Cover Letter

xia.chen@iek.uni-hannover.de
+49 157 8095 9659

Dear CHAI Research Fellowship Selection Committee,

My name is Xia Chen, a computer scientist with a unique background in architecture and design. I am currently completing my Ph.D. under the guidance of Prof. Dr.-Ing. Philipp Geyer at Technical University Berlin and Leibniz University Hannover. Previously, I held a position as an invited visiting scholar at the University of California, Berkeleys Center for the Built Environment, where I worked on investigating how AI embedded with causal inference can conduct counterfactual reasoning to benefit the engineering domain.

My Ph.D. journey began in an interdisciplinary background that has been instrumental in shaping my current research focus, where I was deeply involved in a German foundation research project aimed at designing machine assistance to help designers and engineers make informed decisions. This experience has constantly challenged me to consider how we can embed our prior knowledge to better solve engineering tasks and, more importantly, how machines can capture and align with uniquely human features such as analogy, reasoning, preference, and values of social factors. Throughout my Ph.D., I've encompassed my research into my dissertation, titled "*Beyond Predictions: Alignment between Prior Knowledge and Machine Learning for Human-Centric Augmented Intelligence*," emphasizes aligning human intelligence with machine capacities to enhance, rather than undermine, our capabilities and autonomy. The core content is organized into three essential objectives:

1. **Decision-Making Processes Alignment**: Constructing a fundamental framework to align with users' complex decision-making processes, inspired by the estimation mechanisms of human nervous systems.
2. **Methodological Paradigms Alignment**: Proposing a paradigm to systematically embed prior knowledge and domain insights into data-driven models, addressing domain-specific challenges through three levels of knowledge integration, which draws on theories of human cognitive development (Piagetian stage theory) to construct models:
    a. *Modeling Knowledge for System Description*: Utilizing explicit knowledge to enhance machine learning models' interpolation capabilities through data augmentation and feature engineering.
    b. *Inductive Logic and Disentanglement for Extrapolation*: Using inductive logic and disentangled system compositionality to modify the modeling process, enabling predictions beyond observed data.
    c. *Abstract Reasoning and Deductive Logic*: Engaging in abstract thinking to allow machine learning models to solve complex problems and answer "what-if" questions.
3. **Interaction Pattern Alignment**: Improving human-computer interaction and communication patterns, exploring a symbiotic framework, and investigating the potential information exchange.

Within this scope, my coding experience includes comprehensive data analysis skills and advanced machine learning techniques in supervised, unsupervised, reinforcement, and semi-supervised learning, using Python and R.

During my Ph.D. in Germany, I was a data science lecturer in masters and bachelors courses for four years. I independently led two fundamental research projects funded by the German Research Foundation and the German Federal Ministry of Education with follow-up funding proposal preparation, coordinating efforts across various European research units and industrial partners. These roles have provided me with substantial interdisciplinary experience in coordinating different roles and strong communication skills that enable me to engage in community-building and outreach events effectively.

Having recently completed my visiting scholarship at UC Berkeley, I was deeply impressed by the campuss free academic environment and the encouraging, friendly research atmosphere. Ive long admired CHAIs commitment to developing provably beneficial AI systems and it aligns perfectly with my research goals. Im excited by the prospect of collaborating with world-class researchers to address some of the most critical challenges in AI development. You can find my one-page statement of research interest below:

One-page statement of interest

My future research plan is threefold, aimed at advancing AI in alignment with human intelligence. I believe it aligns closely with CHAIs mission and could contribute significantly to the work on the foundations of rational agency, causality, value alignment, and human-robot cooperation:

- **Knowledge-Integrated Machine Learning (KIML) Advancement**: Building on my dissertation, I will further develop methods for prior information embedding and domain knowledge integration in a bottom-up approach. This work aims to propose and refine the hierarchical structure of knowledge integration (direct modeling knowledge, inductive logic, deductive abstract reasoning) in data-driven models for the community in applied science. I plan to explore new factors besides compositionality entanglement in prior knowledge for aligning machine learning with human cognition in specific domain problems, contribute to CHAIs focus on value alignment by developing AI systems that are more attuned to human values and expertise, and apply these approaches to specific domain problems, contributing to the broader field of AI for Science and potentially addressing CHAIs interest in multi -agent perspectives.

- **Biologically-Inspired Dynamic Neural Architectures**: Built on my dissertation's machine assistance framework mechanisms, as well as the philosophy of alignment and analogy. I aim to address current limitations in AI by exploring more biological plausible model, such as spiking neural networks, combined with mortal dynamic network architectures to mitigate catastrophic forgetting and enhance learning from small sample sizes, develop architectures that incorporate the time dimension and network self-organization mechanisms in information processing from a perspective of complexity science. I will develop a prototype beyond the typical von Neumann architecture and explore how these dynamic structures can contribute to CHAIs work on the foundations of rational agency and causality, as well as advance interests in AI capabilities and robust inference and planning through the novel integration of storage and computing in neural networks.

- **Cognitive Alignment for Human-AI Collaboration**: Ultimately, I envision developing an intelligent system that contains ergonomics to better construct machine assistance for human-AI symbiosis in an informational construct akin to a cybernetic mindset. Key aspects of this research include investigating techniques to enable AI systems to adapt and evolve in response to human feedback and changing conditions, along with advanced human-computer interaction patterns (such as one of my previous research by mapping human electroencephalogram signals with personal emotions and implicit preference, the third part from my dissertation) that facilitate seamless collaboration between humans and AI systems from information theory perspectives. This objective should be closely combined with scientific discovery, engineering, and social sciences application, with which I hold strong connections and rich experience from my background to translate theoretical models into practical, scalable technologies. I aim to enhance the transparency and interpretability of AI decision-making processes, exploring ways to align AI systems with human values and ethical considerations in real-world applications.

Thank you for considering my application. I look forward to the opportunity to discuss how my research interests, experience, and enthusiasm for AI alignment topics can contribute to CHAIs important work.

Sincerely,

Xia Chen