Accenture – Generative AI for Drug Discovery

TIN-ACC-020

About the Organisation
Accenture Labs are Accenture’s dedicated arm of research and development. We are focused on solving critical business problems with advanced technology, bringing fresh insights and innovations to our clients, helping them capitalize on dramatic changes in technology, business and society. Our dedicated team of technologists and researchers work with leaders across the company to invest in, incubate and deliver breakthrough ideas and solutions that help our clients create new sources of business advantage. At Accenture Labs Dublin, based at The Dock, we focus on Artificial Intelligence technologies and systems like explainable AI, knowledge discovery and computational creativity applied to business areas as diverse as healthcare, life science, workforce, supply chain and consumer products.

Analysis Methods
Deep learning, Machine learning, Large Language Models

Data Source
Publicly available data sets. Internal private data sets.

Broader Objectives
To create techniques that can discover potential drug candidates and predict their characteristics and potential interactions.

Role Description and Responsibilities
Generative AI are widely used in many subfields of AI and Machine Learning and more recently have come to the public’s attention through large language models (LLM). Generative models based on GANs and diffusion models have been employed in a creative manner to generate new media (images, music) but they have also been applied to areas such as drug discovery (especially small molecules) and data synthesis for privacy. LLMs have gained widespread attention through the latest version of ChatGPT. However, it's important to note that even before this, LLMs had already made significant contributions to areas such as speech recognition, machine translation, and part-of-speech tagging. Although BERT and GPT models have been utilized to understand molecular structures, their impact on drug discovery has not been substantial thus far.

Despite the advancements made by deep learning methods in drug discovery, several challenges remain. One of these challenges involves computationally predicting interactions between drugs and targets. When dealing with peptides or proteins instead of just small molecules, a lack of labeled data becomes a significant hurdle in training models. However,
by treating peptides and protein sequences as text, there is potential for large language models to have an impact in this domain. Furthermore, recent developments in deep learning have allowed insights into the relationship between small molecules and genetic perturbations, enabling the prediction of effective drug combinations to combat complex diseases. This underscores the potential of constructing generative AI models that combine multiple related sources of information, such as drug compounds and gene expression.

The research intern will join a team working on deep learning for drug discovery. They will be in charge of designing, implementing, and evaluating novel principled ways to tackle one or more research problems listed above, with the ultimate goal of discovering more useful drug candidates (small molecules and large molecules). The intern is expected to explore and experiment with a range of techniques from prior art, propose original research, and implement ideas that will be validated with the research team in Accenture Labs Dublin.

**Expected Outcomes**

The research intern will work with AI researchers to select and develop models with an application for drug discovery. Expected deliverable: prototype and a publication submission to a major academic conference.

**Supervision and Mentorship**

On-site supervision will be by Jer Hayes and other members of the research team in the Accenture Labs, Dublin.

**Person Specification**

- Being enrolled in a PhD programme in Computer Science, Bioinformatics, Computational Biology, Genomics or related field;
- Strong academic background in machine learning/deep learning methodology, especially with deep generative networks;
- Strong proficiency in at least one programming language, ideally Python, and relevant toolboxes: Scikit-Learn, Pytorch, Keras;
- Strong proficiency in working with software version control, e.g., Git;
- Interest in solving real-world scientific problems and in acquiring commercial experience;
- Demonstrable experience of performing well both when working part of a team and individually;
- Ability to pause your PhD for the duration of the internship and return subsequently;
- Ability to work creatively and analytically in a problem-solving environment;
- Excellent verbal and written communication in English.

**Internship Logistics**

This internship will be based in Accenture, Dublin - The Dock, 7 Hanover Quay. Please note that it is **not** possible for this internship to be performed remotely and it requires relocation to Dublin, Ireland.

The start date is expected to be September 2023, although this date is negotiable.

The duration will be 6 months.
The remuneration will be pro-rated, based on an equivalent annual salary of €40k. This is a full-time position, and we regret that we are unable to consider part-time applications.

When applying for this role, please include your resume. Your resume should show any relevant links to illustrate programming experience (e.g. GitHub handle) and scientific accomplishments (e.g. Google Scholar, dblp, arXiv links, personal homepage) if not already captured in the application form.