What is AI Literacy? Competencies and Design Considerations

Duri Long
Georgia Institute of Technology
Atlanta, GA, USA
duri@gatech.edu

Brian Magerko
Georgia Institute of Technology
Atlanta, GA, USA
magerko@gatech.edu

ABSTRACT
Artificial intelligence (AI) is becoming increasingly integrated in user-facing technology, but public understanding of these technologies is often limited. There is a need for additional HCI research investigating a) what competencies users need in order to effectively interact with and critically evaluate AI and b) how to design learner-centered AI technologies that foster increased user understanding of AI. This paper takes a step towards realizing both of these goals by providing a concrete definition of *AI literacy* based on existing research. We synthesize a variety of interdisciplinary literature into a set of core competencies of AI literacy and suggest several design considerations to support AI developers and educators in creating learner-centered AI. These competencies and design considerations are organized in a conceptual framework thematically derived from the literature. This paper's contributions can be used to start a conversation about and guide future research on AI literacy within the HCI community.

Author Keywords
AI literacy; AI education; AI for K-12; artificial intelligence; machine learning; computing education

CSS CONCEPTS
• **General and reference~Surveys and overviews**
• **Social and professional topics~Computing literacy**
• **Computing methodologies~Artificial intelligence**

INTRODUCTION
Artificial intelligence is becoming increasingly integrated in user-facing technologies. However, algorithms on common platforms can be opaque to users, who often do not recognize they are interacting with AI [10,54,55]. These misconceptions can limit people's ability to effectively use, collaborate with, and act as critical consumers of AI [57]. Widely held misconceptions about AI can also lead to misdirected regulatory action [124] and public letdown if expectations for development are not met [57].

CHI '20, April 25–30, 2020, Honolulu, HI, USA

ACM 978-1-4503-6708-0/20/04…$15.00
https://doi.org/10.1145/3313831.3376727

Design and education both play a role in contributing to public misunderstandings about AI. Black-box algorithms (i.e. algorithms with obscured inner-workings) can cause misunderstandings about AI [55]. On the other hand—even with more transparent technologies—a lack of technical knowledge on the part of the user can lead to misconceptions [25]. There is a clear need for a better understanding of this space from the perspectives of both learners and designers.

Researchers in the HCI community have begun to address public misconceptions of AI by investigating how people make sense of AI (e.g. [46]) and exploring how to design more understandable technology (e.g. [67]). However, there is a need for additional research investigating what new competencies will be necessary in a future in which AI transforms the way that we communicate, work, and live with each other and with machines. We refer to this set of competencies as *AI literacy*.

Emerging research is exploring how to foster AI literacy in audiences without technical backgrounds. Within the past year, companies have pursued initiatives to broaden AI education to underrepresented audiences in an effort to increase workforce diversity [5,148], educators have published guides on how to incorporate AI into K-12 curricula [145], and researchers are exploring how to engage young learners in creative programming activities involving AI [45,79,132,146,149]. The "AI for K12" working group is currently developing a set of standards for K-12 classrooms to determine what each grade band should know about AI [130]. The group has also identified five "big ideas" of AI to guide the standards development: 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 with humans is a substantial challenge for AI developers"; and 5) "AI applications can impact society in both positive and negative ways" [130].

The five "big ideas" of AI provide a strong foundation for future research on fostering AI literacy. However, most of the research on AI education for non-technical learners has just been published within the last year. In contrast, AI as a field has been active since the 1950s, and there is a variety of existing research (scattered across disciplines and venues) that could contribute to understanding what competencies should be included in a definition of AI literacy and how to better design educational experiences that foster AI literacy.

We engaged in an exploratory review of literature with the goal of distilling key ideas from various fields that could inform our understanding of how learners make sense of AI. We organize these key ideas in a conceptual framework that we thematically derived from the literature. The main contributions of this paper are a concrete definition of AI literacy and a related set of competencies and design considerations. This framework is not intended to be an exhaustive summary of the literature; rather, it is a set of key ideas/provocations we distilled from the literature that can serve as inspiration and initial guidelines for the design of future learning experiences centered on AI literacy. We present this framework as the start of a conversation, with the expectation that it will shift, grow, and spark debate in the future as more research is conducted in the field.

The next section of this paper presents a definition of AI literacy in the context of a broader discussion of literacy as a concept and how it has been applied in various related disciplines. We then present a conceptual framework—consisting of AI literacy competencies and design considerations—that we derived by conducting a review of a variety of interdisciplinary research related to AI.

DEFINING AI LITERACY

The term *literacy* as it was originally construed refers to *the ability to express ourselves and communicate using written language*. Fostering more widespread literacy has historically had political and emancipatory consequences, broadening access to knowledge and the ability for people to share and communicate ideas [61]. The notion of literacy has more recently been applied to defining skill sets in a variety of disciplines that have the same potential to enable expression, communication, and access to knowledge. Some examples include *digital literacy* (i.e. competencies needed to use computational devices [14]), *computational literacy* (i.e. the ability to use code to express, explore, and communicate ideas [40]), *scientific literacy* (i.e. "an appreciation of the nature, aims, and general limitations of science, coupled with some understanding of the more important scientific ideas" [86]); and *data literacy* (i.e. "the ability to read, work with, analyze, and argue with data as part of a broader process of inquiry into the world" [36]).

We 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.* The competencies and design considerations outlined in the remainder of this paper provide a more specific understanding of the contents of this skillset.

AI literacy is clearly related to other previously defined literacies in related fields. We see these relationships manifesting in several ways. Digital literacy is a prerequisite for AI literacy, as individuals need to understand how to use computers to make sense of AI. Computational literacy, however, is not necessarily a prerequisite for AI literacy. Understanding how to program can inform and aid in making sense of AI and is certainly necessary for AI developers. However, programming can also be a major barrier to entry for learners, and we argue that most individuals interacting with AI in their daily lives will not need to know how to program it. In this paper we define a set of skills that can aid in understanding AI that do not require learners to know how to write code. Scientific literacy can similarly inform AI literacy (particularly understanding machine learning practices [117]) but is not a required prerequisite. Finally, data literacy is closely related to the AI subfield of machine learning, and therefore certain data literacy competencies overlap with AI literacy competencies defined in this paper.

METHODOLOGY

We conducted an exploratory review of interdisciplinary literature in order to define a) a detailed set of AI literacy competencies for learners and b) design considerations for developers of learner-centered AI. Due to the limited amount of existing peer-reviewed literature on AI education and the variety of research in related fields that can inform AI education, we did not conduct a traditional systematic literature review. Our methods were instead more closely aligned with an approach called *scoping studies*, which

> aim to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available...especially where an area is complex or has not been reviewed comprehensively before [9].

In scoping studies, researchers do not place "strict limitations on search terms, identification of relevant studies, or study selection at the outset" and "the process is not linear but iterative" [9]. The goal of a scoping study is typically to identify all relevant literature "regardless of study design" as well as to identify gaps in the literature [9].

Our review was guided by two key research questions: 1) What do AI experts think non-technical learners should know about AI? and 2) What existing perceptions and misconceptions do non-technical learners have when interacting with AI?. The literature we reviewed in response to these two questions included 150 documents (Table 1). The first author conducted the literature review, consulting the second author for feedback and relevant expertise.

Year published		**Venue Type**	
Before 2000	8	Conference papers	53
2000 - 2009	43	Journal papers	38
2010 - 2017	55	Books	15
2018 - 2019	44	Other grey literature	44

Table 1: Breakdown of papers reviewed by year and venue type

We began our literature review by searching for papers related to AI education by closely following updates on the AI4K12 mailing list and reading papers by researchers currently active in the field, searching the proceedings of the

2008 AAAI AI Education Colloquium and proceedings of several post-2016 conferences including AAAI, AI Ed, CHI, and IDC; and searching Google Scholar and the ACM digital library. Search terms used were iteratively revised and included: “AI education”, “learning about AI”, “teaching AI”, “AI literacy”, “ML literacy”, “understanding ML” “understanding AI”, “AI for K-12”, “AI university”, “AI courses”, “AI school”, “AI informal learning”. We also searched for papers related to using robotics for AI—not CS—education. We focused on reviewing papers on AI education for non-technical learners, although we reviewed several papers on university courses. After identifying an initial set of papers, we reviewed the reference lists to find additional literature. This entire search yielded 18 papers and 8 projects related to AI education for non-technical learners and 14 papers on AI education for CS undergraduates.

Since our initial search revealed only a few recent papers on the topic of AI education for non-technical learners, we expanded our search to related fields and “grey literature” (i.e. literature that is not peer-reviewed). We examined 14 public syllabi from accredited universities in the USA for classes related to artificial intelligence (4), machine learning (6), cognitive science (2), and robotics (2). We looked at the contents of popular AI textbooks (3), seminal writings in AI research (10), papers related to AI ethics (22) and explainable AI (10), and polls on public perceptions of AI (9). We also explored peer-reviewed literature on perceptions of AI (23) by searching for papers with terms such as “perceptions of AI”, “misconceptions AI”, “AI in the home”, “interactions with AI”, “AI in media”. Finally, we reviewed select survey-style papers on related forms of literacy (e.g. digital, data, scientific literacy) (6), and looked at papers on CS education (13) relating to the AI education literature to see if there was support for these findings in a more established field.

We thoroughly read papers focused on AI education for learners without technical backgrounds. We read the abstracts and skimmed the contents of papers focused on AI education for experts. We also thoroughly read grey literature on AI education and perceptions of AI as well as the select papers from related fields. In a running document, we listed key ideas from each paper and grouped them based on similarity, drawing connections between the literature. We distilled competencies and design principles from this list by asking three questions: 1) does this reflect our definition of AI literacy?; 2) is this supported by numerous sources in the literature?; and 3) is this a useful guideline for designers and educators?. We then sorted the design considerations and competencies into thematic groups.

CONCEPTUAL FRAMEWORK

Our literature review resulted in a conceptual framework composed of five different overarching themes, which we frame as questions about AI: *What is AI?*; *What can AI do?; How does AI work?*; *How should AI be used?*; and *How do people perceive AI?*. These themes provide the structure for the remainder of the paper—for each theme, we include a set of competencies and design considerations. After each competency, we list supporting references.

What Is AI?

Defining what AI is can be confusing even for experts [116,124], as the term has evolved over the course of many years. Figuring out what AI is can be even more complex for individuals without a technical background, as AI is often overblown and conflated with other areas of computing in popular media. Many people think that AI is synonymous with robotics [57,138,145], and artifacts that do not achieve human-level intelligence are often discounted as being “not AI” (a phenomena referred to as the *superhuman human fallacy* [18]). AI is also often obscured on commonly used platforms—as a result, many users do not realize when they are interacting with AI [10,54,55,73]. The ability to recognize AI (Competency 1 (Recognizing AI)) is a critical skill necessary for informed interactions with AI.

Established definitions of AI can aid learners in understanding what AI is. Nilsson defines AI as “that activity devoted to making machines intelligent…[where] intelligence is that quality that enables an entity to function appropriately and with foresight in its environment” [100]. However, Schank notes that definitions of intelligence can differ depending on the researcher and their approach to understanding AI [116]. He suggests that there are two main goals to AI research—to “build an intelligent machine” and to “find out about the nature of intelligence” [116]. He then proposes a set of traits that comprise general “intelligence”—communication, world knowledge, internal knowledge, intentionality, and creativity—emphasizing that the ability to learn is the most critical component of intelligence [116].

Brooks provides a contrasting definition, taking a bottom-up approach to understanding intelligence [21]. He suggests that developing human-level intelligence is too lofty a goal, and instead we should focus on understanding intelligence incrementally, starting with simple levels of intelligence (e.g. that of an insect). Brooks argues that by excluding tasks such as perception and motor response and conducting experiments in controlled environments, AI researchers are abstracting away the most challenging components of intelligence [21]. He suggests instead developing “completely autonomous mobile agents” that are capable of perceiving, acting, and pursuing a set of goals in a dynamic environment [21]. These agents would not be capable of human-level intelligence at first but would autonomously operate in the real world.

Others have synthesized perspectives on intelligence into summative definitions. Russell and Norvig describe intelligence in terms of *thinking* or *acting* either *humanly* (i.e. based on an empirical understanding of human intelligence) or *rationally* (i.e. based on mathematical principles) [115]. Goel and Davies characterize AI as the intersection of three disciplines—cognitive systems, robotics, and machine learning [64]. These definitions suggests that it is important for learners to be able to examine what it means to be intelligent (Competency 2 (Understanding Intelligence)). Activities like comparing AI devices [69] and AI vs. human abilities [145] have been used to promote this understanding. Taken in conjunction with recent calls for broadened AI curricula [117,145], these definitions of intelligence also suggest the importance of understanding that AI is interdisciplinary (Competency 3 (Interdisciplinarity)).

Each one of the three areas of AI has produced "narrow AI", or AI that is intelligent within a particular domain, but "general AI", or AI that rivals human intelligence across multiple domains, has yet to be achieved [64]. This distinction has implications for understanding AI and its capabilities, suggesting Competency 4 (General vs. Narrow).

> **Competency 1 (Recognizing AI)**
> Distinguish between technological artifacts that use and do not use AI.
> *Supporting References:* [10,18,54,55,57,73,116,124,138,145]
>
> **Competency 2 (Understanding Intelligence)**
> Critically analyze and discuss features that make an entity "intelligent", including discussing differences between human, animal, and machine intelligence.
> *Supporting References:* [21,64,69,100,115,116,125]
>
> **Competency 3 (Interdisciplinarity)**
> Recognize that there are many ways to think about and develop "intelligent" machines. Identify a variety of technologies that use AI, including technology spanning cognitive systems, robotics, and ML.
> *Supporting References:* [64,115,117,145]
>
> **Competency 4 (General vs. Narrow)**
> Distinguish between general and narrow AI.
> *Supporting References:* [57,58,64]

What Can AI Do?

Consumer polls indicate that people's trust in AI is heavily task-dependent [10,106]. Having accurate knowledge of AI's ability to complete different types of tasks can therefore help people to make more informed decisions about how to use and when to trust AI. While AI is good at detecting patterns in large amounts of data, doing repetitive tasks, and making decisions in controlled environments, humans currently remain better at most tasks requiring creativity, emotion, knowledge transfer, and social interaction. Understanding the current capabilities of AI—and that there are still many open questions in AI research (the fifth "big idea" of AI [130])—can help users in making more informed decisions. In addition, individuals will likely be more well-equipped to leverage the different capabilities of AI and humans to solve problems if they understand AI's strengths and weaknesses (Competency 5 (AI's Strengths & Weaknesses)).

AI is rapidly changing and in order to plan for the future, make long-term policy decisions, and evaluate potential consequences, it is important for individuals to consider not just what AI *can* do in the present, but also what AI *could* do in the future. One way of fostering this skill is by creating *design fictions* (i.e. fictional scenarios about what designed artifacts may exist in the future and what effects those artifacts will have on the world) [88]. Design fictions have been used as a tool for exploring the effects of AI on future cities with citizen stakeholders [143], for understanding children's perceptions of AI devices [43], and in K-12 AI ethics education [6]. The ability to imagine "future AI" can enable individuals to creatively explore novel ideas, consider the values inherent in a technology, and critically evaluate the long-term effects a technology may have on the world (Competency 6 (Imagine Future AI)).

> **Competency 5 (AI's Strengths & Weaknesses)**
> Identify problem types that AI excels at and problems that are more challenging for AI. Use this information to determine when it is appropriate to use AI and when to leverage human skills.
> *Supporting References:* [10,22,106,124,125,130]
>
> **Competency 6 (Imagine Future AI)**
> Imagine possible future applications of AI and consider the effects of such applications on the world.
> *Supporting References:* [6,43,143,145]

How does AI work?

Many people self-report that they know little about AI [138]. Despite this, people often develop "folk theories" (i.e. "informal theories...to perceive and explain how a system works") to explain AI algorithms [55]. These theories, whether accurate or not, shape the nature of user interaction and experience [55]. A better understanding of how AI works can help people to form more accurate mental models of the systems they interact with. For this reason and others, most existing research on AI education in university and K-12 environments is focused on communicating how AI works.

We conducted a review of topics covered in university syllabi for ML [26,87,89:229,90,98,123], AI [30,72,78,113], cognitive science [63,112], and robotics [12,82] courses by writing down a list of all topics covered in the schedules. We also listed learning goals outlined in AI education initiatives for K-12 audiences [5,45,69,130,142]. Topics ranged from high-level concepts (e.g. learning, kinematics, planning) to specific implementations (e.g. Bayesian networks, Markov models). Most syllabi were targeted at CS majors, and many of the K-12 initiatives also required some prerequisite math, statistics, or CS knowledge. This level of prerequisite knowledge may make such courses and their content inaccessible to groups who could benefit from AI literacy, such as children interacting with AI in their homes or adults using AI in the workplace. For this reason, we focus on the higher-level concepts and "epistemological practices" [117] in the syllabi rather than on implementation details of specific algorithms. We review work spanning all three areas of AI—cognitive systems, robotics, and ML.

Cognitive Systems

Cognitive systems—or AI systems that are modeled after theories about the human mind [64]—are used in a variety of application domains, including WordNet, IBM's Watson, expert systems, and cognitive tutors. Most cognitive systems syllabi cover topics related to knowledge representations, planning, decision-making, problem-solving, and learning.

Knowledge representations model the world in a way that is understandable to a computer [92]. For example, an image is represented as a matrix of float values in which each value represents the color of a pixel. The average user interacting with AI likely does not require an in-depth understanding of how to implement knowledge representations (the focus of many university courses). However, a conceptual understanding of representations (one of the "big ideas" of AI [130]) could aid users in understanding how computers represent knowledge and in recognizing that some knowledge is always lost in a representation of the world [92] (Competency 7 (Representations)).

Cognitive systems use many strategies for *planning*, *decision making*, *problem solving*, and *learning*. Users likely do not need to understand all of these strategies in detail, but a high-level understanding of how computers make decisions can aid in interpreting and understanding algorithms [29] (Competency 8 (Decision-Making)). *Explainable AI* (i.e. AI that provides the user with explanations of why it delivered a particular outcome) is one way of helping users learn about agent reasoning. Many of these systems are intended for expert users, but recent research has started to use explainable AI to aid novices in understanding how AI works. Strategies employed in these contexts include providing interactive demonstrations and visualizations (e.g. [27,65,126,150]), having learners test hypotheses in simulation environments (e.g. [48,128]), presenting explanations using storytelling techniques (e.g. [50]), and providing explanatory debugging capabilities (e.g. [83]). These strategies can be utilized when designing learning interventions (Design Consideration 1 (Explainability)). However, it is important to consider how many components of a system to explain. Research has shown that exposing the "inner-workings" of too many components can overwhelm users [109], whereas too few can inhibit learning [71].

Competency 7 (Representations)
Understand what a knowledge representation is and describe some examples of knowledge representations.
Supporting References: [30,72,78,92,113,130]

Competency 8 (Decision-Making)
Recognize and describe examples of how computers reason and make decisions.
Supporting References: [29,30,72,78,113]

Design Consideration 1 (Explainability)
Consider including graphical visualizations, simulations, explanations of agent decision-making processes, or interactive demonstrations in order to aid in learners' understanding of AI.
Supporting References: [27,43,48,50,65,83,126,128,150]

Machine Learning

Machine learning (ML) is an important tool in a wide variety of disciplines ranging from social media to healthcare. However, little research has explored how to teach ML, which arguably has more in common with scientific practice in disciplines like chemistry or physics than deterministic approaches to AI in cognitive systems and robotics [117]. Some recent work is beginning to investigate how to teach ML to individuals without a CS background (e.g. [45,125,145]). Some of these initiatives focus on teaching non-experts how to implement ML algorithms; others focus on communicating more high-level practices such as data gathering and preparation, model selection, training, testing, and prediction [117,145] (Competency 9 (ML Steps)).

Sulmont et al. have begun to explore what misconceptions students without a background in CS or statistics have in introductory university ML courses [125]. Many students assume that computers think like humans and want to make connections between human theories of cognition and machine learning [125] (supporting Competency 2 (Understanding Intelligence)). Students are also often surprised that ML requires human decision-making and is not entirely automated (suggesting Competency 10 (Human Role in AI)). Finally, students often have difficulty identifying the limits of ML and identifying constraints that may make ML unsuitable for solving a particular problem (supporting Competency 5 (AI's Strengths & Weaknesses)).

Research suggests that one way of dispelling student misconceptions about ML is to engage in embodied interaction. Sulmont et al. and others suggest having students physically enact algorithms in order to understand them in a more concrete way (Design Consideration 2 (Embodied Interactions)) [45,69,125]. This tactic has also been used in CS education [2]. More broadly, embodied hands-on experimentation with AI has been used as an approach in a variety of AI education initiatives (e.g. [45]), including projects in which learners can train ML models to analyze their athletic moves and gestures [4,146].

Research on data literacy education can also inform our understanding of how to design ML-related learning interventions. Prado and Marzal define a set of competencies for data literacy (e.g. the "ability to critically assess data and their sources") [107]. The importance of these competencies to understanding ML suggests that knowledge of basic data science concepts is a component of AI literacy (Competency 11 (Data Literacy)). Recognizing when personal data is being used to train ML and interpreting the results of algorithms in the context of the data they were trained on are two particularly relevant data literacy issues for AI. Research suggests that it is important for learners to understand that computers learn from their data [68,130] (Competency 12 (Learning from Data)) and that learners should be able to critically examine data with "skepticism and interpretation" [68] (Competency 13 (Critically Interpreting Data)).

A variety of tactics can be used to promote critical engagement with data and ML. Hautea et al. suggest having young learners creatively engage with data that is collected about them online [68]. D'Ignazio and Sulmont et al.

encourage educators to carefully select the datasets they use in class, favoring datasets that are low-dimensional when initially introducing concepts [125]; datasets that are "messy" (i.e. not cleaned and neatly categorizable) when demonstrating issues of bias [36]; and incorporating personally relevant datasets that learners can easily relate to and understand [36]. Finally, D'Ignazio suggests writing "data biographies" (i.e. contextual explanations of datasets and their origins) as a way of helping learners better understand the limitations and origins of data [36] (Design Consideration 3 (Contextualizing Data)).

Competency 9 (ML Steps)
Understand the steps involved in machine learning and the practices and challenges that each step entails.

Supporting References: [45,117,125,145]

Competency 10 (Human Role in AI)
Recognize that humans play an important role in programming, choosing models, and fine-tuning AI systems.

Supporting References: [22,125]

Competency 11 (Data Literacy)
Understand basic data literacy concepts such as those outlined in [107].

Supporting References: [36,68,107]

Competency 12 (Learning from Data)
Recognize that computers often learn from data (including one's own data).

Supporting References: [36,68,107,130]

Competency 13 (Critically Interpreting Data)
Understand that data cannot be taken at face-value and requires interpretation. Describe how the training examples provided in an initial dataset can affect the results of an algorithm.

Supporting References: [6,36,68,107,130,145]

Design Consideration 2 (Embodied Interactions)
Consider designing interventions in which individuals can put themselves "in the agent's shoes" [45] as a way of making sense of the agent's reasoning process. This may involve embodied simulations of algorithms and/or hands-on physical experimentation with AI technology.

Supporting References: [2,45,46,69,71,76,103,125]

Design Consideration 3 (Contextualizing Data)
Encourage learners to investigate who created the dataset, how the data was collected, and what the limitations of the dataset are. This may involve choosing datasets that are relevant to learners' lives, are low-dimensional, and are "messy" (i.e. not cleaned or neatly categorizable).

Supporting References: [36,68,107,125,130]

Robotics

The third branch of AI is *robotics*, or AI systems that can physically act on and react to the world. Most existing research on robotics education uses robotics as a context to teach design thinking [77,94] mathematics [77,131], physics [77], computational thinking [42,66,131], or software engineering [131]. Some research explores how to use robotics to teach AI concepts such as: sensors and integrating sensing, perception, and action [94,115]; representations that are used to localize and guide robot movement [42,131]; decision making, search, and planning algorithms necessary to plan robot action [81,84,105,131]; using ML (especially vision) to make sense of sensorial input [114,131,132]; understanding reactive control [115]; and using effectors and kinematic trees to control a robot's body [115,131].

Many of the AI-related competencies from robotics overlap with ML and cognitive systems. However, concepts such as reactive control and understanding perception and action sensors are specific to robotics. Understanding that AI agents can physically act on and react to the world is an important prerequisite for understanding robotics (Competency 14 (Action & Reaction)). Learning about sensors and their capabilities (one of the "big ideas" of AI [130]) can also aid in understanding how AI devices gather data and interface with the world (Competency 15 (Sensors)).

Competency 14 (Action & Reaction)
Understand that some AI systems have the ability to physically act on the world. This action can be directed by higher-level reasoning (e.g. walking along a planned path) or it can be reactive (e.g. jumping backwards to avoid a sensed obstacle).

Supporting References: [42,115,131]

Competency 15 (Sensors)
Understand what sensors are, recognize that computers perceive the world using sensors, and identify sensors on a variety of devices. Recognize that different sensors support different types of representation and reasoning about the world.

Supporting References: [94,114,115,131,132]

How Should AI Be Used?

There are many ethical questions surrounding how AI should be used, and there has been growing concern surrounding issues such as AI's effect on the job market [57], bias and discrimination in AI [7,24,99], and AI-related data privacy scandals [119]. It is clear that "AI applications can impact society in both positive and negative ways" (the fifth "big idea" of AI) [130]. Recent educator-led initiatives are developing curricula for AI ethics education for non-technical learners [6,145]. Below, we list key ethical issues surrounding AI based on these initiatives, textbooks related to technology and ethics [8,108], and a review of the papers presented at the Fairness, Accountability, and Transparency in ML conference since 2016.

Privacy/surveillance: The amount of personal data that is collected, stored, and analyzed in order for many AI systems to function has raised concerns about user privacy [119], government surveillance [56], and data security [35].

Employment: Advances in automation have reduced the need for human workers while also increasing productivity, an issue that has generated concern long before AI [11]. However, advancements in AI have heightened concerns about technology replacing the human workforce [75,144].

Misinformation: The spread of misinformation and "fake news" has been exacerbated by AI algorithms on social media and search engines that promote "clickbait" articles and create "filter bubbles" [104].

Singularity/concern about harm to people: The idea of "the singularity"—or the time when machine intelligence surpasses human intelligence [85]—has been popularized in science fiction, and many have concerns about AI intentionally causing harm to people [13,144].

Ethical decision making: Most computing ethics syllabi and textbooks emphasize that embedding ethical decision-making strategies in technical systems is a challenging problem [8,108]. Giving decision-making power to AI can result in ethical dilemmas such as the trolley problem [129] or unexpected results due to AI executing actions that people *tell* it to do rather than doing what people *intend* it to do (e.g. a self-driving car driving at 125 mph because it was told to get to the airport "as fast as possible") [35].

Diversity: Diversity in the CS workforce is an issue, and gender diversity in AI is no exception—in 2018, 80% of AI professors and 71% of applicants to AI-related jobs identified as male [122]. Lack of workforce diversity can affect who systems are developed for [33]—a significant issue in AI, where biased algorithms can have pronounced adverse effects on marginalized subgroups [32].

Bias/fairness: Most of the papers in the 2018 FAT ML conference focused on issues related to algorithmic bias (e.g. [118,121]). Algorithmic bias is often directly related to bias present in training datasets. Agents in-the-wild are also able to learn bias and bigotry from human users [99].

Transparency: Many AI algorithms (especially in ML) are black-box and their functionality (and sometimes even existence) can be opaque to users [55]. This can lead to deception and misunderstanding. [55]. The ACM recently defined seven principles relating to algorithmic transparency and accountability as part of its code of ethics, suggesting that additional tactics are needed to address issues of transparency (e.g. developing explainable AI, testing and documenting models, and promoting bias awareness) [3].

Accountability: A major issue with AI being used to make life-altering decisions in areas such as hiring or recidivism is that there is often no way to report algorithmic errors [134], receive feedback on why decisions were made [51], or hold anyone accountable for errors that adversely affect people's lives. The EU's recent GDPR legislation mandates that "data subjects" have the right to challenge decisions made by AI and receive an explanation, but this remains challenging in practice [52].

The current ACM guidelines for undergraduate CS curricula include an ethics course in which students learn about ethical theories and apply them to evaluate technology, focusing on many of the issues described above. Such skills can help both computing professionals and everyday users to identify when it is appropriate use AI (Competency 16 (Ethics)).

AI ethics education initiatives use a variety of interdisciplinary strategies to communicate key ethical concepts, including creating "ethical matrices" to consider values of different stakeholders in technology, imagining future AI and its implications, reflecting on AI representations in popular media and the news, discussing and debating key ethical questions, and engaging in programming activities that spur learners to critically examine algorithms and bias [6,69,93,145]. In informal spaces, artists and researchers have created interactive art experiences that spur participants to question the implications of technologies like facial recognition [34,70].

> **Competency 16 (Ethics)**
> Identify and describe different perspectives on the key ethical issues surrounding AI (i.e. privacy, employment, misinformation, the singularity, ethical decision making, diversity, bias, transparency, accountability).
> *Supporting References:* [3,6,8,35,93,108,130,145]

How Do People Perceive AI?

It is important to understand existing public conceptions of AI in order to develop effective AI literacy interventions that build on prior knowledge. The past several sections have touched on some of these preconceptions, but this section goes into a more in-depth review of research that has focused on how humans perceive and make sense of AI.

Interpreting AI Systems

Humans understand the actions of other agents using *theory of mind*, or our ability to "explain and predict other people's behavior by attributing to them independent mental states, such as beliefs and desires" [62]. However, due to the differences between AI and human reasoning, theory of mind is not always a reliable way of making sense of AI [111]. As a result, misconceptions can arise when interpreting interactions with intelligent systems.

Wardrip-Fruin describes three effects "that can arise in the relationship between the surface appearance of a digital system and its internal operations" [137]. The *Eliza effect* is a misconception that occurs when a system uses simple techniques but produces effects that appear complex [137]. Humans often attribute much more intelligence to these systems than they actually possess. In contrast, the *Tale-Spin effect* refers to a system that has complex internal operations, but externally appears "significantly less complex" [137]. These effects result from a lack of transparency—often it is impossible to discern via interaction how these systems work internally. Finally, the *SimCity effect* refers to "a system that, through play, brings the player to an accurate understanding of the system's internal operations" [137].

Some of these misconceptions may be caused by opaque technologies that obscure functionality. The Turing Test, which has long been used to assess whether an agent is intelligent, is based on the idea that if a computer can fool a person into thinking it is human, it can be considered intelligent. Miller calls machines that masquerade as humans

Turing deceptions and suggests that they may be ethically problematic [97]. For instance, introducing black-box AI decision-making algorithms into popular platforms without informing users can lead to concern and apprehension [55]. Researchers seeking to foster AI literacy may want to avoid misleading tactics like Turing deceptions and black-box algorithms [45]. While black-boxing system components can minimize cognitive overload [71], it can also lead to issues with accountability, bias, and misunderstanding. Balance can be achieved by giving users the option to inspect and learn about system components, explaining only a few components at once, or introducing scaffolding that fades as the user learns about the system (Design Consideration 5 (Unveil Gradually)). It is important to keep in mind that many factors affect how humans interpret explanations, including the framing of an explanation given by an AI agent [31]. Statements that imply agency and intentionality, like "I selected this because it seemed like something you would enjoy," typically lead to higher perceptions of intelligence than technical statements like "I selected this because it was 15% more similar to your previous choices than other options in the decision space" [31].

Mateas further discusses how people make sense of AI, highlighting the role that the AI creator plays in mediating user interpretations. He describes *interpretive affordances,* or "actionable properties of objects in the world" that support "the interpretations an audience makes about the operations of an AI system" [95]. Interpretive affordances help users make sense of a system's operations, understand how to interact with it, and understand the creator's intentions. Interpretive affordances and other strategies that promote transparency can aid in improving user understanding of AI (Design Consideration 4 (Promote Transparency)).

Design Consideration 4 (Promote Transparency)
Promote transparency in all aspects of AI design (i.e. eliminating black-boxed functionality, sharing creator intentions and funding/data sources, etc.). This may involve improving documentation, incorporating explainable AI (Design Consideration 1), contextualizing data (Design Consideration 3), and incorporating design features such as interpretative affordances or the Sim-City Effect.

Supporting References: [3,36,41,45,55,67,95,97,111,137]

Design Consideration 5 (Unveil Gradually)
To prevent cognitive overload, consider giving users the option to inspect and learn about different system components; explaining only a few components at once; or introducing scaffolding that fades as the user learns more about the system's operations.

Supporting References: [41,45,71,109]

Children's Perceptions of AI

Most children do not develop theory of mind until they are 3-5 years old [139], which leads to additional complexities in understanding how children make sense of AI. Research has also shown that early exposure to technology (specifically AI) can shape the way that children think about concepts like what it means to be alive or intelligent [16,133]. Several studies have examined how children make sense of AI systems such as My Friend Kayla [141], AIBO [16], and Siri [46]. This section examines children's perceptions of AI and strategies for helping children better understand AI. Some of these strategies are child-specific and some are more broadly relevant to adult audiences.

Children's perceptions of agent intelligence are dependent on a variety of factors. Children tend to focus on observable characteristics (e.g. success) rather than unobservable ones (e.g. strategy) when assessing agent intelligence [47] [59]. Age also plays a role in shaping perceptions. Children over 8 tend to agree with their parent's assessments of agent intelligence, whereas younger children tend to overestimate intelligence, often perceiving agents to be smarter than themselves [47]. Agent form may also make a difference in perceptions of intelligence. Children generally accept that robots can be intelligent even though they are not alive and do not have brains [16]. However, prior work indicates that children think robots are "ontologically different from other objects, including computers" [46,80], suggesting that children may perceive the intelligence of other types of AI differently (though there is little research on this topic).

Research indicates that children first personify agents and then recognize that they are programmable [47,69,80]. This recognition is foundational for understanding how AI works (Competency 17 (Programmability)), and providing opportunities for learners of all ages to program AI can foster this understanding (Design Consideration 6 (Opportunities to Program). Several recent projects such as Cognimates [44], eCraft2Learn [151], and others [4,146,149] enable young learners to program AI. However, it is important for designers to keep in mind that prerequisite coding skills can be a barrier to entry, especially for children who are still learning how to read [43,69]. Visual and auditory elements [43], fill-in-the-blank code [69], and Parsons problems [53] are some techniques that can reduce this barrier.

Early experiences with technology can improve children's perceptions of agent intelligence [80], and lack of prior experience can inhibit children's ability to accurately assess what types of problems a computer can solve [135]. The influence of factors such as cognitive development, age, and prior experience on perceptions of intelligence should be taken into consideration when designing learning interventions (Design Consideration 7 (Milestones)).

Children often attribute socio-emotional characteristics to AI agents—more so than adults [46]. This is not affected by whether or not children believe the agent is alive [16]. Children have a tendency to personify agents and treat them like humans [46,69,127], and generally perceive agents as being both friendly and trustworthy [46,141]. This suggests that children may overestimate agent capabilities and put a lot of trust in agents. Design Consideration 8 (Critical Thinking) suggests encouraging all learners—but particularly children—to critically examine AI.

Both adult and child perceptions of intelligence and socio-emotional characteristics to AI agents may be affected by cultural upbringing and geographic location [10,45,138]. This suggests the importance of keeping learners' identities and backgrounds in mind (Design Consideration 9 (Identity, Values, & Backgrounds)). Making AI literacy interventions culturally relevant may also have the added benefit of increasing learner interest in AI—research on CS education has found that learning interventions centered around cultural values and personal identities are particularly effective, especially for underrepresented groups [38,49].

Research suggests that social interaction plays an important role in AI learning. Families often learn about AI together, but parents make fewer efforts to aid their children when they are simultaneously trying to learn about novel technologies [15]. Providing scaffolding for parents can aid them in supporting their children's learning [60] (Design Consideration 10 (Support for Parents)). Research has also shown that peer collaboration can be motivating, particularly for underrepresented learners [23,74,91,110,140] (Design Consideration 11 (Social Interaction)).

Children tend to prefer interacting with embodied agents that have social communication abilities [46,76]. Research suggests that social, embodied agents promote collaboration, conversation, and joyful interactions more than other styles of agents [76]. They can also foster learning about AI research on emotional intelligence [120]. Both adults and children also associate more socio-emotional qualities with agents that have faces [39,46]. This indicates that such agents are well-suited for designing engaging learning experiences. However, AI systems we interact with daily are often not social or embodied. A balance needs to be struck between fostering engaging interactions and providing exposure to a variety of forms of AI. This could involve designing social, embodied learning experiences around more common AI systems (Design Consideration 11 (Social Interaction), Design Consideration 2 (Embodied Interactions)).

Building on prior knowledge and interests can also contribute to engaging learning experiences [19]. Research in CS and AI education has shown that leveraging learners' interests in areas like music [91], games [37,96,136,147], or sports [146] can encourage learning, particularly in underrepresented groups (Design Consideration 12 (Leverage Learners' Interests)). Recent research is investigating children's interests in AI. When asked to imagine future AI, kids drew robots, animals, and "things that can play games" [43]. Many wanted AI to do things for them that they did not want to do (e.g. chores). Other desired abilities included conversing and school tasks [43].

Competency 17 (Programmability)
Understand that agents are programmable.
Supporting References: [45,47,79,80]

Design Consideration 6 (Opportunities to Program)
Consider providing ways for individuals to program and/or teach AI agents. Keep coding skill prerequisites to a minimum by focusing on visual/auditory elements and/or incorporating strategies like Parsons problems and fill-in-the-blank code.
Supporting References: [43,45,47,53,69,79,80]

Design Consideration 7 (Milestones)
Consider how developmental milestones (e.g. theory of mind development), age, and prior experience with technology affect perceptions of AI—particularly when designing for children.
Supporting References: [80,135,139]

Design Consideration 8 (Critical Thinking)
Encourage learners—and especially young learners—to be critical consumers of AI technologies by questioning their intelligence and trustworthiness.
Supporting References: [16,46,69,127,141]

Design Consideration 9 (Identity, Values, & Backgrounds)
Consider how learners' identities, values, and backgrounds affect their interest in and preconceptions of AI. Learning interventions that incorporate personal identity or cultural values may encourage learner interest and motivation.
Supporting References: [10,17,38,45,49,138]

Design Consideration 10 (Support for Parents)
When designing for families, consider providing support to aid parents in scaffolding their children's AI learning experiences.
Supporting References: [15,60]

Design Consideration 11 (Social Interaction)
Consider designing AI learning experiences that foster social interaction and collaboration.
Supporting References: [23,39,46,74,76,91,110,120,140]

Design Consideration 12 (Leverage Learners' Interests)
Consider leveraging learners' interests (e.g. current issues, everyday experiences, or common pastimes like games or music) when designing AI literacy interventions.
Supporting References: [19,37,43,91,96,136,146,147]

Perceptions of AI in Popular Media

The previous sections have addressed how people perceive specific AI systems. This section reviews research addressing how the public perceives AI more broadly. Representations of AI in news coverage and popular media both affect and reflect public perceptions about AI [138]. In this section we review public poll data, meta-analyses of news coverage, and representations of AI in other media.

A meta-analysis of New York Times articles has revealed numerous trends in AI-related coverage [57]. Coverage related to AI has generally increased over time, with the exception of the AI "winter" beginning in 1987 and a spike in coverage after 2009 [57]. The sentiment of discussion about AI has become more optimistic over time, although coverage of certain issues has become pessimistic recently (e.g. impact on work, loss of control of AI, ethical concerns) [57]. Polls have also found that there is a significant amount of public concern related to these issues [1,13,75,138]. There are notable gender and age differences in opinions about the

Design Consideration 13 (Acknowledging Preconceptions)
Acknowledge that learners may have politicized/sensationalized preconceptions of AI from popular media and consider how to address, use, and expand on these ideas in learning interventions.
Supporting References: [20,57,58,101,138]

Design Consideration 14 (New Perspectives)
Consider introducing perspectives in learning interventions that are not as well-represented in popular media (e.g. less-publicized AI subfields, balanced discussion of the dangers/benefits of AI).
Supporting References: [20,57,58,101,138]

development of AI—men and younger audiences tend to be more optimistic about AI development than women and older age groups [1,138]. Keywords associated with AI in news coverage have also transformed over time. Some keywords like "robot" were consistently associated with AI across the entire timeline, but others showed shifts in public concern—for instance, *space weapons* was a keyword commonly associated with AI in 1986; *search engines* in 2006; and *driverless vehicles* in 2016. Drawing on current public concerns is a way of leveraging learners' interests (Design Consideration 12 (Leverage Learners' Interests)).

Another meta-analysis found that recent news coverage on AI in the UK has been heavily dominated by industry, with 60% of ~760 articles focusing on industry products, and 12% of articles mentioning Elon Musk specifically [20]. The same analysis also found that AI issues are becoming politicized in the media—right-leaning outlets tend to highlight "issues of economics and geopolitics", whereas left-leaning outlets focus on "issues of ethics of AI" [20]. This suggests Design Consideration 13 (Acknowledging Preconceptions).

Other media such as television, movies, and science fiction can also have effects on perceptions of AI [28,101,138]. Many representations of AI in media are dystopian in nature, in which AI rebels against humanity (e.g. the *Terminator* film series) [58]. In other representations, humans are dominant but the way in which they treat AI is ethically problematic (e.g. *A.I.*). In some instances, AI appears in a benevolent form as a non-central character in a plot about a futuristic universe (e.g. droids in *Star Wars*). AI is most frequently represented in the form of a robot in popular media, and is generally shown as either a mindless killing machine, a complex device (e.g. Rosie the Robot in *The Jetsons*), or as a being with human-level intelligence [58]. AI in media are often treated as equivalents to human protagonists, with their own set of motivations, emotions, and problems (e.g. *Wall-E, Her*). Since AI is often represented as having human-level intelligence (which has not yet been approached in contemporary AI research), it is important for learners to be able to distinguish between AI's abilities in media vs. real life (Competency 5 (AI's Strengths & Weaknesses)). In addition, since media highlights certain types of AI while obscuring others, it is important for educators to share perspectives on AI that may be less well-represented (Design Consideration 14 (New Perspectives)).

Perceptions about Learning AI

Perceptions about AI can affect who seeks out opportunities to learn about AI. High school students who are not interested in studying CS often cite reasons such as the field being particularly demanding, a lack of prior exposure to the subject, and the perception of computers as "mechanical" or "cold", in contrast to more human-centered professions [102]. These perceptions likely also apply to the subfield of AI. Recent research focused on understanding student misconceptions in ML courses has highlighted some additional preconceptions students often hold: 1) believing ML is important, particularly for the job market; 2) hearing of ML through popular, often sensationalized, media; and 3) believing that implementing ML is not accessible without having a background in CS/math [125]. Math in particular is a barrier—students repeatedly self-identify as not able to do math in introductory ML classes [125]. These findings suggest the importance of lowering barriers to entry in AI education (Design Consideration 15 (Low Barrier to Entry)).

Gender may also play a role in shaping perceptions about learning AI. Research has shown that men are much more likely than women to tinker with and program in-home AI devices and that, compared with women, men perceive their tinkering to be more successful [17]. These differences may be a result of perceptions of perceived usefulness of tinkering as an activity—again suggesting the importance of considering learner interests and identity (Design Consideration 12 (Leverage Learners' Interests), Design Consideration 9 (Identity, Values, & Backgrounds)).

Design Consideration 15 (Low Barrier to Entry)
Consider how to communicate AI concepts to learners without extensive backgrounds in math or CS (e.g. reducing required prerequisite knowledge/skills, relating AI to prior knowledge, addressing learner insecurities about math/CS ability).
Supporting References: [102,125]

CONCLUSION AND FUTURE WORK

This paper provides an operational definition of AI literacy. In addition, it distills a set of AI literacy competencies and design considerations from a survey of interdisciplinary literature. It is important to keep in mind that research on AI education is still in its nascent stages. Much of the work we cite was just published in the last two years, and there is still a need for more empirical research in order to build a robust and accurate understanding of what existing preconceptions non-programmers have about AI and what the best practices are for teaching AI to a non-technical audience.

The competencies and design considerations outlined in this paper will likely need to be expanded to accommodate new findings, technologies, and rapidly changing social norms. We encourage researchers and educators in the HCI, AI, and learning science communities to both engage in conversation around the competencies and design considerations in this paper and use them to guide and inspire future empirical and design research on AI literacy.

REFERENCES

[1] 60 minutes. 2016. 60 Minutes/Vanity Fair poll: Artificial Intelligence.

[2] Edith Ackermann. 2004. Constructing knowledge and transforming the world. *A learning zone of one's own: Sharing representations and flow in collaborative learning environments* 1: 15–37.

[3] ACM. 2017. Statement on Algorithmic Transparency and Accountability. Retrieved from https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

[4] Adam Agassi, Hadas Erel, Iddo Yehoshua Wald, and Oren Zuckerman. 2019. Scratch Nodes ML: A Playful System for Children to Create Gesture Recognition Classifiers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*.

[5] AI4All. 2019. AI4All. Retrieved from http://ai-4-all.org/

[6] Safinah Ali, Blakeley H Payne, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education. In *Proceedings of IJCAI 2019*.

[7] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2: 211–36.

[8] Michael Anderson and Susan Leigh Anderson. 2011. *Machine ethics*. Cambridge University Press.

[9] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. *International journal of social research methodology* 8, 1: 19–32.

[10] Arm, Ltd. *AI Today, AI Tomorrow | Global AI Survey Results – Arm*. Arm, Ltd. Retrieved September 6, 2019 from https://www.arm.com/solutions/artificial-intelligence/survey

[11] Stanley Aronowitz and Jonathan Cutler. 2013. *Post-work*. Routledge.

[12] Harry Asada and John Leonard. 2005. Introduction to Robotics. Retrieved from https://ocw.mit.edu/courses/mechanical-engineering/2-12-introduction-to-robotics-fall-2005/syllabus/

[13] British Science Association and others. 2016. *One in three believe that the rise of artificial intelligence is a threat to humanity*.

[14] David Bawden and others. 2008. Origins and concepts of digital literacy. *Digital literacies: Concepts, policies and practices* 30: 17–32.

[15] Laura Beals and Marina Bers. 2006. Robotic technologies: when parents put their learning ahead of their child's. *Journal of Interactive Learning Research* 17, 4: 341–366.

[16] Debra Bernstein and Kevin Crowley. 2008. Searching for signs of intelligent life: An investigation of young children's beliefs about robot intelligence. *The Journal of the Learning Sciences* 17, 2: 225–247.

[17] Alan F Blackwell. 2006. Gender in Domestic Programming: From Bricolage to Séances d'Essayage. In *CHI'2006 Workshop on End User Software Engineering*, 1–4.

[18] Margaret A Boden. 2004. *The Creative Mind: Myths and Mechanisms*. Routledge, New York, NY, USA.

[19] John D Bransford, A Brown, and R Cocking. 1999. How people learn: Mind, brain, experience, and school. *Washington, DC: National Research Council*.

[20] J. Scott Brennen, Philip Howard, and Rasmus Nielsen. 2018. *An Industry-Led Debate: How UK Media Cover Artificial Intelligence*.

[21] Rodney A Brooks. 1991. Intelligence without representation. *Artificial intelligence* 47, 1–3: 139–159.

[22] Meredith Broussard. 2018. *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

[23] Philip Sheridan Buffum, Megan Frankosky, Kristy Elizabeth Boyer, Eric N Wiebe, Bradford W Mott, and James C Lester. 2016. Collaboration and Gender Equity in Game-Based Learning for Middle School Computer Science. *Computing in Science & Engineering* 18, 2: 18–28.

[24] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, 77–91.

[25] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1: 2053951715622512.

[26] Marine Carpuat and Ramani Duraiswani. Introduction to Machine Learning CMSC422. Retrieved from http://www.cs.umd.edu/class/spring2017/cmsc422//schedule0101/

[27] Daniel Smilkov and Shan Carter. Tensorflow — Neural Network Playground. Retrieved January 7, 2020 from http://playground.tensorflow.org

[28] Arjun Chandrasekaran, Viraj Prabhu, Deshraj Yadav, Prithvijit Chattopadhyay, and Devi Parikh. 2018. Do explanations make VQA models more predictable to a human? *arXiv preprint arXiv:1810.12366*.

[29] Arjun Chandrasekaran, Deshraj Yadav, Prithvijit Chattopadhyay, Viraj Prabhu, and Devi Parikh. 2017. It takes two to tango: Towards theory of AI's mind. In *ChaLearn Looking at People Workshop, (CVPR)*.

[30] Moses Charikar. 2019. CS221: Artificial Intelligence: Principles and Techniques. Retrieved from http://web.stanford.edu/class/cs221/

[31] John William Charnley, Alison Pease, and Simon Colton. 2012. On the Notion of Framing in Computational Creativity. In *ICCC*, 77–81.

[32] Alexandra Chouldechova and Max G'Sell. 2017. Fairer and more accurate, but for whom? In *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2017)*.

[33] Kate Crawford. 2016. Artificial intelligence's white guy problem. *The New York Times* 25.
[34] Kate Crawford and Trevor Paglen. 2019. *Training Humans*. Retrieved from http://www.fondazioneprada.org/project/training-humans/?lang=en
[35] Thomas G Dietterich and Eric Horvitz. 2015. Rise of concerns about AI: Reflections and directions. *Commun. ACM* 58, 10: 38–40.
[36] Catherine D'Ignazio. 2017. Creative data literacy. *Information Design Journal* 23, 1: 6–18.
[37] Betsy DiSalvo, Mark Guzdial, Amy Bruckman, and Tom McKlin. 2014. Saving Face While Geeking Out: Video Game Testing as a Justification for Learning Computer Science. *The Journal of the Learning Sciences* 23, 3: 272–315.
[38] Betsy DiSalvo, Jason Yip, Elizabeth Bonsignore, and Carl DiSalvo. 2017. Participatory Design for Learning. In *Participatory Design for Learning*. Routledge, 15–18.
[39] Carl F DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, 321–326.
[40] Andrea A DiSessa. 2001. *Changing minds: Computers, learning, and literacy*. Mit Press.
[41] Zachary Dodds, Christine Alvarado, and Sara Owsley Sood. 2008. Making Research Tools Accessible for All AI Students. In *Proceedings of the AAAI 2008 AI Education Colloquium*.
[42] Zachary Dodds, Lloyd Greenwald, Ayanna Howard, Sheila Tejada, and Jerry Weinberg. 2006. Components, curriculum, and community: Robots and robotics in undergraduate AI education. *AI magazine* 27, 1: 11–11.
[43] Stefania Druga. 2018. Growing Up With AI Cognimates: from coding to teaching machines. MIT.
[44] Stefania Druga. Cognimates. Retrieved January 7, 2020 from http://cognimates.me/home/
[45] Stefania Druga, Sarah T.Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world.
[46] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. Hey Google is it OK if I eat you?: Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children*, 595–600.
[47] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys?: children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 231–240.
[48] Eric Eaton. 2008. Gridworld search and rescue: A project framework for a course in artificial intelligence. In *In the Proceedings of the AAAI-08 AI Education Colloquium, Chicago, IL*.
[49] Ron Eglash, Audrey Bennett, Casey O'donnell, Sybillyn Jennings, and Margaret Cintorino. 2006. Culturally Situated Design Tools: Ethnocomputing from Field Site to Classroom. *American Anthropologist* 108, 2: 347–362. https://doi.org/10.1525/aa.2006.108.2.347
[50] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark Riedl. 2019. Automated Rationale Generation: A Technique for Explainable AI and its Effects on Human Perceptions. *arXiv preprint arXiv:1901.03729*.
[51] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2017. *Decision making with limited feedback: Error bounds for recidivism prediction and predictive policing*.
[52] Ziv Epstein, Blakeley H Payne, Judy Hanwen Shen, Casey Jisoo Hong, Bjarke Felbo, Abhimanyu Dubey, Matthew Groh, Nick Obradovich, Manuel Cebrian, and Iyad Rahwan. 2018. TuringBox: An experimental platform for the evaluation of AI systems. In *IJCAI 2018*, 5826–5828.
[53] Barbara J Ericson, Lauren E Margulieux, and Jochen Rick. 2017. Solving parsons problems versus fixing and writing code. In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*, 20–29.
[54] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 153–162.
[55] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, A Elazari, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300724
[56] Executive Office of the President. 2014. *Big data: Seizing opportunities, preserving values*. The White House Washington, DC.
[57] Ethan Fast and Eric Horvitz. 2017. Long-term trends in the public perception of artificial intelligence. In *Thirty-First AAAI Conference on Artificial Intelligence*.
[58] Robert B Fisher. 2001. AI and Cinema-Does artificial insanity rule? In *Twelfth Irish Conference on Artificial Intelligence and Cognitive Science*.
[59] Jodi Forlizzi and Carl DiSalvo. 2006. Service robots in the domestic environment: a study of the Roomba vacuum in the home. In *Proceedings of the 1st ACM*

SIGCHI/SIGART conference on Human-robot interaction, 258–265.

[60] Natalie Anne Freed. 2012. "This is the fluffy robot that only speaks french": language use between preschoolers, their families, and a social robot while sharing virtual toys. Massachusetts Institute of Technology.

[61] Paulo Freire. 1972. *Pedagogy of the oppressed*. Herder and Herder, New York.

[62] Helen L Gallagher and Christopher D Frith. 2003. Functional imaging of 'theory of mind.' *Trends in cognitive sciences* 7, 2: 77–83.

[63] Ashok Goel. 2014. Georgia Tech CS 6795: Introduction to Cognitive Science. Retrieved from https://www.cc.gatech.edu/classes/AY2017/cs6795_spring/description.html

[64] Ashok Goel and Jim Davies. 2011. Artificial Intelligence. In *Cambridge Handbook of Intelligence* (3rd ed.), R.J. Sternberg and S.B. Kaufman (eds.). Cambridge University Press, New York, 468–484. https://doi.org/10.1145/2063176.2063177

[65] Google. AI Experiments | Experiments with Google. Retrieved January 7, 2020 from https://experiments.withgoogle.com/collection/ai

[66] Michal Gordon, Eileen Rivera, Edith Ackermann, and Cynthia Breazeal. 2015. Designing a relational social robot toolkit for preschool children to explore computational concepts. In *Proceedings of the 14th International Conference on Interaction Design and Children*, 355–358.

[67] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*.

[68] 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.

[69] Clint Heinze, Janet Haase, and Helen Higgins. 2010. An Action Research Report from a MultiYear Approach to Teaching Artificial Intelligence at the K6 Level. In *1st Symposium on Educational Advances in Artificial Intelligence*.

[70] Rafael Lozano Hemmer. 2015. *Level of Confidence*. Retrieved from http://www.lozano-hemmer.com/artworks/level_of_confidence.php

[71] 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*, 415.

[72] Vasant Honavar. 2007. Principles of Artificial Intelligence: Syllabus. Retrieved from http://web.cs.iastate.edu/~cs572/syllabus.html

[73] HubSpot Research. 2016. *Artificial Intelligence Is Here - People Just Don't Realize It*. HubSpot Research. Retrieved September 6, 2019 from https://blog.hubspot.com/news-trends/artificial-intelligence-is-here

[74] Susan P Imberman. 2008. Making nifty assignments niftier and not so nifty assignments nifty with online technologies.

[75] Ipsos. 2017. *Revolution@Work: Fears and Expectations*. Ipsos. Retrieved September 6, 2019 from https://www.ipsos.com/en/revolutionwork-fears-and-expectations

[76] Sooyeon Jeong, Deirdre E Logan, Matthew S Goodwin, Suzanne Graca, Brianna O'Connell, Honey Goodenough, Laurel Anderson, Nicole Stenquist, Katie Fitzpatrick, Miriam Zisook, and others. 2015. A social robot to mitigate stress, anxiety, and pain in hospital pediatric care. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 103–104.

[77] Jeffrey Johnson. 2003. Children, robotics, and education. *Artificial Life and Robotics* 7, 1–2: 16–21.

[78] David Joyner. 2015. CS7637: Knowledge-Based AI: Cognitive Systems Summer 2015 Syllabus. Retrieved from https://docs.google.com/document/d/1qV7SSWhpjeMhCZyyYQJXmW5YYSOXzhF_pyRClDEwI4k/edit

[79] Ken Kahn and Niall Winters. 2017. Child-friendly programming interfaces to AI cloud services. In *European Conference on Technology Enhanced Learning*, 566–570.

[80] Peter H Kahn, Batya Friedman, Deanne R Perez-Granados, and Nathan G Freier. 2006. Robotic pets in the lives of preschool children. *Interaction Studies* 7, 3: 405–436.

[81] Frank Klassner. 2002. A case study of LEGO Mindstorms' suitability for artificial intelligence and robotics courses at the college level. In *ACM SIGCSE Bulletin*, 8–12.

[82] Ross Knepper. 2019. Foundations of Robotics. Retrieved from http://www.cs.cornell.edu/courses/cs4750/2019fa/

[83] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces*, 126–137.

[84] Deepak Kumar and Lisa Meeden. 1998. A robot laboratory for teaching artificial intelligence. *ACM SIGCSE Bulletin* 30, 1: 341–344.

[85] Ray Kurzweil. 2005. *The singularity is near: When humans transcend biology*. Penguin.

[86] Rüdiger C Laugksch. 2000. Scientific literacy: A conceptual overview. *Science education* 84, 1: 71–94.

[87] Yu Liang. 2014. CPSC 4430 Introduction to Machine Learning.

[88] Conor Linehan, Ben J Kirman, Stuart Reeves, Mark A Blythe, Joshua G Tanenbaum, Audrey Desjardins, and Ron Wakkary. 2014. Alternate endings: using fiction to explore design futures. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 45–48.

[89] Tengyu Ma and Christopher Re. 2019. CS 229 Machine Learning Syllabus and Course Schedule. Retrieved from http://cs229.stanford.edu/syllabus.html

[90] Bruce MacLennan. 2017. Introduction to Machine Learning. Retrieved from http://web.eecs.utk.edu/~bmaclenn/Classes/425-528-F17/

[91] Brian Magerko, Jason Freeman, Tom McKlin, Mike Reilly, Elise Livingston, Scott McCoid, and Andrea Crews-Brown. 2016. EarSketch: A STEAM-Based Approach for Underrepresented Populations in High School Computer Science Education. *ACM Transactions on Computing Education (TOCE)* 16, 4: 14.

[92] Arthur B Markman. 1999. *Knowledge Representation*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA.

[93] James B Marshall. 2008. Leveraging the Singularity: Introducing AI to Liberal Arts Students. *Association for the Advancement of Artificial Intelligence*.

[94] Maja J Mataric. 2004. Robotics education for all ages. In *Proc. AAAI Spring Symposium on Accessible, Hands-on AI and Robotics Education*.

[95] Michael Mateas. 2001. Expressive AI: A hybrid art and science practice. *Leonardo* 34, 2: 147–153.

[96] Amy McGovern, Zachery Tidwell, and Derek Rushing. 2011. Teaching introductory artificial intelligence through java-based games. In *Second AAAI Symposium on Educational Advances in Artificial Intelligence*.

[97] Keith W Miller. 2010. It's not nice to fool humans. *IT professional* 12, 1: 51–52.

[98] Raymond Mooney. 2007. Course Syllabus for CS 391L: Machine Learning. Retrieved from https://www.cs.utexas.edu/~mooney/cs391L/syllabus.html

[99] Gina Neff and Peter Nagy. 2016. Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay. *International Journal of Communication* 10: 17.

[100] Nils J Nilsson. 2009. *The quest for artificial intelligence*. Cambridge University Press.

[101] Matthew C Nisbet, Dietram A Scheufele, James Shanahan, Patricia Moy, Dominique Brossard, and Bruce V Lewenstein. 2002. Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology. *Communication Research* 29, 5: 584–608.

[102] Marina Papastergiou. 2008. Are computer science and information technology still masculine fields? High school students' perceptions and career choices. *Computers & education* 51, 2: 594–608.

[103] Seymour Papert. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc., New York, NY, USA.

[104] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

[105] Simon Parsons and Elizabeth Sklar. 2004. Teaching AI using LEGO mindstorms. In *AAAI Spring Symposium*.

[106] Pega. 2018. *What Consumers Really Think About AI: A Global Study*. Pega. Retrieved September 6, 2019 from https://www.pega.com/ai-survey

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

[108] Michael J Quinn. 2010. *Ethics for the information age*. Addison-Wesley Publishing Company.

[109] Mitchel Resnick, Robbie Berg, and Michael Eisenberg. 2000. Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences* 9, 1: 7–30.

[110] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, and Kafai, Yasmin. 2009. Scratch: Programming for All. *Communications of the ACM* 52, 11: 60–67. https://doi.org/10.1145/1592761.1592779

[111] Mark O Riedl. 2019. Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies* 1, 1: 33–36.

[112] Mary Rigdon. 2017. Introduction to Cognitive Science. Retrieved from https://ruccs.rutgers.edu/academics/undergraduate/syllabi/1-201-intro-syllabus/file

[113] Alexander Rush. Syllabus for CS 182 Artificial Intelligence. Retrieved from http://people.seas.harvard.edu/~srush/syllabus.pdf

[114] Ingrid Russell, Zdravko Markov, and Todd Neller. 2006. Teaching AI through machine learning projects. In *ACM SIGCSE Bulletin*, 323–323.

[115] Stuart J Russell and Peter Norvig. 2016. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.

[116] Roger C Schank. 1987. What is AI, anyway? *AI magazine* 8, 4: 59–59.

[117] R Benjamin Shapiro, Rebecca Fiebrink, and Peter Norvig. 2018. How machine learning impacts the undergraduate computing curriculum. *Communications of the ACM* 61, 11: 27–29.

[118] Judy Hanwen Shen, Lauren Fratamico, Iyad Rahwan, and Alexander M Rush. Darling or Babygirl? Investigating Stylistic Bias in Sentiment Analysis.

[119] Matthew Smith, Christian Szongott, Benjamin Henne, and Gabriele Von Voigt. 2012. Big data privacy issues in public social media. In *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, 1–6.

[120] Sara Owsley Sood. 2008. Emotional computation in artificial intelligence education. In *AAAI Artificial Intelligence Education Colloquium*, 74–78.

[121] Sucheta Soundarajan and Daniel L Clausen. Equal Protection Under the Algorithm: A Legal-Inspired Framework for Identifying Discrimination in Machine Learning.

[122] Stanford University. 2018. *Artificial Intelligence Index 2018 Report*. Retrieved from https://aiindex.org

[123] Stanford University. Machine Learning Syllabus. Retrieved from https://www.coursera.org/learn/machine-learning

[124] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, and others. 2016. Artificial intelligence and life in 2030: One hundred year study on artificial intelligence. *Report of the 2015 Study Panel, tech report*.

[125] 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.

[126] Yunjia Sun. 2016. Novice-Centric Visualizations for Machine Learning. University of Waterloo.

[127] Fumihide Tanaka, Aaron Cicourel, and Javier R Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences* 104, 46: 17954–17958.

[128] Dan Tappan. 2008. A pedagogical framework for modeling and simulating intelligent agents and control systems. *Technical Report WS-08-02*.

[129] Judith Jarvis Thomson. 1984. The trolley problem. *Yale LJ* 94: 1395.

[130] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the 2019 Conference on Artificial Intelligence*.

[131] David S Touretzky. 2012. Seven big ideas in robotics, and how to teach them. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*, 39–44.

[132] David S Touretzky. 2017. Computational thinking and mental models: From kodu to calypso. In *Blocks and Beyond Workshop (B&B), 2017 IEEE*, 71–78.

[133] Sherry Turkle. 2005. *The second self: Computers and the human spirit*. Mit Press.

[134] Berk Ustun, Alexander Spangher, and Yang Liu. 2019. Actionable recourse in linear classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 10–19.

[135] Mike Van Duuren, Barbara Dossett, and Dawn Robinson. 1998. Gauging children's understanding of artificially intelligent objects: a presentation of "counterfactuals." *International Journal of Behavioral Development* 22, 4: 871–889.

[136] Scott A Wallace, Ingrid Russell, and Zdravko Markov. 2008. Integrating games and machine learning in the undergraduate computer science classroom. In *Proceedings of the 3rd international conference on Game development in computer science education*, 56–60.

[137] Noah Wardrip-Fruin. 2007. Three Play Effects–Eliza, Tale-Spin, and Sim City. *Digital Humanities*: 1–2.

[138] Weber Shandwick, Inc. 2016. *AI-Ready or Not: Artificial Intelligence Here We Come!* Weber Shandwick, Inc. Retrieved September 6, 2019 from https://www.webershandwick.com/news/ai-ready-or-not-artificial-intelligence-here-we-come/

[139] Henry M Wellman, David Cross, and Julanne Watson. 2001. Meta-analysis of theory-of-mind development: The truth about false belief. *Child development* 72, 3: 655–684.

[140] Linda L. Werner, Brian Hanks, and Charlie McDowell. 2004. Pair-programming helps female computer science students. *Journal on Educational Resources in Computing (JERIC)* 4, 1: 4.

[141] Randi Williams, Christian Vázquez Machado, Stefania Druga, Cynthia Breazeal, and Pattie Maes. 2018. "My Doll Says It's Ok": A Study of Children's Conformity to a Talking Doll. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (IDC '18), 625–631. https://doi.org/10.1145/3202185.3210788

[142] Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. 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*, 447.

[143] Jason Shun Wong. 2018. Design and fiction: imagining civic AI. *Interactions* 25, 6: 42–45.

[144] Baobao Zhang and Allan Dafoe. 2019. *Artificial Intelligence: American Attitudes and Trends*. University of Oxford. Retrieved September 6, 2019 from https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/high-level-machine-intelligence.html#subsecharmgood

[145] Michelle Zimmerman. 2018. *Teaching AI: Exploring New Frontiers for Learning*. International Society for Technology in Education.

[146] 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.

[147] Mike Zyda and Sven Koenig. 2008. Teaching artificial intelligence playfully. In *Proceedings of the AAAI-08 Education Colloquium*, 90–95.

[148] ReadyAI | Empowering all students to improve our world with AI. *ReadyAI*. Retrieved January 7, 2020 from https://www.readyai.org/

[149] Machine Learning for Kids. Retrieved January 7, 2020 from https://machinelearningforkids.co.uk

[150] GAN Lab: Play with Generative Adversarial Networks in Your Browser! Retrieved January 7, 2020 from https://poloclub.github.io/ganlab/

[151] eCraft2Learn - Digital Fabrication and Maker Movement in Education Making Computer supported Artefacts from Scratch. *eCraft2Learn*. Retrieved January 7, 2020 from https://project.ecraft2learn.eu/