

Towards Adversity-Aware Intelligent Agents

Advisor:

Julien Gori email: gori@isir.upmc.fr web: <https://www.isir.upmc.fr/personnel/gori/>

Gilles Bailly email: gilles.bailly@isir.upmc.fr web: www.gillesbailly.fr

Keywords

HCI; User Modeling; Intelligent Agents;

Description

Many intelligent agents such as assistants, artificial teachers, and recommender systems provide suggestions to users. Often these suggestions may be inaccurate or irrelevant [1], which can lead users to become adverse to the agent's suggestions, and overall may lead to a low acceptability of these systems [2]. One idea to create more acceptable agents is for them to actively account for and monitor users' adversity to suggestions, when deciding whether or not to provide a particular suggestion.

However, little is known quantitatively about user adversity, which makes this approach currently unfeasible. The goal of this internship is precisely to fill that gap. More precisely, the goal of the student will be to characterize adversity through several empirical studies. In particular, how does adversity to suggestions evolve over time when no suggestions are provided? How does this adversity evolve when relevant/irrelevant suggestions are provided, and how much does this adversity depend on the past suggestions? It is stressed that these empirical studies will need to have a high ecological validity.

Required skills

- Programming skills
- Interest in user modeling
- Design of experiments

Context

The intern will be hosted at the ISIR Laboratory at Sorbonne Université (Paris).

References

- [1] Matejka, J., Li, W., Grossman, T., & Fitzmaurice, G. (2009, October). CommunityCommands: command recommendations for software applications. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (pp. 193-202).
- [2] Do, H. J., Kong, H. K., Tetali, P., Lee, J., & Bailey, B. P. (2023). To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1-23.