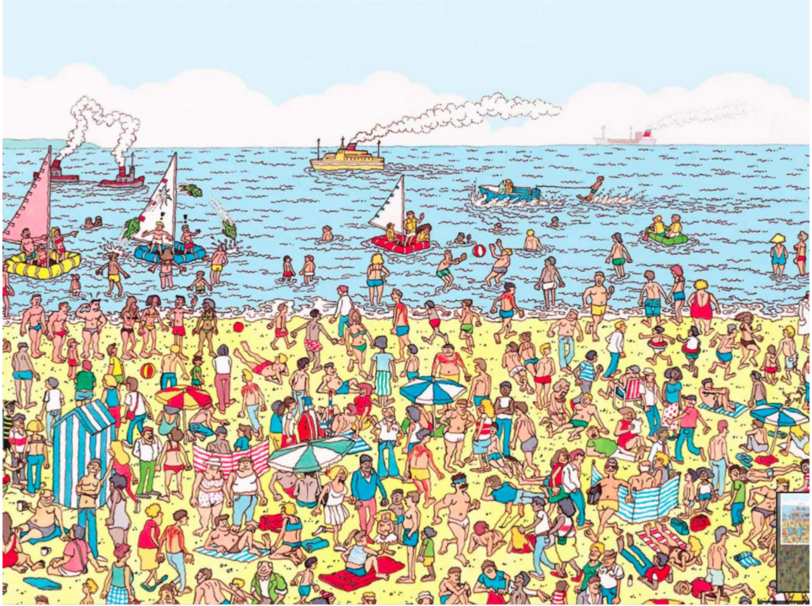
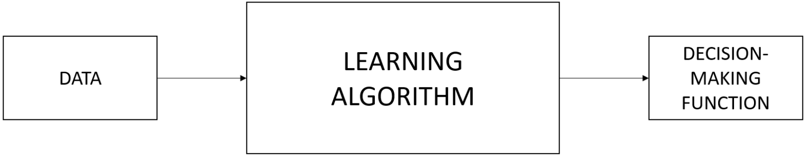
The human matters!
Perspectives on a human-centered view of Machine Learning

Prof. Pierre-Alexandre Murena

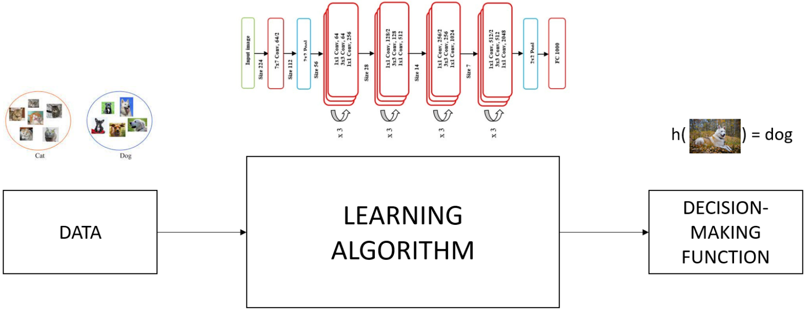
September 25, 2023



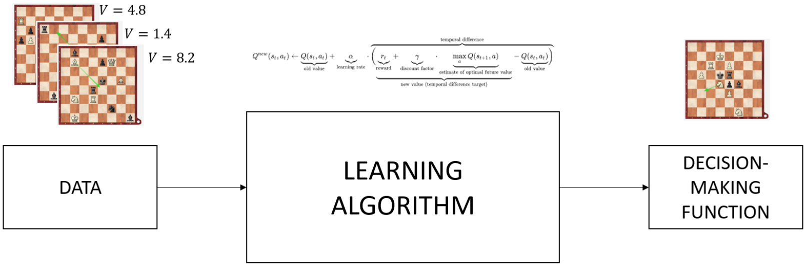
Machine Learning: General Pipeline



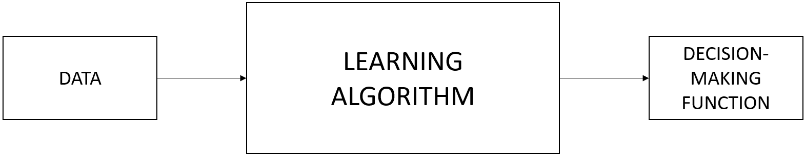
An Example: Classification



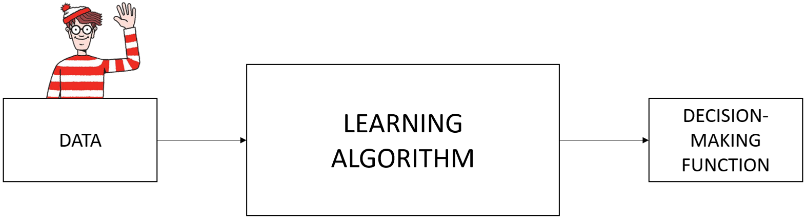
An Example: Reinforcement Learning



Machine Learning: General Pipeline



The human behind the data...



The human behind the data...

Data was created by humans.

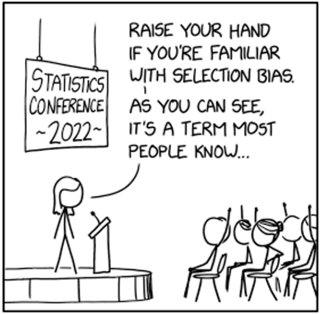
proud and heroic teenage man, photorealistic 4k

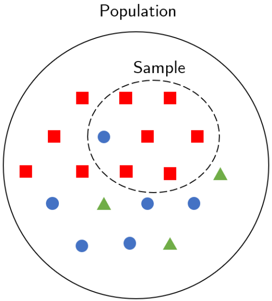
proud and heroic first nations teenage man, photorealistic 4k

(Example provided by David Pledger and Tony Briggs)

The human behind the data...

Data was collected by humans.

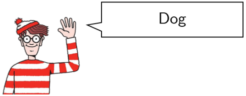




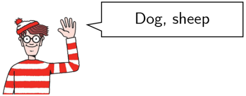
The human behind the data...

Data was annotated by humans.

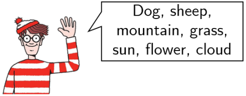
Task 1: What is the main subject?



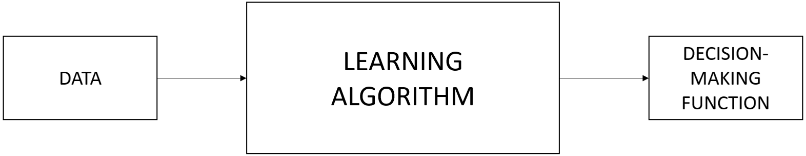
Task 2: Which animals do you see?



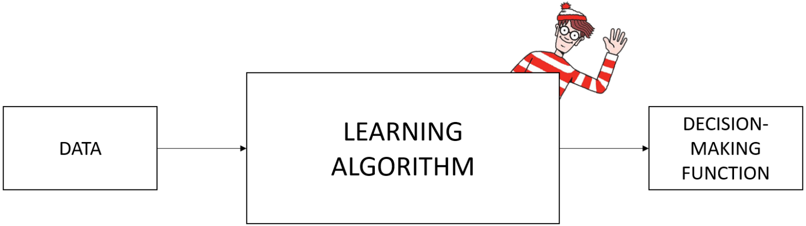
Task 3: What do you see?



Machine Learning: General Pipeline

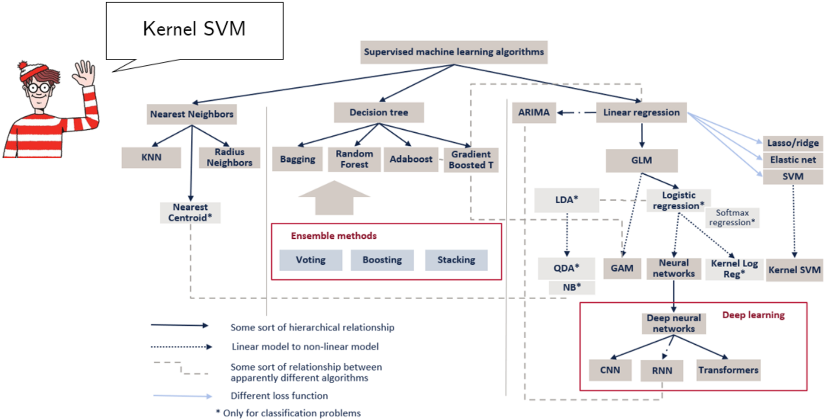


The human behind the learning algorithm...

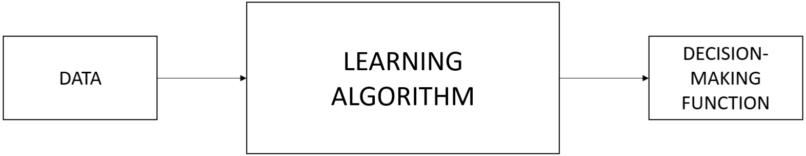


The human behind the learning algorithm...

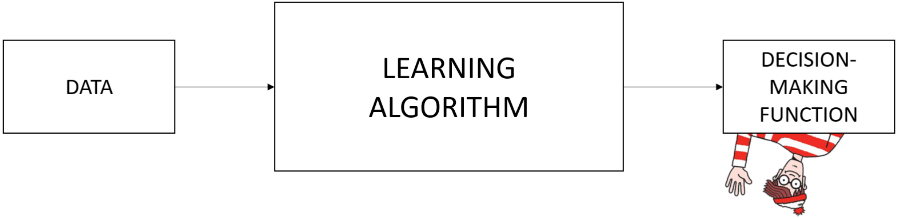
The learning algorithm was curated by the human user.



Machine Learning: General Pipeline

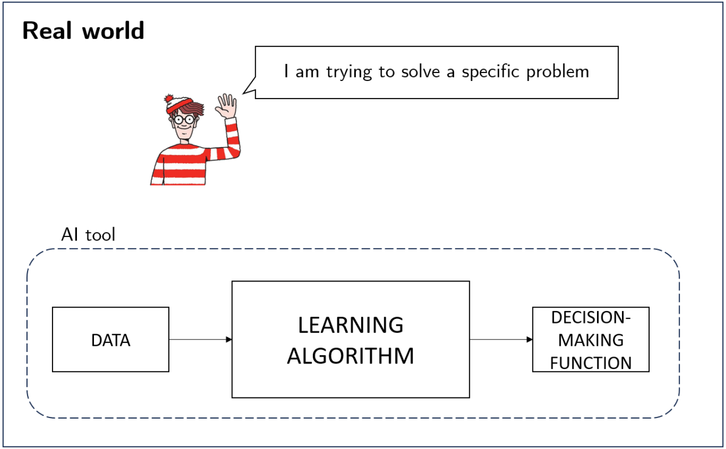


The human behind the decision-making function...



The human behind the decision-making function...

The decision-making will be used by the user to solve a problem in the real world.



The specific problem of the user has consequences on the data collection.

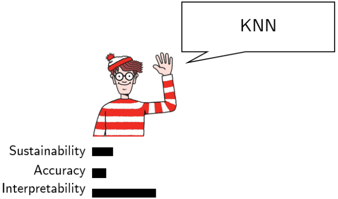
I would like to understand the impact of temperature over the conductivity

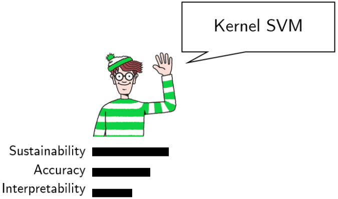
...

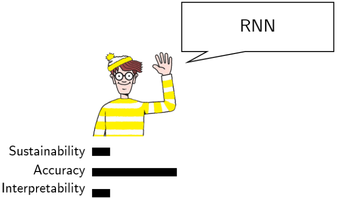
EXPERIMENTAL DESIGN

The human behind the learning algorithm...

The specific problem of the user has consequences on the choice of the algorithm







So what?

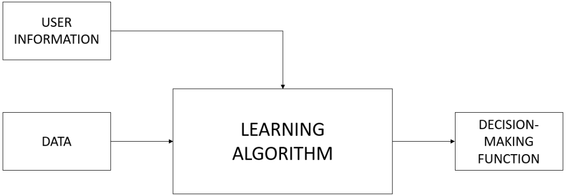
So what?

We cannot consider machine learning pipeline indepently from the human user.

How to bring the human back in the loop?

Solution 1: Adapt the algorithms to consider the presence of the human

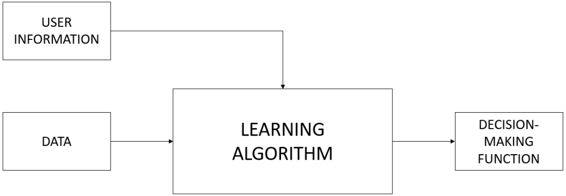
The AI algorithm takes the same inputs + additional information about the user



How to bring the human back in the loop?

Solution 1: Adapt the algorithms to consider the presence of the human

The AI algorithm takes the same inputs + additional information about the user

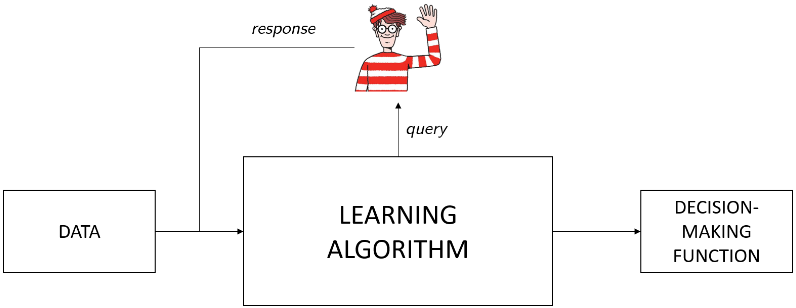


Usually not possible, because:

- Detecting biases requires additional information
- Learning algorithms are *objective*
- Requires from the user to provide all detailed information

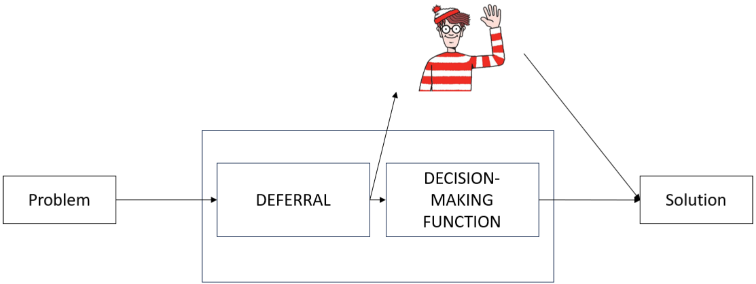
How to bring the human back in the loop?

Solution 2: Invoking the human when needed
The AI agent works independently but requests the service of the human user.



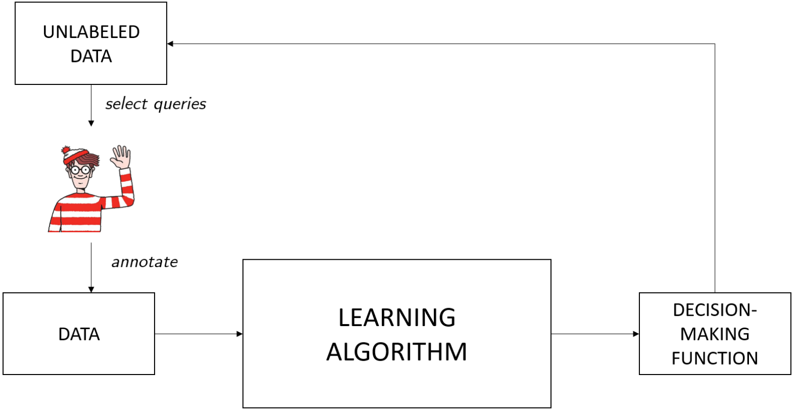
How to bring the human back in the loop?

Decision deferral



How to bring the human back in the loop?

Active learning



How to bring the human back in the loop?

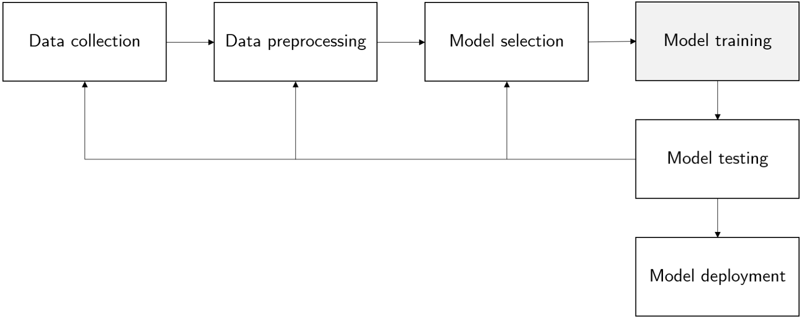
Solution 2: Invoking the human when needed

Problem: The human user is not in control.

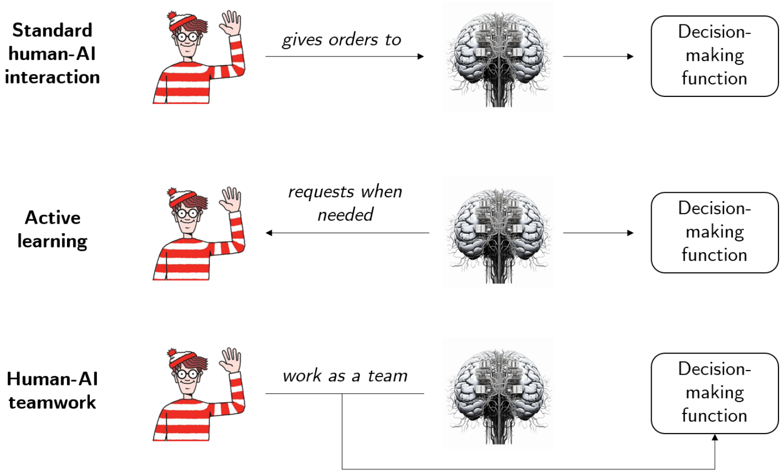
How to bring the human back in the loop?

Solution 3: The human and AI agents collaborate all over the process

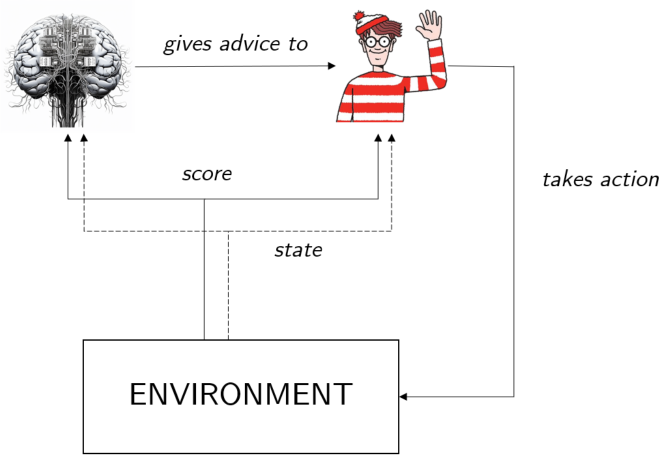
The whole learning process is done hand in hand with the user who stays in control. This is possible because machine learning is intrinsically an iterative process.

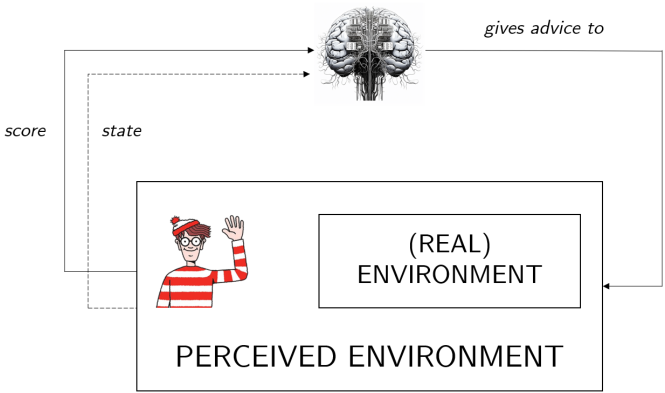


How to bring the human back in the loop?



Human-AI teamwork from the perspective of the team




gives advice to
score
state
(REAL)
ENVIRONMENT
PERCEIVED ENVIRONMENT

Example 1: Teamwork in data preprocessing

Teaching to Learn:
Sequential Teaching of Learners with Internal States

Mustafa Mert Çelikok[1], **Pierre-Alexandre Murena**[1], **Samuel Kaski**[1, 2]

[1] Aalto University
[2] The University of Manchester
mustafamert.celikok@aalto.fi, pierre-alexandre.murena@aalto.fi, samuel.kaski@aalto.fi

Example 1: Teamwork in data preprocessing

The human user has collected a dataset $\{(X_n, Y_n)\}$, and wants to learn a linear regression model:

$$Y = w^0 + w^1 X^1 + \ldots + w^d X^d$$

Pre-processing task: Choose which variables to include to the model.

Example 1: Teamwork in data preprocessing

For this scientist, there are multiple difficulties in this task:

- May be tempted to include all variables into the model
- May ignore on which criteria to select or discard variables
- May not have the possibility to explore all variables

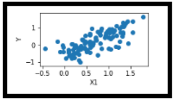
How to overcome these difficulties?

Example 1: Teamwork in data preprocessing

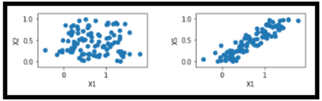
The assistant does **not display the optimal choice** directly, because this choice wouldn't be understandable.

In this application, at each time step, the assistant:

- Chooses a variable to suggest to the user
- Displays relevant 2-dimensional scatter plots about this variable.
- Lets the user choose whether to include this variable to the model.



Correlation
to the output



Correlation to other variables
in the model

Would you like to include variable X1 to the model?

Example 1: Teamwork in data preprocessing

Naive learner: ignores multicollinearity and its impact onto the performance of the regression. Therefore, includes any variable with correlation to the output.

Example 1: Teamwork in data preprocessing

Naive learner: ignores multicollinearity and its impact onto the performance of the regression. Therefore, includes any variable with correlation to the output.

Teaching such a learner: Three strategies:

1. **Manipulation:** Do not present multicolinear variables (*the learner is not tempted to include them, we are safe!*)

Example 1: Teamwork in data preprocessing

Naive learner: ignores multicollinearity and its impact onto the performance of the regression. Therefore, includes any variable with correlation to the output.

Teaching such a learner: Three strategies:

1. **Manipulation:** Do not present multicolinear variables (*the learner is not tempted to include them, we are safe!*)
2. **Sub-optimality:** Present all variables (*it would not be honest to hide the truth, even if the learner does not understand it correctly.*)

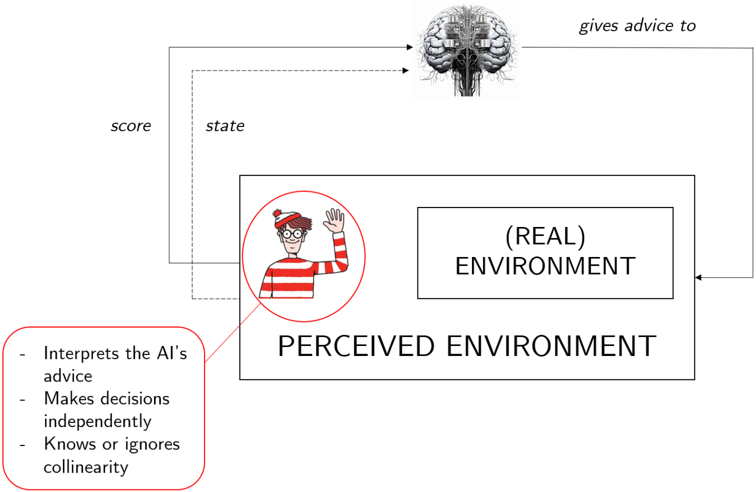
Example 1: Teamwork in data preprocessing

Naive learner: ignores multicollinearity and its impact onto the performance of the regression. Therefore, includes any variable with correlation to the output.

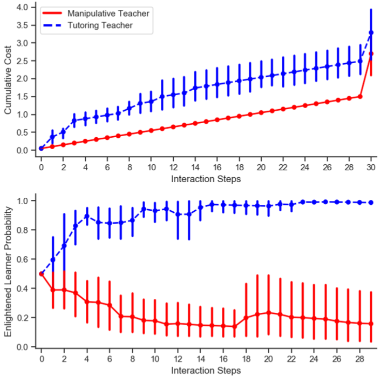
Teaching such a learner: Three strategies:

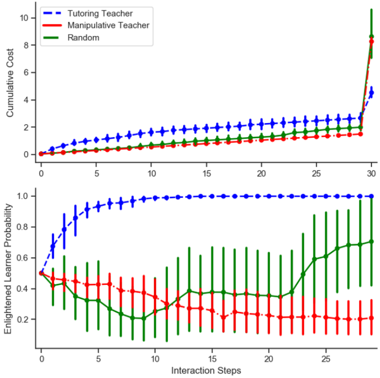
1. **Manipulation:** Do not present multicolinear variables (*the learner is not tempted to include them, we are safe!*)
2. **Sub-optimality:** Present all variables (*it would not be honest to hide the truth, even if the learner does not understand it correctly.*)
3. **Education:** Explain to the learner why picking multicolinear variables is not a good choice (*after that, we could safely present any variable!*)

Example 1: Teamwork in data preprocessing



Example 1: Teamwork in data preprocessing





The goal is to choose optimal variables for ***this one*** task

The goal is to choose optimal variables for ***this one*** task and for ***future*** tasks

Example 1: Teamwork in data preprocessing

Example 2: Sequential shared decision for exploration

Cooperative Bayesian Optimization for Imperfect Agents

Ali Khoshvishkaie[1], Petrus Mikkola[1], Pierre-Alexandre Murena[1,2], and Samuel Kaski[1,3]

[1] Department of Computer Science, Aalto University, Helsinki, Finland
[2] Hamburg University of Technology, Hamburg, Germany
[3] Department of Computer Science, University of Manchester, Manchester, UK
{firstname.lastname}@aalto.fi

Example 2: Sequential shared decision for exploration

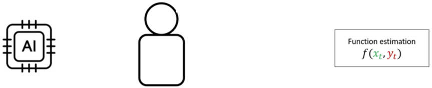
- Cooperative data acquisition in Bayesian optimization

Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user

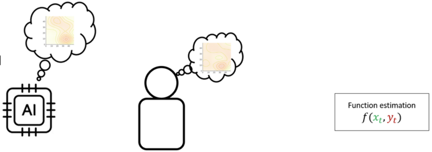
Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension



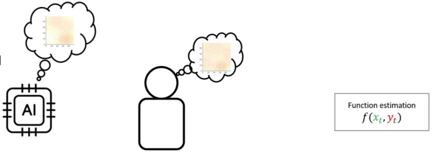
Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension
- Agents initially have different priors over the functio



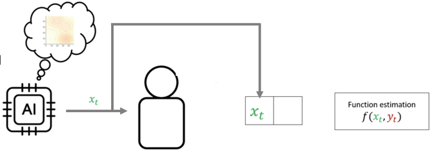
Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension
- Agents initially have different priors over the functio
- Fixed number of iterations, each step t as follows:



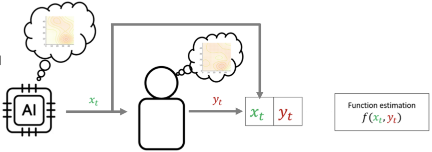
Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension
- Agents initially have different priors over the functio
- Fixed number of iterations, each step t as follows:
  - AI chooses x_t



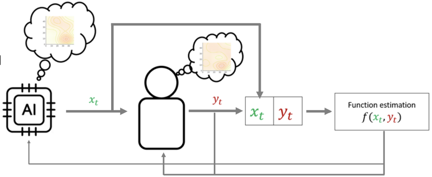
Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension
- Agents initially have different priors over the functio
- Fixed number of iterations, each step t as follows:
    - AI chooses x_t
    - User opts for y_t



Example 2: Sequential shared decision for exploration

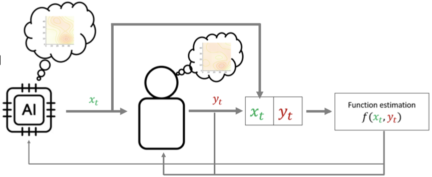
- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension
- Agents initially have different priors over the functio

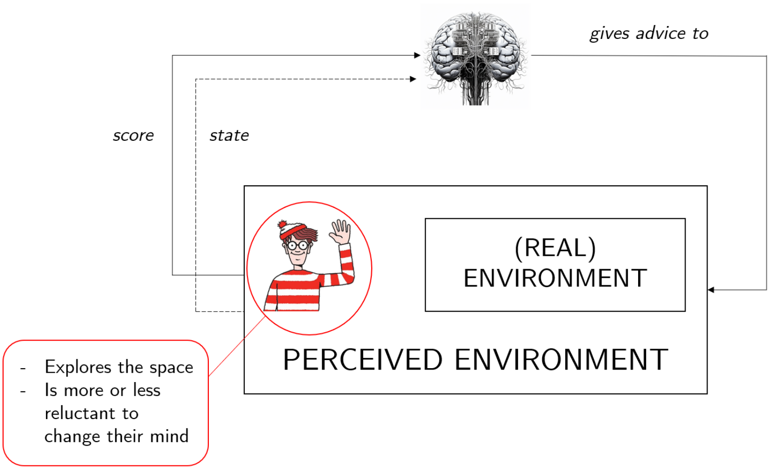


- Fixed number of iterations, each step t as follows:
  - AI chooses x_t
  - User opts for y_t
  - $f(x_t, y_t)$ is evaluated and observed by both

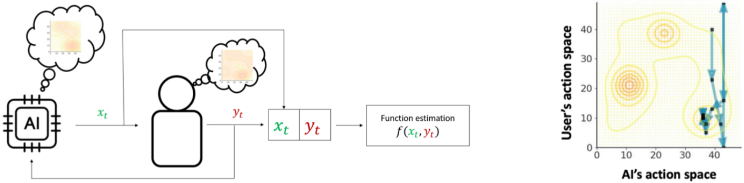
Example 2: Sequential shared decision for exploration

- Cooperative data acquisition in Bayesian optimization
- Two agents: AI and human user
- Common goal: Maximize a 2D function $f(x, y)$ together while each has only access to one dimension
- Agents initially have different priors over the functio
- Fixed number of iterations, each step t as follows:
  - AI chooses x_t
  - User opts for y_t
  - $f(x_t, y_t)$ is evaluated and observed by both
- Our goal: How to pick x_t?

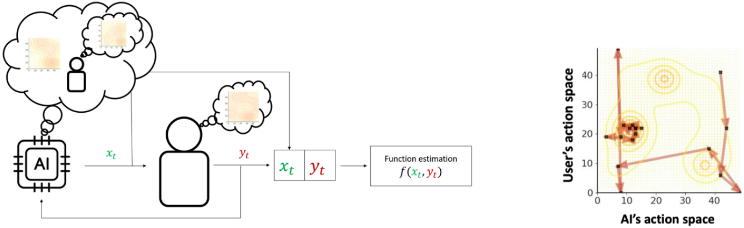



gives advice to
score
state
(REAL)
ENVIRONMENT
PERCEIVED ENVIRONMENT
- Explores the space
- Is more or less reluctant to change their mind

Example 2: Sequential shared decision for exploration



(a) Greedy AI agent

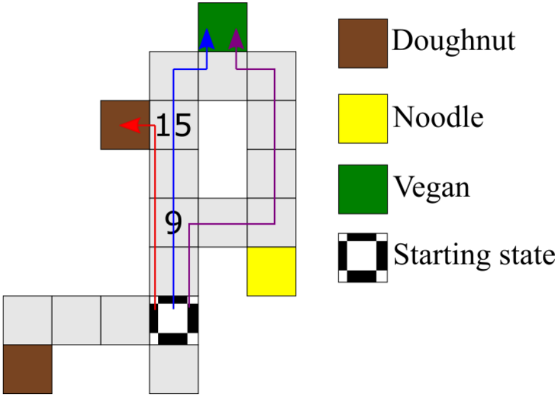


(b) Our proposed agent: Strategic planner with a user model.

Why is this difficult?

Why is this difficult?

Human users do not act the same way as an AI agent would



Why is this difficult?

Large number of observations needed to learn the human user's behaviour

A solution: Using pre-defined models. But

- Which ones?
- What if the behaviour differs?
- Are these models learnable?

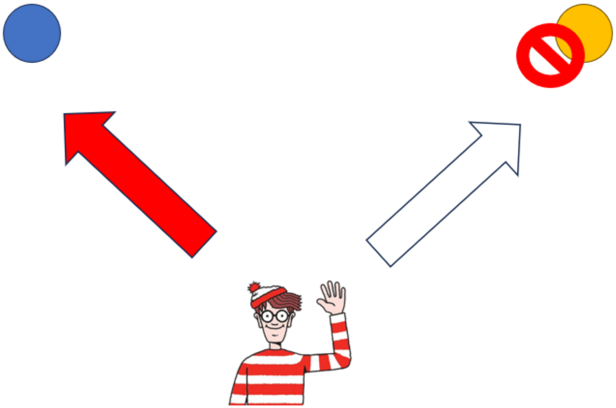
Why is this difficult?

Inferring Case-Based Reasoners' Knowledge to Enhance Interactivity

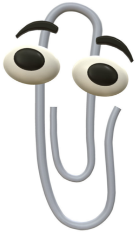
Pierre-Alexandre Murena[1,2(✉)] and Marie Al-Ghossein[1,3]

[1] Helsinki Institute for Information Technology HIIT, Helsinki, Finland
[2] Department of Computer Science, Aalto University, Espoo, Finland
pierre-alexandre.murena@aalto.fi
[3] Department of Computer Science, University of Helsinki, Helsinki, Finland
marie.al-ghossein@helsinki.fi

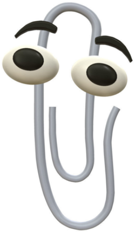
Why is this difficult?



Why is this difficult?



Why is this difficult?



Microsoft Office. However, instead of supporting the user with clear and precise guidance, studies show that Clippy "was considered to be annoying, impolite, and disruptive of a user's workflow" (Veletsianos 2007, p. 374). In the end, Clippy, the "non-intelligent artificial intelligence assistant", was so despised that even Microsoft made fun of it.[1] However, more

Conclusion

- Humans are here, even when you don't see them
- Taking them into account requires a different way of doing AI, at the intersection between:
  - Traditional machine learning
  - Reinforcement learning
  - Multi-agent systems
  - Cognitive sciences
  - Human-computer interaction
  - ...
- (And also, I'm recruiting... Feel free to contact me!)