

Fabio Tollon

fabiotollon@gmail.com

Areas of Specialization: AI ethics, moral philosophy, philosophy of technology

Areas of Competence: free will, philosophy of mind, philosophy of science

EDUCATION AND EMPLOYMENT

- 2023-present **Postdoctoral Researcher at the University of Edinburgh**, funded by UKRI BRAID (Bridging Responsible AI Divides)
 Principle investigators: Shannon Vallor, Ewa Luger
- 2020-2023 **PhD, Bielefeld University** (awarded *summa cum laude*, highest award possible in Germany), funded by the DFG research training group 2073 “Integrating Ethics and Epistemology of Scientific Research”.
 Title: *Agency, Responsibility, and AI: From New Technologies to Old Problems, and Back Again.*
 Supervisors: Prof. Ralf Stoecker, Prof. Thomas Reydon
- 2018-2019 **M.A. in Philosophy, Stellenbosch University, cum laude.**
 Title: *Moral Encounters of the Artificial Kind: Towards a non-anthropocentric account of machine moral agency.*
 Supervisor: Dr. Tanya de-Villiers Botha
- 2017 **B.A. (Hons) in Philosophy, Stellenbosch University, cum laude.**
 Thesis: *Pragmatic Freedom: Belief in Free Will as an Epistemically Innocent Delusion.* **Supervisor:** Dr. Tanya de-Villiers Botha
- 2013-2016 **B.A. PPE (Politics, Philosophy and Economics), Stellenbosch University.**

RESEARCH

Peer Reviewed Journal Articles (Citations: 104, h-index: 6, [Google Scholar Profile](#))

- Bennett, S.J., Catanzariti, B., and **Tollon, F.** (2025) “Everybody knows what a pothole is”: Representations of work and intelligence in AI practice and governance’, *AI & Society*.
- Oimann, A.-K. and **Tollon, F.** (2024a) ‘Responsibility Gaps and Technology: Old Wine in New Bottles?’, *Journal of Applied Philosophy*. Available at: <https://doi.org/10.1111/japp.12763>
- Tollon, F. (2024b) ‘Technology and the Situationist Challenge to Virtue Ethics’, *Science and Engineering Ethics*, 30(10). Available at: <https://doi.org/10.1007/s11948-024-00474-4>.
- Tollon, F. (2023a) ‘Free Will as an Epistemically Innocent False Belief’, *European Journal of Analytic Philosophy*, 19(2), pp. 1–23. Available at: <https://doi.org/10.31820/ejap.19.2.2>
- Tollon, F. (2023b) ‘Reactive agency and technology’, *AI and Ethics*. Available at: <https://doi.org/10.1007/s43681-023-00366-6>.
- Tollon, F. (2022) ‘Responsibility gaps and the reactive attitudes’, *AI and Ethics*. Available at: <https://doi.org/10.1007/s43681-022-00172-6>.
- Tollon, F. (2021a) ‘Artifacts and affordances: from designed properties to possibilities for action’, *AI & Society*. Available at: <https://doi.org/10.1007/s00146-021-01155-7>.
- Tollon, F. (2021b) ‘Designed to Seduce: Epistemically Retrograde Ideation and YouTube’s Recommender System’, *International Journal of Technoethics*, 12(2), pp. 60–71. Available at: <https://doi.org/10.4018/IJT.2021070105>.
- Tollon, F. (2021c) ‘Do others mind? Moral agents without mental states’, *South African Journal of Philosophy*, 40(2), pp. 182–194. Available at: <https://doi.org/10.1080/02580136.2021.1925841>.
- Tollon, F. (2021d) ‘The artificial view: toward a non-anthropocentric account of moral patiency’, *Ethics and Information Technology*, 23(2), pp. 147–155. Available at: <https://doi.org/10.1007/s10676-020-09540-4>.
- Tollon, F.** and Naidoo, K. (2021e) ‘On and beyond artifacts in moral relations: accounting for power and violence in Coeckelbergh’s social relationism’, *AI & Society*. Available at: <https://doi.org/10.1007/s00146-021-01303-z>.

Tollon, F. (2019). Moral Agents or Mindless Machines? A critical appraisal of agency in artificial systems. *Hungarian Philosophical Review* 63(4), pp. 9-23. Available at: <https://philpapers.org/archive/TOLMAO.pdf>

Reports and White Papers

Tollon, F., & Vallor, S. (2025). The Responsible AI Ecosystem: A BRAID Landscape Study. Zenodo. <https://doi.org/10.5281/zenodo.15195686>

Peer Reviewed Book Chapters:

Tollon, F. (2023) 'Social Robots and Relational Capacities'. In: Hakli, R., Mäkelä, P., Seibt, J. (eds.) *Social Robots in Social Institutions*. Frontiers in Artificial Intelligence and Applications, vol. 366. IOS Press.

Tollon, F. (2022a) 'Answerability, Accountability, and the Demands of Responsibility'. In: Pillay, A., Jembere, E., Gerber, A. (eds) Artificial Intelligence Research. SACAIR 2022. *Communications in Computer and Information Science*, vol. 1734. Springer, Cham.

Tollon, F. (2022b) 'Is AI a Problem for Forward Looking Moral Responsibility? The Problem Followed by a Solution'. In: Jembere, E.; Gerber, A.J.; Viriri, S. & Pillay A. (eds), Artificial Intelligence Research. SACAIR 2021 *Communications in Computer and Information Science*, Vol. 1551. Springer: Cham. 307–318.

Book Reviews:

Tollon, F. (2020) 'Transhumanism as a New Social Movement: The Techno-Centered Imagination', *Metapsychology Online Reviews*, Vol. 24 (39).

Tollon, F. (2021a) 'The Oxford Handbook of the Ethics of AI', *Prometheus*, Vol. 37 (2).

Tollon, F. (2021b) 'Action and Interaction', *Metapsychology Online Reviews*, Vol. 25 (16).

Tollon, F. (2021c) 'Cyber Republic', *Prometheus*, Vol. 37, (3).

Tollon, F. (2022) 'Green Leviathan or the Poetics of Political Liberty', *Prometheus*

Tollon, F. (2024a) 'Technology Ethics', *Prometheus*

Tollon, F. (2024b) 'The Handover', *Robonomics*

In-Progress Papers and Under Review:

'Entangling Ourselves with AI: Affirmative Responsibility and the Cultivation of Responsible Agency' (invited book chapter, under review), with Shannon Vallor

'Avoiding Responsibility Monoculture' (under review)

'Ed-tech, role responsibilities and teachers' moral entanglement', with Michal Wieczorek (draft available upon request)

'Vicarious Responsibility and autonomous systems' (draft available upon request)

'Conceptual Engineering and Responsibility gaps', with Enrico Galvagni (Revise & Resubmit, *Inquiry*)

'Assertions from the Margins', with Guido Löhr (under review)

'Conservative Freedom: Dennett's Conceptually Conservative Defense of Free Will' (draft available upon request)

'Reactive Attitudes and "Virtue Theatre"' (with Andrew Rebera)

'Cybersecurity and Responsibility' (with Ann-Katrien Oimann)

'Artificial Systems: Entities or Agents?' (Under review)

Peer-reviewed Conference Presentations:

1. 'Ed-tech, role responsibilities and teachers' moral entanglement', IACAP 2025, University of Twente, the Netherlands, 2-4 July 2025
2. 'Vicarious Responsibility and Autonomous Systems', SPT bi-annual conference, Technical University Eindhoven, the Netherlands, 25-28 June 2025
3. 'Vicarious Responsibility and Autonomous Systems', Penn-Georgetown Digital Ethics Workshop, Georgetown University, 7-8 March 2025.
4. 'Moral Answerability and Relational Agency', South African Conference for Artificial Intelligence Research, University of the Free State, South Africa, 2-6 December 2024.

5. 'Affirmative Responsibility Taking and the Example of Project Maven', AI4Society Symposium Series: AI and the Military, University of Pretoria, Online, 8 November 2024
6. 'Responsible AI: Contested Bodies of Practice', Conference on the History of Philosophy of Technology, Maastricht University, the Netherlands, 28-30 October 2024, Maastricht.
7. 'Responsible AI: Evolving Bodies of Practice', 4TU ESDIT Conference, University of Twente, the Netherlands, 2-4 October 2024
8. 'Responsible AI: Evolving Bodies of Practice', AFINO International Conference, Oslo, 28-30 August 2024.
9. 'Perverse Incentives and Unhealthy Ecologies: R-AI in Undergraduate Education' (with James Garforth), TexnēCon 2024, Northeastern University, 26-27 July 2024.
10. 'Responsibility Gaps: Old Wine in New Bottles', Southern African Philosophical Society Conference (PSSA), Stellenbosch, South Africa, 17-19 January 2024.
11. 'Artificial Systems: Entities or Agents?', CAIR AI4Good, Stellenbosch, South Africa, 16 January 2024.
12. 'Does Blaming Robots Make Them Responsible?', South African Conference for Artificial Intelligence Research, Muldersdrift, South Africa, 4-8 December 2023.
13. 'Artificial Systems: Entities or Agents?', Salzburg Conference for Young Analytic Philosophy (SOPhiA), Salzburg, 6-8 September 2023.
14. 'Free Will as an Epistemically Innocent Delusion', 4th Munich Graduate Conference in Ethics, Munich, 6-7 July 2023.
15. 'Responsibility Gaps: Old Wine in New Bottles', IACAP 2023, Prague, 3-5 July 2023.
16. 'Does AI Undermine Forward-Looking Moral Responsibility', Tech Ethics eXchange Northeast, Harvard, 18-19 May 2023.
17. 'Answerability, Accountability, and the Demands of Responsibility', South African Conference for Artificial Intelligence Research, Stellenbosch, 5-9 December 2022.
18. 'Artificial Systems: Entities or Agents?', Ethics of Socially Disruptive Technologies, Leiden, 6-7 October 2022.
19. 'The Soft Impacts of Emerging Technology', Ethics of Socially Disruptive Technologies, Leiden, 6-7 October 2022.
20. 'Digital Nudging and Climate Change', Technology and Politics, KU Leuven, 19-21 September 2022.
21. 'Social Robots and Relational Capacities', Social Robots in Social Institutions, University of Helsinki, 16-19 August 2022.
22. 'Digital Nudging and Virtue Ethics in the Face of Climate Change', Philosophy of Human Technology-Relations, Aalborg University Copenhagen, 5-7 July 2022.
23. 'Thinking Social Relationism from Africa', Intercultural Ethics and Technology Conference 2022, University of Twente, 27-28 January 2022.
24. 'How, and Why, we are Responsible for the Future of AI', South African Conference for Artificial Intelligence Research, Durban, 6-10 December 2021.
25. 'From Responsibility Gaps to Responsibility Maps', Australasian Postgraduate Philosophy Conference, Australasian Association of Philosophy (online), 20-21 November 2021.
26. 'Artifacts and Affordances: From Designed Properties to Possibilities for Action', Changing Values, Changing Technologies, Delft University of Technology, the Netherlands, 12-13 October 2021.
27. 'Unpredictable Futures: Why, and How, we are Responsible for AI', Philosophy and Theory of Artificial Intelligence. Chalmers University of Technology, Gothenburg, Sweden, 27-28 September 2021.
28. 'Designed to Seduce', UCD Online Conference on the Philosophy of Conspiracy Theories. University College Dublin, Ireland, 1-2 July 2021.
29. 'Do Others Mind? Moral Agents Without Mental States', Philosophy of Human-Technology Relations. University of Twente, The Netherlands, 4-7 November 2020.
30. 'Toward a Coherent Account of Moral Agency', South African Forum for Artificial Intelligence Research, Cape Town, South Africa, 3-6 December 2019.
31. 'Functional Moral Agency'. 4IR: Philosophical, Ethical and Legal Perspectives, University of the Witwatersrand, South Africa, 3-5 September 2019.

Invited Talks and Keynote Presentations:

1. 'Who's Responsible for Responsible AI?', Technical University Munich (TUM) Life Sciences Seminar Guest Lecture, 22 May 2025

2. 'Vicarious Responsibility and Autonomous Systems', Workshop on AI and Responsibility Gaps, Keynote, LMU Munich, 21 May 2025
3. 'Responsibility Gaps and Emerging Technology: Old Wine in New Bottles?', Department of Communication, Northern Illinois University, 21 March 2025
4. 'Responsibility Gaps and Technology: Old Wine in New Bottles?', Uppsala Vienna AI Colloquium, Online, 13 December 2024
5. 'Entangling Ourselves with AI: Affirmative Responsibility and the Cultivation of Responsible Agency', AI Ethics & Society, Edinburgh, 6 November 2024
6. 'Should we entangle ourselves with AI? Promises and limits of vicarious responsibility', INTERACT! workshop "Interaction with intelligent systems", INTERACT! Group, Ruhr University Bochum, 24-25 October 2024.
7. 'Moral Answerability and Technology: Cultivating Responsible Agency in an Irresponsible World', International conference on virtue ethics and technology, KU Leuven, 19-20 September 2024, Leuven (Keynote).
8. 'Responsibility Gaps and Technology: Old Wine in New Bottles', INTERACT! Group, Ruhr University Bochum, 17 May 2024.
9. 'An Affordance Account of Value Embedding in Technology: Why Good Intentions are Not Enough', Cambridge NLP Seminars, 3 March 2023.
10. 'Does AI Undermine Forward-Looking Responsibility?', UCT Department of Philosophy Seminar, 1 March 2023.
11. 'Does AI Pose a Threat to Forward-Looking Moral Responsibility?', ZEPP-Talks, LMU, Germany, 9 November 2022.
12. 'From Designed Properties to Possibilities for Action', MilaNLP Lab, Bocconi University, Italy, 4 November 2022.
13. 'The Ethics and Imperative of Digital Nudging', MKAI September AI Inclusive Forum, 23 September 2021.
14. 'From Responsibility Gaps to Responsibility Maps', Data and Computational Ethics Research Group, Department of Philosophy, Stellenbosch University, 13 May 2021.

TEACHING

2018-present	Guest lecturer at the Creative Academy in Cape Town (4 lectures per year)
2019	TA, Critical Social Theory and Ideology Critique (BA 3 rd year), Stellenbosch University
2017-2019	TA, Philosophy of Science and Epistemology (BA 2 nd year), Stellenbosch University
2018-2019	TA, History of Analytic Philosophy and Philosophy of Mind (BA 3 rd year), Stellenbosch University.
2018	Lecturer, Engineering and Ethics (BEng 3 rd and 4 th year), Stellenbosch University
2018	Lecturer, Science, Ethics and Aesthetics in a Philosophical Perspective (for international students), Summer School Philosophy Elective, Stellenbosch University
2017	Tutor, introduction to Systematic Philosophy (BA 1 st year)
2017	Tutor, Greek Philosophy and the Philosophy of the Middle Ages (BA 1 st year)
2017	Tutor, Moral Philosophy (BA 1 st year)
2017	Tutor, Practical Logic and Critical Thinking Skills (BA 1 st year)

SUPERVISION

Creative Academy Cape Town

Supervision of 3rd year undergraduate student research essays (x3)

University of Edinburgh Natural Language Processing (NLP) Centre for Doctoral Training (CDT)

Co-supervision of an NLP group project on different perspectives on Responsible AI (with James Garforth, Benedetta Catanzariti, and Hannah Rohde)

University of Edinburgh

MSc Supervision (x2, in progress)

Executive Editor

Executive editor for the theme ‘digital ethics’ at *Minds and Machines*.

Philosophical Glossary for AI

Contributed two entries (‘moral responsibility’ and ‘agency’) to an online [AI glossary for policy-makers](#), in collaboration with University of London’s School of Advanced Studies (SAS) and the UK department for Science, Innovation, and Technology (DSIT).

Conference Organization

2022 and 2023 iterations of the Ethics of Technology Early Career Group (ETEG) conference

‘Humane AI’ workshop at the University of Melbourne in 2024

South African Centre for AI Research (SACAIR) conference organizer (Philosophy and Law track), 2023-present

Peer Review

Reviewer for: *AI & Society*, *AI and Ethics*, *Science and Engineering Ethics*, *Journal of Medical Ethics*, *Philosophy & Technology*, *Minds and Machines*, *Topoi*, *Journal of Value Inquiry*, *International Journal of Technoethics*, *Synthese*, *Journal of Applied Philosophy*

Book reviews editor

Journal of Ethics and Emerging Technologies

Community Engagement & Leadership

Vice-Chairperson (2018) and Chairperson (2019), Socratic Society, Stellenbosch University

VISITING POSITIONS

2022-present	Research fellow at the Unit for the Ethics of Technology at Stellenbosch University
2021-present	Research associate at the Center for Artificial Intelligence Research (CAIR), at the University of Pretoria

RESEARCH VISITS

May-June 2024	Visiting Fellow at INTERACT!, Ruhr University Bochum (third party funded)
November 2022	Visiting researcher at the Center for Ethics and Philosophy in Practice (CEPP) at the Ludwig Maximilian University of Munich (third party funded)
Feb-April 2022	Visiting researcher at the Working Group on the Philosophy of Technology at KU Leuven (third party funded)

REFERENCES

Prof. Shannon Vallor: svallor@exseed.ed.ac.uk

Prof. Emma Ruttkamp-Bloem: emma.ruttkamp-bloem@up.ac.za

Prof. Dr. Ralf Stoecker: ralf.stoecker@uni-bielefeld.de (PhD thesis supervisor)

Prof. Lode Lauwaert: lode.lauwaert@kuleuven.be

Dr. Tanya De Villiers-Botha: tdev@sun.ac.za (MA thesis supervisor)