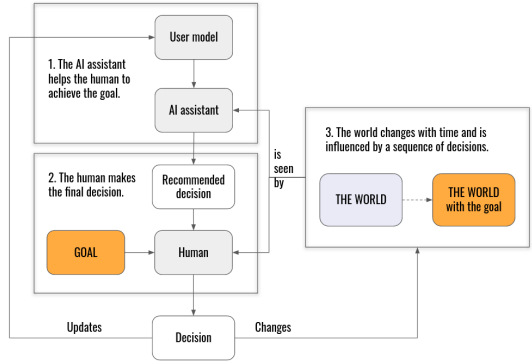
# Collaborative modelling, design and decision making with AI, Part II

Sebastiaan De Peuter, Alex Hämäläinen, and Elena Shaw

March 20th, 2023

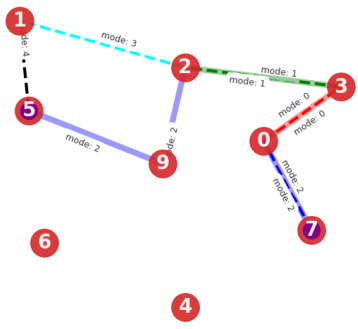
# Tutorial Agenda

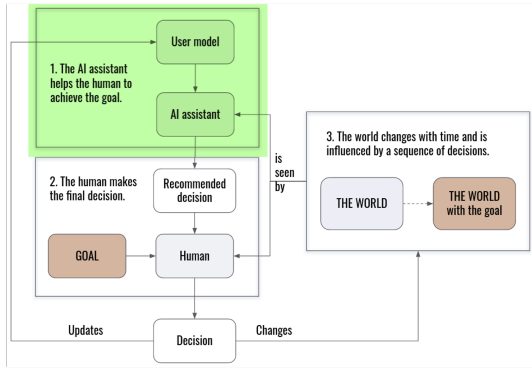
- Introduction to the techniques (20-30 mins)
  - The components
  - The task
- Hands-on Jupyter notebook (remaining time)
  - 3 coding exercises
  - Additional details



# Quick Overview

## The journey planning toy problem



# Drug Design Motivation

This tutorial focuses on multi-objective optimization, but AI-assistance can be more generally applied.

**Tasks with themes**:

- Navigating large combination spaces
- Multi-objective optimization
- Cold start/one-shot learning
- Integrating human/expert input
- Interpretability and explainability

**Applications**:

- Biomarker discovery  gene expression
- Drug synergy
- Computational pathology/tissue analysis
- Small-molecules drug design
  - sub-structure
  - binding sites

# Defining the world

## Overview

**Goal:** to convert the problem of interest into an MDP (or similar)

- Define the state $\mathcal{S}$ and action $\mathcal{A}$ spaces of the problem
- Define the transition $\mathcal{T}$ and objective $\mathcal{R}$ functions

$\mathcal{S}, \mathcal{A}$, and $\mathcal{T}$ are often tied to the task, while $\mathcal{R}$ is typically user-specific.

# Defining the world

## Use case: multimodal travel planning

An anonymous user wants to travel from city A to city B via a personalized journey itinerary catering to their travel preferences:

- A set of 10 locations, fixed starting and end points
  - The locations can be connected via different transportation modes $i \in [1, 5]$
  - Each represented by a graph $G_i = (V, E_i)$
- Different modes have different costs for different people
  - Distance, price, time

**Objective:** find an ordered list of transportation options between start and end points which are optimal w.r.t. the user preferences

# Defining the world

## Example: multimodal travel planning as an MDP

- State space: set of possible travel path configurations between start and end points
  - $\mathcal{S} = \bigcup_n E^n$, where $E = \bigcup_i E_i, \forall i \in [1, 5]$
- Action space: same as state space
  - $\mathcal{A} = \bigcup_n E^n$
- Transition function: updates the current path based on the new plan
  - $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$
- Reward function: user-specific preference weights over distance, price, and time
  - $\mathcal{R} : \mathcal{S} \times \mathbb{R}^3 \to \mathbb{R}$

# User interaction

## Overview

The interaction between our AI and user should be carefully designed. Efficient interaction will be helpful, while bad interaction will get in the way of the user

Examples of different interaction styles:

- Recommendation - reaction
- Joint decision
- Complementary
- Indirect
- etc.

# The User Model

## overview

*The user model is a generative model of human behavior, mapping the current state of the world, the user's parameters, etc... to actions.*

- Important to model interaction with the AI as well
  (e.g. reaction to recommendation)
- Can be used for:
  - *inferring* user parameters from real user data
  - *simulating* user behavior in future situations for training/planning
- Accuracy of the model is clearly important, but a perfect model is likely not needed.

# The User Model

## User Modeling Paradigms

- data-driven - lots of data needed
  - Treat **user modeling as a supervised learning problem**.
    Models used are NNs [7] and HMMs [10].
- model-driven - make assumption to require less data
  - **IRL** [1]
    Assumes people act optimally.
  - **Boltzmann rationality** [5]
    Assumes people are noisily optimal.
  - **information-theoretic bounded rationality** [6]
    Assumes humans trade off utility with the information cost of representing their policy.
  - **computational rationality** [8, 2]
    Assumes human behavior is optimal under cognitive limitations.

# The User Model

## Boltzmann rationality applied to our use case

### Bounded rationality heuristic

$dim(\mathcal{A}) \approx 16.000$, but we bound this by 2 assumptions:

- users have a strong aversion to exactly 1 segment of the journey, e.g. routing via a specific city
- users can propose up to 2 changes to avoid that segment

### Within the user's action space $\mathcal{A}_{user}$:

The utility $\mathcal{U}$ view

$$a^* = \arg\max_a \exp\{\beta \cdot \mathcal{U}(a)\}$$

$$Pr(a;\beta) = \frac{\exp(\beta \cdot \mathcal{U}(a))}{\sum_{a \in \mathcal{A}} \exp(\beta \cdot \mathcal{U}(a))}$$

The cost $\mathcal{C}$ view

$$a^* = \arg\min_a \exp\{-\beta \cdot \mathcal{C}(a)\}$$

$$Pr(a;\beta) = \frac{\exp(-\beta \cdot \mathcal{C}(a))}{\sum_{a \in \mathcal{A}} \exp(-\beta \cdot \mathcal{C}(a))}$$

# The Assistance Problem

## Assistance as a decision problem

To find an optimal policy for the assistant, we define it's task as a sequential decision problem.

- Action space $\mathcal{A}$: actions available to the assistant, including
  - interactions with the user (e.g. recommendations)
  - potentially actions within the world
- State space $\mathcal{S}$: including
  - state of the world (as defined by the task)
  - user parameters and state (unobserved!)
- Transition function $\mathcal{T}$
- Reward function $\mathcal{R}$: user's reward function + potentially other factors

# The Assistance Problem

## Assistance as a decision problem

What kind of sequential decision problem? Depends on the user model and the world.

- Assume user has **no internal state** (reactive)
  - Generalized Hidden Parameter MDP [9, 4] *(only if world is observable)*
    Goal and other factors are hidden parameters of MDP
  - POMDP [13]
    Goal and other factors are hidden part of state
- user has **internal state, but we know how it changes** over time
  - POMDP [13, 12]
    Goal and other factors are hidden part of state
- user has **internal state and we don't know how it changes** over time
  - Bayes-Adaptive POMDP [11, 3]
    Goal and other factors are hidden part of state. Transition function over hidden part of state is uncertain.

# The Assistance Problem

## Assistance in multimodel travel planning

Our toy implementation takes a very simplistic approach to the MDP. From the perspective of the AI,

- Action space $\mathcal{A}$:
  - AI $\mathcal{A}_{AI}$: all valid paths, approximately 16.000
  - User $\mathcal{A}_{user}$: paths that improve the most aversive (i.e. costly) path based on *hidden* preference parameters $\beta^*$

# The Assistance Problem

## Assistance in multimodel travel planning

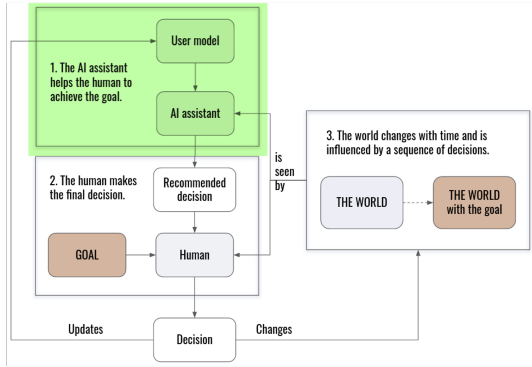
Our toy implementation takes a very simplistic approach to the MDP. From the perspective of the AI,

- Action space $\mathcal{A}$:
  - AI $\mathcal{A}_{AI}$: all valid paths, approximately 16.000
  - User $\mathcal{A}_{user}$: paths that improve the most aversive (i.e. costly) path based on *hidden* preference parameters $\beta^*$
- State space $\mathcal{S}$:
  - World: same as $\mathcal{A}_{AI}$
  - User: posterior distribution over user parameters given observed user action $Pr(\beta|a_{user})$
- Transition function $\mathcal{T}$:
  - World: completely determined by AI action
  - User: completely determined by AI action + posterior $Pr(\beta|a_{user})$
- Reward function $\mathcal{R}$: utility/cost based on user model

# The Assistance Problem

## Optimal assistance in multimodel travel planning

- Tasks 1 + 2: User model components necessary for AI to learn
  - Task 1: Model user behavior as a Boltzmann cost
  - Task 2: Implement posterior inference engine to reason about user preferences based on observed actions
- Task 3: Implement AI-user interaction to simulate how the two agents solve the journey planning task



# Time to code!

Questions?

# Time to code!

Questions?
In groups of 2-4, work through the notebook together.
We're here to help with questions.

# References


[1] Saurabh Arora and Prashant Doshi. "A survey of inverse reinforcement learning: Challenges, methods and progress". In: *Artificial Intelligence* 297 (2021). DOI: 10.1016/j.artint.2021.103500.

[2] Frederick Callaway et al. "Rational use of cognitive resources in human planning". In: *Nature Human Behaviour* 6.8 (2022), pp. 1112–1125.

[3] Mustafa Mert Çelikok, Frans A Oliehoek, and Samuel Kaski. "Best-Response Bayesian Reinforcement Learning with Bayes-adaptive POMDPs for Centaurs". In: *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. Richland, South Carolina, USA: International Foundation for Autonomous Agents and Multiagent Systems, 2022, pp. 235–243.

[4] Sebastiaan De Peuter and Samuel Kaski. "Zero-Shot Assistance in Sequential Decision Problems". In: *Thirty-Seventh AAAI Conference on Artificial Intelligence*. To appear. arXiv:2202.07364. Austin, Texas, USA: AAAI Press, 2023.

[5] Owain Evans and Noah D Goodman. "Learning the preferences of bounded agents". In: *NIPS Workshop on Bounded Optimality*. Vol. 6. 2015, pp. 2–1.

[6] Tim Genewein et al. "Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle". In: *Frontiers in Robotics and AI* 2 (2015), p. 27.

[7] Christopher McComb, Jonathan Cagan, and Kenneth Kotovsky. "Capturing Human Sequence-Learning Abilities in Configuration Design Tasks Through Markov Chains". In: *Journal of Mechanical Design* 139.9 (July 2017). DOI: 10.1115/1.4037185.

[8] Antti Oulasvirta, Jussi PP Jokinen, and Andrew Howes. "Computational Rationality as a Theory of Interaction". In: *CHI Conference on Human Factors in Computing Systems*. New York, New York, USA: Association for Computing Machinery, 2022, pp. 1–14. DOI: doi.org/10.1145/3491102.3517739.

[9] Christian Perez, Felipe Petroski Such, and Theofanis Karaletsos. "Generalized Hidden Parameter MDPs: Transferable Model-Based RL in a Handful of Trials". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. AAAI Press, 2020, pp. 5403–5411.

[10] Ayush Raina, Christopher McComb, and Jonathan Cagan. "Learning to Design From Humans: Imitating Human Designers Through Deep Learning". In: *Journal of Mechanical Design* 141.11 (Sept. 2019). DOI: 10.1115/1.4044256.

[11] Stéphane Ross et al. "A Bayesian Approach for Learning and Planning in Partially Observable Markov Decision Processes". In: *Journal of Machine Learning Research* 12.48 (2011), pp. 1729–1770.

[12] Rohin Shah et al. "Benefits of Assistance over Reward Learning". In: *Workshop on Cooperative AI (Cooperative AI @ NeurIPS 2020)* (2020).

[13] Matthijs TJ Spaan. "Partially Observable Markov Decision Processes". In: *Reinforcement Learning*. Berlin, Germany: Springer, 2012, pp. 387–414. ISBN: 978-3-642-27645-3. DOI: 10.1007/978-3-642-27645-3_12.