Understanding user decisions when interacting with an AI

Advisors:

Baptiste Caramiaux | email: baptiste.caramiaux@lri.fr | web: www.baptistecaramiaux.com
Gilles Bailly | email: gilles.bailly@isir.upmc.fr | web: www.gillesbailly.fr
Quentin Roy

Keywords

Human-Computer Interaction, Artificial intelligence, Decision making, User Expertise, Interpretability, Trust, Transparency

Description

In recent years, machine learning algorithms had an increasing impact on our life. Algorithms are selecting the information we see after a search on Internet, they are recommending music albums and movies, or they are choosing the most appropriate navigation plans in transportation. In other words, these algorithms perform predictions and make decisions or recommendations on our behalf based on a certain number of criteria. Take the example of one driving her car from her home to her office. To avoid traffic, she is using an application able to predict what is the shortest journey from point A to point B. The criteria used to predict the best route, the cost function used to assess what is a better route, or the algorithm itself are not know to the driver and the prediction can sometimes happen to be wrong. Under such uncertainty, when the application recommends a new route, the driver faces the fundamental problem of: either trusting the technology and engaging herself on the new route (which may require more attention and cognitive load), or remaining on the usual route (but potentially arrive later). This simple example shows that interacting with an AI poses several fundamental challenges that have not yet been addressed in the field of Human-Machine Interaction, especially the role of algorithm interpretability [Doshi-Velez et al., 2017] in human decision making in the context of human-AI interaction.

The goal of this internship is to provide insights for the understanding on how users make the decisions when interacting with an AI. More precisely, we want to understand when and why users prefer to rely on their expertise/knowledge rather than the recommendations of an AI.

Previous works in the field of interactive machine learning have looked at ways to improve human trust in ML-based systems, namely by improving the interaction signals between the system and the humans. By enabling more complex human feedback than “yes/no” to the system, the user would better understand the behavior of the system and the system may perform better [Stumpf et al., 2009]. A complementary approach is to enable explanations from the system in order to improve the user understanding of the system behavior which has been assessed in a debugging context [Kulesza et al., 2015]. However, these approaches do not investigate how users will “cooperate”, “delegate”, “negotiate” with the AI (1) once they consider to have learned the behavior of the system, (2) in different contexts such as time pressure, risk management, etc.

In this internship in HCI, the student has to design, implement and conduct user studies to understand how the chosen factors (e.g. AI-based decision technique, user knowledge, practice, task, context, etc.) impact the decision of the users to rely or not on an AI to accomplish a task.

We will work with the student to (1) identify the key literature at the crossroad of HCI and AI, (2) identify the key factors to investigate (we foresee controlling factors such as perceived precision of the system, perceived knowledge of the user, cost of errors and time to make decisions), and (3) design the studies. We anticipate that this work will lead to a publication in a conference such as ACM CHI.

The internship may last from 4 to 6 months. If successful, the work could serve as the foundation for a Phd thesis.

Required skills

- Knowledge about Human-Computer Interaction and/or cognitive science
- Preferably a great interest in Machine Learning and Artificial Intelligence (no technical skills required)
- Programming skills (e.g. python, Java)
- Experience in experiment design, data analysis

Context

The internship will take place at LRI, University of Paris-Saclay, within the Ex)Situ Team. The intern will be supervised by Baptiste Caramiaux, Gilles Bailly and Quentin Roy.

Baptiste Caramiaux is a CNRS research scientist at LRI, University Paris-Saclay, member of the Inria team ex)situ. He is conducting research in interactive machine learning and human-computer interaction. He has also published in more general cognitive science journals (human perception and learning).

Gilles Bailly is part of the HCI group at Sorbonne Université, which has a strong track record at the CHI conference and is part of an exciting and multi-disciplinary laboratory (robotics, machine learning, perception, cognitive science, haptics, social interaction, etc.)

Quentin Roy…

References

[Doshi-Velez et al., 2017] Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

[Kulesza et al., 2015] Kulesza, T., Burnett, M., Wong, W.-K., and Stumpf, S. (2015). Principles of explanatory debugging to personalize interactive machine learning. In Proceedings of the 20th international conference on intelligent user interfaces, pages 126–137. ACM.

[Stumpf et al., 2009] Stumpf, S., Rajaram, V., Li, L., Wong, W.-K., Burnett, M., Dietterich, T., Sullivan, E., and Herlocker, J. (2009). Interacting meaningfully with machine learning systems: Three experiments. International Journal of Human-Computer Studies, 67(8):639–662.