Human–AI Interaction through Computational Understanding of Human Behavior

Hee-Seung Moon · Research Statement

In the realm of cooperative Artificial Intelligence (AI), understanding human behavior is paramount. Observing others, we, as humans, can infer each other's intentions and plan our responses accordingly — an ability I aim to extend to AI. With AI now embedded everywhere, from mobile devices to robots, automobiles, and even more intimate forms such as chatbots, infusing it with a cooperative intelligence akin to human interaction could significantly enhance its effectiveness. This integration promises to aid humans in performing tasks more intuitively and safely, while providing AI systems that can explain their actions comprehensively.

As a researcher at the intersection of Human–Computer Interaction (HCI) and AI, my primary objective is to develop AI systems that, through a profound understanding of human behavior, can computationally assist users to maximize their utility. Here, the "understanding" is defined as follows: Firstly, it involves inferring the underlying processes hidden behind observable human actions — essentially identifying what motivates a person to act in a certain way. Secondly, it requires the ability to plan how this person might behave under different situational contexts, such as those influenced by the interventions of AI systems. This presents a significant challenge. For example, any observed human behavior could stem from many possible underlying factors and processes. If achievable, AI could evolve beyond simply forecasting a person's next action but to understanding the complexities of their inner workings. This will enable AI to determine when and how to optimally intervene with users. The remaining question then is how AI can incorporate and utilize such understanding to deliver beneficial assistance to humans. Herein lies the importance of "computational" modeling. By capturing human behaviors through computational models,

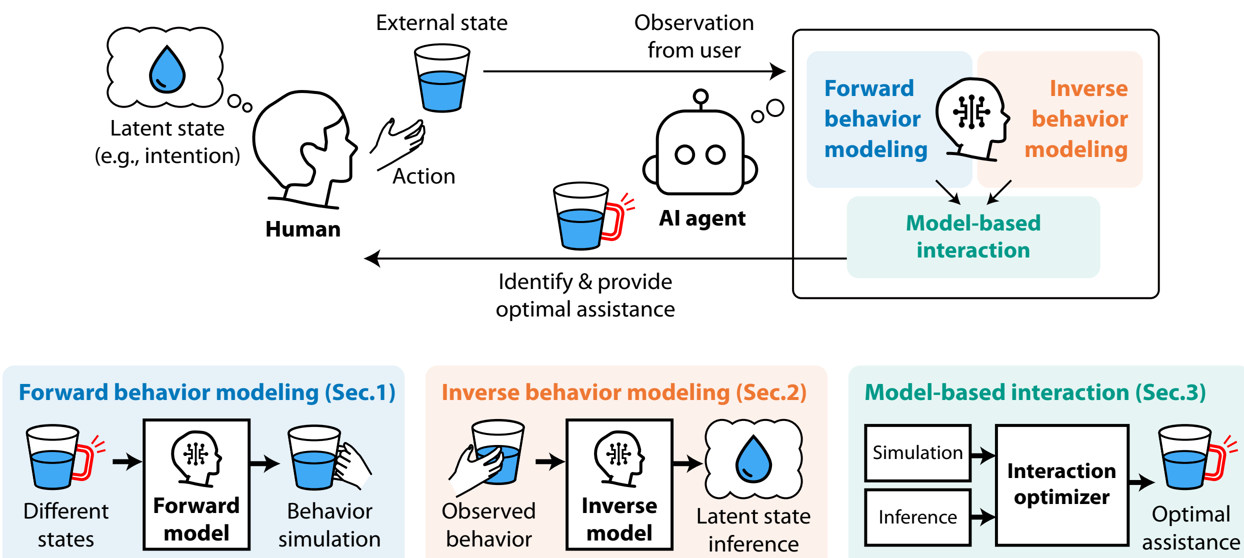

Figure 1: My research investigates the potential of AI to understand user behavior and determine the best forms of assistance. It focuses on three main areas: (1) developing a user behavior model that simulates responses to both internal changes (such as intentions and capabilities) and external interventions (such as those made by AI); (2) using this model to infer the user's internal states based on their actions; and (3) optimizing AI assistance to enhance user utility through such simulation and inference.

AI can leverage its efficient computational power to generate the most beneficial outcomes, ensuring the highest utility for the user. Such outcomes could take the form of actions by cooperative robots, adaptive interfaces, optimized designs, and more.

My research unfolds across three key aspects of building human–AI interaction (Figure 1). First, it delves into the computational modeling of user behavior, aiming to simulate human actions from a blend of latent (e.g., intentions, preferences, capabilities) and external (e.g., given task) states. This approach posits that human behavior is the rational outcome of these factors, seeking to capture and model the complexity of human actions. Second, the focus shifts to the computational inference of a user's latent state from observed behavior. Utilizing recent techniques like neural density estimation, also known as amortized inference, my research explores streamlining the probabilistic inference process, facilitating their immediate application in human–AI interactions. Third, the research investigates the computational strategy for building AI actions based on the user model. This involves determining how AI can decide on subsequent actions and update the user model to provide personalized assistance, based on the insights gathered from observations. Together, these research directions aim to enhance AI's capability to understand and predict human behavior, fostering more intuitive and effective human–AI collaborations.

1 Forward Behavior Modeling

Humans seem to seamlessly transform their intentions into actions, yet this process relies on complex, inner mechanical stages of perception, cognition, and motor function. My research builds on the principle of computational rationality [9], positing that these intricate processes are the result of rational decision-making that connects mechanical actions to intentions. The advent of recent reinforcement learning (RL) techniques empowers this framework, enabling computers to autonomously develop optimal decision-making policies. This is achieved not by relying on extensive human-derived data or explicitly hand-crafted procedural rules, but through defining the goals and rewards that shape task execution. This evolution underpins the development of sophisticated behavioral models, facilitating AI's ability to perform alongside humans through counterfactual reasoning, thus making AI interactions more intuitive and aligned with human behavior.

My research has enhanced user simulations by effectively integrating human variability characterized by diverse latent features across individuals, such as cognitive-physical abilities and internal utility functions. At the core of my research are theory-driven models of human perception and motor execution to reflect such latent features. I've utilized biomechanical models to explore the details of how limb kinematics and motor noise affects these movements, aiming to closely replicate human motion [7]. Then, a key technical aspect of constructing such simulation models is the identification of a policy network adaptable to a range of Markov Decision Processes (MDPs) with different latent feature values. I have developed techniques for the rapid adaptation of policy networks within simulation models to address this multi-task RL problem [5]. By applying these models to HCI tasks, such as point-and-click [5], menu search, touchscreen typing [6], and VR target selection [7], I've demonstrated that user simulations can diminish the reliance on expensive human data collection and showcase enhanced predictive accuracy regarding individual human behavior.

My ultimate objective is to enable AI to universally simulate human behavior across any interactive task. This requires equipping computer models with the capability to replicate the delicate motor skills humans possess, such as the intricate dexterity of fingers, which involves overcoming the challenge of numerous muscle tendons acting in overactuation. Moreover, a comprehensive understanding of how the human mind operates — managing attention, learning new information, and organizing them — is crucial for developing a flexible framework. Additionally, AI must evolve to autonomously determine human-like goals, moving beyond the current necessity of manually setting rewards. The overarching aim is to close the gap between AI's theoretical understanding of human behavior and the intricate reality of human actions, thereby improving the quality of human-AI interactions in a variety of tasks.

2 Inverse Behavior Modeling

Human behaviors emerge from their inner latent states, such as the intention to pick up a cup, but only the explicit actions, such as arranging fingers to grasp and extending the arm towards the cup, are observable. Inverse modeling is the process of deriving these latent insights from observable behaviors, essentially inferring the reasons behind a person's specific actions. My approach involves using simulations to identify the relationship between latent features and observable behavior. Traditionally, this has involved a probabilistic search for latent feature values that yield simulation outcomes closely resembling the observed behavior, which has been suffered from intensive computational cost per inference [2].

My previous work has focused on enabling inference that is both fast enough for real-time applications and robust enough to handle noisy data by probabilistic approaches that account for uncertainty. The technical core of this effort lies in amortizing the traditional simulation-based Bayesian inference process by neural network-based proxy models [6]. Essentially, this involves building and training neural models that immediately estimate the posterior distribution of the latent parameters from observation data. I have applied the neural posterior estimation to several HCI tasks, achieving efficient inference with millisecond-level speed. Based on this, my work includes predicting individuals' perceptional and motor noise from cursor movements during point-and-click tasks [5], analyzing finger motor skills from touchscreen typing performance [6], and inferring intentions for selecting precise targets in VR from prior motions [7]. These works highlight the potential of inverse behavior modeling to significantly improve AI's capability to understand and adapt to human behavior on an individual level.

My research aims to advance personalized interactions through pioneering inverse modeling techniques, which enable AI to extract maximum insights from human behavior with minimal observations. A significant challenge in this quest is minimizing the simulation gap. The inherent imperfections in simulation models, often stemming from incorrect assumptions of humans, can create discrepancies between simulated behaviors and actual human actions, leading to inaccurate inferences. Therefore, my future research will focus on not only enhancing model precision and but also developing an inference methodology robust against these discrepancies, ensuring more accurate and reliable AI interpretations of human behavior.

3 User Model-based AI Interaction

As AI acquires a deeper understanding of human behavior through both forward and inverse modeling, my research goal continues toward harnessing this knowledge to drive computational AI interactions. The objective is to utilize AI's refined insights into human actions and intentions to dynamically enhance user interactions. My research agenda include three key areas, each selected for its unique contribution to realizing this goal:

- **Simulation-based Interaction Optimization**: One of my focus is on enhancing interaction and interface parameters through user simulation and computational optimization process. This includes refining design elements like button sizes or tailoring media content to user preferences, and even adjusting the physical shape of devices for enhanced ergonomics. Simulating users allows AI to test various outcomes in different settings without the costly need for user participation, enabling optimization according to their needs and preferences [10, 8].

- **Human–AI Collaboration via Multi-Agent RL**: I aim to develop strategic cooperation between AI and humans, treating both as agents aiming to maximize user utility. It considers how AI actions, such as intervening with users or adapting interfaces, affect user behavior, acknowledging that such actions can sometimes lead to unexpected negative impacts [11]. By adopting a multi-agent RL framework, the study addresses complex situations, enabling AI to learn how to choose optimal moments and methods for collaboration, tailored to the unique context of each user [1].

- **Fast User Model Update through Interaction**: My research plan involves enabling AI to swiftly update its understanding of user preferences and behaviors based on real-time interaction data. Employing meta-learning techniques [3, 4], this research aims to enable AI systems to adapt promptly to changes in user behavior, ensuring interactions remain personalized and contextually relevant.

Together, these three areas form a comprehensive strategy for applying AI's theoretical understanding of human behavior towards practical, impactful interactions. By addressing each area's distinct challenges and opportunities, my research strives to transform AI systems into a more effective partner in a range of settings, from personal assistance to collaborative work environments, ultimately enhancing the synergy between humans and AI.

References

[1] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. On the utility of learning about humans for human-ai coordination. *Advances in Neural Information Processing Systems*, 32, 2019.

[2] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[3] Hee-Seung Moon and Jiwon Seo. Fast user adaptation for human motion prediction in physical human–robot interaction. *IEEE Robotics and Automation Letters*, 7(1):120–127, 2021.

[4] Hee-Seung Moon and Jiwon Seo. Optimal action-based or user prediction-based haptic guidance: Can you do even better? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2021.

[5] Hee-Seung Moon, Seungwon Do, Wonjae Kim, Jiwon Seo, Minsuk Chang, and Byungjoo Lee. Speeding up inference with user simulators through policy modulation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2022.

[6] Hee-Seung Moon, Antti Oulasvirta, and Byungjoo Lee. Amortized inference with user simulations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2023.

[7] Hee-Seung Moon, Yi-Chi Liao, Chenyu Li, Byungjoo Lee, and Antti Oulasvirta. Real-time 3d target inference via biomechanical simulation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024.

[8] Roderick Murray-Smith, Antti Oulasvirta, Andrew Howes, Jörg Müller, Aleksi Ikkala, Miroslav Bachinski, Arthur Fleig, Florian Fischer, and Markus Klar. What simulation can do for hci research. *Interactions*, 29(6):48–53, 2022.

[9] Antti Oulasvirta, Jussi PP Jokinen, and Andrew Howes. Computational rationality as a theory of interaction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2022.

[10] Sayan Sarcar, Jussi PP Jokinen, Antti Oulasvirta, Zhenxin Wang, Chaklam Silpasuwanchai, and Xiangshi Ren. Ability-based optimization of touchscreen interactions. *IEEE Pervasive Computing*, 17(1):15–26, 2018.

[11] Kashyap Todi, Gilles Bailly, Luis Leiva, and Antti Oulasvirta. Adapting user interfaces with model-based reinforcement learning. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.