Earl T. Barr
UCL Computer Science
Gower Street
London WC1E 6BT
United Kingdom
e.barr@ucl.ac.uk

March 15, 2024

To Whomsoever It May Concern,

I am pleased to endorse Tairan Wang for the summer research program at University College London (UCL). As a Professor of Software Engineering at UCL, I conduct research on application of machine learning and large language models to software engineering problems. My pioneering paper in this area "On the naturalness of software" received over 1300 citations and was recognised by ICSE Most Influential Paper Award. My role also includes serving as the head of UCL's software engineering group.

Tairan has shown a deep passion for research in large language models (LLMs), coupled with a strong initiative for self-directed learning. His engagement with the LLM field is influenced by the teachings of Dr. Sergey Mechtaev, who is a CREST member and my collaborator. Sergey's lectures for the COMP0010 Software Engineering course fuelled Tairan's interest in this field. Exploring various open-source LLM models on GitHub, such as Qwen, Tairan has demonstrated a keen interest in applying these models to real-world challenges, focusing on complex reasoning and code inspection in various benchmark datasets. Tairan also explored recent literature on Prompt Engineering. Tairan's strong performance in Dr. Mechtaev's COMP0010 Software Engineering class, where he ranked as the **top 1 in the cohort of 150**, clearly demonstrates his intellectual capabilities and his aptitude for excelling in sophisticated technological domains.

The research proposal "The Fact Selection Problem in LLMs for Software Engineering," will be supervised by me in collaboration with Dr. Sergey Mechtaev. It aims to refine the efficacy of LLMs for code related tasks. It is based on previous research that I conducted in collaboration with Dr. Sergey Mechtaev to investigate the relationship between the inclusion of bug-related facts in prompts and the performance of large language models (LLMs) in software bug fixing. The research highlighted the important problem of optimal fact inclusion for enhancing LLM's bug-fixing capabilities. Tairan will investigate this phenomenon in a wide context of software engineering, and will develop an automatic, model-agnostic methodology for fact selection, aimed at enhancing LLM performance in software engineering tasks.

I am confident in Tairan Wang's abilities and believe his participation in the summer research program will be immensely beneficial, both to his personal academic growth and for software engineering research of CREST centre. Specifically, our aim is advance the state-of-the-art of prompt engineering in code-related tasks, and to publish a paper in a top ranked conference.

For additional information or further details regarding Tairan Wang's qualifications and this project, please do not hesitate to contact me at e.barr@ucl.ac.uk.

Yours Sincerely,

Earl T. Barr