A journey into the Generative AI and large language models: From NLP to BioInformatics

Ahmed Elnaggar

Technische Universität München, Germany

Editors: François Coste, Faissal Ouardi and Guillaume Rabusseau

Abstract

In the last year, the generative AI field has seen a remarkable breakthrough, specifically the generative ai models and their applications in the natural language processing domain. It has achieved new state-of-the-art results on all public datasets and super human-level chatting capabilities. The backbone of this breakthrough is the large language models, including OpenAI GPT and Google Palm. The advantages of these large language models are that they can effectively capture the semantic, syntactic, grammar, and meaning of characters, words, and sentences from large unlabelled datasets using self-supervised learning. Later it can be used to represent sentences and documents better through embedding or as a zero/multi-shot learning method for many NLP tasks. Fortunately, these models have started to be leveraged in other fields like bioinformatics and biochemistry. This talk will give an overview of the large language models and how it was applied in the Bioinformatics field to boost the performance on many use cases. Furthermore, it will show how high-performance computing and optimized deep-learning software and libraries allowed these models to be faster and more efficient during training and inference.