

ConvCodeWorld: Benchmarking Conversational Code Generation in Reproducible Feedback Environments

Hojae Han (Seoul National Univ.)

SEMINAR ANNOUNCEMENT

- **Date & Time :** *Apr. 2, 2025, 10:00~12:00*
- **Place :** *Room 502, ITBT Bldg., Hanyang University*
- **Abstract:** Large language models excel at code generation, but current benchmarks overlook the diversity of feedback in multi-turn interactions. We introduce CONVCODEWORLD, a benchmark simulating 9 interactive code generation scenarios with three feedback types: (a) compilation, (b) execution with varying test coverage, and (c) verbal feedback from GPT-4o. We also present CONVCODEBENCH, a static version using pre-generated feedback logs, closely matching CONVCODEWORLD's results while reducing costs. Our evaluations reveal that feedback quality significantly affects LLM performance, weaker models can outperform stronger ones with sufficient feedback, and training on specific feedback types may hinder generalization.
- **Bio:** Hojae Han is a Ph.D. candidate at the Language and Data Intelligence Lab, supervised by Seung-won Hwang, in the Department of Computer Science and Engineering at Seoul National University, South Korea. His work lies in applying large language models in code-related tasks. He has published his work in Computer Science venues such as ICLR, ACL, and EMNLP. Hojae received his M.S. and a B.S. in Computer Science from Yonsei University, supervised by Seung-won Hwang. He has experience as a research intern at Snowflake, LG AI Research, Microsoft Research Asia, and Naver Corp. He is a recipient of a NAVER Ph.D Fellowship. For more information, please visit his website at <https://www.linkedin.com/in/hojae-han>.
- **Contact :** kimtaeuk@hanyang.ac.kr (컴퓨터소프트웨어학부 김태욱 교수)