Cosmin Badea

AI researcher · Lecturer, Imperial College London · Founder, Ethicos
cos@ethicos.co.uk · London, UK · linkedin.com/in/cbadea

Researcher in AI safety and artificial moral reasoning. Builds explicit, logic-based frameworks for value-aligned decision-making and works on interpretability for moral AI. Lectures on contemporary philosophy and AI at Imperial. Leads AI and ethics on two European Commission-funded consortia in personalised oncology and early diagnosis.

RESEARCH INTERESTS

Artificial moral agents · scalable oversight through explicit moral reasoning · interpretability for value-aligned systems · meta-decision-making (the "Three Rs") · the interpretation problem in moral AI · philosophy of AI.

EXPERIENCE

Lecturer, Imperial College London *2019 — present*

Department of Computing (2019–2023) · Centre for Languages, Culture and Communication (2023—present)

- Designed and delivered *AI and Ethics* — the first course of its kind in the Department of Computing (launched 2019). Student satisfaction 95.45/100. Seeded multiple PhD collaborations and joint publications.
- Teach *Contemporary Philosophy and AI* at the Centre for Languages, Culture and Communication; grew the course from 10 students to 47 after introducing the AI strand.
- Supervise MEng and MSc research in moral AI, interpretability, and AI safety; several student projects have led to peer-reviewed publications.

Chief AI Officer, FH-EARLY *2024 — present*

European Commission Horizon early-diagnosis consortium

- Lead AI strategy across the consortium: define interpretability and responsible-use requirements for the ML tools in development, and chair the AI working group across clinical and technical partners.

Lead, AI and Ethics, 4PCAN *2023 — present*

European Commission Horizon personalised-oncology consortium · €5.3M

- Design the interpretability and ethics framework for ML-driven cancer-care tools across consortium partners; embed value-aligned decision-making into the consortium's clinical AI work.
- Point person for the consortium on AI safety and responsible-use policy.

Founder and Director, Ethicos *2024 — present*

- Advise research groups, clinical consortia, and AI startups on interpretable, value-aligned AI. Clients include two early-stage AI companies (ongoing technical advisory).

Founder, Imperial-based Research Institute (in formation) *2025 — present*

- Setting up a research institute on AI applied to health, backed by committed international investors. Leading on research direction and AI safety posture.

Co-founder and Acting CEO, Mișcă-te acasă *2020 — 2021*

Digital health platform, founded as a charity at the start of the pandemic

- Co-founded and led a digital platform for better health, originally established as a charity. Acting CEO for the first year of operations; the platform grew rapidly during the pandemic.

Analyst, Morgan Stanley *2012 — 2013*

- Quantitative software engineering in the Fixed Income division (Scala, Java), held concurrently with undergraduate studies at Imperial.

EDUCATION

PhD, Imperial College London *2014 — present*

Department of Computing · Supervisor: Prof Marek Sergot · Thesis pending defence

- Thesis on value-aligned moral agents and rule-based AI. Developed MARS (Multi-valued Action Reasoning System), a framework for agents that deliberate over moral rules explicitly rather than learning them implicitly.

MEng Computing, Imperial College London *2010 — 2014*

First Class Honours

SELECTED PUBLICATIONS

- Bolton, W., Badea, C., Georgiou, P., Holmes, A., Rawson, T. *Developing moral AI to support antimicrobial decision-making*. *Nature Machine Intelligence*, 4, 912–915 (2022).
- Vijayaraghavan, A. & Badea, C. *Minimum levels of interpretability for artificial moral agents. AI and Ethics* (2024).
- Seeamber, A. & Badea, C. *Building morality into an artificial agent: first steps. IEEE Intelligent Systems* (2023).
- Badea, C. & Artus, G. *Morality, machines and the interpretation problem: a value-based, Wittgensteinian approach to building moral agents.* In *Artificial Intelligence XXXIX (SGAI 2022)*, Springer LNAI.
- Badea, C. *Have a break from making decisions, have a MARS: the Multi-valued Action Reasoning System.* In *Artificial Intelligence XXXIX (SGAI 2022)*, Springer LNAI.
- Badea, C. & Gilpin, L. *Establishing meta-decision-making for AI: an ontology of relevance, representation and reasoning*. *AAAI Fall Symposium* (2021).
- Post, B., Badea, C., Faisal, A., Brett, S. *Breaking bad news in the era of artificial intelligence and algorithmic medicine. BMC Medical Ethics* (2022).
- Hindocha, S. & Badea, C. *Moral exemplars for the virtuous machine: the clinician's role in ethical artificial intelligence for healthcare. AI and Ethics* (2021).

SERVICE AND COMMUNITY

- Reviewer Board, *AI and Ethics* (Springer).
- Reviewer, *Frontiers in Health Services*.
- Steering Committee, AAAI Fall Symposium on Cognitive Systems for Anticipatory Thinking (COGSAT).
- Affiliate, Imperial-X Human-AI initiative.

SELECTED RECOGNITION

- Student satisfaction 95.45/100 on the *AI and Ethics* course at Imperial Computing — highest in the department that year.
- Invited keynote speaker, lecturer, and panellist on AI safety, ethics, and philosophy at international academic and industry forums.
- Received offers, after full interview processes, for technical AI and AI strategy roles at a leading technology firm, a research-in-industry position at a major banking group, and a senior-manager role in AI consulting at a Big Four firm; chose to remain in research to focus on AI and Ethics.

TECHNICAL

Languages with industry experience: Java, Scala, Haskell, Prolog, C++, SQL, JavaScript. Track record of acquiring a new language to working fluency within roughly a week.

Methods: symbolic and logic-based AI, rule-based reasoning systems, interpretability for value-aligned systems, philosophy of AI.

Languages spoken: English (native-level), Romanian (native), French (working).