

Joe Edelman

Contact Information

- Email: joe@meaningalignment.org
- Location: Berlin, Germany
- Twitter: @edelwax

Research Focus

- Philosophy of choice and agency
- Mechanism design, social choice, and AI alignment
- Human values elicitation for AI systems
- Metrics design for meaningful social interactions
- Impact of technology on social fabric and individual meaning

Selected Publications and Talks

- Edelman, J., Klingefjord, O., & Lowe, R. (2024). What are Human Values, and How Do We Align AI to them? <https://arxiv.org/abs/2404.10636>
- Edelman, J. (2023). Practical Techniques to Align Markets with Meaning. <https://www.youtube.com/watch?v=hZpKdfbrd6o>
- Edelman, J. (2022). Values, Preferences, Meaningful Choice. <https://github.com/jxe/vpm/blob/master/vpm.pdf>
- Edelman, J. (2021). Values-Based Data-Science and Design. <https://textbook.sfsd.io>
- Edelman, J. (2017). Nothing to be Done. <https://medium.com/what-to-build/nothing-to-be-done-bfe2ce71a3a2>
- Edelman, J. (2016). Is Anything Worth Maximizing. <https://nxhx.org/maximizing>
- Edelman, J. (2014). Choicemaking and the Interface. <https://nxhx.org/Choicemaking/>

Selected Professional Experience

Meaning Alignment Institute

Co-founder, Principal Investigator | Jun 2023 - Present Berlin, Germany

- Lead research initiatives on aligning AI with human values
- Develop methodologies for eliciting and synthesizing human values into training models for AI
- Pioneer democratic processes for collecting values from diverse groups to build "moral graphs" for policy-making and LLM tuning

School for Social Design

Co-Founder | Jan 2017 - 2022 Berlin, Germany (Part-time)

- Wrote "Values-Based Data Science and Design" textbook
- Developed curriculum for values-based design and metrics
- Trained metrics and user research people at all major tech cos on values-based data science

Independent Research Funded by Patrick Collison and Stripe

Independent Researcher | 2015 - 2020

- Delivered the influential talk "Is Anything Worth Maximizing?", which led to recommender system changes at Facebook, Apple, and Google.
- Developed the philosophical foundations of new business metrics and design methods.

Center for Humane Technology (CHT)

Co-founder | 2013 - 2015

- Coined the term "Time Well Spent" and developed related metrics adopted by teams at Facebook, Google, and Apple
- Designed what later became "Screen Time" and "Downtime" in iOS.
- Briefed policymakers on trends towards political polarization, social isolation, declining trust, and internet addiction.

CouchSurfing International

Reputation Data / Company Metrics | 2006 - 2008 New Zealand

- Developed meaning-based product-success metrics, which became the north star for the organization and the ultimate judge of feature roll-outs
- Contributed to almost 10 million real-world positive experiences between strangers, and 400,000 close friendships
- Developed community and trust algorithms, reputation system, etc

Education

- **B.A. in Computer Science & Modeling**
 - Institution: Dartmouth College
 - Location: Hanover, NH, USA
 - Year: 1999

Influential Collaborations and Mentors

- Tech: Alan Kay, Terry Winograd, Bill Verplank (Interval Research), Casey Fenton (CouchSurfing), Howie Shrobe, Marvin Minsky (MIT), Bret Victor, Rob Ochshorn