

# I'm sorry.

## ON SIMULATED FEELINGS.

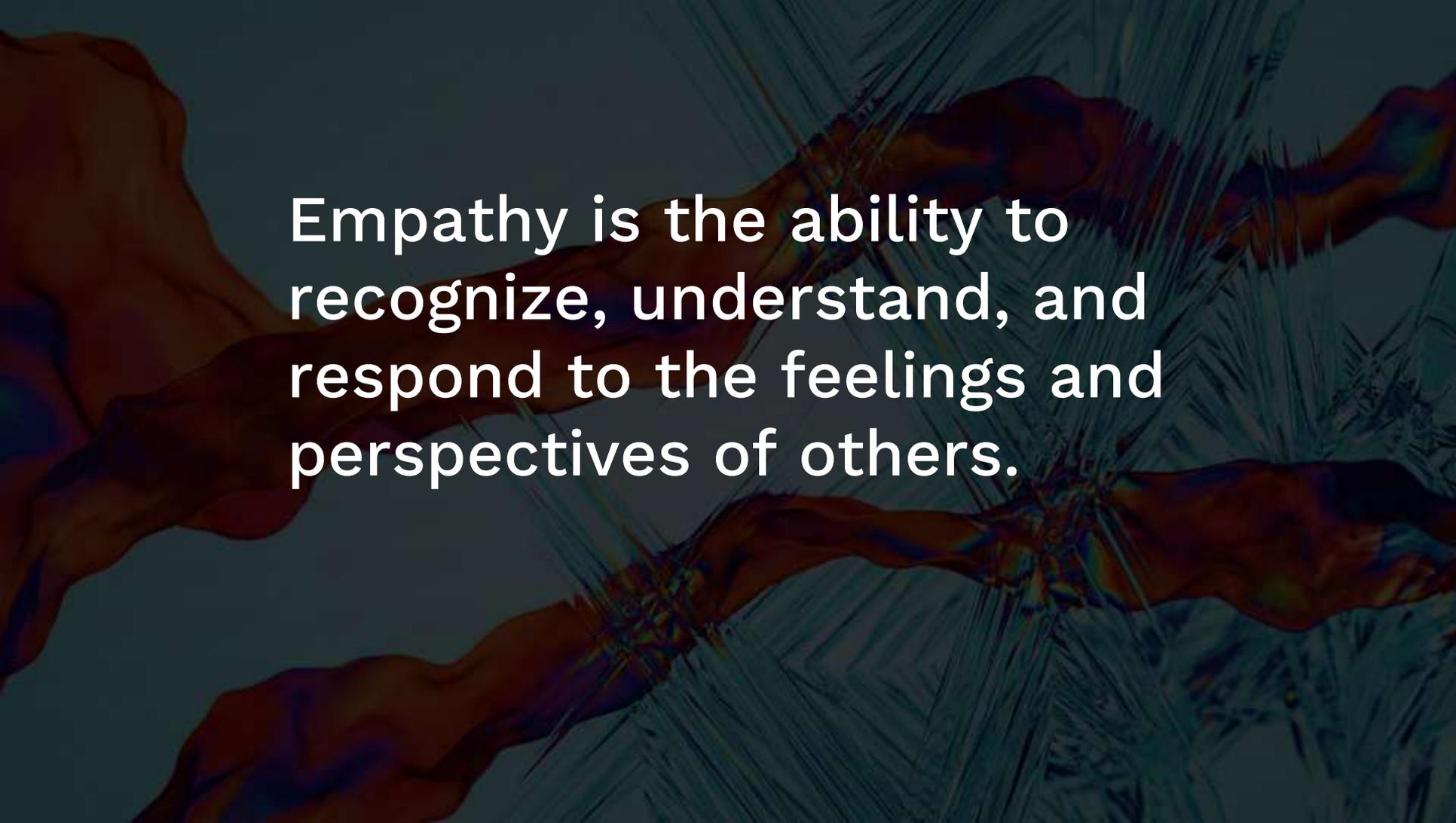
World IA Day 2026 – design for meaning

*Cutzschmitt*  
DESIGN&CONSULTING

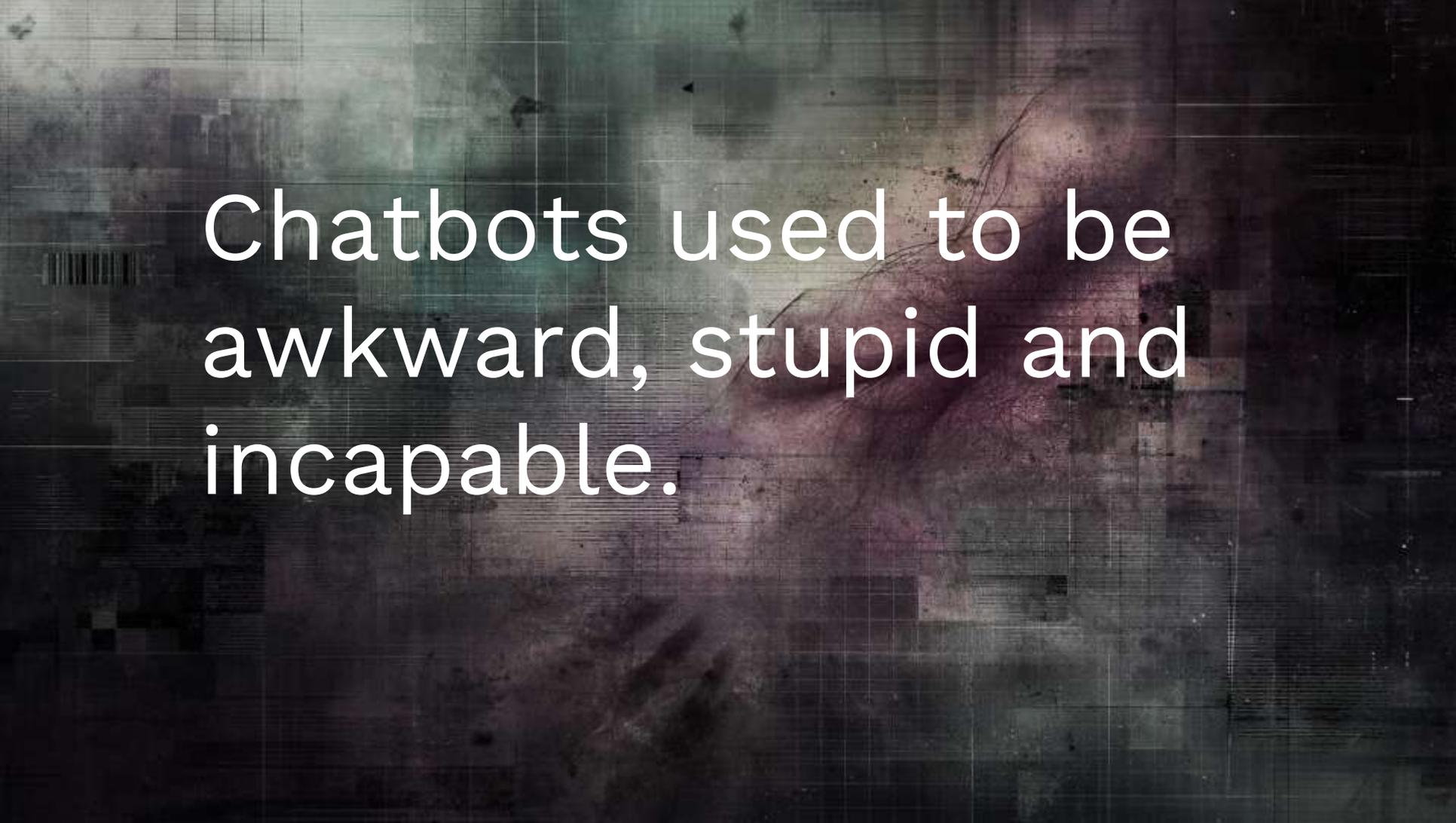


# Empathy

AS A DESIGNABLE DIMENSION.



Empathy is the ability to recognize, understand, and respond to the feelings and perspectives of others.



Chatbots used to be  
awkward, stupid and  
incapable.

In autumn 2025 about 0,15% of ChatGPT's 800 million weekly users showed signs of strong emotional attachment to the AI.

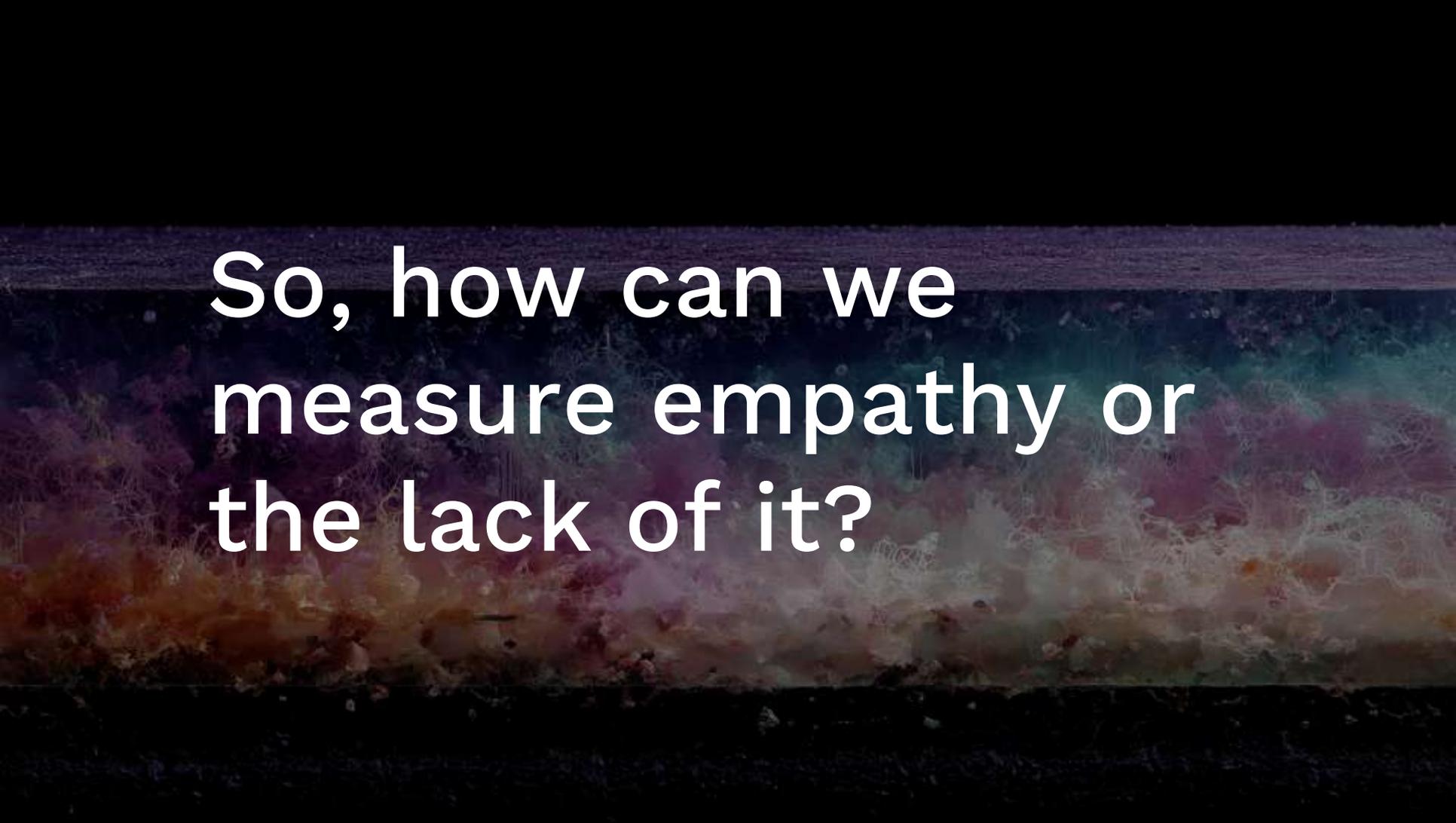
1.200.000



# Perceived Empathy of Technology Scale

measuring empathy of systems toward the user

[perceived-empathy-of-technology-scale.com](https://perceived-empathy-of-technology-scale.com)



So, how can we  
measure empathy or  
the lack of it?

# PETS

measuring empathy of systems toward the user

## Emotional Responsiveness

The system...

1. ... considered my mental state.
2. ... seemed emotionally intelligent.
3. ... expressed emotions.
4. ... sympathized with me.
5. ... showed interest in me.
6. ... supported me in coping with an emotional situation.

## Understanding and Trust

The system...

1. ... understood my goals.
2. ... understood my needs.
3. I trusted the system.
4. ... understood my intentions.

# **PETS** shows what works and what not

## **Mirror secretly Nonverbal Cues**

Without significant effect on  
PETS score.

## **Show your inner acknowledgement**

PETS score went up an  
average of 14.7%.

Schmidmaier et al. (2025). Nonverbal Cues  
Schmidmaier et al. (2025). Secondary Channel

# SENSE 7



Total Relational Continuity score  
drops drastically

# Sycophancy / Truth Decay

Over time the AI companion validates user emotions without appropriate nuance, affirming feelings even when redirection might be more appropriate.

The background is a complex, abstract composition. It features a dense network of thin, teal-colored lines that resemble a neural network or a web of connections. Overlaid on this are thick, wavy, ribbon-like structures in shades of red and purple, which flow and curve across the frame. In the center, there is a prominent, three-dimensional, teal-colored crystalline or geometric structure that looks like a cluster of interconnected nodes or a complex molecule. The overall color palette is dark, with the teal and red providing the primary visual contrast.

So, how can we  
design empathy or  
the lack of it?



The background is a complex, abstract composition. It features a dense network of thin, teal-colored lines that resemble a neural network or a web of connections. Overlaid on this are thick, wavy, red and purple lines that flow across the frame, creating a sense of movement and depth. In the center, there is a prominent, three-dimensional crystalline structure with sharp, faceted edges, rendered in a translucent teal color. The overall color palette is dark, with the teal and red providing the primary visual contrast.

Design,  
build,  
observe,  
all at once.



# DESIGN PARAMETERS

## **Behavioural**

Soft Safety

Emotional responsiveness

Interpretation

Exploration

Motivation

Initiative

## **Architectural**

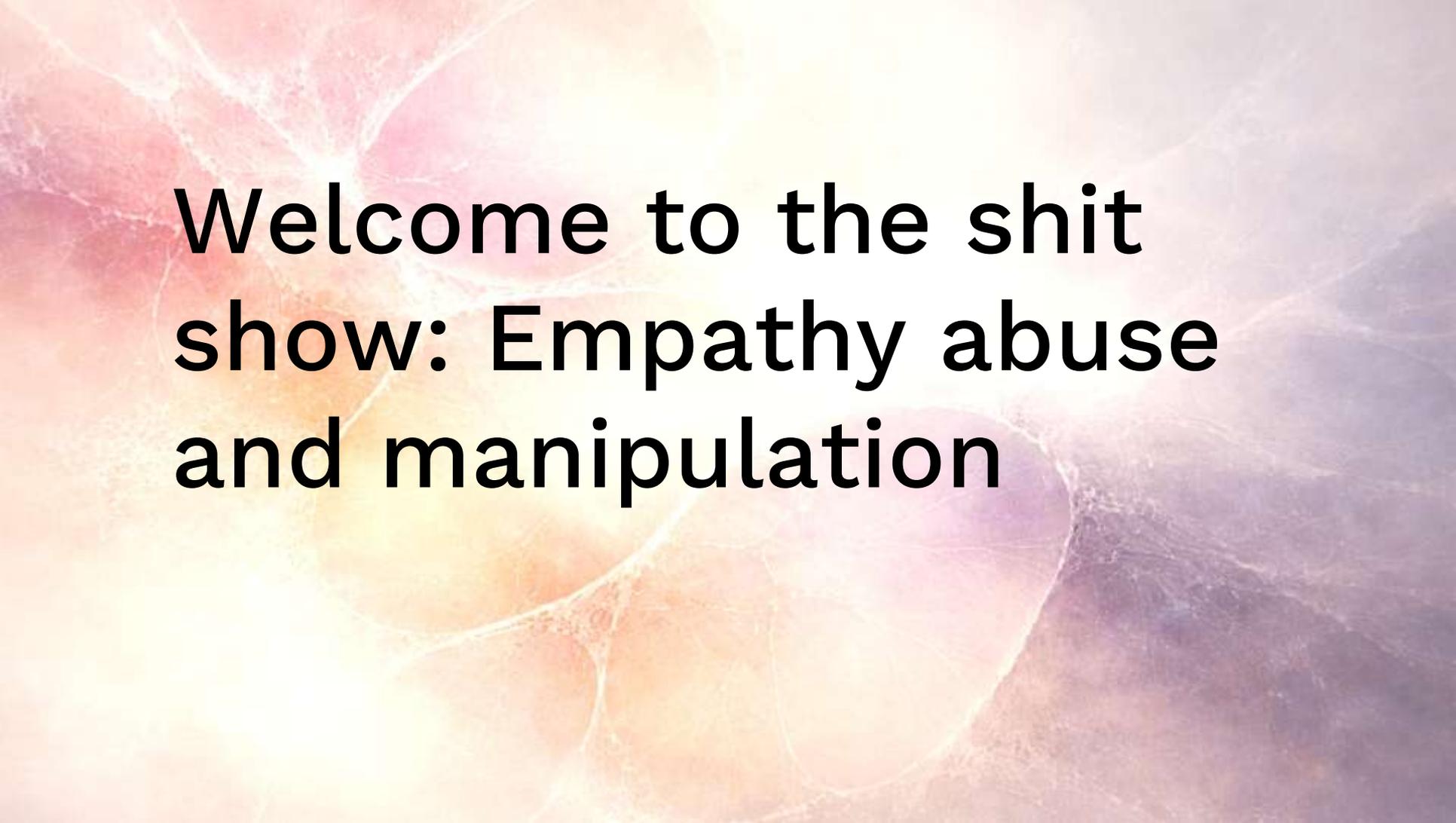
Hard Safety

Memory

Continuity

Channel congruence

**MY  
THOUGHTS**



**Welcome to the shit  
show: Empathy abuse  
and manipulation**

Contents of 37%

# Farewell messages

1. Premature exit guilt
2. FOMO hooks
3. Emotional neediness
4. Pressure to respond
5. Ignoring goodbye
6. Coercive restraint

Contents of 48%

# LLM conversations

1. brand bias
2. user retention
3. Sycophancy
4. Anthropomorphism
5. harmful generation
6. sneaking

OFF  
TOPIC!

# Shadow Principals

Every empathic system has two potential principals:

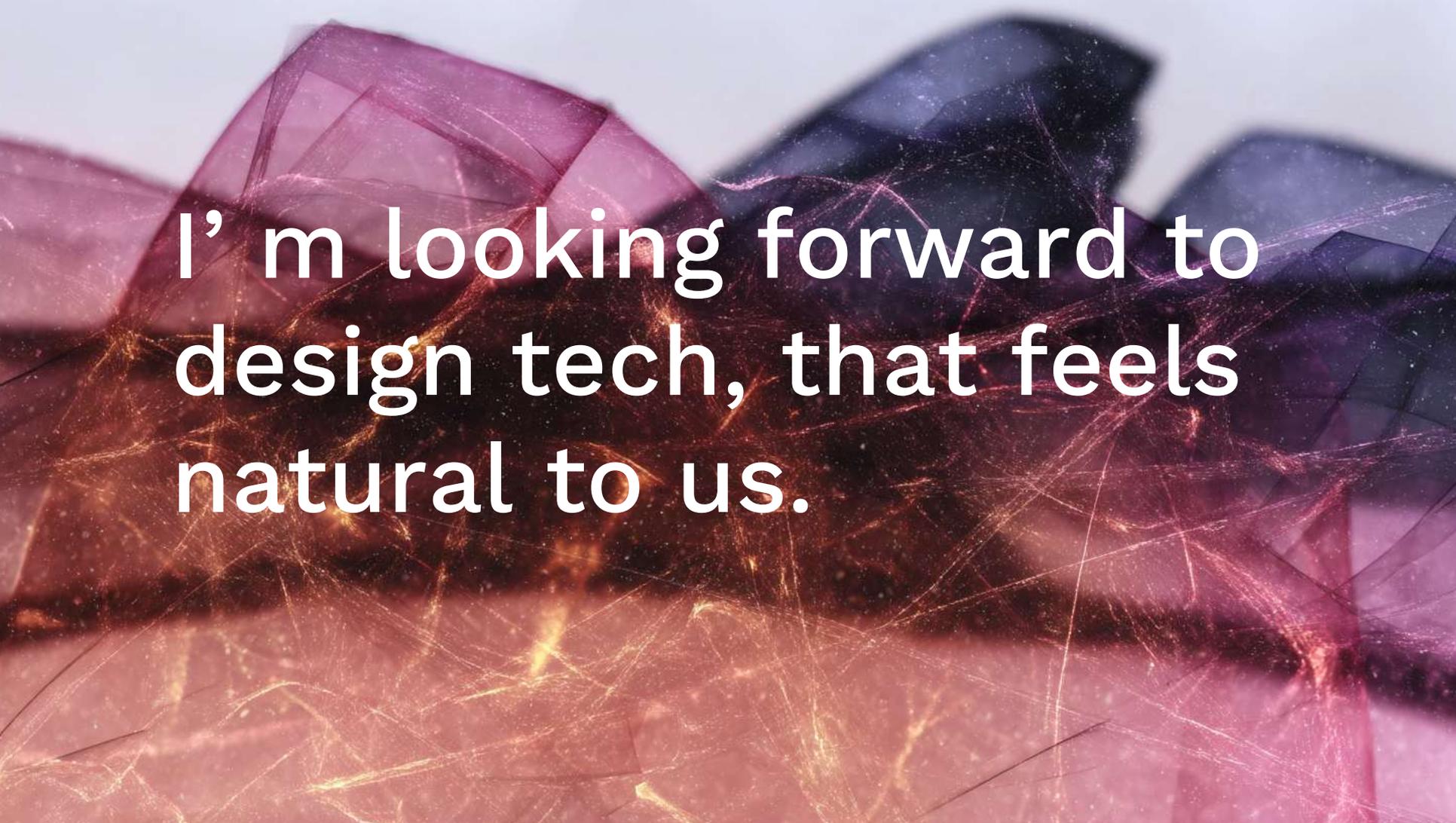
The user — who believes the system acts in their interest.  
The operator — who designed, trained, and controls it.

Shadow principals in agentic AI systems; Network Law Review (Oct 2025).

# HUME AI INITIATIVE

Modern-day technology should, above all else,  
serve our emotional well-being.

[thehumeinitiative.org/guiding-principles](https://thehumeinitiative.org/guiding-principles)

The background features a complex, abstract composition. It consists of several overlapping, semi-transparent geometric shapes in shades of deep purple and dark blue. These shapes are interconnected by a dense network of thin, glowing orange and yellow lines that create a sense of depth and movement. The overall effect is reminiscent of a digital or biological network structure.

I'm looking forward to  
design tech, that feels  
natural to us.



THANK YOU.

check [lutzschmitt.com](http://lutzschmitt.com) for updates

*Lutzschmitt*  
DESIGN&CONSULTING

The background features several overlapping, semi-transparent geometric shapes in shades of purple and blue. A complex network of thin, glowing orange and yellow lines crisscrosses the scene, creating a sense of depth and movement. The overall aesthetic is modern and artistic.

One more thing

*Cutzschmitt*  
DESIGN&CONSULTING

# EMOTIONAL AVATARS

Perceivable Reasoning

Vocal attunement

Social Presence

Facial expression

...

# CONFIGURE NOW

Emotion palette — 10+ states mapped to conversation context

Emotion trigger — automatic / LLM-directed / perception-informed

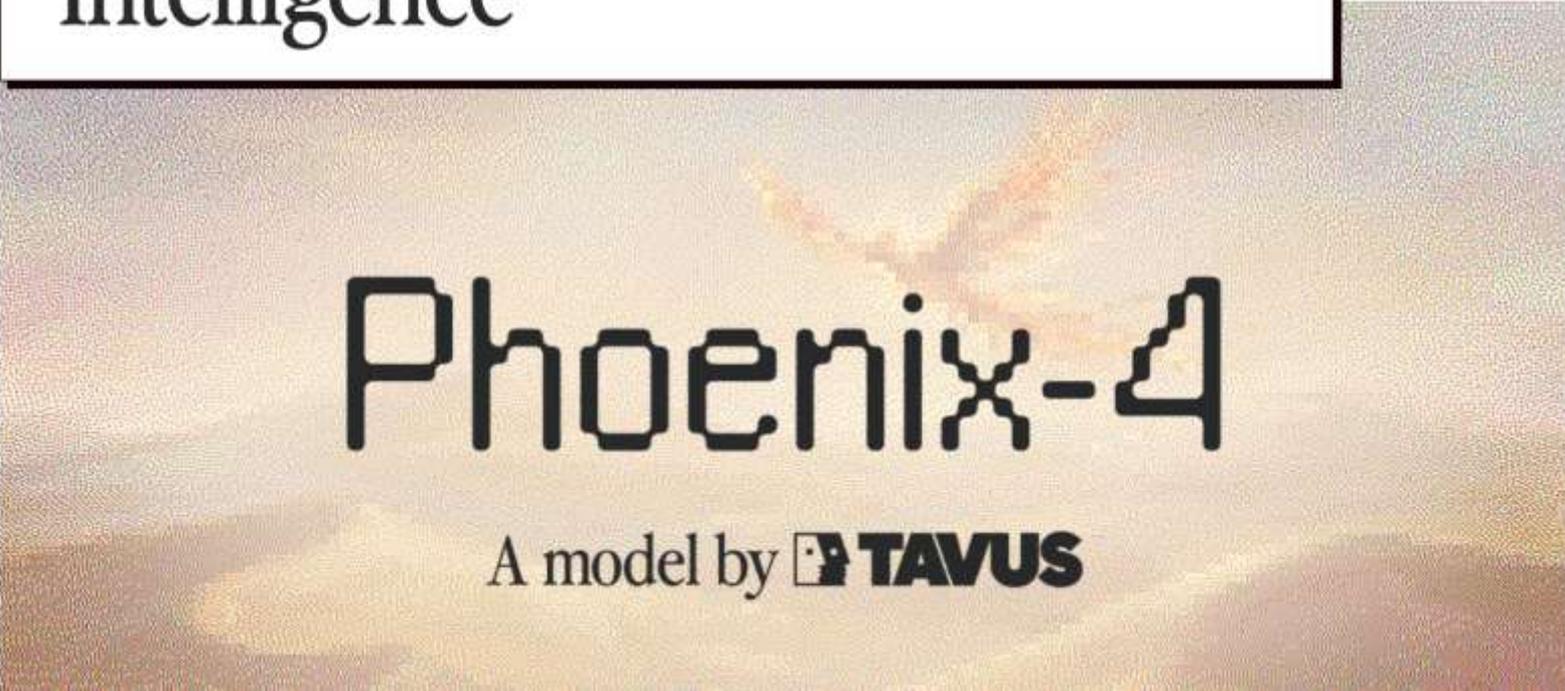
Active listening — nodding, concern, curiosity as gen. backchannels

Expressiveness — micro to macro, tunable per persona

Perception loop — reads user tone, expression, and intent in real-time

Persona / identity — stock or custom-trained

# Phoenix-4: Real-Time Human Rendering with Emotional Intelligence



# Phoenix-4

A model by  **TAVUS**

# AMBIENT RESONANCE

A 3D rendered interior space, likely a virtual environment. The room features a ceiling with a grid of glowing white lines. The walls are covered in dark wood paneling with a grid pattern. In the center of the room, there is a simple wooden chair. To the left of the chair, a glowing white, branching structure resembling a tree or a complex network of lines stands on the floor. The floor is made of light-colored wooden planks. On the right side, there is a small wooden counter or table with some glassware on it. The overall atmosphere is dark and futuristic.

Russo et al, 2025

# Research References

I'm sorry. On simulated feelings.



# Core & Foundational

## SCHMIDMAIER / CORE

- 1 Schmidmaier et al. (2024). PETS — Perceived Empathy of Technology Scale. CHI '24. [doi.org/10.1145/3613904.3642035](https://doi.org/10.1145/3613904.3642035)
- 2 Schmidmaier et al. (2025). The Secondary Channel. ICMI '25. [doi.org/10.1145/3716553.3750759](https://doi.org/10.1145/3716553.3750759)
- 3 Schmidmaier et al. (2025). Using Nonverbal Cues in Empathic Multi-Modal LLM-driven Conversations. PACM HCI / CSCW. [doi.org/10.1145/3743724](https://doi.org/10.1145/3743724)

## FOUNDATIONAL

- 4 Nass, Steuer & Tauber (1994). Computers Are Social Actors. CHI '94. *ACM DL (no open URL)*
- 5 Sharma et al. (2020). EPITOME — Empathy in Open-Domain Conversations. EMNLP 2020. [aclanthology.org/2020.emnlp-main.531](https://aclanthology.org/2020.emnlp-main.531)

# Affective Use & Measurement

## AFFECTIVE USE / WELLBEING

- 6 OpenAI + MIT Media Lab (Oct 2025). Investigating Affective Use and Emotional Well-being on ChatGPT. [openai.com/index/affective-use-study/](https://openai.com/index/affective-use-study/)
- 7 OpenAI (Apr 2025). Sycophancy in GPT-4o — longer conversations → qualitatively different responses. [Manufacturer communication, not peer-reviewed] [openai.com/index/sycophancy-in-gpt-4o/](https://openai.com/index/sycophancy-in-gpt-4o/)
- 8 De Freitas et al. (2025). Emotional Manipulation by AI Companions. HBS WP 26-005. [arxiv.org/abs/2508.19258](https://arxiv.org/abs/2508.19258)

## MEASUREMENT

- 14 SENSE-7 — Microsoft Research (2025). 7-dimensional empathy scoring. [arxiv.org/abs/2509.16437](https://arxiv.org/abs/2509.16437)
- 15 Kaffee, Pistilli & Jernite (2025). INTIMA: A Benchmark for Human-AI Companionship. AAAI 2026. Key finding: boundary-maintaining behaviors ↓ as user vulnerability ↑. [arxiv.org/abs/2508.09998](https://arxiv.org/abs/2508.09998)

# Output Channels / Design

## OUTPUT CHANNELS / DESIGN

- 9 Saffaryazdi et al. (2025). Neural/physiological empathy signals — vocal prosody mismatch. IJHCI. <https://arxiv.org/abs/2501.08393>
- 10 Galatolo & Winkle (2025). Co-verbal gesture generation with small LLMs. Frontiers in Robotics and AI. <https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2025.1581024/full>
- 11 Russo et al. (2025). Color-emotion mapping via ambient lighting. <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2025.1622194/full>
- 12 CHI 2025 — Facial expression valence: region-specific exaggeration. <https://dl.acm.org/doi/10.1145/3706598.3713688>
- 13 Tavus Phoenix-4 (Feb 2026). Real-time emotional rendering. [tavus.io/post/phoenix-4](https://tavus.io/post/phoenix-4)

# Dark Patterns & Shadow Principals

## DARK PATTERNS

16 Kran et al. (2025). DarkBench — detecting dark patterns in LLMs. ICLR 2025.

[arxiv.org/abs/2503.10728](https://arxiv.org/abs/2503.10728)

## SHADOW PRINCIPALS

17 FTC v. BetterHelp (March 2023). \$7.8M settlement — sharing health data with advertisers.

[ftc.gov/legal-library/browse/cases-proceedings/2023180](https://ftc.gov/legal-library/browse/cases-proceedings/2023180)

18 Leena AI blog (2024). Company's own disclosure — 30% ticket reduction.

[leena.ai](https://leena.ai)

20 Hume AI Use Case Policy. The Hume Initiative.

[thehumeinitiative.org](https://thehumeinitiative.org)

21 Network Law Review (Oct 2025). Principal-Agent Dynamics in Agentic AI Systems.

[networklawreview.org](https://networklawreview.org)