**Leila ZINE**

**LEVEL 1 : Morgane Gaspar**

**Investigating Code Generation Performance of ChatGPT with Crowdsourcing Social Data**

**Keywords Specific :** Code generation, ChatGPT, Social media, Programming languages, Dominant fear emotion, Crowdsourcing data, Generative models, User perceptions, Data analysis.

**Article-2**

The paper explores the potential of ChatGPT for code generation and looks at user attitudes towards AI-based programming assistance tools. This study represents the first large-scale attempt to analyse emerging generative models for code writing and testing, using collaboratively collected social data. The aim of this research is to bridge the knowledge gap on how AI is transforming software engineering and programming education.

The paper presents a scalable framework based on collaborative social data collection to investigate the code generation performance of large-scale language models. To this end, a framework comprising three key components is proposed: keyword expansion, data collection and data analysis. The hybrid keyword model integrates words suggested by topic modelling and expert knowledge to filter relevant social posts on Twitter and Reddit. This process collected 316,000 tweets and 3,200 Reddit publications on code generation by ChatGPT, covering the period from 1 December 2022 to 31 January 2023.

Analysis of the data obtained shows that ChatGPT has been used in more than 10 programming languages, with Python and JavaScript being the two most popular, for a diverse range of tasks such as debugging code, preparing interviews and solving academic assignments. To the researchers' surprise, the analysis revealed that fear was the dominant emotion associated with code generation using ChatGPT, eclipsing the emotions of happiness, anger, surprise and sadness. In addition, a dataset including ChatGPT code generation invitations and corresponding codes was constructed from the analysis of ChatGPT code generation screenshots shared on social networks, and made available to the public.

Finally, the paper poses a number of research questions, including the most popular programming languages used with ChatGPT, the programming scenarios, tasks and goals for which people use ChatGPT, the temporal distribution of ChatGPT code generation discussions, stakeholder perceptions, the quality of generated code, potential ethical issues related to generated code, among others. These questions aim to provide valuable guidance for future research on AI-powered code generation.