The Limits of Artificial Agency

Joanna Bryson
Hertie School of Governance
Berlin, Germany

ABSTRACT

Decades ago, Pattie Maes (perhaps among others[1]) first argued for the use the word 'agent' to describe both animals and 'animats' — that is, to easily express the unified concerns of artificial and natural intelligence. Now those of us with experience in building agentic systems seem only able to gape in horror as ordinary users blithely give password and credit card control to systems no one has validated.

What are the limits of the agency we can attribute to AI? Are humans really no more than 'stochastic parrots', and do AI systems even rise to that level? What does it mean to be collaborator? What does it take to establish trust? How do we stop people from being stupid or evil about AI?

In this talk, I lay out the ontology for understanding what qualifies for not only agency [4], but legal and moral agency [1, 3]. I describe how trust is built in a community [6], and collaboration built between partners [5]. I also touch on the massive legislative effort the European Union has made to ensure that AI is deployed in ways that do not destabilise our world [2].

I do this all in about thirty minutes, so we can have plenty of time for the discussion and debate that builds our society, and gives us collective agency.

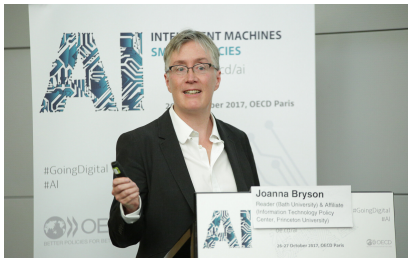
KEYWORDS

Ethics, Agency, Regulation, Security

ACM Reference Format:
Joanna Bryson. 2026. The Limits of Artificial Agency. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 1 page. https://doi.org/10.65109/

BIOGRAPHY

Joanna J Bryson has been Professor of Ethics and Technology at Hertie School, Berlin since February 2020. She is globally recognised for expertise in intelligence broadly, including AI policy and impacts. Her original academic focus was behavioural ecology, using AI for scientific simulations of intelligence. During her PhD on systems engineering of AI, she observed the confusion generated by anthropomorphised AI, leading to her first ethics publication "Just Another Artifact" in 1998. In 2010 her work in AI ethics was first recognised by a policy body when she was invited to participate in the UK research councils' Robot Ethics retreat, where she coauthored the UK's (EPSRC/AHRC) "Principles of Robotics," the world's first national-level AI ethics soft policy. Her present research focus is the impact of intelligent technology on economies, security, and human cooperation. She also studies transparency for and through AI systems, technological impacts on power, interference in democratic regulation, the future of labour, redistribution, and digital governance more broadly. She consults frequently on policy and science including to government entities in Germany, the UK, the EU (EP/EC), US, Singapore, Switzerland, and Canada; transnational organisations including Unesco, the UN, OSCE, OECD, CoE, EuroMed; NGOs such as the Red Cross, Chatham House, IEEE, WEF. In 2020, Germany nominated her to the Global Partnership of AI, where she chaired an AI Governance committee.



ACKNOWLEDGMENTS

I deeply appreciate this invitation, and thank the committee involved.

REFERENCES

[1] Joanna J. Bryson. 2018. Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology* 20, 1 (01 March 2018), 15–26. https://doi.org/10.1007/s10676-018-9448-6
[2] Joanna J Bryson. 2023. Human Experience and AI Regulation: What European Union Law Brings to Digital Technology Ethics. *Weizenbaum Journal of the Digital Society* 3, 3 (2023).
[3] Joanna J. Bryson, Mihailis E. Diamantis, and Thomas D. Grant. 2017. Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law* 25, 3 (01 Sep 2017), 273–291. https://doi.org/10.1007/s10506-017-9214-9
[4] Joanna J. Bryson and Brendan McGonigle. 1998. Agent Architecture as Object Oriented Design. In *The Fourth International Workshop on Agent Theories, Architectures, and Languages (ATAL97)*, Munindar P. Singh, Anand S. Rao, and Michael J. Wooldridge (Eds.). Springer, Providence, RI, 15–30.
[5] Katie D. Evans, Scott A. Robbins, and Joanna J. Bryson. 2025. Do We Collaborate With What We Design? *Topics in Cognitive Science* 17, 2 (2025), 392–411. https://doi.org/10.1111/tops.12682 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12682
[6] Paul Rauwolf and Joanna J. Bryson. 2018. Expectations of Fairness and Trust Co-Evolve in Environments of Partial Information. *Dynamic Games and Applications* 8, 4 (01 Dec 2018), 891–917. https://doi.org/10.1007/s13235-017-0230-x

[1] If anyone has her 1987 VUB AI Lab Tech report, I'd love to see it!

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/