

Ethics Check

Creating with Purpose

Grade Level: Suitable for 6-12

Suggested Time: 1 Hour

Overview

Students explore how AI developers communicate transparency and accountability through model cards and responsible use cards. They examine real-world examples, then apply ethical thinking to their own AI projects by creating the appropriate documentation. The lesson closes with a discussion on explainability and transparency in AI systems.

Overview	1
AI Learning Priorities	2
Knowledge and Skills Topics	2
Materials and Resources Needed	2
Differentiation Suggestions	3
Evidence of Understanding	4
Key Concepts	4
Activity Overviews.....	4
Sequence, Session Flow, and Facilitation Prompts.....	5
Additional Resources	12

AI Learning Priorities

Humans and AI: Humans are responsible for designing AI systems that are fair, transparent, and accountable to the people they affect.

Representation and Reasoning: The data used to train AI models shapes what those models can and cannot do, including who they may or may not work well for.

Machine Learning: Machine learning models should be documented so that others can understand how they were trained, what data was used, and how accurate they are.

Ethical AI: Developers have a responsibility to identify and communicate the limitations and potential biases of the AI systems they create.

Societal Impacts of AI: AI systems that lack transparency or explainability can affect individuals and communities in ways that are difficult to detect or challenge.

Knowledge and Skills Topics

- What a model card is and why it matters
- The difference between a model card and a responsible use card
- How to identify bias, limitations, and intended use in an AI system
- The concepts of explainability and transparency in AI

Materials and Resources Needed

- ☐ Student devices with a web browser and internet access. This works best on a laptop rather than a phone or tablet.
- ☐ Whitelisted items
 - ☐ [Gemma Model Card](#)
 - ☐ [MediaPipe FaceMesh Model Card](#)
 - ☐ [Object Detection Model Card](#)
 - ☐ [Example Responsible Use Card](#)
 - ☐ [Responsible Use Card Template](#)
 - ☐ [Model Card Template](#)

Curriculum Standards

- **ISTE 1.3.b Knowledge Constructor:** Students evaluate the accuracy, validity, bias, origin, and relevance of digital content.
- **ISTE 1.3.d Knowledge Constructor:** Students build knowledge by exploring real-world issues and gain experience in applying their learning in authentic settings.
- **CSTA 2-IC-20:** Compare tradeoffs associated with computing technologies that affect people's everyday activities and career options.
- **CSTA 3A-IC-24:** Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices.
- **CSTA 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.
- **AI4K12 Big Idea 3 Learning:** AI systems learn from data and we must be clear about how that data is used.
- **AI4K12 Big Idea 4 Natural Interaction:** AI systems interact with humans in ways that require transparency.

Differentiation Suggestions

- For students who need support
 - Ensure students are in a group with supportive peers.
 - Provide a sentence starter bank for filling in card sections
- For Multilingual Learners
 - Allow use of translation tools for card completion
 - Provide a visual glossary of key terms (bias, transparency, explainability, accuracy)
- For students who are ready to go further
 - Have students compare their card to a real Google model card and reflect on the differences
 - Ask students to write a short argument for or against requiring model cards by law
 - Assign the extension activities: Survival of the Best Fit or The Moral Singularity

Evidence of Understanding

- Completed model card or responsible use card that accurately reflects their Samsung Solve for Tomorrow technology
- Participation in the group discussion on explainability and transparency
- Ability to explain the difference between a model card and a responsible use card
- Quality of reasoning in identifying limitations or biases in their own project

Key Concepts

- Model Card - A document describing an AI model's intended use, performance, limitations, and potential biases
- Responsible Use Card - A document reflecting ethical AI use based on principles like IBM's 5 Pillars, used when no dataset is involved
- Bias - Unfair skew in an AI system's outputs, often caused by unrepresentative training data
- Explainability - The ability to clearly communicate how an AI system arrived at a result
- Transparency - Openness about how an AI system works, what data it uses, and how decisions are made
- Accountability - Taking responsibility for the outcomes an AI system produces

Activity Overviews

Activity #1: Learn About Model Cards

Purpose: Become familiar with the purpose and style of model cards.

Task: Students will have 5 minutes to review three examples of model cards from Google.

[Gemma Model Card](#)

[MediaPipe FaceMesh Model Card](#)

[Object Detection Model Card](#)

Activity #2: Build Your Card

Purpose: Document your project's ethics.

Task: Students will have 30 minutes to demonstrate the ethical considerations involved in their Solve for Tomorrow technology. If the project leans more toward the creative side and does not rely on a dataset, the student will make a responsible use card. If the project uses data, regardless of the source of the data, the student will create a model card. All students will watch a brief video to see how the model card should be completed.

Sequence, Session Flow, and Facilitation Prompts

Sequence

1. Opening
2. Ethics
3. Activity #1 - Learn About Model Cards
4. Activity #2 - Build Your Card
5. Explainability & Transparency
6. Wrap-Up

Session Flow

1-Hour Class Period

Opening	0:00 - 0:02	Slide 1; Low stakes
Ethics	0:02 - 0:09	Slide 2; Direct instruction
Activity #1	0:09 - 0:18	Slides 3-4
Activity #2	0:18 - 0:43	Slides 5-8
Explainability & Transparency	0:43 - 0:50	Slide 9; Direct instruction
Wrap-Up	0:50 - 1:00	Slide 10

45-Minute Class Period

STOP		Give Pre-Reading the night before
Opening	0:00 - 0:02	Slide 1; Low stakes
STOP		SKIP to Activity #1
Activity #1	0:02 - 0:11	Slides 3-4
Activity #2	0:11 - 0:43	Slides 5-8
STOP		Give homework SKIP to Wrap-Up
Wrap-Up	0:43 - 0:45	Slide 10

Facilitation Prompts

[SLIDE DECK](#)

SLIDE 1 - Title Slide

Say “Up until now, you’ve been learning how to use AI tools, how to prompt them, how to visualize data with them, and how to build things with them. Today, we’re going to zoom out and ask a bigger question: just because we CAN build something with AI - should we? And if we do build it, how do we make sure it’s built responsibly?”

This lesson is called the Ethics Check. And it’s not optional - a completed ethics check is a required part of your Samsung Solve for Tomorrow technology submission. By the end of this class, your team will have produced the documentation that proves you’ve thought carefully about the ethical implications of your own project.

SLIDE 2 - AI is Already Making Decisions About You

Say “Let’s start with getting a job. Many companies - especially large ones - don’t have a human being read your resume first. An AI system reads it and decides whether you even make it to the next round. And here’s the problem: if that model was trained on historical hiring data from a company that mostly hired from certain schools, or certain zip codes, or certain backgrounds - the model learned those patterns. It might reject a fully qualified candidate because

they went to a school the model hasn't "seen" before, or because their name sounds unfamiliar. No one intended that. But it happens.

Now look at healthcare. AI tools are used to help doctors diagnose illness, predict which patients might be at risk for certain diseases, and recommend treatments. That sounds great - until you realize that a lot of medical AI was trained mostly on data from one type of patient. So when the same tool is used on someone from a different demographic group, it performs worse. That's not a hypothetical. There's documented research showing that certain diagnostic AI tools were less accurate for patients with darker skin tones because the training data skewed toward lighter skin. That's a real consequence.

What about music and creative tools? A lot of you are probably already using AI tools to generate images, music, or videos. Here's something worth thinking about: those tools were trained on existing human creative work - often scraped from the internet, often without asking the original creators for permission. So when you use a tool that generates music "in the style of" a particular artist, you're benefiting from that artist's labor. The question of who deserves credit - and who gets paid - is very much unresolved.

Games and recommendations - this one you probably feel in your daily life. Game matchmaking, your YouTube feed, your TikTok algorithm, your Spotify recommendations - all AI. These systems are optimizing for one thing above everything else: engagement. Keeping you watching, playing, scrolling. That's not inherently evil, but it can mean the algorithm pushes content that keeps you hooked rather than content that's actually good for you. That's a design choice baked into the model.

Finally, criminal justice. Some courtrooms in the United States use AI tools to predict whether a defendant is likely to reoffend in the future - and those predictions influence sentencing and parole decisions. A study called the COMPAS investigation found that one of these tools flagged Black defendants as higher risk at nearly twice the rate of white defendants, even when controlling for other factors. That's not a minor bug - that's a system that affects whether someone goes home or stays in prison.

So when we talk about AI ethics - we're not talking about something abstract or philosophical. We're talking about real decisions, affecting real people, right now. And increasingly, you're the ones who will be building these systems.

SLIDE 3 - Activity #1 Learn About Model Cards

Say "Okay - now that we've established why this matters, let's talk about one of the most important tools in AI ethics: the model card.

A model card is a document that describes a machine learning model. It tells you the model's intended use, how well it performs, what its limitations are, and what biases might be present in its training data.

You can think of it as the owner's manual for an AI system. Or - and this is the analogy we'll come back to in a minute - think of it as a nutrition label for AI.

Here's why model cards matter: without them, users of an AI system have no idea what they're really working with. Is this model accurate for people like me? Was it tested on a diverse dataset? What happens when it's wrong? A model card answers those questions.

On the right side of the slide you'll see three real examples from Google. I want you to take a few minutes to click through and look at them briefly - you don't need to read every word.

[Gemma Model Card](#)

[MediaPipe FaceMesh Model Card](#)

[Object Detection Model Card](#)

Notice: what kinds of information do they include? You have 5 minutes. Go.

Okay, what did you notice? What kinds of things were in those model cards?

[Students might mention: accuracy metrics, limitations sections, information about the training data, intended uses vs. out-of-scope uses.]

Good. Now notice that those model cards from Google are very detailed - they're long documents. You are not going to be asked to produce anything like that. What you're creating today is a streamlined, student-friendly version that captures the core ideas. We'll look at exactly what goes in it in a moment.

SLIDE 4 - The Nutrition Label Analogy

Say "Let's talk about why this analogy works so well.

When you pick up a box of cereal, the nutrition label tells you what's inside -

ingredients, serving size, what percentage of your daily needs each nutrient covers. It also tells you what to watch out for: high sodium, added sugar, common allergens.

A model card does the exact same thing for an AI model. It tells users: what is this model designed to do? What data was it trained on? Where might it fail? How accurate is it on a test set?

And here's the key point from the slide: without that label, you're consuming something you know nothing about. Imagine if food companies weren't required to disclose ingredients. You'd have no idea if you were allergic to something, no idea how much sugar you were eating, no idea where the food came from. That's what deploying AI without documentation looks like.

Now look at the right side of this slide. This is important. Whether you're building AI for music, games, healthcare, or anything else - your choices matter.

The card you create today is not busywork. It is your team's formal commitment to building something that is fair, honest, and accountable. It is your way of saying: we thought about who this affects, we know where our data came from, we understand what our model can and can't do.

And to be clear: an ethics check is required for your Samsung Solve for Tomorrow technology project. This is not optional. This is why.

This is what responsible builders do. And that's what you're going to be today.

SLIDE 5 - Model Card Instructional Video

Say "Before you start filling in your card, I want to show you a short video that walks you through how to complete the model card template. This will make the activity much clearer.

[Watch video, approx 6 minutes]

SLIDE 6 - Responsible Use Card

Say "You're going to create the ethics documentation for your Samsung Solve for Tomorrow project. But first, you need to decide which kind of card you need.

There are two options. Let me explain both.

Option 1 - Responsible Use Card. Use this if your project is on the creative side and

does NOT use a dataset that was trained. So for example: if you're using an AI tool like a generator, a chatbot, or an image creator, but you didn't train a model on your own data - this is your card.

The Responsible Use Card is based on IBM's 5 Pillars of Responsible AI. It asks you to think about how you used AI responsibly in your project - things like fairness, transparency, and privacy. You don't have to fill in accuracy percentages or talk about training data, because that's not what your project involves. Instead, you're documenting your choices as a user.

[Example Responsible Use Card](#)
[Responsible Use Card Template](#)

SLIDE 7 - Dataset Model Card

Say "Option 2 - Model Card. Use this if your project uses data in any form. If you trained a model using Machine Learning for Kids, Teachable Machine, or any other platform - or if you downloaded a dataset and used it - this is your card.

The model card documents your model's name, accuracy, limitations, data source, and the features and labels that drive its predictions. It's the formal way of saying: here's exactly how this model works and where it might fall short.

[Model Card Template](#)

SLIDE 8 - Activity #2

Say "Okay - this is your main work time for today.

Regardless of which card you're doing, the steps are the same. Don't rush through this.

SLIDE 9 - Explainability & Transparency in AI

Say "Before we get to our final discussion, I want to make sure we're clear on two important concepts that underpin everything we've talked about today: explainability and transparency. These sound similar, but they mean different things.

Explainability is about being able to communicate how an AI system arrived at a result - even when the outcome isn't what you hoped for.

Think about school. Imagine you got a B on an assignment. If your teacher gave you a rubric at the start, told you exactly what would be graded and how - you

understand why you got a B. You might not be happy about it, but you can see the logic. That's an explainable system. You know how the outcome was produced.

Now imagine a teacher who just handed back a grade with no rubric, no feedback, no explanation. Same B. But now you have no idea what to improve, whether the grading was fair, or whether another student doing similar work got a different result. That's an unexplainable system.

AI that makes decisions without being able to explain them - like loan approvals, hiring decisions, medical diagnoses - creates the same problem at a massive scale.

Transparency is about openness around where the AI's source material came from and whether the process is consistent and fair for everyone.

The analogy on the slide uses AP exams. An AP exam is the same for every student across the country. The curriculum is public. The test is standardized. That's transparent - you know what to expect, and the playing field is level.

Compare that to an elective class where each teacher designs their own curriculum, grades differently, and covers different topics. There's nothing wrong with that - but if you're trying to compare students across schools, you don't have a transparent system.

Transparency in AI means: is this model's process visible and consistent? Who built it, what data did they use, and would another team using the same data get the same result?

And here's the connection back to model cards: creating one is an act of both explainability and transparency. You're saying: here's what our model does, here's how it does it, here's where it might fail. That's what responsible AI documentation looks like.

SLIDE 10 - Group Discussion

Say "We're going to close today with a question. There's no right answer here.

Do you think AI systems that make decisions should have user protections that ensure privacy, transparency, and explainability? Feel free to dig into the sub-questions or go straight into the main one.

These are real debates happening right now in policy and in the tech industry.

Additional Resources

- Survival of the Best Fit <https://www.survivalofthebestfit.com/>
- The Moral Singularity <https://themoralsingularity.org/>
- "[Machine Bias](#)" - ProPublica - The original investigative journalism piece on the COMPAS criminal justice algorithm. Directly supports the criminal justice example discussed in Slide 2 and is accessible to high school students.
- [Gender Shades Project](#) - MIT Media Lab - Joy Buolamwini's research showing facial recognition systems performed significantly worse on darker-skinned women. Short, visual, and hits hard for students.

Created in partnership with the Mark Cuban Foundation, markcubanai.org