

JOHN DORSCH

johndorsch@gmail.com

[Google Scholar](#)

[Homepage](#)

[ORCID](#)

[Open Science Framework](#)

[LinkedIn](#)

CURRENT POSITIONS

- 2024 – Present: **Postdoctoral Researcher**, Center for Environmental and Technology Ethics, Prague (CETE-P). Institute of Philosophy, Czech Academy of Sciences. Funded by Horizon Europe, 101086898.
- 2026 – Present: **Lecturer**, Charles University, Center for Economic Research and Graduate Education.

PREVIOUS POSITIONS

- 2022 – 2024: **Postdoctoral Researcher**, Ludwig-Maximilians-Universität München (LMU), Munich Interactive Intelligence Initiative (MI3); Cognition, Values, and Behaviour Lab, funded by the Bavarian Research Institute for Digital Transformation. Project title: *Digital Workplace and Human-AI Assisted Training*.
- 2018 – 2022: **Lecturer**, University of Edinburgh (UoE), School of Philosophy, Psychology, and Languages Sciences (PPLS). Philosophy, Psychology and Informatics Group.

EDUCATION

- 2018 – 2022: **PhD** in Philosophy of Cognitive Science. Awarded Nov 2022. University of Edinburgh. Supervisors: Prof. Tillmann Vierkant & Dr. Dave Ward. Examiners: Prof. Andy Clark & Dr. Alistair Isaac. Passed without corrections (*summa cum laude*). Thesis: *Embodied Metacognition*. [UoE Archive](#).
- 2016 – 2018: **Master's** (research) Philosophy (GPA 3.95; Note: 1,1). Universität Tübingen, Center for Integrative Neuroscience. Supervisor: Prof. Hong Yu Wong. Thesis: *Phenomenology of Meaning Experience*.
- 2012 – 2016: **Bachelor's** (honours) Philosophy (GPA 3.7; Note: 1,5). Universität Tübingen, Center for Integrative Neuroscience. Supervisor: Prof. Hong Yu Wong. Thesis: *Phenomenology of the 'Aha!' Experience*.
- 2001 – 2004: **Associate's** Computer Science (GPA 3.5). Keiser University, Florida, USA.

PUBLICATIONS

- **Dorsch, J.** (2026). The Collective Risk Structure of AI Welfare: Rethinking Consciousness, Vulnerability, Suffering, and Moral Standing in the More-Than-Human World. *Proceedings of the AISB Convention 2026. Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, 2026. <https://doi.org/10.5281/zenodo.20964411>
- **Dorsch, J.**, & Moll, M. (2026). Explainable and Human-Grounded AI for Decision Support Systems: The Theory of Epistemic Quasi-Partnerships. In: Müller, V.C., Dung, L., Löhr, G., Rumana, A. (eds) *Philosophy of Artificial Intelligence*. Synthese Library, vol 533. Springer, Cham. https://doi.org/10.1007/978-3-032-10073-3_5
- **Dorsch, J.** (2026). *Origins and Future of Self-Knowledge: Epistemic Agency, 4E Metacognition, and Artificial Intelligence*. Studies in Brain and Mind, vol. 25. Springer Nature. <https://doi.org/10.1007/978-3-032-16161-1>
- Schadd, M., Berman, D.S., Chen, C., Cohen, M., **Dorsch, J.**, Gegov, A., Moll, M., Rose, O., Rösner, A., Saastamoinen, K., Schiller, T., Strand, A., Vasankari, L., & Rempel, M. (2026). Reinforcement Learning for Decision Support in Defense and Security: A Systematic Review. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-026-07205-x>
- **Dorsch, J.**, Goddu, M. K., Nave, K., Vierkant, T., Coeckelbergh, M., Gürtler, P., Urban, P., Spang, F., & Moll, M. (2025). Against AI Welfare: Care Practices Should Prioritize Living Beings Over AI. *AI Magazine*. <https://doi.org/10.1002/aaai.70016>
- Moll, M. & **Dorsch, J.** (2025). A Systematic Review of Human-Centered Explainability in Reinforcement Learning: Transferring the RCC Framework to Support Epistemic Trustworthiness. *Human-Intelligent Systems Integration*. <https://doi.org/10.1007/s42454-025-00084-w>
- **Dorsch, J.** (2025). Mindshaping and AI: Will mindshaping a robot create an artificial person? In T. Zawidzki, T. & R. Tison (Eds.), *The Routledge Handbook of Mindshaping* (pp. 406-417). Routledge. AAM available at <https://philpapers.org/rec/DORMAA-3>
- **Dorsch, J.**, & Deroy, O. (2025). The impact of labeling automotive AI as trustworthy or reliable on user evaluation and technology acceptance. *Nature Scientific Reports*, 15, 1481. <https://www.nature.com/articles/s41598-025-85558-2>
- **Dorsch, J.**, & Deroy, O. (2024). Quasi-Metacognitive Machines: Why We Don't Need Morally Trustworthy AI

and Communicating Reliability is Enough. *Philosophy & Technology*, 37, 62.

<https://doi.org/10.1007/s13347-024-00752-w>

– **Dorsch, J.** (2023). Are Noetic Feelings Embodied? The Case for Embodied Metacognition. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2023.2197937>

– **Dorsch, J.** (2022). Hijacking Epistemic Agency: How Emerging Technologies Threaten our Wellbeing as Knowers. *Proceedings of the AAAI and ACM on Artificial Intelligence, Ethics, and Society*, 5.

<https://doi.org/10.1145/3514094.3539537>

– **Dorsch, J.** (2022). Embodied Metacognition: How we feel our hearts to know our minds. Doctoral Dissertation, University of Edinburgh. *Edinburgh Research Archive*. <https://dx.doi.org/10.7488/era/2763>

– De Preester, H. & **Dorsch, J.** (2021). Descartes's Passions of the Soul, Epistemic Affect and the Challenges for Interoception Research in Emotions. *Danish Yearbook of Philosophy*, 54 (1), 65-92.

<https://doi.org/10.1163/24689300-bja10021>

– **Dorsch, J.** (2018). *Prescribing Proficiency: Ethical Medical Practice in English-Speaking Contexts*. German American Institute (DAI), Tübingen. (Course book for teaching cultural competency and ethical medical practices in patient encounters, published by the DAI and actively used in their curriculum).

– **Dorsch, J.** (2017). On Experiencing Meaning: Irreducible Cognitive Phenomenology and Sinewave Speech. *Phenomenology and Mind* 12, 218-227. https://doi.org/10.13128/PHE_MI-21120

– **Dorsch, J.** (2016). Irreducible Cognitive Phenomenology and the 'Aha!' Experience. *Phenomenology and Mind* 10, 108-121. https://doi.org/10.13128/PHE_MI-20095

Critical Notices

– **Dorsch, J.** (2024). Epistemic and Moral Agency in a Crisis of Trust? A Critical Notice of The Tinkering Mind *International Journal of Philosophical Studies*. <https://doi.org/10.1080/09672559.2024.2437269>

– **Dorsch, J.** (2024). Can You See a Ganzfeld? A Critical Notice of The Unity of Perception *International Journal of Philosophical Studies*. <https://doi.org/10.1080/09672559.2024.2360782>

Work forthcoming

– **Dorsch, J.** (forthcoming). Denial of Artificial Agency: AI Might Soon Be Entitled to Moral Concern. In T. Hribek (Ed.), *Theoretical Problems for the Philosophy of AI*. Springer.

– **Dorsch, J.**, Coeckelbergh, M., & Vierkant, T. (forthcoming). Does AI Undermine Democratic Deliberation? How Engagement Algorithms Hijack Epistemic Agency. *Synthese*.

– **Dorsch, J.**, Moll, M., & Deroy, O. (forthcoming). Why We Need Epistemically, Not Morally, Trustworthy AI: Decision Support and the Need for Collective Accountability. *Philosophy & Technology*.

– **Dorsch, J.** (forthcoming). Why AI Cannot Be a Bona Fide Moral Agent (Yet): Trustworthiness, Moral Normativity, and Giving a Damn. *American Philosophical Quarterly*.

– Schadd, M., Berman, D.S., Chen, C., Cohen, M., **Dorsch, J.**, Gegov, A., Moll, M., Rose, O., Rösner, A., Saastamoinen, K., Schiller, T., Strand, A., Vasankari, L., Rempel, M. (2026). Exploiting Reinforcement Learning to Achieve Decision Advantage. NATO Science and Technology Organization (STO), SAS-181 Research Task Group. Pre-release Technical Report. <https://doi.org/10.14339/STO-TR-SAS-181>

Work under review

– **Dorsch, J.** (under review). *Can Metacognition Still Be Truly Meta? Confronting The Twin Triviality Problem*

– **Dorsch, J.** & Kopecký, R. (under review). *Uncanny Valley Emerges in Middle Childhood: Evidence from Two Studies on Likeability Judgments*.

– **Dorsch, J.**, Kim, M., Burianová, E., & Nimrichter, I. (under review). *How Do People Respond to Evidence of Intergenerational Inequality in Academia? A Content Analysis of a Viral LinkedIn Discussion*.

– **Dorsch, J.**, Bagherianzarat, A., Hvoreký, J. (under review). *Your Chatbot Will Never Wholly Kiss You: Sycophantic AI Companionship and Its Psychological and Social Risks*.

Work in preparation

– Krutský, M., Gürtler P., Peleška, J., Souregus, G., Němeček, J., **Dorsch, J.** (in prep.). *XAI Desiderata for Ethical AI: Insights from the AI Act*.

POLICY REPORTS AND BRIEFS

– Gürtler, P., Yoldas, T., Bellon, J., Kotková, A., McIntyre, A., Mesarčík, M., Šimko, J., Solarova, S., & Dorsch, J. (2026). *Improving Existing EU Policy to Preserve Cognitive Agency: Democracy in AI-Mediated Environments*. Center for Environmental and Technology Ethics — Prague, Institute of Philosophy, Czech Academy of Sciences. <https://doi.org/10.5281/zenodo.21278799>

GRANTS, FELLOWSHIPS, MENTORSHIPS, AND AWARDS

– 2018 – 2022: **PhD Scholarship**, University of Edinburgh, PPLS Scholarship. £100,000.
– 2022: **Associate Fellow of the Higher Education Academy**. University of Edinburgh.
– 2022: **Mentorship on AI Ethics**, University of Oxford, Mentor: Dr. Atoosa Kasirzadeh.
– 2022: **Finalist for Tutor of the Year**, (in top three, university-wide), University of Edinburgh.
– 2019 – 2022: **Nominated Tutor of the Year**, University of Edinburgh. Thrice nominated.
– 2017 – 2019: **Germany Scholarship**, Universität Tübingen. Twice awarded. €8,000.
– 2017: **Best Presentation Award**, Conference of Junior Neuroscientists, Universität Tübingen.
– 2017: **Best Essay Award**, International Competition for Master's Students. University of Edinburgh.

TEACHING ACTIVITIES

– 2026 – Present: **Course Organizer**, Charles University. *Philosophy of Technology and Ethics of AI*
– 2023 – 2024: **Course Organizer**, LMU (taught in German) (I): *Einführung in die Philosophie der Metakognition* (taught for two semesters). (II): *Einführung in die Philosophie der Technologie*.
– 2022: **Lecturer**, LMU: *Cognitive Maps and their Implications for Human-AI Interaction*.
– 2022: **Lecturer**, UoE: *Topics in Mind and Cognition*.
– 2018 – 2022: **Tutor**, UoE: (I): *Logic I*. (II): *Philosophy of Science I*. (III): *Greats: From Plato to the Enlightenment*. (IV): *Mind, Matter and Language*. (V): *Knowledge and Reality*.
– 2012 – 2016: **Tutor**, Universität Tübingen (taught in German): *Oberseminar: Philosophy and the Cognitive Sciences*. (I/II): *Vorlesung: Die großen Gestalten der griechischen Philosophie – Teil I und Teil II*.
– 2016 – 2018: **Course Organizer**, Universität Tübingen. *Ethics in Medical Practice* (taught for six semesters).
– 2010 – 2018: **Course Organizer**, German American Institute. *English as a Foreign Language* (various).
– 2007 – 2010: **Course Organizer**, Universität Tübingen. *Programming in Java and C#* (taught for four years).

SUPERVISION

– **PhD** in Philosophy: Co-Supervisor & Mentor
– 2024 - Present: Czech Academy of the Sciences, Paula Gürtler; Theme: *Political Philosophy of AI*
– **BA (Honours)** in Philosophy: Thesis/Examination
– 2024: LMU München, Fabian Segatto; Theme: *Metacognition, Mindfulness, and Mental Wellbeing*.
– 2024: LMU München, Elias Fleischmann; Theme: *AI, Consciousness, and Personhood*.

PRESENTATIONS (Peer-Reviewed / Since completing PhD)

– Against “AI Welfare” (2025). (I) *Neuroethics and Ethics of AI*, Forschungszentrum Jülich.
– Epistemic Agency in the Age of the Algorithm. (2024-2025). (I): *Responsible AI Group*. Charles Technical University, Prague. (II): *Philosophy, Psychology, and Informatics Group*, University of Edinburgh. (III): *AI & Propaganda Workshop*, Czech Academy of Sciences
– Explainable and Human-Grounded AI for Decision Support Systems: The Theory of Epistemic Quasi-Partnerships. (2024). (I): *NATO STO Group SAS-181*. Ottawa. (II): *Society for the Philosophy of AI*, 5. University of Erlangen.
– Algorithm Aversion: Trustworthy vs Reliable AI. Between-Subject Study on Technology Acceptance. (2024). (I): *Bavarian Research Institute for Digital Transformation*. Bavarian Academy of the Sciences and Humanities, München. (II): *Cognition, Values & Behaviour Research Group*, LMU. (III): *Ethics of AI Colloquium*, LMU.
– Why AI Cannot be Morally Trustworthy (2023). (I): *Workshop on Human-AI Interactions: Peace Research Institute*, Oslo. (II): *NATO: MSCoE Group (North Atlantic Treaty Organization: Modelling & Simulation Centre of Excellence)*. (III): *Cognition, Values & Behaviour Research Group*, LMU. (IV): *socialBRIDGES: Trust: Bridge Between Society and Technology*. Universität Dresden.
– EU Regulations of AI. (2023). *Cognition, Values & Behaviour Research Group*, LMU, München.
– Knowing Ourselves Together Through Embodied Entanglement: The Role of Embodied Confidence on the Path of Self-Knowledge. (2023). *4th International Conference on Philosophy of Mind: 4E's Approach to*

Mind/Brain. Universidade Católica Portuguesa. Braga, Portugal

– Co-Learn: Trust in AI Assistants and Humans in the Learning Loop. (2022/23). *Bavarian Research Institute for Digital Transformation. Bavarian Academy of the Sciences and Humanities, München.*

– Hijacking Epistemic Agency: How Emerging Technologies Threaten Our Wellbeing as Knowers. (2022). *The Fifth Association for the Advancement of Artificial Intelligence & Association for Computing Machinery Conference on Artificial Intelligence, Ethics, and Society, Keble College, University of Oxford.*

– What Do You Feel When You Feel Confident? The Bayesian Brain and Embodied Metacognition. (2022). *Consciousness Club, Wellcome Centre, University College London.*

ACADEMIC SERVICE

– Grant / Funding Review

- External Expert Evaluator, COST (European Cooperation in Science and Technology), 2025–2026:
 - 29444: AI and (Mis)Information Ethics
 - 28839: Ordinary Moral Constructions and Rethinking Ethical AI

– Manuscript Peer Review

– *Synthese, American Philosophical Quarterly, Theoria, Philosophy & Technology, Philosophical Psychology, Topoi, AI & Society, Science and Engineering Ethics, Oxford Intersections, Journal of Applied Philosophy*
– *Frontiers in Psychology, Frontiers in Sociology, Humanities and Social Sciences Communication, Discover Psychology, Discover Artificial Intelligence, Perspectives on Psychological Science, Social Science, Journal of Experimental and Theoretical Artificial Intelligence, npj digital medicine, Nature Scientific Reports, Journal of Medical Education and Curricular Development, Societies, Informatics, Health Care Analysis*

– Outreach & Mentorship

– Mentor, AI Ethics (outreach appointment). *Czech Academy of Sciences, Otevřená věda (“Open Science”)*. Year-long mentorship of three high-school research interns (2026). Project: “When AI Gives Ethical Advice: Studying Large Language Models as Moral Reasoners”.

– Event Organization

- *Sussex Workshop on AI Consciousness and ethics at AISB-2026, University of Sussex*
- *Inaugural Karel Čapek Lecture Series (2025). Invited speaker: Shannon Vallor (University of Edinburgh)*
- *CETE-P Workshop on AI Ethics with Shannon Vallor (2025).*
- *Days of AI/CETE-P Seminar (2025). Invited speaker: Tillmann Vierkant (University of Edinburgh).*
- *Edinburgh Workshop: Mental Action (2019).*
- *Edinburgh Conference: Metacognition (2020).*
- *Edinburgh Work in Progress Seminar (2019–2020).*

LEADERSHIP

- 2025 – Present: **Co-Founder & Host**, Karel Čapek Lecture Series, CETE-P, Czech Academy of Sciences.
- 2024 – Present: **Research Leader**, NATO SAS-181 for XAI; CETE-P for “AI-Welfare”.
- 2023: **Development of Policy Governing Use of AI by Students in Submitted Materials**, LMU Munich.
- 2021: **Achieved Financial Support for PhD Students Impacted by Covid-19**, University of Edinburgh
- 2002 – 2005: **IT Manager**, 8 full-time & 6 part-time computer technicians. Best Buy Inc., Florida, USA.

EXPERIMENTAL SKILLS

- **Experimental Design**: Likert-scales, vignette studies, pre-reg., ethics approval, Qualtrics, Prolific, MTurk.
- **Technical Proficiency**: R Studio, Python, JsPsych.
- **Collaboration & Open Science**: Team-based research, open science principles, transparent methodology.

PROFESSIONAL EXPERIENCE IN IT AND IT CERTIFICATIONS

- 2007 – 2010: **Web Developer**, Zentrum für Datenverarbeitung, University of Tübingen; *back-end development: PHP and SQL; front-end development: HTML and JavaScript.*
- 2006 – 2007: **Server Support Engineer**, Intel Corporation, Oregon, USA, *diagnose and repair disk, memory, and processor incidents; develop and deploy firmware updates.*
- 2005 – 2006: **Network Administrator**, McNerny Holdings, Maryland, USA, *Install, configure, update and*

maintain company network on MS 2003 Server.

– 2001 – 2005: **Computer Technician**, Best Buy Inc., Florida, USA, *diagnose and repair hardware and software issues in MS Windows Operating Systems.*

– **IT Certifications**: Microsoft Certified Professional (3381844); CompTIA: A+, Net + (COMP001004382519).

SCHOLARLY AFFILIATIONS

– 2023 – Present: **Member of NATO STO GROUP: SAS-181**, System Analysis & Studies: *Exploiting Reinforcement Learning to Achieve Decision Advantage*; interdisciplinary team, researching defence systems.

– 2026 – Present: **Member of NATO STO GROUP: IST-235**, Information Systems Technology: *Explainable AI In Defence And Security Context*; interdisciplinary group, researching defence systems.

LANGUAGES

– **Human**: *English*: Native; *German*: C2 (DSH, 2010); *Latin*: Latinum 3.0 (2015); *Spanish*: A2.

– **Machine**: *R*; *Python*, *C#*, *Java*, *PHP*, *SQL*, *LaTeX*, *HTML*: intermediate.