Virtual Therapy: Embodied AI in Mental Health Care

Why?

This lesson tackles a controversial emerging technology that could impact many vulnerable individuals. As artificial intelligence advances, chatbots, virtual avatars, and other AI agents are taking on roles similar to human counselors and therapists. On the surface this seems like it could expand access to support for people struggling with their mental health. However, these are algorithms created by developers and companies, and don't have human understanding and discretion. Ethical mistakes by an AI system could be devastating and tragic. This lesson aims to open students' eyes to subtle but serious pitfalls ahead in an AI augmented world so they can help craft solutions.

Materials Needed	Time needed
● Simulation handout for students	● Approx. 60 - 75 minutes

Objectives

- Students will be able to articulate different ethical perspectives on using AI systems for mental health therapy.
- Students will be able to analyze the risks, biases, and limitations of AI mental health apps.
- Students will be able to evaluate whether, and under what conditions, the use of AI chatbots for student counseling might be ethically acceptable.

Key Concepts & Vocabulary

- **Algorithm**: A set of computational rules or procedures used by artificial intelligence systems to analyze, interpret, and respond to humans.
- **Embodied AI**: Artificial intelligence integrated into computers with physical form or virtual avatars, enabling natural interaction.

Lesson Components

1. **Before You Watch**: Connect lesson to background knowledge of embodied AI and get students' attention
2. **Video**: Show the pedagogy.cloud video explaining the ethical considerations in the topic of AI involved in therapy.
3. **Case Study**: Detail a real-world scenario that relates to the issue of students using apps to interact with "robot therapists."
4. **Simulation**: Lead students through an interactive activity exploring the possible ethical considerations related to a school district acquiring technology for AI therapists to help with the burden of mental health care for their students.
5. **Discussion**: Ask whole-class questions to reflect on experience and consider perspectives.

6. **Assessment**: Verify student understanding with an exit ticket.

1. Before You Watch

Brainstorming Session: Begin with a brief group discussion. Ask, "What do you know about Artificial Intelligence (AI)? How is it used in different fields like healthcare, entertainment, or education?"
Ask, "How much would you trust AI to diagnose health conditions? What about mental health conditions?" (Could have students hold up fingers to represent how much they would trust an AI diagnosis – 0 to 5, with 5 being "completely trust."

While You Watch: Mention these topics and questions for students to look out for as they watch the video:
- What are some reasons schools might turn to AI therapy?
- Identify two possible concerns about this technology.

2. Video Summary

The video explores the complex ethics surrounding the integration of AI chatbots and virtual avatars into mental health counseling. Using narrative case study, it introduces key AI capabilities, such as machine learning that allow algorithms to simulate human conversation. The video highlights promises of improving student access and anonymity, while raising concerns about emotional disconnects, risk detection, and data privacy. The video prompts students to think critically about balancing innovation and ethical precautions. It ends with a call toward responsible development guidelines so technology and ethics can evolve together in service of society's needs.

3. Case Study

Distribute or read Case Study handout.
Summary: A 16-year-old high school student turns to an AI chatbot for text-based therapy when the school's counseling staff is busy. The chatbot, while convenient, may miss critical context clues, lacks a mechanism for involving adults in cases of severe distress, and raises concerns about data privacy and increased isolation. Possible solutions include limiting AI therapy use without human supervision, implementing guidelines for safety and privacy, and considering a blended approach with human counseling.

4. Simulation

Background
The school district is facing budget cuts and looking into purchasing AI mental health chatbots to provide counseling services to students. An advisory board has been created to develop recommendations on whether and how to implement this.

Break students up into groups of approx. 5. (If groups are slightly smaller, consider eliminating roles from the bottom. If groups are slightly larger, include multiple students, counselors, and/or parents.)

Explain roles. Provide students with information on roles and positions.

Roles & Positions
Principal:
- Worried about liability issues and duty of care to students
- Doubts bots can detect serious issues
- Recommends strict guidelines on AI limitations

Counselor:
- Thinks AI could help with mild issues but not crisis situations
- Concerned about impacts on counseling staff
- Suggests hybrid model with human oversight

Student:
- Open to convenience of text-based sessions, but still wants option for in-person counseling
- Finds bots less judgmental than teachers
- Opposed to transcripts of sessions being made available to administrators and parents

Parent:
- Happy about increased access to support
- Uneasy about data privacy issues
- Wants right to review transcripts

Physician:
- Notes promising research on AI therapy effectiveness
- Concerned about lack of oversight and liability issues
- Recommends clinical trial approach before instituting in schools

Sequence of Tasks
1. Superintendent (role played by teacher) opens meeting and explains need for an agreement about certain issues:
    a. What level of AI therapy app should be available to students?
    b. How much interaction should there be between human therapists and AI therapy apps?
    c. How much information should administrators and parents be able to access about students' sessions?
    d. Are there specific types of issues that AI therapist apps should be able to, or not able to, address?
2. Each member makes opening statement on position
3. Group discusses ethical priorities and proper limitations
4. Members collaborate on draft guidelines plan
    a. Provide time for each role to comment on each question, if applicable.
    b. After each conversation, have members attempt to come to some sort of agreement on what should be in place, and any limitations, if applicable.

c. If group members don't all agree on limitations, have them vote on two options.
5. Groups present their conclusions to the class.

5. Discussion

These questions are designed to be used in whole-class discussion. Ask questions that relate most effectively to the lesson.
1. What stood out to you most during the debates between different perspectives? Was there a position you empathized with more?
2. If you could only pick one ethical guideline related to this topic, what would be the most important to you, and why?
3. Should an AI system be allowed to make suggestions directly to students or should a human always be involved? Why?
4. How would you feel about confiding in or taking advice from an AI chatbot for your problems? How do you think your friends would feel about it?
5. Would you prefer to talk with a robot therapist or a human? Why?
6. Do you think AI and human counseling should be separate or integrated?
7. What kind of oversight is needed on how student data is collected and shared by these AI systems? Who should be able to access transcripts? Are there certain situations where you feel humans *should* or *should not* be able to access transcripts?
8. What should happen if an AI system fails to properly refer students to crisis services? Should anyone be held legally liable for missing an opportunity to get help for a student?

6. Assessment

Exit Ticket: Provide a prompt for students to reflect on their learning, such as:
● What do you think is the most significant ethical concern about using AI for mental health therapy?
● Should students have access to AI mental health services without any supervision? Why or why not?
● If your friend was using an AI chatbot like VRtex, what risks or downsides would you warn them about?

Sources to Learn More

● Scholarly article discussing various ethical issues related to virtual therapy - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6532335/
● Article discussing the background, promise, and concerns about AI therapy - https://www.technologyreview.com/2021/12/06/1041345/ai-nlp-mental-health-better-therapists-psychology-cbt/
● Article that addresses pros and cons, and situations where AI therapy works well and poorly - https://www.npr.org/sections/health-shots/2023/01/19/1147081115/therapy-by-chatbot-the-promise-and-challenges-in-using-ai-for-mental-health

Case Study: Embodied Therapy App

Background Information
Sarah is a 16-year old high school student struggling with anxiety and depression. Her school offers free counseling services, but the one counselor is always extremely busy. Sarah decides to sign up for VRtex, an AI chatbot that provides text-based therapy. The bot uses algorithms to have conversations, track Sarah's mood, and provide coping strategies.

Problem Analysis
While VRtex is convenient, it also poses risks. The app may fail to pick up on context clues and assess Sarah's condition accurately. Without human oversight, it can't involve parents/teachers if Sarah describes self-harm. The app also collects sensitive data with unclear privacy protections. And over-reliance on VRtex versus human connections could worsen Sarah's isolation.

Possible Solutions
Some argue VRtex should not be used for teens without any human supervision. Others think that with disclaimers on its limitations, VRtex can provide some initial mental health support if traditional services aren't available. Guidelines could also restrict unsafe content recommendations and require better data protections. In an ideal setting, VRtex would complement human counseling.

Conclusion
AI therapy apps hold promise but require safeguards. Students should discuss appropriate oversight, how to uphold safety and privacy, and the pros/cons of blended human and AI counseling.

Questions
- What is your current opinion about using AI chatbots like VRtex for mental health support in schools?
- Should AI be used only as a complement to human counseling, or do you believe there's a scenario where AI could effectively stand alone?

Simulation Activity

Roles & Positions

Principal:
- Worried about liability issues and duty of care to students
- Doubts bots can detect serious issues
- Recommends strict guidelines on AI limitations

Counselor:
- Thinks AI could help with mild issues but not crisis situations
- Concerned about impacts on student counseling skills
- Suggests hybrid model with human oversight

Student:
- Open to convenience of text-based sessions, but still wants option for in-person counseling
- Finds bots less judgmental than teachers
- Opposed to transcripts of sessions being made available to administrators and parents

Parent:
- Happy about increased access to support
- Uneasy about data privacy issues
- Wants right to review transcripts

Physician:
- Notes promising research on AI therapy effectiveness
- Concerned about lack of oversight and liability issues
- Recommends clinical trial approach before instituting in schools

Video Script for Animator

Hello Young Innovators! Today we're discussing the ethics of AI in mental healthcare.
Title screen

Artificial intelligence is being integrated across healthcare. One domain it's now entering is mental health counseling and therapy for conditions like anxiety, depression, trauma, or eating disorders.
[Socrat narrating throughout, facing viewers. Wide view of a futuristic hospital lobby with holograms, robots, and glowing lights representing technological connections. Example - https://i.imgur.com/kGHxJ2y.png]

Driven by complex algorithms, embodied AI applications aim to simulate human conversation using speech recognition and natural language processing. Companies claim these AI systems could expand access to support by serving as virtual counselors.
[Socrat standing in a Futuristic healthcare setting, with robots and images of AI approximations of humans. Example - https://i.imgur.com/L9ixADZ.png]

But should we hand over such a delicate, personal aspect of healthcare to machines? Ethicists argue both sides. Some focus on positives, such as anonymity for stigmatized issues. Others flag risks, such as emotional disconnects and crises missed.
[Socrat standing off to the side. Still in healthcare setting, focus on one person (adult female) in the lobby, who pulls cell phone out and turns it on. Zoom in on phone to show AI-looking face that pretends to talk. Example - https://i.imgur.com/cGu4fkb.png Picture on phone starts to glitch out, symbolizing disconnects.]

Students access AI therapy for different reasons. School districts may turn to AI therapy options for their students in order to expand access while keeping costs controlled. At the same time, companies are marketing AI services directly to students. As innovations march ahead, treatment guidelines and oversight mechanisms lag behind.
[Socrat stands in a school lobby with chairs. Students are talking to each other, some looking at phones.Example - https://i.imgur.com/lhotxOp.png]

Meet Noah - increasingly anxious and withdrawn. He messages VRtex, an AI chatbot that provides mental health support through automated text conversations.
[Party scene, people talking and laughing. Noah (approx. 16 years old, skinny high school student–could have somewhat darker-toned skin if desired) sits on a couch,

looking sad, and pulls out his phone. Socrat stands on the other side of the couch and looks over at Noah. Example - https://i.imgur.com/eDxa2ZY.png]

AI chatbots hold promise for many people. They provide anonymity that encourages openness for some. They also offer more convenient access for students that can't get appointments with overburdened school counselors.
[Scene change to waiting room with multiple students looking anxious, bored, or checking phones. A door shows a counselor's schedule that features a calendar with a lot of markings on most of the days. (Socrat could turn the page of the calendar and see that most of the days have several marks on them.) Calendar example - https://i.imgur.com/8JsD8FP.png]

But algorithms also have blindspots. Without human nuance, they may miss serious risk signals and can't involve parents or teachers appropriately. There are also data privacy issues.
[Scene change to a computer desk with nobody sitting at it. Socrat narrating from the side. Monitor shows a chat-type program has received message, "I'm feeling hopeless." Computer replies "Error," and then gets a ... indicating that the person is typing. Example image: https://i.imgur.com/TG7pLbP.png but better if it's zoomed in a little more]

Let's see how Noah is doing. At first, Noah found VRtex's coping strategies and non-judgmental approach helpful. But over time, its preset responses failed to meet his struggles. Yet with no human involvement, risks went unseen.
[Back to Noah scene, zoomed in on him on the couch – Noah reacts negatively to something shown on the screen, and then leans his phone forward, showing screen with a blinking red X]

Over-reliance on AI bots risks making teens more isolated, and fails to build interpersonal skills. Guidelines and oversight are needed to ethically integrate this technology. Figuring out what limits to set is important. Students' health and lives are in the balance.
[Zoom out, showing more students who talk to each other, smiling and laughing. They ignore Noah. Noah looks around.]

The future offers incredible promise to expand mental health access through AI innovation. Through education and thoughtful policymaking, technology and ethics can evolve together. Working together, we can manage the risks in order to serve society's mental health needs.

[Back to school lobby, where students now look up from their phones and smile at each other.]

Let's discuss: If the principal at your school asked your opinion on whether you would trust an AI therapist, how would you respond?

[Discussion Question: If the principal at your school asked your opinion on whether you would trust an AI therapist, how would you respond?]

Question text appears on screen

Video Script for Narrations

Hello Young Innovators! Today we're discussing the ethics of AI in mental healthcare.

Artificial intelligence is being integrated across healthcare. One domain it's now entering is mental health counseling and therapy for conditions like anxiety, depression, trauma, or eating disorders.

Driven by complex algorithms, embodied AI applications aim to simulate human conversation using speech recognition and natural language processing. Companies claim these AI systems could expand access to support by serving as virtual counselors.

But should we hand over such a delicate, personal aspect of healthcare to machines? Ethicists argue both sides. Some focus on positives, such as anonymity for stigmatized issues. Others flag risks, such as emotional disconnects and crises missed.

Students access AI therapy for different reasons. School districts may turn to AI therapy options for their students in order to expand access while keeping costs controlled. At the same time, companies are marketing AI services directly to students. As innovations march ahead, treatment guidelines and oversight mechanisms lag behind.

Meet Noah - increasingly anxious and withdrawn. He messages VRtex, an AI chatbot that provides mental health support through automated text conversations.

AI chatbots hold promise for many people. They provide anonymity that encourages openness for some.They also offer more convenient access for students that can't get appointments with overburdened school counselors.

But algorithms also have blindspots. Without human nuance, they may miss serious risk signals and can't involve parents or teachers appropriately. There are also data privacy issues.

Let's see how Noah is doing. At first, Noah found VRtex's coping strategies and non-judgmental approach helpful. But over time, its preset responses failed to meet his struggles. Yet with no human involvement, risks went unseen.

Over-reliance on AI bots risks making teens more isolated, and fails to build interpersonal skills. Guidelines and oversight are needed to ethically integrate this technology. Figuring out what limits to set is important. Students' health and lives are in the balance.

The future offers incredible promise to expand mental health access through AI innovation. Through education and thoughtful policymaking, technology and ethics can evolve together. Working together, we can manage the risks in order to serve society's mental health needs.

Let's discuss: If the principal at your school asked your opinion on whether you would trust an AI therapist, how would you respond?