

Ethics Check: Homework

Creating with Purpose

Explainability & Transparency in AI

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.

Prompts

Do you think AI systems that make decisions should have user protections that ensure privacy, transparency, and explainability?

1. What are the benefits of requiring companies to publish model cards?
2. Why might a company choose not to be transparent about how their AI works?
3. Can you think of a real example where a lack of AI transparency caused harm?

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