



Postdoc Positions in Rational Machine Learning

Deadline: September 10th, 2025

The Rational Intelligence Lab (https://ri-lab.org/) at CISPA Helmholtz Center for Information Security in Saarbrücken, Germany (https://cispa.de/en) is seeking postdoctoral researchers under the 2025 call for the Marie Sklodowska-Curie Actions (MSCA 2025) Postdoctoral Fellowships.¹ The candidates will work in the areas situated at *the intersection of machine learning, statistics, and economic theory* with the goal of enhancing the rationality, efficiency, and reliability of intelligent systems.

For more information on our past and current research, please visit our publications page at https://ri-lab.org/pubs/. You can also check out the core values we are adhering to at https://ri-lab.org/values.

What We Look For. Candidates for this position are expected to hold an excellent PhD (on the closing date for applications) in Computer Science, Mathematics, Statistics, Economics, Physics, or a related discipline, with a strong record of publications and a demonstrated interest in publishing in top-tier machine learning venues such as ICML, NeurIPS, EC, AAMAS, AAAI, ICLR, and JMLR.

We are especially interested in candidates with a strong mathematical background or empirical researchers with strong programming skills (or both) who are eager to better understand and improve the behavior of complex AI systems—such as large predictive models, e.g., recommendation systems, and Generative AI (GenAI), e.g., diffusion models, large language models, as well as how they are deployed. As AI models grow in scale and capability, new training paradigms have emerged, including pre-training, fine-tuning, and model merging. At the same time, these systems are reshaping how predictions and generations are consumed in downstream tasks, accelerating scientific discovery and supporting high-stakes decision-making. Despite the revolutionary advancements in machine learning model training and deployment, the profound real-world implications of this paradigm shift have yet to be adequately investigated. Hence, our overarching goal is to develop principled, mathematical frameworks for understanding the sometimes unpredictable behavior of these systems, with the hope to ultimately improve their generalisability, safety, security, reliability, and trustworthiness.

Specific topics of interest include, but are not limited to:

- Mathematical Abstraction and Formalization of General-Purpose AI Systems: How can we rigorously define and measure "abstraction" within AI systems? Can we formulate a mathematical theory of concept learning and generalization that explains how these systems acquire and apply abstract knowledge, and how this differs from human cognition?
- Causal and Counterfactual Analyses of Complex AI Systems: How can we design "causally aware" AI systems that can explicitly reason about interventions and their consequences, rather than merely recognizing correlations?
- Game-Theoretical Analyses of Human-AI Collaborations: How can we model human-AI interaction as a multi-agent game, where both humans and AI agents have potentially differing

¹https://marie-sklodowska-curie-actions.ec.europa.eu/calls/msca-postdoctoral-fellowships-2025

objectives, information asymmetries, and learning capabilities? What equilibrium concepts (e.g., Nash equilibrium, correlated equilibrium, cooperative game theory) are most relevant for analyzing stable and beneficial collaborations?

- **Incentive-Aware Analysis of Complex AI Systems**: *How can we formally specify and align the incentives of complex AI systems with human values and societal goals, especially when these systems operate autonomously in open-ended environments?*
- Uncertainty Representation and Quantification in Complex AI Systems: How can formal methods from probability theory and statistical inference be leveraged to provide guarantees on the reliability and safety of AI systems under conditions of inherent uncertainty, particularly in safety-critical applications?

Other topics are also welcome as long as they align with our broader mission of advancing rational intelligence. Candidates and the PI will jointly identify the precise research topic.

We are committed to fostering a diverse workforce and strongly encourage applications from individuals in underrepresented communities, including but not limited to women, people of color, LGBTQ+ individuals, persons with disabilities, and those from economically disadvantaged backgrounds.

What We Offer. Beyond the opportunity to engage in a world-class research environment, the candidates will have the opportunity to collaborate with other research groups at CISPA (https://cispa.de/en/research) and with partners within the Helmholtz AI Cooperation Unit (https://www.helmholtz.ai/). We strongly encourage collaborations with neighboring research institutes at the Saarland Informatics Campus (https://saarland-informatics-campus.de/en/) such as Saarland University (UdS), Max Planck Institute for Informatics (MPI-INF), Max Planck Institute for Software Systems (MPI-SWS), Helmholtz Center for Pharmaceutical Research (HIPS), Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), as well as the Max Planck Institute for Intelligent Systems (MPI-IS) in Tübingen. International collaborations will be highly appreciated.

Our group is part of the ELLIS Unit Saarbrücken (https://ellis.eu/units/saarbrucken).

About CISPA. The CISPA Helmholtz Center for Information Security provides a unique work environment that offers the advantages of a university department and a research laboratory alike. As the latest member of the Helmholtz Association, the largest research organization in Germany, CISPA has embarked on a mission: to rethink the digitalized world of the future from the ground up and make it safer through innovative cutting-edge research. In the medium term, the center will grow to more than 800 employees with not less than 60 Faculty and research group leaders. Faculty receive extremely competitive institutional funding, enjoy academic freedom, and build and lead their team of young researchers, and are granted the opportunity to teach graduate and undergraduate courses.

CISPA is located in Saarbrücken, in the tri-border area of Germany, France, and Luxembourg. We maintain an international and diverse work environment and seek applications from outstanding researchers worldwide. **The working language is English**. A command of German is not required for a successful career at CISPA.

How To Apply. The first step is to check whether or not you are eligible for the MSCA Postdoctoral Fellowships.² The eligible candidates can then apply by sending (1) a **cover letter** and (2) a **CV** to muandet@cispa.de. Your cover letter should clearly articulate why you believe our group is the ideal place for you and how your research experience is relevant to this position. If the fit is right, we will collaborate on the proposals.

Candidates should prepare proposals well in advance of the September 10, 2025 deadline.

 $^{^{2}} https://marie-sklodowska-curie-actions.ec.europa.eu/actions/postdoctoral-fellowships/6-steps-to-prepare-your-application$