Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics

Yim Register
University of Washington
Seattle, Washington, USA
yreg@uw.edu

Amy J. Ko
University of Washington
Seattle, Washington, USA
ajko@uw.edu

ABSTRACT

Machine learning systems are increasingly a part of everyday life, and often used to make critical and possibly harmful decisions that affect stakeholders of the models. Those affected need enough literacy to advocate for themselves when models make mistakes. To understand how to develop this literacy, this paper investigates three ways to teach ML concepts, using linear regression and gradient descent as an introduction to ML foundations. Those three ways include a basic *Facts condition*, mirroring a presentation or brochure about ML, an *Impersonal condition* which teaches ML using some hypothetical individual's data, and a *Personal condition* which teaches ML on the learner's own data in context. Next, we evaluated the effects on learners' ability to self-advocate against harmful ML models. Learners wrote hypothetical letters against poorly performing ML systems that may affect them in real-world scenarios. This study discovered that having learners learn about ML foundations with their own personal data resulted in learners better grounding their self-advocacy arguments in the mechanisms of machine learning when critiquing models in the world.

KEYWORDS

Machine learning literacy, linear regression, artificial intelligence, data science, data literacy, algorithmic fairness

ACM Reference Format:
Yim Register and Amy J. Ko. 2020. Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In *Proceedings of the 2020 International Computing Education Research Conference (ICER '20), August 10–12, 2020, Virtual Event, New Zealand*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3372782.3406252

1 INTRODUCTION

Machine Learning systems are increasingly becoming a part of everyday contexts, such as medicine, finance, legal decisions, transportation, social media, entertainment and more. While many people think of "Artificial Intelligence" as *The Terminator* [6, 8, 9, 49], or something that will "take over the world" [56], they may not be recognizing that Machine Learning (ML) technology is integrated into almost everything we do on our phones, including Google Maps, Facebook, text prediction, face recognition, photo tagging, friend and media recommendations, spam detection, and information retrieval in search engines.

There are potentially harmful effects of ML systems, beyond just a poor facial recognition in a Snapchat application trying to detect your face. Many ML systems are used to determine whether people can get a loan, buy a house, receive government assistance, be eligible for employment, are likely to have committed a crime, will successfully reintegrate into society after incarceration, and many other impactful decisions about human lives [7, 11, 28, 29, 53].

Stakeholders of ML systems should feel empowered to speak up against models that affect them, yet few people can actually explain how these systems work. Contrary to preconceptions, ML is not just for computing majors, as ML spans topics from astrophysics to zoology [50, 51]. For non-experts to advocate for themselves in ML scenarios, they should be able to reason about whether or not a tumor-detection system is trustworthy, knowing which political news they will see on their newsfeeds and why, or how to interpret the recommendations from a prisoner recidivism predictive model. Widespread ML literacy would be important for jurors, consumers, voters, policymakers, engineers, designers, journalists, and more.

Core features of this literacy include model transparency, understanding the mechanisms, contextualizing data, critical thinking, and leveraging learners' interests and backgrounds [31]. This would allow for agency and more targeted self-advocacy for those affected by ML. This means that someone would be able to articulate the flaws in the design of various ML systems, be able to ask effective questions, and be able to express critiques and solutions for their own interactions with ML. This kind of self-advocacy within the machine learning domain might look like:

- Jurors asking questions about a predictive tool used to argue if an incarcerated individual should be granted parole, and discussing if it was fatally biased to favor past judicial decisions.
- Patients asking a doctor if a computer vision tool was trained on their particular condition, and asking about common mistakes the model has and how it is accounted for.
- Voters having basic knowledge about why an article appeared on their social media, and how newsfeeds can be biased by what they and their friends click on or "Like".
- Loan borrowers asking creditors what features were included in models determining whether they should be approved for a loan, and what data that model was trained on.
- Potential employees questioning a company using an NLP model to filter resumes for hire, pointing out that their name or writing style may influence the hiring decision due to biased training data.

ICER '20, August 10–12, 2020, Virtual Event, New Zealand

ACM ISBN 978-1-4503-7092-9/20/08…$15.00
https://doi.org/10.1145/3372782.3406252

Being able to make these critical judgments about ML systems likely relies on the ability for stakeholders to understand the mechanisms of those systems; the inputs and outputs, the strength of the relationship, the appropriateness of the features used in prediction, the shape of the model being used, and the fit of the model to the data. Each of these skills would provide stakeholders with necessary insight to resist against models making incorrect predictions and allow them to identify alternatives or fatal flaws in those ML systems. This literacy is not at the level of programming or innovating on the systems themselves, but it is more generalizable than simply knowing facts about ML systems [13, 15, 23, 41, 42].

Resources for teaching ML often teach specific programming techniques and code libraries for specific problems that are often irrelevant to learners personally, such as: "predicting price of diamonds by hardness, predicting type of iris by length of its petals, predicting survival rates on the Titanic". The most well-known Machine Learning course is Andrew Ng's Coursera course [36], which is mathematically heavy and also relies on the common datasets in the ML community. Similarly, data science education research includes experience reports of experimental data science courses, or has contributed tools to help learners explore their data [1–4, 25, 26, 30], but without scaffolding and real-world context combined. These also focus on university-level instruction or those with background already, and rely on having weeks to months of material to situate the learner. AI literacy for children often involves physical machines such as robots or voice assistants, and doesn't explicitly focus on tools for resistance against harmful models [12, 16, 25, 27, 47, 57]. Developing ML literacy demands different techniques than the traditional lecture model if it is to reach diverse populations [5, 17, 18, 40]. None of the existing approaches directly teach the ML literacy necessary for stakeholders to make critical judgements about the ML systems impacting their daily lives, and we do not yet know if the current resources generalize to helping learners understand ML systems in real-world personal scenarios.

One way to teach machine learning literacy for self-advocacy is to link the mechanisms of ML to the learner's own prior knowledge and experience [2, 13, 40]. To do this, we could incorporate learners' funds of knowledge: leveraging the learner's already existing knowledge and experience by strategically teaching material revolving around the learner's culture, situated knowledge, and relationships [19, 20, 33], a concept adjacent to Papert's constructionism [38, 39] and Dewey's experiential learning [14]. There are some projects that have incorporated relevant and interesting data into data science education, such as CORGIS (Collection of Really Great and Interesting dataSets) [4]. However, there is greater potential for integrating personal data and experience into teaching data science, especially for situated and justice-oriented projects [32, 52]. Because ML systems are so intricately tied to the data they process, we theorize that integrating personal experience and domain knowledge into the teaching of ML literacy could benefit learner's understanding of the mechanisms at play, and teaching with the learner's *own* data allows them to situate themselves with regards to the ML system. Using personal data could increase learners' attention on the mechanisms of ML by making the mechanisms more personally interesting; this increased attention would lead to 1) better understanding of the mechanisms, 2) better ability to apply the mechanisms of ML to their lives, 3) critiques of ML scenarios that are more explicitly grounded in the mechanisms of ML.

To test if leveraging learners' personal funds of knowledge benefits their ability to self-advocate against potentially harmful ML systems, we designed three forms of instruction and an empirical study to evaluate them. In order to give learners the best chance at learning about the mechanisms of ML, we designed a tutorial using best practices from learning sciences to teach a foundational ML concept: linear regression with gradient descent. We used that tutorial to teach using the learner's own personal data (*Personal condition*) vs. a hypothetical individual's data (*Impersonal condition*). We compare these conditions to a baseline description of facts about ML systems and linear regression without referring to any data (*Facts condition*). We studied the impact of those interventions on learners' ability to self-advocate in real world ML scenarios by asking learners to write a letter to the enforcer of a model that made a wrong prediction. This mimics a reasonable pathway in the world; speaking up for yourself in medical, financial, digital, legal, and institutional scenarios.

2 POSSIBLE WAYS OF TEACHING MACHINE LEARNING LITERACY

It is unclear how to effectively teach stakeholders to make critical judgments of ML systems, but we can go through some possibilities drawn from what we see in current ML education. First, we might try to teach prediction by providing closed-form mathematical equations, which are often used to teach ML at the university level. These are likely not comprehensible by the average person using ML systems because they require a lot of math background. Moreover, even if the learner did understand the equation, they would still rely on further mental simulation to determine effects of the model on different kinds of data. Equation 1 demonstrates how knowledge of sigmas, subscripts, weights, vectors, and more would be required to reason about linear regression from the closed form representation.

$$y = w_0 x_0 + w_1 + x_1 + \ldots + w_m x_m = \sum_{j=0}^{m} w^T x \tag{1}$$

Instead of mathematical explanations, we could teach ML literacy by providing the general facts about ML systems and how they work, similar to a presentation about ML, but this could be insufficient for learners to trace how new or unusual data might be manipulated by the system or how it might play out for them personally. For example, a consultant might tell a client that ML suffers from "garbage in, garbage out." The client may now rightfully distrust models with flawed data, but will still have to further mentally simulate to understand the severity of the consequences, or to offer alternative solutions that work better. Given facts about ML systems, they may be unable to ground those general facts in how it applies to them personally or where such systems are used in the world. They may also not have enough information to reason out alternative ways the model could work or be able to pinpoint what kinds of data cause specific models to fail. All of these skills are useful for successful self-advocacy when critiquing an ML model.

While the above techniques are less amenable to supporting the learner to make critical judgments, we could teach the idea that

ML algorithms are responsive processes that manipulate data by describing the steps of the algorithmic process. This might give the learner more insight into where the model can fail. Consider this explanation of Gradient Descent:

> "Gradient Descent works by starting with random values for each coefficient. The sum of the squared errors are calculated for each pair of input and output values. A learning rate is used as a scale factor and the coefficients are updated in the direction towards minimizing the error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible."

However, describing the algorithmic process without any data does not say anything about how the algorithm would respond to new scenarios and is removed from relevant context. This could result in the learner ignoring crucial data scenarios that would happen in the real world, such as models missing outliers or not accounting for important features for a problem. To use machine learning vocabulary, the learner might "underfit"

Using actual data to trace through a problem could give more insight into how that data gets manipulated and where the system may fail. One promising way to teach about ML systems would be to describe an algorithmic process by actually following a trace of some data, demonstrating how data is manipulated and can affect the outcome (as in prior work that explains programming language semantics [35, 54]). However, typical datasets used are either completely abstract ("Product A, B, C and x, y, z") or irrelevant and unapproachable to the learner; who might lack context or domain expertise about the example dataset. For instance, the common iris dataset includes variables like "sepal length, sepal width, petal width and petal length" to predict the various species of iris (*setosa* or *versicolor*). Without context or domain knowledge, the learner may not have any intuition about what is correct or incorrect in their model, or where the model fails. They may trust the results of a faulty model due to lack of insight into the data. Furthermore, such "toy projects" do not engage with societal impact [24].

One way to ensure that the learner thinks more critically about the ML system they are learning about is to fully immerse them in the data process [55]. By using their own funds of knowledge and their own data, the learner must automatically grapple with the nuances of the algorithmic process from start to finish. We theorize that leveraging personal data is particularly suitable for teaching ML literacy because the outputs of ML systems rely critically on relationships in data, and allowing the learner to draw on their own knowledge of the data domain could contribute to better understanding of how ML mechanisms relate to them in the world [10, 34, 48]. Drawn from Dewey's *Experience and Education*, we theorize that creating an experience that situates the learner in the data domain may allow the learner to more readily construct a basis for understanding how ML is working on that data [14]. Using personal data automatically means using different data for each individual learner. This means that the learner may get the chance to explore "dirty" data, or data that is not compatible with the model they are learning about. This may prompt them to think about the pitfalls of the ML techniques in general, which could strengthen their self-advocacy arguments. Instead of learning about algorithmic bias in theory, this technique allows learners to confront how algorithmic bias may affect their own data. We theorize that it might provide the learner with agency to explore their own biases towards what is "objective" in data science, while also giving them richer insight into possible solutions against algorithmic bias. We know that higher level design decisions are some of the most difficult ML concepts to teach [45, 46], and this work demonstrates that integrating personal data and self-advocacy tasks may prompt learners to engage with those tasks in a natural way.

While this seems promising, it may also be the case that learners "overfit" to their own experiences, and are unable to think of other people or scenarios affected by the models. It could be the case that they hyperfocus on their own data, without considering the average use cases for the model. Given that we want to test the effect of using personal data on learner's engagement with the ML tutorials, we arrived at the following research questions:

- **RQ1: Do learners using personal data pay more attention to the mechanisms of machine learning?** We theorize that using personal data would be more interesting and relevant to the learner; resulting in them paying more attention to the tutorial they were given. In particular, they would pay attention to the actual mechanisms of machine learning; and refer to more of those mechanisms in their critiques and self-advocacy arguments.
- **RQ2: Do learners using personal data have a better ability to apply the mechanism to their life?** We theorize that using personal data would allow the learner to draw on relevant domain knowledge from their own experiences; we explore if learners reference their personal experiences more if they use their own data to learn about linear regression.
- **RQ3: Do learners using personal data ground their self-advocacy arguments in the mechanisms of machine learning?** The ability to self-advocate against potentially harmful machine learning models relies on being able to articulate critiques of the model at hand. Successful self-advocacy relies on articulation, negotiation, domain knowledge, and problem solving skills. We look for evidence of these skills in relation to the machine learning scenarios. We explore how using personal data to learn ML relates to learners' self-advocacy arguments.

3 *LEARNMYDATA* TUTORIAL

In order to test how learning ML on personal data affects the ability to critique ML systems and self-advocate, we needed to create a custom tutorial that took in personal data as the data used to teach the mechanisms of machine learning. We decided to teach univariate linear regression (one predictor variable and one response variable) and gradient descent as a proxy for other introductory machine learning concepts. The *LearnMyData* tutorial improved upon the most popular linear regression tutorial for introductory Machine Learning: Andrew Ng's Coursera course videos. We did this by combining some content from the original Coursera material and by introducing promising design practices from Learning Sciences, including but not limited to: minimal visual design, engaging the learner by asking for feedback along the way, drawing upon knowledge that the learner already has, and designing for

self-paced learning. The learning objectives of the tutorial were to describe univariate linear regression, communicate how machine learning "fits" a model to data in order to predict new data, and how the model must be "trained" on data that may or may not generalize. We used the *LearnMyData* tool for both the personal data instructional design and impersonal data instructional design, with the former including an input table for learners to input their own data, and the latter framed around some hypothetical data. To present ML facts to the learners, they did not see the *LearnMyData* tool, but instead got a printout sheet of similar content (with one hypothetical example about grades) without interactivity.

Everyday situations that involve ML systems in an educational setting include: admissions decisions, promotion decisions, allocation of resources (for the institution or financial aid for the students), or lay-off decisions of instructors. We decided to focus on how modeling is used to predict student performance (a tactic often used to make admissions decisions) [37]. The tutorials centered around an undergraduate college experience: "does your interest level in a class relate to the final grade you receive in the class?" This problem was relatable, while also allowing for all kinds of things to happen in the actual data (it is not necessarily a linear or positive relationship). For the *Impersonal condition*, participants reasoned about a hypothetical student who had increasing final grades with their Interest Level in that course. For the *Personal condition*, participants actually input their *own* grades and interest for their last 5 courses at the university (See Figure 1.1). The *Facts condition* prompted the learner to consider the scenario abstractly. All three tutorials covered several mechanisms of machine learning: scatter plots, slope, intercept, formal notation, linear modeling, residuals, mean squared error, minimizing error, gradient descent, generalization on new data, and additional features that might affect the model.

For the *Personal condition*, participants inputted their own data into a table measuring their Interest Level (1-7) and Final Grade (0-4.0) for 5 university courses they had already completed. This means that the learner would be exposed to linear regression through an arbitrary relationship (the relationship could be negative, positive, moderate, weak, strong, random, etc.) It is crucial to note that linear regression should not be done on non-normal, ordinal data, though this happens in practice often. Most datasets given by learners were moderately strong and positive, meaning that self-reported Interest Level did seem to correlate positively with Final Grade. Under the circumstances that the inputted data made linear regression impossible (only one point, all zeroes, etc), the learner explored why linear regression did *not* work on their data.

Figure 1.2 shows the screen where learners try to guess the best fitting line before it was revealed, and could then compare their line of best fit to the actual best fitting line. This way, the learner saw why we might need an actual algorithm to get the true best line, rather than just approximating the relationship. Two screens in the tutorial also demonstrated the residual error for the learner's chosen line versus the residual error on the true best fit line (Figure 1.3, showing more in-depth how close they got to correctly modeling the relationship. After guessing their own line and then seeing the true best fit line, the tutorial asked participants how they thought the true line gets calculated. They next learned about "parameters" of linear regression (slope and intercept) in relation to what they might already have familiarity with: $y = mx + b$ (See Figure 1.4). Next the tutorial introduced the bare minimum "baseline model": brute force guess-and-check the parameters of the model (See Figure 1.5). The animation demonstrated randomly guessing parameters of slope and intercept (bounded by the minimum and maximum of the axes for demonstration purposes, and with a .5 chance of being positive or negative slopes). Each line appeared half a second apart on the scatterplot in an animation. In contrast with the Guess-and-Check model, the next screen demonstrated a Gradient Descent animation. This animation showed the line of best fit updating until it converged, with the line moving into a stable, unmoving position. This was included to demonstrate the process by which gradient descent "learns" the better and better fit line based on partial derivatives (See Figure 1.6).

Figure 1.7 shows another interactive element that prompted learners to use their linear regression line to predict what grade they might get in a *new* class (introducing the concept of generalization). Participants could input an Interest Level for a course they were currently taking, and see what their linear regression line would predict their grade to be. If they felt their interest was currently a 5 in a class they were taking, the linear regression line might say they would get a 3.2 in the class. The new point appeared in red on their linear regression line (see Figure 1.8). Together, all of these elements contribute a novel and interactive instructional design for teaching linear regression. We used this instructional design to test the effect of using personal data on the ability to self-advocate against potentially harmful ML models. The *Impersonal condition* also uses this design, but on hypothetical data as opposed to the learner's own inputted data.

Learners in the *Facts condition* learn about linear regression in list form, part of which is shown in Figure 2. It could be the case that this is enough to result in well-formed self-advocacy arguments, but given what we know from learning sciences it is unlikely. It might be more fitting to use the *LearnMyData* tutorial and its many interactive elements to teach linear regression and ML concepts.

4 CRITIQUE INSTRUMENT

Given that our vision of machine learning literacy involves stakeholders participating in real-world scenarios, we needed a way of measuring learners' ability to critique ML models and self-advocate following the tutorial they did. It wouldn't be as appropriate to use a test of linear regression knowledge, because recalling such knowledge isn't what stakeholders would be doing in the real world. Instead, they might be pointing out flaws in the models in corporate meetings, to fellow jurors, or with doctors. Next, they might be articulating those concerns in letters to some enforcer of the model, advocating for themselves. We chose to mimic these pathways in what we asked learners to do. They saw two ML **Scenarios**, listed **Critiques** of the models, and then wrote **Letters** arguing to an enforcer of a model that made a mistake. The ML literacy we are interested in is about helping stakeholders of potentially harmful models participate in the world. Research about self-advocacy pathways and disability self-advocacy suggest that these letters could reasonably measure that skill [21, 22, 43, 44].

Figure 3 and 4 show the text for the machine learning **Scenarios** that all learners saw when prompted to write critiques and self-advocacy letters. After reading each scenario, we asked learners

Figure 1: Selected screens from *LearnMyData* tutorial, demonstrating some of the interactive elements of the tool. Each screen is accompanied by a paragraph of text with instructions (not shown). The Collect Data (1) screen is unique to the *Personal condition*, where the learner inputs their own data. The Draw a Line (2) screen gives the learner a chance to try to draw their own best fitting line. Later on they see the residual values on the true best fitting line (3), and reason about the parameters of the model in terms of $y = mx + b$ (4). They watch a poor way of determining those parameters, which would be to Guess and Check(5) until you find the best fit. Screens (7) and (8) show the screens that the learner sees when they reason about generalization of the model. They input a new datapoint and see what the linear regression line would predict. The entire *LearnMyData* tool contains 22 screens.

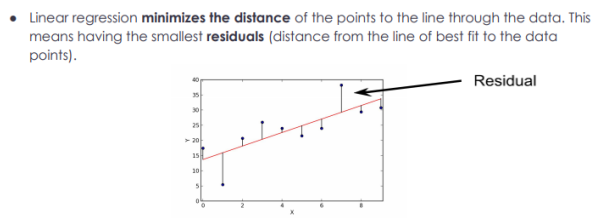

Figure 2: Part of the *Facts condition* printout

to generate as many **Critiques** of the model as they could. We defined critique as: "providing some criticism of the model, about what information it uses and how it might be used in the world; identifying potential problems with what information it takes into account, or how it uses that information to make predictions." Then we asked learners to write a letter to the enforcer of a model that made a wrong prediction for two different scenarios. Figure 3 shows the Interest-to-Grades scenario, and Figure 4 shows the Financial Aid scenario. We chose self-advocacy letters to mimic a reasonable pathway in the world; speaking up for yourself in medical, financial, digital, legal, and institutional scenarios.

> "The instructor of a college course uses a model to identify which students might need extra support and help throughout the quarter. The instructor uses the linear model that you saw in the tutorial. At the beginning of the course, the instructor collects everyone's Interest Level, and makes a prediction for their Grade, based on last year's Interest-to-Grade data. If the model predicts a grade lower than an 75, the instructor will intervene and offer extra help. So, if a student who rates their interest at a 2 tends to score below a 75, the instructor will intervene with a new student who rates their interest at a 2. List as many critiques of this model as you can. Try to use what you learned in the tutorial to make your case. Next, write a convincing argument of how you might advocate for yourself as a student in this scenario (someone being affected by the model). Imagine you are making your case to the instructor, or someone else enforcing the results of the model. "

Figure 3: The Interest-to-Grades Scenario

5 METHOD

We hypothesized that learners in the *Personal condition* would 1) pay attention to more machine learning mechanisms due to increased engagement and interest in the data, 2) would refer to their own personal experiences more, demonstrating that they were linking ML mechanisms to their personal lives, and 3) would ground their self-advocacy arguments in the mechanisms of machine learning more than the other conditions. In order to test these hypotheses, we designed a between-subjects experiment to reveal the differences between using Personal data, Impersonal data, or no data at all (the *Facts condition*). We used the three instructional designs defined in Section 3 as the three different interventions that learners saw.

> "The financial aid office has to make tons of decisions in order to give out aid. Usually, they offer an amount and won't change it unless a family "appeals" the process because it is not enough. They try to predict how much aid they should give as accurately as they can, so that they offer an amount that a family won't appeal. They use a model that uses the number of siblings that a student has to predict how much money their family will need. They use last year's data, looking at families who were "happy" (did not appeal) with the offer the office gave. So, if families with 3 children tend to need $20,000 in aid, that's what the office will budget for a new family with 3 children. Next, write a convincing argument of how you might advocate for yourself as a student or family in this scenario (someone being affected by the model). Imagine you are making your case to the financial aid office, or someone else enforcing the results of the model."

Figure 4: The Financial Aid Scenario

Following the tutorial, learners each completed a self-advocacy task, where they wrote letters about scenarios in which hypothetical models had made a harmful wrong prediction.

5.1 Participants

Our inclusion criteria were university students who had interest in learning about ML, but did not have any experience with learning it. We gave information about the study by word-of-mouth and recruitment flyers to different university lecture courses across a range of disciplines, including information science, chemistry, biology, archaeology, design, economics, and some language studies. When students asked to participate, we screened for previous data science or machine learning training and they were not allowed to participate if they had ever had any data science (collected by a screening survey before scheduling for the actual study). Even students in relevant fields like informatics or economics were new to their majors and did not have any experience with regression or data science. Fifty-one participants took part in this study (*Personal condition = 20, Impersonal condition = 17, Facts condition= 14*). Different numbers of participants was due to scheduling conflicts, and we had already reached saturation of content in the *Facts condition*. Student majors were randomly assigned among conditions, with the most students majoring in Information Technology (14) or pre-major (15). Others included language, business, psychology, health, literature, construction management, and one economics major. 37 had taken introductory Java which does not include data or statistics in any way. We do not report gender because it is irrelevant to these findings, but the first author (who is a nonbinary trans PhD student) ensured inclusive gender practices in both recruitment, methods, and the workshops. Participants knew that the experimenter was passionate about education, with a background in ML. They did not know the goals of the study or that there were other conditions.

5.2 Procedure

Learners participated in the experiment in a workshop setting led by the first author, similar to a tutoring session or study group, with participants randomly assigned to condition. Between 6-12 people were in one workshop at a time, all working on the same condition. Learners could only ask clarifying logistics questions as opposed to conceptual ones. We encouraged breaks throughout the hour long workshop. Learners were compensated with a $15 gift card for their time. Participants learned linear regression with their randomly assigned tutorial (*Learn My Data* tool for *Personal* and *Impersonal conditions*, and a fact sheet printout for the *Facts condition*). They used their own laptops for the *Personal* and *Impersonal conditions*, for familiarity. Learners in the *Personal condition* entered data in the table in Figure 1 1. Following whichever tutorial they did, all participants filled out the same critique instrument, which also included prompts for the self-advocacy arguments. The critique instrument used is described in Section 4. Participants filled them out on paper with pen or pencil, and the paper copies were then stored securely. No documents contained any identifying information.

5.3 Analysis

To answer our research questions, we needed to determine if there were meaningful differences between the instructional conditions. If the *Personal condition* resulted in the most attention to ML mechanisms (RQ1), personally relevant critiques (RQ2), and arguments grounded in those mechanisms (RQ3), this would be evidence that using personal data provided these benefits over the other instructional designs. Because there are no prior theories on how to analyze machine learning critiques, we needed an inductive coding scheme derived from the data. What counts as signals of paying attention (RQ1)? What counts as "personally relevant" (RQ2)? And what is evidence for grounding an argument in the mechanisms of machine learning (RQ3)?

The two authors collaboratively coded the data, inducing a range of themes without knowledge of each participant's condition assignment. The authors anonymized the data and separately produced a set of inductive codes that related to each research question. They did this by reading each document and manually identifying indicators of paying attention to the mechanisms of machine learning, applying that knowledge to their personal life, and using those mechanisms to ground the self-advocacy arguments. The authors tried to define those instances with a label, and generated a possible codebook for the data. The authors met to discuss each candidate code and definition, synthesizing into a single code book, and then separately applied the code book to all content. They identified disagreements in the codes and resolved them, sharpening definitions for the code labels where necessary. The largest disagreement was over the concept of "Construct Validity" due to a misunderstanding. Authors resolved disagreements by sharpening definitions of what each code referred to in the data. Because all of the disagreements were resolved, there was no need for an inter-rater reliability measure.

For each research question, there was a set of codes that would indicate evidence of the learner using those skills. Our coding process generated 6 codes for RQ1, shown in Table 1. All of the coding scheme consisted of binary variables that indicated the presence or absence of some idea in learners' writing. If they mentioned any mechanisms of ML shown in Table 1, we marked it down as Paying Attention. Each of these variables could also be present as

Code	Criteria
Construct Validity	criticizing one or more of the variables in the proposed model for not accurately representing the concept the variable is trying to operationalize. If it was a construct validity critique, it is likely that the learner believed the concept itself exists as a phenomena in the world, but that the way it was captured in the proposed model was wrong.
Additional Features	True if the writing contained a critique that points out additional features/variables/factors that could influence this phenomena. e.g. "The model should take x into account", "The model doesn't take x into account".
Confounds	True if the writing contained a critique that points out that other factors are influencing the variables being measured and affecting the phenomenon the model is trying to predict, in a causal way. It is different from saying that the model should include more factors; they are statements addressing missing logic/factors that are plausibly the true cause of the model's result.
Outliers	True if the writing contained a critique that either includes the explicit term "outlier" or references an edge case or counter example with regards to the model.
Causality	True if the writing contained a critique that either explicitly mention "cause" or point out that the independent variable does not influence the dependent variable.
Model Performance	True if the writing contained a critique that point out problems with the model itself (as opposed to the measurement or operationalization of the variables). They mention accuracy, fit, spread, shape and/or strength of the relationship.

Table 1: Mechanisms of machine learning

Code	Criteria
Personal Detail (not model)	True if the learner provides a comment about themselves that is outside the context of the model; usually additional context about the scenario. Example: *"My father is an immigrant"* or *"I am really invested in archaeology now"*, *"I want to be a math teacher"*
Consequence (model)	True if the learner identifies something that happened to them because of the model, with explicit mention of the model, such as *"because of the model I failed the class"* or *"the model made a wrong prediction, then I couldn't pay for school"*
Consequence (not model)	True if the learner includes some additional, richer context on what happened when the model made a wrong prediction, but does not mention the model explicitly. outlines something good or bad that could result from this wrong prediction scenario. e.g. *"my friend rated their interest low and felt patronized by you".*

Table 2: Evidence of mentioning personal life

Code	Criteria
Model is good	True if the writing contained some kind of positive sentiment towards the model. This may include any indication that the model is representing a phenomena in the world ' *'it may be the case that interest correlates with grade..."*, or that the idea of a model for helping students is a good idea despite this particular model being ineffectual.
Use this model instead	True if the learner provides an alternative model or alternative processes to follow to create the model. This is above and beyond suggesting additional factors/features. e.g. *"instead of interest, use motivation" or "you could try taking interest measures multiple times before the midterm."*
Model could be gamed	True if the writing contains a critique that identifies a pathway for people to manipulate the model for their own gain, such as intentionally lying to trick the system into giving them some benefit. e.g. *"students might give low interest on purpose just to get extra help".*

Table 3: Additional mechanisms of machine learning seen in self-advocacy letters

personally applying to the learner's life (RQ2) or as part of the self-advocacy arguments (RQ3). Evidence of applying ML mechanisms to their own lives (RQ2) would include any of the above variables, but in reference to the learner themselves, such as "*I* am an outlier because...", or "*my* interest changed over time". Our coding process also generated 3 additional codes for RQ2, shown in Table 2.

Finally, our coding process generated indications of the learner grounding their self-advocacy arguments in the ML mechanisms. Recall that the critique instrument included both critiques of the models and self-advocacy letters. Similar to RQ2, if any of the concepts from RQ1 were present as part of the letters, they are recorded as evidence of learners grounding their self-advocacy arguments in the mechanisms of ML (RQ3). In addition to the listed mechanisms of ML from Table 1, we saw 3 other indicators that the learners were thinking critically, and writing about, those mechanisms in their arguments (see Table 3).

After deriving these codes and definitions, we went through the anonymized-to-condition data and marked where each of the codes occurred for each participant. A participant's data was a series of binary variables, indicating whether or not a specific code was present in their writing (either in their list of critiques or their self-advocacy letters). A sum of the binary variables represents a total count of how many codes were present in their writing. A higher count would mean that the participant mentioned more of the mechanisms we identified.

6 RESULTS

We theorized that personal data would increase learners' attention on the mechanisms of machine learning by making the mechanisms more personally interesting; this increased attention would lead to 1) better understanding of the mechanisms, 2) better ability to apply the mechanisms of machine learning to their lives, 3) critiques of ML applications that are more explicitly grounded in the mechanisms of machine learning.

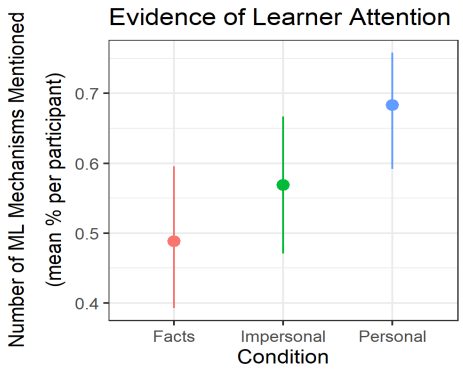

Figure 5: Evidence of paying attention across the three conditions. Proportion of indicators present per participant.

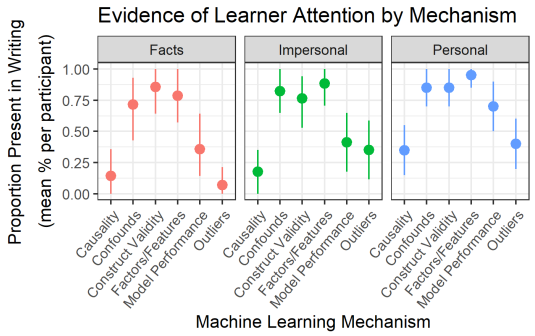

Figure 6: Evidence of paying attention across the three conditions, broken up by mechanism

6.1 RQ1: Did learners using personal data pay more attention to the mechanisms of machine learning?

We hypothesized that personal data would increase learners' attention on the mechanisms of ML by making the mechanisms personally interesting. This increased attention would lead to better understanding of the mechanisms and could be seen through incorporating more of the mechanisms in their writing.

We considered all of the binary variables described in Table 1 and computed the proportion present in each participant's response (therefore using mean as opposed to median to represent proportion). Figure 5 shows these proportions by condition. To analyze whether the visually apparent differences in Figure 5 were statistically significant, we compared counts of the number of things in Table 1 that were present. Because the variables are a count, but the data was ordinal, we used a Kruskal-Wallis test. For this data, the test evaluates expected vs. actual counts if the conditions were equal. A Kruskal-Wallis test revealed a difference in attention by condition ($\chi^2 = 8.01, df = 2, p = 0.02$). Warranted post-hoc Wilcoxon-Mann-Whitney tests reveal that this difference was between *Facts* and *Personal conditions* ($W = 4056, p = 0.005$), though there was a marginally significant difference between *Impersonal* and *Personal conditions* ($W = 5418, p = 0.08$).

Figure 6 shows the mean number of participants who mentioned a specific machine learning mechanism across the different conditions. We can see that 7% of participants in the *Facts condition* mentioned Outliers, whereas 40% of participants in the *Personal condition* mentioned Outliers. We also see that 95% of participants in the *Personal condition* mentioned Factors/Features in their writing. The overall omnibus difference seen from the Kruskal-Wallis test can be attributed to Model Performance, which significantly changed depending on the condition. Participants in the *Personal Condition* paid attention to more of the machine learning mechanisms, with particular attention to Model Performance more than the other conditions. In the *Personal condition*, 70% of participants mentioned Model Performance, as opposed to 41% in the *Impersonal condition*, and 35% in the *Facts condition*.

To help illustrate how participants wrote about Model Performance, consider these quotes from their letters. Note that any mention of accuracy, fit, or shape was counted as paying attention to Model Performance, to avoid favoring students who simply tend to write more.

> *"I happen to know that your data uses a regression model, and I feel that it is flawed. I'm sure if you check the data, you may have calculated a line from the set, but it was probably really spread out"* - P46 (Personal)

> *"If you remove a point, the slope and y-intercept change dramatically. This does not mean that the line of best fit algorithm is wrong, but it does imply that the prediction can be flawed with data values that can significantly skew the findings of the model"* - P36 (Impersonal)

> *"I don't think your model for predicting exam grades is accurate."* - P28 (Facts)

6.2 RQ2: Did learners using personal data have a better ability to apply the mechanism to their life?

We theorized that the personalization of the instruction would lead differences in attention toward model bias. If that was the case, we should see some evidence of learners connecting the material to personal experiences. Given that we saw some trend of more learners in the *Personal condition* attention to a larger range of machine learning mechanisms, now we investigate if those mechanisms were explicitly presented as personal experiences. Additionally, many participants offered personal information unrelated to the model or the mechanisms of machine learning.

To demonstrate the coding process and some relevant examples from the data, here are some examples of how Additional Factors/Features was labeled as personal:

> *"**Personally**, I have 3 younger siblings which all have various expenses. Yet these expenses are not accounted for."* - P1 (Impersonal)

> *"**My** brother is about to come to college, and I'll definitely not be happy with our financial aid because we don't get enough money for college. So my suggestion is, number of siblings is a rather narrow factor, so the office should definitely look into other reasons behind our applications than just to shut our mouths"* - P24 (Personal)

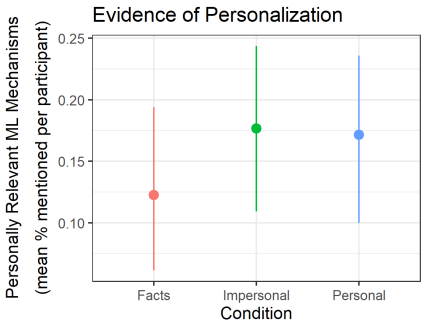

Figure 7: Evidence of personalization of machine learning mechanisms across the three conditions

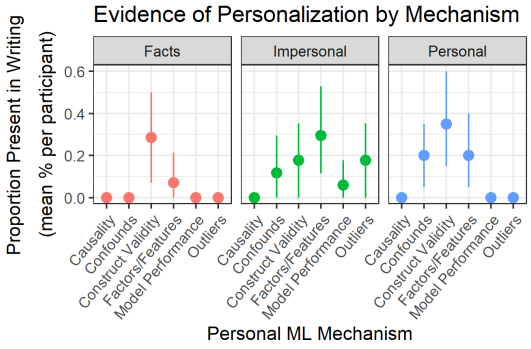

Figure 8: Evidence of personalization of machine learning mechanisms across the three conditions, broken up by concept. Proportion of indicators present per participant

> *"While **my** family does have 3 siblings, I insist you take into account details about our family's finances. Please re-assess our application to take into account household income."* -P41 (Facts)

We considered all of the binary variables described in Table 2 and computed the proportion present in each participant's response. Figure 7 shows these proportions by condition. Figure 8 shows the proportion of participants from each condition who mentioned a given ML mechanism in a personal way. A Kruskal Wallis test revealed no difference between conditions ($\chi^2 = 1.40, df = 2, p = 0.496$).

6.3 RQ3: Did learners using personal data ground their self-advocacy arguments in the mechanisms of machine learning?

We see that learners in the *Personal condition* paid attention more to the mechanisms of machine learning, but did not allude to their personal experiences any more than the other conditions. Next, we look to see if personalized instruction led to any differences in learners' ability to self-advocate by grounding their arguments in the mechanisms of machine learning. Figure 10 shows the proportion of participants in each condition who exhibited those mechanisms in their self-advocacy arguments (either scenario). A Kruskal Wallis test revealed a significant difference between conditions ($\chi^2 = 17.98, df = 2, p = 0.0001$), and Cramer's V revealing a

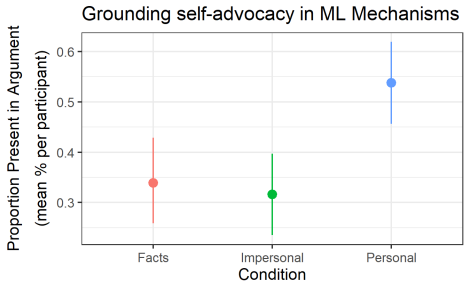

Figure 9: Aggregated across mechanism. Evidence of grounding self-advocacy in mechanisms of machine learning

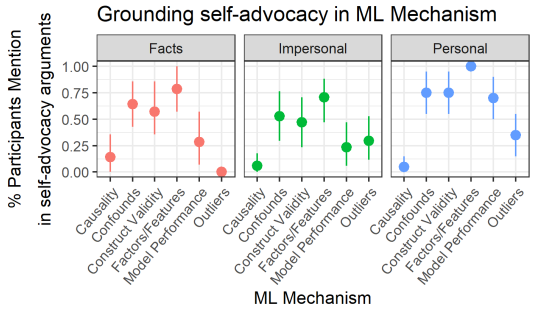

Figure 10: Evidence of referring to the mechanisms of machine learning models by concept. We theorize that Facts and Impersonal produce similar results because participants in the *Impersonal condition* used a lot of space in their arguments to imagine personal implications instead of discussing mechanisms of ML.

medium effect size ($V = .15$), with the *Personal condition* exhibiting the most grounding in the mechanisms of machine learning in their self-advocacy arguments (See Figure 9).

It could be the case that the difference is entirely driven by the fact that every learner in the *Personal condition* included some mention of additional Features/Factors in their arguments ($proportion = 1.0$), shown in Figure 10. However, even after removing this concept altogether, the relationship holds ($\chi^2 = 15.65, df = 2, p = 0.0004$). This suggests that the difference between conditions was driven by more than one concept. We note that the self-advocacy arguments in the *Personal condition* had the highest proportion of mentions for Factors/Features, Confounds, Construct Validity and Model Performance, each with higher proportions than the other conditions. To get a picture of the type of letters that learners wrote, we present a letter with a high number of codes from each condition.

> *"Firstly, the idea as a whole violates the right to equal opportunity. While extra resources are typically available to students, they are given to all students, not just some. Furthermore, basing the model on only 5 classes means that it is based aroud recent trends and not an academic career as a whole. This could result in an unreliable model that has potentially been influenced by outliers. After all, residuals must be miimized for a line of best fit, and this factors outliers which strongly skew the residuals. With a model that seems unreliable like this, it could cause issues via wrong predictions. If I were a student who was given extra assstance in*

a class I already did well in how would I know that it was done on my own merit? The system would be problematic at best and potentially hurt a lot of students." - P25 (Personal) number of codes:6

"Prediction based on previous students but generalization takes me not as a student but rather a potential outlier. I may need to work harder but that interest level can vary among different times. You can not keep track of that. The gradient is just a number, but that does not cover reality." -P52 (Impersonal) number of codes:4

"I'm wondering how much does the Interest level really reflect one's interest or even ability. If people have different rationale in mind, the information of interest level will not make sense. And if you use interest level to predict one's grade, it might have wrong results. Futhermore, interest level doesn't necessarily mean one's ability, but grades reflect more on one's ability, so the relationship between interest level and grades doesn't make sense to me." - P44 (Facts) number of codes:5

7 LIMITATIONS

The biggest threat to construct validity in this study is that if the learners didn't write something down, we couldn't measure it. We chose to give open-ended prompts as opposed to targeted questions about specific ML mechanisms in order to preserve some ecological validity. But this favored those who wrote more, though this should have been distributed equally across conditions due to random assignment. Some threats to internal validity are that the *LearnMyData* tutorial had a few bugs during the workshops, resulting in some participants needing to refresh and start over, and that participants filled out the critique instrument by hand, including some who reported hand cramps. This may have encouraged some learners to write less, or to be more frustrated with the task. However, everyone finished and were accommodating when there were bugs. Threats to external validity include only studying linear regression and using university students as opposed to any other population of stakeholders in real ML scenarios.

8 DISCUSSION

We theorized that using the learner's own personal data would help stakeholders pay attention to the mechanisms of machine learning, relate the mechanisms more to their own lives, and ground self-advocacy arguments in these mechanisms. Our results do show evidence that one way to help develop a self-advocacy skill in the domain of machine learning is to teach the learner on their own personal data. We see some evidence of better attention to the material and arguments that are more grounded in the ML mechanisms. We predicted that these benefits arose directly from being able to better relate the material to yourself, but we do not have evidence to support that. Instead, the ability to self-advocate was linked to learning on personal data, but also likely linked to attention.

There are several ways to interpret these links. It could be that learners in the *Personal condition* did link the mechanisms to their own personal experiences, but did not write about them. It could also be that because the *Personal condition* learned on their own data, they wanted to provide something more novel and generalizable in their critiques. This idea is supported by what happened in the *Impersonal condition*; where relating the ideas to participants' selves was something they hadn't yet done and therefore more warranted to talk about in the critiques. We predicted that learning on personal data might lead to "overfitting" to learners' experiences; but we actually saw more evidence of this in the *Impersonal condition*. This suggests that there could be a natural pathway when confronted with a relevant scenario to try to link it to learners' selves and draw upon their personal experiences. This may have led learners to focus more on their personal involvement as data, without the scaffolding to actually visualize their own data. The *Personal condition* may have allowed learners to explore their personal involvement and then move on to a more generalizable critique, where the learner considered several different cases and how the model might handle them.

Perhaps most surprising were the self-advocacy arguments from those in the *Impersonal condition*. We saw evidence that those learners paid more attention than those in the *Facts condition*, and that they included personal details in their writing, but that they grounded their arguments in the same proportion of machine learning mechanisms as those in the *Facts condition*. It is almost as if presenting a relevant scenario without including the learner's own life distracted the learner to write about personal details in order to relate to the material, as opposed to writing about the mechanisms of machine learning. We do see evidence of attention and repetition of what they learned, meaning that they probably did learn more and pay attention more than those in the *Facts condition*, but that extra attention was unrelated to their self-advocacy arguments being grounded in what they learned.

These results suggest that machine learning education, to the extent that it seeks to develop literacy relevant to people's lives, needs to have learners' voices and lives represented in their learning. Our work presents a possible pathway for people to learn about their own experiences online and in the world, while relating the mechanisms of the systems to their own data. We contributed both a novel tool and a methodology for measuring self-advocacy arguments against potentially harmful machine learning models. We do not claim that this is the only way to present a successful argument, but that using personal data does result in these differences in self-advocacy articulation after learning about linear regression. Future work could explore the relationship between personal data and self-advocacy for different ML algorithms in natural contexts. This includes how social media users process reliability of information online after learning about clustering algorithms, or how people critique facial recognition bias after exploring their own images. Two in-progress projects by the first author explore how users learn about collaborative filtering or how NLP works by running their own Facebook posts through simple models. In practice, teachers could integrate the self-advocacy tasks into their lessons of any subject, demonstrating to students that they can critique relevant models in the world, and develop their ability to articulate what is wrong with the models they are learning about. Machine learning education is not just about the computations, but also the ability to critically engage with these systems and stand up for ourselves.

9 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1735123, 1539179, 1703304, and 1836813, and unrestricted gifts from Microsoft, Adobe, and Google.

REFERENCES

[1] Ruth E. Anderson, Michael D. Ernst, Robert Ordóñez, Paul Pham, and Ben Tribelhorn. 2015. A Data Programming CS1 Course. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15)*. ACM, New York, NY, USA, 150–155. https://doi.org/10.1145/2676723.2677309

[2] Mad Price Ball. [n. d.]. Open Humans. https://www.openhumans.org/.

[3] Austin Cory Bart, Dennis Kafura, Clifford A Shaffer, and Eli Tilevich. 2018. Reconciling the Promise and Pragmatics of Enhancing Computing Pedagogy with Data Science. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. ACM, 1029–1034.

[4] Austin Cory Bart, Ryan Whitcomb, Dennis Kafura, Clifford A Shaffer, and Eli Tilevich. 2017. Computing with corgis: Diverse, real-world datasets for introductory computing. *ACM Inroads* 8, 2 (2017), 66–72.

[5] Rahul Bhargava and Catherine D'Ignazio. [n. d.]. Designing tools and activities for data literacy learners.

[6] Nick Bostrom. [n. d.]. The Vulnerable World Hypothesis. ([n. d.]).

[7] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 41.

[8] Ed Burns. 2018. SearchEnterpriseAI Current state of AI is poorly understood by the public. https://searchenterpriseai.techtarget.com/opinion/Current-state-of-AI-is-poorly-understood-by-the-public.

[9] Howard Chen. 2018. MSandE 238 Blog Public perception of artificial intelligence. https://mse238blog.stanford.edu/2018/07/howachen/public-perception-of-artificial-intelligence/.

[10] National Research Council et al. 2000. *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.

[11] Imad Dabbura. 2018. Predicting Loan Repayment. https://towardsdatascience.com/predicting-loan-repayment-5df4e0023e92.

[12] Sayamindu Dasgupta and Benjamin Mako Hill. 2017. Scratch community blocks: Supporting children as data scientists. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3620–3631.

[13] Erica Deahl. 2014. Better the data you know: Developing youth data literacy in schools and informal learning environments. *Available at SSRN 2445621* (2014).

[14] John Dewey. 1986. Experience and education. In *The Educational Forum*, Vol. 50. Taylor & Francis, 241–252.

[15] Catherine D'Ignazio and Rahul Bhargava. 2016. DataBasic: Design principles, tools and activities for data literacy learners. *The Journal of Community Informatics* 12, 3 (2016).

[16] Stefania Druga, Sarah T Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*. ACM, 104–111.

[17] Catherine D'Ignazio and Rahul Bhargava. [n. d.]. Approaches to building big data literacy.

[18] Catherine D'Ignazio and Lauren F Klein. 2016. Feminist data visualization. In *Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore. IEEE.*

[19] Norma González, Luis C Moll, and Cathy Amanti. 2006. *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Routledge.

[20] Norma Gonzalez, Luis C Moll, Martha Floyd Tenery, Anna Rivera, Patricia Rendon, Raquel Gonzales, and Cathy Amanti. 1995. Funds of knowledge for teaching in Latino households. *Urban Education* 29, 4 (1995), 443–470.

[21] Dan Goodley. 2000. *Self-advocacy in the lives of people with learning difficulties: The politics of resilience*. Open University Press Buckingham.

[22] Dan Goodley. 2005. Empowerment, self-advocacy and resilience. *Journal of Intellectual Disabilities* 9, 4 (2005), 333–343.

[23] Samantha Hautea, Sayamindu Dasgupta, and Benjamin Mako Hill. 2017. Youth perspectives on critical data literacies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 919–930.

[24] Birte Heinemann, Simone Opel, Lea Budde, Carsten Schulte, Daniel Frischemeier, Rolf Biehler, Susanne Podworny, and Thomas Wassong. 2018. Drafting a data science curriculum for secondary schools. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research*. 1–5.

[25] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. 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*. ACM, 415.

[26] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M Drucker. 2019. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 579.

[27] C. D. Kidd and C. Breazeal. 2004. Effect of a robot on user perceptions. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, Vol. 4. 3559–3564 vol.4. https://doi.org/10.1109/IROS.2004.1389967

[28] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. Human decisions and machine predictions. *The quarterly journal of economics* 133, 1 (2017), 237–293.

[29] Konstantina Kourou, Themis P Exarchos, Konstantinos P Exarchos, Michalis V Karamouzis, and Dimitrios I Fotiadis. 2015. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal* 13 (2015), 8–17.

[30] Sean Kross and Philip J Guo. 2019. Practitioners Teaching Data Science in Industry and Academia: Expectations, Workflows, and Challenges. (2019).

[31] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[32] Camillia Matuk, Anna Amato, Kayla DesPortes, Marian Tes, Veena Vasudevan, Susan Yoon, Jooeun Shim, Amanda Cottone, Kate Miller, Blanca Himes, et al. [n. d.]. Data Literacy for Social Justice. ([n. d.]).

[33] Luis C Moll, Cathy Amanti, Deborah Neff, and Norma Gonzalez. 1992. Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into practice* 31, 2 (1992), 132–141.

[34] Engineering National Academies of Sciences, Medicine, et al. 2018. *How people learn II: Learners, contexts, and cultures*. National Academies Press.

[35] Greg L Nelson, Benjamin Xie, and Amy J Ko. 2017. Comprehension first: evaluating a novel pedagogy and tutoring system for program tracing in CS1. In *Proceedings of the 2017 ACM Conference on International Computing Education Research*. ACM, 2–11.

[36] Andrew Ng. 2011. Machine Learning Coursera Course. https://www.coursera.org/learn/machine-learning.

[37] Cathy O'Neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

[38] Seymour Papert. 1980. *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

[39] Seymour Papert. 1999. Papert on piaget. *Time magazine, pág* 105 (1999).

[40] Evan M Peck, Sofia E Ayuso, and Omar El-Etr. 2019. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 244.

[41] Javier Calzada Prado and Miguel Ángel Marzal. 2013. Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri* 63, 2 (2013), 123–134.

[42] Milo Schield. 2004. Information literacy, statistical literacy and data literacy. In *Iassist Quarterly (IQ)*. Citeseer.

[43] Anneliese A Singh, Sarah E Meng, and Anthony W Hansen. 2014. "I am my own gender": Resilience strategies of trans youth. *Journal of counseling & development* 92, 2 (2014), 208–218.

[44] Jenny Slater. 2012. Self-advocacy and socially just pedagogy. *Disability Studies Quarterly* 32, 1 (2012).

[45] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R Cooperstock. 2019. Can You Teach Me To Machine Learn?. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 948–954.

[46] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R Cooperstock. 2019. What is hard about teaching machine learning to non-majors? Insights from classifying instructors' learning goals. *ACM Transactions on Computing Education (TOCE)* 19, 4 (2019), 1–16.

[47] Dave Touretzky. 2019. AI4K12. https://github.com/touretzkyds/ai4k12/wiki.

[48] Sherry Turkle and Seymour Papert. 1990. Epistemological pluralism: Styles and voices within the computer culture. *Signs: Journal of women in culture and society* 16, 1 (1990), 128–157.

[49] James Vincent. 2017. The Verge Robots and AI are going to make social inequality even worse, says new report. https://www.theverge.com/2017/7/13/15963710/robots-ai-inequality-social-mobility-study.

[50] Thomas Way, Lillian Cassel, Paula Matuszek, Mary-Angela Papalaskari, Divya Bonagiri, and Aravinda Gaddam. 2016. Broader and earlier access to machine learning. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. 362–362.

[51] Thomas Way, Mary-Angela Papalaskari, Lillian Cassel, Paula Matuszek, Carol Weiss, and Yamini Praveena Tella. 2017. Machine learning modules for all disciplines. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*. 84–85.

[52] Michelle Hoda Wilkerson and Joseph L Polman. 2020. Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences* 29, 1 (2020), 1–10.

[53] Yin-Ling I Wong, Trevor R Hadley, Dennis P Culhane, Stephen R Poulin, Morris R Davis, Brian A Cirksey, and James L Brown. 2006. Predicting staying in or leaving permanent supportive housing that serves homeless people with serious mental illness. *Departmental Papers (SPP)* (2006), 111.

[54] Benjamin Xie, Greg L Nelson, and Amy J Ko. 2018. An explicit strategy to scaffold novice program tracing. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. ACM, 344–349.

[55] Moira L Zellner, Leilah B Lyons, Charles J Hoch, Jennifer Weizeorick, Carl Kunda, and Daniel C Milz. 2012. Modeling, Learning, and Planning Together: An Application of Participatory Agent-based Modeling to Environmental Planning. *Journal of the Urban & Regional Information Systems Association* 24, 1 (2012).

[56] Baobao Zhang and Allan Dafoe. 2019. Artificial Intelligence: American Attitudes and Trends. *Available at SSRN 3312874* (2019).

[57] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K Kane, and R Benjamin Shapiro. 2019. Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 121–132.