Anka Reuel

anka.reuel@stanford.edu | Stanford, US | linkedin.com/in/ankareuel/ | ankareuel.com

Education

- **Ph.D. in Computer Science, Stanford University** *2022–2027 (projected)*
  Ph.D. Minor in Political Science
  Current GPA: 4.0/4.0
  Advisors: Prof. Mykel Kochenderfer & Prof. Sanmi Koyejo
- **MSE in Computer and Information Science, University of Pennsylvania** *2020–2022*
  GPA: 3.97/4.0 (top 5%)
  Major field of study: AI
  Advisor: Prof. Lyle Ungar
- **M.Sc. in Management & Strategy, London School of Economics** *2015–2016*
  Grade: Merit
  Major fields of study: Game Theory & Finance
- **B.Sc. in Economic Sciences, University of Hagen** *2011–2015*
  Grade: 1.4/6.0 (top 1%)
  Major field of study: Operations Research & Game Theory
- **International Baccalaureate, Stiftung Louisenlund** *2011–2013*
  Grade: 40/45 (top 5% worldwide)

Awards

- Stanford Interdisciplinary Graduate Fellowship, 2024-2027
- Stanford HAI Graduate Fellow, 2023-2024
- UPenn Outstanding Academic Award, 2022
- Hans Weisser Scholarship, 2020 – 2022
- DAAD Scholarship, 2020 – 2022
- Gen-ZEO Top Talent Under 25 Award, 2019
- Young Titans Scholar, 2018
- German National Academic Foundation Scholarship, 2013 – 2017

Research Interests

Quality and validity of model evaluations, technical responsible AI, organizational AI governance, international AI governance, vulnerabilities in critical AI systems

Professional Activities

- Geopolitics & Technology Fellow, Harvard Kennedy School/Belfer Center for Science and International Affairs, Sep 2024 – ongoing
- Fellow & Lead Writer for the AI Chapter, Stanford Emerging Technology Review, Stanford Hoover Institution, Jun 2024 – ongoing
- Moderator, AI Governance Day at the 2024 AI for Good Summit, International Telecommunication Union, May 2024
- Lead Researcher, Responsible AI Chapter, Stanford AI Index, Oct 2023 – ongoing
- Founding Member, Center for AI Risks & Impacts (KIRA), Germany, April 2023 – ongoing
- AI and Equality Initiative Research Affiliate, Carnegie Council for Ethics in International Affairs, US, Summer 2023

Publications

---

1. **A. Reuel***, Hardy A.*, Smith, C., Lamparth, M., Kochenderfer, M. (2024). BetterBench: Assessing AI Benchmarks, Uncovering Issues, and Establishing Best Practices. Under review at *2024 Conference on Neural Information Processing Systems*.
2. **A. Reuel***, Bucknall, B.*, Casper, S., Fist, T., Soder, L., Aarne, O., Hammond, L., Ibrahim, L., Chan, A., Wills, P., Anderljung, M., Garfinkel, B., Heim, L., Trask, A., Mukobi, G., Schaeffer, R., Baker, M., Hooker, S., Solaiman, I., Luccioni, A. S., Rajkumar, N., Moës, N., Ladish, J., Guha, N., Newman, J., Bengio, Y., South, T., Pentland, A., Koyejo, S., Kochenderfer, M. J., & Trager, R. (2024). Open Problems in Technical AI Governance. *arXiv preprint arXiv:2407.14981*.
3. **A. Reuel**, Soeder, L., Bucknall, B., & Undheim, T. A. (2024). On The Importance of Technical Research and Talent for AI Governance. *2024 International Conference on Machine Learning*. **Accepted as oral – top 1.5% of papers**
4. **A. Reuel** & Ma, D. (2024). Fairness in Reinforcement Learning: A Survey. *AAAI AI, Ethics & Society 2024*.
5. Rivera, J.-P.*, Mukobi, G.*, **A. Reuel***, Lamparth, M., Smith, C., & Schneider, J. (2024). Escalation Risks from Language Models in Military and Diplomatic Decision-Making. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 836-898.
6. **A. Reuel*** & Undheim, T. A.* (2024). Generative AI Needs Adaptive Governance. Under review at *Digital Policy, Regulation and Governance*.
7. Undheim, T. A. & **A. Reuel** (2024). A Literature Review of AI Governance Trends, 2020-2024. Under review at *AI & Society*.
8. Trager R., Harack, B. **A. Reuel**, Carnegie, A., Heim, L., Ho, L., Kreps, S., Lall, R., Larter, O., Ó hÉigeartaigh, S., Staffell, S., & Villalobos, J. (2023). International Governance of Civilian AI: A Jurisdictional Certification Approach. *arxiv:2308.15514*.
9. Nie, A., **A. Reuel**, & Brunskill, E. (2023). Understanding the Impact of Reinforcement Learning Personalization on Subgroups of Students in Math Tutoring. *International Conference on Artificial Intelligence in Education*, pp. 688–694.
10. Schuett, J.*, **A. Reuel***, & Carlier, A. (2023). How to Design an AI Ethics Board. *AI & Ethics*.
11. Lamparth, M., & **A. Reuel** (2023) Analyzing And Editing Inner Mechanisms Of Backdoored Language Models. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*.
12. **A. Reuel**, Peralta, S., Sedoc, J., Sherman, G., & Ungar, L. (2022). Measuring the Language

of Self-Disclosure across Corpora. *Findings of the 60th Annual Meeting of the Association for Computational Linguistics 2022*.

13. **A. Reuel**, Koren, M., Corso, A., & Kochenderfer, M. (2021). Using Adaptive Stress Testing to Identify Paths to Ethical Dilemmas in Autonomous Systems. *Proceedings of the AAAI-22 Workshop on Artificial Intelligence Safety*.

Teaching Experience

- Head Teaching Assistant, CIS 5220 Deep Learning, University of Pennsylvania, 2022
- Teaching Assistant, NETS 213, Crowdsourcing and Human Computation, University of Pennsylvania, 2021

Service

- Stanford Computer Science PhD Student-Applicant Support Program Reviewer, 2023
- Fairness of Algorithms Working Group Member, ThinkTech e.V., Germany, 2021 – 2022
- Management & Strategy Society President, LSE Student Union, UK, 2015 – 2016
- Mentor for Careleavers, Brückensteine e.V. and others, Germany, 2014 – 2023
- Volunteer Teacher, Various NGOs, Honduras, Colombia & Ecuador, November 2011 – September 2017

Professional Affiliations

Member of the Association for Computing Machinery, Stanford Intelligent Systems Laboratory, Stanford Trustworthy AI Research, Stanford Institute for Human-Centered AI