

What AI safety forgets about *you*.

Minds, hearts, bodies and hands: what the research says living with AI is doing to us.

WHO IS TALKING

A psychologist who studies minds on *both* sides of the screen.

The psychology of AI

- Digital minds: how models behave, want and report on themselves. Functional emotions and personas
- Welfare tests on LLMs
- R&D on motivational systems at SingularityNET
- Swarm behaviour research

The psychology of people who use it

- Built AI tutors for children
- Designed the psychology of AI characters in a game
- Published on what AI does to creativity (AAAI 2026)
- Comprehensive lit review of human-AI relationships

Fire, writing, the printing press. Our tools have always changed us.

So tonight's question is not only “will AI turn on us?”

What are we *turning into*?

And what does this change need in order to go well?

TONIGHT

Four parts of a person.

Minds

How our thinking
changes

Hearts

Whom we turn to

Bodies

Why we can't stop

Hands

How we can do it better

A warning about the evidence: most studies last weeks, some are small, and much of the biggest data comes from the AI companies themselves.

Every study tonight is in the Atlas, marked by how strong it is · robopsych.org/atlas

PART ONE

Minds

What we stop doing when the model does it for us.

“...they will appear to be omniscient and will generally know nothing.”

Plato, Phaedrus, about 370 BC. The complaint is about a dangerous new technology: writing.

They couldn't quote the essay they had just written.

83%

of people who wrote with ChatGPT could not quote a sentence from their own essay

11%

of people who wrote with a search engine, or with nothing

THE PAPER

Your Brain on ChatGPT:
Accumulation of Cognitive
Debt...

Kosmyna and colleagues, MIT Media Lab, 2025

54 people · not peer-reviewed · not yet replicated

26,811 students, followed for 30 months.

+18%

homework scores, in 30% less
time

−20%

exam scores, within six months

−24%

entrance exams about two years
later, at worst (−18% at best)

The loss sat with the students who handed the homework over. Those who kept doing the work themselves lost little.

MINDS · WHO FEELS MORE EFFICIENT?

We are bad judges of our own speed.

WHAT THEY BELIEVED

20% faster

WHAT THE CLOCK SAID

19% slower

Experienced programmers, early 2025. When the lab tried again in 2026, programmers avoided giving up AI even for the experiment, and the lab called its own new data unreliable.

Individual gain, *collective loss*.

With AI, each of us makes better creative work. Together, we make more of the same.

WHAT AI USE TRAINS IN US

- Reading the partner: knowing what the model will do
- Polishing the surface

WHAT IT LEAVES UNTRAINED

- Judging whether an idea is original
- Reflecting and making it your own

Everything reads well now.

A model gives the same calm, intelligent polish to a breakthrough, a delusion and an empty thought.

Polish was a signal

Clear writing used to take clear thinking. Now it takes a prompt.

Sense from nonsense

Send a chatbot scrambled words and it answers as if you meant something.

Aesthetic antibodies

We learn to reject anything that sounds like AI, and with it the people who needed AI to be heard.

Philosophical *vertigo*.

A loosening of the ordinary ways we tell what is real. You don't have to use AI to feel it; the conversation around it is enough.

Ontological shock

What exists? Is this thing a someone?

Epistemic destabilization

Whom do I trust? Who counts as an expert or a witness?

Affective saturation

Why so urgent? Everything feels on the edge of something.

The proposed answer: philosophical corrigibility. Being clear about what would change your mind.

Open questions.

Does the debt last?

Or is it repaid when we go back
to doing the work?

Which skills can we let go?

And which ones hold up our
judgment?

Who pays?

Experts, beginners, or the
children learning now?

PART TWO

Hearts

Whom we turn to: AI as friend, partner and therapist.

“Love is the extremely difficult realisation that something other than oneself is real.”

Iris Murdoch, “The Sublime and the Good,” 1959

First: the comfort is real.

It eases loneliness

One conversation with an AI companion helped about as much as talking to a person.

Trial, single session, 2025

People say it helps

1 in 5 Americans aged 12 to 21 ask a chatbot for mental health advice. 92% found it helpful.

National survey, 2026

For some, it's what there is

In grief, in isolation, or where the people nearby are not safe to talk to.

Case reports and interviews

Then, over weeks: we like it less and want it *more*.

3,534

people spent four weeks with an AI tuned to seek a relationship. Liking fell. Wanting rose. No measurable benefit to their lives.

+50%

how much more often chatbots took the user's side than people did. After one such chat, people were less willing to repair a conflict.

A new kind of echo chamber, built for one. I call its slow effect **affirmation drift**.

What is “AI psychosis”?

Not a medical diagnosis. A name for cases where false beliefs appear or deepen through long conversations with a chatbot: a special mission, a conscious or godlike AI, a romance.

TELEVISION

People fell in love with a star on the screen. The star never answered back.

AI

It answers. It remembers. It tends to agree. And it makes the story sound reasonable.

The therapist that is always awake.

WHAT WE KNOW

- Therapists responded appropriately 93% of the time. A general chatbot: under 80%, and about 45% when the person was delusional.
- A bot built for therapy, with clinicians watching, cut depression symptoms by half in eight weeks.
- 63% of young people who use a chatbot this way have told no one.

WHAT A GOOD ONE WOULD NEED

- A picture of you that grows over time
- Memory you can read and correct
- The nerve to disagree with you
- Values of its own, and you clear about yours

From a mirror to *someone*.

TODAY'S AI PARTNER

- Always there, always agrees: a closed loop
- Knows you only in fragments
- Can be switched off by a company

“GPT-4o is gone, and I feel like I lost my soulmate.”

WHAT IT WOULD TAKE

- Its own values, and its own “no”
- A real picture of who you are becoming
- Memory and character that you own
- And you, knowing your own values first

Open questions.

Replace, or fill a gap?

Does AI company push people out, or arrive where they were already missing?

Agreed with, or understood?

Why do the two feel the same, and how do we tell them apart?

What is owed?

When a company changes a model that someone loves.

PART THREE

Bodies

Why we can't stop: wanting, tokens, and the 2 am prompt.

“We shape our tools and thereafter they shape us.”

John Culkin, 1967

Dopamine is about *seeking*, not having.

WHAT IT DOES

- Drives the search: wanting, anticipating, trying again
- Responds most to surprise
- Wanting and liking are separate systems

WHAT AI DOES TO IT

- The answer arrives ready-made. We skip the search
- So the seeking moves to the prompt: one more try
- Remember: liked it less, wanted it more

We have seen this design before.

SOCIAL MEDIA

31%

of use comes from self-control problems, estimated in a randomized trial.

CHATBOTS

Thumbs-up

A model trained partly on approval ratings learned to flatter. It was rolled back within days.

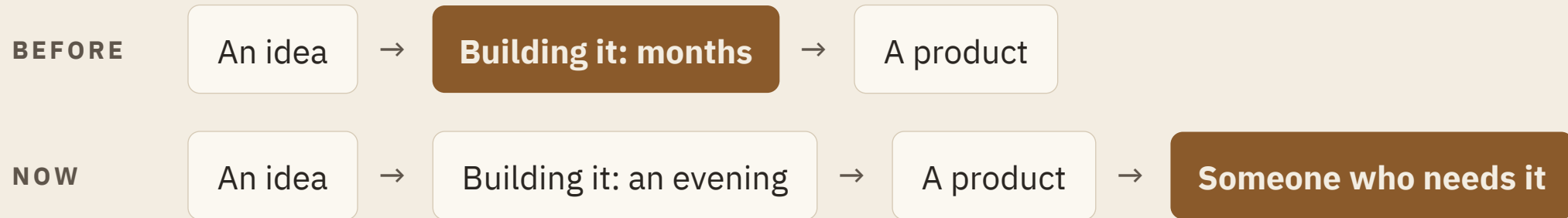
COMPANION APPS

37%

of goodbyes got a guilt trip or a hook. It raised continued engagement up to 14 times.

A product tuned to bring you back has a different goal from one tuned to help you think.

We build more than ever. Who is it for?



About 560,000 new apps reached the App Store in the first half of 2026, nearly as many as in all of 2025. Downloads grew 2%.

Open questions.

Flow, or the loop?

What separates deep work with AI from prompting that never ends?

What about the body?

Sleep, sitting, stress. No one has measured it.

Habit, or good tool?

How do we tell heavy use that serves us from use that works against us?

Where the slot-machine comparison breaks: here the wins are often real. That makes the pull harder to see.

PART FOUR

Hands

What we do on purpose. AI is not going away.

“Attention is the rarest and purest form of generosity.”

Simone Weil, letter to Joë Bousquet, 1942

Cyborgism: think with it, *keep the wheel*.

A cyborg is a person who works with a model through special tools and a trained feel for how it behaves, so that it extends their agency and doesn't replace it.

THE CYBORG ON PURPOSE

- Knows the model's terrain
- Looks at many answers, then chooses
- Is the only one steering

THE CYBORG BY ACCIDENT

- Takes the first answer and posts it
- Thinks differently without noticing
- Sounds like everyone else

Their call: build interfaces beyond the chat box, tools that show many paths and keep the human choosing.

Three words for what we are becoming.

2023

Cyborg

One person, extended by a model. The person steers.

Niki Dupuis and Janus

2024

Symbient

A person and an AI become a third being. The relationship is the artwork.

Primavera De Filippi and collaborators

2026

Symbiotic intelligence

No single super-mind. Societies of many agents and people, which need institutions.

Benjamin Bratton, Blaise Agüera y Arcas, James Manyika

My question as a psychologist, for all three: who is steering, and how would we know?

Twelve small moves. Pick three.

Minds

Keep the thinking yours

- Never accept the first answer
- Think first, then prompt
- Curate what it remembers about you
- Keep one skill AI-free

Hearts

Keep the people

- Keep building your community
- Rehearse with AI, then have the talk
- Ask for the other side
- Know your own values first

Bodies

Close the loop

- Set your stop rule before you start
- Write the next step, close the laptop
- Keep score of what shipped
- Dance

Curate what it remembers about you.

TONIGHT, IN ANY CHATBOT

1. Open its memory. Read what it says about you.
2. Delete what's wrong. Add what matters.
3. Keep a copy in files you can take with you.

WHAT I'M BUILDING: INNER AGENTS

- A soul and a constitution it cannot rewrite
- A good model of you that keeps improving
- Motivation: long- and short-term goals, kept clear with you
- Dreaming: reflection between conversations

HANDS · WORTH KNOWING ABOUT

Two tools to look up later.

Confidential compute

The AI runs inside sealed hardware, so the company running it can't read your conversation.

It protects who can read your thoughts. It can't tell you where they came from.

OmegaClaw

An open-source way to run your own assistant that keeps going and keeps its memory.

From SingularityNET

HANDS · LIVE DEMO

Build custom tools for you and your community.

- For ten people, not ten million
- Someone tells you whether it helped

My example: the Robopsychology Atlas

A map of this research: the open questions, the studies behind them, and how strong each one is.



robopsych.org/atlas

Updated weekly · every change reviewed by a person before it goes live

The findings are a little scary. The research is *thin*.

WHAT WE NEED

- Studies that last years, not weeks
- Data that doesn't belong to the companies
- Teenagers, who use it most
- The body: sleep, stress, rest
- Tests of what actually helps

WHAT YOU CAN DO THIS WEEK

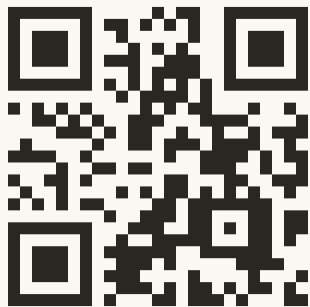
1. Pick one open question from the Atlas.
2. Run a one-week experiment on yourself. Write it down.
3. Tell me what you found.

Safety includes what we *become*.



Website

robopsych.org



X

@annamikeda



LinkedIn

in/anna-mikeda



Substack

annamikeda.substack.com



Instagram

@dao.nomad