differently shaped 0's, it did not improve at recognising 1's drawn as lines. One participant identified the issue; "We haven't made any more 1's that are completely straight. Maybe if we did, it would be better"

In the second task, participants chose to train a model for recognising drawings of the sun from drawings of the moon. The participants seemed more engaged in this task, and began on their own to explore the MLM. During the Data Gathering and Representation step, the participants tried to align their drawings to improve the model by drawing suns with 7 beams of light around it and drawing only waxing crescent moons. The system was able to recognise most drawings, but struggled with some and was not very certain in its predictions. The participants experimented with different placements of the drawings inside the Evaluator after having experienced that the system was not able to recognise a drawing of a sun that had been awkwardly placed. Initiated by one of the authors, participants also began experimenting with different lighting conditions by opening and closing curtains at a nearby window and blocking light from reaching the Evaluator with their bodies and hands. After having found a drawing that was particularly difficult for the Evaluator to correctly classify, one participant wondered what would happen if they were to train the model using the exact drawing and having it be classified afterwards; "If it has seen it before, it should be 100 [percent certain of the classification]".

In general, participants reported that they found the system easy and engaging to use. However, they struggled with understanding what the Model artefact represented and what its functionality was. When asked what their understanding of the model was, one participant guessed that they worked like "memory cards", storing photos of the drawings and moving them between the Trainer and the Evaluator. Although this understanding is close to reality, they had not realised that the model only had abstract knowledge of the data, and did not interpret *moving the model* as separating the model from the data as we had intended. Participants also struggled with feeding data to the Trainer, although this seemed to be the result of a system failure of the clutch mechanism, which would not always release the paper into the box. The participants quickly found workarounds for the issue and were not noticeably frustrated, but it did slow down the interaction with the Trainer.

C.6 Discussion

In this section we discuss the qualities of tangible approaches for ML learning tools, starting from The Machine Learning Machine (MLM), and how the findings from the pilot study illustrates the qualities of MLM, but also makes clear that future work is needed.

C.6.1 Qualities of a Tangible Approach to Engaging with ML

The ML process steps students are guided through with MLM are similar to other existing (and some very often-used) learning tools such as AlpacaML [246], Gest [92], ML4Kids and Teachable Machine, and its novelty and contributions lay in how students interact with ML processes and concepts. The MLM does not provide the same interactive, efficient tool to process a broad variation and large quantities of data proviced by Teachable machine and ML4Kids. It also does not embody the data generation to the same degree as AlpacaML and Gest and it does not contextualize ML into another subject such as AlpacaML (ML in athletics). The MLM takes another approach by making ML processes more tangible than they are in the real world, and puts them into the classroom to spur curiosity and reflection among students. It meets students with big industrial looking tools, which require them to generate every piece of data by hand and manually *feed* it to the ML model. Similarly, the evaluation of models is not automated, but is conducted manually, one drawing at a time. The system may seem unnecessary and laborious as other digital tools can achieve the same with less effort, but as Marshall, Price, and Rogers argue: *"If extra information is gained about a domain or if students' interpretations are guided constrained by manipulating physical materials, then tangible systems might offer advantages over other kinds of learning environments."* With the MLM, learners produce real-world, analogue data, with all the kinks and quirks that come with it. The MLM might not recognise a drawing of a cow if it has been drawn with a slightly bolder pen or in a different colour. In this way, the MLM embodies the richness of real-world ML applications and the messiness of collecting data about the real world. The MLM presents ML as a *slow technology* [81]. It requires students to use a slowly moving, laborious system, and asks them to make time and space to assemble around ML, to explore it together, and to take their time to reflect on details; *"Why was this model classified wrongly, while this one wasn't? Maybe its because its too narrow?"*.

On one hand, the interaction with the MLM may be experienced as pointless by students, because the tools are inefficient and can only be used to produce models with very few use-cases and little meaning in the real world. On the other hand, the physical interaction can make ML less abstract than more digitised systems and the slowness can make more room for understanding what is going on and for reflecting on how small changes in the data can impact the models' predictions.

C.6.2 Qualities of an Iterative Process to Engaging with ML

A core design rationale of the MLM is to support students' engagement with ML through an iterative process. During the one hour pilot study, participants managed to train two different models with several iterations for each model. These were partly supported by participants' own reflections and partly supported by challenges introduced by the authors. The design rationale is related to Marshall, Price, and Rogers's argument that effective tangible user interfaces for learning require transitions between ready-to-hand and present-at-hand [143]. During the pilot study, we saw how iterations were most often triggered by reflections, and the MLM being present-at-hand. As an example, the iterative process served to engage the students in reflecting on limitations in the data-set and seeking to improve the performance of the model through providing additional training data, which was carefully crafted based on limitations in the previous collection of drawings. After the first round of training data, one of the authors made a few drawings, which were deliberately made slightly different than the rather homogeneous first training data set, and asked the students to evaluate them with their current model. The low success rate of the evaluations of these drawings triggered students' reflections on sample versatility (variety of the data-set) and they added drawings with larger variation, re-trained the model, and experienced a higher success rate for the evaluations of the author's drawings with the improved model. Figure C.7 illustrates how additional drawings were created to enhance sample versatility. In a second case, the students added additional training data to the model, but experienced a worse performance of the Evaluator. This triggered reflections on which other aspects could influence the performance as well as a few quick iterations, where lighting conditions and the positions of drawings were changed to see how both factors impacted the performance. The pilot study indicates that the MLM supports relatively fast iterations on the MLM process and that this serves to engage the students in exploring and reflecting upon ML. Future studies will serve to establish this finding further and also help towards a more systematic exploration of how to trigger reflections in such processes.

C.6.3 Challenges in Relating Activities around the MLM to Everyday Life and Wider Implications of ML for Individuals and Society

The MLM is deliberately designed as a generic machine which can be appropriated for different purposes, cases and aspects of ML, as described in section 4.4. In the pilot study, this turned out to be a challenge in terms of clearly connecting the ML process of the MLM to real-life applications, which participants knew from their everyday lives. We expect this to be the cause of the generic task of drawing 1's and 0's with no context supplied. Future work will explore a more deliberate

Figure C.7: Participants look at the data-set used for evaluating their model and make new drawings to improve the model through a greater variety in the data-set

framing of the MLM through different activities around the current MLM. We also plan to design and implement an additional MLM component which will allow students to bring their models out of the classroom and into their everyday lives using a mobile device and subsequently bring their explorations back into the classroom through a series of teaching activities supporting reflection over time.

C.6.4 Limitations

Due to the Covid-19 pandemic, we were not able to recruit participants for our pilot study as we would have normally done. Instead we opted for a small, preliminary study with two persons. This means, that the findings cannot be generalised, and the findings should instead be used to inform future designs and evaluations. This also means that one of the author knew in advance both participants and that the participants knew each other. We do not believe this affected the results of the study.

The design of the MLM is limited in its scalability. Although several models can be trained, the two devices are major bottlenecks when trying to engage an entire classroom of students. This was not an issue in the pilot study with only two persons, but we suspect that it will be a limitation in a larger evaluation.

C.7 Conclusion and Future Work

The Machine Learning m´Machine (MLM) is designed to complement existing learnings tools for teaching ML in classrooms and to expand existing research regarding what students should know about ML and how we can design tools and meaningful activities that allow them to engage with it. The design of and design rationale behind the MLM represent a novel approach to making ML explorable and graspable through tangible user interfaces. The MLM achieves this by allowing students to engage with ML through an iterative process of Problem Formulation, Data Gathering, Training and Evaluation. A preliminary pilot study has been conducted, with the MLM, which demonstrates the qualities of an iterative approach for engaging students in understanding and reflecting about ML and the qualities of a tangible approach to teaching ML in terms of shifting between being ready-to-hand and present-at-hand [143] and of presenting ML slowly and deliberately. Future work is needed into exploring how the MLM can be integrated in teaching activities to support more reflections on the implications of ML in their everyday life. In addition, work is needed to make the MLM scalable to a full classroom. We have, already made plans for evaluating the MLM with a larger group of students in order to validate preliminary findings and expand on the functionality of the MLM, to further explore its potential. In order make the MLM more scalable we see two paths forward. First, an additional component could be developed to supplement the existing devices, which would allow students to bring their models into the world using their smartphones. Another way would be to allow students to use their own devices throughout the process of training and evaluating their models. This would entail designing a frame in which students could insert a device, e.g, a smart phone, and use the phone's camera for taking pictures of their drawings. We plan to investigate both directions in future work.

C.8 Acknowledgements

We would like to thank our colleague Daniel Graungaard for helping us in the early stages of building the MLM. We would also like to thank Simon Møller Christensen and the entire ChomskyLab team at the Department of Computer Science, Aarhus University for making sure we had an ample supply of materials and machinery to finish the MLM. Additionally a big thank you to everyone in the Center for Computational Thinking and Design at Aarhus University[9] for your comments and insights during the design process. Finally, this work was supported by a research grant (#28831) from VILLUM FONDEN, and could not have happened without it.

[9]https://cctd.au.dk/

Paper D

Towards a Model of Progression in Computational Empowerment in Education

Magnus Høholt Kaspersen, Daniel Graungaard, Niels Olof Bouvin, Marianne Graves Petersen, & Eva Eriksson

Abstract

In this paper, we present a model for progression in computational empowerment in education. Computational empowerment (CE) expands computational thinking (CT) by adding a focus on empowering citizens to critically engage with technology, but currently lacks an articulation of what characterises progression towards CE. Through combining aims from computational thinking with computational empowerment, and structure progression using the SOLO taxonomy, we take the first steps towards a model for understanding and articulating progression in computational empowerment. The model has been applied in the analysis of four CE-focused research projects. Based on the analysis, we propose that computational empowerment is a matter of reaching a high competency level regarding computational concepts, computational practices and reflexivity regarding the effect of technology in one's own life and in society. Finally, by formulating examples of learning goals, we illustrate how the model can be used by teachers and researchers to articulate, determine and compare CE learning goals, so that learning goals are aligned and complement each other from one stage to the next.

D.1 Introduction

Computational Thinking (CT) [234] has been widely recognised as way to encourage new generations to create technology and to pursue careers in STEM subjects (science, technology, engineering and math). However, in recent years, CT has been criticised for being too narrowly focused on developing students' technical and generative skills, and for neglecting critical skills for students to decode the role of technology in our everyday lives and society [113]. This critique has led to multiple proposals for expanding CT to encompass broader, more critical perspectives [107, 213].

One of these proposals is Computational Empowerment (CE) [50, 107]. Dindler, Smith, and Iversen [50] defines CE as *"the process in which children and youth, as individuals and groups, develop the skills, insights and reflexivity needed to understand digital technology and its effect on their lives and society at large, and their capacity to engage critically and curiously with the construction and deconstruction of technology"*. The authors demonstrate, through digital fabrication activities in a Danish school context [203], how CE might look in practice, but provide limited reflections on how successful these activities were in computationally empowering the participants. They find that students improved their understanding of digital fabrication technologies, gained hands-on experience with digital fabrication technologies, found the work with such technologies motivating, and had initial signs of *design literacy*; the competencies to work *"creatively with technology, and complex problem solving"* [203].

However, Dindler, Smith, and Iversen [50] and Iversen, Smith, and Dindler [107] do not provide much analysis of the relation between the results and the activities that might be used to inform future learning activities and interventions. We argue, that if CE is to gain traction, researchers and teachers first need ways to relate the aims of CE with specific learning goals. Tools, methods, and practices for teaching and assessing design and programming are quite well developed, but the same cannot be said for teaching and assessing children's abilities to analyse, understand and critically reflect on the role of technology, especially when it comes to resources that teachers can use in their daily practice [50]. While CE is clearly defined in terms of its agenda and goals, it is less clear what steps teachers need to take, to take students from being users of technology to becoming computationally empowered citizens. This has led us to formulate the following research question:

RQ: *What characterises progression in computational empowerment in education?*

In this paper, we take the first steps towards a model of progression in CE that allows researchers and educators to articulate learning goals in computational empowerment that are aligned and complement each other from one stage to the next. We do so by analysing four on-going research projects, which all adhere in some way to the CE agenda. For the analysis, we developed an early model for relating concrete learning goals to progression in CE, by combining aspects from CT, CE, and the learning sciences. We found that while the model was not able to provide an all encompassing analysis of the research projects, it provided the involved researchers with a language to express how their projects align with CE. In addition, the analysis allowed us to compare and discuss strengths and shortcomings of the research projects across different aspects. Finally, we suggest that it can help design new activities and projects based on the CE agenda.

The paper is structured in the following way: First, we present work related to what characterises CT and CE skills as well as progression in teaching. Next, we present the CE progression model. Subsequently, we apply it to four research projects to exemplify how it is useful as an analytic tool. Finally, we discuss the use of the model both as an analytic as well as a generative tool.

D.2 Related Work

In order to move towards a model of progression in CE, in the following, we present work related to characterising computational skills and progression in learning activities for children and young people.

D.2.1 Characterising Computational Skills

The notion of computational empowerment has recently grown out of a need to address limitations in conceptions of CT in terms of its lack of focus on the wider implications of the role of technologies in society and personal lives [107] [123]. Kinnula and Iivari [123] makes the point that even though projects seeks to support empowerment, this might imply different perspectives including functional, educational, democratic and critical perspectives. Encompassing these perspectives Iversen, Smith, and Dindler define three core aspects of CE [107]: *"1) engaging creatively in technology development, 2) understanding the role of digital technology in society, and 3) reflectively and critically understanding the role of technology in one's own life."*. The first aspect encompasses traditional CT skills, which multiple authors have tried to characterise. Wing's own conceptualisation of CT [233, 234] consists of several aspects including problem reformulation, recursion, problem decomposition, abstraction and systematic testing, which all relate to the first aspect of CE outlined above. Similarly, Perković et al. [175] present a framework for CT, which they base on Denning's principles of computing [49]; computation, communication, coordination, recollection, automation, evaluation and design. They argue that to develop a full CT curriculum, these principles should be addressed.

Weintrop et al. [227] introduce a taxonomy of CT (for mathematics and science) consisting of data practices, models and simulation practices, computational problem solving practices and systems thinking practices. Barr and Stephenson [12] include data organisation, data analysis, automation, efficiency and generalisation while Bers et al. [16] include abstraction, generalisation, and trial and error activities.

Common for these frameworks is, that they focus on computational concepts and practices related to computing, and little on the wider perspectives on technology related to the implications of the use of digital technology for personal life as well as society, the second and third aspect of CE mentioned above. Adopting a wider perspective, Brennan and Resnick Brennan and Resnick propose three key dimensions of CT, namely computational *concepts*, computational *practices* and computational *perspectives*. Concepts refer to concrete computational aspects such as abstraction, automation and so forth. Practices refer to the practices programmers develop as they work with e.g. general problem solving skills and data practices. Together, concepts and practices encompass the CT skills presented above, and focus on allowing students to *engage creatively in technology* *development*, i.e. the first aspect of CE. Unlike other CT frameworks, Brennan and Resnick go beyond this and introduce *perspectives* as a set of skills leading to "a shift in the understanding of oneself or the world" similar to the more reflective second and third aspects of CE. As an example, to assess students working with Scratch (so called *Scratchers*) against these concepts, Brennan and Resnick propose three approaches; project analysis, artefact-based interviews, and design scenarios. They find, that while the three approaches are able to effectively assess the students' knowledge and use of computational concepts and practices, they are not effective at evaluating perspectives, and it seems they struggle in defining exactly what role perspectives should play in CT.

To summarize, both Iversen, Smith, and Dindler [107] and Kinnula and Iivari [123] have quite broad understandings of CE related to the wider implications of digital technology, but represent a more narrow conception related to technology development, whereas the more CT-oriented perspectives have vague conceptions relating to the wider implications of digital technology but quite rich conceptions relating to the creation of digital technology. But none of these conceptions of CE have notions of progression. We argue that in order to move towards a model for learning goals and progression for computational empowerment, CE and CT approaches should be brought together, but also be complemented with notions regarding progression in learning activities.

D.2.2 Characterising Progression in Learning Activities

A key challenge in teaching computational thinking is supporting *progression* of knowledge [58], and even more so in computational empowerment. A learning progression is a sequence of subskills that needs to be mastered to reach a curricular aim [181], and is characterised by starting with something simple and moving on to something more advanced, applying the knowledge in new ways. Learning is an individual activity: each of us learns at a different pace and has different cognitive abilities [235]. Thus, progression can be hard to measure, as it includes both cognitive and more observable processes. In mathematics, progression is characterised by learning one new concept while building on understanding the previous concept [235]. However, we do not yet know much about what characterises the development from basic to complex forms of many of the 21st century skills such as computational thinking [122], and even less so CE. Thus, we do not yet have a model of best practice for how to guide teachers in what to expect from learners at different levels of skill, and how they can make progress.

We see a shift in what learners need in order to fulfil their potential, from the previous focus on knowledge, skills, and competencies [172] to a focus on knowledge, skills, attitudes, and values, meaning encompassing not only knowing and doing, but also *becoming* [9, 10, 166]. Attitudes and values are an acknowledgement that competencies are more than knowledge and skills. They refer to the principles and beliefs that influence one's choices, judgements, behaviours, and actions in regards to the individual, society, and environment. This is aligned with the thinking behind computational empowerment, where perspectives on one's own life and on society are vital aspects, and where critical thinking is a major driver.

Hattie raises attention to the need for setting challenging learning intentions, being clear about what success means, and an attention to learning strategies for developing conceptual understanding about what teachers and students know and understand [87]. However, in defining and analysing the goals of a specific activity, a general structural framework for evaluating learning outcomes can be useful as it enables us to compare learning goals between different subject areas and learning activities. One taxonomy based on the outcome of teaching is the Structure of the Observed Learning Outcome (SOLO) [17]. The SOLO Taxonomy is a five-tier hierarchical framework for structuring learning outcomes. *"SOLO describes a hierarchy where each partial construction [level] becomes a foundation on which further learning is built"* [17, p.41].

The SOLO taxonomy differs from, e.g., the BLOOM taxonomy in being based on observable outcomes rather than internal cognitive processes [17, 24]. The goals of the SOLO taxonomy include providing a tool for defining curriculum objectives, intended learning outcomes, and evaluating learning outcomes based on these objectives [17]. The SOLO taxonomy can be used to identify and describe what learners are doing, explain how well it is going and predict what they should do next [94]. Progression can be defined as moving up in SOLO levels, from pre-structural, to uni-structural, multi-structural, relational, and up to extended abstract level as the highest level. The first two levels refer to developing surface knowing and the latter two levels refer to developing deeper knowing. Surface learning refers to studying without much reflecting on either purpose or strategy, learning many ideas without necessarily relating them and memorising facts and procedures routinely. Deep learning refers to seeking meaning, relating and extending ideas, looking for patterns and underlying principles, checking evidence and relating it to conclusions, examining arguments cautiously and critically, and becoming actively interested in course content [89].

Each level in the SOLO taxonomy is represented by a number of verbs that can be used to formulate learning goals. For instance Multistructural is illustrated by Combine, Describe, Enumerate, Perform serial skills, list, and Extended abstract is illustrated by Reflect, Evaluate, Theorize, Hypothesize, Generalize, Predict, Create, Imagine. The SOLO taxonomy offers a hierarchical and linear structure and offers a good measure for describing progression in student competence, and is not specific to a certain subject or context.

SOLO offers a powerful model to illustrate the distinction between surface and deep in the structure of observed learning outcomes [89], it can guide in clarifying what success means [87], it can be used in assessment tools for teaching and learning [88], but most importantly, SOLO can be used to create a common language of learning in any curriculum area, and to help students of any age adopt a growth mindset when learning [94].

D.3 Towards a model of CE Learning Goals

Through a theoretical grounding based in computational thinking and computational empowerment, we present an early model of progression in CE, where we seek to integrate aspects of CE with slightly wider conceptualisations regarding the development of digital technology, using the common language of learning from the SOLO taxonomy, see Figure D.1. As a starting point, we look to the three aspects of CE as presented by Iversen, Smith, and Dindler [107]. *"1) engaging creatively in technology development, 2) understanding the role of digital technology*

	Uni-structural	Multi-structural	Relational	Extended Abstract
Concepts				
Practices				
Perspectives in one's own life				
Perspectives in society				

Figure D.1: Our Computational Empowerment model

in society, and 3) reflectively and critically understanding the role of technology in one's own life." As argued above, these overlap with the three categories of CT skills presented by Brennan and Resnick [33]*"1) computational concepts, 2) computational practices 3) computational perspectives"* but the emphasise different aspects. To better cover the aspects presented by Iversen, Smith, and Dindler, we propose to expand the categories by Brennan and Resnick to four: *Computational Concepts, Computational Practices, Computational Perspectives in One's Own Life, and Computational Perspectives in Society.*

These categories allow for a more detailed classification of technical knowledge and skills than the original aspects of CE, as well as a finer grained understanding of the attitudes and values associated with being computationally empowered and taking a critical stance towards technology than allowed by the framework of Brennan and Resnick [33].

Having identified four categories to make distinctions between CE skills, we also need a distinction between different levels of progression in these skills. We have chosen to use the SOLO taxonomy as a common language [17], as it is a widely used tool, which provides educators with common ground for formulating learning goals that can be compared across subjects [236]. However, we acknowledge that there are other taxonomies that could be used as a common language, such as e.g. the BLOOMs taxonmy [24], and we are aware of that the SOLO taxonomy has been criticised for being simplistic and reductionist [116, 236]. Still, SOLO is a powerful tool [89], which provides an established and a common language for articulating and discussing learning goals [94], and that has been applied to many different school subjects from poetry and history [18] to science [44] and math [88], but to the best of our knowledge never before to computational empowerment. This research provides a first exploration of the use of the SOLO taxonomy in the context of CE.

By combining the four categories of CE skills and the levels of competency described in the SOLO taxonomy, we arrive at an emerging model of progression in CE, as seen in Figure D.1. We propose this model as a common language for discussing and comparing CE projects, for developing and determining learning goals in new CE projects, and for establishing coherent CE research programmes.

D.4 Method

To answer the research question on what characterises progression in computational empowerment, we have combined related work in computational thinking, computational empowerment, and progression in learning, in order to develop an initial model, as described in the above section. In order to validate the model and investigate its applicability, we have further applied the model in the analysis of four research projects from a larger project portfolio at our university research centre [57]. The projects have in common that they relate to and/or focus on computational empowerment in children and young people.

The study consists of interviews with leading researchers for each project, while deploying the model to position and reflect upon each project. The four projects were included in order to represent a diversity of CE perspectives. The projects range from high school teacher training to primary school children in special needs education. For an overview of participants, see Table D.1.

The project *FabLab@School.dk* is aimed at empowering primary school students by engaging them in creating and imagining with technology so that they can critically discuss their role in society. In the *Computational Thinking in Math and Science* (CTiMNAT) project, high-school teachers are trained to ensure that their students work with a particular phenomenon related to their subject both in itself and through code. In *Computational Thinking in Humanities, Arts and Social Sciences* (MCTIG), a number of code based teaching activities are co-designed with high school teachers. Finally, the *Collaborative Information Technology in special Education* (CITE) project focus on children in special education, and aims to make the children reflect on their own educational tools, and also that they can have a say in the development of digital tools used in their education.

D.4.1 Data Collection and Analysis

One of the authors interviewed the leading researchers in each project. The interviews consisted of two parts: The first part of the interview was a semi-structured interview to understand the project activities, and to discuss the main purpose of the project, and how it relates to CE. In the second part of each interview, the interviewer collaborated with the respondents to fill in the learning goals from the project in the CE model. Each interview lasted for around one hour and took place in the respondents' offices. Some of the interviews were conducted in the researchers native language and translated, some were conducted in English. All the interviews were recorded and later transcribed.

For the analysis, we first analysed data from interviews for how the projects and their learning goals were positioned in relation to CE.We then mapped the learning goals of the project activities in relation to concepts, practices, perspectives in one's own life, and perspectives in society, using the SOLO taxonomy. Two of the researchers re-categorised and synthesised some of the learning goals based on the description given in the interviews, according to the following principles: Given the hierarchy of the SOLO taxonomy, some projects had intermediate goals causally leading to a final goal, e.g., a goal of learning to download and install some software leading to learning to use the software. For such learning goals, only the high-level goal was included in the analysis. Additionally, some projects included activities rather than learning goals, such as using a particular program.

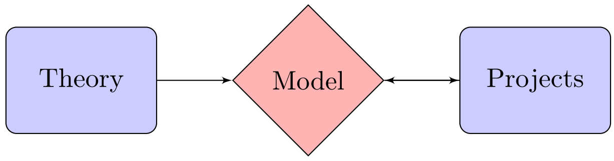


Figure D.2: Overview of method, where the resulting model stem from projects and theory, and where the model has been applied to the projects.

Project	**Level**	**Participants**	**Interview**
Fablab@school	Primary school	1.100 teachers 11.000 children	1 researcher (F)
CTiMNAT	High school	200 teachers 4.000 children	1 researcher (F)
MCTIG	High school	22 teachers 350 children	1 researcher (M)
CITE	Special education Grade 5-6	2 teachers 2 pedagogues 8 children	2 researchers (F)

Table D.1: Overview of projects, educational level, participants and number of people who have been interviewed.

These were also removed. Finally, several of the projects feature parallel learning goals in a single category of computational skills. E.g., the CTiMNAT project includes both *learning the methods of a particular design model* and *learning to apply knowledge of design processes in different subjects* as two separate learning goals within practices. Since these are seen as parallel learning goals on different levels and not as causal, they were both included in the analysis.

Having analysed the four projects individually, we reflect upon blind spots, common trends and challenges across the projects.

Finally, we articulate how the model can be used to derive general CE learning goals at all levels for a teaching activity focused on designing with technology.

D.5 Research Projects

In this section, we present four research projects using the emerging model presented in section D.3. For each project, we present a short project description followed by a summary of the analysis with examples of learning goals and their categorisation using the model.

D.5.1 #1: Fablab@School

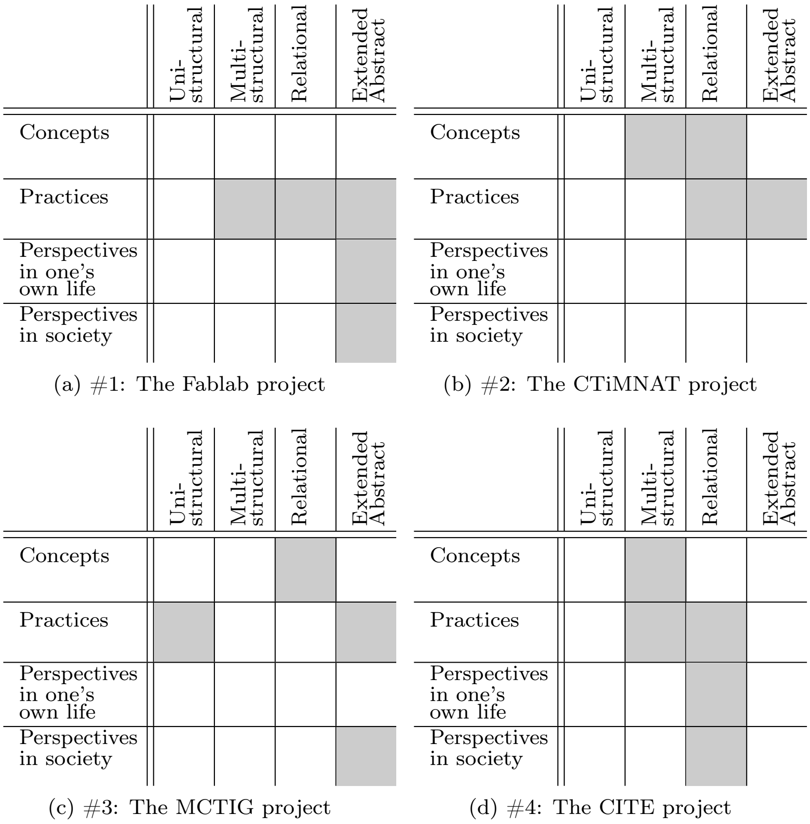
The Fablab@School project explored the core challenges and potentials of integrating digital fabrication technologies into the context of the Danish educational system [107] . The purpose was to create a sustainable educational initiative, and to give teachers the tools to integrate design processes and digital making technologies into their teaching. It was carried out between 2013 and 2017 by a small interdisciplinary research team as a collaboration between three Danish municipalities. The project was based on the global fablab@school project developed by the Transformative Technologies Learning Lab at Stanford University. The experiment involved 1100 teachers and 11.000 students from different Danish primary schools. Fablab@school sought to empower teachers as well as students, although we for this analysis focus solely on the latter group. Regarding students, the aim was to empower them by engaging them in creating and imagining with technology in a way where its role in society could be critically examined. An overview of the CE learning goals in the project can be seen in Figure D.3a.

We did not find any explicit learning goals concerning CT *Concepts* in the fablab@school project. In the *Practices* category, the fablab@school project has several learning goals. Among these were *the students can carry out an iterative design process on their own case and can explain and reason about their choices in the process.* Learning to conduct a process must be considered a practice, and is *Relational* as it requires an understanding of the correlation between design process model and children's own designs. In the *Perspectives in one's own life* category, this project has the following learning goal: *Students should understand how technologies influence them in their everyday life.* As it requires the students to transfer the knowledge they have gained from working with their own case to an application and consideration of the technology that they themselves encounter in their own lives, this places this goal, in the SOLO taxonomy, at the *Extended abstract* level. Finally, at the *Perspectives in society* level, the learning goal was *Students should understand how the technology they created in their process could be integrated into society.* As before, this fits with the *Extended abstract* category.

D.5.2 #2: Computational Thinking in Math and Science (CTiMNAT)

This is an ongoing project working with high-school math or STEM teachers. The teachers are trained to, by themselves, create teaching material in Agent Based NetLogo [214] that revolves around a particular phenomenon found in the subjects that they are teaching, and that can be used together with students. The project is based on the Coding, Modeling & Content (CMC) approach [154] which in turn ensures that the students work with the modelled phenomenon both in itself and through code. At the point of writing, 67 teachers had participated in the project. An overview of the CE learnings goals in the project, as defined for high-school students, can be seen in Figure D.3b.

One of two learning goals in the *Concepts* category is that students should be able to *understand selected parts of code.* This requires the students to know certain concepts about programming and is thus categorised as concepts learning goal. Since it requires students to know certain programming concepts and to recognise these in code, it is placed in the *Multistructural* category of the SOLO taxonomy. Another learning goal, relating to *Practices* is that student should be able to *make changes to the code that changes the model's behaviour related to subject specific terms.* This learning goal is intended to let the students reflect on knowledge gained from the subject and from that imagine how they can create a new behaviour for the modelled based on this knowledge. This learning goal requires the students to go through a process of experimenting with the code and changing it. The goal describes that the students are able to carry out a process and therefore is considered as a practice. As it requires the students to



(a) #1: The Fablab project

(b) #2: The CTiMNAT project

(c) #3: The MCTIG project

(d) #4: The CITE project

Figure D.3: The learning goals in the model applied to four research projects.

use existing knowledge to extend the model it is categorised as *Extended abstract* in the SOLO taxonomy.

D.5.3 #3: Computational Thinking in Humanities, Arts and Social Sciences (MCTIG)

This project teaches high-school teachers from the humanities, arts, and social sciences to use NetLogo models for teaching in their subject. Through a process of co-creation with the teachers, some teaching activities based around models are created. Throughout the project, focus has been to adapt and adjust models, and model-based teaching methods to the Danish high schools. Teachers are in the Scandinavian school tradition involved in curriculum development, and co-developing alternatives for curricular activities together with the teachers is a necessary process in order to create new activities in a sustainable way, where they will be used in classrooms after the project has finished. An overview of the CE learning goals in the project as defined for high-school students, can be seen in Figure D.3c.

The MCTIG project has two learning goals in the *Practices* category, one being: *Students should be able to make and justify subject-specific changes to* *the NetLogo code, that changes the behaviour of the model.* After analysing and working with a model, students apply their existing knowledge about the subject (e.g., Social Studies) and based on this they alter the model. The learning goal is about a process wherein the students construct code and compare the result to subject specific theory. This is a way of working, and must be categorised as a practice. As it requires the students to both use the model and to use existing knowledge from the subject and apply this knowledge to changes of the model. Through a process of hypothesis and creation, the learning goal is categorised as *Extended abstract* in the SOLO taxonomy. In the *Perspectives in society* category, MCTIG has the following learning goal: *Students should be able to reflect on the validity of models in society and realise that they are built on assumptions.* Students are given an agent based model by their teacher which simulates a real phenomenon. By working with these models, students should be able to understand the limitations of models and transfer these to other types of models in society such as economical models or climate models. The learning goal is categorised as *Perspectives in society*, since it requires the students to use the perspectives from the teaching activity to reason about technology use in society. This learning goal is on the *Extended abstract* level in the SOLO taxonomy as it requires the students to reflect and hypothesise about new models encountered in society based on the knowledge gained through working with a specific NetLogo model.

D.5.4 #4: Collaborative Information Technology in Special Education (CITE)

The project worked with one class of children from special education (seven children) and the teaching personnel connected to the class. The purpose of the project was to explore and develop collaborative technologies such as e.g. games in co-design with children [57]. From a computational empowerment perspective, the aim was help the children understand that the tools they use can impact their learning, and that they can impact and modify games: Games are not only for consumption, they are also something that has been designed by other people for a purpose. An overview of the CE learnings goals in the project can be seen in Figure D.3d.

The CITE project has one learning goal in the *Perspectives in one's own life* category: *Children should understand that they can shape technology.* This learning goal describes the knowledge that the children is to have after involvement in the project activities. It is categorised as *Perspectives in one's own life*, since it describes how the activities should shape the children's understanding of themselves and technology. It is classified as *Relational* as it describes the children's relation to technology in their own life, and in order to actually have achieved this learning goal the children would have to understand different processes in game development and relate them to each other and their own abilities to influence game design and affect the design of a game. In the *Perspectives in society* category, CITE has the following learning goal: *Children should understand that technology is not necessarily complete, but follows an iterative process.* The goal is about shaping children's view of what technology is, and how it affects society, which places the goals in the *Perspectives in society* category. This learning goal is *Relational* since it is about relating the process children have taken part in

	Uni-structural	Multi-structural	Relational	Extended Abstract
Concepts		#2;4	#2;3	
Practices	#3	#1;4	#1;2;4	#1;2;3
Perspectives in one's own life			#4	#1
Perspectives in society			#4	#1;3

Figure D.4: Projects summarised in heatmap. Numbers refer to each project.

during the experiment with technology encountered elsewhere in society.

Together, these four projects give an indication of how current CE projects relate to the broad goals of CE. We can see that a single project rarely encompasses all aspects of CE, and looking at the heatmap of all projects in Figure D.4, we see that learning goals seem to gravitate towards more advanced levels of the SOLO taxonomy. While it is understandable that research projects aim to qualify their work in regards to high levels of competencies in their participants, we argue, that it might be difficult for a teacher to know how to structure intermediate learning goals in their own teaching based on these projects. Below, we discuss what to consider when striving to computationally empower students, and we provide an example of how the model can support teachers and researchers in defining subject specific learning goals for computational empowerment.

D.6 Discussion

In this paper, we have provided the first steps towards a model of progression in Computational Empowerment (CE) in education. The model is developed by combining aspects of computational thinking with computational empowerment, using the structured of the SOLO taxonomy. To qualify the model, we have analysed four research projects with a focus on educational activities and learning goals connected to CE. Based on this analysis, we will here discuss the strengths and weaknesses of the model, and address the research question: *what characterises progression in computational empowerment in education?*

Dindler, Smith, and Iversen state that computational empowered is when "people are empowered to autonomously and critically engage in the development of digital artefacts" [50]. Although we do not disagree with this, we think that there is a lack in what characterises progression in CE in order to be able to teach and assess it.

By combining attitudes and values with knowledge and skills in line with OECD [166], we argue that in order to be computational empowered, you need to reach the relational and extended abstract levels as in the SOLO taxonomy [17] in concepts, practice, perspectives in one's own life and perspectives in society. When teaching for computational empowerment, you need to support the learners in both knowing and doing, as well as in becoming [9]. This approach acknowledges that principles and beliefs influence one's choices, judgements, behaviours and actions in regards to the individual, society and environment, and is vital for an autonomous critical reflective engagement in the development of digital artefacts.

	Uni-structural	Multi-structural	Relational	Extended abstract
Concepts	Students will be able to **recognise and identify** generative concepts, tools, methods, and materials for designing with technology.	Students will be able to **list and describe** generative concepts, tools, methods, and materials for designing with technology.	Students will be able to **compare and contrast** different generative concepts, tools, methods, and materials for designing with technology.	Students will be able to **create** new technological artefacts using generative concepts, tools, methods, and materials for designing with technology.
Practices	Students will be able to **recognise and identify** procedures for designing with technology.	Students will be able to **perform serial activities** for designing with technology using generative tools, methods, and materials.	Students will be able to **integrate or analyze** several procedures for designing with technology.	Students will be able to **create** new generative tools, methods, and materials for designing with technology.
Persp./Own life	Students will be able to **recognise** that different tools, methods and materials for designing with technology can have substantial impact on their lives.	Students will be able to **list and describe** possible impacts on one's own life of different tools, methods and materials for designing with technology.	Students will be able to **relate** to different tools, methods and materials for designing with technology.	Students will be able to **critically reflect** on personal tradeoffs between different tools, methods and materials for designing with technology.
Persp./Society	Students will be able to **recognise** that different tools, methods and materials for designing with technology can have substantial societal impact.	Students will be able to **list and describe** possible societal impact of different tools, methods and materials for designing with technology.	Students will be able to **relate** to different tools, methods and materials for designing with technology on a societal level.	Students will be able to **critically reflect** on societal impact of different tools, methods and materials for designing with technology.

Table D.2: Examples of learning goals in computational empowerment with a focus on designing with technology. Persp/Life is short for Perspectives in one's own life, and Persp/Society is short for Perspectives in society

As presented earlier, a learning progression is a sequence of subskills that needs to be mastered to reach a curricular aim [181], and is characterised by starting with something simple and moving on to something more advanced, applying the knowledge in new ways. However, in computational empowerment, we do not yet have a model of best practice for how to guide teachers in what to expect from learners at different levels of skill, and how they can make progress.

Analysing the results from Figure D.4, we can see that CE relies heavily on practices, and that CE focuses more on the higher levels of the SOLO taxonomy: multi-structural, relational and extended abstract levels [17]. This goes in line with that critical thinking is a major driver within CE [50], as critical thinking is an extended abstract competence. To further inspire teachers and researchers, when defining their subject and learner specific CE learning goals, we suggest examples of learning goals for all competency levels of computational empowerment, here specifically targeting *designing with technology*, see Table D.2. The suggested examples of learning goals were formulated by mapping verbs from each level in the SOLO taxonomy, to concepts, practices and perspectives deriving from CE and CT, and serve as an illustration of how the model could be applied in practice.

It should be noted that these learning goals in *designing with technology* do not cover all aspects of CE, but rather serve as an example to illustrate and articulate progression in accordance with the model. In order to teach CE, these learning goals need to be complemented with other learning goals, such as e.g. envisioning the consequences of a technology, understanding indirect and direct stakeholders of a technology, identify and evaluate the implicit and explicit values of a technology, reflect on the role of the designer, etc.

D.6.1 Limitations of the Model

The model has several limitations, that should be addressed in order to move towards a deeper understanding of progression in CE. First, we found the vertical axis to be useful. It allowed us to classify the learning goals of the research projects in meaningful ways, that could be compared both within and between projects. We did, however, have some issues using the horizontal axis. The SOLO taxonomy implies that learning goals and thus learning can be divided into four categories, which arguably is a reductionist view on learning [116]. We witnessed this in the interviews, where it was sometimes difficult to agree on the categories of different learning goals. There are however, also arguments for using the SOLO Taxonomy; it is already widely used in education and is a familiar tool for teachers [236]. Further, its simplicity allows learning goals to be formulated precisely and in ways that make them comparable between projects and classes. Secondly, we only included projects, for which we had access to researchers with first-hand experience, similarly to the approach taken in [47]. We considered including experiments from other researchers. However, from our interviews, we found that CE learning goals for an activity is often implicit and therefore requires first-hand access to data and researchers involved in the experiment. We do however suggest, that other researchers interested in empowering youth and children apply the model to their own projects.

We further acknowledge the importance of social aspects of learning, and the need for differentiating by taking each individual's needs, interests, and goals into consideration, and the individual child's ability for regulation of cognition, socio-emotional factors and behaviour. However although such aspects are not currently part of the suggested model, we highly encourage the teachers and researchers to consider such aspects when adopting the model in their own subject and learner specific contexts.

Finally, we acknowledge that there is still no common clear-cut understanding of what the E in CE mean. However, the proposed model can be seen as a first step towards understanding progression in computational empowerment in education, and hopefully make a contribution to the existing body of literature.

D.6.2 Future Work

In the previous sections we have demonstrated the model's ability to analyse projects and shed light on the coherence of research programmes. We see several directions in which the model could enhance our understanding of CE in education. The generative abilities and its usefulness to help teachers determine CE learning goals for their teaching should be explored. Similarly, the model is a first step towards a tool for researchers to look beyond CE learning goals of specific projects and begin to form coherent portfolios of research projects with different focus points. Finally, work is needed that go beyond learning goals and look toward evaluating computational empowerment in children. We see the model as a first step towards formulating learning goals for CE that lend themselves to being evaluated and assessed.

D.7 Conclusion

In this paper, we have addressed the research question on what characterises progression in computational empowerment. The contribution is an initial model of progression in computational empowerment in education. The model has been developed by merging aims from computational thinking with computational empowerment, and the progression has been structured using the SOLO taxonomy. The model has been applied in the analysis of four different research projects. To illustrate how the model can be used to formulate learning goals, examples of learning goals for *designing with technology* have been suggested. It is the hope of the authors that this initial model of progression in computational empowerment, as illustrated by examples of learning goals, can serve as a guide to on the one hand teachers in what to expect from learners at different levels of competency, and on the other hand support researchers when defining projects and research questions. Looking forward, we invite other researchers and practitioners to critique, revise and discuss the model.

D.8 Acknowledgements

The work is part of Centre for Computational Thinking and Design and is funded by AUFF - Aarhus University Research Foundation AUFF-E-2017-7-5 and Villum Foundation.

Paper E

Remarkable AI: Designing for Computational Empowerment in Use of AI Systems

Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, John Zimmerman, Ole Sejer Iversen, & Marianne Graves Petersen

Abstract

In recent years AI-enabled systems have become truly pervasive in our daily lives, but their AI components often remain difficult to comprehend and even notice. This challenges how we design such systems. A recent paper presents Unremarkable AI as an approach to making successful AI systems through integrating AI-based decision support unobtrusively into professionals' routines. This paper proposes that designers work with a spectrum between Unremarkable and Remarkable AI and we present the position of Remarkable AI, echoing previous discourse around Unremarkable and Remarkable Computing. Remarkable AI entails exposing and explaining its use of AI, as well as allowing users to interrogate and negotiate this. Remarkable AI is based on three case studies of learning tools for AI education and, we provide three design strategies. To contrast Unremarkable and Remarkable AI, we redesign an unremarkable system to be remarkable. Finally, we discuss how to determine the right level of remarkableness.

E.1 Introduction

AI systems are now a large part of most people's everyday lives, driving technologies that we use multiple times a day such as social media, search engines, music and video streaming services. However, with their proliferation has followed concerns on how AI systems are also affecting our societies negatively [164] in ways that are opaque and difficult to understand [1, 164].

With AI's promises of progress and concerning implications there is a tension between designing powerful, seamless systems and making sure their users are empowered and in control. This is reflected in the fields of computing and AI education, which in recent years has highlighted the need for students to become Computationally Empowered [50, 107], as well as the need to include critical aspects of AI and its use in AI curricula [137, 216]. Further, this echoes a long-standing debate in HCI about dichotomies and continua regarding to what extend the inner workings and/or values of technologies should be exposed to users. One of these discussions surrounded the notion of Remarkable [178] and Unremarkable [215] Computing. As an approach to ubiquitous computing in the home, Unremarkable Computing, as coined by Tolmie et al. [215, p. 404], aimed "to make computational resources that can be unremarkably embedded into routines and augment action", referencing Weiser's [229] seminal view on *the computer for the 21st century.* In response, Petersen [178] suggested *Remarkable Computing* as a complementary approach, arguing that as much as digital technologies are able to support people's routines, they are also increasingly becoming "objects of lifestyle and identity" [178, p. 1446], and that such objects are and should be remarkable.

In 2019, Yang, Steinfeld, and Zimmerman [238] presented an argument similar to that of Tolmie et al., namely the notion of *Unremarkable AI*; an approach to designing AI systems, specifically for supporting clinical decision making, that aims for them to be "subservient to the day-to-day decision-making" of their users. They motivate this by the failure of other clinical decision support tools to successfully be incorporated into doctors daily routines.

AI and ubiquitous computing are analogous in that they share a central tenet, namely the "interest in building technologies that make sense and respond in a sensible way to the complex dynamics of human environments" [133], and (as least in regards to machine learning) a dependence on data collection and analysis. However, this also leaves Unremarkable AI subject to similar critiques to those delivered by Petersen to Unremarkable Computing.

In addition, concerns have been raised regarding the increasing infusion of AI-enabled decision support systems in domains such as insurance, banking, school administration, which largely black-boxes the underlying AI models, and leave both the professional users as well as other stakeholders largely unaware of the underlying assumptions and potential biases in these systems and without any feedback loops which allows for corrections and improvements to the systems [164]. We see these as examples of potentially harmful unremarkable AI systems, which are integrated into professional routines, largely without opportunities for professionals or stakeholders to critically inspect and reflect about the underlying assumptions of such systems [196]. Urged by this, and inspired by previous unremarkable/remarkable debates, we propose the need for Remarkable AI as a complementary perspective to Unremarkable AI [238], and aim to unfold a continuum between these two outer points. Inspired by previous discussions in HCI such as the seamless/seamful, invisible/visible, black-box/glass-box [93] debates, we argue that these distinctions are useful, if not necessary, for staging a dialogue about the underlying assumptions and values of the systems we design, and we argue that such a debate is particularly needed in the context of how we design AI-enabled systems.

In this paper we explore how to determine the right level of remarkableness in an AI-system, by presenting Remarkable AI as one end of a remarkableness-in-AI continuum, with Unremarkable AI on the other end. Remarkable AI is focused on making users aware that they are interacting with an AI system, exposing what the AI system is doing, how it is doing it, and why, as well as empowering them in reflecting and deciding on the possible consequences AI has on their lives. Note that we use the notion of the "user" throughout the paper to refer to anyone directly or indirectly interacting with AI systems, but are aware that the term is contested (see e.g., [13, 191]).

Remarkable AI came to exist as we drifted [130] in our design work in AI Education (see e.g., [19, 20, 118, 120]), based on the realization that we might teach students how to spot an AI system in the wild and how it works (sometimes against their interests), but providing them with the means to make changes according to their self-interests, necessitates changes with regards to how end-user AI-enabled systems are designed. In an effort to demystify our process leading to Remarkable AI [243], we will detail this process below.

To qualify Remarkable AI, we look to existing approaches for designing for reflection [196], to the field of computing education, which has a rich body of work regarding opening up complex technologies to increase users' agency and reflexivity in emerging technologies such as AI [50, 107, 114], and to Participatory Design [28, 29] for its political focus on values and power dynamics in systems design.

Based on this, we present two high-level design sensitivities for Remarkable AI, namely *Exposing & Explaining AI* and *Interrogating & Negotiating the Role of AI*. We argue that these sensitivities are central, not only for learning about AI, but also for designing Remarkable AI systems that retain users' agency and empower them in reflecting and acting on the possible consequences of such systems in their lives.

To further delimit the design-space, we present three case studies based on our own research projects [19, 119, 120], and extract three concrete strategies for design Remarkable AI systems from them. All of these projects are in the context of AI education, but were designed with the explicit aim of computationally empowering students with regards to AI, and exposing different critical aspects of AI, and thus share similar goals with Remarkable AI. Although these prototypes all feature supervised ML, we argue that Remarkable AI is applicable to all data-driven AI-applications, since they are based on similar high-level choices [56]. Based on this analysis, we present three concrete design strategies; 1) Foregrounding the ML Process to Expose AI Components, 2) Iterative & Safe Experimentation to Explain AI, 3) Facilitating Discussions and Reconfiguration to allow Interrogation and Negotiation.

The remainder of the paper is structured as follows; first, we present related works on the history and usefulness of discussions around dichotomies and continua in HCI, as well as approaches to designing for reflection. Second, we present Remarkable AI as a design approach to AI systems that aim to expose their inner workings to their users, including how our own work in AI education lead to this framing of AI systems design. Next, we analyse three of our own research projects, and present the three design strategies. To further aid designers, we apply these strategies in a redesign of the clinical decision support tool presented by Yang, Steinfeld, and Zimmerman [238]. Finally, we discuss the need for considering remarkableness in AI systems design, how to determine the right level of remarkableness, as well as our method for delimiting the Remarkable AI design-space and the limitations of Remarkable AI.

E.2 Background

The HCI community has a rich history of using dichotomies and continua to discuss and inform the design of interactive technologies and the field as a whole. An early example of this can be seen in the heated but seminal discussion between Newell and Card [155, 156], and Carroll and Campbell [40] about how methodically strict the use of psychology in HCI research aught to be.

Another central tension is the dichotomy of seamful/seamless design of ubiquitous systems. This idea originated from Weiser [229]'s vision of "The Computer for the 21st Century"; a vision in which computers weaved ubiquitously into our work lives. This lead to a push for designing *seamless* technologies [42]. However, aiming for seamlessness risks "sacrificing the richness of each tool in order to obtain bland compatibility" [42]. Weiser himself later qualified the notion of seamlessness by arguing for seamful design [228], which "emphasizes configurability, user appropriation, and revelation of complexity, ambiguity, or inconsistency" [102]. 25 years later, Inman and Ribes [102] reviewed the research that sprung from these notions and found that they should not be treated as opposites but instead as complementary, each appropriate for different use cases and contexts.

Similarly, we see Remarkable AI as complementary to Unremarkable AI [238]. Remarkable AI is appropriate in contexts where reflexivity on the use of AI is necessary and where Unremarkable AI might lead to skewed power dynamics between stakeholders or loss of agency of users, although Unremarkable AI might not always lead to this (and Remarkable AI is deemed to bring its own issues in certain cases).

To inform Remarkable AI we look to approaches to designing for reflexivity. Sengers et al. [196]'s Reflective Design is an influential approach, drawing on several other traditions within HCI and Design such as critical and speculative design [4, 54], and reflection-in-action [195], to present a comprehensive approach to designing technologies for critical reflection. They argue that "for those concerned about the social implications of the technologies we build, **reflection itself should be a core technology design outcome for HCI**" (emphasis in the original). They propose that designing for critical reflection entails exposing otherwise hidden aspects and making these "available for conscious choice". To demonstrate their approach, they present two case studies embodying Reflective Design.

We agree with the goals of Reflective Design, but argue that its scope is not adequate for dealing with emerging technologies such as AI; since AI can be too complex for end-users to immediately grasp, providing opportunities for conscious choice is not enough. In addition, a reflexive approach to designing AI systems should include opportunities for users to inspect and learn how the system's AI-component operates and what effect it has on the system, themselves and their surroundings, and then allow users to make conscious choices regarding their interaction with the system.

Amershi et al. [2] suggest 18 guidelines for designing AI-infused systems "that people can understand, trust, and can engage with effectively". These cover initial interaction with the system (G1: Make clear what the system can do, G2: Make clear how well the system can do what it can do), during interaction (e.g., G5: Match relevant social norms, and G6: Mitigate social biases), when something goes wrong (e.g., G8: Support efficient dismissal, and G11: Make clear why the system did what it did), and interaction over time (e.g., G16: Convey the consequences of user actions, and G18: Notify users about changes). However, as mentioned by Amershi et al. themselves, "an AI system may adhere to each of these guidelines and yet impact people's lives or livelihoods in a consequential manner". Remarkable AI thus can be seen as a complement to these guidelines that aim to empower users with regards to such impacts.

Another approach to helping users understand how AI systems work is everyday algorithm auditing as suggested by Shen et al. [198]. Algorithm auditing is an umbrella term used for describing methods of uncovering issues in AI systems [148, 190], and everyday algorithm auditing specifically is a method that relies on everyday, situated users of AI systems reporting on their lived experiences with AI systems and the issues that might arise during use. Shen et al. suggest that compared to more systematic and centralized ways of auditing, everyday auditing promises to uncover issues specific to different, often marginalized user groups that might be overlooked if auditing in centrally controlled. Because everyday auditing relies on the collective actions of several users, the approach also "require less algorithmic expertise on the part of those organizing and enacting the audit". However, some knowledge, and not least interest, is still required by users, and importantly, the systems to be audited require that systems, at least to some degree, expose their AI component for users to uncover possible issues.

The aim of designing systems to expose their AI components has in recent years spawned a community around Explainable AI (xAI). xAI addresses how to ensure that models in AI systems become more explainable (see [1, 78]), by producing explanations alongside model predictions in order to foster an understanding of their behavior [146]. However, xAI often fails to consider the needs of non-expert end-users [147], or the different needs of different users, and its focus on explainability mostly does not extend beyond model predictions to include the process by which models come to be. In order to empower end-users of AI systems, explanations of the entire process should be presented [20, 56].

This notion of exposing the entire AI process has been a focus in recent research on teaching AI. Bilstrup et al. [20] systematically reviewed research on learning tools for teaching machine learning (ML) and found that most tools focus on the beginning and end of the process and only a few tools address the inner processes, such as data representation, and choice of model type, and learning algorithm. Similarly, Long and Magerko [137] argue that one of the core competencies for navigating a world "in which AI transforms the way that we communicate, work, and live with each other and with machines" is understanding each step in the process of creating ML models. More broadly, Computational Empowerment [50, 107] is an approach to designing learning experiences for children that aim to empower them "to make critical and informed decisions about the role of technology in their lives". Dindler, Smith, and Iversen [50] argue that there are two main ways of engaging with technology when trying to understand it, namely through *coding* (i.e., constructing, making, etc.) and through *decoding* (i.e., observing, analyzing, etc.).

Computational Empowerment was born from the marriage of Computing Education and Participatory Design (PD), and we find that PD too has an important role to play in formulating Remarkable AI. Since its beginning more that 50 years ago, PD has been occupied with involving users in design decisions about the use of technology in their work [28, 29, 161] and, later, everyday [26, 177] lives. Remarkable AI is inspired by PD's rich catalog of methods for including users in the design process and for allowing them to negotiate technology's role in their lives.

E.3 Remarkable AI

Remarkable AI aims to make the AI component of such systems present for their users in order to expose and explain their inner workings, and empower users to interrogate and negotiate the role of AI in these systems and its effect on their lives.

In the education field, a major issue is what happens when students leave the classroom. We might succeed in AI Education in our quest for Computational Empowerment [50, 107], but if everyday AI-enabled applications do not follow, our victory might not be as impactful as we wish it to be: *"Freedom and morality are meaningless unless they are to be enjoyed among people who are free and moral. Hence, the ethics of the individuals in the AI world is influenced by the architects of the ecosystem within which they enact..."* [212, p. 225]. It was this realization that led us to consider remarkableness as a frame, that can, and should, inform AI systems design, and which fostered a drift from AI Education to end-user design. We find it imperative that designers of AI-enabled systems heed the calls to consider aspects of fairness, accountability and transparency in such systems, and that they design their systems with an appropriate level of remarkableness at the right time.

In order to do so, the AI component of such systems must first be made perceivable to users, since things that we cannot perceive, cannot become present in our lives. In Unremarkable *Computing*, this is addressed through routines; "artifacts that are implicated in routines can be perceptually available yet practically invisible in use" [215]. In other words, in Unremarkable Computing the apparent perceptibility of an artifact does not matter; what is important is its perceptibility *in use*. Presence in Unremarkable AI [238] plays a somewhat more sophisticated role. Yang, Steinfeld, and Zimmerman [238] argue that while Unremarkable AI should not usually interfere with users' routines, it should be "present enough to slow decision making down" when the prediction is at odds with the users' assumptions.

As a response to Unremarkable Computing, Petersen [178] points out, that no interactive technology is inherently invisible in use; it must first be appropriated and learned, before it can fade to the background. However, as the proliferation of AI systems has shown us, that can no longer be said to be true. Surely, the front-end of an interactive system is not inherently invisible, but the back-end might be, and this is where AI is most-often situated. Unremarkable AI does in fact highlight that a prediction is being made [238, Fig. 1], but its aim of not interfering with users' routines means that this might be ignored. Remarkable AI complements this approach by aiming for reflection and empowerment.

This leaves two opportunities for AI to become present for the users of systems that integrate it. First, we might teach users how to tell if an AI system is being deployed behind the scenes and to understand what it is doing, and why. This is the approach taken in computing education research (e.g., [20, 53, 137]). While we agree that this step is necessary to prepare coming (and current) generations for a future in which AI will play an increasing role, if everyday technology exclusively integrates AI in unremarkable ways, users might be left with little power of the their use of such technologies, apart from outright not using them.

Second, designers of AI systems can design their products in ways that exposes and explains their AI components, and allow users to interrogate and negotiate the role of AI in the system, as well as its effects. Below, we expand how Remarkable AI aims to support designers in doing so, by looking to research in computing education on teaching AI, since this community has produced several examples of exposing and explaining AI systems through use. Obviously, these approaches are presented in the context of learning, but as we will show, they contain important lessons to be learned for designing end-user systems. Additionally, xAI plays a role in providing explanations about how AI systems produce predictions, and is an aspect to consider when designing Remarkable AI system, but as argued above, it often does not cover the entire process and thus is not itself enough.

In the following, we present two high-level design sensitivities derived from the literature, and argue why designers should pay attention to these when designing AI systems to be remarkable.

E.3.1 Learning through Use: Exposing & Explaining AI

In Petersen [178]'s Remarkable Computing a central aspect that sets it apart from its Unremarkable counterpart is the notion of Learning through use; that appropriating and learning to use a technology is an inherent trait that should be designed for. Remarkable AI extends this notion, and argues that to understand the potential impacts of an AI system some level of understanding of its inner components is necessary, and that this should be designed for, by exposing and offering explanations of these components through use.

The notion of *coding* from Computational Empowerment [50] is particularly useful here. For example, consider how we typically interact with a search engine: We enter a search query (most likely in our web-browsers' search bar, i.e. not on the web site of the search engine), scroll through the results of the query and click any link that sparks our interest. If we instead try to code (in the broadest of meanings) our own, we are forced to pay attention to how they work behind the scenes; how queries are interpreted, how web-sites are crawled and indexed to become available for searching and, crucially, how results are ordered using AI, and which data are collected about the users. Coding then makes us look at and understand technology in a new light.

This is not to suggest, that AI systems should require users to code them from scratch, but if aiming to be remarkable they should provide users with opportunities to dive below their front-end, to change and customize system parametres and to observe the effect this has on the outcome in a safe environment. This approach is similar to the notion of *infrastructuring* [180, 206], but focuses on exposition and explanation rather than appropriation. Exposition and explanation is, however, not enough since it does not necessarily provoke reflection on the wider implications of a system's use of AI.

E.3.2 Interrogating and Negotiating the Role of AI

Therefore, the second aspect of a Remarkable AI system is providing opportunities for users to interrogate and negotiate the role AI plays in the system and how this affects its stakeholders and context of use.

Again, Computational Empowerment provides a useful notion, namely *decoding*. Here, technologies are analyzed in depth with regards to their purpose, the values they imbue and their potential and current impact. Drawing on this, we suggest that systems designers should incorporate ways of decoding their use of AI with regards to, e.g., what values guide how data was collected, analyzed, and used in the system, what its criteria of success are, but also which stakeholders are involved/implicated in which aspects of the system, who has control over how the system's predictions are used, etc.

We are also inspired by Reflective Design [196], which provides several design principles for fostering reflections in end-users. It argues, that reflective designs should support users in reflecting on their own lives, by e.g., "offering users new ways of experiencing and reflecting on their activities" and "to offer up new choices that may not have been in the user's awareness". Further, they argue, that "technology should support skepticism about and reinterpretation of its own working". Users should be the ultimate authority on the activities they are performing with the system and systems should be designed to encourage reflection on this and to allow users to reject using the system, if they disagree with its values and/or purposes.

Finally, Remarkable AI argues that users should be involved in negotiating the values and power dynamics built into the system. Here we look to PD for its efforts of doing so, more broadly, in the design of interactive technologies. In their paper on the future of PD, Bødker and Kyng [30] called for visions of approaches to "escape the iron grip of the big corporations" with regards to their, now all the more so, ubiquitous use of data and AI with no opportunity for users to intervene. Remarkable AI is an attempt to provide designers with an approach for developing AI systems that utilize AI's numerous promises while aiming to avoid its many pitfalls by, among other things, drawing on PD methods for inviting users into the engine room and and negotiating how the use of AI plays out.

Remarkable AI aims to guide designers to design systems that expose and explain their AI components and invite users to interrogate and negotiate AI's role in their lives. It proposes doing so by allowing users to dive below a system's front-end to alter and experiments with it's configuration and ultimately the results of it's AI component, and by offering up opportunities for users to reflect and act on the values, stakeholders and purposes of the system. Below, we present three use cases from AI Education, which we have designed with similar aims, and discuss how the lessons learned in each case can inform the design of Remarkable AI systems.

E.4 Case Studies

In the previous sections we motivated the need for considering the spectrum between Remarkable AI and Unremarkable AI. In this section, we will analyse three studies from our own research to qualify and demonstrate how Remarkable AI can be designed for. These case studies all stem from the Education field, and are learning technologies designed to expose the inner workings of ML and trigger reflections in students using them. Indeed it is our experiences from these projects that led to our drift towards end-user design, and in turn the cases share similar goals with Remarkable AI, although their designs are particular to Education.

We chose this method as an experimental way of qualifying the problem-space of designing for Remarkable AI. By drawing on the different-yet-similar problem-space of AI Education, we highlight the similarities and differences between the contexts, with the aim of offering a delimination of the boundaries of Remarkable AI, and proposing concrete design strategies to designers when dealing with the remarkableness of their AI systems.

To analyse the cases, the first author (who was involved in designing all three cases) used the two design sensitivities above as themes, deductively analyzing the cases with regards to how they address these. Based on this analysis, the same author then compiled a set of design strategies, which were then discussed and refined in collaboration with two other authors, who were also involved in the research projects.

E.4.1 VotestratesML

VotestratesML (VML) [119] is a collaborative learning tool for exploring ML and its use in democratic elections (see Figure E.1) for use in a social studies classroom. Students collaborate on building ML models for predicting voter behavior using data collected about the 2015 <Anonymized> general election. VML was designed to allow students to interrogate the impact ML is already having on our societies. The tool exposes how machine learning is used to predict the behavior of voters, and allows students to discuss different ways of using it in political campaigns. VML does not explicitly explain how each of its AI components work, but instead allow students to iteratively explore AI concepts such test and training data, different model types, etc., and the system is designed such that it can never be wrongly configured.

In terms of designing for Remarkable AI, we take notice of how VML was explicitly designed to make societal impacts remarkable by framing ML as a *political* tool rather than a *computational* tool, and by highlighting the similarities between the tool and real world applications. Further, VML aims to foster free exploration by never allowing students to bring it to a faulty state and get stuck.

E.4.2 The Machine Learning Machine

The machine learning machine (MLM) [120] was designed to make every (simplified) step of the process of creating a ML model explicit and *physical* (see Figure E.2). Students draw sketches on A5 paper and "feed" it to a *Trainer*-machine, that digitizes the sketch and trains a model on the sketches its fed. After training a model, students can move their physical model-artifact to an

Figure E.1: To the left: A screenshot of the VML interface, to the right: students using VML during the in classroom study.

Figure E.2: Images from the MLM pilot study. A) Users drawing circles to create a data set. B) Users training a model using the Trainer-machine. C) Users testing their model using the Evaluator-machine.

Evaluator-machine that lets them use their model to recognize newly drawn sketches. By moving data creation, analysis, training and evaluation into the physical world, students are invited to explore the process and make on-the-fly hypotheses about the system they are creating.

In terms of designing for Remarkable AI, from MLM we take notice of how the entire process from data collection to the use of the model is exposed to students, and how this allows users of the MLM to systematically explore each element in said process and what effect it has on the next step. Further, while the MLM does not explicitly explain what happens in each step, it allows students to iterate on their models, come up with explanations for what happens and then test these out by strategically trying to improve their system through re-training the model.

E.4.3 ML Ethics Workshop

Finally, we take a look at a ML ethics workshop (see Figure E.3) designed to engage students without prior knowledge of ML in discussions on concrete ethical issues, by scaffolding them in designing a pseudo-functional ML-system using different types of cards and templates to guide their design [19]. This activity is

Figure E.3: Students designing a ML system using the Ethics workshop materials

initiated by a design brief provided by the teacher, e.g. to design an AI-enhanced mobile app to combat loneliness in adolescents. Next, students are handed a set of data cards that they are asked to discuss. These cards describe different categories of data that might be used in a ML model such as "health" or "smart home". After openly discussing the data cards, students ideate on the design brief and pick an idea to work on. Subsequently, students describe their envisioned system design using a paper template and selecting relevant data cards and describing specific instances in the broader categories. Once they have described a system, students are handed a single ethics card selected by their teacher, that describes an ethical issue relevant to their idea, e.g., "Privacy — is the system violating the privacy of you and your users? How?". Next, they are asked to discuss and *redesign* their system based on concrete issues they identify through their discussion of the ethics card. Finally, students present their system and discussion to each other, and a shared discussion is had it the classroom of each project.

In terms of designing for Remarkable AI, of notice here is how the workshop is able to engage participants in interrogating and negotiating the impact of an AI system of their own design, even though they have little to no prior knowledge of AI or ML. Additionally, the specific design brief in the workshop was chosen due to its close relation to students' own lives. This was done to better allow students to relate to the issues they find during their discussions.

E.5 Design Strategies for Remarkable AI

Above, we have presented two design sensitivities for Remarkable AI, namely designing to expose and explain AI and to interrogate and negotiate its role in our lives and in society. While the three cases have been designed for educational purposes, we have found three design strategies for designing end-user systems that can be extracted from them. Where the design sensitivities are high-level sensitivities to discuss and consider in a design process, these design strategies are concrete suggestions to designing for the sensitivities. Note that while the three strategies presented below are what we have been able to extract from our case studies, different and complementary strategies might be extracted from looking at other work.

E.5.1 Foregrounding the ML Process to Expose AI Components

The first strategy that is deployed across these systems, is to foreground the ML process in a structured way, as to expose different components that go into creating an AI system. With the MLM, students create their own data, manually load it into the model and then manually test the resulting output. VML opts instead to use fixed real-world data, but allows students to explore the data, select what data they want to include, and further to experiment with different model types and their configuration. While students do not create a functional ML system in the ethics workshop, they are asked to consider what data is relevant to their system and which predictions or inferences it makes.

By foreground the ML process to users of AI systems, they first and foremost are made aware that a computer is making decisions that impact their use of it or the decisions they make based on their interactions with it. While they might not understand each step in the process, it is at least apparent to them. We see this as the first step towards users being able to call the system and/or its use into question.

E.5.2 Iterative & Safe Experimentation to Explain AI

To aid students in understanding AI, several of the cases above allow experimentation without the any risk of getting stuck. VotestratesML is designed to allow students to iteratively try any combination of data, models type, and model parameters and the interface adapts to provide feedback to students regarding which choices are valid given their current configuration. Similarly, MLM allows users to iterate on their model to improve it. Here, users can take advantage of the physical data points by spreading them out and looking at their similarities and differences to make hypotheses about what inferences the model is making, and then test their hypotheses by strategically making new data-points and testing or training their data on these.

For end-user systems, it might not always be appropriate to allow users to experiment freely with their AI components, but systems could allow their users to do so in a sandbox environment without affecting the actual system. This would allow users to get a sense of what happens to (their) data in the system, and how different choices made by its developers or maintainers affect them. This functionality could be further expanded by the inclusion of xAI functionality, although this was not the case in the above systems.

E.5.3 Facilitating Discussions and Reconfiguration to allow Interrogation and Negotiation

Finally, to scaffold students in negotiating the use of AI systems and their effect on their lives, the systems invite for discussion and reconfiguration. An integral part of both the Ethics workshop and VML is that the systems and ML models produced by students are scrutinized and discussed by a wider audience. Additionally, in the Ethics workshop, students are explicitly asked to redesign their system to address the issues they have encountered. While this is often difficult for the students, it highlights the trade-offs between, e.g., privacy and effectiveness made by real-world designers and developers of AI systems.

While a completely redesign of an end-user system by the user might not be desirable, systems could be designed to reflexively engage their users in configuring the systems to their preference. This might entail allowing users to choose between privacy preserving data collection or increased accuracy of predictions, or to let users choose between models that are advanced but opaque or simple but explainable. Further, this strategy could be deployed to design systems where multiple stakeholders are affected by their use of AI to help them negotiate how it's deployed. By highlighting how these systems' use of AI interweave with their users' daily lives and surroundings, users are supported in interrogating the systems' roles in these.

E.6 Making an Unremarkable Clinical Decision Support Tool Remarkable

To exemplify how the Remarkable AI design strategies might be used by designers of end-user systems, we present a redesign of the clinical decision support tool (DST) by Yang, Steinfeld, and Zimmerman [238]. We stress that we do not believe that this redesign would result in a better DST, rather it serves to highlight the complementary nature of Remarkable and Unremarkable AI, as well as their differences. A sketch of this Remarkable DST-redesign can be found in Figure E.4.

E.6.1 An Unremarkable DST System

The original unremarkable DST system by Yang, Steinfeld, and Zimmerman [238] was designed to "not constrain clinicians' decision making flow *except when it needs to*" [238, p. 3] (emphasis in original). Decision support systems fail when they move from labs to clinics largely because they do not meet the needs of clinicians nor to they fit within the busy and chaotic context of clinical practice. The system is designed to be used in patient evaluation meetings, where clinicians meet to decide the best course of action for a patient. To ease the adoption of the system, it automatically pulls in patient data and is able to generate slides for this meeting with the relevant information. The output of the system's AI component is implemented as line chart in the top right corner of the display: "The subtlety was a deliberate choice toward achieving the right level of unremarkableness" [238, p. 3]. The majority of mechanical implant decisions are textbook, and the clinicians have very little uncertainty about what to do. This means that most of the time, the AI system adds no value, it simply tells clinicians what they already

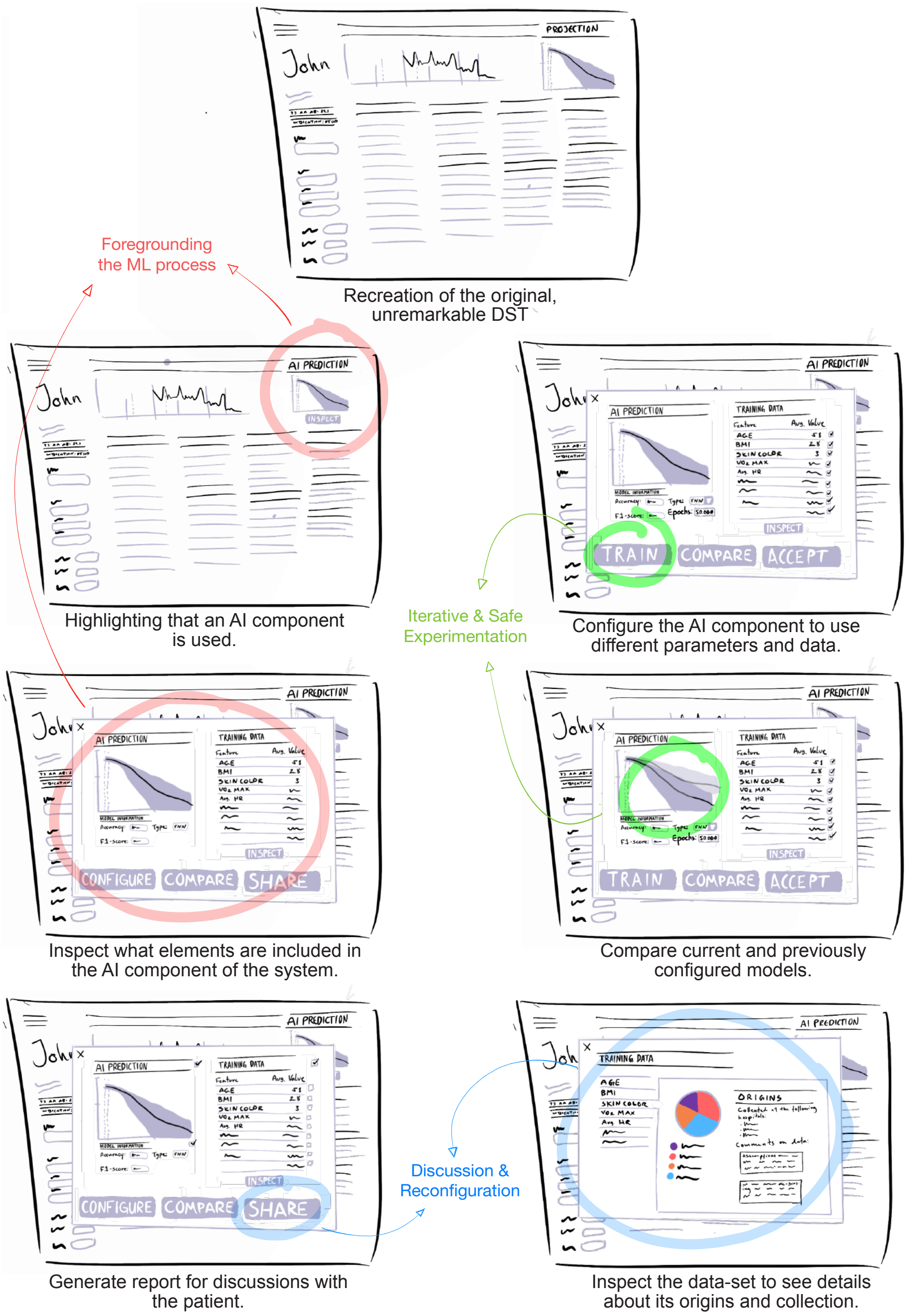


Figure E.4: Sketches detailing the original and redesigned DST system know, and thus was designed to allow clinicians to glance over the information, and ignore it if they agree with its assessment. Sometime there is disagreement within a clinical team. When this happens between mid-levels (nurses, residents, medical students) and attending physicians, the hierarchical nature of current practice makes it hard for lower status workers to challenge higher status workers. When the AI system suggests actions different than the high status workers, it provides a way for lower status workers to suggest this case might need more consideration. When there is disagreement, the unremarkable interface adds a little bit of friction to slow decision making down. The design wants to mostly be out of the way of the experts, but to occasionally add a little friction and slow decisions for cases that are likely less text book. A sketch recreating the original design can be seen at the top of Figure E.4

E.6.2 Foregrounding the ML Process

A DST designed for Remarkable AI, might expose the ML process by foregrounding it in the interface. First, this would mean highlighting that the projection is made by an AI system in the first place. Further, it might entail providing doctors the opportunity to inspect this projection by, e.g., exploring the data-set that the model is trained on, and what optimization-objectives it aims for. The data-set could be visualized to provide doctors an overview of how the data is demographically distributed and how possible imbalances in the data are dealt with. Insights into the model's optimization objectives could provide doctors with information about possible shortcomings such as what margin of error is allowed in the model's predictions and how it balances precision versus accuracy.

E.6.3 Iterative & Safe Experimentation

Before making their decision, doctors could be provided an interface allowing them to experiment with the input data of the model (possibly in collaboration with the patient or other doctors) to filter or balance out different features that might affect the model's predictions such as skin color, age or sex, and to inspect if and how this might change the outcome of the model. Further, doctors could be asked to provide their own "prediction" before seeing the prediction of the AI system. This would invite doctors to reflect on differences in the projections and on whether and how they themselves or the system might be biased.

E.6.4 Facilitating Discussion and Reconfiguration

The original unremarkable DST is already designed to present the AI projection as an integral part of doctors' routines and context. This could be expanded to include the context of the patient as well. For example, the system could ensure that if an AI projection was considered in doctors decisions, this is reflected in their patient records. Further, the patient should be made aware and consented before their data is included in the further improvement of the AI system, and perhaps even be included in discussing and (re)configuring if and how AI is used in decisions made regarding their treatment. Some patients might object to an AI system diagnosing them, preferring the decision be made by humans irregardless that it might be less precise. A remarkable DST could include a patient interface, that allows them to monitor how doctors use its AI component or to pre-configure their preference towards its use.

Finally, and perhaps separate to a concrete decision situation, doctors could routinely be asked to reflect on how the inclusion of AI-driven decision support tools changes the craft of being a medical doctor.

E.7 Discussion

As we see it, Remarkable AI and Unremarkable AI [238] are two ends of a continuum, delimiting the design-space for remarkableness in AI-enabled application, which come with each their strengths and weaknesses. To introduce the idea of this continuum, we have drawn on previous work in HCI regarding Remarkable [178] and Unremarkable Computing [215]. We did so to highlight how contrasting approaches can foster constructive discussions in the HCI-community, and since the criticisms of Unremarkable Computing by Petersen [178] are as applicable to Unremarkable AI [238]. However, as mentioned earlier, because AI works in the background and a user might never realize that they are using a AI-enabled system, Remarkable AI is further reaching than its Computing counterpart.

Below, we discuss the necessity, we see, for considering remarkableness in everyday AI-enabled applications, how to find the right level of remarkableness, as well as our method for exploring the new design-space of remarkableness in AI systems design and the limitations of Remarkable AI.

E.7.1 Designing a Right Level of Remarkableness

As mentioned above, we fully recognize that Remarkable AI is not a catch-all solution for designing AI-enabled systems. Instead, we see Remarkable and Unremarkable AI as delineating a continuum of AI systems design, in which designers will need to ask themselves (and their users) important questions regarding the nature of their needs and wants for a system. As much as designers might want to support and/or augment our users' decision-making, all systems are, arguably, transformative in the sense that they necessarily influence users [225].

In recognizing this, an important question to raise is whether it is acceptable for a system to do so indirectly, or to which degree users should be confronted with and offered control over this influence. This of course depends on the application-area and not least the values of the designer. The case of DSTs are an interesting one, since both sides of the argument could be made; Yang, Steinfeld, and Zimmerman [238] argue that a DST should support clinicians work-practices and only be noticed when their predictions "add value", in which case the Unremarkable approach is most appropriate. However, if one begins to question what value is added and whom decides that it's valuable, the case is less clear cut; the patient might not wish to have AI play a role in their decision-making, or clinicians might wish to know more about how the DST makes its predictions, what data these are based on, and if biases is present, before trusting it. We are not suggesting that a remarkably designed DST would be more effective or indeed that clinicians would be more inclined to use it (in fact the opposite might be true), but to point out the political and ethical questions that AI-enabled systems rise and that we believe should be addressed in any design process.

As a helpful analogy to guide designers in finding the right level of remarkableness, we suggest looking at how, we use epistemology when teaching, e.g., a history class. In early grades, we teach children important events throughout our history, but leave out how we know the details of these historic events. Later, in an appropriate grade, we then introduce epistemology, e.g., understanding the difference between primary, secondary, and tertiary sources, etc. To know when to introduce these ideas, educators need to consider when students' can be expected to become critical thinkers, and on what level of abstraction they can be expected to think. Similarly, designers of AI systems need to consider whether or not the stakeholders of their systems would benefit from a remarkable design or not. Determining a right level of remarkableness is a design choice, which designers should engage with, and which answer should be informed by an understanding of the system's stakeholders, context of use, etc. Indeed, it might be the case that a system clearly would not benefit from being remarkable and that designing it to be so would only subtract from the value it provides (arguably, one example of this could be ML-powered, on-device enhancement of smartphone photography). However, we stress that the right level of remarkableness is highly contextual, and as such designers should engage rigorously with determining it, even if the AI system's use-case seems inconsequential at surface-level.

While we do see Remarkable and Unremarkable AI as two end-points in a continuum, we see no reason why they could not be combined in different parts of a timeline and/or context of use. It might be that clinicians at first wish to inspect the AI-component of their DST, but as they come to trust it they might be less inclined to do so; a sort of appropriation-through-use instead of Petersen [178]'s notion of "learning-through-use". Or alternatively, clinicians might use this DST in an unremarkable way, observe that it is helpful, and then feel the wish to unfold its inner workings.

E.7.2 Method Reflection & Limitations

We find that our method of exploring the design-space of Remarkable AI through cases from a different domain, AI Education, was helpful for understanding and delimiting this new problem space, and we suggest that other designers, dealing with a novel design-space might benefit from a similar process. This method is, arguably, particularly useful for rapid exploration of the boundaries of new, and yet-to-be defined problem-spaces. The idea is similar to methods for reviewing and delimiting emergent design-spaces previously used in HCI [182, 219], however it differs from these by looking forward to a yet unknown design-space, rather that trying to characterize one that is emergent. This however, requires that designers have intimate knowledge of and access to research projects which in someway can inform the new design-space.

We do not intend for all AI-enabled systems to remarkable, but rather that designers consider the remarkableness of these systems. As mentioned above, there are systems, which would not benefit from being Remarkable; Remarkableness adds a layer of complexity and potentially slow down decision processes. This might be inappropriate for systems where speed is a determining factor, or where tasks are so minute or marginal that adding this complexity would simply not make sense. Similarly, it asks that all users have, at least to some degree, an understanding of AI. This is of course what many in AI Education are working towards (e.g. [137, 216], however, we are not there yet, and might not be for quite a while. Regarding professionals, the design of remarkable AI systems should also go hand in hand with lifelong education for professionals, who may benefit from insights gained from advanced analysis of data. In addition, the re-configuring part of Remarkable AI sets demands for the availability of models and data sources which may often not be readily available nor interchangeable.

E.7.3 Future Work

As these ideas are still new, and continuously growing, we invite other researchers to join in maturing the idea of remarkableness in AI-enabled systems design, and of striking the balance between unremarkable and remarkable AI. Above we have presented both high-level design sensitivities derived from the literature as well as concrete design strategies from our case studies. Neither of these are comprehensive in the design-space, and our wish is to see research both in suggesting new and/or other high-level sensitivities as well as other design strategies. Finally, we invite other HCI researchers to utilize our design strategies in their own work for designing Remarkable AI systems.

E.8 Conclusion

In this paper, we have highlighted the need for considering remarkableness in AI systems design, and have detailed Remarkable AI as one end of the remarkableness continuum, with Yang, Steinfeld, and Zimmerman [238]'s Unremarkable AI at the other. To qualify Remarkable AI we have drawn in existing literature on designing for reflection, computing education and to participatory design. Based on the literature we suggest two design sensitivities for Remarkable AI systems; Exposing & Explaining AI and Interrogating & Negotiating the Role of AI, and through analyzing design cases from the AI education domain, we contribute design strategies for designing remarkable AI. Finally, we emphasize that designing the right level of remarkableness must be considered in each specific design case, and what are suitable levels for a specific case might also change over time and with specific user groups. Finally, we invite other researchers to join the further exploration of the design space between unremarkable and remarkable AI.

Bibliography

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda". en. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, Apr. 2018, pp. 1–18. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174156. URL: https://dl.acm.org/doi/10.1145/3173574.3174156 (visited on 09/29/2022).

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. "Guidelines for Human-AI Interaction". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–13. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300233. URL: https://doi.org/10.1145/3290605.3300233 (visited on 03/23/2022).

[3] Nick Anstead. "Data-Driven Campaigning in the 2015 United Kingdom General Election". In: *The International Journal of Press/Politics* 22.3 (2017). Publisher: SAGE Publications, pp. 294–313. ISSN: 1940-1612. DOI: 10.1177/1940161217706163. URL: http://journals.sagepub.com/doi/10.1177/1940161217706163.

[4] James Auger. "Speculative design: crafting the speculation". In: *Digital Creativity* 24.1 (Mar. 2013). Publisher: Routledge _eprint: https://doi.org/10.1080/14626268.2013.767276, pp. 11–35. ISSN: 1462-6268. DOI: 10.1080/14626268.2013.767276. URL: https://doi.org/10.1080/14626268.2013.767276 (visited on 11/28/2022).

[5] Z.H. Aung, S. Sanium, C. Songsaksuppachok, W. Kusakunniran, M. Precharattana, S. Chuechote, K. Pongsanon, and P. Ritthipravat. "Designing a novel teaching platform for AI: A case study in a Thai school context". English. In: *Journal of Computer Assisted Learning* 38.6 (2022), pp. 1714–1729. ISSN: 0266-4909. DOI: 10.1111/jcal.12706.

[6] Anne Louise Bang and Mette Agger Eriksen. "Experiments all the Way in Programmatic Design Research". en. In: *ARTIFACT* III.2 (2014), p. 14.

[7] Anne Louise Bang, Peter Gall Krogh, Martin Ludvigsen, and Thomas Markussen. "The Role of Hypothesis in Constructive Design Research". In: *The Art of Research 2012: Making, Reflecting and understanding*. Issue: 4. Helsinki, 2012, pp. 1–11.

[8] Liam Bannon, Jeffrey Bardzell, and Susanne Bødker. "Reimagining participatory design". en. In: *Interactions* 26.1 (Dec. 2018), pp. 26–32. ISSN: 1072-5520, 1558-3449. DOI: 10.1145/3292015. URL: https://dl.acm.org/doi/10.1145/3292015 (visited on 11/21/2022).

[9] Ronald Barnett. *Engaging the curriculum in higher education*. Maidenhead, England New York: Society for Research into Higher Education Open University Press, 2005. ISBN: 0-335-21289-1.

[10] Ronald Barnett. "Knowing and becoming in the higher education curriculum". In: *Studies in Higher Education* 34.4 (May 2009). Publisher: Informa UK Limited, pp. 429–440. DOI: 10.1080/03075070902771978. URL: https://doi.org/10.1080/03075070902771978.

[11] Solon Barocas and Andrew D Selbst. "Big data's disparate impact". In: *California Law Review* 104 (2016).

[12] Valerie Barr and Chris Stephenson. "Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community?" In: *ACM Inroads* 2.1 (Mar. 2011), pp. 48–54. ISSN: 21532184. DOI: 10.1145/1929887.1929905. URL: https://dl.acm.org/doi/10.1145/1929887.1929905.

[13] Eric P. S. Baumer and Jed R. Brubaker. "Post-Userism". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. event-place: Denver, Colorado, USA. New York, NY, USA: Association for Computing Machinery, 2017, pp. 6291–6303. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025740. URL: https://doi.org/10.1145/3025453.3025740.

[14] Beth T Bell. "Understanding adolescents". In: *Perspectives on HCI research with teenagers*. Springer, 2016, pp. 11–27.

[15] Tim Bell, Frances Rosamond, and Nancy Casey. *Computer science unplugged and related projects in math and computer science popularization*. ISBN: 9783642308901 ISSN: 03029743 Pages: 398–456 Publication Title: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Volume: 7370. 2012. DOI: 10.1007/978-3-642-30891-8_18. URL: https://link.springer.com/chapter/10.1007/978-3-642-30891-8_18.

[16] Marina Umaschi Bers, Louise Flannery, Elizabeth R. Kazakoff, and Amanda Sullivan. "Computational thinking and tinkering: Exploration of an early childhood robotics curriculum". In: *Computers and Education* 72 (2014). Publisher: Elsevier Ltd, pp. 145–157. ISSN: 03601315. DOI: 10.1016/j.compedu.2013.10.020. URL: http://dx.doi.org/10.1016/j.compedu.2013.10.020.

[17] John Biggs. *Teaching for quality learning at university*. 2nd ed. The Society for Research into Higher Education and Open University Press, Buckingham, 2003. ISBN: 0-335-21168-2.

[18] John B. Biggs and Kevin F. Collis. *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. en. Google-Books-ID: xUO0BQAAQBAJ. New York: Academic Press, 1982. ISBN: 978-1-4832-7331-0.

[19] Karl-Emil Kjær Bilstrup, Magnus H. Kaspersen, and Marianne Graves Petersen. "Staging Reflections on Ethical Dilemmas in Machine Learning: A Card-Based Design Workshop for High School Students". In: *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. DIS '20. New York, NY, USA: Association for Computing Machinery, July 2020, pp. 1211–1222. ISBN: 978-1-4503-6974-9. DOI: 10.1145/3357236.3395558. URL: https://doi.org/10.1145/3357236.3395558 (visited on 10/31/2022).

[20] Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Simon Enni, Ira Assent, and Marianne Graves Petersen. "From Demo to Design in Teaching Machine Learning (in press)". In: *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. Seoul, South Korea: ACM, 2022.

[21] Thomas Binder and Johan Redström. "Exemplary Design Research". In: *Design Research Society Wonderground International Conference 2006*. 2006.

[22] Gro. Bjerknes, Pelle Ehn, Morten Kyng, and Kristen Nygaard. *Computers and democracy : a Scandinavian challenge*. Gower Pub Co, 1987. isbn: 978-0-566-05476-1. url: http://www.worldcat.org/title/computers-and-democracy-a-scandinavian-challenge/oclc/614994092?referer=di&ht=edition.

[23] Paulo Blikstein. "Gears of our childhood: constructionist toolkits, robotics, and physical computing, past and future". In: *Proceedings of the 12th International Conference on Interaction Design and Children*. IDC '13. New York, NY, USA: Association for Computing Machinery, June 2013, pp. 173–182. isbn: 978-1-4503-1918-8. doi: 10.1145/2485760.2485786. url: https://doi.org/10.1145/2485760.2485786 (visited on 12/20/2022).

[24] Benjamin S Bloom, Max D Englehart, Edward J Furst, Walter H Hill, and David R Krathwohl. "Taxonomy of Educational Objectives: Handbook I Cognitive Domain". In: *New York, New York: David McKay Company* (1956), pp. 144–45.

[25] Herbert Blumer. "What is Wrong with Social Theory?" In: *American Sociological Review* 19.1 (1954). Publisher: American Sociological Association, Sage Publications, Inc., pp. 3–10. issn: 0003-1224. doi: 10.2307/2088165. url: https://www.jstor.org/stable/2088165 (visited on 05/23/2022).

[26] Susanne Bødker. "When second wave HCI meets third wave challenges". en. In: *Proceedings of the 4th Nordic conference on Human-computer interaction changing roles - NordiCHI '06*. Oslo, Norway: ACM Press, 2006, pp. 1–8. isbn: 978-1-59593-325-6. doi: 10.1145/1182475.1182476. url: http://portal.acm.org/citation.cfm?doid=1182475.1182476 (visited on 03/16/2022).

[27] Susanne Bødker, Christian Dindler, and Ole Sejer Iversen. "Tying Knots: Participatory Infrastructuring at Work". en. In: *Computer Supported Cooperative Work (CSCW)* 26.1 (Apr. 2017), pp. 245–273. issn: 1573-7551. doi: 10.1007/s10606-017-9268-y. url: https://doi.org/10.1007/s10606-017-9268-y (visited on 11/28/2022).

[28] Susanne Bødker, Pelle Ehn, Joergen Knudsen, Morten Kyng, and Kim Madsen. "Computer support for cooperative design (invited paper)". en. In: *Proceedings of the 1988 ACM conference on Computer-supported cooperative work - CSCW '88*. Portland, Oregon, United States: ACM Press, 1988, pp. 377–394. isbn: 978-0-89791-282-2. doi: 10.1145/62266.62296. url: http://dl.acm.org/citation.cfm?doid=62266.62296 (visited on 06/20/2022).

[29] Susanne Bødker, Kaj Grønbæk, and Morten Kyng. "Cooperative Design: Techniques and Experiences From the Scandinavian Scene". en. In: *Readings in Human–Computer Interaction*. Elsevier, 1995, pp. 215–224. isbn: 978-0-08-051574-8. doi: 10.1016/B978-0-08-051574-8.50025-X. url: https://linkinghub.elsevier.com/retrieve/pii/B978008051574850025X (visited on 06/20/2022).

[30] Susanne Bødker and Morten Kyng. "Participatory Design that Matters - Facing the Big Issues". In: *ACM Transactions on Computer-Human Interaction* 25.1 (Feb. 2018), 4:1–4:31. issn: 1073-0516. doi: 10.1145/3152421. url: https://doi.org/10.1145/3152421 (visited on 02/03/2022).

[31] Susanne Bødker, Morten Kyng, Pelle Ehn, John Kammersgaard, and Yngve Sundblad. "A utopian experience: On design of powerful computer-based tools for skilled graphic workers". In: *Computers and democracy*. Gower Publishing, 1987, pp. 251–278.

[32] Leanne Bowler, Amelia Acker, Wei Jeng, and Yu Chi. "“It lives all around us”: Aspects of data literacy in teen's lives". en. In: *Proceedings of the Association for Information Science and Technology* 54.1 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.2017.14505401004, pp. 27–35. ISSN: 2373-9231. DOI: 10.1002/pra2.2017.14505401004. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pra2.2017.14505401004 (visited on 12/07/2022).

[33] Karen Brennan and Mitchel Resnick. "New frameworks for studying and assessing the development of computational thinking". en. In: *Proceedings of the 2012 annual meeting of the American Educational Research Association*. Vancouver, Canada, 2012, p. 25.

[34] Brian Broll, Shuchi Grover, and Derek Babb. "Beyond Black-Boxing: Building Intuitions of Complex Machine Learning Ideas Through Interactives and Levels of Abstraction". In: *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 2*. Vol. 2. ICER '22. New York, NY, USA: Association for Computing Machinery, Aug. 2022, pp. 21–23. ISBN: 978-1-4503-9195-5. DOI: 10.1145/3501709.3544273. URL: https://doi.org/10.1145/3501709.3544273 (visited on 12/22/2022).

[35] Geoffrey Brown. *The information game: Ethical issues in a microchip world*. Humanities Press International Inc., 1990.

[36] Jenna Burrell. "How the machine 'thinks': Understanding opacity in machine learning algorithms". In: *Big Data & Society* 3.1 (2016). Publisher: SAGE Publications Sage UK: London, England, p. 2053951715622512.

[37] Carole Cadwalladr and Emma Graham-Harrison. "The Cambridge analytica files". In: *The Guardian* (2018), pp. 6–7. URL: https://www.theguardian.com/news/series/cambridge-analytica-files.

[38] Anne-Emmanuèle Calvès. "Empowerment: The History of a Key Concept in Contemporary Development Discourse". EN. In: *Revue Tiers Monde* 200.4 (2009). Place: Paris Publisher: Armand Colin, pp. 735–749. ISSN: 9782200926038. DOI: 10.3917/rtm.200.0735. URL: https://www.cairn-int.info/journal-revue-tiers-monde-2009-4-page-735.htm.

[39] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. "Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–8. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3382839. URL: https://doi.org/10.1145/3334480.3382839 (visited on 12/22/2022).

[40] John M. Carroll and Robert L. Campbell. "Softening Up Hard Science: Reply to Newell and Card". In: *Human–Computer Interaction* 2.3 (Sept. 1986). Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1207/s15327051hci0203_3, pp. 227–249. ISSN: 0737-0024. DOI: 10.1207/s15327051hci0203_3. URL: https://www.tandfonline.com/doi/abs/10.1207/s15327051hci0203_3 (visited on 06/14/2022).

[41] John M. Carroll and Wendy A. Kellogg. "Artifact as theory-nexus: hermeneutics meets theory-based design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '89. New York, NY, USA: Association for Computing Machinery, Mar. 1989, pp. 7–14. ISBN: 978-0-89791-301-0. DOI: 10.1145/67449.67452. URL: https://doi.org/10.1145/67449.67452 (visited on 11/22/2022).

[42] Matthew Chalmers and Ian MacColl. "Seamful and Seamless Design in Ubiquitous Computing". en. In: *Workshop at the crossroads: The interaction of HCI and systems issues in UbiComp*. 2003, p. 9.

[43] Po-Hsuan Cameron Chen, Yun Liu, and Lily Peng. "How to develop machine learning models for healthcare". In: *Nature materials* 18.5 (2019). Publisher: Nature Publishing Group, p. 410.

[44] Kevin F. Collis and H. A. Davey. "A technique for evaluating skills in high school science". In: *Journal of Research in Science Teaching* 23.7 (Oct. 1986). Publisher: Wiley, pp. 651–663. DOI: 10.1002/tea.3660230707. URL: https://doi.org/10.1002/tea.3660230707.

[45] Matthew B Crawford. *Shop class as soulcraft: An inquiry into the value of work*. Penguin, 2009.

[46] Peter Dalsgaard and Christian Dindler. "Between theory and practice: bridging concepts in HCI research". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 1635–1644. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557342. URL: https://doi.org/10.1145/2556288.2557342 (visited on 11/22/2022).

[47] Peter Dalsgaard and Kim Halskov. "Designing urban media façades: Cases and challenges". In: *Conference on Human Factors in Computing Systems - Proceedings*. Vol. 4. New York, New York, USA: ACM Press, 2010, pp. 2277–2286. ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753670. URL: http://portal.acm.org/citation.cfm?doid=1753326.1753670.

[48] Xiaozhou Deng, Danli Wang, Qiao Jin, and Fang Sun. "ARCat: A tangible programming tool for DFS algorithm teaching". In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children, IDC 2019*. Vol. 19. New York, New York, USA: Association for Computing Machinery, Inc, June 2019, pp. 533–537. ISBN: 978-1-4503-6690-8. DOI: 10.1145/3311927.3325308. URL: http://dl.acm.org/citation.cfm?doid=3311927.3325308.

[49] Peter J Denning. "Great principles of computing". In: *Communications of the ACM* 46.11 (2003), pp. 11–15. DOI: https://dl.acm.org/doi/10.1145/948383.948400.

[50] Christian Dindler, Rachel Smith, and Ole Sejer Iversen. "Computational empowerment: participatory design in education". In: *CoDesign* 16.1 (2020). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15710882.2020.1722173, pp. 66–80. DOI: 10.1080/15710882.2020.1722173. URL: https://doi.org/10.1080/15710882.2020.1722173.

[51] Stefania Druga. *Growing up with AI - Cognimates: from coding to teaching machines*. Backup Publisher: Massachusetts Institute of Technology Pages: 1–204. 2018. URL: http://hdl.handle.net/1721.1/120691.

[52] Stefania Druga, Nancy Otero, and Amy J. Ko. "The Landscape of Teaching Resources for AI Education". In: *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*. ITiCSE '22. New York, NY, USA: Association for Computing Machinery, July 2022, pp. 96–102. ISBN: 978-1-4503-9201-3. DOI: 10.1145/3502718.3524782. URL: https://doi.org/10.1145/3502718.3524782 (visited on 12/22/2022).

[53] Stefania Druga, Sarah T. Vu, Eesh Likhith, and Tammy Qiu. "Inclusive AI literacy for kids around the world". In: *Proceedings of FabLearn 2019*. FL2019. New York, NY, USA: Association for Computing Machinery, Mar. 2019, pp. 104–111. ISBN: 978-1-4503-6244-3. DOI: 10.1145/3311890.3311904. URL: https://doi.org/10.1145/3311890.3311904 (visited on 10/26/2022).

[54] Anthony Dunne and Fiona Raby. *Speculative everything: design, fiction, and social dreaming*. Cambridge, Massachusetts; London: The MIT Press, 2013. ISBN: 978-0-262-01984-2.

[55] Simon Enni. "Deliberation and Dissemination in Machine Learning". PhD Thesis. Aarhus University, 2021.

[56] Simon Enni and Ira Assent. *Learning by Design: Structuring and Documenting the Human Choices in Machine Learning Development*. arXiv:2105.00687. May 2021. DOI: 10.48550/arXiv.2105.00687. URL: http://arxiv.org/abs/2105.00687 (visited on 01/28/2023).

[57] Eva Eriksson, Gökçe Elif Baykal, Staffan Björk, and Olof Torgersson. "Using Gameplay Design Patterns with Children in the Redesign of a Collaborative Co-located Game". In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, 2019, pp. 15–25.

[58] Eva Eriksson, Carl Heath, Peter Ljungstrand, and Peter Parnes. "Makerspace in school—Considerations from a large-scale national testbed". In: *International Journal of Child-Computer Interaction* 16 (June 2018). Publisher: Elsevier BV, pp. 9–15. DOI: 10.1016/j.ijcci.2017.10.001. URL: https://doi.org/10.1016/j.ijcci.2017.10.001.

[59] Eva Eriksson, Ole Sejer Iversen, Gökçe Elif Baykal, Maarten Van Mechelen, Rachel Smith, Marie-Louise Wagner, Bjarke Vognstrup Fog, Clemens Klokmose, Bronwyn Cumbo, Arthur Hjorth, Line Have Musaeus, Marianne Graves Petersen, and Niels Olof Bouvin. "Widening the Scope of FabLearn Research: Integrating Computational Thinking, Design and Making". In: *Proceedings of the FabLearn Europe 2019 Conference*. FabLearn Europe '19. event-place: Oulu, Finland. New York, NY, USA: ACM, 2019, 15:1–15:9. ISBN: 978-1-4503-6266-5. DOI: 10.1145/3335055.3335070. URL: http://doi.acm.org/10.1145/3335055.3335070.

[60] Julian Estevez, Gorka Garate, and Manuel Graña. "Gentle Introduction to Artificial Intelligence for High-School Students Using Scratch". In: *IEEE Access* 7 (2019). Conference Name: IEEE Access, pp. 179027–179036. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2956136.

[61] Jerry Alan Fails, Allison Druin, Mona Leigh Guha, Gene Chipman, Sante Simms, and Wayne Churaman. "Child's Play: A Comparison of Desktop and Physical Interactive Environments". In: *Proceedings of the 2005 Conference on Interaction Design and Children*. New York, NY, USA: Association for Computing Machinery, 2005, pp. 48–55. ISBN: 1-59593-096-5. DOI: https://doi.org/10.1145/1109540.1109547.

[62] Daniel Fitton, Janet C C. Read, and Matthew Horton. "The challenge of working with teens as participants in interaction design". en. In: *CHI EA '13 CHI '13 Extended Abstracts on Human Factors in Computing Systems*. Paris, France: ACM Press, 2013, p. 205. ISBN: 978-1-4503-1952-2. DOI: 10.1145/2468356.2468394. URL: http://dl.acm.org/citation.cfm?doid=2468356.2468394.

[63] Kirstin Fontichiaro and Jo Angela Oehrli. "Why Data Literacy Matters". en. In: *Knowledge Quest* 44.5 (2016). Publisher: American Association of School Librarians ERIC Number: EJ1099487, pp. 21–27. ISSN: 1094-9046. URL: https://eric.ed.gov/?id=EJ1099487 (visited on 12/07/2022).

[64] Tom Forester and Perry Morrison. *Computer ethics: cautionary tales and ethical dilemmas in computing*. Mit Press, 1994.

[65] Mark Frank, Johanna Walker, Judie Attard, and Alan Tygel. "Data Literacy - What is it and how can we make it happen?" In: *The Journal of Community Informatics* 12.3 (Sept. 2016). Number: 3. ISSN: 1712-4441. DOI: 10.15353/joci.v12i3.3274. URL: https://ojs.uwaterloo.ca/index.php/JoCI/article/view/3274 (visited on 12/07/2022).

[66] Christopher Frauenberger, Marjo Rauhala, and Geraldine Fitzpatrick. "In-Action Ethics". In: *Interacting with Computers* 29.2 (Jan. 2016), pp. 220–236. ISSN: 0953-5438. DOI: 10.1093/iwc/iww024. URL: https://doi.org/10.1093/iwc/iww024.

[67] Paulo Freire. *Pedagogy of the Oppressed*. Routledge, 1978. ISBN: 978-0-429-33953-0.

[68] Batya Friedman and David G. Hendry. "The Envisioning Cards: A toolkit for catalyzing humanistic and technical imaginations". In: *Conference on Human Factors in Computing Systems - Proceedings* (2012). ISBN: 9781450310154, pp. 1145–1148. DOI: 10.1145/2207676.2208562.

[69] Gerald Futschek and Julia Moschitz. "Learning algorithmic thinking with tangible objects eases transition to computer programming". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 7013 LNCS. ISSN: 03029743. Springer, Berlin, Heidelberg, Oct. 2011, pp. 155–164. ISBN: 978-3-642-24721-7. DOI: 10.1007/978-3-642-24722-4_14. URL: https://link.springer.com/chapter/10.1007/978-3-642-24722-4_14.

[70] Bill Gaver and John Bowers. "Annotated portfolios". In: *Interactions* 19.4 (July 2012), pp. 40–49. ISSN: 1072-5520. DOI: 10.1145/2212877.2212889. URL: https://doi.org/10.1145/2212877.2212889.

[71] William Gaver. "What should we expect from research through design?" en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin Texas USA: ACM, May 2012, pp. 937–946. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2208538. URL: https://dl.acm.org/doi/10.1145/2207676.2208538 (visited on 09/29/2022).

[72] Maliheh Ghajargar, Jeffrey Bardzell, Alison Marie Smith-Renner, Kristina Höök, and Peter Gall Krogh. "Graspable AI: Physical Forms as Explanation Modality for Explainable AI". In: *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI '22. New York, NY, USA: Association for Computing Machinery, Feb. 2022, pp. 1–4. ISBN: 978-1-4503-9147-4. DOI: 10.1145/3490149.3503666. URL: https://doi.org/10.1145/3490149.3503666 (visited on 12/23/2022).

[73] Michail Giannakos, Iro Voulgari, Sofia Papavlasopoulou, Zacharoula Papamitsiou, and Georgios Yannakakis. “Games for Artificial Intelligence and Machine Learning Education: Review and Perspectives”. In: *Non-Formal and Informal Science Learning in the ICT Era*. Ed. by Michail Giannakos. Singapore: Springer Singapore, 2020, pp. 117–133. ISBN: 978-981-15-6747-6. DOI: 10.1007/978-981-15-6747-6_7. URL: https://doi.org/10.1007/978-981-15-6747-6_7.

[74] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. “Dark patterns in proxemic interactions: A critical perspective”. In: *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS* (2014). ISBN: 9781450329026, pp. 523–532. DOI: 10.1145/2598510.2598541.

[75] Jean M. Griffin. “Constructionism and De-Constructionism as Complementary Pedagogies”. In: *Constructionism 2018*. Issue: August. 2018, pp. 225–237.

[76] Shuchi Grover and Roy Pea. “Computational Thinking in K–12: A Review of the State of the Field”. In: *Educational Researcher* 42.1 (2013). _eprint: https://doi.org/10.3102/0013189X12463051, pp. 38–43. DOI: 10.3102/0013189X12463051. URL: https://doi.org/10.3102/0013189X12463051.

[77] Jonathan Grudin. “AI and HCI: Two Fields Divided by a Common Focus”. en. In: *AI Magazine* 30.4 (Sept. 2009), p. 48. ISSN: 0738-4602, 0738-4602. DOI: 10.1609/aimag.v30i4.2271. URL: https://aaai.org/ojs/index.php/aimagazine/article/view/2271 (visited on 09/29/2022).

[78] David Gunning. *Explainable artificial intelligence (xai)*. Tech. rep. 2017. URL: https://nsarchive.gwu.edu/sites/default/files/documents/5794867/National-Security-Archive-David-Gunning-DARPA.pdf (visited on 01/28/2023).

[79] Mark Guzdial. “Education Paving the way for computational thinking”. In: *Communications of the ACM* 51.8 (Aug. 2008), pp. 25–27. ISSN: 0001-0782. DOI: 10.1145/1378704.1378713. URL: https://doi.org/10.1145/1378704.1378713 (visited on 11/10/2022).

[80] Lars Hallnäs and Johan Redström. *Interaction Design: Foundations, Experiments*. eng. Textile Research Centre, Swedish School of Textiles, Unversity College of Borås and Interactive Institute, 2006. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-21279 (visited on 11/07/2022).

[81] Lars Hallnäs and Johan Redström. “Slow Technology – Designing for Reflection”. en. In: *Personal and Ubiquitous Computing* 5.3 (Aug. 2001), pp. 201–212. ISSN: 1617-4909. DOI: 10.1007/PL00000019. URL: http://link.springer.com/10.1007/PL00000019 (visited on 03/23/2022).

[82] Kim Halskov and Peter Dalsgård. “Inspiration card workshops”. In: *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*. Vol. 2006. 2006, pp. 2–11. ISBN: 1-59593-341-7. DOI: 10.1145/1142405.1142409.

[83] Erica Rosenfeld Halverson and Kimberly Sheridan. “The Maker Movement in Education”. In: *Harvard Educational Review* 84.4 (2014), pp. 495–504. DOI: 10.17763/haer.84.4.34j1g68140382063. URL: https://doi.org/10.17763/haer.84.4.34j1g68140382063.

[84] Susanne Hambrusch, Christoph Hoffmann, John T. Korb, Mark Haugan, and Antony L. Hosking. “A multidisciplinary approach towards computational thinking for science majors”. In: *ACM SIGCSE Bulletin* 41.1 (Mar. 2009), pp. 183–187. ISSN: 0097-8418. DOI: 10.1145/1539024.1508931. URL: https://doi.org/10.1145/1539024.1508931 (visited on 11/10/2022).

[85] Elin Irene Krog Hansen and Ole Sejer Iversen. "You are the real experts!: Studying teenagers' motivation in participatory design". en. In: *IDC '13 Proceedings of the 12th International Conference on Interaction Design and Children*. New York, New York, USA: ACM Press, 2013, pp. 328–331. ISBN: 978-1-4503-1918-8. DOI: 10.1145/2485760.2485826. URL: http://dl.acm.org/citation.cfm?doid=2485760.2485826.

[86] Idit Harel and Seymour Papert. *Constructionism*. Ablex Publishing, 1991.

[87] John Hattie. *Visible Learning*. Routledge, Nov. 2008. DOI: 10.4324/9780203887332. URL: https://doi.org/10.4324/9780203887332.

[88] John Hattie and Gavin TL Brown. *Cognitive processes in asTTle: The SOLO taxonomy*. Ministry of Education, 2004.

[89] John A. C. Hattie and Gregory M. Donoghue. "Learning strategies: a synthesis and conceptual model". en. In: *npj Science of Learning* 1.1 (Aug. 2016). Number: 1 Publisher: Nature Publishing Group, pp. 1–13. ISSN: 2056-7936. DOI: 10.1038/npjscilearn.2016.13. URL: https://www.nature.com/articles/npjscilearn201613 (visited on 11/21/2022).

[90] Martin Heidegger. *Being and time*. State University of New York Press, 1996.

[91] Joseph R. Herkert. "Ways of thinking about and teaching ethical problem solving: Microethics and macroethics in engineering". In: *Science and Engineering Ethics*. Vol. 11. ISSN: 13533452 Issue: 3. Kluwer Academic Publishers, 2005, pp. 373–385. DOI: 10.1007/s11948-005-0006-3.

[92] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. "Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. event-place: Glasgow, Scotland Uk. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–11. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300645. URL: https://doi.org/10.1145/3290605.3300645.

[93] Cindy E. Hmelo and Mark Guzdial. "Of Black and Glass Boxes: Scaffolding for Doing and Learning". In: *Proceedings of the 1996 International Conference on Learning Sciences*. ICLS '96. event-place: Evanston, Illinois. International Society of the Learning Sciences, 1996, pp. 128–134. ISBN: 1-880094-23-1.

[94] Pam Hook. *First steps with SOLO taxonomy: applying the model in your classroom*. Laughton, United Kingdom Strawberry Hills, NSW, Australia Invercargill, New Zealand: Essential Resources, 2016. ISBN: 1-77655-255-5.

[95] Kristina Höök. "Steps to take before intelligent user interfaces become real". In: *Interacting with Computers* 12.4 (Feb. 2000). Conference Name: Interacting with Computers, pp. 409–426. ISSN: 1873-7951. DOI: 10.1016/S0953-5438(99)00006-5.

[96] Kristina Höök and Jonas Löwgren. "Strong concepts: Intermediate-level knowledge in interaction design research". In: *ACM Transactions on Computer-Human Interaction* 19.3 (Oct. 2012), 23:1–23:18. ISSN: 1073-0516. DOI: 10.1145/2362364.2362371. URL: https://doi.org/10.1145/2362364.2362371 (visited on 11/22/2022).

[97] Michael S. Horn, Erin Treacy Solovey, R. Jordan Crouser, and Robert J.K. Jacob. "Comparing the use of tangible and graphical programming languages for informal science education". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: Association for Computing Machinery, Apr. 2009, pp. 975–984. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518851. URL: https://doi.org/10.1145/1518701.1518851 (visited on 11/10/2022).

[98] Matthew B. Hoy. “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants”. In: *Medical Reference Services Quarterly* 37.1 (Jan. 2018). Publisher: Routledge _eprint: https://doi.org/10.1080/02763869.2018.1404391, pp. 81–88. ISSN: 0276-3869. DOI: 10.1080/02763869.2018.1404391. URL: https://doi.org/10.1080/02763869.2018.1404391 (visited on 11/30/2022).

[99] Felix Hu, Ariel Zekelman, Michael Horn, and Frances Judd. “Strawbies: Explorations in tangible programming”. In: *Proceedings of IDC 2015: The 14th International Conference on Interaction Design and Children*. New York, New York, USA: Association for Computing Machinery, Inc, June 2015, pp. 410–413. ISBN: 978-1-4503-3590-4. DOI: 10.1145/2771839.2771866. URL: http://dl.acm.org/citation.cfm?doid=2771839.2771866.

[100] Netta Iivari. “Empowering children to make and shape our digital futures – from adults creating technologies to children transforming cultures”. In: *The International Journal of Information and Learning Technology* 37.5 (Jan. 2020). Publisher: Emerald Publishing Limited, pp. 279–293. ISSN: 2056-4880. DOI: 10.1108/IJILT-03-2020-0023. URL: https://doi.org/10.1108/IJILT-03-2020-0023 (visited on 01/17/2023).

[101] Netta Iivari, Sumita Sharma, Leena Ventä-Olkkonen, Tonja Molin-Juustila, Kari Kuutti, Jenni Holappa, and Essi Kinnunen. “Critical agenda driving child–computer interaction research—Taking a stock of the past and envisioning the future”. en. In: *International Journal of Child-Computer Interaction* 32 (June 2022), p. 100408. ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2021.100408. URL: https://www.sciencedirect.com/science/article/pii/S2212868921000957 (visited on 09/29/2022).

[102] Sarah Inman and David Ribes. “"Beautiful Seams": Strategic Revelations and Concealments”. en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–14. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300508. URL: https://dl.acm.org/doi/10.1145/3290605.3300508 (visited on 06/14/2022).

[103] Ole Sejer Iversen and Christian Dindler. “Sustaining participatory design initiatives”. In: *CoDesign* 10.3-4 (July 2014). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15710882.2014.963124, pp. 153–170. ISSN: 1571-0882. DOI: 10.1080/15710882.2014.963124. URL: https://doi.org/10.1080/15710882.2014.963124 (visited on 01/23/2023).

[104] Ole Sejer Iversen, Christian Dindler, and Elin Irene Krogh Hansen. “Understanding teenagers’ motivation in participatory design”. en. In: *International Journal of Child-Computer Interaction* 1.3-4 (Sept. 2013), pp. 82–87. ISSN: 22128689. DOI: 10.1016/j.ijcci.2014.02.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S221286891400004X.

[105] Ole Sejer Iversen and Rachel Charlotte Smith. “Scandinavian participatory design”. In: *Proceedings of the 11th International Conference on Interaction Design and Children - IDC ’12*. New York, New York, USA: ACM Press, 2012, p. 106. ISBN: 978-1-4503-1007-9. DOI: 10.1145/2307096.2307109. URL: http://dl.acm.org/citation.cfm?doid=2307096.2307109.

[106] Ole Sejer Iversen, Rachel Charlotte Smith, and Christian Dindler. “Child as Protagonist: Expanding the Role of Children in Participatory Design”. In: *Proceedings of the 2017 Conference on Interaction Design and Children*. IDC ’17. New York, NY, USA: Association for Computing Machinery, June 2017, pp. 27–37. ISBN: 978-1-4503-4921-5. DOI: 10.1145/3078072.3079725. URL: https://doi.org/10.1145/3078072.3079725 (visited on 12/21/2022).

[107] Ole Sejer Iversen, Rachel Charlotte Smith, and Christian Dindler. "From computational thinking to computational empowerment: a 21st century PD agenda". In: *Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1*. 2018, pp. 1–11.

[108] Jakob G. Jensby, Anders E. Pedersen, and Peter Brøndum. *Politikbogen*. 3rd ed. Forlaget Columbus, 2017. ISBN: 978-87-7970-380-3. URL: https://forlagetcolumbus.dk/produkt/politikbogen/produkt/vis/.

[109] Sonam Joshi. *Teenage girl coders from Mumbai slum are building apps to solve local problems*. 2016. URL: https://mashable.com/2016/03/29/mumbai-dharavi-girls-coding-apps/?europe=true.

[110] Yasmin Kafai, Chris Proctor, and Debora Lui. "From Theory Bias to Theory Dialogue: Embracing Cognitive, Situated, and Critical Framings of Computational Thinking in K-12 CS Education". In: *Proceedings of the 2019 ACM Conference on International Computing Education Research*. ICER '19. event-place: Toronto ON, Canada. New York, NY, USA: Association for Computing Machinery, 2019, pp. 101–109. ISBN: 978-1-4503-6185-9. DOI: 10.1145/3291279.3339400. URL: https://doi.org/10.1145/3291279.3339400.

[111] Yasmin Kafai, Chris Proctor, and Debora Lui. "From theory bias to theory dialogue: embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education". In: *ACM Inroads* 11.1 (Feb. 2020), pp. 44–53. ISSN: 2153-2184. DOI: 10.1145/3381887. URL: https://doi.org/10.1145/3381887 (visited on 11/25/2022).

[112] Yasmin B Kafai and Mitchel Resnick. *Constructionism in practice: Designing, thinking, and learning in a digital world*. Routledge, 2012.

[113] Yasmin B. Kafai and Quinn Burke. "The social turn in K-12 programming: moving from computational thinking to computational participation". In: *Proceeding of the 44th ACM technical symposium on Computer science education*. SIGCSE '13. New York, NY, USA: Association for Computing Machinery, Mar. 2013, pp. 603–608. ISBN: 978-1-4503-1868-6. DOI: 10.1145/2445196.2445373. URL: https://doi.org/10.1145/2445196.2445373 (visited on 09/29/2022).

[114] Yasmin B. Kafai, Chris Proctor, and Debora A. Lui. "Framing Computational Thinking for Computational Literacies in K-12 Education". en. In: *Weizenbaum Conference 2019 "Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life"*. Publisher: WI - Weizenbaum Institute for the Networked Society Version Number: 1. 2019. DOI: 10.34669/WI.CP/2.21. URL: https://www.ssoar.info/ssoar/handle/document/62615 (visited on 09/29/2022).

[115] Ken Kahn, Rani Megasari, Erna Piantari, and Enjun Junaeti. "AI programming by children using Snap! block programming in a developing country". en. In: *Thirteenth European Conference on Technology Enhanced Learning* 11082 (2018). ISBN: 9783319985725 Publisher: Springer. ISSN: 0302-9743. URL: https://ora.ox.ac.uk/objects/uuid:9a82b522-6f9f-4c67-b20d-be6c53019b3b (visited on 12/22/2022).

[116] Peter Kahn. "Critical perspectives on methodology in pedagogic research". In: *Teaching in Higher Education* 20.4 (May 2015). Publisher: Routledge _eprint: https://doi.org/10.1080/13562517.2015.1023286, pp. 442–454. ISSN: 1356-2517. DOI: 10.1080/13562517.2015.1023286. URL: https://doi.org/10.1080/13562517.2015.1023286 (visited on 11/21/2022).

[117] Michael Karlberg. “The Power of Discourse and the Discourse of Power: Pursuing Peace Through Discourse Intervention”. In: *International Journal of Peace Studies* 10.1 (2005). Publisher: International Peace Research Association (IPRA), pp. 1–25. ISSN: 1085-7494. URL: https://www.jstor.org/stable/41852070 (visited on 12/20/2022).

[118] Magnus Høholt Kaspersen and Karl-Emil Kjær Bilstrup. “VotestratesML: Civics as a Vehicle for Teaching Machine Learning”. In: *Constructionism*. 2020.

[119] Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, Maarten Van Mechelen, Arthur Hjort, Niels Olof Bouvin, and Marianne Graves Petersen. “VotestratesML: A High School Learning Tool for Exploring Machine Learning and its Societal Implications”. In: *FabLearn Europe / MakeEd 2021 - An International Conference on Computing, Design and Making in Education (FabLearn Europe / MakeEd 2021)*. ACM, New York, NY, USA, 2021, p. 14. DOI: https://doi.org/10.1145/3466725.3466728.

[120] Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen. “The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning”. In: *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI ’21. Association for Computing Machinery, Feb. 2021, pp. 1–12. ISBN: 978-1-4503-8213-7. DOI: 10.1145/3430524.3440638. URL: https://doi.org/10.1145/3430524.3440638.

[121] Irni Eliana Khairuddin, Corina Sas, and Chris Speed. “BlocKit: A physical kit for materializing and designing for blockchain infrastructure”. In: *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*. New York, NY, USA: Association for Computing Machinery, Inc, June 2019, pp. 1449–1462. ISBN: 978-1-4503-5850-7. DOI: 10.1145/3322276.3322370. URL: https://dl.acm.org/doi/10.1145/3322276.3322370.

[122] Helyn Kim and Care Esther. *Learning progressions: Pathways for 21st century teaching and learning*. 2018. URL: https://www.brookings.edu/blog/education-plus-development/2018/03/27/learning-progressions-pathways-for-21st-century-teaching-and-learning/ (visited on 01/28/2023).

[123] Marianne Kinnula and Netta Iivari. “Empowered to Make a Change: Guidelines for Empowering the Young Generation in and through Digital Technology Design”. In: *Proceedings of the FabLearn Europe 2019 Conference*. FabLearn Europe ’19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–8. ISBN: 978-1-4503-6266-5. DOI: 10.1145/3335055.3335071. URL: https://doi.org/10.1145/3335055.3335071 (visited on 09/29/2022).

[124] Marianne Kinnula, Netta Iivari, Tonja Molin-Juustila, Topi Leinonen, Eetu Mansikkamäki, Toni Käkelä, and Martti Similä. “Cooperation, Combat, or Competence Building – What Do We Mean When We Are ‘Empowering Children’ in and through Digital Technology Design?” en. In: *Thirty eighth International Conference on Information Systems* (2017), p. 21.

[125] Marianne Kinnula, Netta Iivari, Sumita Sharma, Grace Eden, Markku Turunen, Krishnashree Achuthan, Prema Nedungadi, Tero Avellan, Biju Thankachan, and Rucha Tulaskar. “Researchers’ Toolbox for the Future: Understanding and Designing Accessible and Inclusive Artificial Intelligence (AIAI)”. In: *Proceedings of the 24th International Academic Mindtrek Conference*. Academic Mindtrek ’21. New York, NY, USA: Association for Computing Machinery, Sept. 2021, pp. 1–4. ISBN: 978-1-4503-8514-5. DOI: 10.1145/3464327.3464965. URL: https://doi.org/10.1145/3464327.3464965 (visited on 01/17/2023).

[126] Clemens Klokmose, James Eagan, Siemen Baader, Wendy Mackay, and Michel Beaudouin-Lafon. "Webstrates: demonstrating the potential of Shareable Dynamic Media". In: *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*. New York, New York, USA: ACM Press, 2016, pp. 61–64. ISBN: 978-1-4503-3950-6. DOI: 10.1145/2818052.2874325. URL: http://dl.acm.org/citation.cfm?doid=2818052.2874325.

[127] Ilpo Koskinen and Peter Gall Krogh. "Design Accountability: When Design Research Entangles Theory and Practice". In: *International Journal of Design* 9.1 (Apr. 2015), pp. 121–127. ISSN: 19913761. URL: https://search.ebscohost.com/login.aspx?direct=true&db=aft&AN=109190283&site=ehost-live (visited on 11/07/2022).

[128] Ilpo Koskinen, John Zimmerman, Thomas Binder, Johan Redstrom, and Stephan Wensveen. *Design Research Through Practice: From the Lab, Field, and Showroom*. Elsevier, Sept. 2011. ISBN: 978-0-12-385503-9.

[129] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental evidence of massive-scale emotional contagion through social networks". In: *Proceedings of the National Academy of Sciences* 111.24 (June 2014). Publisher: Proceedings of the National Academy of Sciences, pp. 8788–8790. DOI: 10.1073/pnas.1320040111. URL: https://www.pnas.org/doi/full/10.1073/pnas.1320040111 (visited on 11/30/2022).

[130] Peter Gall Krogh and Ilpo Koskinen. *Drifting by Intention: Four Epistemic Traditions from within Constructive Design Research*. en. Design Research Foundations. Cham: Springer International Publishing, 2020. ISBN: 978-3-030-37895-0 978-3-030-37896-7. DOI: 10.1007/978-3-030-37896-7. URL: http://link.springer.com/10.1007/978-3-030-37896-7 (visited on 09/29/2022).

[131] Peter Gall Krogh, Thomas Markussen, and Anne Louise Bang. "Ways of Drifting—Five Methods of Experimentation in Research Through Design". en. In: *ICoRD'15 – Research into Design Across Boundaries Volume 1*. Ed. by Amaresh Chakrabarti. Smart Innovation, Systems and Technologies. New Delhi: Springer India, 2015, pp. 39–50. ISBN: 978-81-322-2232-3. DOI: 10.1007/978-81-322-2232-3_4.

[132] Daria J. Kuss and Mark D. Griffiths. "Social Networking Sites and Addiction: Ten Lessons Learned". en. In: *International Journal of Environmental Research and Public Health* 14.3 (Mar. 2017). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 311. ISSN: 1660-4601. DOI: 10.3390/ijerph14030311. URL: https://www.mdpi.com/1660-4601/14/3/311 (visited on 11/28/2022).

[133] Lucian Leahu, Phoebe Sengers, and Michael Mateas. "Interactionist AI and the promise of ubicomp, or, how to put your box in the world without putting the world in your box". en. In: *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*. Seoul, Korea: ACM Press, 2008, p. 134. ISBN: 978-1-60558-136-1. DOI: 10.1145/1409635.1409654. URL: http://portal.acm.org/citation.cfm?doid=1409635.1409654 (visited on 02/18/2022).

[134] Irene Lee, Fred Martin, Walter Allan, Jeri Erickson, Linda Werner, Jill Denner, Bob Coulter, and Joyce Malyn-Smith. "Computational thinking for youth in practice". en. In: *ACM Inroads* 2.1 (Feb. 2012). ISBN: 2153-2184, p. 32. ISSN: 21532184. DOI: 10.1145/1929887.1929902. URL: http://dl.acm.org/citation.cfm?doid=1929887.1929902.

[135] Annabel Lindner, Stefan Seegerer, and Ralf Romeike. "Unplugged Activities in the Context of AI". en. In: *Informatics in Schools. New Ideas in School Informatics*. Ed. by Sergei N. Pozdniakov and Valentina Dagienė. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 123–135. ISBN: 978-3-030-33759-9. DOI: 10.1007/978-3-030-33759-9_10.

[136] Michael Lodi and Simone Martini. "Computational Thinking, Between Papert and Wing". en. In: *Science & Education* 30.4 (Aug. 2021), pp. 883–908. ISSN: 1573-1901. DOI: 10.1007/s11191-021-00202-5. URL: https://doi.org/10.1007/s11191-021-00202-5 (visited on 11/28/2022).

[137] Duri Long and Brian Magerko. "What is AI Literacy? Competencies and Design Considerations". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. event-place: Honolulu, HI, USA. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–16. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376727. URL: https://doi.org/10.1145/3313831.3376727.

[138] Mille Skovhus Lunding, Jens Emil Sloth Grønbæk, Karl-Emil Kjær Bilstrup, Marie-Louise Stisen Kjerstein Sørensen, and Marianne Graves Petersen. "ExposAR: Bringing Augmented Reality to the Computational Thinking Agenda through a Collaborative Authoring Tool". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–14. ISBN: 978-1-4503-9157-3. DOI: 10.1145/3491102.3517636. URL: https://doi.org/10.1145/3491102.3517636 (visited on 01/05/2023).

[139] Ezio Manzini. *Design, when everybody designs: an introduction to design for social innovation*. eng. Design Thinking, Design Theory. Cambridge, Massachusetts ; The MIT Press, 2015. ISBN: 978-0-262-32864-7.

[140] Gary Marcus. "Deep learning: A critical appraisal". In: *arXiv preprint arXiv:1801.00631* (2018). DOI: https://doi.org/10.48550/arXiv.1801.00631.

[141] Radu Mariescu-Istodor and Ilkka Jormanainen. "Machine Learning for High School Students". In: *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*. Koli Calling '19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1–9. ISBN: 978-1-4503-7715-7. DOI: 10.1145/3364510.3364520. URL: https://doi.org/10.1145/3364510.3364520 (visited on 12/22/2022).

[142] Paul Marshall. "Do tangible interfaces enhance learning?" In: *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction*. New York, NY, USA: Association for Computing Machinery, 2007. ISBN: 978-1-59593-619-6. DOI: https://doi.org/10.1145/1226969.1227004.

[143] Paul Marshall, Sara Price, and Yvonne Rogers. "Conceptualising tangibles to support learning". In: *Proceedings of the 2003 conference on Interaction design and children*. IDC '03. New York, NY, USA: Association for Computing Machinery, July 2003, pp. 101–109. ISBN: 978-1-58113-732-3. DOI: 10.1145/953536.953551. URL: https://doi.org/10.1145/953536.953551 (visited on 11/10/2022).

[144] Anna Mavroudi, Monica Divitini, Francesco Gianni, Simone Mora, and Dag R. Kvittem. "Designing IoT applications in lower secondary schools". In: *IEEE Global Engineering Education Conference, EDUCON* 2018-April (2018). ISBN: 9781538629574 Publisher: IEEE, pp. 1120–1126. ISSN: 21659567. DOI: 10.1109/EDUCON.2018.8363355.

[145] Matthew B. Miles and A. Michael Huberman. *Qualitative Data Analysis: An expanded sourcebook (3rd ed.)* London, UK: SAGE Publications, 1994.

[146] Tim Miller. "Explanation in artificial intelligence: Insights from the social sciences". In: *Artificial Intelligence* 267 (2019). Publisher: Elsevier, pp. 1–38.

[147] Tim Miller, Piers Howe, and Liz Sonenberg. *Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences*. Tech. rep. arXiv:1712.00547. arXiv:1712.00547 [cs] type: article. arXiv, Dec. 2017. DOI: 10.48550/arXiv.1712.00547. URL: http://arxiv.org/abs/1712.00547 (visited on 05/19/2022).

[148] Brent Mittelstadt. "Automation, algorithms, and politics| auditing for transparency in content personalization systems". In: *International Journal of Communication* 10 (2016), p. 12.

[149] Preben Holst Mogensen and Randall H. Trigg. "Using Artifacts as Triggers for Participatory Analysis". en. In: *DAIMI Report Series* 21.413 (1992). ISSN: 0105-8517. DOI: 10.7146/dpb.v21i413.6726. URL: https://tidsskrift.dk/daimipb/article/view/6726.

[150] Bill Moggridge and Bill Atkinson. *Designing interactions*. Vol. 17. MIT press Cambridge, MA, 2007.

[151] James H Moor. "What is computer ethics?" In: *Metaphilosophy* 16.4 (1985). Publisher: Wiley Online Library, pp. 266–275.

[152] Simone Mora, Francesco Gianni, and Monica Divitini. "Tiles: A card-based ideation toolkit for the Internet of Things". In: *DIS 2017 - Proceedings of the 2017 ACM Conference on Designing Interactive Systems* (2017). ISBN: 9781450349222, pp. 587–598. DOI: 10.1145/3064663.3064699.

[153] Erin Murphy-Graham. "And when she comes home? Education and women's empowerment in intimate relationships". en. In: *International Journal of Educational Development* 30.3 (May 2010), pp. 320–331. ISSN: 07380593. DOI: 10.1016/j.ijedudev.2009.09.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S073805930900131X (visited on 10/27/2021).

[154] Line Have Musaeus and Peter Musaeus. "Computational Thinking in the Danish High School". In: *Proceedings of the 50th ACM Technical Symposium on Computer Science Education - SIGCSE '19*. New York, New York, USA: ACM Press, 2019, pp. 913–919. ISBN: 978-1-4503-5890-3. DOI: 10.1145/3287324.3287452. URL: http://dl.acm.org/citation.cfm?doid=3287324.3287452.

[155] Allen Newell and Stuart Card. "Straightening Out Softening Up: Response to Carroll and Campbell". In: *Human–Computer Interaction* 2.3 (Sept. 1986). Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1207/s15327051hci0203_4, pp. 251–267. ISSN: 0737-0024. DOI: 10.1207/s15327051hci0203_4. URL: https://www.tandfonline.com/doi/abs/10.1207/s15327051hci0203_4 (visited on 06/14/2022).

[156] Allen Newell and Stuart K. Card. "The Prospects for Psychological Science in Human-Computer Interaction". en. In: *Human–Computer Interaction* 1.3 (Sept. 1985), pp. 209–242. ISSN: 0737-0024, 1532-7051. DOI: 10.1207/s15327051hci0103_1. URL: http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0103_1 (visited on 06/14/2022).

[157] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. "Conceptualizing AI literacy: An exploratory review". en. In: *Computers and Education: Artificial Intelligence* 2 (Jan. 2021), p. 100041. ISSN: 2666-920X. DOI: 10.1016/j.caeai.2021.100041. URL: https://www.sciencedirect.com/science/article/pii/S2666920X21000357 (visited on 10/26/2022).

[158] David W. Nickerson and Todd Rogers. "Political Campaigns and Big Data". In: *Journal of Economic Perspectives* 28.2 (May 2014), pp. 51–74. ISSN: 0895-3309. DOI: 10.1257/jep.28.2.51. URL: http://pubs.aeaweb.org/doi/10.1257/jep.28.2.51.

[159] Hilarie Nickerson, Catharine Brand, and Alexander Repenning. "Grounding Computational Thinking Skill Acquisition Through Contextualized Instruction". In: *Proceedings of the eleventh annual International Conference on International Computing Education Research*. ICER '15. New York, NY, USA: Association for Computing Machinery, Aug. 2015, pp. 207–216. ISBN: 978-1-4503-3630-7. DOI: 10.1145/2787622.2787720. URL: https://doi.org/10.1145/2787622.2787720 (visited on 11/10/2022).

[160] Donald A. Norman. "How might people interact with agents". In: *Communications of the ACM* 37.7 (July 1994), pp. 68–71. ISSN: 0001-0782. DOI: 10.1145/176789.176796. URL: https://doi.org/10.1145/176789.176796 (visited on 12/23/2022).

[161] Kristen Nygaard and Olav Terje Bergo. "The Trade Unions - New users of research". In: *Personnel Review* 4.2 (Jan. 1975). Publisher: MCB UP Ltd, pp. 5–10. ISSN: 0048-3486. DOI: 10.1108/eb055278. URL: https://doi.org/10.1108/eb055278 (visited on 06/20/2022).

[162] Keith O'Hara, Douglas Blank, and James Marshall. "Computational Notebooks for AI Education". In: *The Twenty-Eighth International Flairs Conference* (Apr. 2015).

[163] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, and Council on Communications and Media. "The Impact of Social Media on Children, Adolescents, and Families". In: *Pediatrics* 127.4 (Apr. 2011), pp. 800–804. ISSN: 0031-4005. DOI: 10.1542/peds.2011-0054. URL: https://doi.org/10.1542/peds.2011-0054 (visited on 11/28/2022).

[164] Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. en. Google-Books-ID: cbwvDwAAQBAJ. Crown, 2016. ISBN: 978-0-553-41883-5.

[165] Michelle O'Reilly, Nisha Dogra, Natasha Whiteman, Jason Hughes, Seyda Eruyar, and Paul Reilly. "Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents". en. In: *Clinical Child Psychology and Psychiatry* 23.4 (Oct. 2018). Publisher: SAGE Publications Ltd, pp. 601–613. ISSN: 1359-1045. DOI: 10.1177/1359104518775154. URL: https://doi.org/10.1177/1359104518775154 (visited on 11/28/2022).

[166] OECD. *The OECD Learning Compass 2030*. 2019.

[167] Elisaweta Ossovski and Michael Brinkmeier. "Machine Learning Unplugged - Development and Evaluation of a Workshop About Machine Learning". en. In: *Informatics in Schools. New Ideas in School Informatics*. Ed. by Sergei N. Pozdniakov and Valentina Dagienė. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 136–146. ISBN: 978-3-030-33759-9. DOI: 10.1007/978-3-030-33759-9_11.

[168] Sofia Papavlasopoulou, Michail N. Giannakos, and Letizia Jaccheri. "Empirical studies on the Maker Movement, a promising approach to learning: A literature review". en. In: *Entertainment Computing* 18 (Jan. 2017). ISBN: 18759521, pp. 57–78. ISSN: 18759521. DOI: 10.1016/j.entcom.2016.09.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1875952116300301.

[169] Seymour Papert. *Constructionism: A new opportunity for elementary science education*. Massachusetts Institute of Technology, Media Laboratory, Epistemology and Learning Group, 1986.

[170] Seymour Papert and Idit Harel. “Situating Constructionism”. In: *Constructionism* 36.2 (1991), pp. 1–11. URL: https://scholar.googleusercontent.com/scholar.bib?q=info:YbUPvtd1K6oJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAXIJjl_S92vEnukYHIJ6r3wnBt98Isf5u&scisf=4&ct=citation&cd=-1&hl=en.

[171] Seymour A Papert. *Mindstorms: Children, computers, and powerful ideas.* Basic books, 2020.

[172] European Parliament and the Council. *Recommendation of the European Parliament and of the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning.* Publication Title: Official Journal of the European Union. 2008.

[173] Frank Pasquale. *The black box society: The secret algorithms that control money and information.* Harvard University Press, 2015.

[174] Fernando Perez and Brian E. Granger. *Project Jupyter: Computational Narratives as the Engine of Collaborative Data Science.* Tech. rep. UC Berkeley & Cal Poly, 2015. URL: http://archive.ipython.org/JupyterGrantNarrative-2015.pdf.

[175] Ljubomir Perković, Amber Settle, Sungsoon Hwang, and Joshua Jones. “A framework for computational thinking across the curriculum”. In: *ITiCSE'10 - Proceedings of the 2010 ACM SIGCSE Annual Conference on Innovation and Technology in Computer Science Education* (2010). ISBN: 9781605588209, pp. 123–127. DOI: 10.1145/1822090.1822126.

[176] Walt L Perry. *Predictive policing: The role of crime forecasting in law enforcement operations.* Rand Corporation, 2013.

[177] Marianne Graves Petersen. “Designing for Learning in Use of Everyday Artefacts”. In: *DAIMI Report Series* 32.568 (Jan. 2002). ISSN: 2245-9316, 0105-8517. DOI: 10.7146/dpb.v32i568.7125. URL: https://tidsskrift.dk/daimipb/article/view/7125 (visited on 06/20/2022).

[178] Marianne Graves Petersen. “Remarkable computing: the challenge of designing for the home”. en. In: *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04.* Vienna, Austria: ACM Press, 2004, p. 1445. ISBN: 978-1-58113-703-3. DOI: 10.1145/985921.986086. URL: http://portal.acm.org/citation.cfm?doid=985921.986086 (visited on 02/17/2022).

[179] Jean Piaget. “Piaget's theory”. In: *Piaget and His school - A Reader in Developmental Psychology.* Ed. by B. Inhelder, H.H. Chipman, and C. Zwingmann. New York, New York, USA: Springer, 1976, pp. 11–23.

[180] Volkmar Pipek and Volker Wulf. “Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology”. In: *Journal of the Association for Information Systems* 10.5 (May 2009), pp. 447–473. ISSN: 15369323. DOI: 10.17705/1jais.00195. URL: https://aisel.aisnet.org/jais/vol10/iss5/1/ (visited on 06/17/2022).

[181] James Popham. “All about accountability / the lowdown on learning progressions”. In: *Eductional Leadership* 64.7 (2007), pp. 83–84.

[182] Majken K. Rasmussen, Esben W. Pedersen, Marianne G. Petersen, and Kasper Hornbæk. “Shape-Changing Interfaces: A Review of the Design Space and Open Research Questions”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '12. event-place: Austin, Texas, USA. New York, NY, USA: Association for Computing Machinery, 2012, pp. 735–744. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2207781. URL: https://doi.org/10.1145/2207676.2207781.

[183] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, and Yasmin Kafai. "Scratch: programming for all". en. In: *Communications of the ACM* 52.11 (Nov. 2009), pp. 60–67. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/1592761.1592779. URL: https://dl.acm.org/doi/10.1145/1592761.1592779 (visited on 12/20/2022).

[184] Mitchel Resnick and Brian Silverman. "Some reflections on designing construction kits for kids". In: *Proceedings of the 2005 conference on Interaction design and children*. IDC '05. New York, NY, USA: Association for Computing Machinery, June 2005, pp. 117–122. ISBN: 978-1-59593-096-5. DOI: 10.1145/1109540.1109556. URL: https://doi.org/10.1145/1109540.1109556 (visited on 11/10/2022).

[185] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. "On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. FAT* '19. New York, NY, USA: Association for Computing Machinery, Jan. 2019, pp. 140–149. ISBN: 978-1-4503-6125-5. DOI: 10.1145/3287560.3287580. URL: https://doi.org/10.1145/3287560.3287580 (visited on 01/10/2023).

[186] Drew Roselli, Jeanna Matthews, and Nisha Talagala. "Managing Bias in AI". In: *Companion Proceedings of The 2019 World Wide Web Conference*. WWW '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 539–544. ISBN: 978-1-4503-6675-5. DOI: 10.1145/3308560.3317590. URL: https://doi.org/10.1145/3308560.3317590 (visited on 01/04/2023).

[187] Daniele Rotolo, Diana Hicks, and Ben R. Martin. "What is an emerging technology?" en. In: *Research Policy* 44.10 (Dec. 2015), pp. 1827–1843. ISSN: 0048-7333. DOI: 10.1016/j.respol.2015.06.006. URL: https://www.sciencedirect.com/science/article/pii/S0048733315001031 (visited on 01/05/2023).

[188] Robin Roy and James P. Warren. "Card-based design tools: A review and analysis of 155 card decks for designers and designing". In: *Design Studies* 63.Figure 1 (2019). Publisher: Elsevier Ltd, pp. 125–154. ISSN: 0142694X. DOI: 10.1016/j.destud.2019.04.002. URL: https://doi.org/10.1016/j.destud.2019.04.002.

[189] Stuart J. Russell, Peter Norvig, and Ernest Davis. *Artificial intelligence: a modern approach*. 3rd ed. Prentice Hall series in artificial intelligence. Upper Saddle River: Prentice Hall, 2010. ISBN: 978-0-13-604259-4.

[190] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. "Auditing algorithms: Research methods for detecting discrimination on internet platforms". In: *Data and discrimination: converting critical concerns into productive inquiry* 22 (2014), pp. 4349–4357.

[191] Christine Satchell and Paul Dourish. "Beyond the User: Use and Non-Use in HCI". In: *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*. OZCHI '09. event-place: Melbourne, Australia. New York, NY, USA: Association for Computing Machinery, 2009, pp. 9–16. ISBN: 978-1-60558-854-4. DOI: 10.1145/1738826.1738829. URL: https://doi.org/10.1145/1738826.1738829.

[192] Marie-Monique Schaper, Rachel Charlotte Smith, Mariana Aki Tamashiro, Maarten Van Mechelen, Mille Skovhus Lunding, Karl-Emil Kjæer Bilstrup, Magnus Høholt Kaspersen, Kasper Løvborg Jensen, Marianne Graves Petersen, and Ole Sejer Iversen. "Computational empowerment in practice: Scaffolding teenagers' learning about emerging technologies and their ethical and societal

impact". en. In: *International Journal of Child-Computer Interaction* 34 (Dec. 2022), p. 100537. ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2022.100537. URL: https://www.sciencedirect.com/science/article/pii/S2212868922000551 (visited on 12/05/2022).

[193] Alexander Scheidt and Tim Pulver. "Any-Cubes: A Children's Toy for Learning AI: Enhanced Play with Deep Learning and MQTT". In: *Proceedings of Mensch und Computer 2019*. MuC'19. New York, NY, USA: Association for Computing Machinery, Sept. 2019, pp. 893–895. ISBN: 978-1-4503-7198-8. DOI: 10.1145/3340764.3345375. URL: https://doi.org/10.1145/3340764.3345375 (visited on 12/22/2022).

[194] Donald A. Schön. *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. Pages: xvii, 355. San Francisco, CA, US: Jossey-Bass, 1987. ISBN: 978-1-55542-025-3.

[195] Donald A. Schön. *The reflective practitioner. How professionals think in action*. London: Temple Smith, 1983.

[196] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. "Reflective design". en. In: *Proceedings of the 4th decennial conference on Critical computing between sense and sensibility - CC '05*. Aarhus, Denmark: ACM Press, 2005, p. 49. ISBN: 978-1-59593-203-7. DOI: 10.1145/1094562.1094569. URL: http://portal.acm.org/citation.cfm?doid=1094562.1094569 (visited on 03/10/2022).

[197] Pratim Sengupta, John S. Kinnebrew, Satabdi Basu, Gautam Biswas, and Douglas Clark. "Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework". en. In: *Education and Information Technologies* 18.2 (June 2013), pp. 351–380. ISSN: 1573-7608. DOI: 10.1007/s10639-012-9240-x. URL: https://doi.org/10.1007/s10639-012-9240-x (visited on 11/10/2022).

[198] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. "Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors". In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW2 (Oct. 2021), 433:1–433:29. DOI: 10.1145/3479577. URL: https://doi.org/10.1145/3479577 (visited on 12/23/2022).

[199] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholoas Diakopoulos. "Grand Challenges for HCI Researchers". In: *Interactions* 23.5 (Aug. 2016). Place: New York, NY, USA Publisher: Association for Computing Machinery, pp. 24–25. ISSN: 1072-5520. DOI: 10.1145/2977645. URL: https://doi.org/10.1145/2977645.

[200] Jenni AM Sidey-Gibbons and Chris J Sidey-Gibbons. "Machine learning in medicine: a practical introduction". In: *BMC medical research methodology* 19.1 (2019). Publisher: Springer, p. 64.

[201] Shourya Pratap Singh, Ankit Kumar Panda, Susobhit Panigrahi, Ajaya Kumar Dash, and Debi Prosad Dogra. "PlutoAR: An Inexpensive, Interactive And Portable Augmented Reality Based Interpreter For K-10 Curriculum". In: (Sept. 2018). URL: http://arxiv.org/abs/1809.00375.

[202] Zoe Skinner, Stacey Brown, and Greg Walsh. "Children of Color's Perceptions of Fairness in AI: An Exploration of Equitable and Inclusive Co-Design". en. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, pp. 1–8. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3382901. URL: https://dl.acm.org/doi/10.1145/3334480.3382901 (visited on 12/23/2022).

[203] Rachel Charlotte Smith and Ole Sejer Iversen. “Participatory design for sustainable social change”. In: *Design Studies* 59 (2018). Publisher: Elsevier Ltd, pp. 9–36. ISSN: 0142694X. DOI: 10.1016/j.destud.2018.05.005. URL: https://doi.org/10.1016/j.destud.2018.05.005.

[204] Rachel Charlotte Smith, Ole Sejer Iversen, and Mikkel Hjorth. “Design thinking for digital fabrication in education”. en. In: *International Journal of Child-Computer Interaction*. Digital Fabrication in Education 5 (Sept. 2015), pp. 20–28. ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2015.10.002. URL: https://www.sciencedirect.com/science/article/pii/S2212868915000203 (visited on 01/06/2023).

[205] Pieter Jan Stappers and Elisa Giaccardi. “Research through Design”. In: *The Encyclopedia of Human-Computer Interaction, 2nd edition*. Ed. by M. Soegaard and R. Friis-Dam. 2017.

[206] Susan Leigh Star and Karen Ruhleder. “Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces”. en. In: *Information Systems Research* 7.1 (Mar. 1996), pp. 111–134. ISSN: 1047-7047, 1526-5536. DOI: 10.1287/isre.7.1.111. URL: http://pubsonline.informs.org/doi/abs/10.1287/isre.7.1.111 (visited on 06/17/2022).

[207] Erik Stolterman and Mikael Wiberg. “Concept-Driven Interaction Design Research”. en. In: *Human–Computer Interaction* 25.2 (May 2010), pp. 95–118. ISSN: 0737-0024, 1532-7051. DOI: 10.1080/07370020903586696. URL: https://www.tandfonline.com/doi/full/10.1080/07370020903586696 (visited on 11/22/2022).

[208] Rune Stubager. *Folketingsvalget 2015-oprør fra udkanten*. Djøf Forlag, 2017.

[209] Simone Stumpf, Vidya Rajaram, Lida Li, Weng Keen Wong, Margaret Burnett, Thomas Dietterich, Erin Sullivan, and Jonathan Herlocker. “Interacting meaningfully with machine learning systems: Three experiments”. In: *International Journal of Human Computer Studies* 67.8 (2009). Publisher: Elsevier, pp. 639–662. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2009.03.004. URL: http://dx.doi.org/10.1016/j.ijhcs.2009.03.004.

[210] Latanya Sweeney. “Discrimination in online ad delivery”. In: *Queue* 11.3 (2013). Publisher: ACM, p. 10.

[211] Danny Tang. “Empowering novices to understand and use machine learning with personalized image classification models, intuitive analysis tools, and MIT App Inventor”. eng. Accepted: 2019-12-05T18:05:06Z. Thesis. Massachusetts Institute of Technology, 2019. URL: https://dspace.mit.edu/handle/1721.1/123130 (visited on 12/22/2022).

[212] Petros Terzis. “Onward for the Freedom of Others: Marching beyond the AI Ethics”. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. FAT* ’20. event-place: Barcelona, Spain. New York, NY, USA: Association for Computing Machinery, 2020, pp. 220–229. ISBN: 978-1-4503-6936-7. DOI: 10.1145/3351095.3373152. URL: https://doi.org/10.1145/3351095.3373152.

[213] Mike Tissenbaum, Josh Sheldon, Lissa Seop, Clifford H. Lee, and Natalie Lao. “Critical computational empowerment: Engaging youth as shapers of the digital future”. In: *2017 IEEE Global Engineering Education Conference (EDUCON)*. ISSN: 2165-9567. Apr. 2017, pp. 1705–1708. DOI: 10.1109/EDUCON.2017.7943078.

[214] Seth Tisue and Uri Wilensky. “NetLogo: A Simple Environment for Modeling Complexity”. In: *Presented at the International Conference on Complex Systems*. 2004. URL: http://www.ccl.sesp.northwestern.edu/papers/netlogo-iccs2004.pdf.

[215] Peter Tolmie, James Pycock, Tim Diggins, Allan MacLean, and Alain Karsenty. "Unremarkable computing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '02. New York, NY, USA: Association for Computing Machinery, Apr. 2002, pp. 399–406. ISBN: 978-1-58113-453-7. DOI: 10.1145/503376.503448. URL: https://doi.org/10.1145/503376.503448 (visited on 02/17/2022).

[216] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. "Envisioning AI for K-12: What Should Every Child Know about AI?" en. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (July 2019). Number: 01, pp. 9795–9799. ISSN: 2374-3468. DOI: 10.1609/aaai.v33i01.33019795. URL: https://ojs.aaai.org/index.php/AAAI/article/view/5053 (visited on 10/26/2022).

[217] David Touretzky, Christina Gardner-McCune, and Deborah Seehorn. "Machine Learning and the Five Big Ideas in AI". en. In: *International Journal of Artificial Intelligence in Education* (Oct. 2022). ISSN: 1560-4306. DOI: 10.1007/s40593-022-00314-1. URL: https://doi.org/10.1007/s40593-022-00314-1 (visited on 01/06/2023).

[218] Ari Tuhkala, Marie-Louise Wagner, Nick Nielsen, Ole Sejer Iversen, and Tommi Kärkkäinen. "Technology Comprehension". In: *FabLearn Europe'18*. Issue: June. 2018, pp. 72–80. ISBN: 978-1-4503-5371-7. DOI: 10.1145/3213818.3213828.

[219] Laia Turmo Vidal, Hui Zhu, Annika Waern, and Elena Márquez Segura. "The Design Space of Wearables for Sports and Fitness Practices". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–14. ISBN: 978-1-4503-8096-6. DOI: 10.1145/3411764.3445700. URL: https://doi.org/10.1145/3411764.3445700 (visited on 01/10/2023).

[220] Kirsikka Vaajakallio and Tuuli Mattelmäki. "Design games in codesign: as a tool, a mindset and a structure". In: *CoDesign* 10.1 (2014). Publisher: Taylor & Francis, pp. 63–77. DOI: 10.1080/15710882.2014.881886. URL: https://doi.org/10.1080/15710882.2014.881886.

[221] Maarten Van Mechelen, Gökçe Elif Baykal, Christian Dindler, Eva Eriksson, and Ole Sejer Iversen. "18 Years of Ethics in Child-Computer Interaction Research: A Systematic Literature Review". In: *Proceedings of the Interaction Design and Children Conference*. IDC '20. event-place: London, United Kingdom. New York, NY, USA: Association for Computing Machinery, 2020, pp. 161–183. ISBN: 978-1-4503-7981-6. DOI: 10.1145/3392063.3394407. URL: https://doi.org/10.1145/3392063.3394407.

[222] Maarten Van Mechelen, Line Have Musaeus, Ole Sejer Iversen, Christian Dindler, and Arthur Hjorth. "A Systematic Review of Empowerment in Child-Computer Interaction Research". en. In: *Interaction Design and Children*. Athens Greece: ACM, June 2021, pp. 119–130. ISBN: 978-1-4503-8452-0. DOI: 10.1145/3459990.3460701. URL: https://dl.acm.org/doi/10.1145/3459990.3460701 (visited on 09/29/2022).

[223] Maarten Van Mechelen, Marie-Louise Wagner, Gökçe Elif Baykal, Rachel Charlotte Smith, and Ole Sejer Iversen. "Digital Design Literacy in K-9 Education: Experiences from Pioneer Teachers". In: *Interaction Design and Children*. IDC '21. New York, NY, USA: Association for Computing Machinery, June 2021, pp. 32–42. ISBN: 978-1-4503-8452-0. DOI: 10.1145/3459990.3460696. URL: https://doi.org/10.1145/3459990.3460696 (visited on 12/21/2022).

[224] Henriikka Vartiainen, Tapani Toivonen, Ilkka Jormanainen, Juho Kahila, Matti Tedre, and Teemu Valtonen. "Machine learning for middle-schoolers: Children as designers of machine-learning apps". In: *2020 IEEE Frontiers in Education Conference (FIE)*. ISSN: 2377-634X. Oct. 2020, pp. 1–9. DOI: 10.1109/FIE44824.2020.9273981.

[225] Peter-Paul Verbeek. "Materializing morality: Design ethics and technological mediation". In: *Science, Technology, & Human Values* 31.3 (2006). Publisher: Sage Publications Sage CA: Thousand Oaks, CA, pp. 361–380.

[226] Xiaoyu Wan, Xiaofei Zhou, Zaiqiao Ye, Chase K. Mortensen, and Zhen Bai. "SmileyCluster: supporting accessible machine learning in K-12 scientific discovery". In: *Proceedings of the Interaction Design and Children Conference*. IDC '20. New York, NY, USA: Association for Computing Machinery, June 2020, pp. 23–35. ISBN: 978-1-4503-7981-6. DOI: 10.1145/3392063.3394440. URL: https://doi.org/10.1145/3392063.3394440 (visited on 12/22/2022).

[227] David Weintrop, Elham Beheshti, Michael Horn, Kai Orton, Kemi Jona, Laura Trouille, and Uri Wilensky. "Defining Computational Thinking for Mathematics and Science Classrooms". In: *Journal of Science Education and Technology* 25.1 (2016). Publisher: Springer Netherlands, pp. 127–147. ISSN: 15731839. DOI: 10.1007/s10956-015-9581-5.

[228] Mark Weiser. "Creating the invisible interface: (invited talk)". In: *Proceedings of the 7th annual ACM symposium on User interface software and technology*. UIST '94. New York, NY, USA: Association for Computing Machinery, Nov. 1994, p. 1. ISBN: 978-0-89791-657-8. DOI: 10.1145/192426.192428. URL: https://doi.org/10.1145/192426.192428 (visited on 06/14/2022).

[229] Mark Weiser. "The Computer for the 21st Century". In: *SCIENTIFIC AMERICAN* (1991), p. 13.

[230] Frances Westley and Nino Antadze. "When scaling out is not enough: Strategies for system change". en. In: *Social Frontiers: Social Innovation Research Conference*. London, 2013, p. 11.

[231] Randi Williams, Hae Won Park, and Cynthia Breazeal. "A is for Artificial Intelligence: The Impact of Artificial Intelligence Activities on Young Children's Perceptions of Robots". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–11. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300677. URL: https://doi.org/10.1145/3290605.3300677 (visited on 12/22/2022).

[232] Randi Williams, Hae Won Park, Lauren Oh, and Cynthia Breazeal. "Popbots: Designing an artificial intelligence curriculum for early childhood education". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 9729–9736.

[233] Jeannette M Wing. "Computational thinking and thinking about computing". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366.1881 (2008). Publisher: The Royal Society London, pp. 3717–3725.

[234] Jeannette M. Wing. "Computational thinking". In: *Communications of the ACM* 49.3 (Mar. 2006), pp. 33–35. ISSN: 0001-0782. DOI: 10.1145/1118178.1118215. URL: https://doi.org/10.1145/1118178.1118215 (visited on 11/28/2022).

[235] Jeannette M. Wing and Dan Stanzione. "Progress in computational thinking, and expanding the HPC community". In: *Communications of the ACM* 59.7 (June 2016). Publisher: Association for Computing Machinery (ACM), pp. 10–11. DOI: 10.1145/2933410. URL: https://doi.org/10.1145/2933410.

[236] Cathy S.P. Wong. "Views on the adoption and implementation of the SOLO taxonomy". In: *Proceedings of the First International Conference on Enhancing Teaching and Learning through Assessment*. Springer, 2007, pp. 4–15.

[237] Kwok-Fu Wong. "Empowerment as a panacea for poverty - old wine in new bottles? Reflections on the World Bank's conception of power". en. In: *Progress in Development Studies* 3.4 (Oct. 2003). Publisher: SAGE Publications India, pp. 307–322. ISSN: 1464-9934. DOI: 10.1191/1464993403ps067oa. URL: https://doi.org/10.1191/1464993403ps067oa (visited on 11/24/2022).

[238] Qian Yang, Aaron Steinfeld, and John Zimmerman. "Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes". en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–11. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300468. URL: https://dl.acm.org/doi/10.1145/3290605.3300468 (visited on 02/17/2022).

[239] Karen Yeung. "'Hypernudge': Big Data as a mode of regulation by design". In: *Information Communication and Society* 20.1 (2017). ISBN: 1369-118X, 1468-4462, pp. 118–136. ISSN: 14684462. DOI: 10.1080/1369118X.2016.1186713. URL: https://doi.org/10.1080/1369118X.2016.1186713.

[240] Hatice Yıldız Durak. "What would you do without your smartphone? adolescents' social media usage, locus of control, and loneliness as a predictor of nomophobia". eng. In: (Feb. 2018). Accepted: 2019-03-13T10:46:53Z Publisher: Addicta: The Turkish Journal on Addictions. URL: https://acikerisim.bartin.edu.tr/handle/11772/963 (visited on 11/28/2022).

[241] Tal Z Zarsky. "Automated prediction: Perception, law, and policy". In: *Communications of the ACM* 55.9 (2012). Publisher: ACM, pp. 33–35.

[242] John Zimmerman and Jodi Forlizzi. "The Role of Design Artifacts in Design Theory Construction". en. In: *Artifact* 2.1 (Nov. 2008). Number: 1, pp. 41–45. ISSN: 1749-3471. DOI: 10.1080/17493460802276893. URL: https://scholarworks.iu.edu/journals/index.php/artifact/article/view/3915 (visited on 11/21/2022).

[243] John Zimmerman, Aaron Steinfeld, Anthony Tomasic, and Oscar J. Romero. "Recentering Reframing as an RtD Contribution: The Case of Pivoting from Accessible Web Tables to a Conversational Internet". en. In: *CHI Conference on Human Factors in Computing Systems*. New Orleans LA USA: ACM, Apr. 2022, pp. 1–14. ISBN: 978-1-4503-9157-3. DOI: 10.1145/3491102.3517789. URL: https://dl.acm.org/doi/10.1145/3491102.3517789 (visited on 09/29/2022).

[244] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. "An analysis and critique of Research through Design: towards a formalization of a research approach". In: *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. DIS '10. New York, NY, USA: Association for Computing Machinery, Aug. 2010, pp. 310–319. ISBN: 978-1-4503-0103-9. DOI: 10.1145/1858171.1858228. URL: https://doi.org/10.1145/1858171.1858228 (visited on 11/07/2022).

[245] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R. Benjamin Shapiro. "Youth making machine learning models for gesture-controlled interactive media". en. In: *Proceedings of the Interaction Design and Children Conference*. London United Kingdom: ACM, June 2020, pp. 63–74. ISBN: 978-1-4503-7981-6. DOI: 10.1145/3392063.3394438. URL: https://dl.acm.org/doi/10.1145/3392063.3394438 (visited on 01/18/2022).

[246] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K. Kane, and R. Benjamin Shapiro. "Youth Learning Machine Learning through Building Models of Athletic Moves". en. In: *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. Boise ID USA: ACM, June 2019, pp. 121–132. ISBN: 978-1-4503-6690-8. DOI: 10.1145/3311927.3323139. URL: https://dl.acm.org/doi/10.1145/3311927.3323139 (visited on 01/18/2022).

[247] Shoshana Zuboff. *The age of surveillance capitalism: the fight for the future at the new frontier of power*. Profile Books, 2019.