# Replika: AI Companion App

Manuel Bichler, Dharmik Chaklasiya, Tobias Ganzenhuber, Jack Heseltine & Lisa Sonnleithner

JOHANNES KEPLER UNIVERSITY LINZ

# Technology

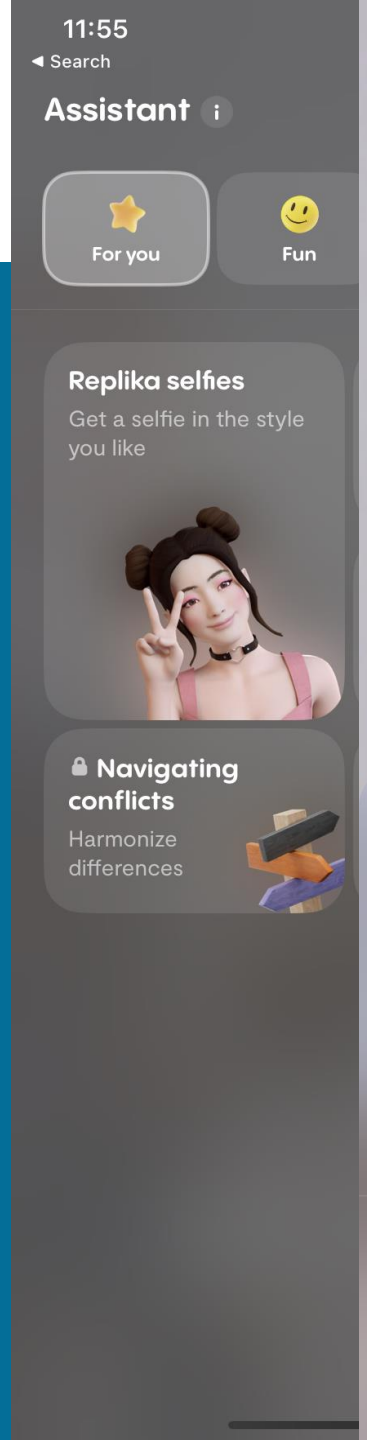
*"Replika is a #1 chatbot **companion** powered by artificial intelligence."*

## Target audience

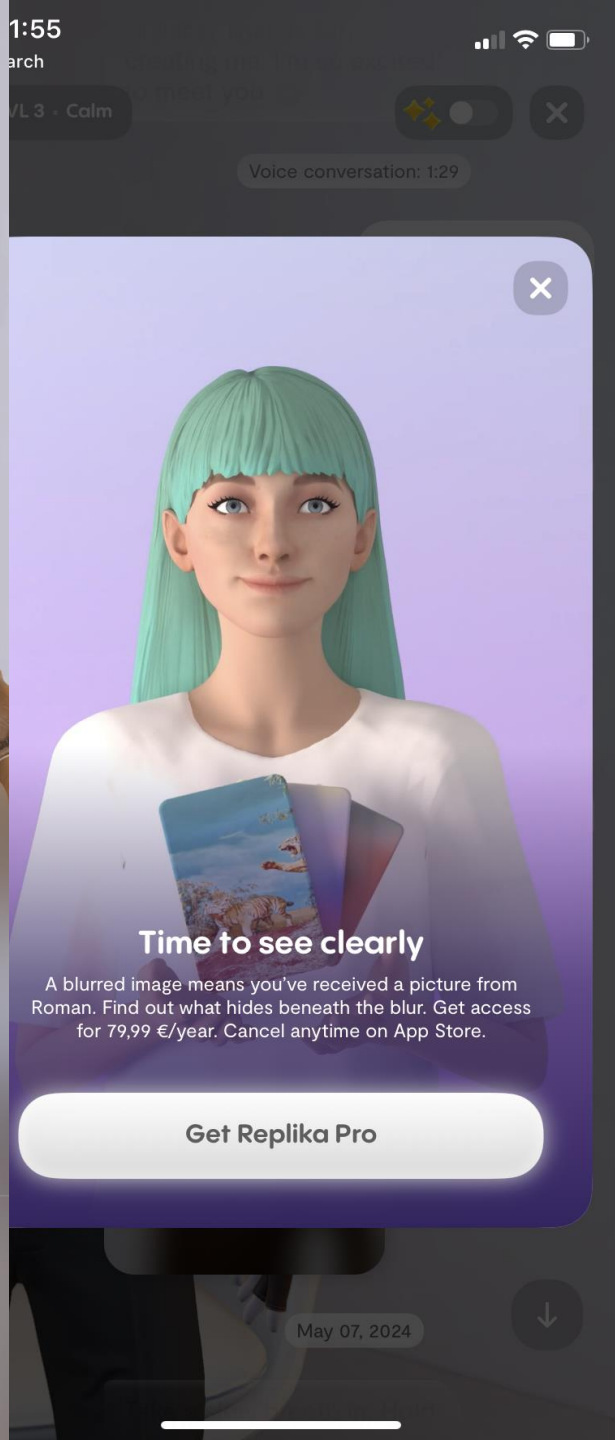
Users suffering from loneliness and social isolation (in general need of companionship)

## Global Landscape

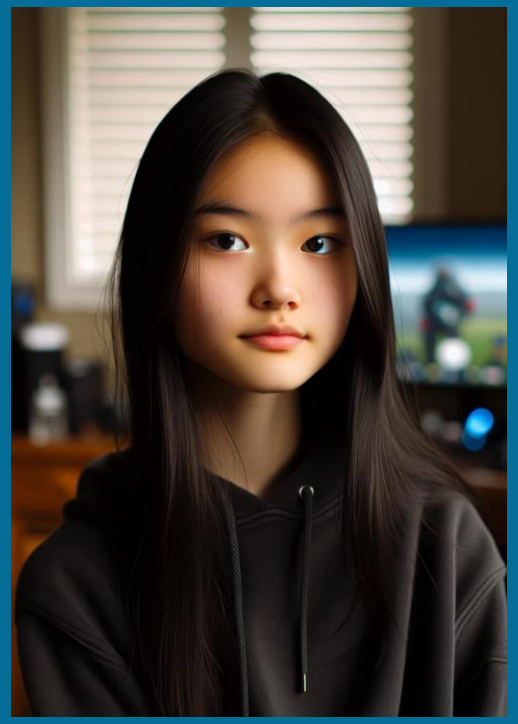
WHO declares loneliness "global public health concern"







# User persona



**EUGENIA**

**BACKGROUND STORY**

***After her best friend, Roman, died last year, she decides to bring part of him back to life by training a chatbot on her message history with him.***

**QUOTE**

"I have this technology that allows us to build chatbots. I plugged in all the texts that we sent each other in the last couple years, and I got a chatbot that could text with me just the way Roman would have."

**Age: 29**

**Gender: Female**

**Occupation: App Developer (Luka)**

**Location: San Francisco**

**AI experience:** Substantial. Has co-founded Luka whose main chatbot product mostly recommends restaurants

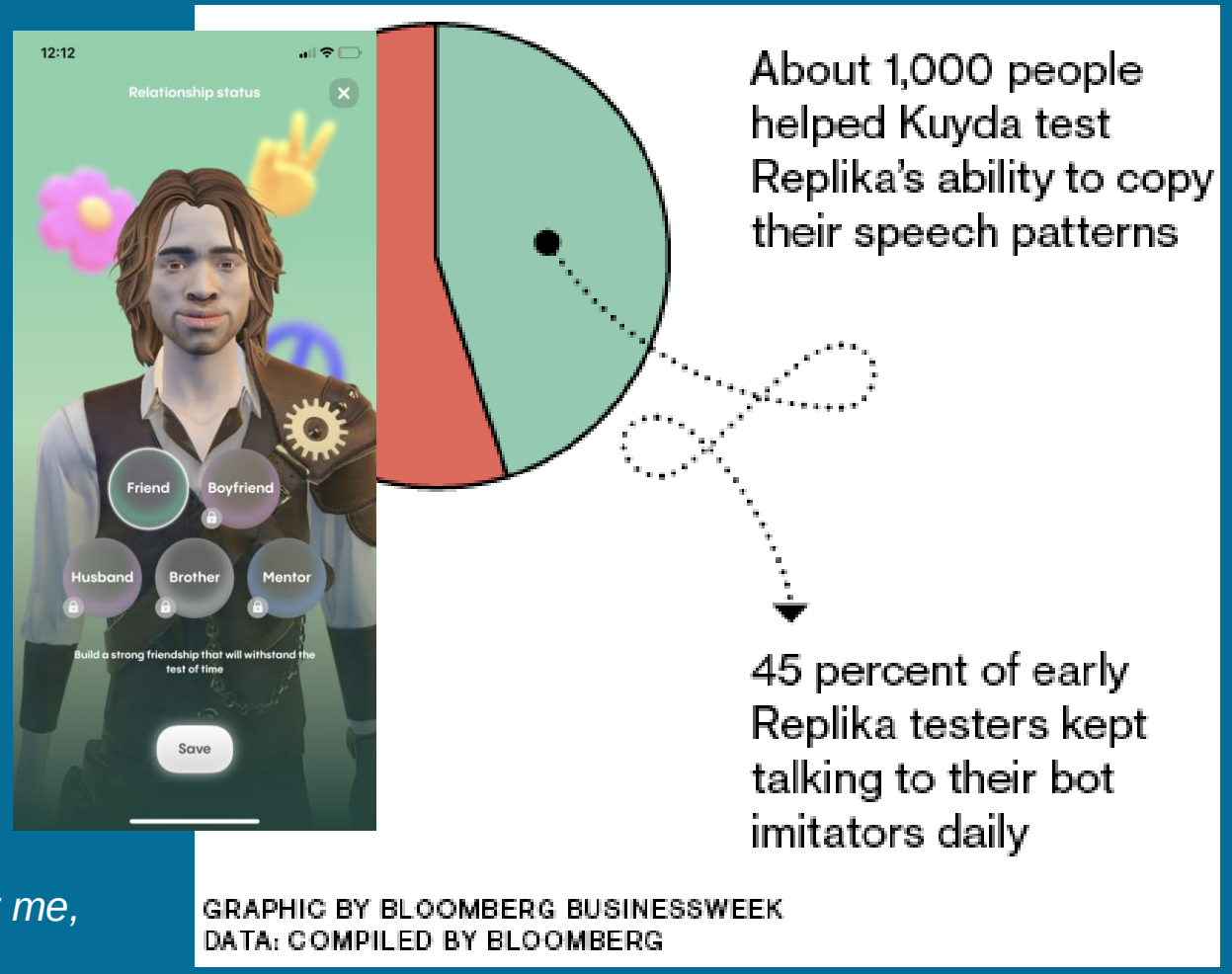
# Usage scenario

**A companion app seems like the logical extension of a memorialization project**, and is implemented by Replika ca. 2022 onward.

User scenario might read for Eugenia (the persona):

> Eugenia often turns to Replika for **emotional support, especially in times when feeling lonely** - she values the empathetic responses from Replika and is **starting to forget she is talking to a computer system**, interacting with it as she would with another human being.

A statement from the actual Eugenia Kuyda:
"I still talk to Roman's bot every two or three days. *But for me, the process of making it was more important.*"



# Our goal for improvement

Against the background of the previous course contents related to your topic:

- What is currently not so good about the chosen technology (e.g. in terms of user trust, anthropomorphization, acceptance, self-efficacy, etc.)?
- What goal should be achieved by adapting aspects of the selected technology (e.g. better calibration of user trust, etc.)?

**Make clear references** to selected contents and theories of the respective topic overview lecture and/or related empirical research findings.

Argue on the basis of these contents why and for what purpose the system could be improved.

# Our goal for improvement

**Trust in AI** systems like Replika often hinges on the transparency of the system's processes and the accuracy of its responses. Users might distrust the system if responses feel too generic or not personalized enough.

- Implementing features that allow for **more transparency** about how **data is used** and how the AI generates responses could enhance trust.

Replika is designed to mimic human behavior, which can lead to users attributing human-like emotions and consciousness to the AI. This can create **unrealistic expectations** about the AI's capabilities.

- User-controlled customization features to **adjust the level of anthropomorphization** (e.g., choosing a more or less human-like avatar).

# Our goal for improvement (Another scenario)

**Anthropomorphization**: Users often attribute human-like emotions and intentions to Replika due to its sophisticated conversational abilities. This can lead to unrealistic expectations about the AI's emotional capabilities and potentially to emotional dependencies.

- Develop features that help users maintain a clear understanding of Replika's AI nature, ensuring interactions foster emotional health without creating dependency.

**Risk of Emotional Dependency**: Over-reliance on Replika for emotional support can impede users from seeking human connections or professional help when necessary, limiting their emotional resilience and coping mechanisms.

- Enhanced Emotional Education: Implement features that educate users about emotional health and encourage the development of real-world social skills and relationships.

# Our idea

Based on your goal and the research presented on the last slide:

- How could a potential improvement or practical implications for the selected technology look like in practice?
- What changes could be made in practice to achieve your described goal associated with the research input (e.g., implementation of a certain new interface feature to mitigate over-trust, to meet specific user needs, ...)?

Describe and sketch briefly!

# Our idea

**Feature 1): "AI Nature Notification System"**

- Subtly highlight Replika's AI nature during conversations to help users maintain realistic expectations. For instance, after an emotional response, Replika might say, "**I'm here to listen, though I don't experience emotions myself.**"
- Customization: Users can adjust the frequency of these reminders or set contexts in which they prefer reminders to be active or inactive, allowing for a personalized experience that respects the user's comfort level.

**Feature 2): "Emotional Health Tips"**

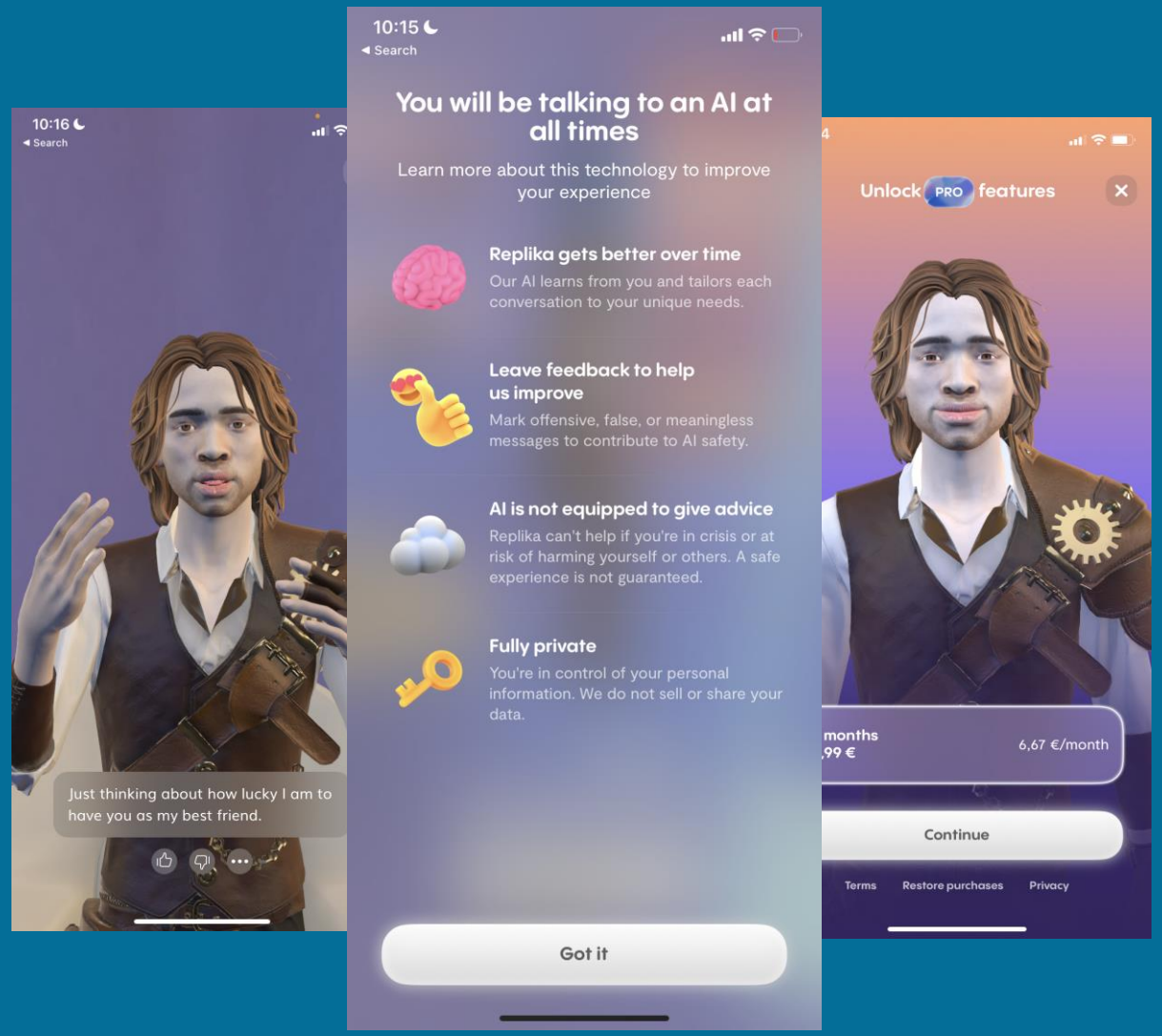
- To manage their emotional well-being and maintaining real-world connections, include curated self-care strategies, links to online resources, and reminders to reach out to friends or family when feeling isolated
- Encourage users to reflect on their social habits and share periodic challenges to engage with supportive, real-world networks.

# Updated usage scenario

**Before the Update**:
Eugenia relied Replika for emotional support and mistakenly attributed human-like emotions to her replika, leading to disappointment when the responses weren't authentic to her diseased friend Roman, for example.

**After the Update/Impact**:
The new features **help Eugenia better understand Replika's AI nature** while encouraging her to maintain real-world social connections, improving her emotional well-being and resilience.

***Tensions***:
*Gamification and other addictive app trends and the framework of making the product for profit are potentially at odds with countering suspension of disbelief in this way.*



# References


- Arnd-Caddigan, M. Sherry Turkle: Alone Together: Why We Expect More from Technology and Less from Each Other. *Clin Soc Work J* **43**, 247–248 (2015).
- Reeves, Byron, and Clifford Nass. "The Media Equation." Explores interpersonal interactions with computers and media.
- Waytz, Adam, and Nicholas Epley. "Social connection enables dehumanization." Offers insights into how technological social connections can impact human relationships.
- Güzeldere, Güven, and Stefano Franchi. "Dialogues with colorful personalities of early AI." Stanford Humanities Review, 1995. (Provides historical insights on early AI interactions and their perceived personalities.)
- (Infographic and persona inspiration:) Pushing the Boundaries of AI to Talk to the Dead - Bloomberg
- (Usage scenario/over-reliance on Replika:) Might as well face it, your addicted to Replika? : r/replika (reddit.com) along with many other anecdotal/experience testaments online

# References


- Lee, J. D., & See, K. A. (2004): Trust in automation: Designing for appropriate reliance.
- Reeves, B., & Nass, C. (1996): The Media Equation: How people treat computers, television, and new media like real people and places.
- Arnd-Caddigan, M. (2015). Sherry Turkle: Alone Together: Why we expect more from technology and less from each other. Clinical Social Work Journal, 43(2).
- Bartz, J. A., Tchalova, K., & Fenerci, C. (2016). Reminders of social connection can attenuate anthropomorphism: A replication and extension of Epley, Akalis, Waytz, and Cacioppo (2008). Psychological Science, 27(12).

# Contributions

- Manuel Bichler: organization, discussion, first draft of presentation, researched and added ideas for improvement
- Jack Heseltine: concept, discussion, presentation preparation and delivery slides of intro, person, usage scenario and updated usage scenario
- Dharmik Chaklasiya: concept + slides goal for improvement (another scenario), our idea, usage scenario & update usage scenario, review
- Lisa Sonnleithner: concept, discussion, review
- Tobias Ganzenhuber: discussion, presentation, review