On the Assessment of Generative AI in Modeling Tasks: An Experience Report with ChatGPT and UML*

Javier Cámara[1], Javier Troya[1], Lola Burgueño[1], and Antonio Vallecillo[1]

ITIS Software, Universidad de Málaga {jcamara,jtroya,lolaburgueno,av}@uma.es

Keywords: Large Language Models, ChatGPT, Software models, Modeling languages, UML
Published in: Software and Systems Modeling, Vol. 22, No. 3, pp.781–793, 2023
Impact Factor: JCR 2.0 - Q3 - Position: 62/108 - Area: Computer Science / Software Engineering
DOI: https://doi.org/10.1007/s10270-023-01105-5

Abstract. Most experts agree that Large Language Models (LLMs), such as those used by Copilot and ChatGPT, are expected to revolutionize the way in which software is developed. Many papers are currently devoted to analyzing the potential advantages and limitations of these generative AI models for writing code. However, the analysis of the current state of LLMs with respect to software modeling has received little attention. In this paper, we investigate the current capabilities of ChatGPT to perform modeling tasks and to assist modelers, while also trying to identify its main shortcomings. Our findings show that, in contrast to code generation, the performance of the current version of ChatGPT for software modeling is limited, with various syntactic and semantic deficiencies, lack of consistency in responses, and scalability issues. We also outline our views on how we perceive the role that LLMs can play in the software modeling discipline in the short term, and how the modeling community can help to improve the current capabilities of ChatGPT and the coming LLMs for software modeling.

* Supported by the Spanish Government (FEDER/Ministerio de Ciencia e Innovación–Agencia Estatal de Investigación) under projects PID2021-125527NB-I00 and TED2021-130523B-I00 and Universidad de Málaga.