
AI Accountability Teacher Toolkit

Introduction

This resource provides some background information and key concepts about Artificial Intelligence (AI) technologies that can be useful to educators who are interested in better understanding AI. The resource also introduces educators to Pulitzer Center reporting on Information and Artificial Intelligence which can support educators in teaching more effectively about AI in the classroom.

Pulitzer Center's Information and Artificial Intelligence initiatives support journalism and audience engagement on in-depth [AI accountability](#) stories that examine governments' and corporations' uses of predictive and surveillance technologies to guide decisions in policing, medicine, social welfare, the criminal justice system, hiring, and more. Pulitzer Center-supported AI reporting explores with nuance how AI systems are designed, sold, and deployed in communities around the world. As more of these systems are deployed in schools and communities, educators are grappling with questions about if and how to teach about them and whether student and teacher use should be regulated. This tool illuminates how educators can utilize AI accountability reporting as a tool to inform their decisions and actions and this tool will help educators make this critical reporting more accessible and meaningful to students.

Key Concepts

- Data Footprints and Digital Citizenship
- Understanding Types of AI
- The Evolution of AI
- The Marketing Behind AI
- The Role of AI Journalism
- Considering Classroom Impact: Safety and Privacy
- Considering Classroom Impact: Equity and Accountability

Instructions

For each Key Concept,

1. Review the background information provided on the page and respond to the following reflection questions
 - a. Why is this a key concept to understand in my work as an educator?
 - b. What parts of this concept do I believe are important for students to understand and engage with?
2. Read the Pulitzer Center news stories highlighted in the infographic and respond to the following reflection questions:
 - a. How does this reporting expand my understanding of the key concepts?
 - b. What actions am I inspired to take in my work after engaging with this reporting?

Key Concepts for Understanding AI

Data Citizenship and Digital Footprints

Summary

Your digital footprint is your unique trail of data from traceable activities on the internet and on digital devices. This includes your interactions with others online. Your digital footprint can inform AI technologies accessing your data.

Background

Digital Citizenship is the responsible and ethical use of digital technology. Students can maximize their digital citizenship by respecting their own privacy and that of others, utilizing online etiquette when interacting with others, and remaining aware of cyber security and data scraping when sharing their information online.

- **Digital Literacy:** When accessing digital resources, including AI tools, remember to assess the accuracy, reliability and bias reflected in these resources. It is tempting to be distracted by relevance and presentation of information, especially when information feels easily accessible. Consider:
 - The purpose of the information and its intended audience
 - The presence of facts, citations, and clear authorship in the resource
 - The potential biases of the resource and its creators
 - The reliability of the resource and its creators
- **Privacy:** When using digital resources, model reading through website and app privacy settings. These settings help users control what personal information website owners can access, who they share the information with, and how they use this information. Additionally, though students may be accustomed to using personal information like names and birthdays to sign into school accounts, they should avoid this when using other tools and resources. Best practice is to use strong and unique passwords with a combination of letters, numbers, and symbols.
- **Netiquette:** Online etiquette is important. When using digital devices users should consider that once you post, comment or share information online it can exist even beyond your attempts to delete it. Students should know not to post anything online they would not say in person and should be aware that their posts and interactions can impact their reputation, offline relationships, and digital footprint.

Your digital footprint consists of the trail of data you leave when using digital devices, applications and/or search engines. There are two types of a digital footprint: active or passive.

- An active footprint is data that you intentionally share online: your social media, online comments, shopping preferences, photos, video and location data you voluntarily share.
- A passive footprint is data collected with your indirect input: your browsing history, website tracking, IP address, device information, tax records and social security number.

Reporting Highlight

[Peering Into the Black Box by Arijit Sen](#): A reporting project that demonstrates how university student's digital footprints were fed into an AI surveillance tool at their schools.

Digital Footprint and Citizenship Resources

- [Common Sense Education](#): Free lesson and resources for grades K-2, 3-5, 6-8, and 9-12 focused on digital citizenship.
- [Pulitzer Center](#): Free lesson plans and resources for grade K-12 that encourage critical thinking, media literacy, and communication.
- Additional References: [edutopia.org](#); [verywellmind.com](#)

Understanding Types of AI

Summary

AI is more than ChatGPT and even GPT, like all of the AI that exists today, is Limited or Narrow in capability. It's most useful to describe AI technologies according to function: reactive, predictive, or generative.

Background

Reactive and Predictive AI

- These are tools that utilize patterns to perform very specific tasks. Reactive AI, like a chess-playing computer works only with presently available data and cannot recollect previous decisions. More advanced Predictive AI, such as virtual assistant chatbots, can use both past- and present-moment data to find patterns and develop outcomes.

Generative AI

- Generative AI tools like ChatGPT and Midjourney generate content to match patterns learned from data input. The tools can't distinguish accurate from inaccurate information and if missing data, will attempt to fill in information. Technologists often do not know which patterns the tool is following to generate content.

Better Understanding the AI We Have

- Some questions you and your students can ask to better understand and evaluate an AI tool include:
 - How is this tool being marketed? What is it supposed to do?
 - How will I know if the tool has accurately completed this task? How important is accuracy to the work I need to do?
 - What data is required for this tool to complete the task? Is there information shared about how the tool was trained?
 - How will this tool use my data and any information I put into it?
 - If this tool requires memory, how and where is it storing past data input?

Conceptualizing More Advanced AI

- Theory of Mind AI and Self Aware AI are titles for types of AI tools that don't yet exist. Falling under the categories of General (AGI) or Super AI, these tools, if ever developed, would theoretically be able to understand and model the thoughts, intentions, and emotions of humans, animals, and other AI. In the case of Self-Aware AI theory, the AI tool would have the ability to understand its own internal conditions, thus developing emotions, needs, and beliefs.

Sources

- <https://www.ibm.com/think/topics/artificial-intelligence-types>
- <https://www.educative.io/blog/generative-ai-vs-predictive-ai>

The Evolution of AI

Summary

The history of AI is intertwined with the history of data, statistics, and computer science. The AI technologies of today trace their roots back to advancements in machine learning, as a means of encrypting and decoding information during World Wars I and II.

Background

Chronology: The History and Evolution of AI

- 1930s
 - 1933: The Polish Cipher Bureau use cryptanalysis to break the German Enigma machine.
 - 1936: Alonzo Church and Alan Turing introduce their theory on computable functions.
- 1940s
 - 1943: Warren McCulloch and Walter Pitts introduce the MCP model neuron which will contribute to the development of artificial neural networks.
- 1950s
 - 1950: Alan Turing releases test for machine intelligence.
 - 1952: Claude Shannon makes maze-solving Theseus mouse.
 - 1956: Dartmouth Conference where John McCarthy coins the term “Artificial Intelligence.”
- 1960s
 - 1966: Joseph Weizenbaum creates ELIZA, widely considered the first chatbot.
 - 1969: Stanford AI Center develops Shakey the Robot to function independently in realistic environments.
- 1970s-1980s
 - The late 1970s to early 1990s mark an “AI Winter” with a gap between expectations and technological advancement.
- 1990s
 - 1989: Convolutional Neural Networks demonstrated to recognize handwriting.
 - 1995: Richard Wallace introduces Alice the Chatbot.
 - 1997: Chess playing program, DeepBlue, beats champion.
- 2000s
 - 2000: Cynthia Breazeal and MIT AI Lab’s Kismet robot reads and mimics emotions.
 - 2002: Roomba vacuum is introduced using AI technology.
 - 2009: ImageNet visual database introduced.
- 2010s

- 2011: the question-answering machine, IBM Watson, is introduced.
- 2011: Apple introduces Siri feature with natural language processing.
- 2014: Amazon's virtual assistant Alexa becomes a primary tool on devices.
- 2016: DeepMind's AlphaGo defeated top Go player Lee Sedol.
- 2017: humanoid robot Sophia the Robot is recognized as a citizen of Saudi Arabia.
- 2018: Cimon the robot built for International Space Station.
- 2020s
 - 2021: OpenAI releases text-to-image model, DALL-E.
 - 2022: OpenAI releases ChatGPT.
 - 2024: AI research and increased conversation about its impacts and regulation continue.

Sources

- ["Data: Past, Present, and Future" Course timeline](#) by Chris Wiggins and Matthew L Jones
- [The History of AI](#) on Coursera.org
- [TechTarget.com](#)

The Marketing Behind AI

Summary

The goals and effectiveness of AI marketing strategies has regularly shifted over time. There have been periods of excitement over AI advancements being "the next big thing," and periods of disillusionment and realism from audiences, funders, and developers.

Background

- 1950s and 60s - Visionary Excitement
 - Read the 1955 [Dartmouth Conference Proposal](#) a "2 month, 10 man study of AI"
 - They asked, "can machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves?"
- 1970s - Disillusionment - AI Winter #1
 - Read the [1966 report on Automatic Language Processing Advisory Committee Report on Language and Machines](#)
 - "Perhaps our attitude might be different if there were some pressing need for machine translation, but we find none".
- 1980s - More Hype
 - Read the [Department of Defense's Strategic Computing Initiative written in 1983](#).
 - "Advances in "expert system" technology now enable the mechanization of the practical knowledge and the reasoning methods of human experts in many fields."
- 1990s - AI Winter #2
 - The Strategic Computing Initiative is ended by the Defense Advanced Research Projects Agency in 1993 and the government reduces investment in AI projects. (McCorduck, Pamela (2004), *Machines Who Think* (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 1-56881-205-1, pp. 426–432)
- Late 90s-2000s - Resurgence

- Read: IBM Deep Blue -first computerized world chess champion and development of Amazon Recommendations Systems - personalized shopping recommendations
- 2010-Present - Integration and Ethical Considerations Arise
 - AI is marketed as an integral part of daily life with voice response apps like Siri and personalized advertising.
 - There is a growing emphasis on responsible and ethical AI.
- 2022- Pulitzer Center pushes AI news beyond marketing with the AI Accountability Network
 - The Network seeks to address the knowledge imbalance on artificial intelligence that exists in the journalism industry and to create a multidisciplinary and collaborative ecosystem that enables journalists to report on this fast-evolving topic with skill, nuance, and impact.

Reflection Questions for the Present Day

1. Who benefits from hyping up AI?
2. Are we headed towards another AI winter?
3. Throughout history why was AI marketed with the promise of great advances without evidence?
4. Historically, who has marketed AI?

Artificial Intelligence is a Global Issue

Summary

AI is a topic that connects the world, often intersecting with other systemic issues like racial justice and health. Leverage AI news stories as a lens to explore a range of stories about the impacts of AI on a local and global scale.

Reporting Highlights

- AI & Environment
 - "Microsoft's Hypocrisy on AI" by Karen Hao for *The Atlantic*
- AI & Migration
 - "Eye on the Wall: Refugees and 'Smart' Borders" by Lydia Emmanouilidou for *Aljazeera, Deutsche Welle, Monocle and Computer Weekly*
- AI & Labor Rights
 - "India's Gig Workers and Algorithms" by Varsha Bansal for *Wired, Rest of the World and the MIT Review*
- AI & Gender Equality
 - "There Is No Standard': Investigation Finds AI Algorithms Objectify Women's Bodies" by Hilke Schellman for *The Guardian*
- AI & Governance
 - "This Algorithm Could Ruin Your Life" by Gabriel Geiger, Evaline Schot and Matt Burgess for *Wired*
- AI & Democracy
 - "The Start-Up Busting the Voting Machine Monopoly" by Spenser Mestel for *Undark*

Considering Classroom Impact: Safety and Privacy

Summary

When considering the impact of AI technologies in your classroom and community it is important to consider student safety and privacy.

Background

The role of Pulitzer Center reporting on Information and Artificial Intelligence is to analyze the impact of emerging technologies on people, including benefits, risks and harms.

Reporting Highlight

Read any or all of the following Pulitzer Center reporting projects before engaging with the questions. These resources and questions support a framework for leading discussions about artificial intelligence that centers people and impact:

There Goes the Neighborhood by Lam Thuy Vo for *The Markup*, *MIT Technology Review* and *Just Tech*

Tracked by Garance Burke, Juliet Linderman, Martha Mendoza and Michael Tarm for *The Associated Press*

Reflection Questions

- Summarize the AI tool(s) being described in the reporting projects. Based on your AI learning, how would you categorize the tool(s)?
 - Reactive AI, Predictive AI, Generative AI
- How does the tool work? What kind of data does it require and where is the data from?
- What are the stated benefits and drawbacks of utilizing the tool? Can you think of others?
- Is the tool designed to automate tasks or eliminate work? How does using this tool redefine human work and labor?
- Who is impacted by the tool? Is the impact of applying this tool different for different groups of people?
- How accurate or effective are the tools? What harm was caused by any inaccuracies?
- Did bias play a role in the development, deployment, and evaluation of the tool? Did the tool improve or further drive systemic inequalities?
- How might a tool like this impact your own classroom or community?

Considering Classroom Impact: Equity and Accountability

Summary

When considering the impact of AI technologies in your classroom and community it is important to consider student safety and privacy.

Background

The role of Pulitzer Center reporting on Information and Artificial Intelligence is to analyze the impact of emerging technologies on people, including benefits, risks and harms.

Reporting Highlight

Read any or all of the following Pulitzer Center reporting projects before engaging with the questions. These resources and questions support a framework for leading discussions about artificial intelligence that centers people and impact:

The Philippines and the AI Boom by Karol Llagan for PumaPodcast, Commoner, and the Philippine Center for Investigative Journalism

AI Colonialism by Karen Hao for MIT Technology Review

Reflection Questions

- Summarize the AI tool(s) being described in the reporting projects. Based on your AI learning, how would you categorize the tool(s)?
 - Reactive AI, Predictive AI, Generative AI
- How does the tool work? What kind of data does it require and where is the data from?
- What are the stated benefits and drawbacks of utilizing the tool? Can you think of others?
- Is the tool designed to automate tasks or eliminate work? How does using this tool redefine human work and labor?
- Who is impacted by the tool? Is the impact of applying this tool different for different groups of people?
- How accurate or effective are the tools? What harm was caused by any inaccuracies?
- Did bias play a role in the development, deployment, and evaluation of the tool? Did the tool improve or further drive systemic inequalities?
- How might a tool like this impact your own classroom or community?