

Andreas Haupt

Curriculum Vitæ

 a.haupt@me.com
 andyhaupt.com
 [indraos](#)
 [indraos](#)
 [Andreas Haupt](#)

Academic Degrees

09/2019-09/2024 **PhD in Engineering-Economic Systems**

Custom interdisciplinary degree with coursework in Economics and Computer Science
Fields: Microeconomic Theory, Market Design, Industrial Organization, Reinforcement Learning
Committee: Alessandro Bonatti, Dylan Hadfield-Menell, Eric Maskin, David Parkes
MIT Presidential Fellowship

10/2017-07/2019 **BS in Computer Science**, Goethe-Universität, Frankfurt, Germany

10/2012-08/2018 **MS in Economics, MS in Mathematics, BS in Mathematics**, Rheinische Friedrich-Wilhelms Universität, Bonn, Germany

Villigst Foundation Full Scholarship

Professional Experiences

10/2024-current **Human-Centered AI Postdoctoral Fellow**, Stanford Departments of Computer Science and Economics, {cs, economics, digitaleconomy, hai, siepr, crfm, ai}.stanford.edu

- working with Erik Brynjolfsson (Economics) and Sanmi Koyejo (Computer Science)
- teaching (Co-Primary Instructor) CS329H: Machine Learning from Human Preferences

06/2023-08/2023 **Fellow**, Federal Trade Commission, Washington D.C., USA, ftc.gov

- drafted background memos on European and U.S. regulatory and self-regulatory approaches to AI and effects of large-scale AI systems on competition (G7 and ICN)
- investigated consumer protection of emerging AI chatbots

01/2021-05/2021 **Trainee**, European Commission, Directorate-General for Competition, Unit C.6: Data Markets and Ecommerce, Brussels, Belgium, commission.europa.eu

- drafted sections of the voice assistants sector inquiry
- rebutted empirical and theoretical economic and technological arguments in the case against Amazon Inc. regarding Amazon Marketplace's data sharing with third-party sellers.

08/2017-06/2019 **High School Teacher**, Teach First Deutschland, Groß-Gerau, Germany, teachfirst.de

- Taught Computer Science and Mathematics at a high school with a professional focus.

Selected Talks

2025 NYU, Carnegie Mellon University, Lindau Nobel Laureate Meeting, Harvard Kennedy School

2024 Microsoft Research New England, Harvard Law School, far.AI

2023 Harvard Business School, Google Deepmind London, MPI Berlin, MPI Tübingen

Selected Honors

2025 Rising Stars in Data Science

2025 Lindau Grant, Lindau Nobel Laureate Meetings

2025 SAP Grant, Heidelberg Laureate Forum

2019 Presidential Fellowship (MIT)

Selected Publications

Ian Gemp, Andreas Haupt, et al. (July 2025). "Convex Markov Games: A New Frontier for Multi-Agent Reinforcement Learning". In: *Forty-second International Conference on Machine Learning*. URL: <https://openreview.net/forum?id=yIfCq03hsM>.

Andreas Haupt et al. (July 2025). "Position: AI Should Not Be An Imitation Game: Centaur Evaluations". In: *Forty-second International Conference on Machine Learning Position Paper Track*. URL: <https://openreview.net/forum?id=LkdH35003E>.

Haiwen Li, Soham De, Manon Revel, Andreas Haupt, et al. (Sept. 2025). "Scaling Human Judgment in Community Notes with LLMs". In: *Journal of Online Trust and Safety* 3.1. DOI: 10.54501/jots.v3i1.255. URL: <https://tsjournal.org/index.php/jots/article/view/255>.

Rylan Schaeffer, Joshua Kazdan, Yegor Denisov-Blanch, Brando Miranda, Matthias Gerstgrasser, Susan Zhang, Andreas Haupt, et al. (Dec. 2025). "Position: Machine Learning Conferences Should Establish a "Responses and Critiques" Track". In: *Thirty-ninth Annual Conference on Neural Information Processing Systems*. URL: <https://openreview.net/forum?id=DS1XSAPvKs>.

Sang Truong, Andreas Haupt, et al. (May 2025). *Machine Learning from Human Preferences*. Stanford University: Living Textbooks Initiative. URL: <https://mlhp.stanford.edu/>.

Stephen Casper, Carson Ezell, Charlotte Siegmann, Noam Kolt, Taylor Lynn Curtis, Benjamin Bucknall, Andreas Haupt, et al. (Feb. 2024). "Black-Box Access is Insufficient for Rigorous AI Audits". In: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '24. New York, NY, USA: Association for Computing Machinery, pp. 2254–2272. DOI: 10.1145/3630106.3659037.

Phillip Christoffersen, Andreas Haupt, et al. (Oct. 2024). "Formal contracts mitigate social dilemmas in multi-agent reinforcement learning". In: *Autonomous Agents and Multi-Agent Systems* 38.2. Previously appeared in AAMAS, Code available at <https://github.com/Algorithmic-Alignment-Lab/contracts>, p. 51. ISSN: 1573-7454. DOI: 10.1007/s10458-024-09682-5. URL: <https://doi.org/10.1007/s10458-024-09682-5>.

Andreas Haupt et al. (July 2024a). "Certification Design for a Competitive Market". In: *Proceedings of the 25th ACM Conference on Economics and Computation*. EC '24. New Haven, CT, USA: Association for Computing Machinery, p. 852. ISBN: 9798400707049. DOI: 10.1145/3670865.3673593.

— (Oct. 2024b). "Risk preferences of learning algorithms". In: *Games and Economic Behavior*. ISSN: 0899-8256. DOI: <https://doi.org/10.1016/j.geb.2024.09.013>. URL: <https://www.sciencedirect.com/science/article/pii/S089982562400143X>.

Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen McAleer, Andreas Haupt, et al. (July 2024). "Steering No-Regret Learners to a Desired Equilibrium". In: *Proceedings of the 25th ACM Conference on Economics and Computation*. EC '24. New Haven, CT, USA: Association for Computing Machinery, pp. 73–74. DOI: 10.1145/3670865.3673536.

Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen Marcus McAleer, Andreas Haupt, et al. (Dec. 2023). "Computing Optimal Equilibria and Mechanisms via Learning in Zero-Sum Extensive-Form Games". In: *Thirty-seventh Conference on Neural Information Processing Systems*. URL: <https://openreview.net/forum?id=yw1v4RqvPk>.

Andreas Haupt et al. (July 2022). "Contextually Private Mechanisms". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22. **Revise and Resubmit at the American Economic Review**. Boulder, CO, USA: Association for Computing Machinery, p. 1144. ISBN: 9781450391504. DOI: 10.1145/3490486.3538259.

References

Erik Brynjolfsson

Professor of Economics
Stanford University
✉ erikb@stanford.edu

Sanmi Koyejo

Professor of Computer Science
Stanford University
✉ sanmi@cs.stanford.edu

David Parkes

Professor of Computer Science
Harvard University
✉ parkes@seas.harvard.edu

Sendhil Mullainathan

Professor of Economics & EECS
Massachusetts Institute of Technology
✉ sendhil@mit.edu

Eric Maskin

Professor of Economics & Mathematics
Faculty of Arts and Sciences
Harvard University
✉ emaskin@fas.harvard.edu

Dylan Hadfield-Menell

Professor of AI and Decision-Making
Department of EECS
Massachusetts Institute of Technology
✉ dhm@csail.mit.edu

Selected Organized and Moderated Events

- | | |
|-----------------------------|---|
| Human-AI Collaboration | Workshop at ACM Economics and Computation (2025) |
| Algorithms & Centralization | Weeklong Seminar for the Villigst Foundation in Germany, jointly with Zoë Hitzig, OpenAI (2024) |
| EU AI Act | "Technical Perspectives on the EU AI Act", with Gabriele Mazzini, AI Act team, European Commission; Boaz Barak, Harvard University; Sarthak Agarwal, former OpenAI (2023) |
| Audits | "Algorithmic Audits in Economic Contexts", with Cathy O'Neil, ORCAA; Christo Wilson, Northeastern University; Alex Mackay, Harvard Business School (2022) |

Selected Leadership

2024-26 Stanford HAI Global AI Regulation Affinity Group, Co-Chair
2022-23 MIT Science Policy Initiative, Vice-President
2021-23 MIT AI Ethics & Policy Group, Co-Chair