AI Toolkit: Libraries and Essays for Exploring the Technology and Ethics of AI

Levin Ho[1], Morgan McErlean[1], Zehua You[1], Douglas Blank[2], and Lisa Meeden[1]

[1]Computer Science Department, Swarthmore College
[2]Comet ML, Head of Research
{lho2, mmcerle1, zyou1, lmeeden1}@swarthmore.edu, doug.blank@gmail.com

Abstract

In this paper we describe the development and evaluation of AITK, the Artificial Intelligence Toolkit. This open-source project contains both Python libraries and computational essays (Jupyter notebooks) that together are designed to allow a diverse audience with little or no background in AI to interact with a variety of AI tools, exploring in more depth how they function, visualizing their outcomes, and gaining a better understanding of their ethical implications. These notebooks have been piloted at multiple institutions in a variety of humanities courses centered on the theme of responsible AI. In addition, we conducted usability testing of AITK. Our pilot studies and usability testing results indicate that AITK is easy to navigate and effective at helping users gain a better understanding of AI. Our goal, in this time of rapid innovations in AI, is for AITK to provide an accessible resource for faculty from any discipline looking to incorporate AI topics into their courses and for anyone eager to learn more about AI on their own.

1 Introduction

AI's impact on society is at an all time high. The One Hundred Year Study on AI states that "the field's successes have led to an inflection point: it is now urgent to think seriously about the downsides and risks that the broad application of AI is revealing" (Littman et al. 2021, p. 71). Furthermore, the report argues that the AI research community has a crucial role to play, sharing "important trends and findings with the public in informative and actionable ways, free of hype and clear about the dangers and unintended consequences along with the opportunities and benefits" (Littman et al. 2021, p. 71).

Currently, a number of excellent resources exist to help explain AI, including blogs such as Janelle Shane's AI Weirdness (Shane 2016) and YouTube channels such as Grant Sanderson's 3Blue1Brown (Grant Sanderson 2015). Also there are numerous toolkits to help build AI such as the Open-Source AI Cookbook (Hugging Face 2024), Fast AI (fast.ai 2024) and Machine Learning Complete (Nyandwi 2023). What is novel about the AITK project with respect to these existing resources and tools? Let's begin by describing AITK.

AITK consists of both a set of Python libraries and a set of Jupyter notebooks. The Python libraries are designed to make it easy to experiment with key AI frameworks such as deep learning and robotics. The code focuses on allowing users to visualize the inner workings of these frameworks, opening up the black box. For example, when building a neural network, users can display the network's structure and visualize the network's activations across all layers based on any desired input values. But how can novices utilize these tools without knowing how to program? The Jupyter notebooks provide a way for novices to interact with this code without having to write it themselves. A notebook consists of a series of cells, where each cell contains either executable code or formatted text. These notebooks enable users to interactively execute and modify prewritten code, interspersed with accompanying explanation, allowing them to gain a richer understanding of AI and how it functions. One can think of Jupyter notebooks as essentially computational essays, and they have transformed the way science is now communicated (Perkel 2021).

Our aim with the AITK project is to provide a resource for better understanding AI and its risks that is free, open-source, accessible to novices, and fundamentally interactive. Blogs and YouTube channels provide informative analysis of AI, but they lack the ability for users to try out the concepts on their own. Existing educational toolkits, including those using Jupyter notebooks, are generally about teaching users to build their own AI and so require substantial coding ability. As a result, they are not well-suited for novices, or for those mainly wishing to become better informed about AI. The AITK project bridges the gap between explanations and experimentation, by providing both together as computational essays that were written with novices in mind.

The rest of this paper is organized as follows. First, we describe the types of computational essays that are available in AITK, and provide details about one example notebook. Next, we discuss how AITK has already been piloted at a number of different institutions in humanities courses focused on the theme of responsible AI. Then we describe how AITK can help address the new focus of Computer Science curricula on the societal impacts of AI. Because our goal is that AITK should be accessible to anyone, we also conducted usability testing and describe both the design and results of this testing. Based on the results, we reflect on ways
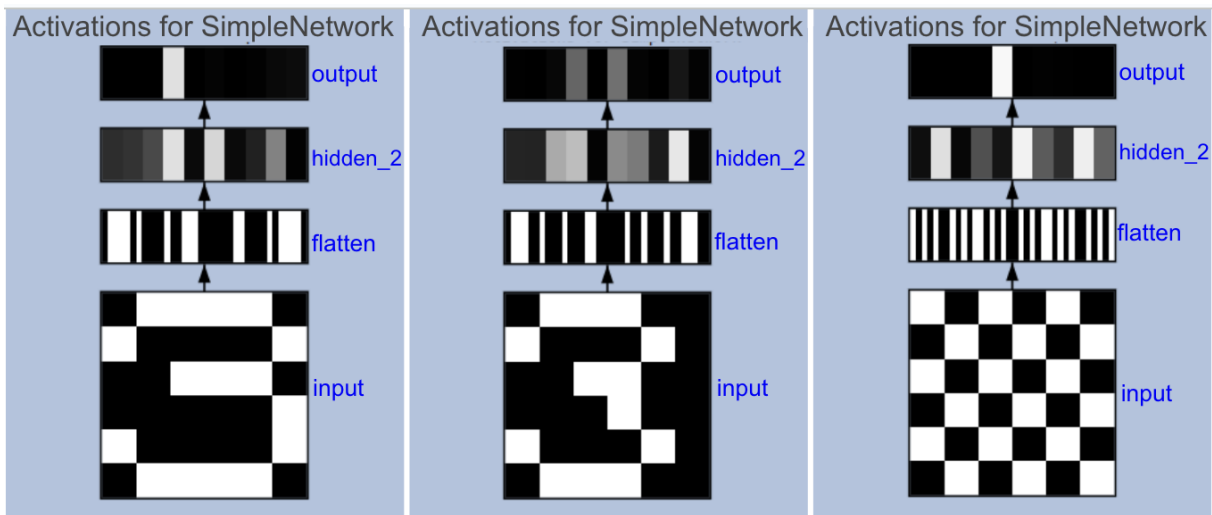
Figure 1: AITK allows users to visualize the activations within a network. On the left, the network has correctly recognized the image as the digit 3. In the center, the network is unsure of whether the image represents the digit 3 or 5. On the right, the network has incorrectly classified the checkerboard image as the digit 4.

to improve AITK. Finally, we summarize the paper in the Conclusions.

2   Description of AITK

AITK began as a collaborative effort between three computer science faculty members, Douglas Blank, James Marshall, and Lisa Meeden, built on many years of experience teaching AI at small liberal arts colleges, and based on several previously created open source projects (Blank et al. 2006; O'Hara, Blank, and Marshall 2015). It was released as an open source project on GitHub[1] in 2021. It underwent significant additions and upgrades in 2024 with the help of three undergraduate students (all co-authors on this paper).

The notebooks within AITK are organized into three main categories: Neural Networks, Generative AI, and Robotics. There is also an Advanced category that is most appropriate for users with a deeper background in computer science and AI. A suggested sequencing is provided for where to begin and how to proceed within each category, see Table 1.
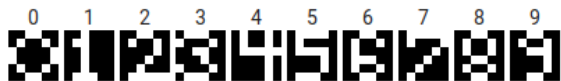

Figure 2: A sample of some of the 6x6 digit images used in the *Basic Neural Networks* notebook.

Example: *Basic Neural Networks* Notebook

Because neural networks and deep learning are the foundation of much of the latest innovations in AI, we recommend beginning with the Neural Networks category, and specifically with the notebook *Basic Neural Networks*, which fo-
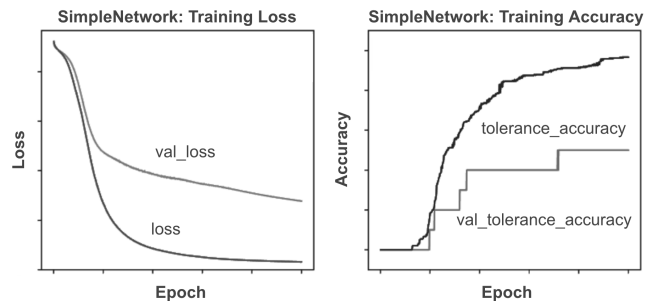

Figure 3: AITK automatically generates graphs summarizing the network's progress during training at reducing loss (at left) and improving accuracy (at right).

cuses on a small, simple dataset of 6x6 digits, see Figure 2. The task in this notebook is to train a neural network to correctly classify the images as representing the digits 0-9. The notebook introduces fundamental concepts from machine learning such as categorization, training and validation sets, generalization, and overfitting.

One of the key features of AITK is that it provides numerous visualizations to make the the concepts being introduced more concrete. For example, AITK allows the user to view a depiction of the network. The sequence of layers in the network are shown as boxes connected by arrows, and activation values within each layer are encoded in grayscale, with black representing 0, white representing 1, and gray for values in between, see Figure 1. Users can see from this visualization that the network has a two-dimensional input layer, which is flattened, and passed through one hidden layer to the output layer. During training, AITK also produces graphs showing the network's progress at both reducing loss and improving accuracy, see Figure 3.

---

[1]https://github.com/ArtificialIntelligenceToolkit/aitk

	Neural Networks	Generative AI	Robotics
Begin	*Basic Neural Networks*	*Word Embedding*	*What is it like to be a robot?*
Next	*Categorizing Faces*	*Nano GPT*	*Braitenberg Vehicles*
	Data Manipulation	*Image Generation*	*Seek Light*
Further	*Analyzing Hidden Representations*	*Transformer*	*Subsumption*
	Structure of Convolutional Networks		

Table 1: Sequencing of AITK notebooks by category

Users are invited to draw their own 6x6 digit and test whether the trained network is able to correctly categorize it, demonstrating the idea of generalization to novel data. In addition, the notebook explores how the network responds to other patterns, such as a checkerboard, and shows that the network may "recognize" it as a particular digit (see Figure 3, right), which is clearly problematic. This leads to a second experiment where random images are added to the training set (replacing all of the images of the digit zero). We then test if we can train the network to recognize both when an image is **not** a digit (see Figure 4 for examples of images that are not digits) and when an image is one of the digits one through nine. Although, this experiment is only partially successful, the second version of the network is typically able to learn that the checkerboard is **not** a digit.

Figure 4: A sample of 6x6 random images generated in the *Basic Neural Networks* notebook for training on what is **not** a digit.

Before using this notebook, many users have incorrect intuitions about how neural networks operate. They often expect the network to correctly classify every image, never making mistakes (as shown in the center and right of Figure 1). They are often surprised that each training run of the network may yield different outcomes, due to the random initialization of the weights. And by changing the data set and the task in the second experiment, they begin to recognize the importance of the training data in shaping the outcome (this theme is explored in greater depth in the notebook *Data Manipulation*). As AI is used in more and more critical applications, gaining these insights plays an important role in better informing the public about how AI works and its potential dangers.

Next we will explore how AITK notebooks, such as the one discussed in this section, have been piloted in a variety of college courses.

3 Piloted in Humanities Courses

Starting in 2021, the National Humanities Center brought together faculty from fifteen institutions to develop courses to engage students on their respective campuses to think more deeply about the ethical issues surrounding AI. This initiative was called the *Responsible Artificial Intelligence Curriculum Design Project* (National Humanities Center 2021).

The cohort of faculty spent two years developing new humanities courses and offered them for the first time during the 2023-24 academic year.

Notebooks from AITK were piloted in a subset of these courses: those developed at Bowdoin College, Davidson College, Duke University, Swarthmore College, and the University of Utah. All of these courses used the *Basic Neural Networks* notebook described earlier. Each faculty member downloaded a copy of the AITK notebook to Google Colab, making it easy for them to customize the notebook for their needs, elaborating on or removing the existing material. Then students made their own copies of the faculty member's notebook and also ran them in Google Colab. The requirements to using one of these notebooks in a course is relatively lightweight: students must have a Google account and a laptop with a web browser. During these pilot studies all students were able to successfully use and complete the notebooks.

Case Study: Swarthmore College

Let's look more closely at one of these course entitled *Ethics and Technology*, co-developed and co-taught by a Computer Scientist and a Philosopher (Meeden and Thomason 2024) at Swarthmore College. The prerequisite for this course was an introductory course in either Computer Science or Philosophy, leading to a diverse mix of student backgrounds. Students were expected to write philosophy papers, and to complete labs related to AI, where the labs were based largely on AITK notebooks. Initially, the CS students were nervous about the papers, while the Philosophy students were worried about the labs.

This course began with primers on both ethics and AI. The notebook *Basic Neural Networks* was used during the AI primer. Throughout the semester, the students read chapters from Kate Crawford's book *Atlas of AI* (Crawford 2022). The notebook *Categorizing Faces* was used to illustrate concepts from her chapters entitled "Classification" and "Affect". Other readings were also included to delve into additional topics not covered by Crawford. For example, we discussed embodiment and read Thomas Nagel's well-known paper *What is it like to be a bat?* (Nagel 1974) and paired this with the AITK notebook *What is it like to be a robot?*.

In the final course evaluations, students were asked "How did the lab component of the course complement your understanding of the course's content?" Here is a sample of their responses:

- "The lab component was very helpful. It was quite lightweight (people without a lot of experience will understand)."

- "The lab was super helpful to put our arguments in class into context."
- "I enjoyed the hands-on nature of the labs and the supplemental questions helped with my understanding."

Some of the students with CS background wished the labs had been more substantial:

- "Loved the labs—maybe more CS intensity can be introduced. It was a fun way to gain insights."
- "As a CS major they were quite easy, that being said they served a purpose."
- "I loved the labs! I wish they had gotten a little more technical, had us create neural networks with slightly less hand holding, but I feel they gave me a really good understanding of how ML works."

We were encouraged that the humanities majors in this course found the AITK notebooks to be accessible and that the CS majors were still engaged despite wanting to do more technical work.

After the successful use of AITK notebooks in this undergraduate course in spring 2024, we piloted them a second time in a four-week mini-course on AI in summer 2024 for faculty and staff at Swarthmore College, which was sponsored by the Teaching and Learning Commons. There was a diverse mix of participants in this mini-course including faculty and lecturers from departments in the Humanities, Social Sciences, and Sciences, as well as staff from the Library, Admissions, Student Affairs, Advancement, Technology Services, and Career Services. At the end of this course, we asked the participants to reflect on the use of the AITK notebooks for lab work. Here are a few of their responses:

- "The notebooks were terrific, going back later to read everything and run the code again really reinforced the topics."
- "I was really intimidated by the idea of doing labs, but in reality they weren't scary at all. I like that I am able to save and refer back to the notebooks."

In this section, we've seen how AITK notebooks have been successfully used in both humanities courses at multiple institutions and for a more general audience in a summer course on AI. Next let's consider how AITK notebooks could address new emerging requirements for CS undergraduate curricula to focus on the societal impacts of AI.

4 Highlighting Societal Impacts of AI

Approximately every 10 years the ACM and IEEE, the primary professional organizations of computer science, issue updated curricular recommendations for undergraduate CS education. The most recent update, called CS2023, "highlights how important it is to understand and assess the broader societal impacts and implications of AI methods and applications, including issues in AI ethics, fairness, trust, and explainability" (Eaton and Epstein 2024). To reflect this, a new core knowledge unit has been added entitled "Applications and Societal Impact" with a suggested 3 hours of instruction.

We believe that a number of AITK notebooks offer an excellent resource for helping to meet these requirements. For example the following notebooks focus on the societal impacts of AI:

Data Manipulation Examines how the relative numbers within each class of a data set biases the outcomes. It highlights the "Gender Shades" work that revealed biases in facial recognition software (Buolamwini and Gebru 2018).

Word Embedding Describes how word embeddings are formed and discusses how they learn biases implicit in training texts such as the association of particular careers with a specific gender (Bolukbasi et al. 2016).

Image Generation Explains the process of image generation and concludes by revealing the biases that emerge in generated images, for example in typical skin tones produced for the prompt "lawyer" vs the prompt "criminal" (Nicoletti and Bass 2023).

Given the diverse audiences from both the humanities and the sciences that might be interested in using AITK, we decided to conduct a usability study.

5 Usability Study

We had two goals in conducting the study. Our first goal was to observe how new users to AITK were able to navigate the GitHub website in order to find a notebook of interest to them. To this end, our user testing facilitators were trained by our UX Librarian who provided various materials to prepare them, such as (Blakiston 2024). Our second goal was to assess the value of our notebooks as educational tools.

We recruited a total of 17 participants who were divided into three groups based on their experience level with CS and GitHub. We had 6 novice users who had never used GitHub and only one of whom had taken an introductory CS course. We had 6 intermediate users who had used GitHub and had taken at least one intermediate CS course. We had 5 experienced users, three were CS or math majors with several upper-level CS courses or some other relevant AI/ML experience and two were CS faculty members.

The usability test lasted for at most one hour and participants received a $25 gift card for their time. Their task was described by the facilitator as follows:

> We'll be doing a website usability session: I'll describe some situations for which information from the toolkit will be helpful and you'll explore the toolkit to look for it. We'll be testing the toolkit—not you– in order to see if the toolkit is intuitive to use. It will help us to watch you navigate, especially if you are not able to find information, because that will tell us how to make the toolkit easier to use. Here is a topic to explore. Go through any notebooks necessary to gain knowledge in this topic.

The landing page of the AITK website contains a link to a short summary of the notebook topics and another link to our suggested sequencing through the notebooks. Some participants read the summary and sequencing information first
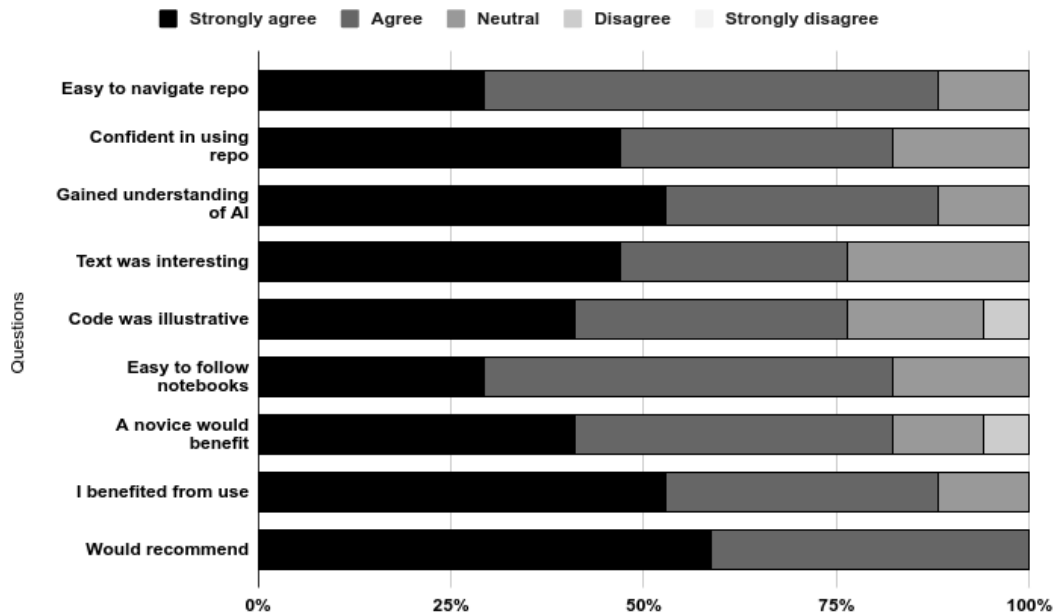
Figure 5: User testing responses indicate that AITK is easy to use and accessible for most novices.

while others chose to click directly on the notebooks folder and began searching around for a notebook with their assigned topic. About half of the participants were able to successfully navigate to their desired notebook on the first try, while the other half needed some guidance.

Once a participant reached the notebook of interest, they were asked to use the notebook, executing the cells within it. For participants who had never used a notebook before, there were some challenges. Several wondered whether they needed to actually read through the code and understand it (this wasn't our intent). In some cases, participants forgot to execute a code block, and then later code blocks failed due to missing dependencies, and they didn't understand how to fix it based on the error messages.

In contrast, during our pilot studies in courses, students had been guided through obtaining the notebook they needed and help was on hand when using the notebooks, so these issues had not arisen. It was useful to discover that for novices working on their own, both navigating the website and using the notebooks could be challenging.

At the end of the usability testing, participants were asked to complete a survey. They rated their agreement with a series of statements using a 5-point Likert scale: Strongly agree, Agree, Neutral, Disagree, or Strongly Disagree. The statements they were surveyed about were:

- The GitHub repo was easy to follow and the sequencing of notebooks was logical to me.
- I am confident in my ability to navigate this GitHub repo and use the notebooks on my own.
- I feel more confident in my understanding of the topic I reviewed.
- I found the text sections of the notebook interesting.

- I found the code sections (running the code cells, visualizations, etc.) engaging and illustrative.

- I found the notebook accessible and easy to follow.

- I believe a novice (a person who doesn't have a lot of background knowledge) would benefit from using this tool.

- I believe that my time using this toolkit has benefited me.

- Based on the content I saw, I would recommend this toolkit to others.

A summary of the user testing results are provided in Figure 5. For every statement at least 75% of the participants strongly agreed or agreed with them.

For two of the statements, about the code sections being illustrative and novices benefiting, a single participant disagreed in each case. Based on this result and on our observations during user testing, we plan to provide more scaffolding for novices as they use the notebooks. For example, we will periodically remind users that when an error occurs, it is likely due to forgetting to execute a previous code cell, and we will suggest that they go back and rerun cells from the beginning. We also plan to explain that users do not need to understand the code sections; they simply need to run them to see the results.

We were heartened to see that the notebooks did succeed as an educational tool, with 87% of the participants agreeing that they understood a topic better after using AITK and that they benefited from its use. Also, 100% of the participants agreed that they would recommend AITK to others.

6 Conclusions

In this paper we described the Artificial Intelligence Toolkit (known as AITK), an open-source project consisting of both Python libraries and Jupyter notebooks. Our goal for this project is to provide a resource for better understanding AI that is free, accessible to novices, and fundamentally interactive. We argued that though many other resources exists for exploring AI, they are either too advanced for novices, or are simply explanations, with no opportunity for interaction. AITK bridges the gap between experimentation and explanation by providing both components together within computational essays that were written with novices in mind. To illustrate this, we provided details on one notebook that introduces users to key concepts related to neural networks and machine learning.

Next, we described how AITK has been successfully piloted in humanities courses developed around the theme of responsible AI at a number of institutions involved in a National Humanities Center initiative. The fact that humanities students, who for the most part had no computer science background, were able to successfully use these notebooks demonstrates that they are accessible to novices. We argued, that AITK could also help computer science faculty address the latest ACM/IEEE curriculum requirements for a new core knowledge unit that focuses on AI applications and their societal impacts. We noted that AITK is a lightweight addition to any course, only requiring students to have a Google account, and a laptop with a web browser to be able to run the notebooks within Google Colab. Finally, we conducted a usability study of AITK with a mix of novice, intermediate, and advanced participants. Results were overwhelmingly positive, though our observations revealed that there are still ways to improve the toolkit by providing additional scaffolding for novices.

Acknowledgments

References

Blakiston, R. 2024. The UX Cookbook: Usability Testing. https://theuxcookbook.com/usability-testing. Accessed: 2024-12-15.

Blank, D.; Kumar, D.; Meeden, L. A.; and Yanco, H. 2006. The Pyro Toolkit For AI And Robotics. *AI Magazine*, 27: 39–50.

Bolukbasi, T.; Chang, K.-W.; Zou, J.; Saligrama, V.; and Kalai, A. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *30th Conference on Neural Information Processing Systems*.

Buolamwini, J.; and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of Machine Learning Research Conference on Fairness, Accountability, and Transparency*.

Crawford, K. 2022. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Eaton, E.; and Epstein, S. L. 2024. Artificial Intelligence in the CS2023 Undergraduate Computer Science Curriculm: Rationale and Challenges. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*. www.aaai.org.

fast.ai. 2024. fast.ai. https://docs.fast.ai. Accessed: 2024-12-15.

Grant Sanderson. 2015. 3Blue1Brown. https://www.youtube.com/@3blue1brown. Accessed: 2024-12-15.

Hugging Face. 2024. Open-Source AI Cookbook. https://huggingface.co/learn/cookbook/en/index. Accessed: 2024-12-15.

Littman, M. L.; Ajunwa, I.; Berger, G.; Boutilier, C.; Currie, M.; Hadfield, F. D.-V. G.; Kitano, M. C. H. C. I. H.; Levy, K.; Lyons, T.; Mitchell, M.; Shah, J.; Sloman, S.; Vallor, S.; and Walsh, T. 2021. Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. https://ai100.stanford.edu/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence-ai100-2021-study. Accessed: 2024-12-15.

Meeden, L.; and Thomason, K. 2024. Ethics and Technology Course Syllabus. https://www.cs.swarthmore.edu/~meeden/phil27/s24/index.php. Accessed: 2024-12-15.

Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review*, 83(4).

National Humanities Center. 2021. Responsible Artificial Intelligence Curriculum Design Project. https://nationalhumanitiescenter.org/education-programs/faculty-programs/responsible-artificial-intelligence-undergraduate-curriculum-design-project/. Accessed: 2024-12-15.

Nicoletti, L.; and Bass, D. 2023. Humans Are Biased. Generative AI is Even Worse. *Bloomberg*.

Nyandwi. 2023. Machine Learning Complete. https://github.com/Nyandwi/machine_learning_complete/tree/main. Accessed: 2024-12-15.

O'Hara, K.; Blank, D.; and Marshall, J. 2015. Computational Notebooks for AI Education. In *The Florida Artificial Intelligence Research Society (FLAIRS-2015)*. FLAIRS.

Perkel, J. M. 2021. Ten Computer Codes that Transformed Science. *Nature*, 589.

Shane, J. 2016. AI Weirdness. https://www.aiweirdness.com/. Accessed: 2024-12-15.