Andreas Haupt

AI Institute Fellow-in-Residence, Schmidt Sciences
Digital Fellow, Stanford Digital Economy Lab
Technology and Human Rights Fellow, Harvard Kennedy School

h4upt@stanford.edu andyhaupt.com Scholar indraos indraos andreas_h0wpt

I study how algorithmic systems represent and are evaluated against diverse human preferences, using methods from microeconomic theory, structural econometrics, and reinforcement learning.

Appointments

2026–2027 **Schmidt Sciences** **AI Institute Fellow-in-Residence**
AI Institutes

2026–2027 **Stanford Digital Economy Lab** **Digital Fellow**
Stanford Institute for Human-Centered AI

2026–2027 **Harvard Kennedy School** **Technology and Human Rights Fellow**
Harvard Kennedy School

2024–2026 **Stanford University** **Human-Centered AI Fellow; Stanford Impact Labs Fellow**
Departments of Computer Science and Economics
HAI, DEL, StatsML, CRFM, STAIR
- Worked with Erik Brynjolfsson (Economics) and Sanmi Koyejo (Computer Science)

Education

2019–2024 **MIT** **Ph.D., Engineering-Economic Systems**
Computer Science and AI Lab; degree conferred February 2025.
Committee: Alessandro Bonatti, Dylan Hadfield-Menell, Eric Maskin, David Parkes.
Fields: microeconomic theory, market design, industrial organization, reinforcement learning.
Dissertation: *The Economic Engineering of Personalized Experiences*.

2017–2019 **University of Frankfurt** **B.S., Computer Science**

2015–2018 **University of Bonn** **M.S., Economics (with distinction)**

2014–2017 **University of Bonn** **M.S., Mathematics**
Year 1 at EPF Lausanne

2012–2014 **University of Bonn** **B.S., Mathematics**

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

Eric Maskin
Professor of Economics and Mathematics
Harvard University
emaskin@fas.harvard.edu

Helen Nissenbaum
Professor of Information Science
Cornell Tech
hn288@cornell.edu

Dylan Hadfield-Menell
Professor of AI and Decision-Making
MIT
dhm@csail.mit.edu

Honors and Awards

2026	Best AI Scientist Award, ICML Workshop on AI for Science (for *Bruno*)
2025	Rising Star in Data Science
2025	Young Scientist, 8th Economics Lindau Nobel Laureate Meeting
2025	Young Researcher, 12th Heidelberg Laureate Forum
2024	Price Theory Summer Camp, Becker Friedman Institute
2019	Presidential Fellowship, MIT
2012–2018	Full Scholarship, Villigst Foundation

Publications

‡ indicates equal contribution or alphabetical author order, † equal supervision.

Welfare, Improvability, and Variance: A Principal-Agent Approach to Optimal Benchmark Item Aggregation
A. Haupt‡, J. Hartenstein‡, A. Reuel, M. Kochenderfer, S. Koyejo
Advances in Neural Information Processing Systems (Evaluations and Datasets Track), 2026.
AI Evaluation · Aggregation · Market Design · PDF · HTML

Reward Bias Substitution: Single-Axis Bias Mitigations Redirect Optimization Pressure
M. Lamparth, D. Fein, **A. Haupt**, M. Hussing, M.J. Kochenderfer
Advances in Neural Information Processing Systems, 2026.
Preferences · PDF

Position: AI Development Should Prioritize Cognitive Security
B. El, S. Su, A. Pappu, P. Yin, J. Heng, E. Heng, R.Z. Wang, **A. Haupt**†, J. Zou†
Advances in Neural Information Processing Systems (Position Paper), 2026.
Preferences · PDF

Contextually Private Mechanisms
A. Haupt‡, Z. Hitzig‡
American Economic Review 116 (9), p. 3223-3262 (Lead Article), 2026.
Privacy · Market Design · PDF · HTML

Machine Learning from Human Preferences
S. Truong‡, **A. Haupt**‡, S. Koyejo‡
Princeton University Press (forthcoming)
Preferences · HTML

Don't Walk the Line: Boundary Guidance for Filtered Generation
S. Ball, **A. Haupt**
International Conference on Machine Learning, 2026.
Multi-Agent · PDF

Bruno: An AI Product Manager for Scientists
A. Haupt‡, P. Ranganathan‡
ICML Workshop on AI for Science (Best AI Scientist Award), 2026.
Work · HTML · VIDEO

The Collapse of Heterogeneity in Silicon Philosophers
Y. Shi, **A. Haupt**
ACM Conference on Fairness, Accountability, and Transparency, 2026.
AI Evaluation · Preferences · PDF · CODE

Discussion of "The ICML 2023 Ranking Experiment" by Su et al.
A. Haupt, S. Koyejo
Journal of the American Statistical Association 121 (554), p. 853-854, 2026.
AI Evaluation · Aggregation

Position: ML Conferences Should Establish a Refutations and Critiques Track
R. Schaeffer, J. Kazdan, Y. Denisov-Blanch, B. Miranda, M. Gerstgrasser, S. Zhang, **A. Haupt**, I. Gupta, E. Obbad, J. Dodge, et al.
Advances in Neural Information Processing Systems (Position Paper), 2025.
AI Evaluation · PDF

Scaling Human Judgment in Community Notes with LLMs
H. Li, S. De, M. Revel, **A. Haupt**, B. Miller, K. Coleman, J. Baxter, M. Saveski, M.A. Bakker
Journal of Online Trust and Safety 3 (1), 2025.
AI Evaluation · Aggregation · Platforms · PDF

AI should not be an imitation game: Centaur evaluations
A. Haupt, E. Brynjolfsson
International Conference on Machine Learning (Position Paper), 2025.
AI Evaluation · Work · PDF

Convex Markov Games: A New Frontier for Multi-Agent Reinforcement Learning
I. Gemp, **A. Haupt**, L. Marris, S. Liu, G. Piliouras
International Conference on Machine Learning, 2025.
Multi-Agent PDF

Risk Preferences of Learning Algorithms
A. Haupt[‡], A. Narayanan[‡]
Games and Economic Behavior 148, p. 415-426, 2024.
Multi-Agent Preferences PDF

Formal Contracts Mitigate Social Dilemmas in Multi-Agent Reinforcement Learning
A. Haupt[‡], P. Christoffersen[‡], M. Damani, D. Hadfield-Menell
Autonomous Agents and Multi-Agent Systems 38, Article 51, 2024.
Multi-Agent Market Design PDF CODE

Steering No-Regret Learners to a Desired Equilibrium
B.H. Zhang, G. Farina, I. Anagnostides, F. Cacciamani, S.M. McAleer, **A. Haupt**, A. Celli, N. Gatti, V. Conitzer, T. Sandholm
ACM Conference on Economics and Computation, 2024.
Multi-Agent Market Design PDF

Certification Design for a Competitive Market
A. Haupt[‡], N. Immorlica[‡], B. Lucier[‡]
ACM Conference on Economics and Computation, 2024.
Market Design PDF

Black-Box Access is Insufficient for Rigorous AI Audits
S. Casper, C. Ezell, C. Siegmann, N. Kolt, T.L. Curtis, B. Bucknall, **A. Haupt**, K. Wei, J. Scheurer, M. Hobbhahn, et al.
ACM Conference on Fairness, Accountability, and Transparency, 2024.
AI Evaluation PDF

Computing Optimal Equilibria and Mechanisms via Learning in Zero-Sum Extensive-Form Games
B. Zhang, G. Farina, I. Anagnostides, F. Cacciamani, S. McAleer, **A. Haupt**, A. Celli, N. Gatti, V. Conitzer, T. Sandholm
Advances in Neural Information Processing Systems, 2023.
Multi-Agent Market Design PDF

Understanding Multi-Homing and Switching by Platform Drivers
X. Guo[‡], **A. Haupt**[‡], H. Wang, R. Qadri, J. Zhao
Transportation Research Part C: Emerging Technologies 154, 2023.
Platforms Work PDF

Towards Psychologically-Grounded Dynamic Preference Models
M. Curmei[‡], **A. Haupt**[‡], B. Recht, D. Hadfield-Menell
ACM Conference on Recommender Systems, 2022.
Preferences PDF

The Optimality of Upgrade Pricing
D. Bergemann[‡], A. Bonatti[‡], **A. Haupt**[‡], A. Smolin[‡]
Web and Internet Economics, 2021.
Market Design PDF

Working Papers and Preprints

TaskNet: Building Occupational Task Inventories from Worker Interviews at Scale
J. Wang[‡], **A. Haupt**[‡], I. Robinson, P. Lewis, E. Brynjolfsson[†], D. Yang[†]
SSRN Preprint 7670419, 2026.
Work AI Evaluation PDF

AISPA: User-Centric System Prompt Auditing for Large Language Model Applications
X. Lin, S. Zhu, S. Yang, Z. Zhang, H. Zhang, Y. Zhao, C. Qian, T. Wang, Z. Zhang, Z. Yuan, D. Wu, J. Wu, Y. Si, J. Liu, B. Bi, R. Mahari, T. South, D. Greenwood, Z. He, R. Bommasani, S. Kazinnik, **A. Haupt**, S. Marro, E. Brynjolfsson, A. Pentland, J. Pei
Preprint, 2026.
AI Evaluation Platforms PDF

General Preference Reinforcement Learning
M. Umer, M.A. Mohsin, A. Bilal, A. Chaudhry, **A. Haupt**, S. Koyejo, E. Fox, J.M. Cioffi
Preprint, 2026.
Preferences PDF

Latent Adversarial Regularization for Offline Preference Optimization
E. Jiang, Y.J. Zhang, Y. Xu, **A. Haupt**, N. Amato, S. Koyejo
Preprint, 2026.
Preferences PDF

Preference Measurement Error, Concentration in Recommendation Systems, and Persuasion
A. Haupt
Allied Social Science Associations Meeting, 2025.
Preferences Platforms PDF

The Self-Preferencing Prohibition as a Fairness Provision
A. Haupt
ACM Symposium on Computer Science and Law (Poster), 2025.
Platforms Market Design PDF

Platform Preferencing and Price Competition I: Evidence from Amazon
O. Hartzell[‡], **A. Haupt**[‡]
SSRN Preprint 5126918, 2025.
Platforms PDF

Recommending to Strategic Users
A. Haupt, D. Hadfield-Menell, C. Podimata
Symposium on the Foundations of Responsible Computing, 2023.
Preferences Platforms PDF

Opaque Contracts
A. Haupt, Z. Hitzig
Preprint, 2023.
Privacy Market Design PDF

Theses

The Economic Engineering of Personalized Experiences
A. Haupt
Ph.D. Dissertation, Massachusetts Institute of Technology, 2025.
Preferences PDF

Multi-Agent Influence Diagrams and Commitment
A. Haupt
B.S. Thesis, Goethe Universität Frankfurt, 2019.
Multi-Agent PDF

Voting with Restricted Communication
A. Haupt
M.S. Thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2018.
Market Design Aggregation PDF

A Data Application of Graphon Theory
A. Haupt
M.S. Thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2017.
Platforms PDF

Die Integrality Ratio der Subtour-Relaxierung
A. Haupt
B.S. Thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2014.
PDF

Organized Events

SimulacraBench
Organizers: **A. Haupt**
Conference on Neural Information Processing Systems, Sydney, 2026. AI Evaluation HTML

The Future of Authorship
Organizers: **A. Haupt**, K. Neubert, E. Brynjolfsson
Stanford Digital Economy Lab, Stanford, November 2026. Work

LLM Incentives
Organizers: Y. Alimohammadi, K. Donahue, S. Fish, **A. Haupt**, C. Li, G. Mantegazza
ACM Conference on Economics and Computation, Rome, July 2026. Market Design HTML

H+AI@EC: Workshop on Human-AI Collaboration
Organizers: **A. Haupt**
ACM Conference on Economics and Computation, Stanford University, July 2025. Work HTML

Algorithms and Centralization
Organizers: **A. Haupt**, Z. Hitzig
Weeklong seminar, Villigst Foundation, Germany, 2024. Market Design Platforms HTML

Technical Perspectives on the EU AI Act
Organizers: **A. Haupt**
MIT AI Ethics and Policy Group, with G. Mazzini (European Commission), B. Barak (Harvard), and S. Agarwal, 2023. AI Evaluation HTML

Algorithmic Audits in Economic Contexts
Organizers: **A. Haupt**
MIT AI Ethics and Policy Group, with C. O'Neil (ORCAA), C. Wilson (Northeastern), and A. MacKay (Harvard Business School), 2022. AI Evaluation Platforms HTML

Policy Experience

2023 **Federal Trade Commission** **Summer Fellow**
Office of International Affairs
Bureau for Competition
- Drafted background memos on European and U.S. regulatory and self-regulatory approaches to AI and on effects of large-scale AI systems on competition (G7 and ICN)
- Investigated consumer protection issues of emerging AI chatbots

2021 **European Commission** **Atypical Trainee**
Directorate-General for Competition, Unit C.6: Data Markets and E-commerce
- Drafted sections of the sector inquiry into voice assistants
- Rebutted economic and technological arguments in the case against Amazon regarding Marketplace data sharing with third-party sellers

Teaching

2024–2026 **Stanford University** **Co-Primary Instructor**
CS329H: Machine Learning from Human Preferences. Textbook: *Machine Learning from Human Preferences* (with S. Truong and S. Koyejo), forthcoming with Princeton University Press.

2017–2019 **Berufliche Schulen Gross-Gerau** **Teach First Deutschland Fellow**
Full-time high school teaching of mathematics and computer science

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

Leadership and Service

2026–2027 | Area Chair, ICLR, ICML, and AAAI
2024–2026 | Co-Chair, Stanford HAI Affinity Group on Global AI Governance
2021–2023 | Co-Chair, MIT AI Ethics and Policy Group
2022–2023 | Vice-President, MIT Science Policy Initiative