The Seminars on "Information Technology Outlook" – PhD Programme in Computer Science and Mathematics

Prof. Cristina Conati
Professor of Computer Science and
Distinguished Scholar
University of British Columbia

📅 Wednesday September 27, 2023
11:30-12:30

📍 Sala Consiglio – 7th floor
Dept. Computer Science

Towards Personalized Explainable AI

The AI community is increasingly interested in investigating explainability to foster user acceptance and trust in AI systems. However, there is still limited understanding of the actual relationship between AI explainability, acceptance and trust, and which factors might impact this relationship. I argue that one such factor relates to user individual differences, including long-term traits (e.g., cognitive abilities, personality, preferences) and short-term states (e.g., cognitive load, confusion, emotions). Namely, given a specific AI application, different types and forms of explanations may work best for different users and even for the same user at different times, depending to some extent on their long-term traits and short-term states. As such, our long-term goal is to develop personalized XAI tools that adapt dynamically to the relevant user factors. In this talk, I focus on research investigating the relevance of long-term traits in XAI personalization. I will present a general methodology for this investigation and examples of how we applied it to understand the importance of personalized XAI in an intelligent tutoring system and a recommender system. I discuss how to move forward from these insights and research paths that should be explored to make personalized XAI happen.

Cristina's research is at is at the intersection of Artificial Intelligence (AI), Human-Computer Interaction (HCI), and Cognitive Science, with the goal to create AI systems that can both perform useful tasks and be well accepted by their users. A key aspect of this endeavor is enabling AI systems to predict and monitor relevant properties of their users (e.g., states, skills, needs, emotions) and personalize the interaction accordingly, in a manner that maximizes both task performance as well user satisfaction. Toward this goal, Cristina is especially interested in investigating how to enable AI technology to strike the right balance between providing accurate predictions and decision-making while maintaining transparency, user control, and trust. For more details on current and past projects, see https://hai.cs.ubc.ca/