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 academic background in engineering and architecture 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, Berkeley's Center for the Built Environment, where I worked on investigating how AI embedded with causal inference for counterfactual reasoning benefits the engineering domain.

My Ph.D. journey began in an interdisciplinary setting 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 methodologically to consider how we can embed our prior knowledge into data-driven methods to better solve engineering tasks and, more importantly, how machines can capture and align with uniquely human intelligence features such as analogy, reasoning, preference, and values of social factors. I encompassed my research throughout my Ph.D. 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 flexible or extrapolative predictions.
    c. *Abstract Reasoning and Deductive Logic*: Engaging in abstraction modeling directly from data instead of pattern finding allows data-driven models to solve complex problems or answer "what-if" questions.
3. **Interaction Pattern Alignment**: Improving human-computer interaction and communication patterns for potential information exchange bandwidth expansion, exploring a symbiotic framework.

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 lecturer in master's and bachelor's data science 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 to transfer innovative outcomes to practical solutions. 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 campus's free academic environment and the encouraging, friendly research atmosphere. I've long admired CHAI's commitment to developing provably beneficial AI systems and it aligns perfectly with my research goals. I'm excited by the prospect of collaborating with world-class researchers to address some of the most critical challenges in AI development. Please find my one-page statement of research interest below:

One-page statement of interest

My future research plan is threefold, aimed at advancing data-driven methods in alignment with human intelligence to construct as assistance, as I explicitly state in the future direction section at the end of my dissertation. I believe it aligns closely with CHAI's mission and could contribute significantly to the work on the foundations of rational agency, causality, value alignment, and human-robot cooperation:

- **Exploration of Fundamental Elements for Knowledge Integration**: Building on my dissertation, I will further develop methods for prior information embedding and domain knowledge integration from 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 studying the fundamental building blocks of human intelligence and creativity for aligning machine learning with human cognition in specific domain problems, I plan to investigate the interaction among elements such as scaling, recursion, and emergence mechanisms, which could lead to the creation of new ideas by combining familiar concepts. For instance, in understanding how to blend compositional and recursive learning to spark new concepts, causality provides the glue that gives them coherence and purpose and unlocks creative design. This research will contribute to CHAI's 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 CHAI's interest in multi-agent perspectives.
- **Biologically-Inspired Dynamic Neural Architectures**: Built on the machine assistance framework mechanisms from my dissertation, as well as my research philosophy of alignment and analogy. I aim to address current limitations in AI by exploring more biologically plausible models, such as spiking neural networks, combined with mortal dynamic network architectures. This approach will focus on mitigating current AI bottlenecks in catastrophic forgetting. It contrasts sharply with natural cognitive systems, where new learning rarely completely disrupts or erases previously acquired information. I will develop architectures that incorporate the time dimension and network self-organization mechanisms to process information from a complexity science perspective. My goal is to create a prototype beyond the typical von Neumann architecture, exploring how these dynamic structures can contribute to CHAI's work on the foundations of rational agency and causality. This research will advance interests in AI capabilities and robust inference through the novel integration of storage and computing in neural networks, while also addressing the integration challenges with neuroscience and cognitive science.
- **Cognitive Alignment for Human-AI Collaboration and Beyond**: Ultimately, I envision developing an intelligent system that incorporates 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 that facilitate seamless collaboration between humans and AI systems from information theory perspectives. For instance, one of my previous research projects aims to map human electroencephalogram signals with personal emotions and preferences. This would be a key focus on aligning AI systems not only with human knowledge but also with implicit factors in societal ethics, values, and morals. This objective should be closely combined with scientific discovery, engineering, and social sciences application, with which I hold solid 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 CHAI's important work.

Sincerely,

Xia Chen