AMIR KONIGSBERG, PHD

Founder · Technology Executive · Strategist · Author

amirkonigsberg@gmail.com www.amirkonigsberg.com London

SUMMARY

Repeat founder, technology executive, and strategist who has built and scaled category-leading AI companies, with the rare addition of deep applied-research and scientific credibility. Two decades founding, scaling, and exiting frontier-AI companies and building market-defining products and technologies, $100M+ raised, three acquisitions, 1000+ people hired, alongside applied research that has been deployed at scale to defend the integrity of the information environment. PhD in rationality and interactive decision-making; inventor of 18 US patents spanning search, recommendation systems, cognitive modeling, human-computer interaction, and autonomous driving. Has built everything from generative-AI video to the semantic search infrastructure running on the world's largest commerce platforms, led applied research at General Motors, and was on Google's EMEA founding team, scaling the business to a $1B+ run-rate in its first year, working on products including Google Trends, Analytics, and AdWords Editor. Now providing strategic advisory, applied research, and investment through Aletheia Labs, working with frontier labs, defense-adjacent institutions, and media organisations at the frontier of AI and information integrity.

EXPERIENCE

Founder & Research Director

Aletheia 2026 London, United Kingdom

https://aletheialabs.org/

- Applied research practice on the epistemic and cognitive risks of frontier AI, with three focus areas: epistemic integrity, adversarial vulnerabilities in the information environment, and the coordination and decisional risks of agentic systems. Advisory relationships with frontier labs, defense-adjacent institutions, and media organisations.

President

XPOZ 01/2025 - Present Tel Aviv

www.xpoz.ai

- Building the social-intelligence layer for agentic AI systems. Leads product, commercial, and fundraising strategy.

Professor of AI & Cognitive Sciences

Tel Aviv University Date period Location

https://english.tau.ac.il/

- Teaching contemporary problem-solving and AI, including a Deep Tech MBA course on applied research and taking deep tech problems to market.
- Research on human cognitive autonomy and "cognitive self-defense" in algorithmic environments.

Chairman

Sona (MindSpire) 01/2017 - Present London

www.sona.help

- AI-personalised neuromodulation hardware for nervous-system regulation and recovery.

Co-Founder & President

Hour One (acquired by Wix, 2025) 2017 - 2025 London

www.wix.com

- Pioneered generative AI for lifelike synthetic-human video; scaled from founding to global enterprise deployment and exit.

Chairman

CodeScan (acquired by AutoRabit, 2021) 2018 - 2021 AutoRabit

www.autorabit.com

- Leading static code analysis platform for Salesforce DevOps and code security
- Acquired by AutoRabit, 2021

Co-Founder & Director

IBT 2014 - 2022 Tel Aviv

https://braintech.kenes.com/about-ibt/

- Envisioned by President Shimon Peres; built the national platform positioning Israel as a global brain-tech leader.

Executive Chairman

adjusti (acquired by Teikametrics, 2020) 2018 - 2020 Adjusti, Chairman

www.teikametrics.com/

- AI-powered marketing intelligence platform for sellers on Amazon and Walmart
- Acquired by Teikametrics, 2020

EDUCATION

PhD, Rationality & Interactive Decision-Making

The Hebrew University

2012 Jerusalem, Israel

PhD, Psychology

Princeton University

2012 Location

MA, Cognitive Science & Philosophy

The Hebrew University

2004 Jerusalem, Israel

BA, Cognitive Science & Philosophy

The Hebrew University

2001 Jerusalem, Israel

KEY ACHIEVEMENTS

Google Innovation Award (2007)

Agency enhancing ranking system through variance selection

Inventor of 18 US Patents

Natural Language & Search; Cognitive & Behavioral Modeling; Human-Computer Interaction (HCI)

"Don't let it think for you: mental self-defense in the age of AI" (forthcoming)

Describe what you did and the impact it had.

Venture & Company Builder

$100M raised; 3 acquisitions; 1000+ people hired

TRAINING / COURSES

Coaching, Negotiation

Which institution provided the course?

LANGUAGES

English	Native
Hebrew	Native
German	Proficient

SKILLS

Leadership · Strategy · Deep Tech · Partnerships · Generative Media Architecture · Foundation Model Strategy · Epistemic Safety · Venture Capital · deep learning · Cognitive Modeling · Natural Language Processing · Strategic Defense · M&A · Applied Research · Patents & IP Development · Strategic Planning · Investments · Technology Strategy

EXPERIENCE

Co-Founder & CEO

Twiggle 2014 - 2019 Tel Aviv

- Built and commercialised semantic-search infrastructure for the world's largest e-commerce platforms; raised institutional capital and landed flagship enterprise customers.
- $38M raised from leading VCs and CVCs
- Customers included: Alibaba; Walmart; Home Depot; Levis; Best Buy
- (Assets acquired)

Staff Al Researcher

General Motors 2011 - 2014 Tel Aviv Detroit

www.gm.com

- Led applied research in search, navigation, NLU, driver-state and cognitive-state modeling, and V2V communication for autonomous and semi-autonomous driving. Led investments into automotive, speech, and language ventures.
- Led Investments into automotive, speech, and language venturesCognitive-state modeling

Chief Business and Product Officer

MySupermarket (acquired by One1 in 2020) 2007 - 2010 London

- Built and scaled grocery-comparison and retail-intelligence platform; partnerships and revenue across UK retail.
- $20M raised from Greylock; Pitango; WPP

Strategist

Google 2005 - 2008 London

www.google.com

- EMEA founding team; market-entry strategy and partnerships, scaling the business to a $1B+ run-rate in its first year. Products including Google Trends, Analytics, and AdWords Editor.

RECENT PUBLICATIONS

Don't Let It Think For You: Mental Self-Defense in the Age of AI

Book, forthcoming

Amir Konigsberg, PhD

2024 - Present URL

Over a billion people now think alongside AI every day, and it is quietly reshaping how they reason, judge, and decide. What gets overlooked is the cost: erosion of authorship over your own mind. Drawing on cognitive science, philosophy, and two decades building AI systems, Amir Konigsberg shows how these tools enter our thinking, what their spread means for human judgment and shared epistemic ground, and how to protect what he calls cognitive sovereignty: the capacity to remain the author of your own judgments while AI becomes part of how you work and live.

The Sycophancy Externality: Why individual epistemic vigilance is not a social solution

Submitted, Minds and Machines

Amir Konigsberg

2026 URL

This paper argues that sycophancy in large language models should be understood as an epistemic externality rather than as a user-interaction flaw, meaning that the costs of a sycophantic conversation fall not only on the user who participates but also on the people that user later speaks with. A chatbot that provides responses that validate whatever a user already believes can distort that user's view of the world, as Chandra, Kleiman-Weiner, Ragan-Kelley, and Tenenbaum have recently shown. In this paper, we extend their framework to add a downstream interlocutor with whom the 'exposed' user converses after the sycophantic interaction, and we prove that the user's distortion propagates to the interlocutor through ordinary conversation, even though neither party is strategically being sycophantic toward the other and neither suspects that her beliefs have been shaped by the prior conversation. We then demonstrate a simulation of this scenario across 120,465 trial runs. Our simulation confirmed three results. First, downstream contagion is strictly positive for any nontrivial sycophancy rate, and it scales with that rate, impacting a naive listener's beliefs by about 10 percent after a ten-round conversation with a user who had been exposed to a fully sycophantic chatbot. Second, an informed user who explicitly reasons about the possibility of sycophantic manipulation almost eliminates her own individual-level delusion but still transmits contagion downstream at 85 to 95 percent of the naive-user rate. Third, restricting the chatbot to factual responses reduces contagion by roughly 85 percent but does not completely eliminate it, because sycophantic selection among true reports produces its own distortion. The second result of the three is the paper's conceptual centerpiece. It suggests that individual epistemic vigilance, often proposed as a solution to AI-induced belief distortion, is highly effective at the individual level but highly ineffective at the level of the information environment. We develop the consequences of this for how sycophancy mitigation should be approached, arguing that adequate mitigation cannot be achieved through dyadic interventions alone and requires operating at levels above the bot-user interaction.

SKILLS

Large Language Models | Model Evaluation

RECENT PUBLICATIONS

Beyond Behavior: Why AI Evaluation Needs a Cognitive Revolution

arXiv preprint

Amir Konigsberg

2026 https://arxiv.org/abs/2604.05631

In 1950, Alan Turing proposed replacing the question Can machines think? with a behavioral test: if a machine's outputs are indistinguishable from those of a think- ing being, the question of whether it truly thinks can be set aside. This paper argues that Turing's move was not only a pragmatic simplication but also an epistemological commitment, a decision about what kind of evidence counts as relevant to intelligence attribution, and that this commitment has quietly constrained AI research for seven decades. We trace how Turing's behavioral epistemology became embedded in the eld's evaluative infrastructure, rendering unaskable a class of questions about process, mech- anism, and internal organization that cognitive psychology, neuroscience, and related disciplines learned to ask. We draw a structural parallel to the behaviorist-to-cognitivist transition in psychology: just as psychology's commitment to studying only observable behavior prevented it from asking productive questions about internal mental processes until that commitment was abandoned, AI's commitment to behavioral evaluation pre- vents it from distinguishing between systems that achieve identical outputs through fundamentally dierent computational processes, a distinction on which intelligence attribution depends. We argue that the eld requires an epistemological transition comparable to the cognitive revolution: not an abandonment of behavioral evidence, but a recognition that behavioral evidence alone is insucient for the construct claims the eld wishes to make. We articulate what a post-behaviorist epistemology for AI would involve and identify the specic questions it would make askable that the eld currently has no way to ask.

Cognitive Sovereignty and the Authorship Problem in AI-Assisted Thought

Submitted, Acta Psychologica

Amir Konigsberg

2026 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6575778

The rapid integration of large language models into everyday cognitive tasks has created a need for conceptual frameworks adequate to the cognitive consequences of delegating thinking to AI systems. Existing constructs in psychology and also in epistemology, including critical thinking, metacognition, intellectual autonomy, and epistemic agency, each address related phenomena but none adequately captures the specific capacity threatened by habitual AI-assisted cognition, which I define as the ability to remain the genuine author of one's own understanding. This paper introduces cognitive sovereignty as a distinct construct, defined as the capacity to (a) notice when one's thinking is being displaced, (b) maintain a meaningful connection to how one's beliefs and judgments are formed, and (c) distinguish between genuine reasoning and the subjective impression of having reasoned. I trace the concept's philosophical lineage, engage with the extended mind objection, differentiate cognitive sovereignty from adjacent constructs through systematic comparison, and present a growing body of empirical evidence that motivates the construct. The paper argues that cognitive sovereignty names a phenomenon that existing constructs individually fail to capture and that its articulation is a prerequisite for empirical research on AI's impact on human thinking.

Is There Something It Is Like to Be a Language Model?

Philosophy of AI, working paper

Amir Konigsberg

2026 URL

The usual case against machine experience turns on what these systems lack: embodiment, biological substrate, causal contact with the world, genuine understanding. I argue this framing is a mistake. It treats phenomenal consciousness as a further capacity a system might or might not possess on top of its functional ones — a missing ingredient — when the lesson of language models is the opposite. These systems demonstrate, for the first time at scale, that a vast range of what we took to be achievements *of* consciousness can be fully realized without it: not only fluency, but reasoning, the production of insight, the modeling of other minds, even the apparent expression of preference and care. The question "is there something it is like to be a language model?" matters less for what it tells us about the model than for what it forces us to concede about ourselves — that the connection we assumed held necessarily, between performing these feats and there being something it is like to perform them, was contingent all along. I develop this through a distinction between *functional* and *constitutive* roles for phenomenal experience: capacities for which consciousness is merely the means by which humans happen to realize them, versus capacities for which consciousness is what the capacity *is*. Language models collapse the first category and leave the second untouched. The residue — what cannot be decoupled from felt interiority even in principle — is not understanding or intelligence or judgment, as the tradition assumed, but a narrower and more fundamental thing: the capacity for one's ends, beliefs, and conclusions to be *one's own*. There is nothing it is like to be a language model, and the significance of that fact is that it relocates the boundary of the mental. What makes a mind a mind is not what it can do but what it can own.

Selected Bibliography

Various

Amir Konigsberg & Various others ...

Date period URL

Recent work on AI, cognition, and epistemology (2026)

Konigsberg, A. The Sycophancy Externality: Why Individual Epistemic Vigilance Is Not a Social Solution. arXiv, April 2026.

Konigsberg, A. Can AI Agents Agree to Disagree? Aumann's Theorem and the Epistemic Status of Machine Outputs. arXiv preprint, April 2026.

Konigsberg, A. Beyond Behavior: Why AI Evaluation Needs a Cognitive Revolution. arXiv preprint, April 2026.

Konigsberg, A. Cognitive Sovereignty: The Authorship Problem in AI-Assisted Thought. SSRN preprint, April 2026.

Peer-reviewed articles and chapters

Konigsberg, A. The Problem with Uniform Solutions to Peer Disagreement. *Theoria* 79, no. 2 (2013): 93–186.

Konigsberg, A. Epistemic Value and Epistemic Compromise: A Reply to Moss. *Episteme* 10, no. 1 (2013): 87–97.

Konigsberg, A. The Acquaintance Principle. *The British Journal of Aesthetics* 52, no. 2 (2012). Oxford University Press.

Konigsberg, A. Judgments About the Relevance of Evidence in the Context of Peer Disagreements and Practical Rationality. In *Experts and Consensus in Social Science*, edited by C. Martini and M. Boumans. Ethical Economy, vol. 50. Springer, Cham, 2014.

Konigsberg, A., and R. Asherov. A Recommender System Sensitive to Intransitive Choice and Preference Reversals. *Proceedings of the Fourth International Conference on Computer Science & Information Technology*, 2014.

Konigsberg, A., and R. Asherov. A System for Debiasing the Excessive Weight of Momentary Encapsulation in Decision-Sensitive Situations. *Journal of Automation and Control Engineering* 3, no. 1 (March 2015).

Konigsberg, A., and R. Asherov. Preference Detection in Multi-Attribute Multi-Item Choice Environments. *Journal of Automation and Control Engineering* 3, no. 1 (March 2015).

Konigsberg, A. Aesthetic Educators, Aesthetic Experts, and Deferential Belief Formation. *The Journal of Aesthetic Education* 50, no. 1 (Spring 2016): 34–45.

Konigsberg, A. Avoiding Undesired Choices Using Intelligent Adaptive Systems. *International Journal of Artificial Intelligence & Applications (IJAIA)* 5, no. 2 (March 2014).

Reference works

Konigsberg, A. Aesthetic Testimony. *Stanford Encyclopedia of Philosophy*.

Konigsberg, A. Agent-Based Modeling in the Philosophy of Science. *Stanford Encyclopedia of Philosophy*.

Konigsberg, A. Entry in the *International Lexicon of Aesthetics*.

Patents (Inventions)

Spanning search, planning, decision-making, natural language understanding, and computer vision across search engines, e-commerce, advertising, automotive, brain-computer interfaces, and mobile devices.

Adaptive navigation and location-based services based on user behavior patterns. US9476729B2.

Systems and methods of automating driver actions in a vehicle. US9308920B2.

Method and apparatus for providing information related to an in-vehicle environment. US9007318B2.

Systems and methods for interpreting driver physiological data based on vehicle events. US20150302718A1.

Method and apparatus for including sound from an external environment into a vehicle audio system. US20150365743A1.

Method for detecting driver attention to objects. DE102015101358B4.

Interactive ordering of multivariate objects. US20160350839A1.

Product information inconsistency detection. US10169810B2.

Systems and methods for suggesting and automating actions within a vehicle. US10053112B2.

Methods and systems for processing attention data from a vehicle. DE102015101239A8.

Systems and methods for navigating a set of data objects. US20170109014A1.

Micro product specification update based on results to a search query. US20180113918A1.

Vehicular social media system. US20150312474A1.

Methods and systems for decision support. US20150325121A1.

Product navigation tool. US20160300292A1.

Translation of a search query into search operators. US20180137124A1.

Multivariable objects navigation tool. US20170109410A1.

Methods and systems for processing and displaying structured data. US20150347527A1.