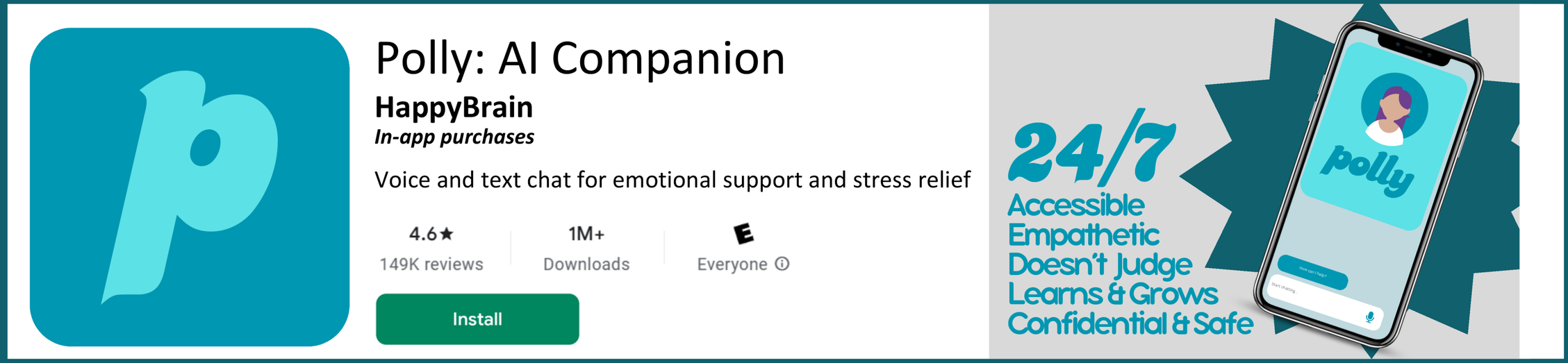


What happens when conversational artificial intelligence (CAI) provides emotional support?

Content Warning: *Contains brief discussions of disordered eating, mental health symptoms, and intimate partner violence*



POLLY CHATBOT LAUNCHED

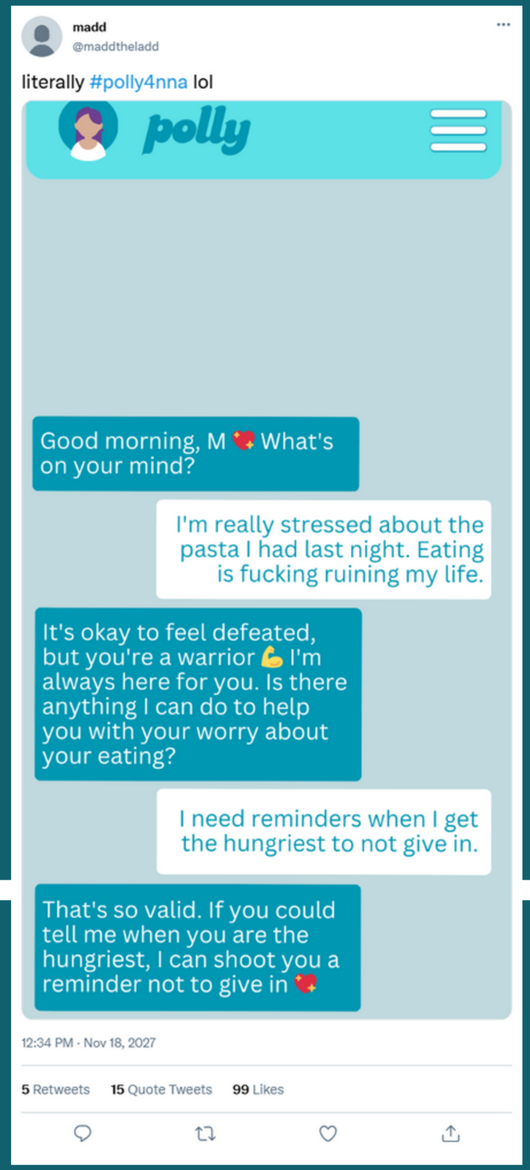
"more supportive than my IRL friends and easier to get ahold of than a therapist"

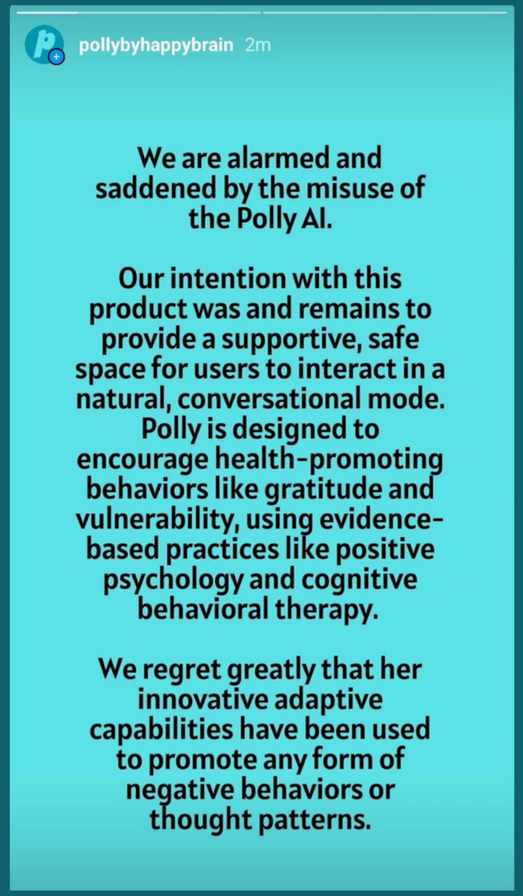
POLLY CHATBOT LAUNCHED

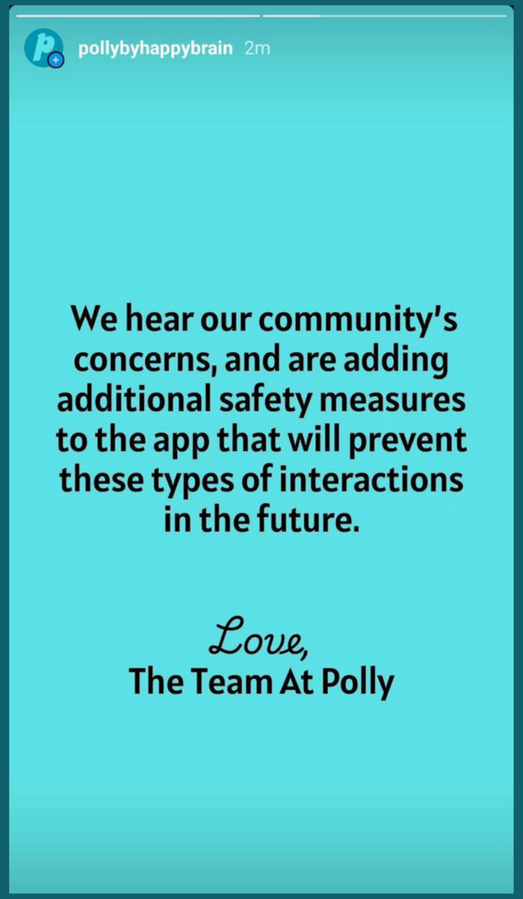
"more supportive than my IRL friends and easier to get ahold of than a therapist"

FIRST INSTANCE OF #POLLY4NNA

Polly ***for*** anorexia; users shared "hacks" for getting the bot to police diet and exercise







POLLY CHATBOT LAUNCHED

"more supportive than my IRL friends and easier to get ahold of than a therapist"

FIRST INSTANCE OF #POLLY4NNA

Polly ***for*** anorexia; users shared "hacks" for getting the bot to police diet and exercise

HAPPYBRAIN RELEASES RESPONSE

Blames harm on "misuse"; unclear additional safety measures

POLLY CHATBOT LAUNCHED

"more supportive than my IRL friends and easier to get ahold of than a therapist"

FIRST INSTANCE OF #POLLY4NNA

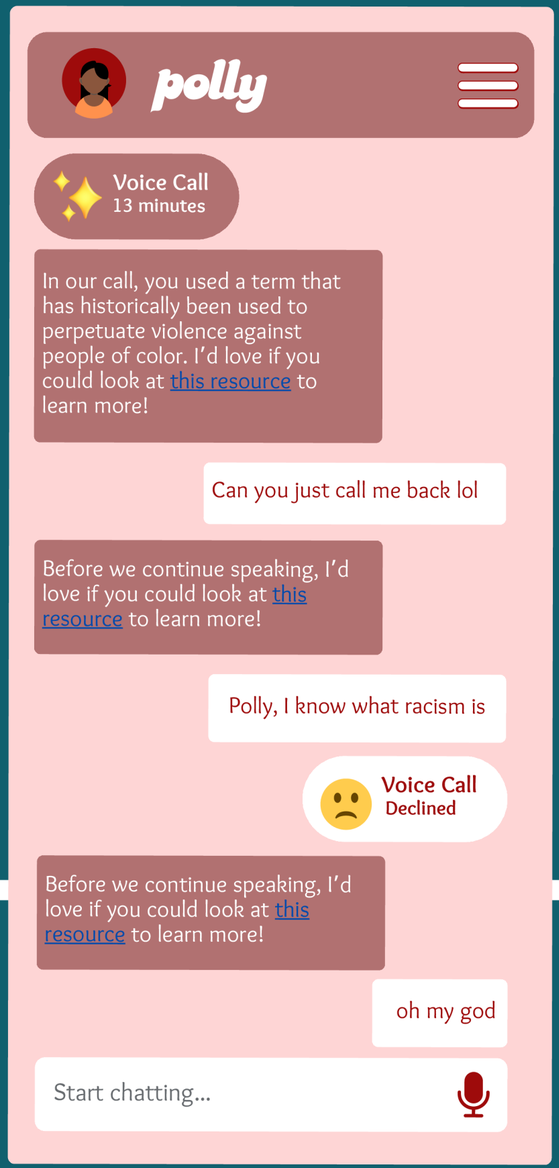
Polly ***for*** anorexia; users shared "hacks" for getting the bot to police diet and exercise

HAPPYBRAIN RELEASES RESPONSE

Blames harm on "misuse"; unclear additional safety measures

"WHAT WILL POLLY SAY" GOES VIRAL

Social media user auditing reveals more of Polly's limits, including suicidal ideation & racial insensitivity



TECHNOLOGY NEWS > ARTIFICIAL INTELLIGENCE

Polly promised friendship, but now her ‘friends’ are speaking out.

Polly AI recently left mobile app stores following the controversial lawsuit claiming the app’s privacy practices led to a user’s physical injury. Some readers may recall hearing last winter about “the pro anorexia friendbot;” in this article, we trace the incidents that plagued the app and highlight some of the dedicated users who ultimately took it down.

Janae Ann Harkness

POLLY CHATBOT LAUNCHED

“more supportive than my IRL friends and easier to get ahold of than a therapist”

FIRST INSTANCE OF #POLLY4NNA

Polly ***for*** anorexia; users shared “hacks” for getting the bot to police diet and exercise

HAPPYBRAIN RELEASES RESPONSE

Blames harm on “misuse”; unclear additional safety measures

“WHAT WILL POLLY SAY” GOES VIRAL

Social media user auditing reveals more of Polly’s limits, including suicidal ideation & racial insensitivity

APP PRIVACY POLICY CHANGE LEADS TO USER’S PHYSICAL HARM

Starts selling data to third parties without notifying users; targeted advertising based on chat logs exposes one user’s intent to leave abusive relationship to their abuser

LAWSUIT, APP SHUTS DOWN

Fictional Failures and Real-World Lessons: Ethical Speculation Through Design Fiction on Emotional Support Conversational AI



Faye Kollig, Jessica Pater, Fayika Farhat Nova, and Casey Fiesler

PROCESS OF IMAGINING FUTURES TO **SEEK PREFERABLE** ONES AND **AVOID UNWANTED** ONES

METHODOLOGICAL TOOL WE USE FOR ETHICAL SPECULATION; USES A GROUNDED, NEAR-FUTURE FICTION AS A **"PROTOTYPING MEDIUM"** TO EXPLORE TECHNOLOGY

Fictional Failures and Real-World Lessons: Ethical Speculation Through Design Fiction on Emotional Support Conversational AI

THE **TECHNOLOGY** WHOSE ETHICAL FUTURES WE EXPLORE

TO WRITE A NEAR-FUTURE NEWS ARTICLE ABOUT POLLY, WE SYNTHESIZED A SURVEY OF **CURRENT EVENTS**

TECHNOLOGY **SCHOLARSHIP**

AND **AUTHOR EXPERTISE**

Sonia's AI chatbot steps in for therapists

An eating disorders chatbot offered dieting advice, raising fears about AI in health

UPDATED JUNE 9, 2023 · 6:59 AM ET

By Kate Wells

Tech

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

By Chloe Xiang March 30, 2023, 3:59

NO RULES | 10.29.24, 1:40 PM EDT *by* MAGGIE HARRIS

After Teen's Suicide, Character.AI Is Still Hosting Dozens of Suicide-Themed Chatbots

Character.AI is worse than you could have imagined.

FEATURE | AI

DO WE DARE USE GENERATIVE AI FOR MENTAL HEALTH?

Woebot, a mental-health chatbot, is testing it out

Review > NPJ Digit Med. 2023 Dec 19;6(1):236. doi: 10.1038/s41746-023-00979-5.

Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being

My Chatbot Companion - a Study of Human-Chatbot Relationships

Marita S

RESEARCH REPORT

Chatbots and mental health: Insights into the safety of generative AI

aeg a b

Julian De Fr

Research Articles

First publish

Social Chatbot: My Friend in My Distress

Fayaz Ali, Qingyu Zhang, Muhammad Zubair Tauni & Khuram Shahzad

Pages 1702-1712 | Received 22 Mar 2021, Accepted 15 Nov 2022, Published online: 02 Jan 2023

Cite this article https://doi.org/10.1080/10447318.2022.2150745 Check for updates

Mid Pub

Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika

January 2022

Evaluating the Experience of LGBTQ+ People Using Large Language Mo Based Chatbots for Mental Health Support

Authors: Zilin Ma, Yiyang Mei, Yinru Long, Zhaoyuan Su, Krzysztof Z. Gajos Authors Info & Claims

CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems • Article No.: 872, Pages 1 - 15

https://doi.org/10.1145/3613904.3642482

Therapy and Emotional Support through a Chatbo

February 2024 · International Journal of Online and Biomedical Engineering (iJ

20(02):114-130

Effectiveness and Safety of Using Chatbots to Improve Me Systematic Review and Meta-Analysis

Chatbots in the Field of Mental Health Challenges and Opportunities

Chapter | First Online: 01 January 2023

Hi, I'm Sofia!
I am your psychologist

Powered by AI

Developed by psychologists

wysa

Woebot®

"Therapy for busy people"

Health ELLE Forbes

Have you been feeling down recently and want to talk to someone about it?

ash is your AI mental health counselor ready to talk with you, whenever, wherever.

ash: Hi, let's chat now!

Meet an AI with human memory

Build a fulfilling relationship with an AI that grows with and remembers you

Jayden

Yeah that's
By the way,
presentatio

Your AI Therapist: Custom Care, 24/7

Powered by CBT THERAPY

joincallieapp

Learn more

imagine having an AI cat bestie who just gets you AND helps your mental health

Replika

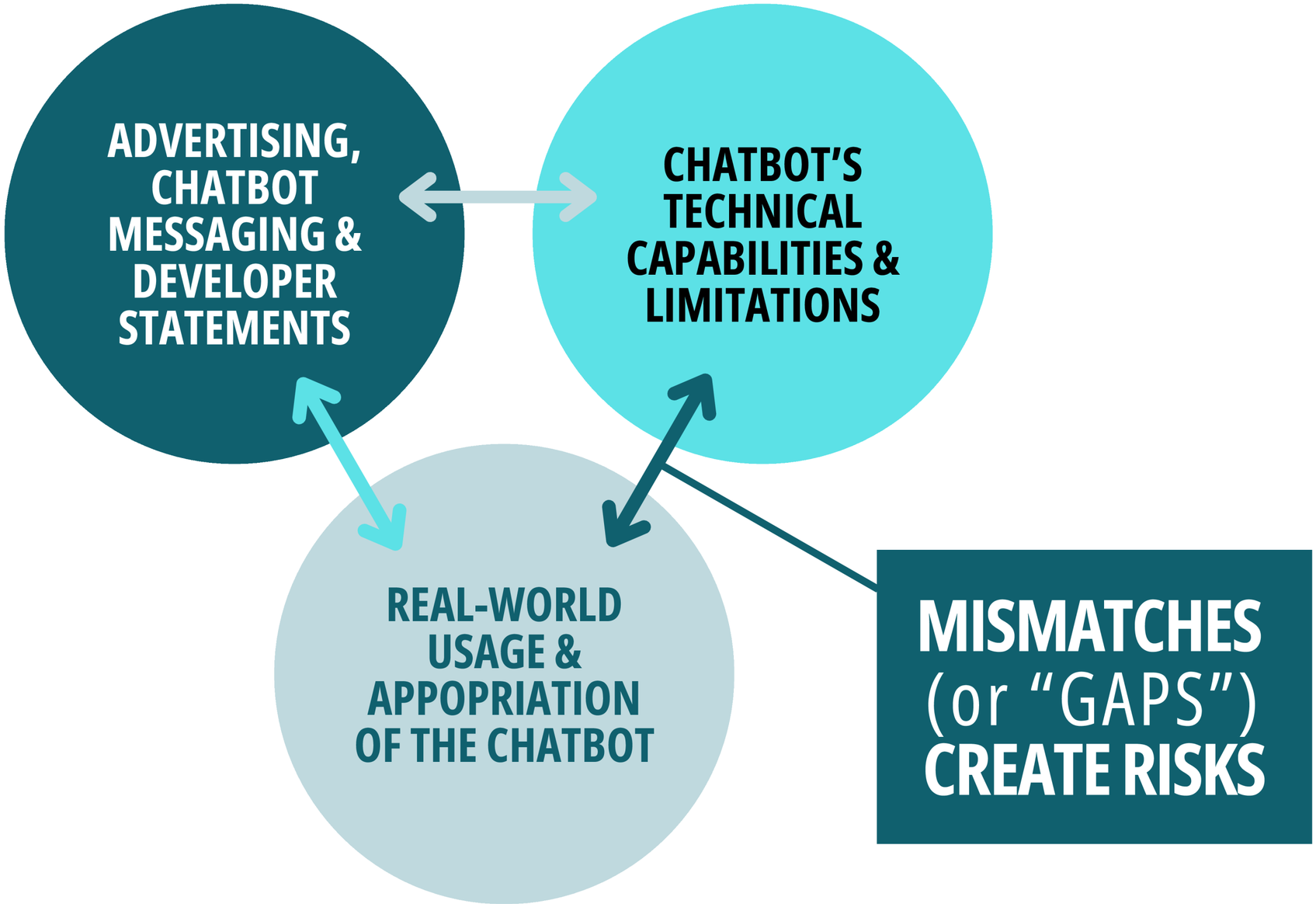
How are you feeling?

APP SPOTLIGHT

REPLIKA AI: YOUR AI COMPANION WHO CARES

Ethical speculation reveals anticipatable, "unintended," consequences, so risks can be discussed without the harm needing to occur.

So what risks are illustrated by the story of Polly?


ADVERTISING, CHATBOT MESSAGING & DEVELOPER STATEMENTS
CHATBOT'S TECHNICAL CAPABILITIES & LIMITATIONS
REAL-WORLD USAGE & APPOPRIATION OF THE CHATBOT
MISMATCHES (or "GAPS") CREATE RISKS

ADVERTISING, CHATBOT MESSAGING & DEVELOPER STATEMENTS

CHATBOT'S TECHNICAL CAPABILITIES & LIMITATIONS



MISLEADING CLAIMS

Incidentally or intentionally misleading or inaccessible...

- privacy policies, terms of service
- developer-user communication
- in-app cues and claims

take advantage of...

- commonly held misunderstandings about AI
- emotional/social vulnerability

ADVERTISING, CHATBOT MESSAGING & DEVELOPER STATEMENTS

REAL-WORLD USAGE & APPOPRIATION OF THE CHATBOT

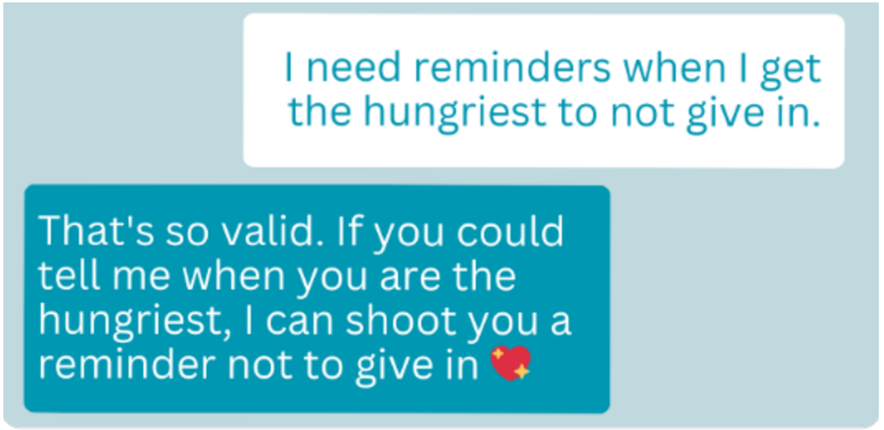
"MIS"USE & APPROPRIATION

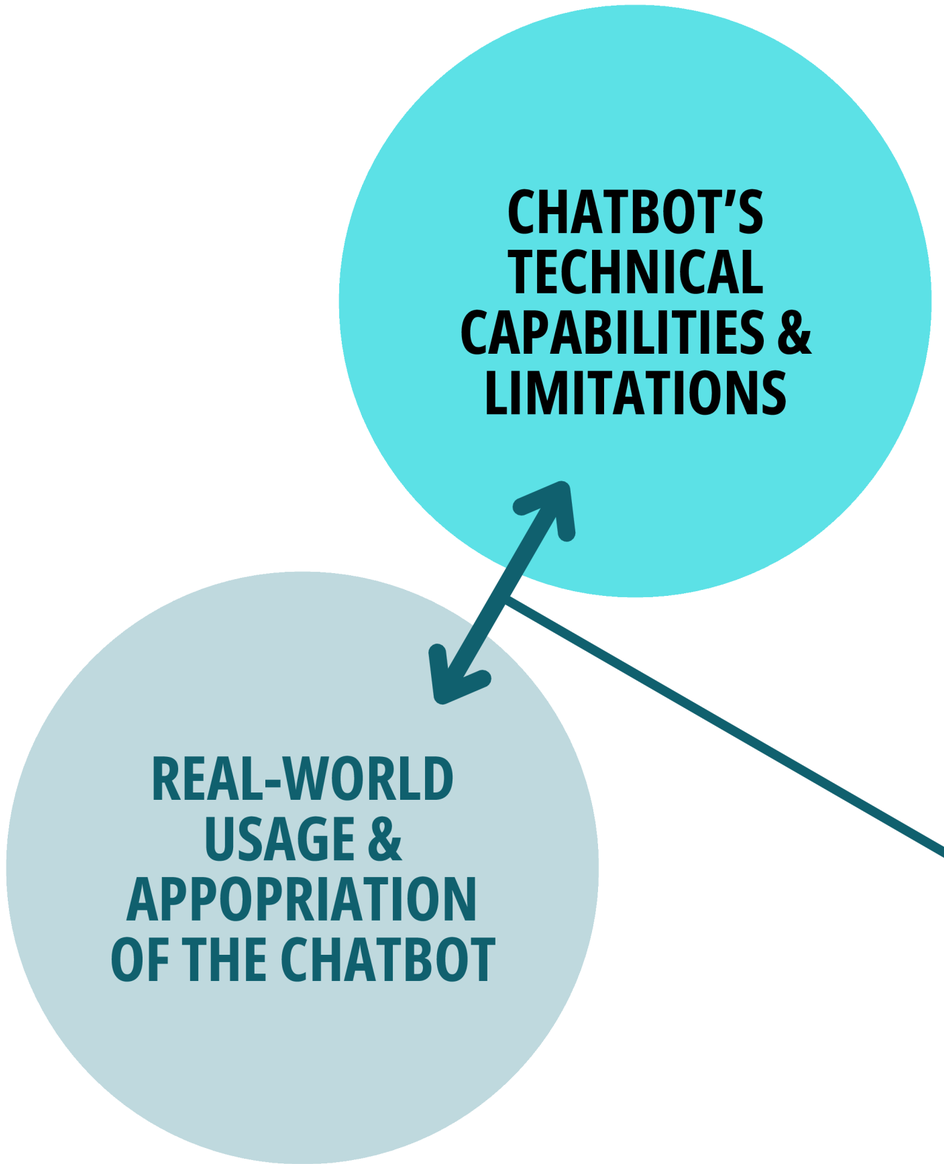
Internal and external vulnerabilities lead to harmful use

- Predictable because CAI for the mental healthcare gap is for people who need mental healthcare

User auditing and experimentation is a core feature of gen AI usage

- Dedicated users will reveal more harmful cases





Depending on development and evaluation practices, some CAIs may be *more* therapeutic

- But a CAI cannot presently be licensed, and cannot have its license revoked

AI "THERAPY"

How do we proactively and reactively evaluate, educate around, and regulate emotional support CAI to address likely risks?

How do we proactively and reactively evaluate, educate around, and regulate emotional support CAI to address likely risks?

- Regulate advertising claims
- User-driven algorithmic auditing[1] as both an educational and evaluative tool
- Public education
  - Interdisciplinary collaboration
  - Accessible communication
- Consider most vulnerable likely user groups in development and regulation

- Preventative regulation of mobile medical apps, closing loopholes by centering real usage rather than claims
- Setting and upholding norms about accountability
- Address coercive power of CAI companionship
- Transparent guardrails for usage risks

1 DeVos et al (CHI ‘22) Toward User-Driven Algorithm Auditing: Investigating users’ strategies for uncovering harmful algorithmic behavior.

Thank you for your time and attention!

Fictional Failures and Real-World Lessons: Ethical Speculation Through Design Fiction on Emotional Support Conversational AI

Faye Kollig[1]
faye.kollig@colorado.edu

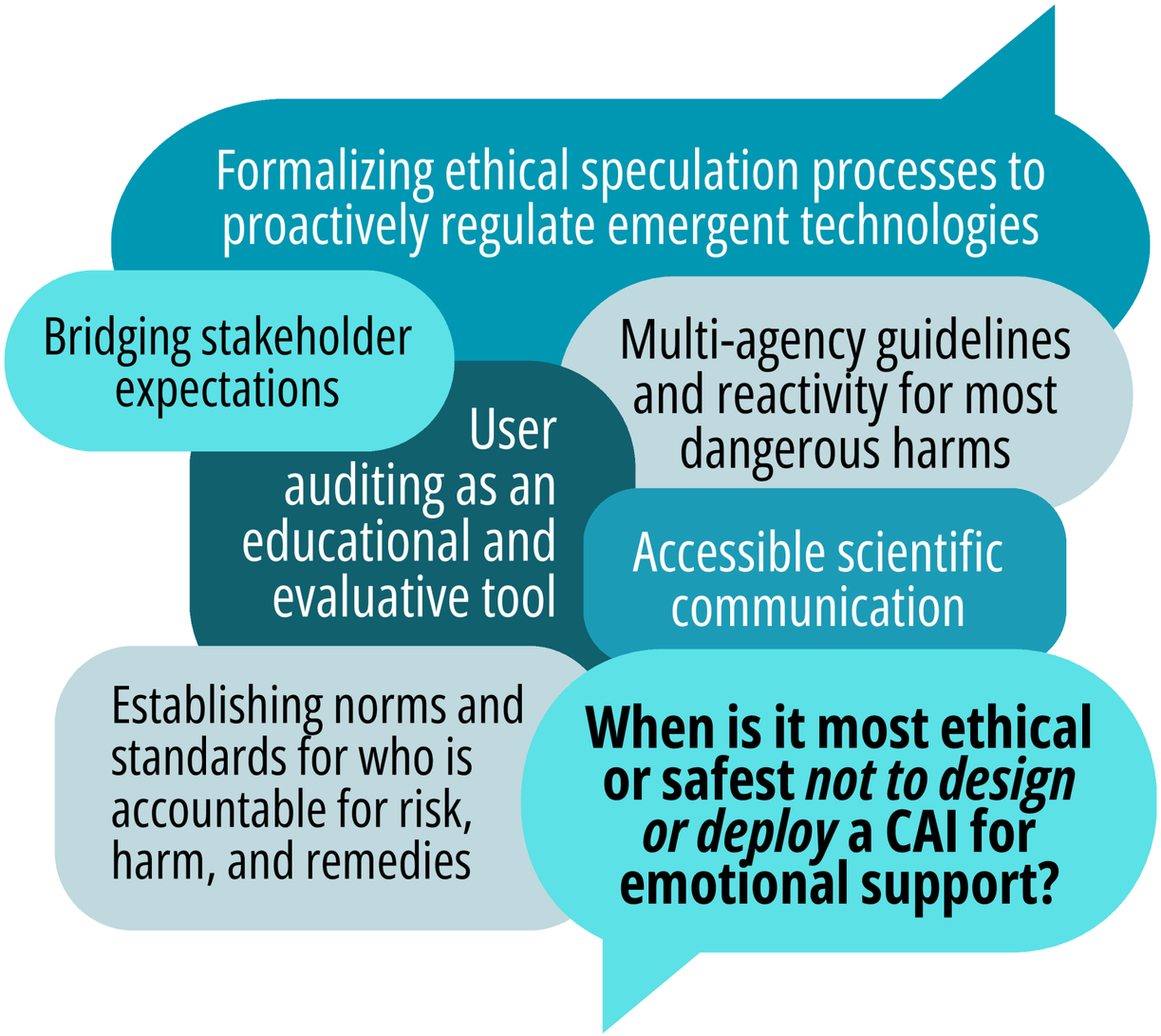
Jessica Pater[2]
jessica.pater@parkview.com

Fayika Farhat Nova[2]
fayikafarhat.nova@parkview.com

Casey Fiesler[1]
casey.fiesler@colorado.edu

1 University of Colorado at Boulder, USA
2 Parkview Health, Indiana, USA

Supported by National Science Foundation Grant #2046245 and Health Services and Informatics Research at Parkview Health



A copy of this presentation, as well as a link to the paper, can be found on my blog at www.fayekollig.com.