AI Chemistry Student Project – PubChem Summaries

We are looking for a motivated student to explore the use of large language models (LLMs) to summarize chemistry knowledge. The goal of the project is to generate accurate and human-readable summaries for chemicals based on rich textual data extracted from public repositories such as PubChem for display on our local PubChemLite collection. This work will contribute to enhancing the usability and accessibility of chemical knowledge for both researchers and the public by powering platforms like PubChemLite and 3D Molecules.

The selected student will:
• Explore and fine-tune state-of-the-art LLMs for summarizing chemical data
• Evaluate summarization accuracy and investigate hallucination mitigation strategies
• Collaborate with researchers from cheminformatics and AI backgrounds
• Potentially engage with an AI-focused startup for technical support

This project is a collaboration between Prof. Emma Schymanski's Environmental Cheminformatics group at the Luxembourg Centre for Systems Biomedicine and Dr. Zhiqiang Zhong (AI & ML) and Prof. Jun Pang from the Computer Science Department in the Faculty of Science, Technology and Medicine; all at the University of Luxembourg. It can be structured as a part-time student job, internship, or a Master's thesis, depending on the student's background and availability.

Your profile…
• Background in computer science, cheminformatics, bioinformatics, data science, or a related field
• Interest in applying machine learning and AI to scientific problems
• Experience with Python and libraries such as PyTorch, HuggingFace Transformers, or similar
• Familiarity with natural language processing and summarization techniques is an advantage
• Curiosity-driven, capable of working independently and in interdisciplinary teams
• Previous exposure to chemical or biological data is a plus, but not required

In short….
A unique opportunity to apply state-of-the-art AI methods to a real-world chemical knowledge platform, with high impact and interdisciplinary collaboration potential. The project can be tailored to a part-time job (UL students), internship, or Master's thesis.

How to apply….
Please send your CV, a brief motivation letter, and any relevant project or code samples to Dr. Zhiqiang Zhong (zhiqiang.zhong@uni.lu), with the subject line: *AI Chemistry Student Project – PubChem Summaries*.

Further information

For further information, please contact Dr. Zhiqiang Zhong (zhiqiang.zhong@uni.lu). Please use the job title mentioned above in the subject line of your email.