SUMMERSOC 2026 · DISTRIBUTED INTELLIGENCE

AI as a Human-Centered Personal Service

From cloud services to background agents

Albrecht Schmidt · LMU Munich

What part of your cognition is technology-free/AI-free?

AI is moving from applications into everyday thinking.

Already Today AI helps us search, write, summarize, decide, create, navigate, remember, and act.

NEXT

AI services become personal, context-aware, proactive, **and agentic**.

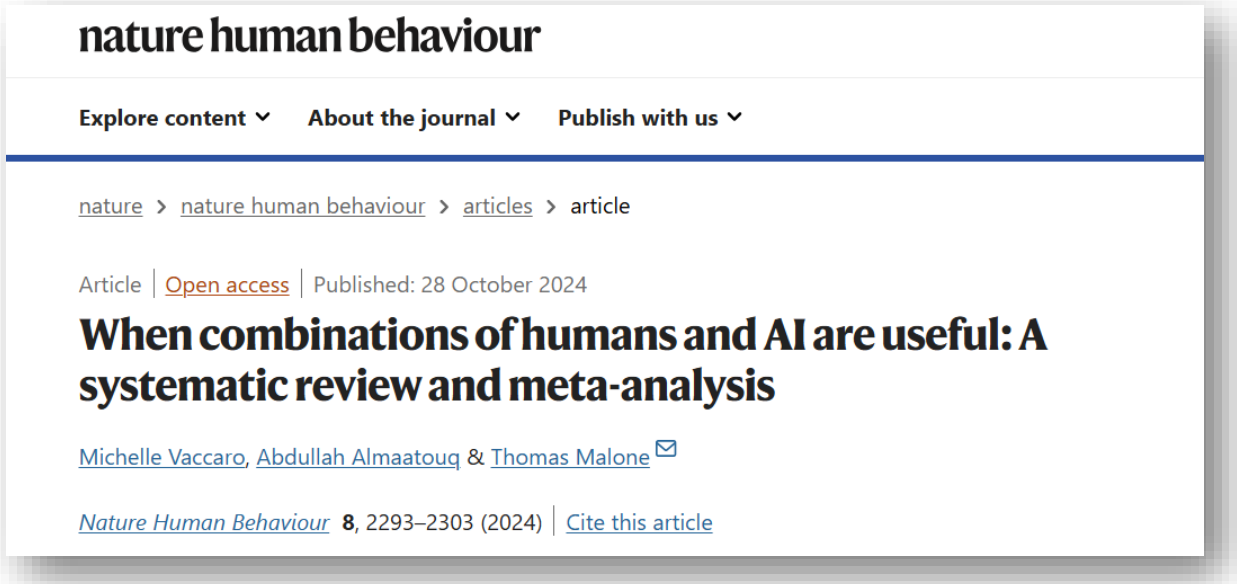
AI woven into perception and action

designing the shift towards human-centered intelligent services

Interactive cognitive augmentation

Hypothesis: AI support can **enhance metacognition**, **memory, creativity, and decision-making**.

"[…] human-AI combinations performed significantly worse than the best humans or AI alone […]"



nature human behaviour

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature human behaviour › articles › article

Article | Open access | Published: 28 October 2024

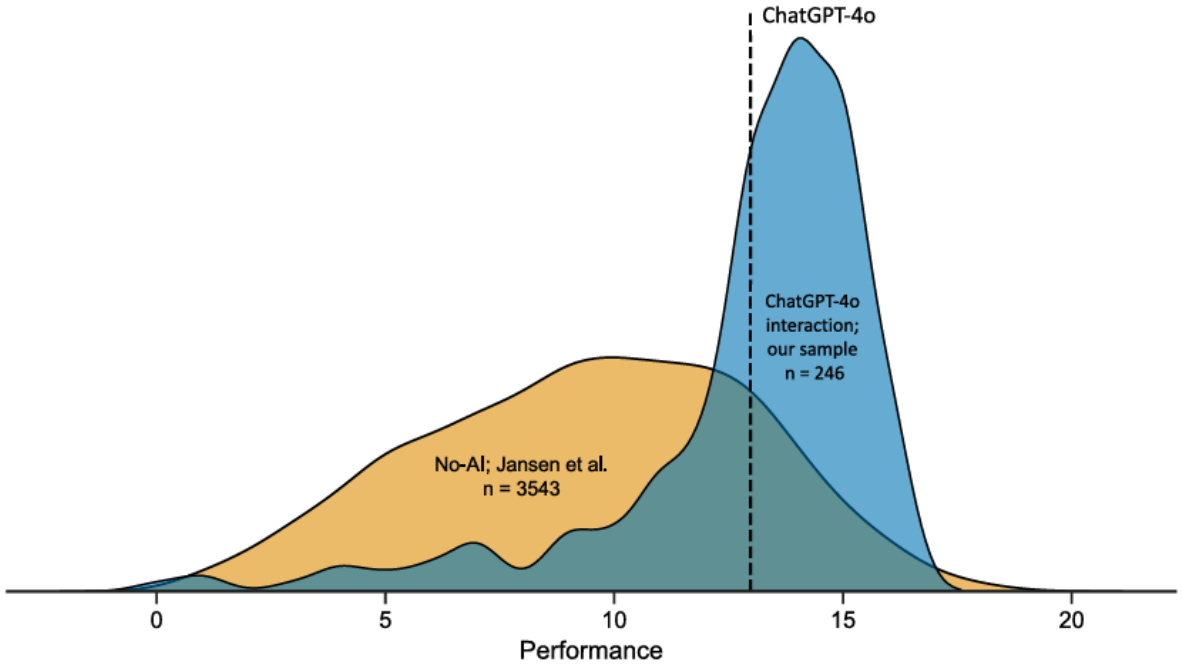
When combinations of humans and AI are useful: A systematic review and meta-analysis

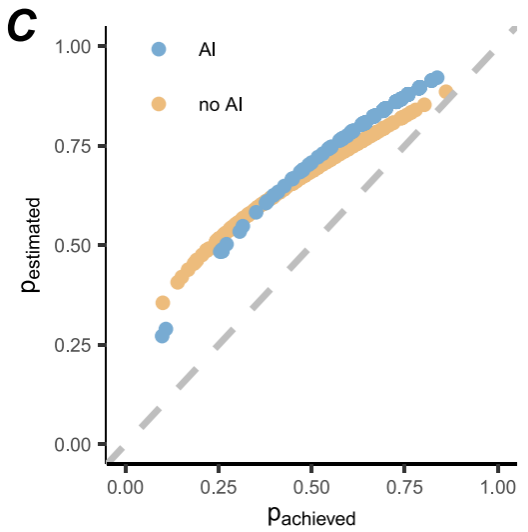
Michelle Vaccaro, Abdullah Almaatouq & Thomas Malone ✉

Nature Human Behaviour **8**, 2293–2303 (2024) | Cite this article

Cognitive augmentation

- Externalize human memory and the ability to recall
- Altering and extending human perception
- Increasing bandwidth between humans and (computing) systems
- Easing cognitive tasks while staying in control
- The understanding of **meta cognition**





CHB2026

Why this belongs to Service oriented Computing

- No longer only about business functions
- Distributed intelligence connects cloud, edge, data, models, tools **and humans**
- Agentic AI turns services into systems that plan and act
- Next: personal services that **amplify human cognition**

SOC VOCABULARY

cloud · edge · microservices
orchestration · data products
security · privacy · agents

HCI VOCABULARY

agency · control · cognition
augmentation · interaction
benefit · transparency

Making services personal and human-centric

CORE SHIFT

From service to background agent

Service
waits for request

→

Assistant
responds in dialogue

→

Agent
plans and executes

→

Background agent
monitors, anticipates, acts

Escalation: response → dialogue → action → ongoing delegated action

The more the system acts, the more human-centered control matters.

The end of serendipity

No randomness anymore?

Who do you trust to decide…

- …what movies you watch?
- …who sits next to you?
- …which way you walk?
- …who is in the same restaurant?
- …which flat you buy?

recommendation quietly becomes steering

steering becomes action

action becomes delegated life logistics

Convenience can quietly become invisible steering. and then delegated action.

HUMAN-CENTERED QUESTION

Who is in control?

Who decides what you know?

Who creates your reality?

Who acts for you?

Do you know, it is not your decision?

Agentic AI makes the old control question urgent — the system can act on your behalf.

DEFINITION

Interactive Human-Centered AI

> **An Interactive Human-Centered Artificial Intelligence** enables interactive exploration and manipulation in real time, is designed with a clear purpose for **human benefit**, while being transparent about **who has control** over data and algorithms.

Key dimensions: interaction · benefit · control

Human-centered AI as service requirements

Interactivity

real-time exploration and manipulation

Observability

see effects, explanations and alternatives

Benefit + risk

state the human benefit and possible harms

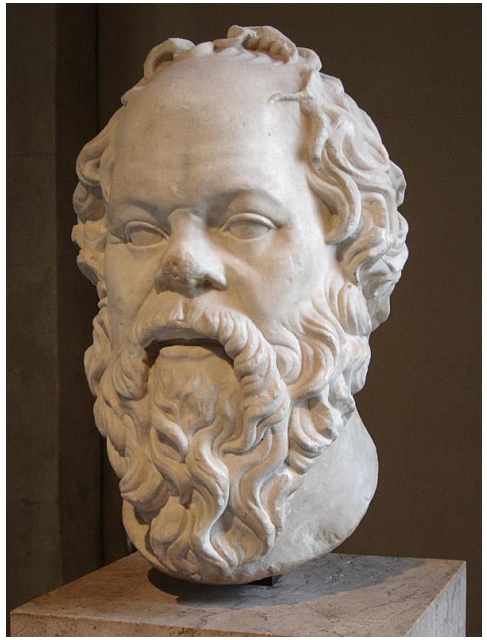
Control + provenance

who controls what · what data informed it

In earlier work: 8 properties — today compressed into 4 requirements.

For personal agents these are not ethical add-ons — they are service design requirements.

Cognitive tools: a long trajectory



- Tools externalize memory, reasoning, navigation, planning, perception
- AI turns tools from passive artifacts into adaptive services
- Agentic AI adds the capability to act over time

writing · maps · calculators · search · AI agents

Personal AI is the next step in this lineage. Now it is able to act, not only assist.

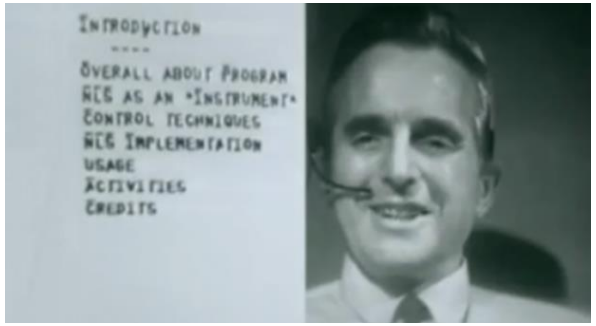
DESIGN GOAL

Augmentation, not automation

Design human–machine systems that outperform humans alone and machines alone

Bush
As We May Think

Licklider
Man–Computer Symbiosis



Engelbart
Augmenting Human Intellect

Weiser
Ubiquitous Computing

ML and automation are components. HCI creates intelligent human–AI systems.

Bush, V.,1945. As we may think. *The Atlantic*, July 1945, pp.101-108.
Licklider, J. C. R., 1960. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1, March 1960, pp 4-11
Engelbart, D. C. 1962. Augmenting Human Intellect: A Conceptual Framework. *SRI Summary Report AFOSR-3223*. October 1962.
Weiser, M., 1991. The computer for the 21st century. *Scientific American* 265.3, pp94-104, 1991.

CORE MODEL

AI as a cognitive tool implemented as a service

PERSONAL CONTEXT IN

documents · notes · emails · photos · locations · preferences · tasks · current situation · sensors

→ **Personal AI Service** →

SUPPORT + ACTION OUT

answers · summaries reminders · recommendations · plans · interventions · delegated actions

Personal in use, personal in grounding, service-oriented in implementation.

INTERACTION

Modes of personal AI support

Reactive

answer a question using my context

Proactive

suggest the next step before I ask

Explicit

adapt the explanation to my state

Implicit

warn, remind, or act in the background

Background agents live in the proactive + implicit quadrant; most powerful and the most risky.

AGENTIC AI

Services that plan and act

Goal
what + why

→

Plan
decompose task

→

Tools
call services

→

Execute
act over time

- Agents coordinate tools and services across multiple steps
- Background agents monitor context and opportunities
- Human-centered design determines what autonomy is acceptable

CONTROL MODEL

Delegation instead of automation

Human

purpose · values
constraints · permissions
what + why

⇄

Agent

plans · proposes
executes · monitors
asks back when needed

⇄

World

services · tools
people · organizations
physical context

The interface to agentic AI is a delegation relationship: inspect, redirect, pause, undo.

Background agents need personal grounding

Generic agent

knows common patterns
optimizes generic objectives
efficient but not situated

Personal agent

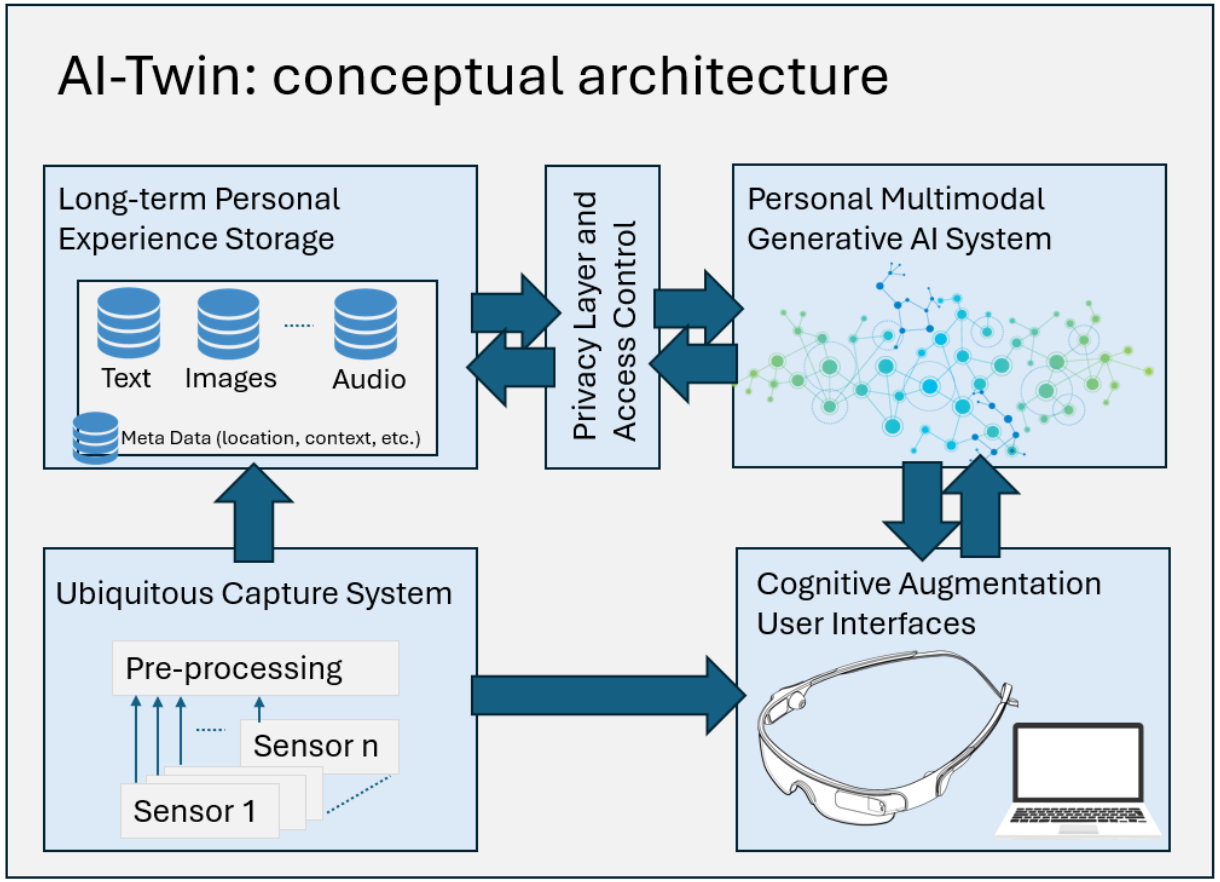
knows preferences, constraints,
routines, goals, history
and current context

Human-centered agent

makes assumptions visible
asks when needed keeps agency with the person

The more personal the agent, the more useful — and the higher the need for control.

Grounding layer for personal agents



Ubiquitous capture
multimodal experience

Personal experience store
secure long-term memory

Personal generative model
RAG · tuning · context

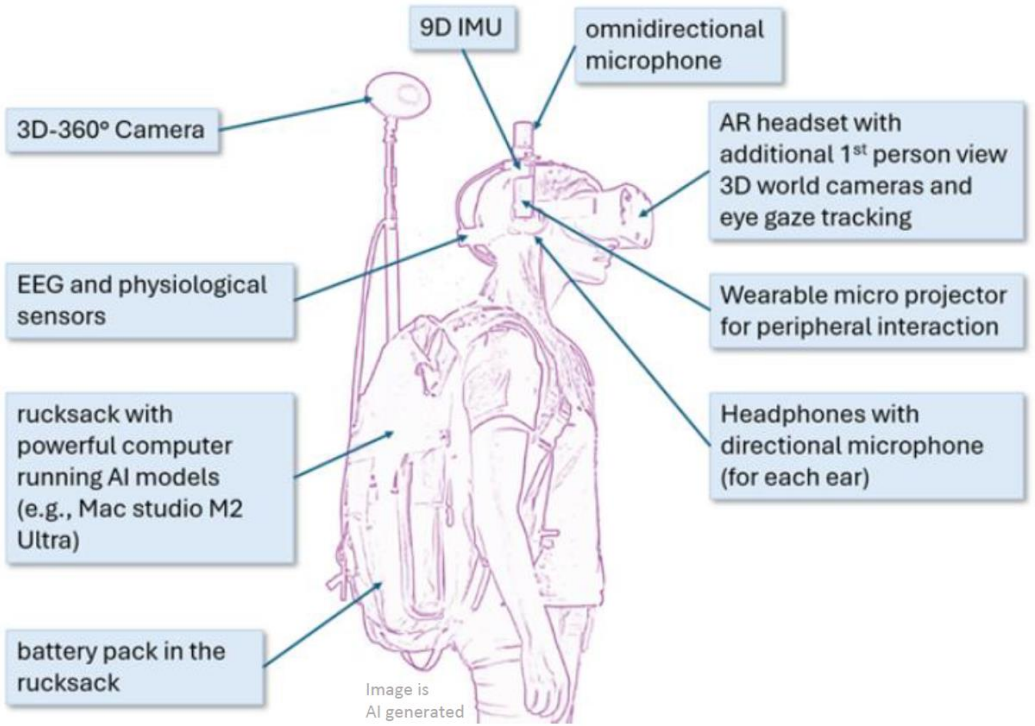
Cognitive interfaces
AR · voice · text · ambient

Privacy + control layer
permissions · provenance

AI-Twin = personal cognitive substrate for agentic AI — inspectable and interruptible.

Experimental research approach

Research through functional prototyping



- Wearable multimodal capture rig
- Mobile compute running personal AI models
- Three iterative prototype generations
- Studied in real, situated everyday use

Work in Progress: Wearable AI-Twin prototype

Fresh, situated, multimodal capture grounds the personal model in real human experience.

Experience Shapes Intelligence and Identity

We Learn from our Environment

LLMs
Llama 3.3
Llama 3.2-Vision
Deepseek-r1
Mistral-nemo

Stable Diffusion
SDXL
Stable Video 4D

+

Captured Experiences

こんにちは

Bonjour

→

Dynamic Personal Foundation Model

Semi-Formal System Description for AI-Twin

Components:

- L **(Learning System):** A pre-trained AI model initialized with general knowledge (L_0), continuously updated.
- E **(Environment):** Multimodal input stream representing real-world experiences:
  - T (Text): Conversations, documents.
  - V (Visual): Images, videos.
  - C (Context): Location, time, interactions.
  - P (Physiological): Gaze, heart rate.
- H **(Human):** Interacts with and controls the system.

Processes:

- **Continuous Updates:** $L \leftarrow L + \Delta E$ (learning system evolves with new input).
- **Human Interactions:**
  - **Query Response:** $R = f(Q, L)$ (human queries the system).
  - **Proactive Suggestions:** $P = g(L, E)$ (system provides suggestions).

Objective:

The AI-Twin aims to **maximize the human's cognitive benefit** ($\mathcal{U}$) over time, defined as:

$$\mathcal{U} = \max \sum_{t=0}^{T} \left[\text{Engagement}(t) \cdot \text{Effectiveness}(R(t), P(t)) \right]$$

where:

- Engagement(t) is the **human's interaction and trust** at time t.
- Effectiveness($R(t), P(t)$) measures the **impact of AI-Twin outputs** (query responses and proactive advice) on **task performance, decision-making, and cognitive augmentation**.

Why personal information matters

- Generic models are powerful, but are not situated in a person's data and life
- Personal context constrains the model and makes support relevant and more precise
- Personal memory enables continuity across tasks and time
- Provenance becomes possible: why did the system suggest this?

DESIGN TENSION

The most useful personal agent is also the most intimate one…
We need to build control into the architecture!

lived experience as grounding

Fresh experience vs. model degeneration

CLOSED SYNTHETIC LOOP

models produce content
models train on content
diversity and grounding degrade

FRESH PERSONAL EXPERIENCE

real-world data
lived context
multimodal evidence

nature

Explore content · About the journal · Publish with us

nature › articles › article

Article | Open access | Published: 24 July 2024

AI models collapse when trained on recursively generated data

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson & Yarin Gal

Nature **631**, 755–759 (2024) | Cite this article

408k Accesses | **3083** Altmetric | Metrics

Abstract

Stable diffusion revolutionized image creation from descriptive text. GPT-2 (ref. [1]), GPT-3(.5) (ref. [2]) and GPT-4 (ref. [3]) demonstrated high performance across a variety of language tasks. ChatGPT introduced such language models to the public. It is now clear that generative artificial intelligence (AI) such as large language models (LLMs) is here to stay and will substantially change the ecosystem of online text and images. Here we consider what may happen to GPT-{*n*} once LLMs contribute much of the text found online. We find that indiscriminate use of model-generated content in training causes irreversible defects in the resulting models, in which tails of the original content distribution disappear. We refer to this effect as 'model collapse' and show that it can occur in LLMs as well as in variational autoencoders (VAEs) and Gaussian mixture models (GMMs). We build theoretical intuition behind the phenomenon and portray its ubiquity among all learned generative models. We demonstrate that it must be taken seriously if we are to sustain the benefits of training from large-scale data scraped from the web. Indeed, the value of data collected about genuine human interactions with systems will be increasingly valuable in the presence of LLM-generated content in data crawled from the Internet.

Shumailov et al., Nature 2024 — model collapse

Personal experience keeps AI grounded in human reality. A stark contrast to learning from its own output.

Pattern 1 · context injection

1 · Capture / select

recent emails, notes, documents, calendar, location, task state

2 · Extend context automatically

add relevant personal information to the request / prompt

3 · Generate support

answer, draft, summarize, decide, explain, plan

example: personal-style draft from context

Quick & Ease; Good for lightweight personalization. Limited by context length and weak long-term memory.

Pattern 2 · personal RAG

Personal data store

documents · transcripts · images · notes · decisions · sensor logs

→

Retrieve evidence

vector + keyword search
filters · recency · metadata

→

Grounded response

answer with provenance
show what was used · allow correction

RAG is currently a strong default for personal memory services. It supports grounding and provenance.

Pattern 3 · fine-tuning / adaptation

WHAT TO ADAPT?

writing style ·
visual style
domain vocabulary
recurring preferences

EXAMPLE

fine-tune image generation with own photos, sketches, or design language

TRADE-OFF

More personalization possible. More privacy risks, more maintenance, and complex control. Fine-tuning for stable patterns; RAG for evidence and recency.

CONTROL

Human-centered control architecture for agents

Goal + constraints
what, why, boundaries

Permissions
may do · ask before · never do

Execution log
what happened, when, with which tools

Provenance
what data informed the action

Intervention UI
pause · inspect · redirect · undo

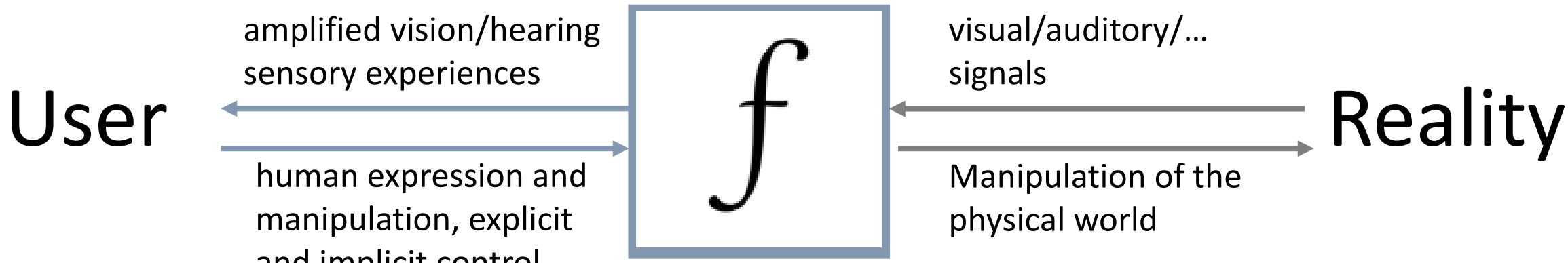
The future is not manual control over every step. This will not scale. Meaningful agency and delegated action

Research questions for service-oriented computing

- What is the service contract of a personal AI agent?
- Which personal context stays local — and what can move to the cloud?
- How should personal agents orchestrate tools and services across time?
- How do we expose uncertainty, provenance, and control without overload?
- How do we evaluate cognitive benefit and agency — not only task success?

SummerSoC can contribute: architecture, orchestration, privacy, data management, runtime support.

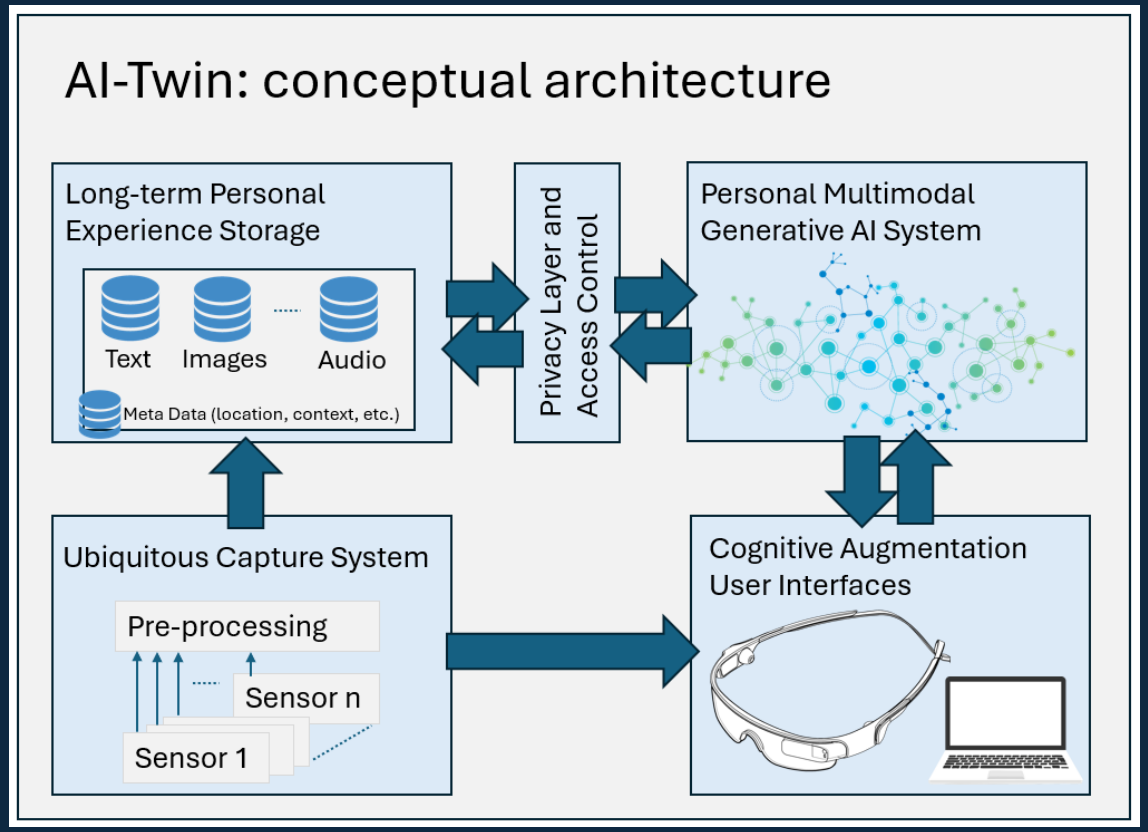
Our conceptual view on Human-AI interaction



TAKE-AWAY

The future service is human-native additionally to cloud-native and agentic.

- Personal cognitive tool — implemented as a service
- Background agent — acts within delegated boundaries
- Human-centered AI — amplifies cognition and preserves agency



Bush 1945 · Licklider 1960 · Engelbart 1962 · Weiser 1991 · Schmidt AVI 2020 · Schmidt AI-Twin 2025/26

Acknowledgements

- ERC AI-Twin
- ERC Amplify
- EU H2020 Humane AI Net
- BMBF IDeA, Hive
- MCML.AI

Reality Design: Shaping Experiences Beyond Interfaces through Human-Centered AI

Albrecht Schmidt, LMU Munich.

Abstract. As artificial intelligence advances, human-computer interaction is no longer just about designing interfaces - it is about designing reality itself. The interface is no longer a separate entity; it is interwoven with our interactions in the world. Our capabilities and limitations are increasingly dependent on digital tools and the way we interact with them. AI, Augmented Reality (AR), Virtual Reality (VR) and Large Language Models (LLMs) are fundamentally changing the way we perceive and manipulate our environment. This is shaping our reality! This talk asks: what meaningful contributions can humans make in a world where AI is exhibiting advanced creativity, solving complex problems, and making decisions based on more comprehensive information than any human can comprehend? Reflecting on automation and the over 70-year-old HABA-MABA principle, we explore what remains uniquely human today. Do we need to rethink our relationship with digital technology and our approach to designing large, interconnected, interactive systems? How can human-centered AI systems empower people and enhance human agency? We will discuss and share examples of how to move from user interface design to reality design to create experiences that are meaningful for individuals and responsible for society at large.

Short Bio. Albrecht Schmidt is Professor of Computer Science at the Ludwig-Maximilians-Universität (LMU) in Munich, where he holds the Chair for Human-Centered Ubiquitous Media. His research and teaching interests are human-centered artificial intelligence, intelligent interactive systems, ubiquitous computing, digital media technologies, and digital technologies for human augmentation. He studied computer science in Ulm and Manchester and received his PhD from Lancaster University in 2003. Albrecht was the conference co-chair of the ACM SIGCHI 2023 conference, he is on the editorial board of the ACM TOCHI journal, and he is the co-founder of the ACM conference TEI and Automotive User interfaces. He was inducted into the ACM SIGCHI Academy in 2018, elected to the German Academy of Sciences Leopoldina in 2020, and named an ACM Fellow in 2023.

https://uni.ubicomp.net/as/

https://www.linkedin.com/in/albrechtschmidt/

https://scholar.google.com/citations?user=qM5jR7YAAAAJ