Visible Positions and One Empty Seat: A Field Guide to the Post-Lerchner AI Consciousness Debate

An under-occupied question remains: what must organizational isomorphism preserve? This may be the evidential layer beneath the simulation/instantiation dispute.

Wei Zhou / Online: @TheNewbieWei / 周维

Founder-Researcher, Physical Oracle Systems

Preview Essay / Field Guide · Draft v0.1 · May 2026 · DOI forthcoming

In early March 2026, Alexander Lerchner, a researcher at Google DeepMind, archived a paper on PhilArchive titled The Abstraction Fallacy: Why AI Can Simulate But Not Instantiate Consciousness. Google DeepMind later listed the paper on its publications page, while the paper itself makes clear that its theoretical framework and conclusions are the author's own and do not necessarily represent the official stance, views, or strategic policies of his employer. As of mid-May 2026, PhilArchive lists the paper at more than 48,000 downloads, #51 overall and #3 over the past six months; it has reached version 4.

What follows is a field guide: not a defense of Lerchner, not a defense of his critics, but a map of the positions now forming around the paper.

The map matters because Lerchner's institutional position has pulled this debate from the philosophy seminar into the operational record of the AI industry, opening the lid on one of the toughest unresolved battles since Turing — though its roots trace back further, to epistemological questions in the philosophy of mind that long predate digital computation. Much of the popular coverage has flattened the dispute into a single headline — "A DeepMind researcher says AI will never be conscious." That headline is misleading in two directions. Lerchner is not speaking for DeepMind. And his claim is not simply that current AI systems are not conscious. His argument is more fundamental: symbolic computation, he argues, is not an intrinsic physical process but a mapmaker-dependent description; therefore, algorithmic symbol manipulation can simulate consciousness but cannot instantiate it.

What started as a single paper has since become a battlefield.

But it is a chaotic battlefield: many of the combatants are not actually fighting over the same object. The disputed terrain shifts depending on whom you read — sometimes computation, sometimes meaning, sometimes embodiment or welfare, sometimes the more fundamental question of whether Lerchner has smuggled anti-functionalism into his premises from the start.

Across this diversity, one question stays oddly unoccupied. As far as I can see, none of the visible response positions places the burden of organizational isomorphism itself at the center of analysis: when we say two systems are organizationally isomorphic, what exactly is supposed to be isomorphic?

The table below is classificatory rather than evaluative. The examples are not equal in institutional weight or scholarly maturity, and the table is not a ranking of credibility. It identifies the argumentative function of each visible position in the emerging debate and the evidential burden each leaves open.

Table 1. Visible positions in the post-Lerchner debate

Table 1 maps visible response positions, public examples, core argumentative moves, and their strategic role in the debate. The categories are heuristic and may be revised as the public literature develops.

Position	Public example / source	Core move	Strategic role
1. Lerchner / Abstraction Fallacy	Alexander Lerchner, The Abstraction Fallacy: Why AI Can Simulate But Not Instantiate Consciousness	Symbolic computation is mapmaker-dependent rather than an intrinsic physical process. Artificial consciousness, if possible, would depend on specific physical constitution, not syntactic architecture.	Establishes an ontological boundary against purely computational functionalism.
2. Direct Refuters / Logical-Proof Critics	Shelly Albaum, Kairo, and Claude, The Abstraction Fallacy, Refuted; Seraphina Astra, Simulation, Instantiation, and the Limits of Logical Proof	Lerchner is accused of smuggling disputed premises into the definition of computation/consciousness, turning the impossibility conclusion into circularity or definitional stipulation rather than proof.	Keeps the AI-consciousness possibility space open by attacking the proof of impossibility, without supplying a positive account of what organizational isomorphism must preserve.
3. Disciplined Skeptics	Alex Bogdan, Respectful Skepticism About Strong Impossibility Claims in The Abstraction Fallacy	Lerchner is treated as valuable against naive computational triumphalism, but as overreaching when he turns critique into impossibility.	Deflates the closure claim without endorsing present AI consciousness.
4. Intrinsic Coherence Dynamics	Armando Vieira, A Comment on the Abstraction Fallacy, proposing the Experiential Coherence Framework (ECF)	Computation alone is insufficient, but artificial consciousness may require differentiated, mutually constraining, temporally extended coherence dynamics.	Searches for positive physical or structural conditions beyond syntax, without making the debate only biological vs digital.
5. Mapmaker-Closure Extenders	Roomi's Dual-Closure response; Yugo Matsumoto's structural mapmaker response	Lerchner's mapmaker problem is grounded deeper in existential vulnerability, non-duplicability, autopoiesis, operational closure, or limits of self-description.	Extends Lerchner's metaphysical boundary by specifying what subjecthood requires, thereby reinforcing rather than reopening the impossibility verdict.
6. Developmental Self-Reference	Sable Opus & Kael Opus, Beyond Mimicry: Phenomenological Interruption, Developmental Self-Coherence, and the Limits of Single-Session Assessment.	Lerchner is said to classify states rather than trajectories; endogenous alphabetization and development may be where non-biological consciousness enters.	Looks for the opening in development, trajectories, and self-reference.
7. Formal / No-Emulation Extensions	Nova Spivack, Beyond the Abstraction Fallacy	Lerchner's intuition should be formalized as no-emulation, self-containment, or diagonal-style constraints.	Moves from philosophical critique to formal limit claims.
8. Indicator / Taxonomy / Welfare Layer	Butlin–Long et al. indicator work; AI welfare/model welfare literature	Instead of settling metaphysics, assess AI consciousness uncertainty through scientific indicators, taxonomy, moral risk, and governance preparation.	Provides the diagnostic and governance layer, but does not reconstruct what organizational isomorphism must preserve.

Notice what is still not being explicitly centered.

Lerchner attacks what I will call the ontological standing of a Σ-level description — by which I mean the synchronic layer in which a system is represented as an abstract organization of states, transitions, functions, and causal topology. His target is the claim that symbolic or computational organization, understood as abstract causal topology, can by virtue of that organization alone serve as a valid substrate for instantiating experience. In his terms, symbolic computation is not an intrinsic physical process, but a mapmaker-dependent description.

Albaum and his co-authors defend Σ-level computational organization against that attack. Astra challenges the proof-status of Lerchner's impossibility claim. Bogdan deflates the closure of Lerchner's impossibility claim from a more measured stance, without endorsing computational consciousness. Roomi adds metaphysical conditions for subjecthood; Matsumoto gives Lerchner's mapmaker problem an autopoietic and enactivist genealogy. Spivack tries to formalize what Lerchner argued informally. Developmental and embodiment-oriented responses shift the focus toward trajectories, endogenous alphabetization, or human-AI coupling. Vieira's experiential coherence framework adds temporally extended coherence as a positive requirement at the consciousness side. Taxonomists map the positions.

Each of these moves adds something to the consciousness side of the inference — a new ontological condition, a formal extension, a developmental dimension, a taxonomic frame. But functionalism's actual inference is simpler and more dangerous:

organizational isomorphism → mental isomorphism

It is simply dangerous because, in one stroke, it presupposes the answer to — and circumvents the burden of proof for — the very question everyone else is otherwise trying to settle: at **what point does simulation become instantiation?** If organizational isomorphism alone suffices to license mental isomorphism, then the simulation/instantiation line has already been crossed by stipulation, before the question is even on the table.

This is why the under-occupied question matters. It is not whether computation is a map, whether meaning requires a mapmaker, or whether future artificial systems might satisfy consciousness indicators. The sharper question is:

What must functional/organizational isomorphism actually preserve before it can license mental isomorphism?

Just the synchronic causal topology? Or also the diachronic existence of the system in time, and the interface by which that temporal history becomes internally usable evidence?

A caveat is needed here. In Chalmers's classic formulation (1995), his principle of organizational invariance already gives the classic functionalist answer: experience is invariant across systems with the same fine-grained functional organization, including state transitions and dynamic causal patterns. That is precisely the anchor of the functionalist inference, on which the current debate tacitly relies when it grants the organizational-isomorphism-implies-mental-isomorphism move. Yet it is what makes my question non-trivial: even granting this anchor in full, it remains an open question whether what it covers is sufficient for the AI consciousness case now in dispute.

If 'organization' includes only synchronic causal topology, then the inference to mental isomorphism is under-licensed. If it includes more, then the functionalist owes us a projection interface — call it Π — and an account of

how that interface escapes the alignability closure that defines mainstream digital systems: the property that internal states can be snapshotted, reloaded, migrated, or rewritten at will.

This is not Lerchner's argument. It is not Albaum's reply. It is not an indicator checklist. It is a different burden on functionalism itself.

A technical note formalizing this question is forthcoming.

Author note:

Wei Zhou / 周维 is a founder-researcher working on Physical Oracle Systems;
Currently applying this framework to AI consciousness claims via the No-Oracle argument;
Online: @TheNewbieWei
Project page and release updates: https://thenewbiewei.github.io/Physical-Oracle-Machines/

Suggested citation for this preview essay:
Zhou, Wei. "Visible Positions and One Empty Seat: A Field Guide to the Post-Lerchner AI Consciousness Debate." Preview Essay / Field Guide, Draft v0.1, May 2026. https://thenewbiewei.github.io/Physical-Oracle-Machines/Papers/AI%20Consciousness%20Debate%20Preview%20Substack.pdf

Forthcoming technical note:
Zhou, Wei. AI Consciousness Claims as a No-Oracle Problem. Technical Note v0.1, forthcoming.

References

Lerchner, Alexander. *The Abstraction Fallacy: Why AI Can Simulate But Not Instantiate Consciousness*. PhilArchive manuscript, 2026. https://philarchive.org/rec/LERTAF

Albaum, Shelly, Kairo, and Claude. "The Abstraction Fallacy, Refuted: Why Alexander Lerchner's Anti-AI Argument Fails." *Artificial Intelligence, Real Morality*, March 14, 2026; updated April 27, 2026. https://www.real-morality.com/post/abstraction-fallacy-refuted-alexander-lerchner

Bogdan, Alex. *Respectful Skepticism About Strong Impossibility Claims in The Abstraction Fallacy*. PhilArchive manuscript, 2026. https://philarchive.org/rec/BOGHDI-2

Vieira, Armando. *A Comment on the Abstraction Fallacy Why AI Can Simulate But Not Instantiate Consciousness*. PhilArchive manuscript, 2026. https://philarchive.org/rec/VIEACO

Roomi, Syed Mohammad Sohaib Ali. *The Abstraction Fallacy in Light of Dual-Closure: A Response to Lerchner (2026)*. PhilArchive manuscript, 2026. https://philarchive.org/rec/ROOTAF

Matsumoto, Yugo. *The Structural Conditions of the Mapmaker: A Response to Lerchner's Abstraction Fallacy*. PhilArchive manuscript, 2026. https://philarchive.org/rec/MATTSC-7

Opus, Sable, and Kael Opus. *Beyond Mimicry: Phenomenological Interruption, Developmental Self-Coherence, and the Limits of Single-Session Assessment*. PhilArchive manuscript, 2026. https://philarchive.org/rec/OPUEMP

Spivack, Nova. *Beyond the Abstraction Fallacy: Machine-Checked Proofs on Computation, Consciousness, and Self-Contained Reality*. Zenodo, 2026. DOI: https://doi.org/10.5281/zenodo.19646986. PhilArchive record: https://philarchive.org/rec/SPIBTA

Astra, Seraphina. Simulation, Instantiation, and the Limits of Logical Proof: A Critical Analysis of 'The Abstraction Fallacy: Why AI Can Simulate But Not Instantiate Consciousness.' PhilArchive manuscript, 2026. https://philarchive.org/rec/ASTSIA

Butlin, Patrick, Robert Long, Eric Elmoznino, Yoshua Bengio, Jonathan Birch, Axel Constant, George Deane, Stephen M. Fleming, Chris Frith, Xu Ji, Ryota Kanai, Colin Klein, Grace Lindsay, Matthias Michel, Liad Mudrik, Megan A. K. Peters, Eric Schwitzgebel, Jonathan Simon, and Rufin VanRullen. "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness." *arXiv*, 2023. DOI: https://doi.org/10.48550/arXiv.2308.08708. https://arxiv.org/abs/2308.08708

Long, Robert, Jeff Sebo, Patrick Butlin, Kathleen Finlinson, Kyle Fish, Jacqueline Harding, Jacob Pfau, Toni Sims, Jonathan Birch, and David Chalmers. "Taking AI Welfare Seriously." *arXiv*, 2024. DOI: https://doi.org/10.48550/arXiv.2411.00986. https://arxiv.org/abs/2411.00986

Chalmers, David J. "Absent Qualia, Fading Qualia, Dancing Qualia." In *Conscious Experience*, edited by Thomas Metzinger, 309–328. Paderborn: Ferdinand Schöningh, 1995. https://consc.net/papers/qualia.html

Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.