Future applications of generative large language models: A data-driven case study on ChatGPT

Filippo Chiarello [a,d,*], Vito Giordano [a,d], Irene Spada [b,d], Simone Barandoni [c,d], Gualtiero Fantoni [b,d]

[a] Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Italy
[b] Department of Civil and Industrial Engineering, University of Pisa, Italy
[c] Department of Computer Science, University of Pisa, Italy
[d] Business Engineering for Data Science (B4DS) Research Lab, University of Pisa, Italy

ARTICLE INFO

ABSTRACT

This study delves into the evolving role of generative Large Language Models (LLMs). We develop a data-driven approach to collect and analyse tasks that users are asking to generative LLMs. Thanks to the focus on tasks this paper contributes to give a quantitative and granular understanding of the potential influence of LLMs in different business areas. Utilizing a dataset comprising over 3.8 million tweets, we identify and cluster 31,747 unique tasks, with a specific case study on ChatGPT. To reach this goal, the proposed method combines two Natural Language Processing (NLP) Techniques, Named Entity Recognition (NER) and BERTopic. The combination makes it possible to collect granular tasks of LLMs (NER) and clusters them in business areas (BERTopic). Our findings reveal a wide spectrum of applications, from programming assistance to creative content generation, highlighting LLM's versatility. The analysis highlighted six emerging areas of application for ChatGPT: human resources, programming, social media, office automation, search engines, education. The study also examines the implications of these findings for innovation management, proposing a research agenda to explore the intersection of the identified areas, with four stages of the innovation process: idea generation, screening/idea selection, development, and diffusion/sales/marketing.

1. Introduction

In the rapidly evolving landscape of artificial intelligence (AI), generative large language models (LLMs) have emerged as a pivotal innovation. Language models are probabilistic models that enable the processing of natural language through algorithms, and they are the core of the natural language processing (NLP) techniques. The term 'large' refers to the huge number of parameters needed to train these models, whereas the term 'generative' indicates a subset of LLMs specifically designed to generate text. Generative LLMs are at the forefront of numerous applications, ranging from summarisation to translation, typically delivered through dialogue-like interactions with the user (Zhao et al., 2023).

The dynamic of the dialogue eases the interaction of users on the system (Dang et al., 2022). According to the Technology Acceptance Model, perceived ease of use and usefulness are the factors that influence user acceptance of technology (Davis, 1989). Indeed, these systems are engendering an unprecedented impact in terms of user numbers, estimated to be over 100 million in January 2023 (Huang et al., 2023). The rise of generative LLMs has been led by ChatGPT, the openAI conversational system built on the generative pre-trained transformer (GPT) model launched in late 2022. Other relevant generative LLMs are Microsoft's Bing, which focuses on interacting with the web to provide answers based on present information, and Google's PaLM, known for its improved multilingual, reasoning, and coding capabilities. Meta's LLaMA is another contender in the space, with an approach to multi-modal learning that integrates both text and image data.

Despite already reaching millions of users, generative LLMs are still in their infancy, aligning with what innovation literature defines as emerging technologies. These are characterized by radical novelty, rapid growth, coherence, prominent impact, and a degree of uncertainty and ambiguity (Rotolo et al., 2015). Moreover, these new generative models have found applications across a wide spectrum of fields, making them general-purpose technologies (Bresnahan and Trajtenberg, 1995).