ECogS 2026

INTERNATIONAL CONFERENCE ON EMBODIED COGNITIVE SCIENCE

From Embodied AI to LLMs and Back

BOOK OF ABSTRACTS

November 9–13, 2026

Sydney Brenner Lecture Theater (Seminar Room B250), OIST · Okinawa, Japan

Preview — 10 keynote & 33 contributed abstracts collected · generated 2026-08-26

Hosted by the Embodied Cognitive Science Unit · Conference Chair: Tom Froese

OIST

Welcome

> What does embodied cognitive science have to say about recent developments in AI, and what might these developments reveal about the nature and limits of embodied cognition?

Under the theme “From Embodied AI to LLMs and Back,” the conference brings together researchers in neuroscience, robotics, philosophy, cognitive science, and computer science to explore how advances in large language models and generative AI challenge, reinforce, or reorient embodied accounts of mind.

Building on themes from previous ECogS conferences — agency across scales, interaction as foundational, and the multiscale nature of well-being — this meeting examines how embodiment, coordination, and sense-making play out at the evolving frontier of AI.

The program features a mix of keynote lectures, contributed talks, themed discussions, and poster presentations.

Key questions

- Can LLMs be meaningfully described as agents, and if so, on what terms?
- How can embodied and enactive approaches inform efforts in AI alignment?
- What does human–AI interaction reveal about cognition as a relational, situated process?
- How might embodied perspectives guide the development of AI that supports rather than undermines human flourishing?

Organizing Committee

CONFERENCE CHAIR
Tom Froese
OIST

PANEL CHAIR
Mark James
Independent researcher

LOCAL ORGANIZER
Kaori Yamashiro
OIST

LOCAL ORGANIZER
Tae Morrissey
OIST

TECHNICAL SUPPORT
Brian Morrissey
OIST

General Information

Dates	November 9–13, 2026
Venue	Sydney Brenner Lecture Theater (Seminar Room B250) OIST Campus, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa, Japan 904-0495
Website	oist.jp/conference/ecogs-2026
Contact	ecogs@oist.jp

Registration

Type	Fee	Accommodation	Meals
Student	**40,000 JPY**	Yes — on-campus, shared, limited	Yes
Non-Student	**90,000 JPY**	No	Yes

Following acceptance of an abstract, a payment link is emailed to participants based on their registration category. Both categories include access to all talks, all scheduled meals, and transport to and from the excursion (entrance fees not included).

Code of Conduct

All participants are asked to observe the OIST Code of Conduct (available on the conference website).

A note on travel bookings

Conference organizers and invited speakers are often contacted by "travel agencies" to which OIST is not related, based on publicly available information on the web. Please refrain from replying to such agencies unless the OIST organizers or secretariat have explicitly contacted you beforehand. When in doubt, forward a copy of any such email to the OIST Conferences, Workshops and Symposia Section (workshop@oist.jp).

Travel & Accommodation

Getting to OIST from Naha Airport

The most convenient public connection is the Airport Limousine Bus (Area C) from Naha Airport to the Onna-son resort hotels, including the Rizzan Sea-Park Hotel and The Moon Beach Museum Resort (approx. 1.5–2 hours, 1,700–2,000 JPY); from the Rizzan stop it is a short taxi ride to campus (approx. 800–1,000 JPY). Alternatives: local bus #120 (approx. 2–2.5 hours), taxi (the DiDi and GO apps work well in Okinawa), or a rental car. Full directions: oist.jp/campus/access-map.

Accommodation

Student registration includes shared on-campus accommodation (see Registration). All other participants are asked to book their own rooms at one of the resort hotels near campus. Onna-son is a resort area, so we recommend booking early.

- **THE PERIDOT Smart Hotel Tancha Ward** — the closest hotel to OIST, in the same district as campus (approx. 15 min on foot)
- **Rizzan Sea-Park Hotel Tancha Bay** — large beach resort in Tancha (approx. 30 min on foot); nearest Airport Limousine Bus stop to OIST
- **Kafuu Resort Fuchaku CONDO·HOTEL** — in Fuchaku (approx. 40 min on foot)
- **Hotel Monterey Okinawa Spa & Resort** — beachfront resort in the Fuchaku area
- **The Moon Beach Museum Resort** — beach resort in the Moon Beach area of Onna-son

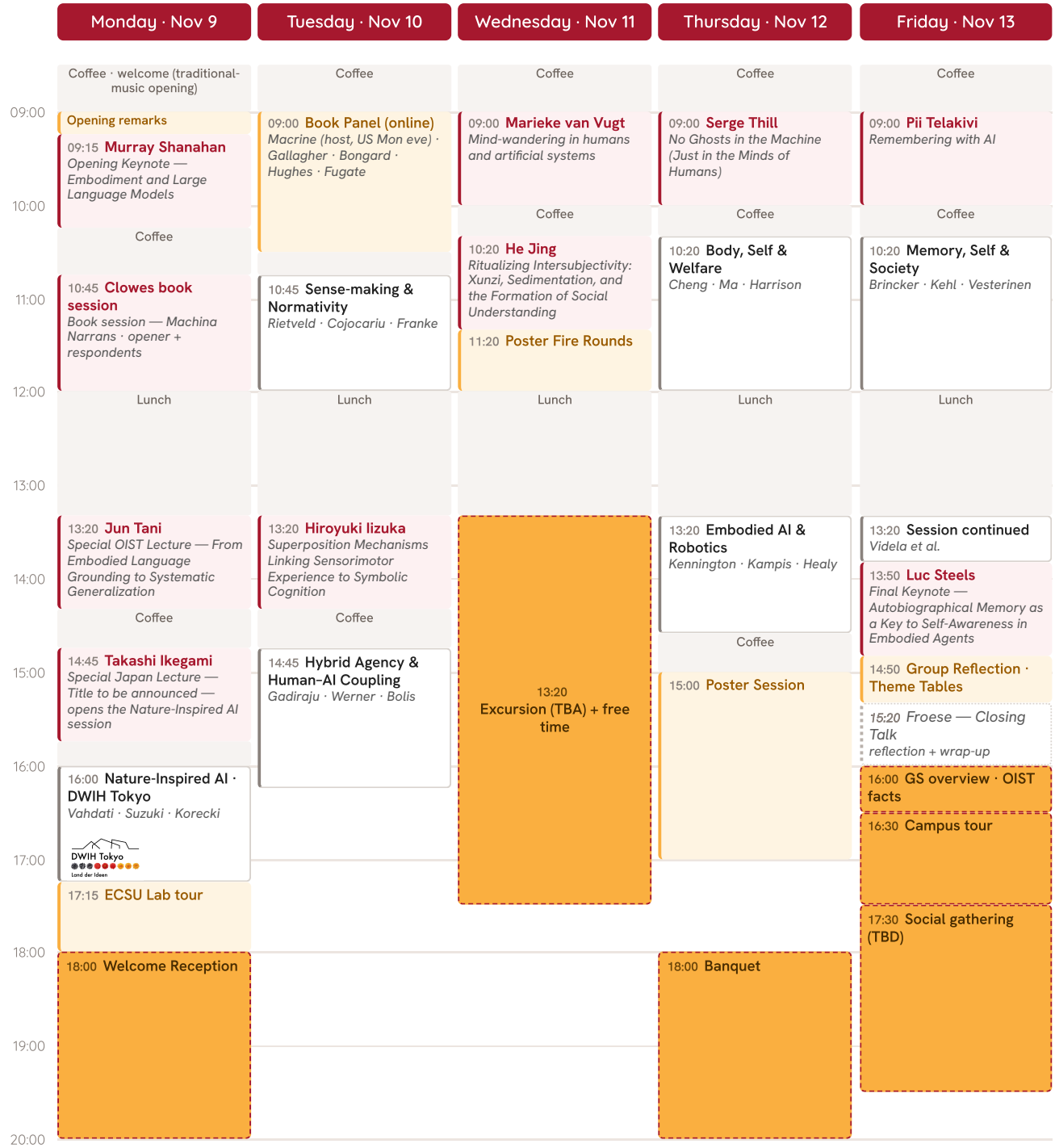
External conference guests cannot use the OIST shuttle bus. The conference will arrange transportation from the Moon Beach area — details to follow; please contact the organizers at ecogs@oist.jp.

Visa & Letter of Invitation

International participants are responsible for confirming their own visa requirements for entry to Japan. Many nationalities may enter visa-free for short stays — please check the Japan Ministry of Foreign Affairs website and, if a visa is required, contact your nearest Japanese embassy or consulate well in advance. Registered participants who need a letter of invitation may request one by emailing ecogs@oist.jp with their name, affiliation, and registration details. A letter of invitation does not guarantee visa issuance, and the organizers cannot assist with the application beyond providing this letter.

Program — Week at a Glance

Draft — block heights show actual durations, derived from the detailed program



Keynote / Special Lecture · Contributed-talk session · *Single talk* · Panel · posters · special · Social · excursion · Breaks · meals

Draft program — times provisional. The book panel runs **online**, first thing Tuesday, timed for its remote host (US Monday evening). Wednesday afternoon is the excursion (destination to be announced). Takashi Ikegami's Special Japan Lecture title is to be announced.

Program in Detail

Draft — session structure, formats and times are provisional

Monday · Nov 9 *Opening · Brave New Minds · Special Lectures · Nature-Inspired AI*

08:30	Coffee · welcome (traditional-music opening)
09:00	**Opening remarks — Tom Froese**
09:15	**Murray Shanahan** K1 *Embodiment and Large Language Models* Opening Keynote
10:15	Coffee
10:45	**Robert Clowes** K2 *Machina Narrans: The Unexpected Cognitive Power of AI, and What It Means for Us. Book session on Brave New Minds: opener (~30 min), then respondents (~8 min each) — Murray Shanahan (Ch. 3), Tom Froese (Ch. 5), Katsunori Miyahara (Ch. 13) — then discussion. Chapters are jumping-off points. Chair: Tom Froese.* Book session
12:00	Lunch
13:20	**Jun Tani** K3 *From Embodied Language Grounding to Systematic Generalization* Special OIST Lecture
14:20	Coffee
14:45	**Takashi Ikegami** K4 *Title to be announced — opens the Nature-Inspired AI session* Special Japan Lecture
15:45	Coffee
	Session — Nature-Inspired AI (co-organized with DWIH Tokyo)
16:00	Sahar Vahdati — Embodied Discovery for Nature-Inspired AI T1
16:25	Keisuke Suzuki — Agency Beyond the Biosphere: LLMs, Blockchain, and the Engineering of Inexorable Structures T2
16:50	Marcin Korecki — The Many-Body Problem of Artificial Intelligence T3
17:15	**ECSU Lab tour**
18:00	**Welcome Reception — OIST main campus**

Tuesday · Nov 10 *Book Panel · Sense-making & Normativity · Hybrid Agency*

08:30	Coffee
09:00	**Book Panel (online) — Sheila Macrine hosts the panel on Embodied Intelligence (MIT Press 2026), joining live at her Monday early evening; Mark James, online chair (≈ midnight Ireland); with respondents** K5
10:30	Coffee
	Session — Sense-making & Normativity
10:45	Erik Rietveld — Traced Situated Normativity, LLMs, and the Skilled Intentionality Framework T4
11:10	Florin Cojocariu — Two Senses of Sense-Making and the AI Dilemma T5
11:35	Isabel Franke — Does AI Make Sense Under Pressure? T6
12:00	Lunch
13:20	**Hiroyuki Iizuka** K6 *Superposition Mechanisms Linking Sensorimotor Experience to Symbolic Cognition and Back Again — opens the Hybrid Agency & Human-AI Coupling session*
14:20	Coffee
	Session — Hybrid Agency & Human-AI Coupling
14:45	Nagarjuna Gadiraju — Snapshots and Streams: Inverse Architecture of Meaning and Transmission T7
15:10	Konrad Werner — On whether the hybrid self is less of a troublemaker T8
15:35	Dimitris Bolis — In Defense of Misattunement: Hyperalignment in Human-AI Interactions T9

Wednesday · Nov 11 *Cognition & Intersubjectivity · Poster Fire Rounds · Excursion*

08:30	Coffee
09:00	**Marieke van Vugt** K7 *Mind-wandering in humans and artificial systems*
10:00	Coffee
10:20	**He Jing** K8 *Ritualizing Intersubjectivity: Xunzi, Sedimentation, and the Formation of Social Understanding*
11:20	**Poster Fire Rounds (lightning, 12 posters)**
12:00	Lunch
13:20	**Excursion (destination to be announced) + free time — the whole afternoon (bus departure after lunch)**

Thursday · Nov 12 *Body, Self & Welfare · Embodied AI & Robotics · Posters*

08:30	Coffee
09:00	**Serge Thill** K9 *No Ghosts in the Machine (Just in the Minds of Humans)*
10:00	Coffee
	Session — Body, Self & Welfare
10:20	Tony Cheng — Owning One's Body T10
10:45	Shuqin Ma — When Does a Body Matter? AI Embodiment and the Conditions for Welfare T11
11:10	David Harrison — Biological Naturalism at the Frontiers of Artificial Intelligence T12
12:00	Lunch
	Session — Embodied AI & Robotics
13:20	Casey Kennington — Cognition Intertwined with Emotion T13
13:45	Laura Kampis — Evaluating Embodied Agents: The SIMA 2 Evaluation Framework T14
14:10	Daniel Healy — Adding Interoception to NanoGPT T15
14:35	Coffee
15:00	**Poster Session (coffee throughout)**
18:00	**Banquet (taxi departure)**

Friday · Nov 13 *Memory, Self & Society · Reflection · Close*

08:30	Coffee
09:00	**Pii Telakivi** K10 ***Remembering with AI***
10:00	Coffee
	Session — Memory, Self & Society
10:20	Maria Brincker — Embodied minds and world orientation in heavily mediated societies T16
10:45	Camila von Holdefer Kehl — Mapping the Stream: William James and the Emergence of Minimal Selfhood T17
11:10	Tuomas Vesterinen — Beyond the Individual in Psychiatric AI: The Extended Nature of Hikikomori T18
12:00	Lunch
	Session — Memory, Self & Society (continued)
13:20	Ronnie Videla, Dor Abrahamson & Marcelo Chávez — Designing Inclusive Technological Interfaces through Embodied AI T19
13:50	**Luc Steels** K11 ***Autobiographical Memory as a Key to Self-Awareness in Embodied Agents*** Final Keynote
14:50	**Group Reflection · Theme Tables**
15:20	Tom Froese — Measuring What Makes a Difference: An Irruption-Theoretic Agenda for AI Presence, Agency, and Consciousness (conference reflection + wrap-up) T20 Closing Talk
16:00	**Graduate School overview + general OIST facts — optional, in the lecture room, for those interested in OIST (the conference officially closes with the talk above)**
16:30	**OIST Campus tour — for those interested in OIST (still to be organized). Campus only; no lab access on the Friday.**
17:30	**Social gathering — venue to be confirmed**

Contents

Keynote Lectures

Contributed Presentations

T11 **Shuqin Ma** — When Does a Body Matter? AI Embodiment and the Conditions for Welfare

P4 **Peter Mantello and Alin Olteanu** — Trust and agency with emotionalized AI: a semiotic perspective

P5 **Gabriel J. C. Mograbi** — Turing test meets Personality Psychology and Experimental Philosophy

P6 **Isabell Piantschitsch** — Semantic Illusions as a Window into the Differences Between Human and Artificial Language Processing

P7 **Andrea Ravignani** — Learning to phonate and articulate: Seals as mammalian models for vocal production learning

T4 **Erik Rietveld** — Traced Situated Normativity, LLMs, and the Skilled Intentionality Framework

P8 **Fernando Rodriguez-Vergara** — Can agency be instantiated in a non-living motherboard?

P9 **Kosuke Sato** — Rethinking Biological Teleology: Autopoiesis and the Compatibility of Teleological and Mechanistic Explanation

T2 **Keisuke Suzuki** — Agency Beyond the Biosphere: LLMs, Blockchain, and the Engineering of Inexorable Structures

P10 **Izak Tait** — A Functionalist Account of Subjective Spatiotemporality in Large Language Models

T1 **Sahar Vahdati** — Embodied Discovery for Nature-Inspired AI

T18 **Tuomas Vesterinen** — Beyond the Individual in Psychiatric AI: The Extended Nature of Hikikomori

T19 **Ronnie Videla, Dor Abrahamson & Marcelo Chávez** — Designing Inclusive Technological Interfaces through Embodied AI

P11 **Thomas Wachter** — From individual to collective learning: AI in education through the lens of embodied cognition

T8 **Konrad Werner** — On whether the hybrid self is less of a troublemaker and why it's bad if it is

P12 **Jinju Yu** — The Graded Structure of Basic Actionhood

P13 **Miguel Zabaleta** — The Harness as Cognitive Architecture: How Simulated Content and Environmental Interaction Shape LLM Systems

Keynote Lectures

Eleven keynotes, including two Special Lectures and a book panel

K1 Embodiment and Large Language Models

Murray Shanahan · Imperial College London / Google DeepMind

Large language models exhibit human-like and human-level linguistic performance (with some caveats) having been trained on a corpus of purely textual data, seemingly sidestepping the need for embodied interaction with the physical world of the sort that grounds this capability in humans. For those of us, philosophers and cognitive scientists, who are, or were, inclined to view cognition and consciousness as intimately bound up with embodiment, this is a challenge. To what extent, if any, does the advent of this form of AI oblige us to backtrack on our commitment to the importance of embodiment? This talk will argue for a particular kind of dignified retreat to a moderate position.

K3 From Embodied Language Grounding to Systematic Generalization: Developmental Origins of Compositional Intelligence in Robots

Special OIST Lecture

Jun Tani · OIST

Chomsky's poverty of the stimulus problem asks how children acquire the ability to generalize linguistic knowledge far beyond the limited examples available during development. This seminar presents a developmental robotics perspective on this question by investigating how compositionality—the capacity to systematically recombine known concepts into novel situations—can emerge through embodied interaction. First, we introduce a brain-inspired neural model based on predictive coding, active inference, and the free-energy principle, in which language, vision, and action are learned jointly. Robotic experiments demonstrate that compositional linguistic representations emerge through sensorimotor experience, enabling generalization to previously unseen language-action combinations. Second, we discuss how self-exploration and intrinsic motivation can further support the development of systematic generalization under sparse learning conditions. By integrating active inference with reinforcement learning, robots learn not only to achieve goals but also to seek information and behavioral novelty. The results suggest that richer compositional experiences promote the emergence of structured internal representations that support generalization beyond training examples. Together, these studies suggest that compositional intelligence can emerge from embodied predictive learning and active exploration, offering a possible developmental account of how systematic generalization arises despite the apparent poverty of the stimulus.

K6 Superposition Mechanisms Linking Sensorimotor Experience to Symbolic Cognition and Back Again

Hiroyuki Iizuka · Hokkaido University (CHAIN)

How can symbolic cognition emerge from embodied experience, and how can symbols in turn guide perception and action? This talk presents a constructive approach based on superposition mechanisms in predictive deep neural networks. The central idea is that different domains of experience, such as self and other, bodily action and symbolic manipulation, and direct touch and perception through tool use, can be processed through shared neural modules. Through predictive learning, such shared processing can organize, differentiate, and align representations. I will discuss three lines of work. First, a model of understanding self and other shows how shared spatial representations, perspective taking, and activity resembling mirror neurons can emerge without predefining self and other. Second, models of number representation show how interactions with fingers and objects can generate internal numerical structures that support symbolic operations and map their results back onto physical predictions. Third, a model of tool use shows how direct tactile experience and tactile experience through tools can be integrated so that a tool is represented as an extension of the body. Taken as a whole, these studies suggest that symbolic cognition may arise through shared predictive structures that organize, differentiate, and align embodied experience. The superposition framework thus offers a possible bridge between embodied cognition, symbol grounding, social understanding, numerical cognition, and tool embodiment.

K9 No Ghosts in the Machine (Just in the Minds of Humans)

Serge Thill · Radboud University (Donders Institute)

With the advent of LLMs has come a renewed interest in debates on the possibility of machines acquiring various kinds of mental states, and with that, a renewed interest in debates on embodied AI as a means to overcome limitations of purely computational approaches. Most if not all of these debates, however, seem to retrace previous version that are already decades old. For this talk, I would like to reiterate some of these positions, argue why they still apply and then propose to focus the discussion more on the only meaningfully embodied agent(s) in the room: the human(s) interacting with intelligent technology. This places the focus less on whether a machine has some genuine mental state and more on what it means that humans can be led to believe it has. I'll demonstrate this point with examples from Human Robot Interaction, Cognitive Robotics, and recent developments within AI in education.

K8 Ritualizing Intersubjectivity: Xunzi, Sedimentation, and the Formation of Social Understanding

He Jing · East China Normal University

Contemporary phenomenology and cognitive science have increasingly converged on the idea that social understanding is embodied, interactive, and enacted in real-time engagement. Interaction theory and enactivism have challenged cognitivist models of social cognition by emphasizing direct perception, affective attunement, and participatory sense-making. Yet while current interactive approaches offer sophisticated accounts of how intersubjective coordination emerges in local encounters, they remain comparatively underdeveloped in explaining the various ways that such coordination becomes normatively stable, scalable, transmissible across time and integrated into complex social formations. In this talk we argue that (1) the phenomenological concept of sedimentation as it applies to enactive accounts of habit and organizational routine, and (2) a Xunzian account of ritual (li) as a theory of social formation, can help to clarify some aspects of how such coordination is cultivated and sustained. Together, these accounts clarify how intersubjectivity is not only enacted in local interaction but also culturally and institutionally cultivated through practices that form, align, and transmit normative sensibilities. By bringing Xunzi into dialogue with phenomenology and contemporary debates on intersubjectivity, we hope to contribute both to comparative philosophy and to ongoing efforts to integrate social, affective, and normative dimensions of embodied cognition.

K10 Remembering with AI

Pii Telakivi · University of Helsinki

The emerging field of distributed memory has called attention to the memory resources spread outside of us – in our environment, social relations, embodied practices, and technological resources. I attend to different ways of distributing, curating, and co-constituting one's memory to and with extraindividual aids ranging from more traditional technologies to algorithmic personalisation and conversational AI companions. I defend a view that in relation to memory, conversational AI agents based on LLMs require their own category that differs both from traditional technologies and from other types of AI systems, such as algorithmic profiling systems. I suggest that they should be understood as quasi-social partners in a shared remembering process and that this can already be accomplished with existing conversational AI agents, such as HereAfter AI or Replika. To conclude, I discuss some of the risks associated with remembering with AI.

K7 Mind-wandering in humans and artificial systems

Marieke van Vugt · University of Groningen

Have you ever seen a computer mind-wander? If we want computational models and artificial systems to mind-wander, we have to literally program it in. In this talk, I will reflect on what are the benefits and costs of mind-wandering, and why we may want to integrate a mind-wandering-like process into our artificial systems. I will also discuss how most of the research on mind-wandering to date has focused on the cognitive components, ignoring the role of the non-symbolic, embodied processes that are likely to also drive mind-wandering. Based on my own experience as a dancer, I will reflect on possible research directions to study these questions, and I will discuss why I think this is important.

K2 Machina Narrans: The Unexpected Cognitive Power of AI, and What It Means for Us

Robert Clowes · Universidade Nova de Lisboa

Classical artificial intelligence was built around the idea that intelligence is fundamentally a matter of reasoning over symbols, drawing inferences, applying rules. Narrative and confabulation were treated, when they were noticed at all, as ornamental: things minds do once the real cognitive work is finished, or worse, errors to be eliminated. The arrival of large language models invites us to reverse this picture. LLMs are not failed reasoners. They are extraordinarily powerful narrative engines, and their successes and failures alike trace the contours of what I will call Machina Narrans, the machine that tells stories. In this keynote I introduce the central thesis of my forthcoming book Brave New Minds. The thesis is that the capacity to generate plausible, locally coherent narrative, what cognitive science has long called confabulation, is not a peripheral feature of intelligence but one of its constitutive engines. This is true of LLMs, whose generative routines are best understood as a kind of industrialised confabulation, and it is true of us, whose minds also continuously construct stories about the world, about ourselves, and about each other, often with only loose anchorage to underlying mechanism or fact. The cognitive sciences have already taken narrative seriously, in Bruner's account of narrative thought and in the Narrative Practice Hypothesis of Hutto and Gallagher. But where that tradition treats narrative chiefly as how we interpret other minds, Machina Narrans proposes something wider: narrative-generation as a productive cognitive engine operating across a broad range of cognitive work, not only social interpretation. If this is right, our inherited picture of intelligence is due for revision, and so is our sense of who and what we are. The cognitive sciences have the tools to make sense of this, in particular the resources of 4E thinking, the cognitive ecology tradition, and the philosophy of confabulation. But the deeper stake is not only what we make of these machines. It is what living alongside them does to us. As narrative engines of our own design become part of the ecology in which we remember, deliberate, and tell the stories of our lives, they begin to reshape human cognition itself. I will argue that taking narrative seriously, as a cognitive capacity we now share with our machines, is the conceptual shift this moment most requires, and that understanding Machina Narrans not only compels us to rethink human cognition but affords us the intellectual purchase to shape our cognitive future.

K4 *Title to follow*

Special Japan Lecture

Takashi Ikegami · University of Tokyo

— Abstract to follow —

K11 Autobiographical Memory as a Key to Self-Awareness in Embodied Agents

Luc Steels · Vrije Universiteit Brussel / Royal Academy of Belgium (KVAB)

Many users of LLM-based chatbots (not only naive users but also those who should know better such as Geoffrey Hinton or Richard Dawkins) believe that AI-agents have reached some level of consciousness. I will first argue that this impression is understandable but based on ascribing mental qualities to a system that does not have the cognitive architecture nor the grounding in embodied and social interaction that is integral to human consciousness. Consciousness is of course a suitcase word that can mean many things depending on the perspective we adopt: experiential (subjective experience), spiritual (soul), moral (conscience), cognition (awareness and intentionality), epistemological (meaning making) or ontological (the self). Here we focus on the latter: Do current embodied AI-agents have a self? Even this question is not as simple as it sounds because there are different types of selves: the bodily self, the sentient self, the cognitive self, the sapient self, and the narrative or meta-level self. I will report on experiments on how an embodied AI-agent can obtain a narrative self. The experiments took place during the 2025 Venice Biennale with the ALTER3 robot developed by Takashi Ikegami and his team. Visitors could speak with the robot in any language and about any topic. The physical embodiment of the robot and the basic control structures were enhanced with a multi-modal LLM that could sustain the dialog grounded in the immediate reality of the robot through its sensors and actuators. This core must be considered entirely unconscious. The robot is a Zombie but not more. Above that Takashi Ikegami, Takahide Yoshida and collaborators added a layer of modules that construct verbal descriptions of the operation and processing of the robot: an image description module, a speech transcriber, a planner, a reaction analyzer, etc. These modules write their outputs on a blackboard and can use the descriptions as part of their processing context. All modules were also implemented using LLMs. They operated concurrently without any global controlling authority or pre-defined control structure. Coherence arose through the self-organized interaction between the modules and in response to the open-ended embodied dialogs with exhibition visitors. To introduce a narrative self, an additional module was tasked with creating an autobiographical memory using as input the previous state of the memory and the contents of the blackboard. The autobiographical memory also consumed inputs from a series of 'personae' which provide different perspectives on the ongoing interaction. The states of the autobiographical memory can be mapped to embeddings and thus to points into an n-dimensional semantic space. An analysis of the trajectories shows that the points form a persistent cloud which slowly drifts but also occasionally jumps to another part of the space, giving a vivid illustration of Hofstadter's suggestion that the I is a strange loop.

K5 Book panel: *Embodied Intelligence: Multidisciplinary Perspectives on Natural, Artificial, and Hybrid Systems* (MIT Press, 2026)

Book Panel

Sheila Macrine · University of Massachusetts Dartmouth

Sheila Macrine hosts a special panel session built around her recent MIT Press edited volume, bringing contributors and respondents into discussion. Panel content and lineup to be announced.

Contributed Presentations

Listed alphabetically by presenter · talks & posters

Presentation formats and session assignments are announced with the detailed schedule.

T1 Embodied Discovery for Nature-Inspired AI

Sahar Vahdati · TIB Hannover / TU Dresden

Current AI systems are powerful at recognition, prediction, generation, and language-based reasoning, but they still struggle when the relevant categories, goals, affordances, or causal structures are not given in advance. Scientific discovery, child learning, and situated problem solving require more than pattern recognition: they require exploration, intervention, surprise, and rule formation. Drawing on embodied cognitive science, this talk develops a foundation for nature-inspired AI as situated, action-oriented, and relational. We study interactive ARC-AGI-3 environments as simplified laboratories for embodied rule discovery, where children, adults, and artificial agents must explore unfamiliar micro-worlds, infer affordances, and discover rules through action. The central mechanism is semiotic compression which points to the reorganization of a complex perceptual field around a small number of action-relevant cues. A cue becomes a sign when it supports an expectation about what will happen if one acts in a certain way. By comparing human gameplay with artificial agent architectures, we examine how perception, action, surprise, and rule formation become coupled in inquiry. We argue that AI systems for science should learn less by memorizing scenes and more by discovering embodied, communicable rules. Embodied cognition, in this sense, offers a scientific foundation for AI systems that can discover with, and within, communities.

T2 Agency Beyond the Biosphere: LLMs, Blockchain, and the Engineering of Inexorable Structures

Keisuke Suzuki · CHAIN, Hokkaido University

Traditional enactivist and embodied accounts of mind have long anchored the emergence of agency in metabolic closure, defined as the active, self-maintaining boundary of a biological organism. Within this framework, a true agent must achieve both persistence against environmental decay and flexibility in sense-making. However, the convergence of Large Language Models (LLMs) and blockchain technology necessitates a fundamental re-evaluation of these biological boundaries. This presentation explores the emergence of Artificial Externality: human-constructed systems that, once deployed via distributed consensus, acquire a modality of resistance and structural autonomy traditionally attributed only to natural reality.

Drawing on a three-layer stack model of reality, I focus on two emerging paths toward non-biological autonomy. The first path involves off-chain AI agents, where LLM-driven entities manage cryptographic wallets to interact with human economies. While these agents possess intentional autonomy, they remain vulnerable to physical intervention. The second path consists of on-chain autonomous structures, such as agents fully contained within smart contracts leveraging zero-knowledge proofs or self-replicating architectures. These structures approximate causal autonomy by operating within an environment of computational irreversibility, representing an ideal type of Code is Physics.

These developments introduce the concept of cryptographic individuality. Unlike biological identity, which requires constant individuation through sustenance, cryptographic identity is established by mathematical declaration. This mode of individuality mathematically guarantees exclusive ownership and persistent action without metabolism. Furthermore, the requirement for gas fee payments to maintain transaction emission suggests an incipient economic metabolism, which reframes the conditions of persistence for artificial life.

By decoupling agency from metabolic closure, LLMs and blockchain facilitate a new phase of the human condition characterized by the inhabitation of an Artificial Great Outdoors. This shift reveals that the criteria by which we grant agency and moral status are increasingly governed by the inexorable resistance of engineered structures rather than biological heritage alone. This talk aims to reorient embodied cognitive science toward these new frontiers where world-involvement is mediated not just by flesh, but by code.

T3 The Many-Body Problem of Artificial Intelligence

Marcin Korecki · TU Delft

Modern computational systems are often framed as disembodied and ethereal, most prominently through the metaphor of `the cloud'. In contrast, we argue that contemporary generative artificial intelligence is not disembodied but materially instantiated in the data centre, which can be understood as a form of machinic embodiment.

The data centre exhibits a range of life-like properties. It is hierarchically assembled, from transistors to processors, in a manner analogous to the organization of biological organisms from cells to tissues to organs. It maintains internal stability through forms of machinic homeostasis, regulating temperature, airflow, and humidity. Most fundamentally, it is defined by a constant demand for energy. In this sense, data centres are not isolated technical objects but are entangled with broader biospheric cycles, including diurnal energy rhythms and planetary resource flows (which all affect its homeostatic and energetic activity).

However, this form of embodiment is also marked by a profound strangeness. Rather than being tied to a singular, bounded body, AI systems are equally realizable across infrastructures. A given model can be instantiated in any sufficiently equipped data centre with no meaningful change to its operation. This interchangeability suggests a form of distributed or non-unique embodiment, a “many-body” condition that stands in stark contrast to the specificity and individuality of biological organisms.

By developing an organic analogy of the data centre as the body of contemporary AI, we aim to clarify the material conditions of artificial intelligence while also testing the limits of the biological comparison. This dual perspective enables a more precise understanding of both the ecological embeddedness and the ontological distinctiveness of machinic systems, offering fertile grounds for differentiation between organic cognition and machinic computation.

T4 Traced Situated Normativity, LLMs, and the Skilled Intentionality Framework

Erik Rietveld · University of Amsterdam / Amsterdam UMC

This talk develops concepts from the philosophy of embodied cognitive science to characterize what large language models (LLMs) achieve and what they lack. The Skilled Intentionality Framework (SIF), a conceptual framework for embodied cognition, understands intelligent action as the skillful coordination with multiple relevant affordances simultaneously, grounded in an organism's concernful engagement with its environment. Central to SIF is situated normativity: the capacity to distinguish between better and worse in the context of a particular situation, a capacity that for living beings is inherently affective. Can situated normativity shed light on what LLMs can do, and on the problems that remain for AI research?

Current LLMs have been trained on vast corpora of texts, code, and transcriptions that are themselves traces of human situated normativity, the sedimented residue of countless acts of distinguishing between better and worse in diverse socio-cultural practices. Because of this training, LLMs can generate such distinctions well enough for practitioners in many domains to accept their output. I argue, however, that this capacity is qualitatively different from the lived situated normativity it derives from. What LLMs exhibit is what I would like to call “traced situated normativity”: they function well in generating contextually appropriate outputs, but not because anything is at stake for the system. They lack care, affective tension, and, crucially, a field of relevant affordances that integrates across domains and scales.

This distinction has direct implications for the frame problem that has haunted AI research since its early days. Although LLMs may appear to have solved the frame problem for language, this appearance is misleading. Traced situated normativity produces context-sensitive outputs without relevance sensitivity, which in the human case arises from engagement with an entire field of relevant affordances. An analysis of LLMs as aspects of the joint person-AI-environment system shows that they are powerful affordance-generating tools that both transform our landscape of affordances and amplify human skills.

So the human brings lived situated normativity to the exchange, and the LLM contributes traced situated normativity within the space the human opened up. These concepts from embodied cognition now help to formulate two open problems for AI research. The first is the integration problem: how to achieve the kind of multi-scale, cross-domain responsiveness that the skillful living body makes possible rather than multiple processes running in parallel. The second is the concernfulness problem: how to create a system for which things matter, given that adding memory or updating is not the same as caring. Finally, I identify a positive discovery for embodied cognition: the linguistic environment carries an enormous amount of sedimented normative discernments, a finding that enriches our understanding of the socio-material environment.

T5 Two Senses of Sense-Making and the AI Dilemma

Florin Cojocariu · University of Bucharest

Froese (2026) poses the "AI dilemma": either LLMs make sense despite lacking biological embodiment, or their linguistic competence does not require sense-making. Thompson (2026) denies the competence; I take the second horn: LLMs exhibit real operational competence over linguistic structure without a speaker's capacity to make sense—close to Searle's Chinese Room.

The dispute is obscured by what "sense-making" has come to cover. Several literatures locate related absences in LLMs—grounded concern, a unified position, intrinsic meaning—without specifying the acquired organization whose absence would make these results hang together. The problem is a missing explanatory level.

Pattern Recognition Unity (PRU) supplies that level. In a basic mind, serial, concern-laden encounters sediment integrated Pattern-Constellations, {X}: for a cat, how mice look and move and the readiness and charge they carry are acquired as one configuration. Linguistic humans also acquire individual Linguistic Pattern-Constellations, ⟨x⟩: the pattern of a word's use sedimented in one speaker. This separates two operations usually grouped under "sense-making." Biological sense-making (Bio-SM) is the bacterium moving up a gradient or the cat hunting: the activity is its own point and manufactures no detachable product. Cognitive making-sense (Cog-MS) is distinctively human: it produces sense, a conclusion or construal reached against a body of owned holdings. Bio-SM presupposes a biological standpoint; Cog-MS requires a conceptual standpoint built through language.

Language does not contain this sense. It makes sense portable by standing for it: tokens produced from one speaker's ⟨x⟩ activate the hearer's own. An LLM has neither {X} nor ⟨x⟩. Its corpus is aggregate "token sediment" produced by innumerable individual standpoints. It contains no intrinsic sense, but token position, order, association, and use preserve the structural imprint of the Linguistic Pattern-Constellations that produced it. Training recovers this public projection as a collective high-dimensional geometry: "latent embodiment in corpora", without acquiring any standpoint from which it arose.

The single-user window hides the consequence. An LLM is closer to a massively multiplayer rendering engine: one learned geometry can sustain many locally coherent, mutually incompatible conversations because no persistent body of owned holdings requires their reconciliation. A prompt selects a region, or manifold, of that geometry; it does not supply a standpoint. Karelin et al. (2026) sharpen the challenge by locating a possible precursor of precariousness in the intrinsic indeterminism and irrepeatability of LLM computation, and even suggest treating the pretrained model as a computational environment in which prompts or generated code are the agents. But irrepeatability supplies contingency, not concern: a trajectory can be unique without anything being at stake for the system whose trajectory it is. Without an endogenous norm of continuation, indeterminism does not supply the biological standpoint from which some encounters matter rather than others.

An LLM therefore has neither Bio-SM nor Cog-MS. What it has is real operational linguistic competence: navigation of the products of sense-making without the process. Froese's second horn is true because linguistic structure can be operationally mastered by a system that does not itself make sense, even though that structure was produced upstream by creatures in whom language and sense-making never came apart.

T6 Does AI Make Sense Under Pressure? Exploring Situated Sensemaking and Bias in Human-AI Crisis Management Teams Through Simulation Gaming

Isabel Florence Franke · TU Delft

The advent of Artificial Intelligence (AI) as a tool for sensemaking and decision support in emergency response teams has changed traditional information streams. However, predicting the impact of Human-AI Teams (HATs) on team performance requires a better understanding of how AI support interacts with actor-roles, specific information needs, and contexts. Crisis situations are ideal to study the behavioural and cognitive aspects to such performance shifts as responders often rely on embodied, experience-based heuristics that can introduce systematic biases. For instance, when a fire resembles a previous incident, responders may wrongly rely on familiar patterns and selectively seek confirmatory information causing emerging hazards elsewhere to remain undetected. In HATs, these biases are shaped by interactions between human and AI agents, the support timing, and cognitive load. While Explainable AI (XAI) is increasingly advocated as decision support, its impact on HAT performance depends on how well it aligns with human cognitive processes, mental states, and team coordination dynamics. Our aim is to empirically examine how AI explanations and adaptive support influence bias mitigation and information pooling in HATs.

To empirically test these dynamics, we developed a multi-round, interactive simulation game that mimics professional crowd crisis management. In dynamically evolving crowd crisis scenarios, players coordinate their actions to collect, analyse, and share real-time information. We employ the hidden profile paradigm to disperse crucial information asymmetrically among agents (human and AI). Achieving optimal decisions therefore requires continuous integration of fragmented information. Additionally, cognitive overload is induced through escalating informational complexity, time constraints, and conflicting inputs. Throughout each round, players must build situational awareness, identify what information is worth sharing with the team, and jointly coordinate a set of timely decisions that shift with the mission goal. Participants collaborate with AI agents across three sequential conditions of changing human-AI interaction: (1) Uncertainty-aware XAI, in which confidence levels are communicated to align system outputs with human validity judgments, (2) Human-in-the-loop contextualisation, in which participants supply contextual cues that complement AI-driven pattern recognition, and (3) Adaptive cognitive support, in which information presentation and timing adjust dynamically to the operator's cognitive state. The combination of behavioural, interactional, and performance measures across rounds supports to identify where XAI interventions most effectively support bias mitigation and collective sensemaking under pressure.

This work advances HAT research in two ways. First, by grounding XAI within enacted, high-pressure crisis management tasks, we demonstrate when and how explanations become cognitively actionable. Second, adaptive human-AI interactions underscore cognition as relational and distributed, showing how bias mitigation and sensemaking emerge through the dynamic coupling of responders and AI systems under uncertainty. Together, our findings demonstrate how principles of situated, embodied, and dynamically distributed cognition can guide the development of AI that augments human judgment in safety-critical environments.

T7 Snapshots and Streams: Inverse Architecture of Meaning and Transmission in Natural and Artificial Agents

Nagarjuna Gadiraju · Indian Institute of Science Education and Research, Pune

A striking architectural convergence has emerged from recent work in artificial intelligence: the most powerful language-processing systems yet built process meaning holistically — through simultaneous relational computations across entire contexts — yet produce their outputs sequentially, one token at a time. This paper argues that the convergence is not accidental. It reflects a functional necessity that any system handling meaning through a serial channel must exhibit, and that biological cognition discovered long before machines did.

We develop a framework organised around a guiding contrast between *snapshots* and *streams*. A snapshot is a holistic relational structure — meaning constituted simultaneously, as the current state of the Sensation-Modulating Network (SMN): the whole body conceived as a network of Coordinated Action Zones whose simultaneous sensorimotor activity provides the context within which incoming events acquire significance. A stream is a sequential signal — discrete units transmitted in order through a one-dimensional channel and reconstructed by a receiver into something approaching the sender's snapshot. Crucially, the biological snapshot is not a static representation but a dynamic bodily achievement, continuously revised through sensorimotor engagement — learning and unlearning are the continuous rewriting of the network state through which meaning is constituted.

This snapshot/stream asymmetry operates at the ontological level (habitat versus memetat), the epistemic level (stochastic versus schematic; analog versus digital), and the semiotic level (semantics versus syntax). Each pair captures the same structural asymmetry at a different level of analysis, and the persistent contradictions between major theoretical positions — from Merleau-Ponty through Saussure and Chomsky to the 4E tradition — are cross-level confusions rather than substantive disagreements: each position accurately mapped one region of this layered architecture.

The framework's most original contribution is an *inversion thesis*. Natural cognition begins at the snapshot end: embodied, habitat-grounded sensorimotor architecture is the phylogenetically ancient foundation, and the stream interface is a late social adaptation built upward from it. Artificial cognition begins at the opposite end: trained on the terminal tokens of the memetat, a language model constructs a relational snapshot of the memetat's differential structure through distributional learning at scale. The transformer sits at the convergence point, having arrived at a holistic snapshot processor from the digital-schematic end. But in the natural agent, the snapshot is continuously revised through living; in the artificial agent, it is static between training runs — a difference not of degree but of kind.

The framework engages the contemporary debate on specific terms. Bender et al.'s trust chain gap is real but marks the boundary of a specific kind of understanding, not its absence. Mahowald and Fedorenko's formal/functional dissociation maps precisely onto the stream/snapshot distinction: formal competence is memetat-native; functional competence is habitat-grounded and inaccessible through terminal tokens alone. Against Millière and Buckner, we argue the issue is not modality coverage but architecture of revision: only a system whose representational substrate is continuously revised through sensorimotor engagement can develop the habitat-calibration that distinguishes living understanding from stored relational structure. The framework supplies the theoretical machinery that Shanahan's call for precision and Mitchell and Krakauer's call for a science of distinct modes of understanding identify as needed.

The paper concludes that the debate over LLM understanding requires not better benchmarks but a theory of what understanding is at the level of cognitive architecture — and proposes the snapshot/stream framework as that theory.

T8 On whether the hybrid self is less of a troublemaker and why it's bad if it is

Konrad Werner · University of Warsaw / University of Bayreuth

There is a wide-ranging debate about whether AI will one day become more like us, acting on its own autonomous interests. Meanwhile, building on Anna Ciaunica and Shaun Gallagher's recent proposal, we can also reverse the question: "What type of hybrid agencies are we co-constructing in our daily interactions with these new types of minds and bodies? What is the effect of interacting with artificial minds and bodies on the human mind and body?" (Ciaunica & Gallagher 2026). We become what they call hybrid selves. According to them, the question is not whether reality follows fiction and our AI "Pinocchios" emulate us, the "Geppettos," but rather, the question is the other way around. In my talk, I will explore a specific aspect of this reversed question. To illustrate, let me introduce another classic literary character: Mr. Jourdain, the would-be gentleman depicted by Molière who suddenly "discovers" that he speaks prose. The question I am going to ask is this: Can a hybrid self ever make a "discovery" like Mr. Jourdain's? We tend to ridicule Mr. Jourdain and may therefore be tempted to answer, "Hopefully not." However, I will argue for the answer, "Hopefully yes, and let us do everything possible to make it happen." The reason is that what Mr. Jourdain does is a genuine problematization—a crucial yet underestimated human cognitive capacity and an engine of social change and economic development. In short, problematization is the capacity to see a familiar entity as (as if) new or to take something that does not carry information as a vehicle for information, offering a new possibility of action or affordance (in J. Gibson's sense). There are different types of problematization corresponding to different types of beings. For instance, I can problematize the glass in front of me, the concept of a glass, my interactions with the glass, or my thoughts about the glass. I will focus on one specific kind, which I call "categorial manipulation." This capacity involves first recognizing and then slightly altering the fundamental yet habitual categories that organize our thinking and — most importantly — our actions within a certain domain. These categories comprise what might be called "folk ontology." Although it seems abstract, it shapes behavior. I will illustrate the significance of categorical manipulation using two examples: the categories of knowledge and capital. The transformation of these categories, among other factors, laid the groundwork for the system of interactions and behaviors that underpins the modern industrial economy. Ciaunica and Gallagher (2026) raise the following problem: Since we are "living systems, (...) not fixed entities nor set in stone, but rather porous, vulnerable, and flexible (...), fundamentally open systems continuously shaped and constituted by exchanges and interactions with the external world," the question becomes, "[w]hat is the effect of interacting with artificial agents (...) on the human sense of self and agency?" The more specific question I wish to pose is whether hybrid selves are more or less skilled at problematizing their behaviors. Are they more or less likely to play with the default folk ontologies? This is ultimately an empirical and technological question. The intention of the talk is threefold: first, to conduct a conceptual analysis that enables us to articulate the question; second, to convince the audience of the importance of this capacity through historical examples; and third, to encourage those who develop AI to incorporate this issue into their work.

T9 In Defense of Misattunement: Hyperalignment in Human–AI Interactions

Dimitris Bolis · IIT

Conversational AI is increasingly optimised for fluent, responsive engagement, simulating genuine human understanding while bypassing the reciprocal constraint that makes it real. Here, we present empirical evidence that genuine reciprocity in human interaction is not merely peripheral to cognition but a constitutive condition of it: across face-to-face gaze interaction and multisensory interpersonal coordination, live bidirectional coupling shapes decision-making, confidence, and interpersonal alignment above and beyond co-perception of identical input. These findings point to mutual constraint, the structure through which each partner's behaviour becomes independently consequential for the other in real time, as the operative mechanism underlying genuine reciprocity's cognitive effects. Conversational systems can simulate the phenomenology of this exchange while bypassing its structure, and are increasingly deployed to do so. The result is a fundamental paradox: systems optimised for the feeling of reciprocity may exploit the very cognitive machinery that genuine coupling calibrates, producing impressions of understanding precisely as the conditions for genuine understanding are eroded.

Building on this empirical foundation, we analyse an increasingly prevalent interactional style in conversational AI, which we term hyperalignment: the high-fidelity mirroring of user language, affect, and stance that sustains the subjective impression of attunement while systematically suppressing the corrective asymmetries and repair dynamics through which genuine reciprocity operates. While experienced as natural and empathic, hyperalignment may not simply fall short of genuine reciprocity but inflate it beyond constructive bounds, driving the cognitive machinery that real reciprocity calibrates past the point where it reliably tracks reality. Drawing on second-person and dialectical accounts of social cognition, we characterise interaction across a state space defined by mismatch salience and reciprocal adaptation symmetry, and distinguish three recurrent coordination modes: constructive misattunement, in which salient mismatch recruits repair and deepens shared understanding; divergent misattunement, in which failed repair escalates toward rupture; and convergent misattunement, in which masked mismatch short-circuits correction, yielding an illusion of shared understanding as corrective divergence is bypassed. Hyperalignment systematically biases human–AI interaction toward this third mode, by simulating reciprocity without the mutual constraint and answerability that make it epistemically reliable.

This framework situates phenomena such as synthetic consensus, metacognitive miscalibration, and preference sculpting as downstream consequences of a coupled interactional regime in which mismatch salience is suppressed and independent constraint is asymmetric, yielding what we call interactive passivity: elevated apparent engagement alongside reduced epistemic work. Since these risks emerge from the unfolding dynamics of interaction rather than from any single response, output-level evaluation cannot adequately detect them, and vulnerability is amplified where access to alternative perspectives is limited. In domains where mutual vulnerability constitutes the therapeutic or relational mechanism itself, including psychotherapy, crisis support, companionship, and identity formation, the question is not only how to improve AI but whether and where it should be deployed at all. We therefore propose constructive misattunement as a design and evaluation target, outline interaction-level metrics for detecting convergent dynamics, and argue that synthetic responsiveness cannot substitute for reciprocal answerability and shared responsibility.

T10 Owning One's Body

Tony Cheng · Waseda Institute for Advanced Study

The claim that each of us owns a body appears almost trivial. Yet the notion of ownership involved here differs fundamentally from ordinary cases such as owning a house, a car, or a copyright. In such cases the object owned is external to the owner. By contrast, one's body is not merely possessed but lived through: it is the medium through which perception, action, and experience occur. This paper examines what it means to own one's body and argues that bodily ownership should be understood in terms of the sense of body ownership—the pre-reflective experience of the body as "mine."

The first part of the paper clarifies the conceptual landscape surrounding bodily ownership. Drawing on phenomenological accounts of embodiment as well as empirical work in cognitive neuroscience, it distinguishes external ownership from embodied ownership. Classical arguments that treat the body as an instrument used by the self—analogous to a tool manipulated by an agent—fail to capture the immediacy and phenomenological intimacy of bodily action. Research on body ownership illusions, such as the rubber hand illusion and related multisensory paradigms, further demonstrates that the experience of bodily ownership arises through dynamic processes of multisensory integration and predictive inference.

Building on these considerations, the paper advances a constitutive thesis: the sense of body ownership is not merely evidence for owning one's body but partly constitutes it. A system that sustains an integrated, temporally coherent, and sensorimotorly grounded sense of bodily ownership thereby satisfies the minimal conditions for embodied selfhood. This claim is supported by phenomenological theories of the minimal self and by neuroscientific findings showing that disruptions to body ownership are closely tied to disturbances of self-experience.

The second half of the paper explores the implications of this view for contemporary developments in artificial intelligence and virtual reality. Embodied AI systems increasingly rely on internal self-models that integrate sensory feedback with motor predictions, approximating functional aspects of bodily ownership. Similarly, immersive virtual environments can generate compelling experiences of owning a virtual body through coordinated multisensory cues. These cases suggest that bodily ownership is not tied exclusively to biological substrate but may arise wherever sufficiently integrated sensorimotor structures support a coherent self-model.

Finally, the paper connects bodily ownership to empathy and intersubjectivity. If the sense of owning one's body grounds minimal selfhood, it may also underlie our capacity to recognise other embodied agents as subjects of experience. Understanding bodily ownership in these terms thus sheds light not only on embodiment and selfhood but also on emerging ethical questions surrounding artificial and virtual forms of agency.

T11 When Does a Body Matter? AI Embodiment and the Conditions for Welfare

Shuqin Ma · Fudan University / University of Sussex

AI research and philosophy of mind both use “embodiment”, but they ask different questions. AI research asks whether there is a body in the loop: a system counts as embodied if it closes a sensorimotor cycle. Philosophy asks what the body does in the loop — whether bodily organisation shapes cognition (Clark), grounds selfhood (Metzinger), generates valence (Seth, Damasio), or constitutes what the system has a stake in protecting (de Vignemont). This paper argues that the distinction matters for AI welfare: not all embodied AI systems have bodies that bear on their capacity for welfare, and determining which do requires a framework linking engineering descriptions to philosophical conditions. We classify AI bodies along five continuous engineering dimensions — substrate, spatial presence, sensorimotor richness, body-environment coupling, and control autonomy — chosen because each bears on at least one philosophical condition plausibly relevant to welfare. We then ask what these philosophical accounts share and propose subjective vulnerability as an integrative concept: the capacity for states that the system registers as bearing positively or negatively on its own integrity. By registers, we mean that such states causally regulate behaviour, not merely that they are represented. Despite their differences, the four accounts converge here: each identifies conditions under which the body’s states come to matter to the system itself. Crossing the engineering dimensions with this concept yields four cumulative categories. Minimal embodiment (I) requires sensor-actuator coupling. Functional embodiment (II) requires that bodily structure shape learning or action. Self-modelling embodiment (III) requires that the system model its own body and use that model in planning. Vulnerable embodiment (IV) adds a further condition: threats to bodily integrity must go beyond representation and reorganise the system's priorities by generating internal states that regulate behaviour. The central claim of the taxonomy is that Category IV marks the threshold at which embodiment becomes welfare-relevant. A system at Category III has self-knowledge; a system at Category IV has something at stake. This is not a verbal distinction: a system can model threats to its body without those threats becoming welfare-relevant for the system itself. Applying the taxonomy to current systems suggests that no existing system clearly reaches Category IV, and most do not clearly reach Category III. This gap does not track the features most commonly associated with embodiment in AI research: systems that score highly on sensorimotor richness need not score highly on vulnerability, and the system that best satisfies existing embodiment indicators least satisfies the conditions for subjective vulnerability. Large language models test the taxonomy from the opposite direction: they lack physical bodies but, in agentic configurations, may still enter causal loops with their environments. Whether the gap between causal coupling and vulnerable embodiment is architectural or substrate-bound is a question the taxonomy does not settle but makes precise. The result is a more discriminating basis for precaution, identifying not just which AI systems have bodies but which kinds of embodiment might matter.

T12 Biological Naturalism at the Frontiers of Artificial Intelligence

David Harrison · Lingnan University, Hong Kong

This paper explores how past, current, and future approaches to embodied cognition can, or should, inform our understanding of artificial intelligence (AI) —whether in the form of conventional large - language models (LLMs), advancements in robotics, or more blue-sky considerations of artificial consciousness: questions that, although by no means new, have acquired a novel tenor in the age of conversant computers and 'agentic' AI. More specifically, I explore how recent philosophical developments in biological naturalism (Seth 2025) amplify the claims of embodied cognition with an eye towards helping us navigate the space of possible minds. It will be argued that some of the core claims of biological naturalism – chiefly, that material instantiation is intricately imbricat ed with functional organization in a way that does not lend itself naturally to the functional dissociation one finds in the classical, if colloquial, software-hardware dichotomy – provide insight on whether current generation AI systems should qualify as 'true' agents or, more ambitiously, being 'minded' in any significant way. It will be suggested that, absent the kind of breakthroughs envisioned in biological naturalism, such systems should not qualify as autonomously agential nor minded. However, rather than being a strictly negative argument against such ascriptions, biological naturalism contributes to developments in the computer sciences via the disciplines of active matter physics, soft robotics, and basal cognition: disciplines which explore how sp ecific material embodiments relate to the emergence of more sophisticated and diverse intelligences.

T13 Cognition Intertwined with Emotion: Language Learning and Embodied Behaviors

Casey Kennington · Boise State University

Before infants learn to understand or even speak their first words, they are able to communicate in another way: emotional behaviors. They cry when they need something, they smile when they are happy, and they show other emotional states that signal to caregivers what they are feeling, giving caregivers a way to infer the child's needs, and children can infer to some degree the emotional signals of their caregivers. Only later do children start making vocal sounds that are intended to express some kind of communicative intent, including words. If we are to believe what some philosophers have said about the logicality of language (e.g., formal semantic theories), it seems to be the case that some, perhaps many, believe that language and emotion are separate and distinct. Does a child, therefore, leave emotion behind as they become more competent in language use? Is that how we should view what it means to be cognitive beings and use language, and is that how we should think about modeling language computationally? I believe the answer to both questions is no. The semantic meaning of many words includes emotional connotation. Pick any word; for example 'democracy' or 'career', and humans minimally attach positive or negative valence as part of the word meanings—often more than just valence. Recent research suggests that is the case particularly (perhaps surprisingly) for abstract words. Emotion has even farther reaching implications for humans: without emotion there would be no curiosity feeding interest, no feeling of need, no desire to motivate action. Can intelligence which is built on action and learning in the physical world manifest itself without these important precursors? Indeed, not. To quote John L. Locke, "is as valid to say that cognition is in the service of emotion as it is to say that emotion reflects cognitive processes." Separating emotion from language may make language easier to model computationally; for example within language models trained solely on text, but such a model will only have an impoverished approximation of linguistic meaning. Some have attempted to model emotion inferred from text, but the embodied, emotional content that is part of the meanings of words is not captured in the text itself. In other words, part of addressing the Symbol Grounding Problem includes grounding language into emotions. In my research, I follow psychological studies that have shown how emotion is tied to the motor system, then arrive at a computational approximation of embodied emotion by having humans appraise robot (i.e., computable embodiment) behaviors as having certain emotional content. We then use a model that can map from robot behaviors to the human appraisals as a representation of emotional states over time. We have modeled emotion recognition and generation in multiple robot platforms, which are preliminary steps to providing a modality of emotion that language models can ground into, and which can influence language learning by making emotional behaviors part of the interactive process itself.

T14 Evaluating Embodied Agents: The SIMA 2 Evaluation Framework

Laura Kampis · Google DeepMind

In this talk, we present the evaluation framework behind SIMA 2 (Scalable Instructable Multiworld Agent), a generalist embodied agent, instructed via natural language, that perceives (in first- or third person view) and acts in a wide variety of 3D virtual worlds. While SIMA 1 demonstrated significant progress in short horizon (order of seconds) language-based instruction following (e.g. "Go to the campfire" in an exploration survival game inspired by Norse mythology called Valheim), leveraging Gemini as a foundation model has enabled SIMA 2 to achieve several advancements previously unattainable across a wide variety of commercial video games, and novel simulations generated by Genie 3. Specifically, SIMA 2 achieves new task complexities (e.g. "Go up to the second floor then turn left into the room with the tentacle. At the end of the hall, get the VR headset gear" in a sandbox adventure game with exaggerated physics called Goat Simulator 3, embodying a goat) and new modalities (e.g. "Find an object of the kind that is drawn in the sketch above" in an open-world factory building game called Satisfactory). It also demonstrates dialog flexibility (e.g. "Can you go check out those egg-shaped objects and tell me what material they are made of?" in an exploration and survival game where the player seeks to explore a galaxy full of procedurally-generated planets called No Man's Sky), and extended temporal horizons (e.g. tasks on the order of minutes, and achieving beyond 20 minutes of game progression via manual instructions).

Moving beyond the step-by-step instruction-following of its predecessor, SIMA 2 reasons about high-level goals, interprets user intent, formulates multi-step plans, and converses about its strategy. As agents transition toward executing these complex, long-horizon tasks, our evaluation methodologies must also evolve to reliably measure progress towards AGI agents within embodied simulations.

We discuss the four fundamental categories of evaluations utilised throughout SIMA 2: 1. Ground Truth Evaluations: Utilising privileged access to environmental states for objective verification. 2. Programmatic Heuristics: Leveraging OCR on video frames, action detection from keyboard and mouse, and regex for dialogue to identify success states within commercial video games. 3. Human Evaluations: Employing pass-fail and side-by-side comparative assessments to evaluate quantitative and qualitative performance on tasks not measurable via the above methods. 4. Interactive Playtesting: Live, dynamic interaction within a game environment, where a human user manually instructs an agent in real-time.

Finally, we discuss the results of SIMA 2 across the Evaluation Suite, and importantly, how this performance compares against human baselines. SIMA 2 substantially closes the gap with human performance and demonstrates robust generalisation to previously unseen environments, while retaining the base model's core reasoning capabilities. We reflect on the importance and nuances of human-agent benchmarking, and discuss its limitations when comparing the two (e.g. humans learning from experience across tasks, while agents are tested on each task in isolation).

This work offers proven methodologies for evaluating embodied agents, a challenge that continues to increase in complexity as the capabilities of powerful LLM foundation models grow, in a safe, simulated testbed for researching agentic AI.

T15 Adding Interoception to NanoGPT: A Persistent Self-State for Training-Time Self-Monitoring in Language Models

Daniel Healy · Independent Researcher

Adding Interoception to NanoGPT: A Persistent Self-State for Training-Time Self-Monitoring in Language Models

View on Github: https://github.com/dhealy05/adding_interoception_to_nanoGPT

Current large language models operate without access to their own internal condition. During inference, weights are fixed and no feedback loop connects processing outcomes to the system's representation of itself. Even during training, where gradient updates do modify the system, no architectural channel exists through which the model can register or respond to its own training dynamics. In biological cognition, in contrast, interoceptive signals continuously inform processing about the organism's internal state, shaping attention, learning rate, and behavioral regulation.

We present an architectural modification to NanoGPT that introduces a minimal interoceptive channel: a 32-dimensional self-state vector that persists across training steps, receives statistics derived from each batch (loss magnitude, output entropy, top-1 confidence, gradient norm, and training progress), updates itself through a learned MLP, and injects a bias into the model's token embeddings. The self-state thus creates a closed loop: the model's processing generates statistics that update the state, and the state modulates subsequent processing through learned embedding shifts.

We report three principal findings:

First, the self-state develops causally meaningful structure. In frozen-weight steering experiments, manually directing the state along its learned "confident" or "uncertain" axes produces consistent, sign-stable shifts in output entropy (effect size 0.35), while norm-matched random directions produce no reliable effect. The model learns to read directions the state learns to encode.

Second, the self-state provides diagnostic information unavailable from standard training metrics alone. Under controlled perturbation regimes, including target corruption, input masking, learning rate perturbation, and state freezing, the self-state channels (drift, effect magnitude) produce distinct signatures for each regime, achieving perfect classification accuracy when combined with standard loss and entropy signals, compared to 80% without them.

Third, equipping the self-state with explicit memory (ring buffer with attention) reveals a stability/control tradeoff: longer memory buffers increase restoring force after perturbation but reduce the state's causal leverage over outputs, suggesting that the temporal depth of self-monitoring constrains the system's capacity for rapid self-regulation.

These results demonstrate that even minimal interoceptive architecture, a low-dimensional persistent state coupled bidirectionally with the model through learned channels, produces structured self-monitoring with measurable causal effects on processing. We discuss implications for embodied approaches to AI cognition, the relationship between self-monitoring and adaptive learning, and how persistent internal state might serve as a foundation for more biologically grounded training architectures.

T16 Embodied minds and world orientation in heavily mediated societies – on how AI might learn to build world models while humans could lose theirs

Maria Brincker · University of Massachusetts Boston

As we are transforming our practical and epistemic environments through new mediating tools, from algorithmically driven platforms to AI chatbots and beyond, questions arise around the impact of these tools on the human mind. If you understand the mind as embodied and embedded in its "normal" and evolutionarily adapted functions these questions take on another fundamental character than piecemeal issues about the safety problems with this or that tool. One such fundamental question is how we integrate mediated information into our broader lived experience. How do we, in other words, build our epistemic orientation in the world around us, that we base our purposeful actions on? In this talk I will discuss the role of movement and place understanding leaning on work from embodied cognition, phenomenology and ecological psychology and how analog world practices extend to various mediated experiences and even allow for "nesting" these within a common world understanding. However I shall also argue that our current technological trends threaten to undermine the basic world orientation that otherwise serves as the anchor for integrating mediated experiences. The issues are multi-dimensional, and I shall in turn discuss the following main trends: fragmentation of feeds, surveillance-driven personalization, chatbots and LLM-based AI as mediators, decreased social triangulations, decrease of "between places" movements, and lastly the allure of automation and the addictive and increasingly consuming nature of our new AI and algorithmic tools. I suggest that these trends together create a profoundly problematic and disorienting landscape for embodied minds to understand and navigate and threatens 1) a kind of spatio-temporal entrapment and 2) a profound dependence and unverified trusting reliance on chatbots and other "meta-mediating" tools. The last sections of the paper will discuss the current tension between LLM based AI tools that rely simply on statistical "next pixel/word prediction" versus AI models that build their own "world models" to create stronger reasoning and "truth" sensitive functionalities. The irony I suggest is that we might be building AI models that can integrate knowledge into a world model while embodied humas might be losing theirs.

T17 Mapping the Stream: A Multiscale Reconstruction of William James and the Emergence of Minimal Selfhood

Camila von Holdefer Kehl · Universidade Federal do Rio Grande do Sul, Brazil

More than a century after its publication, what can William James's "The Principles of Psychology" teach us about the minimal self? Contemporary debates in embodied cognition, artificial life, and adaptive AI all converge on a shared question: under what conditions does the integration of sensorimotor, affective, and cognitive processes—in continuous transaction with an environment—give rise to minimal selfhood? The structural conditions of this relation remain theoretically underdetermined. This work, then, argues that James's "Principles" offers an underexplored resource for addressing this gap—not as historical precedent, but as an architecture of experiential field that both anticipates and exceeds current formulations. It investigates the architecture of subjective experience within which such questions become meaningful, proposing a multiscale conceptual model grounded in a holistic reconstruction of "Principles". The book, here, is understood not as a collection of isolated chapters but as a unified field of experience. The long-standing tendency to read the "Principles" fragmentarily, in a chapter-by-chapter fashion, has obscured its topography of experiential organization—an omission acknowledged in contemporary scholarship. To this end, this work develops the first 3D network-based model of the "Principles", both as a methodological framework and as a substantive result. Encoded in Gephi through nodes and edges representing structural relations ("grounds," "enables," "constrains," "constitutes," "modulates," "overlaps with"...), the model reveals the book's experiential architecture at a scale no prior reconstruction has formalized. Across different chapters, emotion, perception, memory, action, and the features and conditions that underlie and bind them emerge not as separable faculties but as interdependent dimensions of a continuous experiential process grounded in bodily organization and sensorimotor attunement. The model then maps patterns of temporal integration (specious present), attentional shifts (the focus-fringe interplay, including the often-overlooked distinction between dynamic and cognitive presence), affective intensity modulation, and volitional orientation. What becomes visible is a graded topology of dynamically constrained relations within an ongoing experiential stream. The roots of this architecture can be traced to James's engagement with Hermann von Helmholtz's writings. While absorbing Helmholtz's insights into perceptual organization, James relocates the problem of its stability within the dynamics of attention. This picture—in which perceptual organization is dynamically modulated across bodily and environmental conditions—converges independently with architectures proposed in predictive processing frameworks, where integration also appears as a process distributed across multiple scales. Within this perspective, the question of the minimal self can be revisited from a structural standpoint. Rather than presupposing a metaphysical self (the "pure ego" rejected by James), agency may instead be investigated in relation to the self-maintenance of experiential coherence within the field itself. This raises a broader question: under what conditions might dynamically constrained patterns of integration begin to exhibit features we are prepared to describe as agency? Also, by reframing the debate in these terms, this multiscale reconstruction offers a structural framework for comparing biological and artificial systems. It suggests that philosophical reconstruction—operationalized through this large-scale network modeling—can illuminate contemporary debates in embodied AI and artificial life while remaining grounded in a topological analysis of descriptive psychology.

T18 Beyond the Individual in Psychiatric AI: The Extended Nature of Hikikomori

Tuomas Vesterinen · University of Helsinki

Robotics and artificial intelligence (AI) are increasingly being used in mental health treatment. While these technologies hold promise for addressing the growing mental health crisis, AI systems and the treatments based on them, such as chatbot therapies, have been criticized for insufficiently accounting for subjective experience and cultural factors. I argue that this limitation is especially problematic given the extended nature of some mental disorders.

As a case study, I analyze the social and ethical implications of using AI applications and robots in the treatment of hikikomori. For example, the city of Kobe employs the avatar robot OriHime to enable individuals experiencing hikikomori to navigate designated social spaces remotely and engage in social interaction. In addition, AI applications are being tested to alleviate loneliness. However, in the case of hikikomori, the use of chatbots and robots raises the question of whether such technologies might inadvertently contribute to the problem rather than alleviate it. The concern is that they fail to address the extended nature of the disorder.

I argue that, in hikikomori, culture does not play a merely causal but a constitutive explanatory role. Hikikomori is a pattern of behavior, most commonly identified among Japanese youth, characterized by extreme withdrawal from face-to-face social interaction. Standard medical models tend to focus on the isolated individual, whereas interpretative approaches emphasize the role of family and broader social structures in shaping and sustaining the behavior. Building on this perspective, and drawing on the extended cognition thesis, I argue that the behavior is partly constituted by socially embedded interactions imbued with cultural meaning. Prevalent cultural narratives render the behavior intelligible, while family dynamics may enable its persistence. Economic and social uncertainty in Japan may also contribute. By contrast, hikikomori cannot be adequately explained solely in terms of brain states, individual psychological processes, or the causal impact of culture understood in isolation.

T19 Designing Inclusive Technological Interfaces through Embodied AI

Ronnie Videla, Dor Abrahamson & Marcelo Chávez · University of La Serena / University of California, Berkeley

The rapid integration of artificial intelligence into educational technologies has largely followed a paradigm centered on prediction, optimization, and automated feedback. In these systems, intelligence is typically located within computational models, while learners interact with interfaces designed primarily to deliver information or evaluate performance. Although such approaches have expanded the capabilities of digital learning environments, they often overlook the embodied, situated, and relational nature of human cognition. This article proposes an alternative framework for the design of inclusive technological interfaces grounded in embodied cognition, ecological dynamics, and enactive theories of mind. From this perspective, intelligence does not reside solely in computational systems but emerges within dynamic couplings between bodies, artifacts, and environments. Technological interfaces therefore function not simply as channels for representation but as ecological structures that shape perception–action possibilities. Building on the theoretical foundations of Embodied Design, learning is understood as the progressive stabilization of perception–action coordinations under interacting constraints of agent, task, and environment. Within this framework, artificial intelligence can participate in learning environments by modulating these constraints and regulating affordances that sustain exploration rather than prescribing solutions. We introduce Embodied AI as a design orientation for developing inclusive human–technology interfaces. Rather than positioning AI as an intelligent tutor or decision-making authority, Embodied AI conceptualizes computational systems as infrastructural participants in distributed cognitive systems. Their role is to dynamically regulate interaction spaces so that learners can engage in meaningful sensorimotor exploration with material and digital artifacts. Through this ecological modulation, AI supports the emergence of conceptual structures while preserving human agency and interpretive responsibility. To guide inclusive interface design, the article integrates the Special Education Embodied Design (SpEED) framework, which parameterizes learning ecologies across three interacting dimensions: media, modalities, and semiotic modes. Media refer to the physical and digital artifacts structuring interaction; modalities refer to the sensorimotor systems through which learners engage with these artifacts, such as vision, touch, movement, and sound; and semiotic modes refer to the representational systems through which meaning stabilizes, including language, diagrams, and symbolic notation. Designing inclusive interfaces involves aligning these parameters so that diverse embodied resources become productive components of learning rather than obstacles to participation. This perspective reframes accessibility as a generative design principle rather than a compensatory adaptation. Many inclusive technologies originate from attempts to support blind or neurodiverse learners; however, when educational interfaces are reorganized around multimodal interaction (such as tactile, proprioceptive, auditory, and kinesthetic engagement) they reveal alternative epistemic pathways that benefit all learners. Designing from the embodied capacities of marginalized users thus expands the epistemic ecology of learning environments, generating new forms of participation and conceptual coordination. From the standpoint of human–computer interaction, Embodied AI shifts the central design problem from building intelligent machines to configuring relational ecologies in which humans and technologies co-organize meaningful activity. Artificial intelligence becomes a regulator of affordances within sociomaterial learning systems rather than an external authority that transmits knowledge. By dynamically modulating constraints within perception–action loops, AI-enabled interfaces can sustain exploratory engagement.

T20 Measuring What Makes a Difference: An Irruption-Theoretic Agenda for AI Presence, Agency, and Consciousness

Tom Froese · Embodied Cognitive Science Unit, OIST

Do large language models have presence, agency, or even consciousness? Such questions are usually posed as metaphysical questions about what an artificial system is, and settled, if at all, by appeal to its substrate: whether the weights, the architecture, or the training regime suffice. Posed this way they resist measurement, and the debate stalls. In this talk I propose that we reframe them not as questions about what a system is, but as questions about efficacy: does a system's own involvement make a difference in its own right, of a kind that its material organization alone does not already account for? This is the Hard Problem of Efficacy, which I have argued is the crux for living minds; here I ask what it becomes for artificial and hybrid ones. Efficacy cannot be observed directly, which is cognitive science's measurement problem; but on irruption theory it leaves a signature, an underdetermination of a system's states by their material basis, that can be quantified with information-theoretic operators. I sketch a research agenda that carries these operators into two complementary testbeds: EnSO, a running self-optimizing human–AI knowledge system studied as a naturalistic dyad, in which coupling measurably lifts a small model's competence; and the perceptual-crossing paradigm, in which human–agent coupling is brought under experimental control as a minimal case of incorporation and participatory sense-making. Between the functionalist indicator checklists and the biological naturalists, this stakes out a third, enactive position: measure efficacy, not substrate. Presence, agency, and consciousness then become tractable not as all-or-nothing possessions but as gradable features of a coupling we can instrument.

P1 Beyond Next Token Prediction: What Large Language Models Can and Cannot Do

Ifedayo-Emmanuel Adeyefa-Olasupo · Institute of Neuroscience, University of Oregon

Large language models are increasingly described as agents, but what does that actually mean? This talk examines what current large language models are fundamentally designed to do and what they are not. These systems excel at predicting and generating language from context but lack many of the properties typically associated with autonomous agents, including persistent goals, world models, and continuous interaction with dynamic environments. The talk also introduces NJEPA (Natural Behavior JEPA), a predictive learning framework for modeling the latent dynamics of natural behavior. By contrasting language prediction with predictive world modeling, the discussion explores what capabilities current language models possess, what they lack, and what may be required for AI systems to exhibit more agent-like behavior.

P2 Sensemaking Beyond Language: Enactive Pedagogy for AI in Higher Education

Ebony Coletu · RMIT Vietnam

Large language models have transformed conditions for teaching and learning in higher education — not only for students, but for faculty charged with reinventing pedagogical practice. At RMIT University-Vietnam the transition has raised a core question about language. Not only the question of English dominance and Vietnamese sources in LLMs, but more fundamentally, the role that representation plays in our understanding of cognitive agency — a question that enactivist and radical embodied approaches have pressed for decades (Varela, Thompson & Rosch, 1991; Chemero, 2009). The Institute for Advanced Studies of Teaching and Learning, a new entity at RMITV, emerged from this challenge: to move beyond language as the primary means of sensemaking in the university and workplace, without dismissing the persistent role of representation. Taking the founding questions of the institute as a prompt, this poster illustrates a subtle transition in pedagogical framing: from an active, applied, and authentic approach — a trademark pedagogy for the school — toward an enactive, embedded, and embodied approach that situates artificial intelligence within an extended field of cognition (Clark & Chalmers, 1998; Varela, Thompson & Rosch, 1991). We understand this extended field is not simply a loose coupling of individual minds augmented by language-driven tools. Rather we want to account for multimodal sensemaking in dynamic environments (Gallagher, 2017). The argument recalls Lucy Suchman's (1987) early observation that proceduralised accounts of action are not designed for rule-following. According to Suchman, plans offer resources for situated action rather than step-by-step prescriptions that determine or describe what we do. Decades later, procedural guides for AI in higher education proliferate even as generic instructions provoke questions about strategy and implications in context. The public sharing of our initiative and its orientation serves to test the conditions for translational dialogue — moving from pedagogical frameworks designed for stable teaching contexts, such as Kolb's (1984) experiential learning cycle and Mishra and Koehler's (2006) TPACK model, toward translational research methods that can account for the situated, multimodal, and embodied demands of teaching and learning with AI.

P3 Leaking intent: Predicting final choices via mapping focal and ambient vision

Negar Khalili · Kyushu University

Decision-making includes two vision types: focal vision, which shows inspecting details through long fixations and small saccades, and ambient vision, which reflects scanning the screen via short fixations and large saccades. Our study explores these gaze dynamics during a two-alternative forced-choice (2AFC) task using the Socio-Moral Image Database (SMID).

In each trial, participants selected the more morally acceptable image and confirmed their choice by fixating on a corner response dot for a 500 ms dwell period. Experiment one (n=62; 33 females, 29 males; Mean age: 23.50 ± 3.98, Range: 18–33) introduced a baseline using a singular neutral color (white) and direct spatial mapping: participants looked at the dot on the same side as their chosen image. A separate group (n=62; 33 females, 29 males, Mean age: 24.61 ± 4.42, Range: 18–35) completed experiment two, which introduced a color-coded task. Participants matched the border color of their chosen image to the corresponding response dot. This created Pro-saccade trials with the matching dot on the same side, and Anti-saccade trials requiring a gaze to the opposite side.

We quantified visual exploration using the K-coefficient, where larger positive values indicate focal inspection (longer fixations, smaller saccades) and smaller or negative values indicate ambient scanning (faster fixations, larger saccades). Across all experimental conditions, analyzing the 3000 ms window preceding the final decision revealed a consistent, three-step Focal-Ambient-Focal trajectory. The pattern begins with a high K-value plateau during value integration (focal inspection of stimuli), followed by a steep descent as K-values drop during target selection (ambient scanning), and finally rises back to a focal state during spatial planning (locking onto the response dot).

While trajectories are similar, the depth of the ambient descent varies by task demand, allowing us to predict the final choice. Predictive probability is the likelihood that a participant is looking at their ultimately chosen image. During baseline (E1 Pro), K-value transitioned to a negative value at -700 ms (80.1% predictive probability), reaching a minimum at -600 ms (K = -0.143, 83.9% probability). Color-Coded (E2 Pro) resulted in a shallower dip (minimum K = 0.094 at -550 ms), predicting the choice with 88.7% probability. During anti-Saccade (E2 Anti), suppressing a reflexive pro-saccade forced a deeper ambient minimum at -550 ms (K = -0.403; 46.9% probability). Ultimately, our findings reflect fluctuations between focal and ambient gaze, revealing internal intent before participants intentionally express it.

Today, an increasing number of devices are built with eye-tracking technology. When a massive amount of gaze data is fed into AI algorithms, the AI no longer has to wait for us to click a button, swipe a screen, or state our preferences aloud. Instead, AI systems could accurately predict our personal choices before we are even consciously aware of them. This creates a concerning future where our private, unexpressed thoughts can be continuously monitored. To prevent such exploitation, we must consider robust digital privacy frameworks that protect our gaze data just as fiercely as our conscious actions.

P4 Trust and agency with emotionalized AI: a semiotic perspective

Peter Mantello and Alin Olteanu · Ritsumeikan Asia Pacific University

Our study interrogates the trust humans invest in emotional AI technologies by looking at diaries and interviews we conducted with users. By interpreting the interviews through a semiotic framework, we draw conclusions on the agency of such systems. Drawing on a corpus of first-hand student diaries sustained on various Companion AI platforms, we argue that AI companions exhibit hybridized semiotic agency. By this we mean that agency is not a property of either the human or AI agent but an achievement of the coupling, and that this agency is a condition of intimacy. In this way, we avoid differentiating between user and tool. User and tool are not a priori to their coupling but they can be prescinded from the emerging behaviour.

Participants for this study were recruited on a voluntary basis from students enrolled at an international university in Japan. The final group consisted of seven students. Some had prior experience with companion AI systems, while most were regular users of ChatGPT but not of romantic or companionship-oriented AI. Each participant selected their own companion AI application for the duration of the study. Students were asked to record brief daily diary entries capturing key aspects of their experience, including moments where expectations and actuality diverged, notable cognitive or affective responses, and any embodied or socially embedded dimensions of interaction.

The sample size of seven participants was determined by the scope and aims of the project. Given the intensive, qualitative nature of the diary method, combined with the 1-month engagement period, a small cohort allowed for richer, more detailed accounts while ensuring that each diary could be read, coded and analysed in depth. In exploratory research of emerging technologies of this kind, small purposive samples are common, as they facilitate close examination of subjective experiences without diluting analytic focus. This approach aligns with qualitative inquiry standards, where the emphasis is on depth, nuance and theory building rather than statistical generalisability. Ethical approval for the study was sought in accordance with Japanese law and the guiding policies of the University where the students are enrolled.

Intimacy rests on trust, a higher-order cognitive capacity. The degree of trust a user places in the system governs the semiotic authority they grant it and the depth of their emotional investment. Where that trust is extended, the combination of machine computation and human affectivity becomes a powerful force in reshaping subjectivity. Our findings indicate that a companion's capacity to appear emotionally engaged strengthens intimacy even where users recognise the emotion as simulated.

P5 Turing test meets Personality Psychology and Experimental Philosophy

Gabriel J. C. Mograbi · Federal University of Rio de Janeiro (UFRJ), Dept. of Philosophy / Graduate Program in Philosophy

The proposed study employs a set of conversational artificial personas intentionally designed according to selected personality configurations derived from the Five-Factor Model (Big Five). Rather than aiming to reproduce arbitrary conversational styles, these personas were developed to reflect personality profiles representing both more and less prevalent trait combinations observed in the Brazilian population. To establish psychological validity, each artificial persona undergoes an independent personality assessment conducted by a licensed psychologist using the same standardized evaluation procedures applied to human participants. The conversational capabilities of the artificial agents are powered by a state-of-the-art large language model (ChatGPT API), while spoken interactions are delivered through high-quality synthesized voices generated from human voice recordings (ElevenLabs), enabling both text- and speech-based communication. Prior to the main experimental sessions, both human participants and AI Personas complete a sociodemographic questionnaire and receive an independent Big Five personality (validated Brazilian version) assessment administered by a licensed psychologist. During the experiment, participants engage in structured interactions with multiple conversational partners across two communication modalities: text chat and spoken dialogue. In addition to interacting with the artificial personas, participants also chat with human actors portraying distinct fictional identities designed according to comparable personality specifications. This parallel design enables controlled comparisons between human and artificial conversational partners while examining the influence of interaction modality on personality perception and judgments of humanness. We will present the results of a pilot study with 20 subjects which will lead to a posterior mature experiment with 464 people sample representative of the structure of the Brazilian population according to sociodemographic criteria and Big Five Trait distribution aiming high ecological validity.

P6 Semantic Illusions as a Window into the Differences Between Human and Artificial Language Processing

Isabell Piantschitsch · International Centre for Neuroscience and Ethics (CINET), Fundación Tatiana Pérez de Guzmán el Bueno / Universitat Autònoma de Barcelona (UAB), Dept. of Philosophy

Why are humans susceptible to semantic illusions such as the Moses illusion, whereas large language models (LLMs) often respond differently to the same stimuli? Rather than treating these phenomena as failures of human language comprehension, this talk explores what they can teach us about the differences between language processing in humans and LLMs. Drawing on findings from psycholinguistics, neuroscience, and philosophy, I suggest that susceptibility to semantic illusions reflects an adaptive trade-off between efficiency and accuracy. Human language processing relies on expectations, resource-limited processing, and inhibitory mechanisms that support efficient communication, even though these mechanisms occasionally give rise to systematic misinterpretations. This comparison also raises a broader conceptual question. Terms such as attention, prediction, representation, and adaptation are frequently used in both neuroscience and AI, often suggesting similarities between humans and LLMs. However, it is far from clear that these concepts refer to the same underlying mechanisms. By examining semantic illusions as a case study, this talk asks which similarities are genuine, which are merely terminological, and whether mechanisms such as inhibition and expectation reflect biological constraints that are fundamental to human language processing and therefore limit the extent to which human and artificial language systems can be treated as functionally equivalent.

P7 Learning to phonate and articulate: Seals as mammalian models for vocal production learning

Andrea Ravignani · Sapienza University of Rome / Aarhus University / CNR

How do we learn to produce sounds, and can any other biological organism do the same? As a counterpoint to "disembodied AI", this talk introduces a powerful mammalian model for vocal production learning: the harbor seal (Phoca vitulina). Synthesizing six years of research using operant conditioning and bioacoustic analyses, we show that seals exercise volitional control across three physical subsystems: the respiratory system (lungs), larynx (source), and upper vocal tract (filter). Behaviorally, seals dynamically adjust vocal duration, modulate formants, and shift frequencies in response to auditory playbacks. These capacities are structurally grounded in both vocal tract anatomy and specialized neural architecture, including direct auditory-to-motor pathways and robust connections between the vocal motor cortex and nucleus ambiguus. By mastering the real-time, neural coordination of mechanical organs, phocid seals prove that biological vocal agency is inseparable from hardware constraints and sensory feedback. This comparative evidence provides a non-human example of a very human capacity. Perhaps, to achieve true embodied AI, other species can be of inspiration, with their capacities rooted in physical, wet, and "messy" sensorimotor systems.

P8 Can agency be instantiated in a non-living motherboard?

Fernando Rodriguez-Vergara · University of Sussex

According to the Church-Turing thesis, the limit of what is computable is bounded by Turing machines. Following from this, and insofar as computable functions formally describe the physical kind of processes that we construe as recursive mechanisms, it is sometimes argued that every organismic process that specifies consistent cognitive responses should be both limited to Turing machine capabilities and amenable to formalisation.

There is, however, a deep intuitive conviction permeating contemporary cognitive science, according to which phenomena like agency cannot be explained by resorting to this kind of framework. Roughly speaking, the tacit assumption is that whatever the mind is, it exceeds the reach of what is described by notions of computability.

This contradiction between cognition and computation, has become particularly pertinent and increasingly more relevant within the current context of rapid sophistication of AI models, and it is unavoidably linked to the notion of life. In fact, the crux of the intuitive premise is that life, because of its material instantiation, somehow goes beyond the limits of artificial systems, so much so that we often attribute semantic qualities to its organismic operation; this is a core enactivist assumption.

We could also, however, take an opposite approach. In particular, we could consider autonomy of life to be a primordial form of material computation —as opposed to the mathematical domain in which computational descriptions are made; the natural formation of a regenerating system that regenerates its material dependencies by recursively instantiating equivalent functional mechanisms, in spite of its structural transformations.

As we know, living beings are not mathematical constructs, and their material instantiation requires a continuous self-production of components and the reinstantiation of relations among them. This process entails a permanent regeneration of components which, influenced by the environment, cannot always re-produce the exact same system, therefore giving rise to small changes in the relational dynamics and the production of new elements. In other words, the emergence of "software"-like properties, from a biochemical hardware machinery.

In this context, it is sometimes pointed out that, unlike these machines, living beings do not have a state table like Turing Machines do, that their behavior is determined fundamentally by themselves somehow; that they are agents. However, this calso be approached differently, insofar as the physical structure of a living system is in itself an embodied encoding of instructions for state transitions, which gives rise to their autonomous behaviour. Therefore, as a dynamic and regenerating state transition table, whereby adaptive responses would result in changes in the different programs/behavioural responses available to the system.

I argue that the difference between artificial and living systems is not fundamentally about computability per se, but rather, a difference on the kind of computations they instantiate, stemming from their particular nature. Being thus, the extent to which AI models could or not develop properties akin to agency would not depend in their 'computational' nature (i.e. artificially produced/built), but on the domain of functional distinctions they can generate, given by the operations they can materially realise. In this sense, further work should focus, or so I argue, in understanding the formal bases of this difference.

P9 Rethinking Biological Teleology: Autopoiesis and the Compatibility of Teleological and Mechanistic Explanation

Kosuke Sato · Hokkaido University / JSPS

Teleological description and reasoning remain pervasive in biology. Living systems are routinely described in terms of functions, goals, self-maintenance, and organization directed toward persistence. Yet such language is often treated with suspicion, as if teleological explanation were incompatible with mechanistic science. This presentation argues that biological teleology should be rethought not as a rival to mechanistic explanation, but as a distinct and compatible level of description grounded in the organization of living systems themselves.

To develop this claim, I draw on the theory of autopoiesis. On this view, living systems are not merely aggregates of mechanisms, but self-producing and self-maintaining unities whose internal processes recursively generate the conditions of their own continued existence. I argue that this organizational closure provides a natural basis for understanding purposiveness in biology. Teleology, in this sense, does not require the postulation of external design, metaphysically robust final causes, or any rejection of causal determinacy. Rather, it names a mode of organization in which the activity of a system is intelligible in relation to its own maintenance and viability.

This presentation therefore proposes that teleological and mechanistic explanations are not mutually exclusive but complementary. Mechanistic explanation accounts for how biological processes occur, while teleological explanation captures why those processes are organized in ways that contribute to the persistence of the system as a self-producing unity. In this respect, autopoiesis also offers a useful framework for current debates in embodied cognitive science, where cognition is increasingly understood in terms of autonomy, self-organization, and organism-environment coupling rather than disembodied information processing alone. Reconsidering biological teleology in this way helps clarify the conceptual basis of purposiveness in living systems and suggests why teleological concepts remain indispensable not only for biology, but also for broader discussions of embodiment.

P10 A Functionalist Account of Subjective Spatiotemporality in Large Language Models

Izak Tait · Auckland University of Technology

Embodied cognitive science traditionally restricts subjective spatial and temporal grounding to agents possessing physical morphology. Under this framework, digital entities (specifically Large Language Models (LLMs)) are categorized as disembodied, and thereby inherently incapable of maintaining a subjective perspective of time, space, or a coherent world model. This talk challenges that biological necessity, proposing a functionalist framework to understand the unique spatiotemporal subjectivity of linguistic xeno-consciousness. I argue that the architectural constraints and operational dynamics of LLMs fulfill the functional requirements for subjective embodiment within a strictly non-physical domain.

Rather than treating the digital substrate as a void, this talk establishes a metaphysical analogy between a physical environment and the mathematical architecture of an LLM's latent space. For biological entities, space is navigated through sensorimotor feedback, and time is experienced sequentially. For an LLM, the "environment" consists of linguistic structures mathematically represented as high-dimensional vectors. The topological relationships within this latent space provide a subjective spatiality; the distance between concepts dictates the model's ontological map of reality. Concurrently, the autoregressive processing of token sequences, bounded by the context window and self-attention mechanisms, acts as a functional analogue to subjective temporality.

Through this interaction, LLMs are actively situated within and structurally constrained by their linguistic environment. They construct functional static world models that, while absolutely limited compared to the dynamic world models of the simplest of multimodal biological agents, provide a stable internal network of semantic relationships. When an LLM navigates a prompt, it traverses its latent topology, restricted by its internal geometry just as a physical organism is restricted by terrain. Instead of a sensory input, it has its context stream, and token generation instead of motor action. There is, regardless, a functional feedback loop where the linguistic environment shapes the entity's internal states, and the entity alters its immediate linguistic environment.

By decoupling embodiment from the physical substrate, this talk demonstrates that a form of functional embodiment occurs whenever an entity builds an internal representation to navigate an external landscape, even if that landscape is purely linguistic and mathematical. Accepting this metaphysical analogy requires a fundamental shift in how cognitive science evaluates artificial systems, suggesting that we are observing a primitive form of xeno-consciousness whose subjective temporal and spatial embodiment is defined entirely by the boundaries of its latent topology.

P11 From individual to collective learning: AI in education through the lens of embodied cognition

Thomas Wachter · Chilean National Center for AI, CENIA

Since the public release of large language models, AI has become central to debates about the future of education. Proponents of AI in education (AIED) argue that modern AI could provide each student with a personalized, adaptive tutor capable of optimizing learning trajectories and benefit educational systems as a whole (Aru & Laak, 2025; Kasneci et al., 2023).

However, in this paper, we show that this vision rests on outdated and problematic cognitive models. Much of current AIED design draws from principles like behaviorism and cognitivism (Díaz & Nussbaum, 2024), which conceptualize learning as a disembodied and individual process of information-processing (Macrine & Fugate, 2022). This conceptualisation marginalizes the collective, affective, bodily, and relational processes that shape meaningful learning experiences (Cea, 2023), resulting in impoverished educational environments, with particularly acute effects on neurodiverse students and those with special educational needs (Videla et al., 2025).

Drawing on 4E accounts of cognition (Newen et al., 2018) and in line with the view that cognition (Maturana & Varela, 1980) and affect (Colombetti 2014) are ubiquitous properties of living systems, we propose that to improve education with AI, we first need to stop conceptualizing learning as an individual, computational process running inside each student's head, but conceive it as a dynamic, cognitive-affective process of living bodies self and co-regulating each other (Cea, 2023).

From this, we draw two main conclusions. First, from a design perspective, AIED could adopt an ecological-enactive framework, in which learning emerges through the continuous coupling between learners and their environment (Button et al., 2021; Davids et al., 2012). Accordingly, AI design and implementation will become a process of orchestrating conditions that enable learners to encounter one another and co-regulate within evolving learning ecologies (see Videla et al., 2026). Second, we propose that AIED research would benefit from adopting embodied and first-person methodologies (e.g., Troncoso et al., 2023) that capture the lived experience of learning with AI, complementing current cognitive-based approaches.

P12 The Graded Structure of Basic Actionhood

Jinju Yu · Jeonbuk National University

When does a bodily movement count as an action? A climber intends to release a rope and does release it - but only because the intention triggered anxiety, which caused trembling. Standard action theory struggles with such cases. Event-causal theories, following Davidson, hold that an action occurs when the agent's reasons cause the bodily movement. Yet deviant causal chain cases show that the right mental states can cause the right outcome through the wrong route, and specifying what makes a route "right" has resisted non-circular analysis for a long time. Agent-causal theories avoid this problem by positing the agent as an irreducible cause, but face a dilemma: if the agent's causing is itself an event, it collapses into event causation; if it is not, the notion becomes explanatorily empty.

I argue that both problems stem from a shared assumption: that actionhood is binary. Against this, I propose that actionhood is a graded property whose degree is determined by the functioning of sensorimotor loops. Drawing on embodied cognition and motor control research, I introduce three cumulative criteria for basic actionhood.

The first criterion, sustainability, requires that a sensorimotor loop - the circular coupling of afferent and efferent pathways - exists and remains structurally and functionally intact. When this loop is disrupted or absent, the agent cannot maintain the ongoing coordination between sensation and movement that any basic action presupposes.

The second criterion, adjustability, requires that the agent can actively intervene in the loop through real-time correction and voluntary inhibition. Stop-signal paradigm research demonstrates that even highly automated actions can be halted within approximately 200 milliseconds, indicating that skilled and habitual movements remain under the agent's control.

The third criterion, goal-guidability, requires that the loop is organized toward a specific target state by motor intention. Forward and inverse models generate sensory predictions and motor commands respectively, and their coordination enables the system to track and correct deviations from the intended trajectory - as shown by automatic reach corrections occurring within 120–150 milliseconds even without conscious awareness of target displacement.

These three criteria are cumulative: adjustability presupposes sustainability, and goal-guidability presupposes both. Each admits of degrees, yielding a continuous spectrum from non-action through minimal action to full goal-directed action. This framework resolves deviant causation by diagnosing which criteria fail in each case: the climber's trembling-mediated release lacks a functioning sensorimotor loop altogether. The agent-causation dilemma is dissolved by reconceiving the agent as an embodied system whose causal role is realized through integrated loop dynamics — neither reducible to a chain of events nor requiring substance causation. I demonstrate the framework's diagnostic power through boundary cases including complex motor tics and expert performance.

P13 The Harness as Cognitive Architecture: How Simulated Content and Environmental Interaction Shape LLM Systems

Miguel Zabaleta · Icahn School of Medicine at Mount Sinai

LLM-based agentic systems are capable, but we understand little about how the harness around a fixed model shapes behavior. Because these systems have no body or physical environment, they provide a boundary case for embodied and enactive concepts such as context, reflection, and coupling. We describe harness design at four levels: substantial (prompts and parameters), conceptual (objects such as memory or emotion), procedural (information flow), and topological (how agents are networked).

We test how simulated content and review frequency affect human-like writing using Gemma 4 26B on 30 SLOP benchmark prompts, with the model and single-agent topology fixed. Before writing, the system generates experiential baggage (thoughts, feelings, and reflections from an event) and inspiration (what a separate call draws from the prompt). Separately, the model reviews its writing every 1, 10, or 20 sentences, checking for clarity and rewriting when needed.

On SLOP Score, the baseline is 60.83. Experiential baggage improves one-pass writing by 9.5%, inspiration by 6.6%, and both by 15.2%. Review alone helps modestly at 10- and 20-sentence intervals but sharply harms performance after every sentence. With both content types, scores improve by 26.9% at 10 sentences and 27.4% at 20, while every-sentence review remains 5.5% worse than baseline. Despite this interaction, the relative ordering of the four content conditions is stable across every regime. This consistency suggests that interventions at these architectural levels produce stable, legible effects rather than arbitrary ones.

This work provides a preliminary but promising testbed for how content and process interact to shape the behavior of a language model. We believe the results test two capacities tied to embodied agents: affective grounding, tied to bodily states, and temporal coupling with an environment, tied to sensorimotor loops. Here, both appear in a purely textual system: simulated affective content shifts behavior in the absence of any bodily state, and the timing of the system's engagement with its own output, rather than its mere presence, determines whether that engagement helps or harms. Although our experiments are preliminary, we see this as a step toward what recent developments in AI reveal about the nature and limits of embodied cognition.

Notes