

HENRY SHEVLIN

Philosopher & AI Ethicist

Philosopher, Google DeepMind · Associate Director, Leverhulme Centre for the Future of Intelligence

✉ hfshevlín@gmail.com · www.polytropolis.com · www.henryshevlin.com

PROFILE

Philosopher of cognitive science and AI ethicist working on consciousness, machine minds, moral patiency, human-AI relationships, and AI and human flourishing.

- Philosopher at Google DeepMind, working on machine minds, human-AI relationships, and AGI & human flourishing, and Associate Director (Education) at the Leverhulme Centre for the Future of Intelligence, University of Cambridge.
- Co-founder and academic lead of the UK's first master's programmes in AI ethics; author of 20+ peer-reviewed publications.
- Monograph on the ethics of social AI under contract with Cambridge University Press; trade book in development.
- Regular media commentator (BBC, Sky News, France 24, *The Times*) and international keynote speaker. Co-host of two major podcasts, large (40k+) following across X/Twitter and Substack.

EMPLOYMENT

Philosopher , Google DeepMind	May 2026 –
Associate Director (Education), LCFI, University of Cambridge	May 2023 –
Senior Researcher , LCFI, University of Cambridge	Jan 2021 –
Programme Director , Kinds of Intelligence, LCFI, University of Cambridge	April 2023 – June 2026
Course Co-Lead , AI Ethics & Society MSt, LCFI, University of Cambridge	Jan 2021 – April 2023
Researcher , LCFI, University of Cambridge	May 2017 – Jan 2021
Adjunct / Adjunct Assistant Professor , Baruch College, CUNY	Sept 2011 – May 2017

EDUCATION

Ph.D, Philosophy, CUNY Graduate Center (Honors)	Sept 2016
Thesis: "Consciousness, Perception, and Short-Term Memory." Supervisor: Peter Godfrey-Smith. Committee: Ned Block, Eric Mandelbaum, David Papineau, Jesse Prinz.	
BPhil, Philosophy, Trinity College, University of Oxford	June 2009
Thesis: Linking Phenomenal and Access Consciousness. Supervisor: Tim Bayne.	
BA, Classics, St Anne's College, University of Oxford (Double First)	June 2006

AREAS OF SPECIALIZATION & COMPETENCE

Specialization: AI Ethics ♦ Human Flourishing ♦ Mind & Cognitive Science ♦ Comparative Cognition
Competence: Moral & Political Philosophy ♦ Philosophy of Science ♦ Epistemology ♦ Metaphysics

BOOKS

- *Ethics of Social AI*. Monograph under contract, Cambridge University Press Elements (due December 2026).
- Trade book on AI consciousness in development with a Big 5 publisher.

SELECTED ACADEMIC PUBLICATIONS

1. Shevlin (forthcoming). "Aeroplanes also fly: analytic functionalism and the possibility of machine consciousness – a reply to Anil Seth." *Behavioral & Brain Sciences*.
2. Shevlin (2026). "Three Frameworks for AI Mentality." *Frontiers in Psychology*.
3. Axelsson & Shevlin (2026). "Disambiguating Anthropomorphism and Anthropomimesis in Human-Robot Interaction." HRI 2026 (short paper).
4. Shevlin (2026). "The Anthropomimetic Turn in Contemporary AI." *The Artificial Intelligence Revolution* (eds. Fairfax & Rangeley).
5. Shevlin (2026). "Agency & Autonomy in Artificial Systems." *Quaestio*.
6. Shevlin (2024). "Ethics at the Frontier of Human-AI Relationships." *Oxford Handbook of Generative AI*, OUP.
7. Shevlin (2024). "All Too Human? Identifying and Mitigating Ethical Risks of Social AI." *Law, Ethics & Technology*.
8. Shevlin (2024). "Imagination, Creativity, and Non-human Animals." *Oxford Handbook of Philosophy of Imagination and Creativity*, OUP.
9. Shevlin (2024). "Consciousness, Machines, and Moral Status." *Humans and Smart Machines as Partners in Thought* (ed. Strasser).
10. Srivastava et al., incl. Shevlin (2024). "Beyond the Imitation Game (BIG-bench)." *Transactions on Machine Learning Research*.
11. Vervoort, Shevlin, Melnikov & Alodjants (2022). "Deep Learning Applied to Scientific Discovery." *Journal for General Philosophy of Science*.
12. Shevlin (2021). "Rethinking Creative Intelligence." *European Journal for Philosophy of Science*.
13. Shevlin (2021). "General Intelligence: an Ecumenical Heuristic for Artificial Consciousness Research?" *Journal of Artificial Intelligence & Consciousness*.
14. Shevlin (2021). "Non-human Consciousness and the Specificity Problem." *Mind & Language*.
15. Shevlin (2021). "How Could We Know When a Robot Was a Moral Patient?" *Cambridge Quarterly of Healthcare Ethics*.
16. Shevlin (2020). "Which Animals Matter?" *Philosophical Topics*.
17. Shevlin (2020). "Current Controversies in the Cognitive Science of Short-Term Memory." *Current Opinion in Behavioral Sciences*.
18. Crosby & Shevlin (2020). "Defining Artificial Intelligence: a Reply to Wang." *Journal of Artificial General Intelligence*.
19. Shevlin & Friesen (2020). "Pain, Placebo, and Cognitive Penetration." *Mind & Language*.
20. Shevlin, Vold, Crosby & Halina (2019). "The Limits of Machine Intelligence." *EMBO Reports*.
21. Shevlin & Halina (2019). "Apply Rich Psychological Terms in AI with Care." *Nature Machine Intelligence*, 1(4), 165–167.
22. Shevlin (2017). "Conceptual Short-Term Memory: A Missing Part of the Mind?" *Journal of Consciousness Studies*, 24(7–8).

SELECTED PUBLIC WRITING

- Polytopolis (Substack): essays on machine consciousness, human-AI relationships, and games, including "Behaviourism's Revenge" and "The House Elf Problem."
- "What if Artificial Intelligence Saves the Planet?" *The New European* (cover article).
- Essays and criticism in *Aeon*, *Daily Nous*, *iai.tv*, and *Nature Machine Intelligence*.

MEDIA, PODCASTS & PUBLIC PLATFORM

- Co-host, *Conspicuous Cognition* (with Dan Williams) and *Exploring Machine Consciousness* (PRISM) podcasts. Frequent guest on across major science and philosophy podcasts.
- BBC Radio 4 *Today* programme (with Anil Seth, 2026); BBC News, Sky News, France 24, and GB News appearances; commentary for CNN, *The Times*, and *Sunday Times*.
- Direct public following of c. 40,000: ~34k on X/Twitter (@dioscuri) and ~6k subscribers to the *Polytopolis* Substack.

SELECTED INVITED TALKS & KEYNOTES (2023–26)

- Panel on AI, Youth & Relationships, **Rome Conference on AI Ethics & Governance**, Pontifical Academy of Sciences, Vatican (June 2026).
- Opening keynote, **Models of Consciousness 5**, University of Bamberg (Oct 2024).
- Invited talk, **Princeton University Center for Human Values**, “Prospects & Pitfalls for a Science of Artificial Consciousness” (Oct 2023).
- Keynote, **Seoul AI Policy Convention**, Seoul National University (Aug 2025).
- Keynote, **AstraZeneca AI Day** (May 2026).
- Forthcoming keynote, **New Directions in Social Media Research**, St John’s College, Cambridge (Dec 2026).
- Further invited talks and panels include The Next Turing Tests (King’s College Cambridge), the Eleos and Google workshops on AI consciousness and welfare (Berkeley and New York), the Institute of Philosophy (London), the Cambridge Union, and the Finacle Conclave banking keynote (Istanbul).

GRANTS, PRIZES & FELLOWSHIPS

- Berggruen Institute Essay Prize, shortlist (2025; 25 shortlisted from 3,000+ entrants).
- Personal gift of \$50,000 from Google for “LLMs, Empathy & Mentalising” (2022).
- Future of Life Institute Worldbuilding Contest, 2nd-place team (\$10,000) (2022).
- MSt in AI Ethics & Society awarded ‘Best Course in AI’, CogX Awards (2022).
- President’s Excellence Award for Distinguished Teaching, Baruch College (2017).
- Joint winner, Oxford Future of Humanity Institute Thesis Prize (2015); Enhanced Chancellor’s Fellowship & Mellon Scholar, CUNY (2010–15).

ADMINISTRATION & LEADERSHIP

- Co-founder and academic lead of the UK’s first AI ethics master’s programmes (MSt in AI Ethics & Society; MPhil in Ethics of AI, Data & Algorithms); oversight of 100+ graduate students.
- Contributor to the AGI & human flourishing programme at Google DeepMind, on the long-run social and ethical questions raised by advanced AI.
- Unit Editor (2024–26), OUP *Intersections: AI & Society* (AI & Relationships); editorial board, *International Journal of AI and Consciousness*.
- Lead organiser of multiple Cambridge conferences and workshops (Kinds of Intelligence; Varieties of Mind; Ascriptions of Consciousness); founder of the Cambridge Kinds of Intelligence reading group.
- Hiring committees for ten teaching staff, including four UTO-level appointments.

TEACHING

- Designed and delivered eleven postgraduate modules, including Introduction to AI Ethics, Cognitive Science & AI, Anthropomorphism in AI, Ethics of Large Language Models, AI Safety & Existential Risk, Ethics of Human-AI Relationships, and Robot Rights & Artificial Moral Status.
- 30+ master’s dissertations supervised since 2023 across the MSt in AI Ethics & Society, the MPhil in Ethics of AI, Data & Algorithms, and the MPhil in History & Philosophy of Science.
- Extensive Cambridge undergraduate supervising across Philosophy of Mind, Metaphysics, Epistemology, Ethics & Political Philosophy, and Philosophy of Science (Parts IA, IB, and II).
- Five years’ teaching at Baruch College, CUNY (2011–16), spanning 30+ solo-taught lecture courses.

LANGUAGES

Latin & Ancient Greek (proficient) · French & Italian (intermediate) · Japanese, German & Tagalog (elementary)