Probabilistic and Interactive Machine Learning

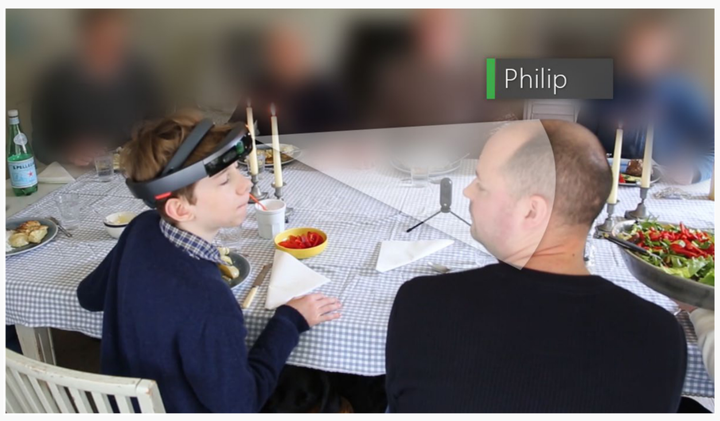
Sebastian Tschiatschek

Motivation & Goals

Efficient & seamless collaboration with intelligent agents

Challenge: Collaboration in the face of

- (significant) mismatch in inputs
- non-aligned goals and constraints
- (complex) large state spaces
- constraints on sample complexity
- inaccurate mental models



Three key research directions

- Reinforcement Learning
  - Reward / constraint inference
  - Exploration & abstraction
- Interactive machine learning
- Probabilistic (generative) models

Reinforcement Learning & Inverse Reinforcement Learning

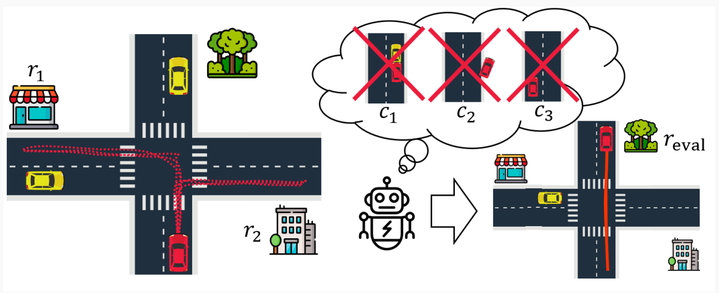
Specifying objectives is hard and the sample complexity for (naively) learning them is often prohibitive

Intelligent agents must leverage potential for generalization and actively seek relevant information

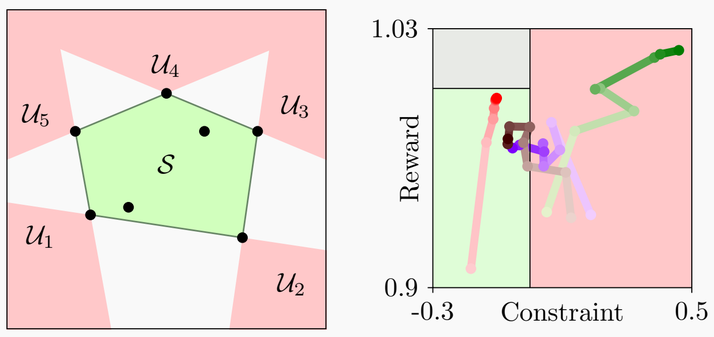
- Need to extend existing formalisms and problem settings to better reflect real-world challenges
- Enable generalization by appropriate choices of *objects that generalize* and learning about them
  ⏭ Requires tailored algorithms and architectures
- Seeking relevant information to learn quickly while enabling more elaborate modes of interaction
  ⏭ Information-directed learning and active information acquisition

Learning Constraints in CMDPs

Constraints might generalize better than rewards



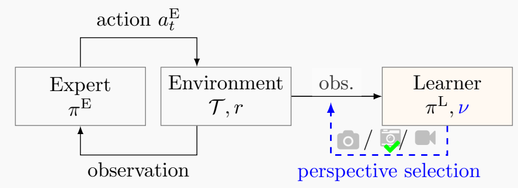
Efficiently learning about constraints and transferring this knowledge across environments



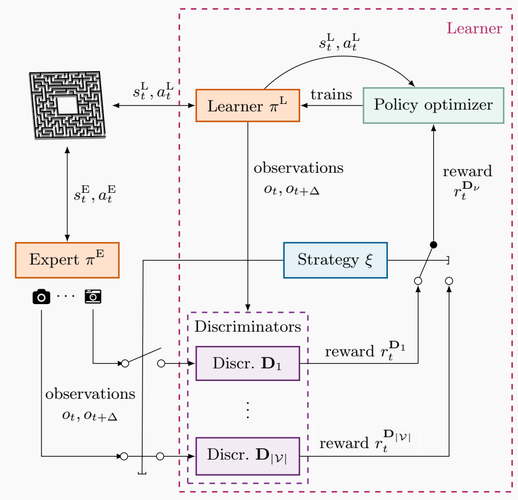
David Lindner, Xin Chen, Sebastian Tschiatschek, Katja Hofmann, Andreas Krause, *Learning safety constraints from demonstrations with unknown rewards*, AISTATS'24.

Active Third-person Imitation Learning

Leveraging different perspectives to faster learn about the reward (also relevant for LLMs)



GAIL based learning architecture



Timo Klein, Susanna Weinberger, Adish Singla, Sebastian Tschiatschek, *Active Third-Person Imitation Learning*. arXiv preprint arXiv:2312.16365, 2024

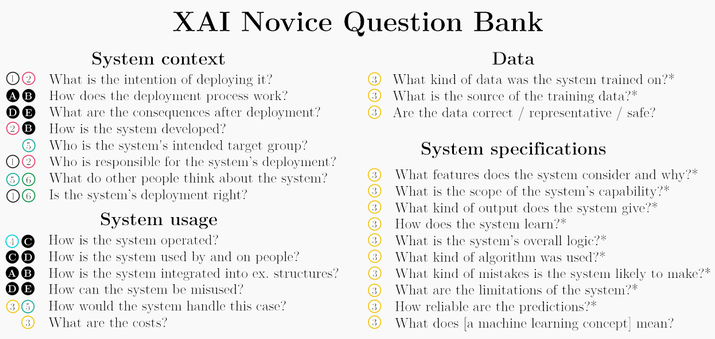
Interacting with People & Society

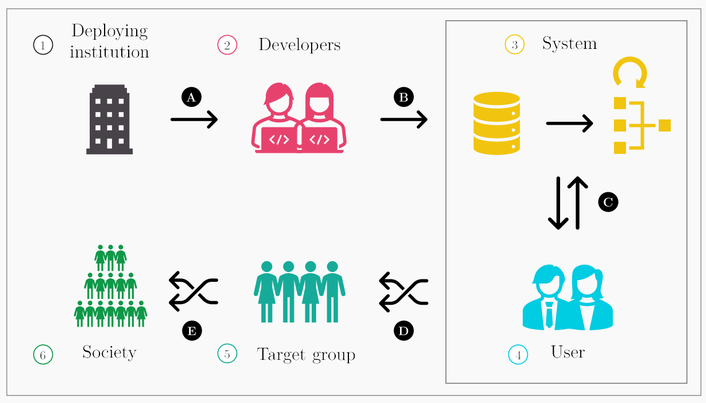
Algorithms' impact on people and society is increasing

AI design and development must account for the involved/affected human stakeholders

- Need to understand algorithms and to understand how people interact with algorithms
- Focus should not be put only on experts but all affected stakeholders
  ⏭ Different information needs
- Stakeholders need to understand relevant aspects of the socio-technical systems to take the right actions
  ⏭ Tailored explanations and accounting for uncertainty

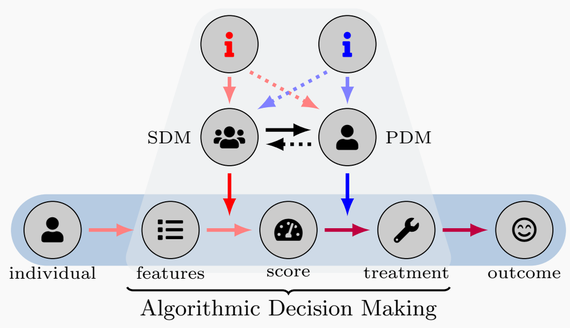
Information Needs of Non-technical Lay People

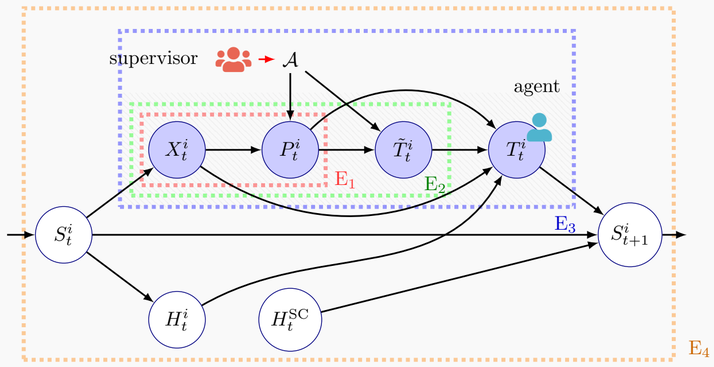




Timothee Schmude, Laura Koesten, Torsten Möller, Sebastian Tschiatschek. *Information That Matters: Exploring Information Needs of People Affected by Algorithmic Decisions*. arXiv preprint arXiv:2401.13324, 2024.

Challenging the Human-in-the-loop





Sebastian Tschiatschek, Eugenia Stamboliev, Mark Coeckelbergh, Laura Koesten, *Challenging the Human-in-the-loop in Algorithmic Decision-making*, Workshop on Humans, Algorithmic Decision-Making and Society @ ICML'24

universität wien