About the Organisation

Accenture Labs BioInnovation is Accenture’s dedicated arm of research and development on Bioinformatics and Artificial Intelligence for healthcare and life sciences. We offer a blend of industry and academic research activities, including an open publication policy and contribution to the open-source community. All supported by our Accenture Labs network of over 200 applied R&D specialists at seven sites worldwide.

Analysis Methods

Biodata, Multi-Omics Data, Language Models, Knowledge Graphs, Graph Machine Learning

Data Source

Accenture’s large dataset of structured real-world clinical records. Public multi-omics datasets (genomic, biological pathways, diseases, drugs, etc). On select projects, patient-level WES/WGS data.

Broader objectives

To design and apply machine learning to generate in silico hypotheses from structured and unstructured biomedical knowledge bases and answer clinically relevant questions.

Role Description and Responsibilities

Recent advances such as Large Language Models, Graph Neural Networks and Neural Graph Databases have been used successfully for knowledge discovery on large structured and unstructured multi-omics knowledge bases [1,2,3,4]. Accenture Labs adopt these methods in pre-clinical drug discovery, precision oncology decision support systems, and hypothesis generation for neurodegenerative disorders.

Despite excellent predictive power, many aspects of these machine learning architectures are still under research scrutiny. The fast-paced Large Language Models community is working on mitigating hallucinations, domain-specific fine tuning, and scaling down such resource-hungry models. Research on Graph Machine Learning include hybrid neuro-symbolic reasoning, time-awareness, few-shot learning, and ensuring human interpretability, just to name a few [1,5]. An area of significant interest is the Large Biomedical Language Models – Graph Machine Learning interplay, to make the most of the two worlds.

The research intern will join the Accenture BioInnovation team and will design, implementing,
and evaluate novel principled ways to tackle one or more research problems listed above. The goal is inferring knowledge from multi-omics knowledge bases in the context of our ongoing research projects (e.g. target identification, adverse events prediction, biomarker discovery, etc). 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 BioInnovation.

Expected Outcomes
Software prototype, technical report, submission to major AI academic conference (open publication policy), opportunity to contribute to our open-source machine learning library [4].

[4] https://github.com/Accenture/AmpliGraph

Supervision and Mentorship
On-site supervision will be by Luca Costabello and other members of the research team in the Accenture Labs, Dublin.

Person Specification

Essential Requirements
- Being enrolled in a PhD programme in Computer Science, Bioinformatics, Computational Biology, Genomics or related field
- Strong knowledge of Machine Learning foundations and mainstream Deep Learning architectures
- Strong scientific Python programming skills (e.g. NumPy)
- Hands-on experience with at least one machine learning framework (e.g. TensorFlow, PyTorch, JAX)
- Familiarity with public multi-omics datasets (genomic, transcriptomic, proteomic, and other biological datasets) and multi-modal clinical data sources (real world data)
- Ability to work creatively and analytically in a problem-solving environment
- 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
- Excellent verbal and written communication in English
- Ability to pause your PhD for the duration of the internship and return subsequently

Optional Requirements
- Previous exposure to *at least one* of these areas: Large Language Models (LLMs), Machine Learning for Knowledge Graphs (e.g. Knowledge Graph Embeddings, Graph Neural Networks), Interpretable and Trustworthy AI.
- Previous exposure to drug discovery and/or genomic medicine projects

**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.