Artificial Intelligence for Future Presidents: Teaching AI Literacy to Everyone

Kate Candon, Nicholas C. Georgiou, Rebecca Ramnauth, Jessie Cheung, E. Chandra Fincke, Brian Scassellati

Yale University, New Haven, CT 06511 USA

{kate.candon, nicholas.georgiou, rebecca.ramnauth, jessie.cheung, chandra.fincke, brian.scassellati}@yale.edu

Abstract

The rapid and nearly pervasive impact of artificial intelligence on fields as diverse as medicine, law, banking, and the arts has made many students who would never enroll in a computer science class become interested in understanding elements of artificial intelligence. Fueled by questions about how this technology would change their own fields, these students are not seeking to become experts in building AI systems but instead are searching for a sufficient understanding to be safe, effective, and informed users. In this paper, we describe a first-of-its-kind course offering, "Artificial Intelligence for Future Presidents" designed and taught during the spring of 2024. We share rationale on the design and structure of the course, consider how best to convey complex technical information to students without the background in programming or mathematics, and consider methods for supporting an understanding of the limits of this technology.

Introduction

Artificial intelligence (AI) has become a pervasive technology that touches an ever-increasing roster of disciplines and professions. It has reshaped how doctors perform diagnoses and keep patient records, how artists create music or imagery, how lawyers seek insights from case law, and how countless other fields conduct their day-to-day functions. This rapid integration means that a much wider range of students with diverse backgrounds are now seeking a basic understanding of AI-based technologies. Fundamentally, our challenge is how to provide these students with clear, accurate information about AI without relying upon the traditional prerequisites of data structures, discrete mathematics, advanced programming, and a strong foundation in mathematics and data science. How do we teach about AI without these skills?

Many instructors in these diverse fields are attempting to resolve this challenge on their own by integrating AI topics and tools into their curricula (Ng et al. 2021). Courses based on using AI technologies have begun to appear across the wide range of disciplines. At our own institution, for example, dedicated AI courses are now offered in the medical, law, and business schools, as well as in several undergraduate programs outside of computer science and data science. However, many of these instructors have neither a background in AI nor a strong understanding of the technical material. While it is unreasonable to expect that these courses would only be taught by AI professionals, preparing students for the rapidly changing landscape of AI-based tools and giving them the ability to understand new tools as they arise is a challenge for both students and instructors.

At the same time, computer science faculty have struggled with both the challenges associated with teaching technical content during the recent pandemic (Birk et al. 2020) and the rapidly increasing enrollments in classes on AI, machine learning, and applications like robotics (National Academies 2018). As a result, there has been limited focus within the AI community on developing courses for non-technical students. Most existing AI courses are designed for aspiring engineers or data scientists, leaving few, if any, that are accessible to a broader audience. This has created a gap in educating non-technical students about both the potential and the realities of AI, as well as addressing pressing questions about AI's impact on their chosen careers and future jobs.

In this paper, we report on the design and implementation of a class focused on bringing reliable and accurate information about AI to a broad audience of students without a background in mathematics, programming, or data science. Our target students are undergraduates, professional school students, and graduate students outside of traditional STEM areas. In short, we attempt to teach future poets, politicians, and historians how to reason about the impact that AI will have on their future.

Our goal is not to teach these students how to design or build new AI-based systems, nor to entice them into becoming computer scientists. Instead, we aim to provide students with a technical understanding through analogies and comparisons, allowing them to grasp how these systems function without delving into complex technical details. This approach equips students to be informed consumers of AI technologies, enabling them to make knowledgeable decisions about what to expect from these systems in the future. In explaining this goal to both students and administrators, we often used the analogy of getting a driver's license. Our goal here was not to teach students how to build a car, but to know enough to be a safe driver– to have an intuition about what the technology can and cannot do and to not be scammed into buying a lemon.

One of the inspirations for this course was the "Physics for Future Presidents" course offered at UC Berkeley, which also served as the basis for Richard A. Muller's book of the same name (2010). In addition to the obvious change in topic coverage, our approach differs from the Berkeley course in three important ways. First, we use a single instructor model rather than inviting multiple faculty to cover multiple topics. This allows us to have a more consistent series of topics that build upon each other. Second, while the Berkeley course seeks to "eliminate policy and politics" by adhering strictly to the science and technology, we attempt to engage students on how AI technology interacts with current policy and discuss matters such as governance and regulation, impacts on electoral processes, and the changing landscapes of ethical use of AI, while still maintaining a balanced discussion around these topics. Finally, while the Berkeley course does presume some basic knowledge of algebra, statistics, and probability, we opted to try to minimize the mathematics within this course. This, of course, prohibits us from covering some critical content, but it also forces us to focus not on teaching the details of how to *build* these systems but rather how to *use and understand* these technologies broadly. This also serves to make the course more attractive to students who feel daunted by the advanced technical nature of the topic, as the promise of "no math and no programming" was very effective in recruiting interested students.

Guiding Principles

Our goal was to create a course that would be broadly accessible to undergraduate, graduate, and professional school students, providing an introduction to key concepts in artificial intelligence. To achieve this, we established five guiding principles to shape the course's design and content.

No prerequisites. Our first principle was that the course needed to be accessible by any student with a general high school education. We relied on students understanding no more than simple algebraic equations and that they would have neither programming experience nor a background in computational thinking. In our discussions with many students who eventually enrolled in the first offering of this course, the mantra of "no math, no programming" was a critical factor in their decision to enroll.

Literate users, not amateur programmers. Rather than training our students to create AI-based tools, we focused on developing their intuition about the capabilities, limits, and trade-offs involved in making decisions about AI-based systems. Upholding this principle throughout the semester was particularly challenging due to the technical backgrounds of our teaching staff, who were more familiar with traditional AI courses centered on algorithms and technical processes. There was a natural tendency to revert to these established approaches. However, we tried to prioritize a broader, more accessible understanding of AI, ensuring that the course content was suitable for non-technical students while still maintaining academic rigor.

Empower users to be critical consumers. As part of developing their literacy of this complex subject, we focused on instilling a healthy skepticism in our students when presented with AI achievements. Given the tendency for media reports to (both deliberately and accidentally) embellish the capabilities of these systems, we felt that students needed to be prepared to question the information being presented to them and to "sanity check" the media stories that they read.

Provide a balanced viewpoint. Although it was tempting to emphasize the positive aspects of AI technology, we aimed to present a balanced view that included both its benefits and its proven and potential risks. While students might have perceived the course staff as having a pro-technology bias—given our backgrounds in computer science and AI research—we made a concerted effort to address and discuss even the most critical and alarmist viewpoints, such as concerns about job displacement by robots or the potential for AI to pose existential threats.

Encourage interdisciplinary relevance. A key principle was to ensure that the course content would resonate with students from a wide range of academic disciplines, not just those with a direct interest in AI or computer science. We designed the course to highlight how AI intersects with fields such as healthcare, law, education, and the arts, encouraging students to critically explore the relevance of AI in their respective areas of study. By connecting AI concepts to real-world applications, we aimed to make the material applicable and engaging for students with diverse professional and academic goals.

While it is challenging to determine the extent to which we were successful in maintaining these principles, we will evaluate our success based on how well students saw the course as maintaining each of these ideals.

Course Structure

In designing the structure of the course, our hope was to allow as much informed discussion as possible. Yet, we realized in our early planning that students would need to have some direct instruction about the capabilities, mechanisms, and especially the limits of AI technologies before being able to engage in these discussions. We also felt that there was value in maintaining a course structure that students were familiar with—that to make artificial intelligence a topic so uniquely distinct from any other course material would be to do a disservice to our attempts to make the topic accessible to everyone.

To that end, we settled on a schedule where lectures were held twice a week, on Mondays and Wednesdays, with 1.5-hour sessions for the entire class. Fridays were reserved for smaller discussion sections, each lasting 1 hour and capped at 15 students. To prepare for these discussions, assignments were due on Thursdays, allowing teaching staff to review student responses and tailor the discussion topics accordingly based on the insights and opinions reflected in their submissions. This structure balanced direct instruction with in-depth, personalized engagement in discussions.

Lectures

The first 15 minutes of each lecture were dedicated to the discussion of a headline about AI from the past one to three days. In many instances, this resulted in coverage of a topic that we had not yet discussed in lecture or discussion sections. The decision to keep to immediately relevant news stories rather than following the course syllabus did at times present challenges as students did not have the relevant information to fully consider these topics. However, this decision also provided two distinct benefits. First, students became much more cognizant of the appearance of AI stories within news sources. After only a few weeks of class, students often began suggesting news stories or asking course staff about an AI-based article that they had read. Second, by working at times with material that they had not yet had classroom-based instruction, students were challenged to practice their careful consideration of information from the media, exercising their skepticism and critical analysis.

Each current events discussion ended with an online poll, using Poll Everywhere[1], which asked students to either write a short open-ended answer or select an option from a set of possible responses. This exercise served three purposes. First, it generated a metric for measuring and tracking individual student attendance over the course of the semester. Second, it allowed course staff to gauge roughly the disposition of the class toward controversial topics so that future discussions could be more finely tuned. Third, as both short answers and multiple-choice selections were shown anonymously during lecture, students saw the distribution of their classmates' opinions and began to understand that even on issues that they felt were straightforward that there might be dissenting opinions.

Lecture topics covered a wide range of material as listed in Table 1 and interwove both technical information with discussions of the ethical, legal, and practical implications of these technologies. Roughly, we divided the semester into four categories of topics: Foundational, Learning, Issues, and Applications.

#	Type	Topic
1	Foundations	Introduction and Course Structure
2	Foundations	10 Things about AI (part 1)
3	Foundations	10 Things about AI (part 2)
4	Foundations	Agents and Environments
5	Foundations	Measuring Intelligence
6	Foundations	Generating Complexity
7	Learning	Machine Learning Basics
8	Learning	Scaling to Neural Networks
9	Learning	Deep Learning Successes
10	Issues	Bias and Failures of ML
11	Issues	Explainability
12	Learning	Large Language Models
13	Applications	Natural Language Processing
14	Applications	Robotics
15	Issues	AI and Job Loss
16	Applications	Human-Robot Interaction
17	Issues	Social Influence
18	Issues	Deep Fakes and Identity
19	Applications	Artificial General Intelligence
20	Applications	Autonomous Vehicles
21	Applications	AI in Medicine
22	Issues	Ethics of AI
23	Issues	Policy Setting for AI
24	Issues	The Future of AI

Table 1: Lecture topics.

Foundations Lectures. The first six topics, spanning three weeks, established some common expectations about what artificial intelligence was, how certain magical-seeming capabilities were generated, and on instilling a sense of what AI could and could not do easily today. The first segment consisted of a two-part lecture series, titled "10 Things about AI," which introduced ten essential concepts that would underpin our discussions throughout the semester:

1. AI is hard to define precisely.
2. Success is judged against science fiction, not science.
3. People build AI for many reasons.
4. AI is an umbrella, not a single community.
5. AI is a moving target.
6. It is easy to over-estimate AI's capabilities.
7. Success is built on decades of work.
8. AI's impact is difficult to predict as it crosses silos.
9. AI is very hot now, but winter will come.
10. AI highlights the ceiling, not the floor.

The remaining lectures in this category focused on why AI systems typically are fragile and cannot easily be combined, how to measure the success of AI, and why it is often easy to generate complex behavior but difficult to view an AI system and interpret its inner structure.

Learning Lectures. Given the immense interest in large language models, generative AI, and other machine learning applications, we scheduled four lectures on machine learning early in the semester. This ensured that students would have a solid foundation in these critical technologies before tackling subsequent topics that depend heavily on them.

These lectures were designed to give students a functional understanding of the capabilities and limits of these technologies without delving into how to technically design and construct a learning system independently. We covered how a machine might "learn" by changing the value of information that it stored, using a single perceptron as a model for how this operation might take place. Students were given a sandbox-style graphical interface to a network of neuron simulations and allowed to see how classification outcomes changed, and were challenged to imagine scaling this system larger and larger to obtain the capabilities of deep networks. Finally, using an abstract functional model of transformers and convolutional networks, students were introduced to some of the remarkable successes of state-of-the-art machine learning models.

Issues Lectures. Although all lectures, even the most technical ones, integrated discussions of societal issues related to the technologies, some topics were organized primarily around specific issues, such as job loss, explainability and transparency, or deep fakes. These issues often spanned mul-

[1]www.polleverywhere.com. Accessed: 2024-09-16.

tiple application domains and involved a variety of technologies. In addition to anchoring these issues in the reality of current AI technologies to help students form realistic expectations about what AI can and might do in the near future, we presented a range of viewpoints on the costs and benefits of these technologies.

Applications Lectures. Applications lectures focused on particular domains where AI-based technologies were currently having a significant impact. From autonomous cars to surgical robots, these topics tended to highlight currently deployed technologies and the near-future extensions of these technologies. For each, we were careful to show that these technologies did not appear magically overnight, even when it might seem that way to the public. Many of these topics were chosen because they aligned with students' interests and also highlighted key issues we intended to explore in the course.

Discussion Sections

In our first semester offering this course, we staffed five discussion sections, each of which contained between 10 and 15 students. All sections were offered on the same day of the week due to university scheduling constraints, which was a challenge for some students but did allow our staff to reliably plan section coverage relative to the lectures and assignments in a consistent manner. Students were offered a choice of section based on the available times. While students were not grouped by experience or by academic interests, this method of scheduling did occasionally create sections that were predominantly representing students who had similar class conflicts. For example, one of our sections was predominantly attended by MBA candidates from the business school. This grouping was unfortunate, as discussions greatly improved when there were a variety of student backgrounds represented within the section.

Each week, course staff met to determine a set of appropriate activities for the discussion section based on the topic coverage of the preceding week. This resulted in a variety of individual session activities, ranging from open discussions to structured debates. For example, after covering the topic of "AI's overestimated capabilities," students engaged in a debate on the real versus perceived impacts of AI in areas such as healthcare and finance, discussing whether AI's actual achievements in these fields matched the hype portrayed in media reports. Following a lecture on "the fragility of AI systems," students participated in a hands-on workshop analyzing case studies of high-profile AI failures, such as the malfunction of autonomous vehicles or biases in facial recognition system, to understand what went wrong, the extent of failure impact, and how these issues could be mitigated. Discussion sections often featured interactive activities. For example, when discussing the more technical details of clean data and supervised learning, students interacted with a simple tool that used a pre-trained neural network to recognize their own handwriting. This allowed students to experiment with the effects of machine learning without any programming experience. These activities were designed to reinforce theoretical concepts, encourage critical thinking, and deepen students' understanding of AI's practical implications and limitations.

Assignments and Grading

We created two different kinds of assignments: problem sets and essays. Unlike traditional engineering courses, our "problem sets" did not ask students to exercise significant problem-solving skills nor did they ask students to build solutions to given problems. Instead, problem sets offered step-by-step walkthroughs for a specific technology and were intended to give students the feeling of working with these technologies directly without asking them to implement novel solutions. Grading for these problem sets were generally an all-or-nothing grade that indicated whether or not the student made a significant effort to complete the walkthrough. In contrast, essays were designed as an opportunity for students to demonstrate their mastery of the material covered in class and were graded in detail based on the quality of the essay content.

Final grades were determined by contributions from the aggregate essay grade (45%), problem set grade (40%), and participation in sections (10%) and lecture (5%). In general, the problem set and lecture participation grades indicated student engagement in the class activities while essay and section participation grades were meant to indicate the quality of the work that the student produced. This method was explained in detail at the start of the semester in order to encourage students to stay involved in class activities.

Problem Sets

Students completed six problem sets in total, each of which focused on giving students an exposure to a particular activity or technology. Most problem sets were designed to take between two and three hours to complete, though some students did become significantly more invested in playing with these models and often spent significantly longer exploring the technologies than was required.

PS #1: Making AI Predictions - The first problem set challenged students to consider their own biases about AI technologies and to critically evaluate the nature of predictions about this technology. We asked students to read a selection of articles that discussed the challenges and common pitfalls in making predictions about AI technologies, including the essay "The Seven Deadly Sins of Predicting the Future of AI" by Rodney Brooks (2017). We then asked students to make a series of predictions about when certain technologies that we would cover later in the course would happen, such as "an AI system will create a pop song that ranks in the top 10 on the Billboard charts" and "a major US city replaces all human-driven taxi services with autonomous cars". Finally, we asked students to consider whether or not they might have fallen victim to one of the "deadly sins" that Brooks spoke of, whether it be misunderstanding exponential growth or relying upon "magical thinking".

PS #2: Generating Complex Behavior - This problem set both demonstrated how easy it was to construct systems that

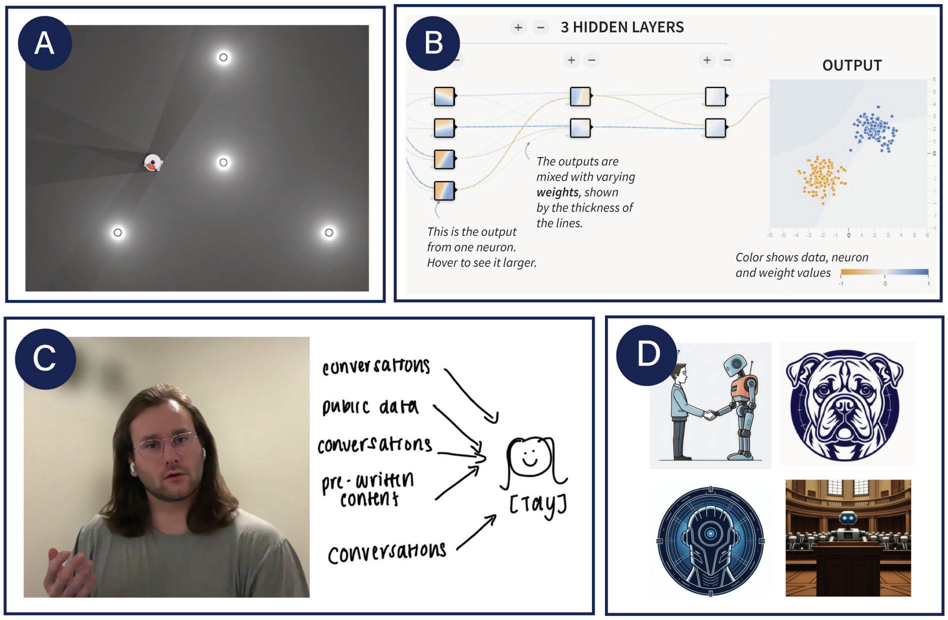

Figure 1: Problem set assignments include (A) a Braitenberg vehicle simulator (de Weerd 2016) used in problem set #2, (B) the Tensorflow sandbox (Smilkov and Carter 2022) used to demonstrate learning in neural networks in problem set #3, (C) video submissions created by a student to explain conversational agents as part of problem set #5, and (D) submissions generated using a diffusion-based image generator when asked to create a logo for this course as part of problem set #4.

generated rich, uncontrolled behavior and how difficult it was to guess the internal structure of an agent by observing its behavior alone. Much of this problem set focused on the elegant examples of agent construction developed by Valentino Braitenberg (Braitenberg 1986). Students were familiarized with a simple 2-wheel, 2-sensor virtual agent that moved about on a flat 2-dimensional plane. By changing connections between sensors and actuators, students saw how very different behavior could emerge. Using a web-based simulator for these Braitenberg Vehicles (de Weerd 2016), students saw that even these simple vehicles might exhibit complex, context-sensitive behavior that was rich and nuanced (see Figure 1-A). However, when asked to reverse this process and identify the inner structure that a mystery vehicle used by observing only the path that the vehicle took, students saw that this analytic approach could be extremely challenging. Finally, students were asked to draw parallels between the complexity of these simple vehicles and the complexity of AI systems that they had seen in recent news stories. Our intent with this assignment was to give students both a taste of the excitement of creation (when the simple vehicles they constructed did something interesting) and the challenge of analyzing the internal structure of an unknown vehicle.

PS #3: Neural Network Sandbox - While we could not reasonably expect students to construct machine learning systems on their own, we did want to give them the experience of working with a small-scale learning system. Using a web-based sandbox developed by Tensorflow (Smilkov and Carter 2022), students were given very detailed, step-by-step instructions on how to set up a small multi-layer neural network, to simulate the activity of that network with a pre-defined data set, and to stop and examine the network to gain some understanding of what happened as different network parameters were altered (see Figure 1-B). While understanding this particular network was not important to finishing the assignment, our intention was to give students a sense of the fragility of these networks and an appreciation for how certain kinds of changes might impact what the network could learn. For example, changing the number of hidden nodes or altering the kind of threshold that the network required to activate would frequently (but not always!) change the learned function. We were sure to provide checkpoints such that students could easily revert to a specific network state without having to restart the assignment from the beginning. We found that having experience with a network at this scale (generally no more than a dozen nodes in 2-4 layers) gave students a rapid appreciation for the scale of more complex deep networks.

PS #4: Large Language Models and Generative AI - In this problem set, we gave students access to three different systems created using foundation models. Our goals were (1) to give students some experience with the capabilities of these systems, (2) to allow students to explore some of the failure conditions of these models, and (3) to provide a basic understanding of what makes these three models different. Using a set of models tuned to our needs and hosted by HuggingFace (http://huggingface.co), students explored a large language model (LLM) tuned for information retrieval and were asked to find some prompts that would return unusual or incorrect responses, a second LLM tuned for conversation and asked to differentiate its performance from the first LLM, and to compare both of these systems with the much simpler Eliza program (Weizenbaum 1966). Students were then given a diffusion-based image generator and asked why repeated queries using the same prompt generated different outcomes and were challenged to create a new logo for this course by engineering their prompt. Figure 1-D shows some example submissions by students.

PS #5: Video Explanations - Inspired by the popular Wired[2] web series which asks experts to explain a technical concept in five levels of difficulty, ranging from young children to subject matter experts, this problem set challenged students to explore ways of explaining a single concept in AI to a variety of different audiences. While the original series varied the audience over a range of expertise levels, this would have been an unfair request for our students; they frequently did not understand the technology in sufficient depth to explain it to a scientist from another discipline, much less to a subject matter expert. Instead, we asked students to explain a concept to different people with a range of experiences and background. Students were asked to record four videos, each 2-3 minutes in length, which explained an AI concept to a child, a peer, a parent, and a grandparent. Students were encouraged to focus on specific interests for each of these groups. For example, for the topic of food delivery robots, a child might be interested in how the robot knows where to go; a peer might wonder about the impact on entry-level food service jobs; a parent might be concerned about the security of an order and whether the food might be tampered with; and a grandparent might be concerned with the long-term impact of these robots on traffic and noise in the neighborhood. These videos were an entertaining way to identify a range of perspectives on a single technology (see Figure 1-C for an example).

PS #6: AI Predictions, Revisited - At the end of the semester, students were presented with the predictions that they had made in the first problem set and given the opportunity to update or change each of these predictions. For any prediction that they chose to change, students were asked to cite something that helped to change their mind. This exercise helped highlight how this course challenged them to update their views on AI-based technologies.

Essays

We asked students to write three essays over the course of the semester. Each essay had a different length requirement (2-4 pages, 4-6 pages, and 6-8 pages approximately), a different topic, and an increasingly demanding evaluation rubric. The essay topics were as follows:

Essay #1: Media Critique - Students were given a choice of three different videos, each produced as a media announcement for a new AI-based technology for a major company in the last 3-5 years. All three of these announcements gave an overly-optimistic view of a technology that still was not in general use (although each of the companies insisted that their technology would be ready for release more than a year prior). Students were asked to provide at least three distinct reasons why a viewer (at the time of the video release) should have been skeptical about the technology. Note that these reasons often did not require a deep understanding of the technology itself, but rather focused on flaws that could be inferred based on how the situations presented were unusual, staged, or uniquely positioned. This exercise had been conducted three times in lecture prior to this assignment, so students had multiple exemplars to emulate. Our hope was to very explicitly develop the skills of critical evaluation of claims about AI that a skeptical consumer would need.

Essay #2: Impact on Jobs - Students were asked to select a single occupation (as defined by the US Bureau of Labor Statistics) and a single specific AI-based technology that either already had or almost certainly will have an impact on that profession. Students were cautioned to be specific in the technology, as something general like "machine learning" would be very difficult to consider as it would have many different impacts on most occupations. Essays were required to identify specific expert sources for any opinions or information that they chose to include in the essay. Again, students had seen us conduct this analysis multiple times in lecture, so they had an exemplar of what was expected. Our goal was for students to focus on a single technology-occupation intersection and to think critically about a variety of information that might be available.

Essay #3: Technology Impacts - Students were asked to follow the development of a single AI-based technology and to consider critically the kinds of impact that this technology would have. They were asked to divide this essay into roughly three equal sections: a detailed description of the development of this technology including landmark instances of demonstrations of the technology, a clear picture of the current state-of-the-art of what this technology can produce, and an exploration of how this technology might impact society. Where the second essay considered a single intersection between technology and occupation, the third essay asked students to consider the societal implications broadly and across silos and disciplinary boundaries. Students were evaluated not only on the accuracy of their depiction of the current technology but also on the clarity of their explanation such that it was accessible to everyone.

Implementation in Spring 2024

The first offering of this course initially attracted 75 for-credit students and 3 auditors. Of the 75 non-auditing students, 67 completed the course and were assigned a final grade. This loss rate was similar to that of other courses without pre-requisites at our institution. This group of students was composed primarily of undergraduates, though there were three graduate students, a dozen professional school students (representing the law school, medical

[2] www.wired.com/video/series/5-levels

Evaluation Prompt	Course Avg	Department Avg	Division Avg
What is your overall assessment of this course? 1-poor, 2-fair, 3-good, 4-very good, 5-excellent	4.0	3.6	3.5
The course was well organized to facilitate student learning. 1-strongly disagree, 2-disagree, 3-neutral, 4-agree, 5-strongly agree	4.1	3.8	3.7
I received clear feedback that improved my learning. 1-strongly disagree, 2-disagree, 3-neutral, 4-agree, 5-strongly agree	3.6	3.5	3.6

Table 2: Questions and scores from our formal assessment.

school, and business school), and one staff member. Approximately 10% of the undergraduates were in STEM-related disciplines, and students represented a wide range of majors that included social sciences, fine arts, and humanities.

We conducted both an informal assessment and a formalized assessment that is standard at our institution. Both assessments were anonymous and were collected before final grades were released. The formal assessment was offered only to enrolled students; 54 of the 67 entrolled students submitted the formal assessment. Three questions were particularly relevant to this course. All three were collected as 5-point scores, and the questions, scales, and results are shown in Table 2. For comparison, average scores across all computer science courses and all science and engineering courses are shown. While these comparisons are not the most apt reference points for a course of this nature, they are the best comparison points available. The formal assessment shows that the class was well received and students considered it to be well organized to support their learning. While we sought to provide timely and constructive criticism to students, we were not as successful as we might have hoped. Informally, we polled students over multiple lectures, asked for feedback in discussion sections, and solicited free-form comments from students through a web-based interface (all submitted anonymously). We aggregated these responses and looked for evidence in the student feedback that we had succeeded in our five guiding principles.

No prerequisites: We saw extensive evidence in the course evaluations that students appreciated being able to approach this material without backgrounds in computing or mathematics. For example, students wrote: "The course truly is an introductory course, as topics are clearly explained and don't require prior technical or even simple background knowledge about the subject matter" and "I would 100% recommend this class. It has been one of my favorite courses... Technology evolves at such a rapid pace, and I find it very interesting to think about the industry impacts and ethical implications. Phenomenal learning experience for people of all academic disciplines. Very accessible. You don't need any technical knowledge to succeed."

Literate users, not amateur programmers: When asked in one of the last lectures of the course, "What is the single most important thing that you learned?", one student responded "Learning how to have conversations about AI and have more realistic expectations".

Empower users to be critical consumers: Many students found that they were better prepared to interpret current stories about AI technology. When asked in the formal assessment, "What knowledge, skills, and insights did you develop by taking this course?", one student responded "I think I learned to be skeptical of AI systems. This means I learned about their capabilities and limitations and not to have this idealized image of AI as well as how AI will develop in the future." Another student responded "[This class] enhanced my critical thinking abilities regarding AI in the news and press releases, to better understand how the technologies work, and how to identify what's real and not regarding the capabilities of AI".

Provide a balanced viewpoint: Students were generally surprised to see how their predictions about AI changed between the first and last problem sets. Most students found topics in which they became more optimistic with their predictions while simultaneously becoming more pessimistic with others. These changes were also seen explicitly in some of the free-response questions. When asked in one of the last lectures, "What is the single most important thing that you learned?", one student responded "AI is not magic — don't overestimate it in the short run, nor underestimate it in the long run" and a second student wrote "It's easy to both underestimate and overestimate the future of AI".

Encourage interdisciplinary relevance: Students focused on an extraordinarily broad range of professions and topics in essays where they chose their own topics. For essay #2, students wrote about how AI might affect zoologists and wildlife biologists, wind turbine service technicians, writers and authors, audio editors, actors, warfare, marketing, lawyers, paralegals, human resources (HR), and many more. In essay #3, students touched on the impact of AI technologies that ranged from ethical considerations for autonomy, impacts on financial institutions, tax implications as AI systems overtook certain jobs, environmental impacts of foundation models, wildlife conservation efforts, and changes in law enforcement as a result of AI.

Lessons learned: Our first implementation of the course was generally successful, but we learned three things to update in the future. First, our essay grading needed more explicit instructions and rubrics. As students had widely different backgrounds, they also had different writing skills and expectations. Second, because we chose headlines based on their coverage not their technical content, we often introduced topics in these short discussions before covering them in depth in the main lecture. In the future, we will select news stories that are more closely aligned with the syllabus. Finally, we will make more of an attempt to recruit students from the professional schools (who use slightly different academic calendars).

Acknowledgments

This work was funded in part by the National Science Foundation (NSF) under grants No. 1955653, 1928448, 2106690, and 1813651. The authors are listed first by role and then alphabetically within each category. The graduate teaching assistants are listed first: K. Candon, N. C. Georgiou, and R. Ramnauth. The following are the undergraduate teaching assistants, J. Cheung and E. C. Fincke. The lead instructor was B. Scassellati.

References

Birk, A.; Dineva, E.; Maurelli, F.; and Nabor, A. 2020. A robotics course during covid-19: Lessons learned and best practices for online teaching beyond the pandemic. *Robotics*, 10(1): 5.

Braitenberg, V. 1986. *Vehicles: Experiments in synthetic psychology*. MIT press.

Brooks, R. 2017. The Seven Deadly Sins of AI Predictions. *MIT Technology Review*. Accessed: 2024-09-16.

de Weerd, H. 2016. Braitenberg Vehicles Simulator. Accessed: 2024-09-16.

Muller, R. A. 2010. Physics for future Presidents. *Science and the Educated American: A Core Component of Liberal Education*, 112.

National Academies of Sciences, Engineering, and Medicine. 2018. *Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments*. National Academies Press.

Ng, D. T. K.; Leung, J. K. L.; Chu, S. K. W.; and Qiao, M. S. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2: 100041.

Smilkov, D.; and Carter, S. 2022. TensorFlow Playground. Accessed: 2024-09-16.

Weizenbaum, J. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1): 36–45.