SORS: "Generative AI for Agile Hardware Development"

Abstract

The integration of generative Artificial Intelligence (AI) into microprocessor design marks a paradigm shift in hardware development. The CyberRio project at RIOS is among one of the earliest attempts of using a Large Language Model (LLM) to create a RISC-V CPU design from scratch. Leveraging the cutting-edge GPT-4, CyberRio deploys AI generated content across the entire CPU design lifecycle, from initial design specifications, RTL implementations to functional verifications. We believe that domain-adapted generative AI models can be highly effective in the field of chip designs.

We are also developing an LLM based framework called Rosemary with various chip design domain knowledge adaptation techniques in tokenization, continued pretraining, model alignment, and retrieval models. Our goal is to create an engineering AI chat assistant that can support engineers in development, a document generator that automatically generates modifiable drafts of technical documentation, a smart tool that can summarize and analyze bugs, as well as generate EDA scripts. Through our explorations with generative AI, we have found that domain-adaptive continued training and fine-tuning models can perform better domain specific tasks than the vanilla LLaMA2 baseline model. This gives us confidence in the potential of domain-specific enhanced LLMs like Rosemary can be more effective in the future agile hardware development.

Short Bio

Prof. Zhangxi Tan is a member of technical steering committee at RISC-V International, an adjunct professor at the Shenzhen International Graduate School?SIGS) of Tsinghua University, and a Co-Director of the RISC-V International Open-Source Laboratory (RIOS Lab). At RIOS, Prof. Tan is leading open-source hardware IP and software development that helps the RISC-V ecosystem world-class. He received his PhD in computer science from UC Berkeley in 2013 supervised by Prof. David Patterson, who is a winner of the 2017 Turing Award. Dr. Tan joined early development of RISC-V at Berkeley. His primary research is computer architecture and networks, microprocessor and VLSI designs, open-source RISC-V technologies and ecosystems, OpenEDA and PDK, non-volatile memory systems, SW/HW co-design and implementation of computer systems.

Prof. Tan is also a successful serial entrepreneur in the Silicon Valley. He was a founding engineer of Pure Storage (NYSE:PSTG), founder of OURS Technology (aquired by Aurora, NASDAQ:AUR). Prof. Tan holds more than 30 US patents in microprocessor designs, flash storage systems and hardware accelerators.

Speakers

Speaker: Prof. Zhangxi Tan. Member of technical steering committee at RISC-V International. Adjunct professor at the Shenzhen International Graduate School (SIGS) of Tsinghua University. Co-Director of the RISC-V International Open-Source Laboratory (RIOS Lab).
Host: Miquel Moretó. High Performance Domain-Specific Architectures - Group Manager. Computer Sciences, BSC.
Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---