

A CONCEPTUAL FRAMEWORK FOR QUALITY CONTROL AND AUTOMATED TESTING OF LARGE LANGUAGE MODELS

Stelian Yolov, Vladimir Valkanov

Abstract. *As large-language models (LLMs) are increasingly made use of across critical applications, efficient quality control and scalable and adaptive testing become essential to ensure reliability, fairness, safety and performance.*

Traditional manual testing is unable to catch up with the constantly evolving nature of LLMs. We shape the architecture that integrates (i) specification of quality measurements (e.g. factual accuracy, coherence, bias mitigation, latency, safety), (ii) using pre-defined patterns, test examples, and templates, creating a broad set of test prompts by automatically filling in templates and including hard-to-handle examples, (iii) evaluation of the LLMs response in controlled environment.

This conceptualization aims to link the theoretical AI evaluation and practical automation strategies and open the way towards standardized methodologies for model verification, clarity, and reliability.

Key words: LLM, Automation Framework, Quality Control, AI, Testing.

Acknowledgments

This study is supported by the project FP25-FMI-010 “Innovative interdisciplinary research in informatics, mathematics and educational pedagogy” at the Paisii Hilendarski University of Plovdiv.

Stelian Yolov¹, Vladimir Valkanov¹

¹ Paisii Hilendarski University of Plovdiv,
Faculty of Mathematics and Informatics,
236 Bulgaria Blvd., 4003 Plovdiv, Bulgaria
Corresponding author: syolov@uni-plovdiv.bg