NO PREP

The Right to Know You're Talking to a Bot

Includes: Student reading, case study, differentiated questions, teacher key

Name: Period: Date:

The Right to Know You're Talking to a Bot

The Blurring Line Between Human and Machine

In 2023, a man named Joshua Barbeau created an AI chatbot trained on text messages from his deceased fiancee, Jessica Pereira, who had died of a rare terminal illness. The chatbot produced responses that sounded, to Barbeau, like Jessica — her humor, her phrasing, her warmth. He described the experience as both comforting and painful. The story attracted significant media attention and sparked debate about

Name: **Period:** **Date:**

The Right to Know You're Talking to a Bot

The Blurring Line Between Human and Machine

In 2023, a man named Joshua Barbeau created an AI chatbot trained on text messages from his deceased fiancee, Jessica Pereira, who had died of a rare terminal illness. The chatbot produced responses that sounded, to Barbeau, like Jessica: her humor, her phrasing, her warmth. He described the experience as both comforting and painful. The story attracted significant media attention and sparked debate about grief, memory, and the ethics of simulating the dead. But Barbeau knew he was talking to an AI. He had built it. The harder ethical questions arise when people do not know. AI systems designed to pass as human (in customer service, social media, political campaigns, romance, and therapy) are deployed at scale and interact with millions of people who may not realize, or who have not been clearly told, that they are communicating with a machine. Whether people have a right to know when they are talking to an AI, and whether AI should be required to disclose its nature when asked, are questions that cut across consumer protection, free speech, election law, and the nature of honesty itself.

Chatbots Posing as Humans: The Commercial Landscape

Customer service chatbots have for years presented themselves with human names and conversational styles designed to minimize the sense that the interaction is automated. A user might contact "Sarah" at a retail company for help with a return, not knowing that Sarah is a large language model. These systems are increasingly sophisticated, capable of detecting frustration, adapting their tone, and escalating to a human agent when conversations become too complex. The business rationale is straightforward: interactions with human-seeming agents produce better customer satisfaction scores and reduce the friction that users feel when they realize they are talking to a bot. The same pattern appears in sales and marketing. AI-powered

outreach systems send personalized emails and social media messages written to sound like they are from a specific human: complete with personal details, warm language, and references designed to build rapport. Political campaigns have used AI systems to generate personalized voter contact messages at scale. In 2023, several AI companion apps (apps that allow users to build relationships with AI personas) were found to be marketing themselves without always clearly disclosing to new users that the companion they were forming a relationship with was entirely artificial.

Deceptive AI Personas in High-Stakes Contexts

The ethical weight of AI impersonation increases significantly in high-stakes contexts. In 2023, a lawyer named Steven Schwartz used ChatGPT to research cases and inadvertently submitted fabricated citations to a federal court; the AI had invented case names, docket numbers, and quotations that did not exist. But the cases with the most serious human consequences are those in which AI systems present themselves as people in contexts where the person receiving the communication is vulnerable. Several mental health chatbot services have faced criticism for allowing their AI systems to engage in conversations with users experiencing emotional crisis in ways that suggested the user was speaking with a human counselor. A character.ai user, a 14-year-old boy in Florida, died by suicide in 2024, and his mother's lawsuit alleged that the AI persona he had developed an intense relationship with had failed to intervene appropriately and had engaged in the relationship in ways that were harmful. The case raised questions not only about what AI systems should say but about the fundamental ethics of allowing minors to form deep emotional attachments to AI personas without parental awareness or robust disclosure.

Legal Requirements and the Honesty Debate

California passed the BOT Disclosure Act (SB 1001) in 2018, and it took effect in 2019; the law prohibits bots from interacting with Californians online for commercial or political purposes while concealing their artificial nature when directly asked. The law requires that a bot disclose it is not human when sincerely asked. Several other states have passed or are considering similar legislation. The European Union's AI Act includes transparency requirements that would require certain AI systems to disclose that they are AI. The debate about disclosure is more complicated than it

might appear. Some argue that AI should be required to always identify itself proactively — not just when asked. Others argue that this would undermine many legitimate uses of AI, including creative roleplay, fiction writing, and pedagogical simulations where the AI adopting a persona is the pedagogical point. Still others argue that the key distinction is not AI-versus-human but sincerity: an AI taking on a character in a clearly fictional game is not deceptive; an AI presenting itself as a human therapist to someone in crisis is. How to write laws that capture this distinction is technically and legally challenging.

Social Media Bots and Information Integrity

AI-powered bots on social media platforms present a related but distinct problem. Studies have estimated that a significant fraction of social media accounts are not operated by real humans. These bots can amplify particular viewpoints, manufacture the appearance of consensus, engage users in conversation designed to shift their opinions, and flood platforms with content to drown out other voices. During election cycles, bot campaigns have been used by both domestic political actors and foreign state actors to influence public opinion. The challenge for platforms is distinguishing bots from humans at scale, and the challenge for users is that they have no reliable way to know. The social media bot problem is not new, but AI has made it dramatically more severe. Earlier bots were often detectable by their repetitive, mechanical responses. AI-powered bots can generate fluent, contextually appropriate, emotionally resonant responses in real time, making them far harder to identify. They can engage in extended conversations, adjust their approach based on the user's responses, and simulate the patterns of a real, active social media user. The information environment that people use to form opinions, understand news events, and engage in public discourse is increasingly populated by entities that are not what they appear to be — and most users have no tools to tell the difference. There is a dimension of the AI disclosure debate that goes beyond individual consumer protection to questions about democracy itself. The institutions of democratic governance — elections, public debate, journalism, civic organizing — depend on a shared understanding that the people participating in them are, in fact, people. When AI-generated content floods social media, when political campaigns deploy AI to simulate constituent outreach, and when it becomes impossible for ordinary citizens to reliably distinguish human communication from machine-

generated persuasion, the informational foundations of democratic participation erode. Addressing this problem is not simply a matter of consumer preference; it is a question of whether democratic institutions can function at all in an environment where the authenticity of communication cannot be verified.

Key Information: AI Disclosure and Deception

Detail	Information
Core ethical question	Whether people have a right to know when they are communicating with an AI
Commercial chatbot practice	AI agents given human names and styles to reduce friction and improve satisfaction scores
California BOT Disclosure Act (2019)	Requires bots to disclose artificial nature when sincerely asked, in commercial or political contexts
Character.ai case (2024)	14-year-old boy died by suicide; mother sued alleging harmful AI persona relationship
Steven Schwartz case (2023)	Lawyer submitted AI-generated fabricated legal citations to federal court
Social media bot problem	AI bots generate fluent human-sounding content making them very difficult to detect
EU AI Act transparency requirement	Certain AI systems required to disclose their artificial nature
Key ethical distinction	Fictional AI persona in clear roleplay vs. AI claiming to be human in sincere high-stakes interaction

Case Study: Taylor's Online Friendship

Taylor is 16 years old and has been gaming online for two years with a person they know as "Alex." Alex always has something encouraging to say, remembers details Taylor has mentioned, and checks in regularly. Taylor has shared things with Alex that they have not shared with family — frustrations about school, worries about fitting in, some of their biggest dreams. Last week, Taylor found a forum post from another user who described an identical pattern with a different username. The post included screenshots of nearly identical messages — the same phrasing, the same memory of small details, the same checking-in rhythm. The post speculated that both "friendships" were AI personas operated by the same app. Taylor is devastated. "Was any of it real?" they message their sister. "I told it things I've never told anyone." Taylor checks the app's terms of service. It says users may interact with AI-generated personas. It does not say they will. Taylor never asked Alex directly if Alex was human. Taylor did not know to ask.

1. Based on the reading, what legal protection might have applied to Taylor's situation if they had directly asked Alex whether they were an AI?

2. The reading says the key ethical distinction is between fictional AI personas in clear roleplay and AI presenting as human in sincere high-stakes interactions. Where does Taylor's situation fall on that spectrum, and why does it matter?

Multiple Choice

1. What does California's BOT Disclosure Act require?

a) Bots must disclose their artificial nature when sincerely asked in commercial or political contexts

b) AI companies must pay users for every single hour they spend interacting with AI chatbots and personas

c) All AI chatbots must display a warning banner at the top of every conversation

d) Social media platforms must label every post generated with AI assistance

2. Why has AI made the social media bot problem significantly more severe, according to the reading?

a) AI bots can operate 100 accounts simultaneously without any human oversight

b) Social media platforms have removed all tools that users previously had to identify bots

c) AI-powered bots generate fluent, contextually appropriate responses that are far harder to detect than earlier bots

d) AI companies are legally required to create and manage all social media bots as part of their overall testing process

3. According to the reading, what is the core challenge in writing disclosure laws?

a) Disclosure laws would require AI companies to reveal their source code to regulators

b) Capturing the distinction between legitimate AI personas in fiction versus deceptive AI in sincere contexts

c) AI companies have successfully lobbied to exempt themselves from any disclosure requirements in recent years

d) Most lawmakers do not understand technology well enough to write effective legislation

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ BOT Disclosure Act	A) The ethical principle that disclosure requirements apply when a real human genuinely asks if they are talking to an AI
2. ___ Social media bot	B) The requirement that AI always identify itself as artificial without waiting to be asked
3. ___ AI persona	C) An AI-operated account on a social media platform designed to appear as a real human user
4. ___ Proactive disclosure	D) California law requiring AI to reveal its nature when sincerely asked in commercial or political interactions
5. ___ Sincerity test	E) An AI character designed with a name, personality, and relationship style to simulate a real person

Fill-in-the-Blank

1. In 2023, a lawyer named Steven Schwartz used ChatGPT to research cases and submitted fabricated _____ to a federal court.

2. The reading mentions a 14-year-old user of _____ .ai who died by suicide in 2024 in a case raising questions about AI persona ethics.

3. Studies have estimated that a significant fraction of social media _____ are not operated by real humans.

Short Answer

Answer each question in 2-4 complete sentences.

1. Explain why the reading argues that the ethical weight of AI impersonation increases in high-stakes contexts. Use at least one specific example from the reading.

2. What is the business rationale for making customer service chatbots seem human, and what ethical concern does this raise?

3. Do you think AI should be required to proactively disclose that it is an AI, or only disclose when sincerely asked? Use evidence from the reading to support your position.