Research Proposal

Ripan Kumar Kundu

My research is situated at the intersection of trustworthy machine learning (ML), human-centered artificial intelligence (AI), and adaptive intelligent systems. Throughout my doctoral studies, I have designed and developed explainable, robust, privacy-aware, and interactive AI systems for complex, high-stakes environments (e.g., extended reality (XR) applications). This experience has provided me with a strong foundation in large language models (LLMs), reinforcement learning (RL), multimodal learning, intelligent decision support, and human-AI interaction. Looking forward, I am particularly interested in developing AI systems that extend beyond task optimization to actively strengthen human thinking, reflection, and collaboration. The Postdoctoral Associate position in MIT EECS and SERC is especially compelling because it enables the pursuit of these goals within an interdisciplinary environment that integrates ML, human-computer interaction (HCI), and the social sciences.

Impact: My research has been published in several high-quality venues, including 4 journals in Transactions on Dependable and Secure Computing (**TDSC**), Transactions on Visualization and Computer Graphics (**TVCG**), **IEEE Access** and 13 peer-reviewed top-quality conference papers in **IEEE ISMAR** 2022 (core rank A*), **IEEEVR** 2023 (core rank A*), **ACM IUI** 2023 (core rank A), **PHM** 2023 (Acceptance rate 21%), **IEEE ISMAR** 2024 (core rank A*), **IEEEVR** 2025 (core rank A*), **IEEE ISMAR** 2025 (core rank A*), **ACM VRST** 2025 (core rank A), and **IEEEVR** 2026 (core rank A*).

At MIT, I intend to pursue a research program focused on AI systems that foster human metacognition and sociality within workplace environments. My primary objective is to design intelligent agents that not only assist individuals in completing tasks but also facilitate the development of reflection, skill acquisition, and social connection over time. I am particularly interested in reinforcement learning with ethnographic rewards (RLER), as it addresses a significant limitation of current AI systems: many valuable human capabilities in real-world work settings, such as articulating reasoning, recognizing uncertainty, mentoring peers, or reframing problems, are tacit, socially embedded, and not captured by standard optimization objectives. I view this direction as a natural extension of my previous work on explainable and adaptive AI, now oriented toward systems that actively enhance human judgment and workplace learning.

One research direction I aim to pursue involves developing reward models for metacognitive and social skill development using rich workplace interaction data. Rather than viewing ethnographic observation solely as a preliminary phase, I seek to integrate qualitative workplace studies and worker co-design directly into the model development process. This approach entails identifying interaction patterns indicative of productive reflection, collaborative knowledge building, and skill growth and translating these patterns into machine-learnable reward signals. A central challenge is to balance specificity and transferability: reward functions must be sufficiently grounded to reflect authentic organizational contexts while remaining generalizable across tasks and domains. My expertise in trustworthy and human-centered AI equips me to address this challenge by designing objectives that are both effective and interpretable.

A second research direction involves constructing adaptive, worker-aware AI agents using LLMs and multi-turn reinforcement learning. I am interested in developing LLM-based systems capable of modeling diverse worker archetypes, interaction styles, and levels of expertise, thereby enabling agents to respond in ways that are tailored to users' evolving needs. These systems could learn when to provide direct assistance, prompt self-explanation, surface missing assumptions, or encourage collaboration. My previous work on adaptive AI systems, interactive reasoning, and LLM-driven support provides a strong foundation for this line of inquiry. More broadly, I am motivated to develop AI systems that function as reflective collaborators rather than static tools.

A third research direction focuses on designing and evaluating interfaces that foster productive friction, reflection, and social learning within workplace workflows. While many AI systems seek to minimize friction, certain forms of friction can be advantageous in environments that prioritize learning and collaboration. I am interested in developing interfaces that prompt users to articulate reasoning, compare alternatives, inspect model reasoning processes, and reflect on potential gaps in understanding. I am particularly interested in evaluating these systems not only by immediate task performance, but also by their impact on downstream learning, confidence calibration, collaboration quality, and long-term skill development.

Beyond my technical research agenda, I place significant value on mentoring and interdisciplinary collaboration. I am eager to contribute to the intellectual community across EECS and SERC, collaborate with scholars in HCI and the social sciences, mentor undergraduate and graduate researchers, and advance research that addresses the social and ethical responsibilities of computing. Ultimately, my long-term vision is to develop AI systems that enhance human agency, and I am enthusiastic about the opportunity to contribute this vision at MIT.