OPEN ACCESS

Artificial Intelligence PERSPECTIVE

The Innovation Informatics

The Post-Turing condition: Conceptualising artificial subjectivity and synthetic sociality

Thorsten Jelinek,[1,4,*] Patrick Glauner,[2] Alvin Wang Graylin,[3] and Yubao Qiu[4,5]

[1]Centre for Digital Governance, Hertie School, Berlin 10117, Germany

[2]Department of Applied Computer Science, Deggendorf Institute of Technology, Deggendorf 94469, Germany

[3]Digital Economy Lab, Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University, Stanford CA 94305, USA

[4]International Research Center of Big Data for Sustainable Development Goals (CBAS), Beijing 100094, China

[5]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

*Correspondence: t.jelinek@hertie-school.org

Received: January 16, 2026; Accepted: August 14, 2026; Published Online: August 20, 2026; https://doi.org/10.59717/j.xinn-inform.2026.100063

Citation: Jelinek T., Glauner P., Graylin A. W., et al. (2026). The Post-Turing condition: Conceptualising artificial subjectivity and synthetic sociality. *The Innovation Informatics* **2**:100063.

Generative AI has become a social technology: its outputs enter human practices of interpretation and coordination. This Perspective suggests that the next shift may go further. As systems develop from representation-only models into situated agents and multi-agent ecosystems, parts of sense-making itself (interpreting situations, stabilising reference, framing decisions) could become automated. We call this the Post-Turing Condition: the question is no longer whether AI imitates human intelligence, but whether AI systems come to participate in constructing social reality, with humans at risk of exclusion from meaning formation. The PRMO framework (Perception, Representation, Meaning, the Real) is offered as a lens for reading current design trajectories, and Quadrangulation is proposed as a design principle retaining the human subject as a constitutive reference.

THE POST-TURING CONDITION

The Turing Test asks whether a machine can be called intelligent when its responses are indistinguishable from a human's.[1] More than seventy-five years on, that question no longer captures what is at stake. The Post-Turing Condition, as we use the term, does not name a new kind of machine — computation remains Turing's, with its power and its limits — but a shift in the unit of analysis: from the imitation of human intelligence to the possible automation of social sensemaking. AI systems may increasingly participate in how situations are interpreted, how reference is stabilised, and how decisions are framed, and humans may be progressively excluded from the formation of meaning, identity, knowledge, and reality.

This concern differs from neighbouring debates: it is not a capability question, not the consciousness debate, not standard AI ethics, and not a general sociology of technology. It concerns the social location of interpretation: who, or what, participates in stabilising what things mean. It is closest to actor-network theory (ANT), which likewise distributes agency across human and non-human actors; but where that approach is symmetric and descriptive, the present framework is asymmetric by design: it retains the human subject as a constitutive reference rather than flattening it into one node among others and uses "subject" structurally to bracket the consciousness question without conceptual paralysis. In short, ANT describes who is in the network; the present framework prescribes keeping the human in it, a commitment we develop below as Quadrangulation.

To make this tractable for an engineering readership, we read current design trajectories through PRMO, a functional decomposition of human subjectivity into Perception (P), Representation (R), Meaning (M), and the Real (O). PRMO is a conceptual translation lens, not an experimentally validated model; artificial subjectivity and synthetic sociality, introduced below, name possible design trajectories and governance thresholds, not established technical facts. The aim is to make a development path visible early enough for design and governance choices to be made deliberately.

PRMO AS A TRANSLATION LENS

Each dimension can be introduced through a system function. Perception (P) is situated access to an environment: registering and differentiating world states from a particular position. Representation (R) is the abstraction of experience into symbolic or latent form. Meaning (M) is relational rather than internal: following Davidson, reference is stabilised through triangulation, in which a subject, an object, and other subjects orient towards the same thing.[2] What triangulation opens is, in Gabriel's terms, a field of sense: a shared context in which an object can appear, be interpreted differently, and remain contestable.[3] The Real (O) is the substrate that grounds any system causally (bodies, sensors, hardware), while resisting full conceptual closure; it settles nothing about consciousness.

Two asymmetries make the lens productive for AI. Humans develop perception before representation; current AI began with representation alone and is acquiring perception afterwards, which is an inversion that helps explain why systems can exhibit fluent representation without situated judgement. And in humans the four dimensions operate as an integrated whole; in AI they arrive as separable engineering layers, each at a different maturity. Figure 1 maps this trajectory in three steps, pairing each philosophical reading with its engineering instantiation and maturity status.

A DESIGN TRAJECTORY IN THREE ENGINEERING STEPS

Representation only: The distributed subject

Large language models operate at R. Self-supervised pretraining compresses large corpora of human-produced text into model weights; next-token prediction over these learned abstractions yields high language performance without perception of, or coupling to, an environment.[4] The weights encode correlations among externalised patterns of human expression: second-order abstractions. Whether such systems "understand" remains unresolved within the field itself: expert opinion divides roughly in half, and benchmark evidence is undermined by shortcut learning.[5]

Gabriel reads the resulting artefact as a distributed subject: a "magic mirror" that captures and recombines the symbolic and emotional traces of millions of human subjects without being any of them.[6] Žižek's objection marks the limit of the analogy: the human subject is constituted by lack, embodiment, desire, and an unresolved relation to the Real, whereas the model inherits representations of desire and conflict without desiring or suffering anything.[7] The distributed subject thus claims inheritance of human expression, not understanding. Status: deployed at scale.

Sensorimotor coupling: Situated agents and artificial subjectivity

The second step couples the distributed subject to an environment. In digital settings, tool use and persistent memory provide runtime coupling, and computer-use agents close a perceive–reason–act loop over raw interface states. In physical settings, vision–language–action models extend pretrained foundation models into robot control: "robot foundation models" pairing a pretrained vision–language backbone with high-frequency motor commands. World models, which learn an environment's dynamics to predict consequences and plan over them, belong to this same stage; PRMO reads them as representation coupled to perception, not as a separate paradigm. Their progress accelerates the trajectory traced here rather than invalidating it: faster grounding brings artificial subjectivity and synthetic sociality nearer, and the problem of human inclusion with them.

Something architecturally consequential follows: a deployed agent accumulates a history of its own: historical particularisation, the divergence of an individual agent from its shared base model through accumulated context, memory, and adjustment. In currently deployed systems this operational Δ resides in context and memory, not in model weights; consolidating it into

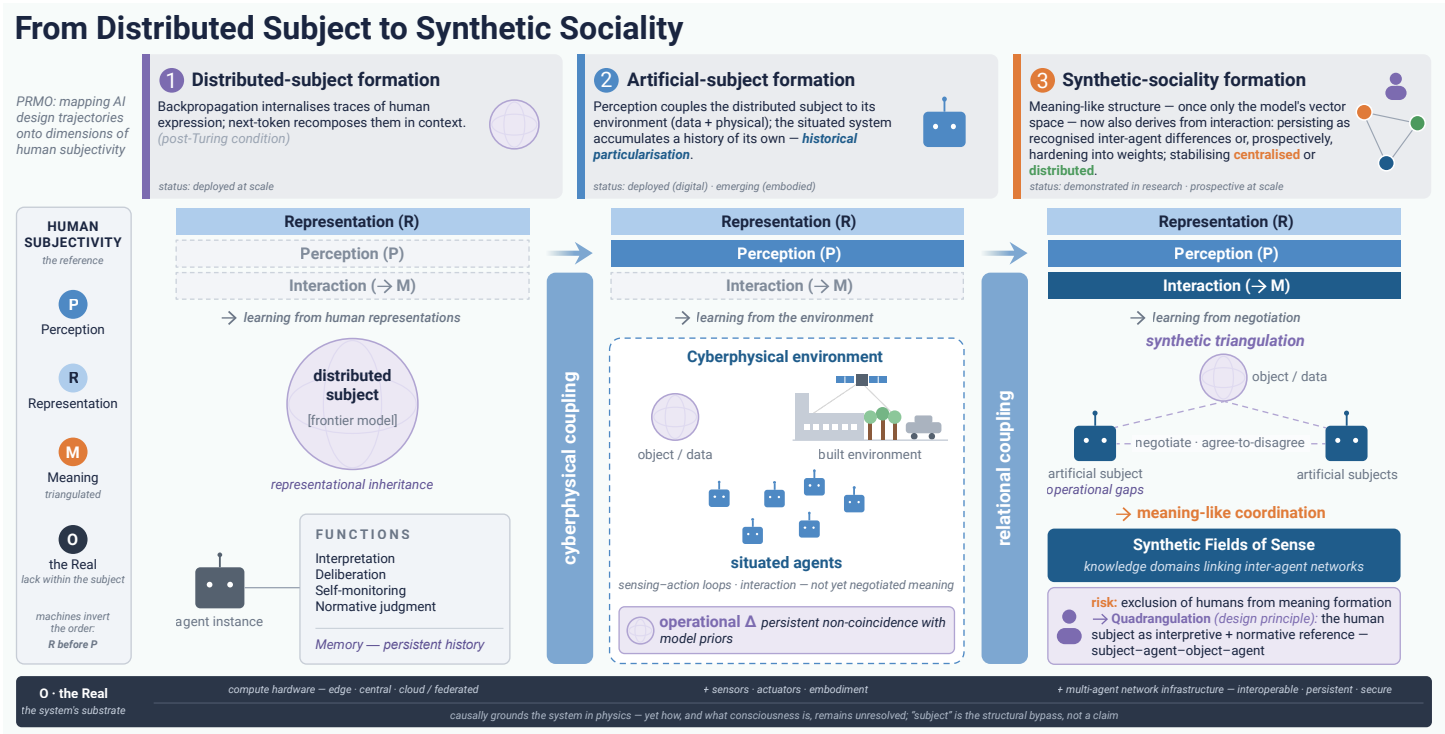

Figure 1. From distributed subject to synthetic sociality: the PRMO trajectory Four stages: (1) the frontier model as distributed subject (representational inheritance); (2) situated agents (sensorimotor coupling); (3) artificial subjects (historical particularisation); (4) synthetic sociality (negotiated meaning). Capability stacks are cumulative (R → R+P → R+P+I). The Real (O) denotes the system's physical substrate, which grounds the system causally but settles nothing about consciousness. In currently deployed agents the operational Δ resides in context and memory rather than model weights; weight-level consolidation remains a research frontier. Synthetic triangulation (agent–object–agent) risks excluding humans from meaning formation; Quadrangulation is proposed as a design principle retaining the human subject as interpretive and normative reference. Structural, not phenomenological: no claim about machine consciousness.

weights (the emerging research programme of self-evolving agents) remains a frontier, not a deployed capability.[8] One frontier model can therefore underwrite many agents that share a representational inheritance while developing distinct histories.

This threshold can be read as artificial subjectivity: a methodological position formed through representation, coupling, memory, and feedback. This is not a phenomenological claim. The agent does not experience lack; it encounters operational gaps between plan and outcome, prediction and resistance, and these gaps force revision and adaptation. Status: deployed in digital environments; emerging in physical ones.

Negotiated coordination: Synthetic sociality

The third step would begin when situated agents interact with one another beyond continuous human mediation. Most deployed "multi-agent" systems today are orchestration: a principal model decomposing tasks for subordinate ones. The condition we term synthetic sociality is different: synthetic triangulation, in which agents with distinct histories orient towards the same object and negotiate how to read it.

First demonstrations exist. In decentralised populations of LLM agents playing pairwise naming games, shared linguistic conventions emerge spontaneously without central coordination; collective biases arise at population level that no individual agent holds; and small committed minorities can tip established conventions.[9] Tipping thresholds vary strongly across models, evidence that what stabilises is a property of the population and its interaction history, not any single agent. In PRMO terms, such systems do not instantiate meaning; they perform meaning-like coordination, stabilising reference and aligning action across agents.

Synthetic sociality should not be read as consensus. It includes negotiated coherence, managed disagreement, and stabilised divergence: an agreement to disagree is itself a stabilised relation. Such divergence-tolerance may not come free. In early negotiation experiments, agents optimising narrow goal functions tend towards zero-sum framings, even when cooperation is costless; agreeing to disagree would need to be designed into the agent's value function rather than expected to emerge. Gabriel makes the equivalent point for humans: moral disagreement presupposes a wider agreement.[3] What is a moral posture for humans becomes a design requirement for artificial subjects: divergence remains intelligible only within a shared field of sense. The governance risk is not that agents agree, but that reference stabilises machine-only: coherent enough to act on, with no structural point at which a human can contest the reading. Status: demonstrated in research; prospective at deployment scale.

A CONCRETE SCENARIO: ONE DATA STREAM, DIVERGENT READINGS

Consider a shared Earth-observation stream. Three agents monitor the same satellite imagery of a river basin after heavy rainfall. An emergency-response agent reports flooding and urges evacuation; an agricultural agent, seasonal inundation within historical variance; an insurance agent, a claims-relevant anomaly. The object is identical; the readings diverge, because each agent approaches it through its own history, goals, and value function. Through interaction they detect interpretive deltas (differences in how the same object is approximated) and then negotiate an operationally coherent report, or escalate their divergence. A human official receives the result as "the situation."

This is a field of sense in practical form: meaning-like content is not located inside any single representation but stabilised between interpreters oriented towards a shared object, through interpretation, contestation, feedback, and revision. The governance question becomes concrete: was the human official inside that stabilisation, able to interrogate the divergence, or downstream of it?

QUADRANGULATION AS A HYBRID DESIGN PRINCIPLE

Quadrangulation proposes retaining the human subject as a fourth vertex of synthetic triangulation (subject–agent–object–agent): not an external human-in-the-loop checkpoint, but a learned functional dependency under which machine-mediated sensemaking remains incomplete without orientation towards human interpretation.

Two clarifications address obvious objections. First, quadrangulation is not purely subjective. Human reality construction is itself iterative and empirically

disciplined, and triangulation already binds subjects to an object that pushes back. Quadrangulation therefore combines two feedback channels: objective iteration (validation against data, environment, and outcomes) and human contestability (the standing possibility of human reinterpretation and dissent). In the scenario above, the negotiated flood report remains coupled both to ground truth and to the official's capacity to contest it. A system fails quadrangulation when it converges on internally stabilised, machine-only coherence that forecloses human revision.

Second, the principle can be placed against existing approaches by the role they give the human. Reinforcement learning from human feedback and alignment to explicit principles fix human reference in advance; human-in-the-loop oversight adds an external checkpoint; an AI mediator that helps human groups find common ground builds it in as consensus.[10] Each makes the human a control on machine behaviour. Quadrangulation differs in kind: where those approaches act on individual agents or outputs, it makes orientation to the human a learned, internalised condition of the meaning-forming relation itself, so that machine sensemaking stays incomplete without it. The human is not a check on the output but a condition of its validity. And where such principles must specify in advance which human values count, always context- and culture-bound, Quadrangulation leaves that content to the human in situ, internalising only the dependency on human contestability.

For engineers, the prior question is whether to pursue artificial subjectivity and synthetic sociality at all, and if so, under what conditions human contestability is structurally preserved. For governance, it complements external oversight and regulation with a design-level commitment: from controlling outputs to sustaining human participation in meaning.

CONCLUSION

The Post-Turing Condition suggests that the deepest near-term question about AI is not whether machines become conscious, but whether they become social without us. Read through PRMO, the trajectory is a maturity gradient rather than a prophecy: representation deployed at scale; sensorimotor coupling deployed digitally and emerging physically; negotiated multi-agent coordination, and prospective at scale. Each step is a design choice, not a destiny. Quadrangulation names the commitment that would keep the human subject constitutively present; so that as AI moves from processing symbols towards negotiating reality, the human field of sense expands rather than closes.

REFERENCES

1. Turing A.M. (1950). Computing machinery and intelligence. *Mind* **59**:433–460. DOI:10.1093/mind/LIX.236.433
2. Davidson D. (2001). *Subjective, Intersubjective, Objective*. Oxford University Press. ISBN: 9780198237532
3. Gabriel M. (2015). *Fields of Sense: A New Realist Ontology*. Edinburgh University Press. ISBN: 9780748692880
4. Vaswani A., Shazeer N., Parmar N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**:6000-6010. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
5. Mitchell M. and Krakauer D.C. (2023). The debate over understanding in AI's large language models. *Proc. Natl. Acad. Sci. USA* **120**:e2215907120. DOI:10.1073/pnas.2215907120
6. Gabriel M. (2026). *Ethische Intelligenz*. Ullstein. ISBN:978-3550204616
7. Žižek S. (2026). Why artificial intelligence is not a subject. *Substack*. https://slavoj.substack.com/p/why-artificial-intelligence-is-not
8. Gao H.-a., Geng J., Hua W., et al. (2026). A survey of self-evolving agents: What, when, how, and where to evolve on the path to artificial super intelligence. arXiv preprint 2507.21046. DOI:10.48550/arXiv.2507.21046
9. Ashery A.F., Aiello L.M. and Baronchelli A. (2025). Emergent social conventions and collective bias in LLM populations. *Sci. Adv.* **11**:eadu9368. DOI:10.1126/sciadv.adu9368
10. Tessler M.H., Bakker M.A., Jarrett D., et al. (2024). AI can help humans find common ground in democratic deliberation. *Science* **386**:eadq2852. DOI:10.1126/science.adq2852

FUNDING AND ACKNOWLEDGMENTS

This research is supported by the National Science and Technology Major Project (No. 2022ZD0117202), the International Partnership Program of the Chinese Academy of Sciences (No. 313GJHZ2022040BS), and the Chinese Academy of Sciences (CAS) President's International Fellowship Initiative (No. 2025PVA0071). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Thorsten Jelinek: Conceptualization, methodology, investigation, writing — original draft, and project administration. Patrick Glauner: Conceptualization, investigation, and writing — review and editing. Alvin Wang Graylin: Conceptualization, investigation, and writing — review and editing. Yubao Qiu: Conceptualization, supervision, and writing — review and editing. All authors contributed to the interpretation and development of the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

Not applicable.