

Knowledge and Technology: AI, Knowledge, and Accountability

Lesson 2: Fact-Finding Task

Focus: *Who is responsible for real-world AI failures?*

Objectives	Critically analyze real-world AI failures. Identify biases, evaluate evidence, and refine arguments about AI accountability.
Homework Preparation Tasks	<p>Case Study Task</p> <ul style="list-style-type: none">• Discussion Prompt: Who is responsible for real-world AI failures?• Divide students into small groups and assign each group a case study related to the topic. Suggestions are listed below. Students will add their evidence to the Kialo discussion from Lesson 1. <p>Each group will:</p> <ul style="list-style-type: none">• Reflect on how these cases connect to the concepts discussed in Lesson 1.• Explore their assigned case using the provided resources (articles, videos, or curated primary sources).• Prepare a short presentation (5–10 minutes). <p>Case Study Options</p> <p>1. Tesla's Autopilot Failure</p> <ul style="list-style-type: none">• Focus: The real-world consequences of AI-assisted driving and the blurred line between human control and AI autonomy.• Task: Examine to what extent Tesla users, the company, and the AI system itself are responsible for accidents involving Autopilot.• Resources:<ul style="list-style-type: none">◦ Tesla's Autopilot and Full Self-Driving linked to hundreds of crashes, dozens of deaths - The Verge◦ Tesla CEO Elon Musk responds to Texas crash in investigation into two deaths <p>2. AI Leadership</p> <ul style="list-style-type: none">• Focus: The responsibilities of AI leaders in anticipating the impact of advanced AI systems, such as AGI.• Task: Evaluate the ethical responsibilities of AI CEOs and developers in preparing for possible future failures of high-stakes AI systems.• Resource:<ul style="list-style-type: none">◦ How OpenAI's Sam Altman Is Thinking About AGI and Superintelligence in 2025 TIME <p>3. User Responsibility</p> <ul style="list-style-type: none">• Focus: Platforms are holding users liable when AI tools generate offensive or false content on social platforms.• Task: Analyze whether it is fair for users to bear responsibility for AI-generated content, especially when they may not fully understand or control how the tools work.• Resource:<ul style="list-style-type: none">◦ Warning: If AI social media tools make a mistake, you're responsible Technology EL PAÍS English <p>4. Medical Note-Taking</p>

	<ul style="list-style-type: none"> • Focus: Doctors are increasingly relying on AI tools to transcribe and summarize patient consultations, aiming to reduce administrative burden and burnout. • Task: Examine who should be held accountable if an AI note-taking tool introduces an error that leads to a medical misdiagnosis or improper treatment. • Resource: <ul style="list-style-type: none"> ◦ Doctors turn to AI for easier medical note-taking
Activities	<ol style="list-style-type: none"> 1. Introduction (5 mins) <ul style="list-style-type: none"> • Recap Lesson 1: <ul style="list-style-type: none"> ◦ Review key arguments from the debate on AI responsibility. ◦ Discussion questions: <ul style="list-style-type: none"> ■ Who should be held accountable when AI systems fail — programmers, users, or the AI itself? ■ How do different contexts (medical, legal, transportation, social media) affect judgments about responsibility? ■ Can an AI system ever be morally or legally responsible? • Present the task's central question: <i>Who should be held responsible when AI systems fail?</i> 2. Bridge to Lesson 2 <ul style="list-style-type: none"> • Explain that students will now investigate real-world examples of AI systems and failures to explore how responsibility is constructed, challenged, and distributed among human and non-human agents. 3. Presentations (30 mins) <ul style="list-style-type: none"> • Students present their case studies to the class. • Students should take note of any useful points from other groups' presentations to use in the Kialo discussion. 4. Recording Findings in a Kialo Discussion (20 mins) <ul style="list-style-type: none"> • Students use their case study and their peers' presentations to update and substantiate their arguments in their Kialo discussion from the previous session, focusing on: <ul style="list-style-type: none"> ◦ Accountability in AI: Who is ultimately responsible when AI systems make mistakes? ◦ Power and ethics: Who has the authority to define responsible AI use — governments, companies, or individuals? ◦ Human vs machine agency: Can an autonomous system be meaningfully "responsible"? ◦ Bias and justification: Are AI errors more forgivable or less forgivable than human ones, and why?
Reflection Questions	<p>Reflection (10 mins): Discuss the following reflection questions in open discussion or exit ticket format:</p> <ul style="list-style-type: none"> • If users are thoroughly informed about an AI tool's limitations but still rely on it, how does that affect their share of responsibility when errors occur? • What biases are evident in these cases? • How does evidence shape our understanding of accountability in AI systems?
Resources	<p>Lesson Slides</p> <p>Kialo discussion: If an AI system makes a mistake, who is responsible: the programmer, the user, or the AI itself?</p> <p>Articles:</p> <p>Elearnia - AI Accountability: Who is Responsible When AI Fails?</p> <p>How OpenAI's Sam Altman Is Thinking About AGI and Superintelligence in 2025 TIME</p> <p>Tesla's Autopilot and Full Self-Driving linked to hundreds of crashes, dozens of deaths - The Verge</p> <p>Tesla CEO Elon Musk responds to Texas crash in investigation into two deaths</p>

	<p>Doctors turn to AI for easier medical note-taking Warning: If AI social media tools make a mistake, you're responsible Technology EL PAÍS English</p> <p>Videos: Ethics of AI: Challenges and Governance Ethics in the Age of AI Davos 2024 World Economic Forum</p>
TOK Concepts	<p>Evidence: How does evidence shape our understanding of accountability in AI systems? Justification: How do the ethical justifications used by companies to implement AI systems hold up in the face of AI failures? Power: In what ways do powerful stakeholders (e.g., tech companies, governments) shape the narrative around AI failures?</p>
Critical Thinking Concepts	<ul style="list-style-type: none"> ● Confronting Biases and Assumptions: <ul style="list-style-type: none"> ○ Identifying Misinformation and Spin: Train students to detect biased or misleading reporting in articles and videos. ○ Questioning Initial Beliefs: Encourage them to challenge their preconceptions (e.g., “I assumed Tesla’s software was foolproof”), asking why they hold these views. ● Exploring Contexts and Expert Opinions: <ul style="list-style-type: none"> ○ Stakeholder Analysis: Have groups consider how various stakeholders (engineers, regulators, consumers) might shift blame or responsibility. ○ Evaluating Sources: Guide students to compare expert perspectives—journalists, tech CEOs, policy makers—to pinpoint credible vs. questionable evidence. ● Responsiveness and Flexibility of Thought: <ul style="list-style-type: none"> ○ Adapting Arguments: Urge students to refine or modify their initial arguments about accountability as they uncover new data. ○ Metacognitive Check-Ins: After reviewing sources, encourage them to note how their thinking has evolved and why. ● Extrapolation and Reapplication of Principles: <ul style="list-style-type: none"> ○ Broadening the Lens: Challenge students to apply lessons from these AI cases (responsibility, bias) to other emerging technologies or controversies. ○ Transferable Frameworks: Emphasize how the same investigative process (collecting evidence, identifying bias, drawing conclusions) can be repeated in various disciplines.