Empowering Students to Navigate the Responsible AI Landscape

2023 ARTIFICIAL INTELLIGENCE ETHICS CURRICULUM FOR HIGH SCHOOL STUDENTS

Curriculum Developed by Shubhi Sinha

IU Center of Execellence for Women & Technology
AI Institute for Intelligence CI with Computational Learning in the Environment

FOREWORD

This guide serves as a teaching resource for educators and outlines an Artificial Intelligence Ethics Curriculum for High School Students. This curriculum was developed as part of the 5-year, NSF-funded Intelligence CI with Computational Learning in the Environment (ICICLE) grant OAC-2112606.

Artificial intelligence (AI) is undoubtedly revolutionizing how we interact with technology, our environment, and each other. As we navigate the rapidly evolving landscape of AI technologies, educators have a pivotal role in shaping the intellectual and ethical minds of future professional leaders. In the near future, high school students will enter industries of work that are already leveraging AI technologies to problem-solve and address challenges. Thus, we must integrate AI ethics literacy in K-12 education to help students develop a technical and ethical worldview.

As you navigate this document, you'll find instructional content and activities that can be implemented in high school classrooms or K-12 STEM camps. We hope this serves as a foundational tool that equips you with useful pedagogical resources, allowing you to facilitate a nuanced discussion of AI ethics in your classroom.

Shubhi Sinha
Indiana University Bloomington, College of Arts and Sciences
Lead Intern at the IU Center of Excellence for Women and Technology
ICICLE Intern

Pictured above are the first cohort of 80 high school students who experienced the AI Ethics Curriculum during a 2023 Pre-College Summer Camp, hosted by the IU Luddy School of Informatics, Computing, and Engineering.

ABOUT ICICLE

The AI Institute for Intelligent Cyberinfrastructure with Computational Learning in the Environment (ICICLE) is led by The Ohio State University and funded by the National Science Foundation. ICICLE aims to democratize Artificial Intelligence (AI) and simplify the application of AI for end users across various domains. The Institute aspires to create a user-friendly, integrated AI environment, fostering interdisciplinary communities and advancing foundational AI and CI. By emphasizing training, innovation, and technology transfer, ICICLE is dedicated to cultivating an AI-enabled workforce while supporting diverse and inclusive companies.

ICICLE focuses on education and outreach through a five year plan, aiming to train the next generation of AI/CI practitioners, increase AI ethics literacy in the field, and address issues related to AI bias and ethics. The phased approach starts with building the ICICLE team culture and extends to the broader community and stakeholders in mid-years. In the final two years, ICICLE reaches K-12 students and career professionals through summer camps and research projects for 11th-12th grade students. Through these efforts, ICICLE aims to enhance the accessibility and democratization of AI in society.

ABOUT THE CENTER

The IU Center of Excellence for Women & Technology is devoted to empowering women in technology, celebrating their achievements, and fostering interdisciplinary excellence through a supportive community. The Center aims to expand opportunities for women in technology fields, and addresses barriers to increase representation at all levels, from students to faculty and staff. With a mission to boost women's confidence and knowledge in technology, the Center is actively working to enhance their technical and leadership skills. This curriculum guide was developed and designed by four interns at the Center: Shubhi Sinha, Jiyoung Kim, Amy Kinney, and Juliette Albert.

CONTENTS

CONTENTS

What is AI?

Artifical intelligence (AI) is a groundbreaking field of computer science that is redefining how we interact with machines, our enviornment, and each other. AI refers to the development of computer systems that can perform tasks that typically require human intelligence. These tasks span a wide spectrum, ranging from simple computations to complex decision-making processes. Notably, what distinguishes AI from other technologies is that AI has the ability to learn and adapt from data and improve its performance and accuracy over time. This adaptability is powered by algorithms that allow AI systems to recognize patterns, extract insights, and make highly-acccurate predictions.

AI is an umbrella term that includes various subfields, such as machine learning, natural language processing, and robotics. Although these applications of AI are industry methods that optimize processes and enhance decision-making, AI-driven technologies are also becoming integral parts of our daily lives. From voice assistants that respond to our commands to recommendation systems that suggest personalized content online, we interact with AI on a daily basis - whether we know it or not.

AI Ethics Literacy

The rise of AI also brings forth complex ethical considerations, often causing unintended and harmful consequences. This is where AI Ethics comes into play. Although there is no universally accepted definition of AI Ethics, industry experts and research professionals have generally defined the term using four main principles: fairness, privacy, trustworthiness and transparency, and accountability. These principles - which also form the framework of our curriculum - have been loosely adopted by companies, businesses, and government conglomerates that want to design, develop, and deploy AI technologies in a responsible manner.

Let's take a closer look at each AI Ethics principle:
1. **Fairness**: AI systems should not cause unfair outcomes or exhibit algorithmic bias towards individuals or groups.
2. **Privacy**: AI systems should protect users privacy and regulate the use of personal and sensitive information.
3. **Trustworthiness and Transparency**: The user and developer should clearly understand the inner workings, or decision-making processes, of an AI system.
4. **Accountability**: Those involved in the lifecycle of an AI technology (design, development, and deployment) should be accountable for its ethical and legal impact.

As AI becomes ubiquitous and its applications permeate almost every aspect of society, AI Ethics literacy in K-12 education becomes even more important. At its core, AI ethics literacy is not just about recognizing the potential of AI; it's about fostering a sense of responsibility and accountability among AI users and developers. By encouraging students to think critically about the ethical impact of AI, educators can mold students to become responsible AI users and gain a deep appreciation for the societal implications that underpin the AI evolution.

CURRICULUM OVERVIEW

Our curriculum is designed to teach high school students about the ethical framework that guides the development and outcomes of responsible AI. We aim to empower students to engage responsibly with AI by (1) critically analyzing its societal and ethical implications and (2) making informed decisions regarding the use of AI technologies. In doing so, we hope to cultivate an ethically-driven mindset among future AI users and developers.

Our Framework

The curriculum consists of five modules. The first module serves as an introduction to AI and AI ethics. The remaining four modules focus on one of the AI ethics principles developed by our framework: fairness, privacy, trustworthiness/transparency, and accountability. This framework was developed as part of the NSF-funded, AI institute for Intelligent CI with Computational Learning in the Environment (ICICLE). The framework is informed by a pedagogical commitment to case based and user-centric issues, real-world implications for ICICLE's use cases, general principles, and critical theory from science and technology studies (STS) on infrastructure.

K-12 Standards (AI4K12, ISTE, CSTA)

Our curriculum is informed by national educational standards developed by: AI4K12 initiative, International Society for Technology in Education (ISTE), and Computer Science Teachers Association (CSTA).

The AI4K12 initiative has developed AI standards for K-12 students, which are summarized as The Five Big Ideas in Artificial Intelligence. Each module addresses one or more of the five core concepts: perception, representation and reasoning, learning, natural interaction, and societal impact.

The ISTE K-12 Standards focus on the knowledge and core competencies necessary for students to thrive in a technology-driven world. Our curriculum centers on two core ISTE standards: Digital Citizen and Knowledge Constructor. Each module addresses one or both of these ISTE standard and sub-standards.

The CSTA K-12 Standards provide a comprehensive outline of the computer science knowledge and skills that students need to be successful at problem-solving, computational thinking, and digital literacy. Each module address one or more of the seven CSTA standards below:
- 3A-IC-24: Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices.
- 3A-IC-25: Test and refine computational artifacts to reduce bias and equity deficits.
- 3A-IC-29: Explain the privacy concerns related to the collection and generation of data through automated processes that may not be evident to users.
- 3A-IC-30: Evaluate the social and economic implications of privacy in the context of safety, law, or ethics.
- 3B-AP-08: Describe how artificial intelligence drives many software and physical systems.
- 3B-IC-27: Predict how computational innovations that have revolutionized aspects of our culture might evolve.
- 3B-IC-28: Debate laws and regulations that impact the development and use of software.

Modules Structure

Each module is divided into two parts: a video and a classroom activity. The video outlines important content related to the AI ethics principle such as definitions, real-world news stories, case-studies, and examples of best practices. The classroom activity is designed to facilitate active discussion on AI ethics through scenario-based questions, thought experiments, and problem-based learning. As you navigate this document, each module includes information from the sections listed below.

Lesson Overview

The lesson overview summarizes the content and learning outcomes from the module. This section will also provide additional information such as the estimated duration to complete the module, the subject and grades that are best-suited for the content, and key vocabulary.

Learning Objectives

The objectives outline the core concepts and key takeaways that students should know after completing the module.

Preparation

The preparation section includes all the online and classroom resources educators will need to conduct the lesson. This will include important links, such as the module video, Activity Powerpoint, and any websites that are used for the lesson. This section also lists out any classroom materials that educators would need to have or prepare beforehand, such as a classroom Computer Projection System or pre-printed handouts.

Note: Each module will contain links to a Video and an Activity PowerPoint. The Activity PowerPoint can be used in the animated format or the PowerPoint format (which includes speaker notes). In addition, some modules will contain a more detailed Activity Instructions document in addition to the Activity PowerPoint.

Resources for Educators

This section outlines external resources that would be helpful for educators who want learn more about the content presented in the lesson plan. This includes links to online articles, blog posts, and videos.

Instructions

The instructions section outlines the main lesson plan in order of presentation. This section includes structured, step-by-step instructions to support educators as they conduct the lesson plan. Some modules will require further instruction than the one provided in this document. When applicable, educators should review the Activity Instructions documents linked under the "Preparation" section of the module.

Curriculum Outline

Table 1. Curriculum Outline.

	Module	Estimated Duration	Agenda
1	AI Ethics: Introduction	30 - 45 min	1. Introductions + Overview 2. Video: AI Ethics Introduction 3. Activity: AI Ethics in Action 4. PBL Student Question Submission
2	Fairness	30 - 45 min	1. Activity: Teachable Machine 2. Video: Fairness
3	Privacy	30 - 45 min	1. AI Ethics in the News 2. Activity: Pwned or Protected? 3. Video: Privacy
4	Trustworthiness & Transparency	30 - 45 min	1. Activity: Thought Experiment 2. Video: Trustworthiness / Transparency
5	Accountability	30 - 45 min	1. Video: Accountability 2. Activity: AI Incident Databse 3. PBL Student Answer Submission

Module 1
AI Ethics: Introduction
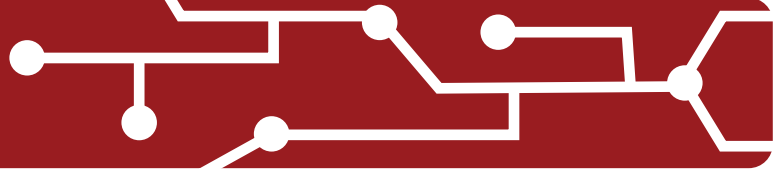
Lesson Overview

This module will provide an overview about AI and AI ethics to teach students about the broad impact of AI on our everyday lives. Students will learn about different types of AI systems and the diverse applications of AI technologies in different industries and fields. Students will also be introduced to the concept of AI ethics through a classroom activity that highlights real news stories and prompts peer-to-peer discussion.

Subject

Mathematics, Computer Science

Target Grades

9 - 12

Estimated Duration

30 - 45 minutes

Vocabulary

- Artificial Intelligence (AI)
- Machine Learning
- Deep Learning
- AI Ethics
- Fairness
- Privacy
- Transparency/ Trustworthiness
- Accountability

Learning Objectives

By the end of the lesson, students will be able to:
1. Identify the diverse applications of AI in everyday life.
2. Understand the ethical and societal consequences of AI technologies.
3. Recognize the four pillars of AI ethics.

Standards

AI4K12 Five Big Ideas in AI

1. Perception: Computers perceive the world using sensors.
3. Learning: Computers can learn from data.
5. Societal Impact: AI can impact society in both positive and negative ways.

ISTE

2b. Digital Citizen: Students engage in positive, safe, legal and ethical behavior when using technology, including social interactions online or when using networked devices.

CSTA K–12 Computer Science Standards

3B-AP-08: Describe how artificial intelligence drives many software and physical systems.

Preparation

To prepare for this module, educators should watch the video, and review the Activity Powerpoint, Activity Powerpoint Speaker Notes, and Activity Instructions. To prepare for the classroom activity, educators should be familiar with the news articles provided in the Activity Instructions.

Important Links

Video
Activity Instructions
Activity Powerpoint
Activity Powerpoint Speaker Notes

Classroom Resources

Computer Projection System (Projector and Screen) to display content for students

Resources for Educators

- AI Education - Khan Academy (Free Course)
- AI for Beginners - Simplilearn (Free Course)
- AI 101 for Teachers - Code.org (Free Course)
- What is AI - Google Cloud
- Difference Between AI, ML, and DL: Using An Incredibly Simple Example - Advancing Analytics
- Teaching AI Ethics - Leon Furze
- AI Ethics Case Studies - Princeton
- The Ethics of Artificial Intelligence - Stanford Encyclopedia of Philosophy

Instructions

1. Getting Started

Activity PPT (Slides 1 - 10)

Instructions
- Instructors should present the introductory slides from the Activity Powerpoint, which will provide examples of AI technologies in different contexts and industries.
- Instructors should also explain the curriculum structure to students: Each module will be 30-45 minutes long. For each module, students will do one or two classroom activities and watch a short video related to a specific AI Ethics pillar.

2. Video | AI Ethics Introduction

Video: AI Ethics Introduction

Instructions
- Play video. The video will discuss the following:
    1. What is AI and how does AI affect our daily lives?
    2. What is the role of data in determining the outcomes of AI systems?
    3. What are the four main pillars of AI ethics?

3. Activity | AI Ethics in Action

Activity PPT (Slides 11-36) Activity PPT Speaker Notes Activity Instructions

Purpose: This activity is meant to facilitate peer-to-peer discussion and help students adopt an ethical mindset when discussing the impact of AI technologies. The news stories chosen for this activity are all recent and relate to a different AI Ethics principle. Instructors don't need to explain the news story in great detail because most of these incidents will be discussed in greater depth later, in relation to a specific AI ethics module.

Instructions
- During this activity, you will present students with five different news stories related to AI Ethics from 2016 to 2022. After describing the news story, the stakeholders affected, and the impact of the AI technology, you will present students with a discussion question that prompts them to pick a side on the issue presented. Depending on the news story, the question will ask students to pick either who is at fault or who is responsible for the consequences or outcome of the AI technology.
- There are two ways to facilitate the discussion question after presenting the news story. First, instructors can have students discuss the news story and their stance on the question with a partner next to them. Next, as a class, instructions can ask students to either stand up or sit down depending on which side they chose. This will allow students to gauge the class's opinion on the question presented. Finally, instructors can ask a few students to share their reasoning for their stance with the class. Note: The Activity PowerPoint and Activity Instructions provided follow this structure for the discussion questions.
- Alternatively, instructors can divide the classroom into small groups. This will allow each group to discuss the news story and the stance they have on the question. Afterwards, the instructor should have groups share their stance with the class.

4. Student Question

Instructions
- After the first lesson, students will create and submit one question that they have about AI or AI ethics. Students can create this question based on what they already know about AI or AI ethics, or what they hope to learn from the course.
- Instructors should have students submit their question via Google Form (or any other type of online form where form responses are accessible to educators).
- Instructors should save these responses since students will re-visit them during the last module.

Module 2
AI Ethics: Fairness
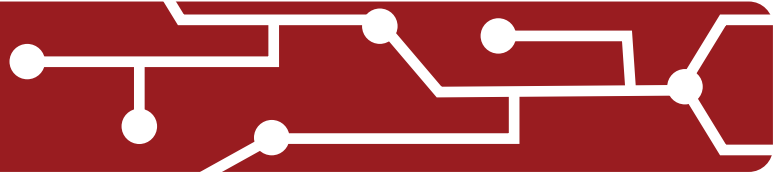
Lesson Overview

This module will focus on how AI systems can cause unfair societal outcomes due to algorithmic bias. Students will learn how algorithmic bias is introduced into AI systems, and the importance of representative and diverse datasets when training models. By exploring real-world case-studies of biased AI technologies, students will recognize the role of fairness in mitigating AI's unfair outcomes and inaccurate predictions.

Subject
Mathematics, Computer Science

Target Grades
9 - 12

Estimated Duration
30 - 45 minutes

Vocabulary
- AI Fairness
- Algorithmic Bias
- Training Data
- Facial Recognition
- Predictive Policing

Learning Objectives

By the end of the lesson, students will be able to:
1. Understand how algorithmic bias gets into data (both directly and by proxy).
2. Understand the importance of using diverse and representative training data for AI systems.
3. Recognize the unfair societal outcomes of biased AI technologies.

Standards

AI4K12 Five Big Ideas in AI

3. Learning: Computers can learn from data.
4. Natural Interaction: Intelligent agents require many kinds of knowledge to interact naturally with humans.
5. Societal Impact: AI can impact society in both positive and negative ways.

ISTE

2b. Digital Citizen: Students engage in positive, safe, legal and ethical behavior when using technology, including social interactions online or when using networked devices.

CSTA K–12 Computer Science Standards

3A-IC-24: Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices.
3B-AP-08: Describe how artificial intelligence drives many software and physical systems.

Preparation

To prepare for this module beforehand, educators should watch the video, and review the Activity Powerpoint, Activity Powerpoint Speaker Notes, and Activity Instructions. To prepare for the classroom activity, educators should also run through the Teachable Machine activity and demo beforehand to decide which activity format is most suitable.

Important Links

Video
Activity Powerpoint
Activity Powerpoint Speaker Notes
Website: Google Teachable Machine
Website: Teachable Machine Demo
Website: Teachable Machine Medium Article

Classroom Resources

Computer Projection System (Projector and Screen) to display content for students.

Activity resources:
- 2-4 bananas at different stages of ripeness.
- Printed photos of bananas at different stages of ripeness and a banana peel.
- Different colored paper.

Resources for Educators

- AI Fairness Introduction - IBM
- What Do We Do About the Biases in AI? - Harvard Business Review
- This is how AI bias really happens - MIT Technology Review
- Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2023 - AI Multiple
- Shedding light on AI bias with real world examples - IBM
- This is how AI image generators see the world - Washington Post
- Examples of AI Bias - Prolific
- Fairness and Bias in Artificial Intelligence - Literary Paper

Instructions

1. Video | AI Ethics: Fairness

Video: AI Fairness

Instructions
- Play video. The video will discuss the following:
    1. How is algorithmic bias integrated into AI systems either directly or by proxy?
    2. Why is the use of quality and representative training data important in producing fair outcomes and accurate predictions?
    3. What are some real-world examples of biased AI technologies?

2. Activity | Teachable Machine

Activity PPT   Activity PPT Speaker Notes   Website: Teachable Machine

Website: Teachable Machine Demo   Website: Teachable Machine Medium Article

Purpose: This activity will use Google's Teachable Mahcine as a tool for demonstrating how different types of data is used to train a machine learning model. This activity will demonstrate that even a simple model can produce inaccurate results based on the quality of data its fed. After this activity, students will have a better understanding of the role of training data in creating fair and accurate AI systems.

Instructions
- Educators should read the Teachable Machine Medium Article for context on the activity and follow along on the Activity Powerpoint for a step-by-step demo.
- For this activity, students or instructors will train a machine learning model by feeding the model data. For our module, the goal is to train the model to correctly recognize a banana's stage of ripeness. Once trained, the model can be tested using different types of visual inputs.
- Instructors should bookmark or save the Teachable Machine Demo website. This website links to a trained model that can be created using this activity. This model is well-trained and robust in recognizing different stages of a banana's ripeness. Instructors can use this as an example to demonstrate or test the model in class.

Note: This activity can be done in any of the three formats listed below.

1. Completed hands-on by students who are divided in groups.
2. Demonstrated by instructors in real-time, in a classroom.
3. Verbally explained by instructors using the provided PowerPoint.

The first two formats require extra resources listed in the "Classroom resources" section. The third format does not require any extra resources. We suggest using the first format since it's more conducive for hands-on and problem-based learning.

Module 3
AI Ethics: Privacy

Lesson Overview

This module will discuss how privacy-preserving AI laws and regulations can help protect users' personal and sensitive information in a data-driven world. Students will learn about the privacy risks associated with AI-powered social media algorithms and Generative AI chatbots. Students will also familiarize themselves with the impact of data breaches and data scraping on user privacy.

Subject
Mathematics, Computer Science

Target Grades
9 - 12

Estimated Duration
30 - 45 minutes

Vocabulary
- AI Privacy
- General Data Protection Regulation (GDPR)
- California Consumer Privacy Act (CCPA)
- The Right to be Forgotten
- Generative AI
- Anonymized Data
- Data Breach
- Sensitive Data
- Personal Data

Learning Objectives

By the end of the lesson, students will be able to:
1. Understand the privacy risks associated with AI-powered search engine and social media algorithms.
2. Understand the privacy risks associated with using Generative AI chatbots.
3. Recognize the privacy-preserving AI regulations that are meant to protect users personal data.
4. Understand how data breaches occur and their impact on user privacy.

Standards

AI4K12 Five Big Ideas in AI

3. Learning: Computers can learn from data.
4. Natural Interaction: Intelligent agents require many kinds of knowledge to interact naturally with humans.
5. Societal Impact: AI can impact society in both positive and negative ways.

CSTA K–12 Computer Science Standards

3A-IC-24: Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices.
3B-AP-08: Describe how artificial intelligence drives many software and physical systems.

ISTE

2a. Digital Citizen: Students cultivate and manage their digital identity and reputation and are aware of the permanence of their actions in the digital world.
2b. Digital Citizen: Students engage in positive, safe, legal and ethical behavior when using technology, including social interactions online or when using networked devices.
2d. Digital Citizen: Students manage their personal data to maintain digital privacy and security and are aware of data-collection technology used to track their navigation online.

Preparation

To prepare for this module beforehand, educators should watch the video, and review the Activity Powerpoint, Activity Powerpoint Speaker Notes, and Activity Instructions. Educators should also visit and familiarize themselves with the two websites that students will be using for this activity beforehand.

Important Links

Video
Activity Powerpoint
Activity Powerpoint Speaker Notes
Website: Have I Been Pwned
Website: Have I Been Trained

Classroom Resources

Computer Projection System (Projector and Screen) to display content for students.

Resources for Educators

- AI Privacy - Ericsson
- AI data scraping: ethics and data quality challenges - Prolific
- Scraping the Web Is a Powerful Tool. Clearview AI Abused It - WIRED
- Generative AI and Its Impact on Privacy Issues - DataGrail
- Generative AI's secret sauce — data scraping— comes under attack - VentureBeat
- OpenAI lawsuit reignites privacy debate over data scraping - CyberScoop
- Google hit with lawsuit alleging it stole data from millions of users to train its AI tools - CNN
- ChatGPT confirms data breach, raising security concerns - Security Intelligence
- How Americans View Data Privacy - Pew Research Center
- Beware the Privacy Violations in Artificial Intelligence Applications - ISACA
- Privacy and artificial intelligence - Literary Publication

Instructions

1. Video | AI Ethics: Privacy

Video: AI Privacy

Instructions
- Play video. The video will discuss the following:
    1. How can AI technologies unknowingly encourage the oversharing of personal data?
    2. What are the privacy risks associated with Generative AI chatbots like ChatGPT?
    3. How have AI technologies such as video surveillance, facial recognition, and Deep Fakes raised privacy concerns?

2. Activity | Pwned or Protected?

Activity PPT   Activity PPT Speaker Notes   Website: Have I Been Pnwed?

Website: Have I Been Trained?

Purpose: This activity will teach students about the impact of data breaches and data scraping on user privacy. Students will visit two websites: haveibeenpwned.com and haveibeentrained.com.

Instructions
For the first website: haveibeenpwned.com.
- The first website focuses on data breaches and their risk to our personal privacy. In this activity, students will type in their personal email address and their school email address on the homepage to see how many data breaches have been associated with both emails.
- Instructors can also ask students to visit the "Who's been pwned" tab of the same website (haveibeenpwned.com/PwnedWebsites). Here, students can browse the list of recent data breaches that have occurred at companies and see what kind of personal information has been exposed.
- Note: The older the email address is, the more data breaches students will find. Instructors can encourage students to also check their parent's email address for data breaches. Also, the school email address should have no data breaches because its protected with more safeguards than our personal email address.

For the second website: haveibeentrained.com.
- The second website focuses on how public image databases are used to train image synthesis AI models. These datasets scrape all public data from the internet without the user's consent or knowledge, which can pose a privacy risk. In this website, students can upload an image of themselves to see whether their face or photos have been unknowingly used to train AI models.
- Before doing the activity, instructors should re-visit a news story from the first module: "Woman horrified to discover her private medical photos were being used to train AI." This news story discusses how a woman named Lapine found that her private medical photos were being used to train AI. Lapine found her private photos in the public dataset by uploading an image of herself on the website haveibeentrained.com. This is the same website that students will use for the second part of the activity. Instructors should explain the connection between the news story and the second part of the activity.
- Note: students can also search their name or search using any text if they are not comfortable uploading photos of themselves to use the reverse-image search feature. However, uploading images will provide more interesting and narrowed results.

Module 4
AI Ethics: Transparency & Trustworthiness

Lesson Overview

This module will explore the role of transparency in building user trust towards AI systems. Students will understand the challenges of Black Box AI in Deep Learning Models and the role of Explainable AI in addressing the Black Box problem. Students will also learn about different types of Black Box technologies in their daily lives and how we can improve transparency in AI systems.

Subject
Mathematics, Computer Science

Target Grades
9 - 12

Estimated Duration
30 - 45 minutes

Vocabulary
- AI Transparency
- AI Trustworthiness
- Deep Learning Model
- Black Box AI / Black Box Algorithms
- Explainable AI

Learning Objectives

By the end of the lesson, students will be able to:
1. Understand the importance of transparency and how it can establish trust between the user and the AI technology.
2. Understand the dangers of Black Box AI such as Deep Learning Models.
3. Recognize that Explainable AI can build transparency and trust in Black Box models.

Standards

AI4K12 Five Big Ideas in AI

2. Representation & Reasoning: Agents maintain representations of the world and use them for reasoning.
4. Natural Interaction: Intelligent agents require many kinds of knowledge to interact naturally with humans.
5. Societal Impact: AI can impact society in both positive and negative ways.

CSTA K–12 Computer Science Standards

3A-IC-24: Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices.

3B-AP-08: Describe how artificial intelligence drives many software and physical systems.

ISTE

2b. Digital Citizen: Students engage in positive, safe, legal and ethical behavior when using technology, including social interactions online or when using networked devices.
3d. Students build knowledge by actively exploring real-world issues and problems, developing ideas and theories and pursuing answers and solutions.

Preparation

To prepare for this module, educators should watch the video, and review the Activity Powerpoint, Activity Powerpoint Speaker Notes, and Activity Instructions. To prepare for the classroom activity, educators should "test-run" the Thought Experiment and read the Blackbox article provided below.

Important Links

Video
Activity Powerpoint
Activity Powerpoint Speaker Notes
Activity Instructions
Website: Blackbox Article

Classroom Resources

Computer Projection System (Projector and Screen) to display content for students.

Resources for Educators

- AI's black box problem: Challenges and solutions for a transparent future - Coin Telegraph
- What is Transparency in AI? - Medium
- AI's mysterious 'black box' problem, explained - University of Michigan
- What is a black box? - The Conversation
- Transparent AI: Unveiling the Black Box for Ethical Decision-Making - LinkedIn

Instructions

1. Activity | Thought Experiment

Activity PPT   Activity PPT Speaker Notes   Activity Instructions   Website: Blackbox Article

Purpose: This activity will teach students about Black Box AI and what it means for an AI system to be non-transparent or untrustworthy. Students will also learn about the concept of a "trade-off" when using AI technologies. Instructors will conduct two thought experiments with students. The first will explore the perspective of an AI user whereas the second will focus on the perspective of an AI developer.

Instructions
For the first Thought Experiment: perspective of AI user.
- This thought experiment explores the accuracy vs. explainability tradeoff that an AI user might face when presented with the choice to use a BlackBox AI system.
- In this quick exercise, students will pretend that they need life-saving surgery. They will have to choose between two options for the surgery: the first option is surgery conducted by an autonomous AI robot and the second option is surgery conducted by a human doctor. Each option comes with a tradeoff: the surgery with the AI robot has a much higher chance of survival, but the AI cannot explain its reasonings or mistakes. On the other hand, the human doctor is able to explain what the surgery entails and any mistakes that were made during the procedure, but is less accurate with a lower chance of survival.
- Instructors can ask students to share answers to each experiment and explain their reasoning.

For the second Thought Experiment: perspective of AI developer.
- This thought experiment explores the tradeoff and challenges that an AI developer might face when presented with a BlackBox model.
- In this exercise, students will pretend to be AI developers who are tasked with deciphering the decision-making process of an unknown AI technology. The unknown AI technology will be represented as a Box that has string lights. In reality, this Box is meant to represent a Deep Learning Blackbox algorithm. Without revealing that the Box follows a Deep Learning model, instructors will slowly provide students with new information about the Box using the Activity Instructions and Activity PPT Speaker Notes. Every time the instructors reveals new information, students will have to decide whether they believe that they have enough information - as an AI developer - to figure out how the box goes from input to output (decision-making).
- At the beginning of the experiment, every student should stand up. After each step when new information is provided, the student should sit down if they believe that they can figure out the decision-making process of the Box.
- By the end of the activity, most students should be sitting down under the assumption that they can answer the question proposed based on all the information provided. However, despite knowing all the information about the box's internal components and wiring, students will learn that the true decision-making process of the box still cannot be determined because it's a Blackbox Deep Learning model. The Deep Learning model represents black box systems, which are not transparent in their decison-making operations.

2. Video | AI Ethics: Transparency/Trustworthiness

Video: AI Transparency/Trustworthiness

Instructions
- Play video. The video will discuss the following:
    1. What makes a machine trustworthy?
    2. What are the dangers of Black Box AI like Deep Learning Models?
    3. What are some examples of Black Box AI technologies in our daily lives?
    4. How can we make Black Boxes more transparent or build trust towards these systems?

Module 5
AI Ethics: Accountability

Lesson Overview

This module focuses on the importance of accountability in the design, development, and deployment of AI technologies. Students will learn about the stakeholders who are responsible for AI's ethical and legal impact using a real-world case-study. Students will also get an overview of Responsible AI practices that are used by companies to scale AI technologies.

Subject

Mathematics, Computer Science

Target Grades

9 - 12

Estimated Duration

30 - 45 minutes

Vocabulary

- AI Accountability
- Stakeholders
- Responsible AI
- AI Code of Ethics

Learning Objectives

By the end of the lesson, students will be able to:
1. Understand the role of stakeholders in addressing accountability when AI systems go wrong.
2. Explain the benefits of responsible AI.

Standards

AI4K12 Five Big Ideas in AI

1. Perception: Computers perceive the world using sensors
5. Societal Impact AI can impact society in both positive and negative ways.

CSTA K–12 Computer Science Standards

3B-AP-08: Describe how artificial intelligence drives many software and physical systems.
3B-IC-28: Debate laws and regulations that impact the development and use of software.

ISTE

2b. Digital Citizen: Students engage in positive, safe, legal and ethical behavior when using technology, including social interactions online or when using networked devices.
3d. Students build knowledge by actively exploring real-world issues and problems, developing ideas and theories and pursuing answers and solutions.

Preparation

To prepare for this module, educators should watch the video, and review the Activity Powerpoint, Activity Powerpoint Speaker Notes, and Activity Instructions. To prepare for the classroom activity, educators should browse through the AI Incident Database and bookmark/save a few AI incidents that might be interesting to the class.

Important Links

Video
Activity Powerpoint
Activity Powerpoint Speaker Notes

Classroom Resources

Computer Projection System (Projector and Screen) to display content for students.

Resources for Educators

- Artificial Intelligence and Ethical Accountability - Educause Review
- Accountability and Responsibility in AI - LinkedIn
- What is AI Accountability? - OECD.AI
- AI Accountability - IBM
- Automation, Ethics And Accountability Of AI Systems - Forbes

Instructions

1. Video | AI Ethics: Accountability

Video: AI Accountability

Instructions
- Play video. The video will discuss the following:
    1. What makes a machine trustworthy?
    2. What are the dangers of Black Box AI like Deep Learning Models?
    3. What are some examples of Black Box AI technologies in our daily lives?
    4. How can we make Black Boxes more transparent or build trust towards these systems?

2. Activity | AI Incident Database

Activity PPT   Activity PPT Speaker Notes   Website: AI Incident Database

Purpose: In this activity, students will learn about the current harmful outcomes of AI technologies in the real-world by browsing the AI Incident Databse. Students will explore the accountability measures taken after a AI technology causes a harmful or inaccurate outcome.

Instructions
- Instructors should first show students how to browse the database and navigate the main page. After opening the AI Incident Database, students should click on the "Discover Incidents" tab, which is the main page that includes the Incident Reports. The Activity PPT Speaker Notes document includes links a few sample AI Incidents that educators can use from this page.

This activity can be student-led, instructor-led, or conducted in small groups:

- Student-Led: Students should browse through the AI Incident Database individually or in small groups. Students should select 1-3 AI incidents that interest them and read about whether any accountability measures were taken after the incident (such as the AI system being shut down, a legal courtcase, or a regulatory consequence). Students can share incidents that they found interesting using the Incident ID. Instructors can then search for any of the shared AI incidents using the Incident ID and review the Incident details with the whole class.
- Instructor-Led: Instructors select a few AI incidents prior to the activity that are relevant and recent. During class, instructors can present these AI incidents to the class.
- Small Group Format: Instructors can divide students in different groups and assign each group a speicfic type of AI technology (i.e. autonomous vehicles, Generative AI chatbots, facial recognition, etc). Students can then work in groups to find the most recent AI incident related to that AI technology and see what steps were taken to address the incident. Groups can then share their findings with the class.

Note: The AI incident database collects and stores a history of incidents when an AI technology has caused harmful outcomes. The database stores a large collection of incident reports, each of which includes news information about the incident, the stakeholders involved, and a timeline on any follow-up steps that were taken to address or resolve the harmful outcome afterwards. Since this is a public database, there might be a few AI Incidents that are not appropriate to share out loud. Instructors can remind students to only share out appropriate and relevant AI incidents.

3. Activity | Student Question

Instructions
- Students will answer the question that they submitted during the first module about AI or AI ethics. Students should submit their question via Google Form or any other type of online form that educators can access.
- Instructors should encourage students to attempt their best to answer their question from Module 1. If students aren't able to answer directly, they can instead submit a response about their favorite, most interesting, or most surprising thing they've learned about AI/AI Ethics through the modules.
- Instructors can compare the questions and responses to guage the effectiveness of the classroom instruction and activities on student learning.

Glossary

Artificial Intelligence (AI):
The ability of machines to perform tasks that would normally require human intelligence, such as visual perception, speech recognition, decision-making, and language translation [1].

Machine Learning:
A subset of AI that involves training algorithms to learn from data and make predictions or decisions without being explicitly programmed [2].

Data Scraping:
A technique in which a computer program extracts data (outputs) generated from another program. This is most used in web scraping, which involves extracting valuable information from a website [3].

Deep Learning:
A type of machine learning that uses artificial neural networks with multiple layers to extract progressively higher-level features from data [4].

AI Ethics:
The study of legitimate and illegitimate uses of AI arising from the design, development, and deployment of AI systems. The AI Ethics framework includes principles such as fairness, privacy, transparency/trustworthiness, and accountability [5].

AI Fairness:
The AI Ethics principle that AI systems should not cause unfair outcomes or exhibit algorithmic bias towards individuals or groups [18].

AI Privacy:
The AI Ethics principle that AI systems should protect users privacy and regulate the use of personal and sensitive information [18].

AI Transparency/Trustworthiness:
The AI Ethics principle that the user and developer should clearly understand the inner workings, or decision-making processes, of an AI system [18].

AI Accountability:
The AI Ethics principle that individuals and organizations that are involved in the lifecycle of AI (design, development, and deployment) should be responsible for the ethical and legal impact of their AI technology [18].

Algorithmic Bias:
Systematic errors or inaccuracies in AI systems that result in unfair treatment or discrimination against individuals or groups based on race, gender, socioeconomic status, or ability [6].

Training Data:
The data used to train AI models before they're deployed [2].

Facial Recognition:
The use of AI systems to identify individuals based on their facial features, often used for security or surveillance purposes [5].

Predictive Policing:
The use of AI systems to predict where crimes are likely to occur or who is likely to commit a crime, often used by law enforcement agencies [5].

Sensitive Data:
Type of personal information that is considered confidential or highly protected under the law, such as political affiliation, medical records, or criminal history [5].

Personal Data:
Information that can be used to identify an individual, such as name, home address, phone number, or email address [7].

Anonymized Data:
Data that has been stripped of identifying information and cannot be linked back to an individual [8].

Generative AI:
An umbrella term for AI systems that can produce original content such as images, music, or text based on patterns learned from training. These systems are trained using extremely large datasets [4].

General Data Protection Regulation (GDPR):
A regulation passed by the European Union in 2016 that sets rules for how companies must protect the personal data of EU citizens [9].

California Consumer Privacy Act (CCPA):
A law passed by the state of California in 2018 that gives California residents certain rights over their personal data held by companies [10].

Right to be Forgotten:
A right in the European Union that states that individuals have a right to get their personal information removed from online databases or search engine results under certain circumstances [11].

Deep Fakes:
Synthetic media created using AI models that is designed to look real but is actually fake [12].

Black Box AI / Black Box Algorithms:
AI systems or algorithms whose inner-workings or decision-making processes are not transparent or understandable by humans [13].

Explainable AI:
Tools and frameworks that allow AI systems to provide explanations for their decisions or predictions in a way that is understandable by both users and developers [14].

Stakeholders:
Individuals or groups who have an interest in the development and deployment of AI systems, including developers, users, regulators, and affected communities [15].

Responsible AI:
The development and deployment of AI systems in a way that is ethical, transparent, accountable, and beneficial for society as a whole [16].

AI Code of Ethics:
A set of principles or guidelines for the ethical development and deployment of AI systems. Several technology companies have started to develop codes of ethics for AI, including Microsoft and IBM [17].

References
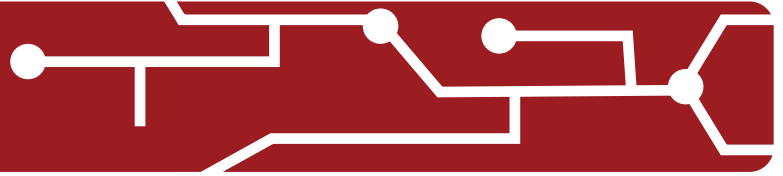
1. Adi Gaskell. (2018, April 18). *Automation, Ethics, And Accountability Of AI Systems.* Forbes. https://www.forbes.com/sites/adigaskell/2018/04/18/automation-ethics-and-accountability-of-ai-systems/?sh=5404b6cb3bc4

2. Artrepreneur. (n.d.). *The Next Rembrandt: Who Holds the Copyright in Computer-Generated Art?* Artrepreneur. https://artrepreneur.com/journal/the-next-rembrandt-who-holds-the-copyright-in-computer-generated-art/

3. BBC News. (2016, April 5). *New Rembrandt to be unveiled in Amsterdam.* BBC News. https://www.bbc.com/news/technology-35977315

4. BBC News. (2021, June 17). *AI-generated medical images pose new privacy concerns.* BBC News. https://www.bbc.com/news/technology-57588270

5. Code.org. (n.d.). Code.org. https://code.org/

6. Cointelegraph. (2021, May 28). *AI's black box problem: Challenges and solutions for a transparent future. Cointelegraph.* https://cointelegraph.com/news/ai-s-black-box-problem-challenges-and-solutions-for-a-transparent-future

7. Cloud Academy. (n.d.). *What is Artificial Intelligence?* Cloud Academy. https://cloud.google.com/learn/what-is-artificial-intelligence

8. David Tuffley. (2021, August 16). *What is a black box? A computer scientist explains what it means when the inner workings of AIs are hidden.* The Conversation. https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when-the-inner-workings-of-ais-are-hidden-203888

9. Dutch Digital. (n.d.). *The Next Rembrandt: Bringing the Old Master Back to Life.* Medium. https://medium.com/@DutchDigital/the-next-rembrandt-bringing-the-old-master-back-to-life-35dfb1653597

10. EDUCAUSE. (2020, July 22). *Artificial Intelligence and Ethical Accountability.* EDUCAUSE Review. https://er.educause.edu/blogs/2020/7/artificial-intelligence-and-ethical-accountability

11. Ericsson. (2022, August 23). *AI and privacy: Everything you need to know.* Ericsson. https://www.ericsson.com/en/blog/2022/8/ai-and-privacy-everything-you-need-to-know

12. Fair Bytes. (n.d.). *What is transparency in AI?* Fair Bytes. https://medium.com/fair-bytes/what-is-transparency-in-ai-bd08b2e901ac

13. Futurism. (n.d.). *Private Medical Photos Are Being Used to Train AI.* Futurism. https://futurism.com/the-byte/private-medical-photos-ai

14. Harvard Business Review. (2019, October 14). *What Do We Do About the Biases in AI?* Harvard Business Review. https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

15. Haas School of Business. (n.d.). *What is Fairness?* Haas School of Business. https://haas.berkeley.edu/wp-content/uploads/What-is-fairness_-EGAL2.pdf

16. IBM. (n.d.). *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks.* IBM. https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/
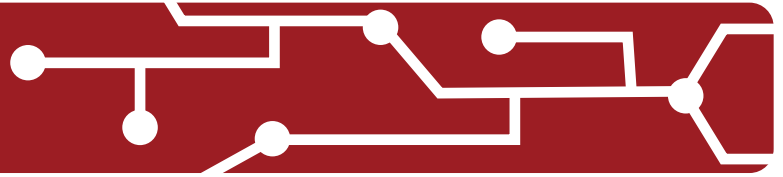
17. IBM. (n.d.). *Accountability.* https://www.ibm.com/design/ai/ethics/accountability/

18. Investopedia. (n.d.). *Artificial Intelligence (AI).* Investopedia. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

19. Jörg Land. (n.d.). *Transparent AI: Unveiling the black box of ethical decision-making.* LinkedIn. https://www.linkedin.com/pulse/transparent-ai-unveiling-black-box-ethical-decision-making/

20. Khan Academy. (n.d.). *AI for Education Unit 1.* Khan Academy. https://www.khanacademy.org/college-careers-more/ai-for-education/x68ea37461197a514:ai-for-education-unit-1

21. Kyle Wiggers. (2022, October 20). *Generative AI's secret sauce: Data scraping under attack.* VentureBeat. https://venturebeat.com/ai/generative-ai-secret-sauce-data-scraping-under-attack/

22. Leon Furze. (2023, January 26). *Teaching AI Ethics.* https://leonfurze.com/2023/01/26/teaching-ai-ethics/

23. Microsoft. (n.d.). *The Next Rembrandt.* Microsoft. https://news.microsoft.com/europe/features/next-rembrandt/

24. NBC News. (2019, November 5). *Self-driving Uber car that hit and killed woman did not recognize jaywalkers.* NBC News. https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

25. OECD. (n.d.). P9: *Accountability.* https://oecd.ai/en/dashboards/ai-principles/P9

26. Paul Veitch. (n.d.). *Accountability and Responsibility in AI: Assigning Age to Systems.* LinkedIn. https://www.linkedin.com/pulse/accountability-responsibility-ai-assigning-age-systems-paul-veitch/

27. Princeton University. (n.d.). *Case Study PDFs. AI Ethics.* https://aiethics.princeton.edu/case-studies/case-study-pdfs/

28. Prolific. (n.d.). *AI Data Scraping: Ethics and Data Quality Challenges.* Prolific. https://www.prolific.com/blog/ai-data-scraping-ethics-and-data-quality-challenges#:~:text=Even%20outside%20of%20consent%20concerns,and%20bias%20in%20their%20output.

29. Shannon Vavra. (2021, November 2). *OpenAI faces lawsuit over privacy, data scraping.* CyberScoop. https://cyberscoop.com/openai-lawsuit-privacy-data-scraping/

30. Simplilearn. (n.d.). *Artificial Intelligence Tutorial.* Simplilearn. https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial

31. Smithsonian Magazine. (2021, October 22). *Lost Edges of Rembrandt's "Night Watch" Are Restored Using Artificial Intelligence.* Smithsonian Magazine. https://www.smithsonianmag.com/smart-news/lost-edges-rembrandts-night-watch-are-restored-using-artificial-intelligence-180978056/

32. The Atlantic. (2018, May 7). *Uber's Self-Driving Car Didn't Malfunction, It Was Just Bad.* The Atlantic. https://www.theatlantic.com/technology/archive/2018/05/ubers-self-driving-car-didnt-malfunction-it-was-just-bad/561185/

33. The Guardian. (2016, April 5). *New Rembrandt to be unveiled in Amsterdam.* The Guardian. https://www.theguardian.com/artanddesign/2016/apr/05/new-rembrandt-to-be-unveiled-in-amsterdam

34. The New York Times. (2020, September 15). *Uber's Self-Driving Cars Were Struggling Before Arizona Crash.* The New York Times. https://www.nytimes.com/2020/09/15/technology/uber-autonomous-crash-driver-charged.html

35. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford, and Karen Hao. (2023). *AI Bias. Annual Review of Statistics and Its Application*, 10(1), 1-31. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9558020/

36. Towards Data Science. (n.d.). *Understanding Bias and Fairness in AI Systems.* Towards Data Science. https://towardsdatascience.com/understanding-bias-and-fairness-in

37. University of Michigan-Dearborn. (n.d.). *AI's mysterious black box problem explained.* University of Michigan-Dearborn. https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

38. *What is the Length of a Banana?* [2023]. (n.d.). Meal Delivery Reviews. https://ageekoutside.com/what-is-the-length-of-a-banana/

39. *What is a black box? A computer scientist explains what it means when the inner workings of AIs are hidden.* (2021, March 17). Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/ethics-ai/

40. *What Is Transparancy in AI?* (n.d.). Levity. https://levity.ai/blog/ai-bias-how-to-avoid

Curriculum Development Team

Project Manager

Shubhi Sinha, *Lead Intern at IU Center of Excellence for Women and Technology*

Content Contributors

Amy Kinney, *Intern at IU Center of Excellence for Women and Technology*

Jiyoung Kim, *PhD student at Indiana University Bloomington*

Juliette Albert, *Designer and Intern at IU Center of Excellence for Women and Technology*

ICICLE and CEW&T Teams

Maureen Biggers, *ICICLE Team*

Beth Plale, *ICICLE Co-Principle Investigator at Indiana University Bloomington*

Sadia Khan, *ICICLE Research Team and PhD student at Indiana University Bloomington*

Michelle Bartley-Taylor, *Director at IU Center of Excellence for Women and Technology*