

# Cognitive Distillation with Parameter-Efficient LLMs: Chain-of-Thought Calibration for Personality Prediction

**Yang Yan**

Zhejiang University – Westlake University Joint PhD Program, Hangzhou, Zhejiang, China

**Lizhi Ma**

Hangzhou Normal University, Hangzhou, Zhejiang, China

**Yu Lu**

Engineering School, Hangzhou, Zhejiang, China

**Renjun Xu**

Zhejiang University, Hangzhou, Zhejiang, China

**Zhenzhong Lan**

Engineering School, Hangzhou, Zhejiang, China

## Abstract

Large language models (LLMs) excel at personality prediction but are often impractical for deployment due to high computational demands. This work introduces cognitive distillation with Chain-of-Thought calibration, a novel framework for transferring structured reasoning from large LLMs to smaller, efficient models. Inspired by cognitive architectures like ACT-R, our method aligns intermediate inference steps using exemplars from single or multiple LLMs. A 1.5B-parameter model distilled through this process surpasses Qwen1.5-110B in predictive accuracy, achieving a 28% improvement in Pearson correlation while using just 1.36% of its parameters. Ablation studies reveal that moderate sampling diversity and multi-model ensembles enhance cognitive skill transfer and construct validity. These findings demonstrate that high-level reasoning can be efficiently and faithfully transferred, enabling psychometrically robust personality assessment in resource-constrained settings. This approach bridges AI and cognitive science, offering a scalable path toward plausible, cognitively grounded language models.