Leveraging Deep Learning and Large Language Models for Chemical Discovery at P&G

Ajay Muralidharan & Matt Barker
Procter & Gamble Corporation

In this presentation, we explore the utilization of deep learning and large language models (LLMs) in the field of chemistry, focusing on the discovery of small molecules, surfactants, and polymers at P&G. We delve into various LLM-based approaches, including finetuning and few-shot learning, to predict molecular properties. Also, we compare LLM-based approaches with traditional deep learning methods, highlighting their current strengths and weaknesses. To exemplify the effectiveness of deep learning, we present a case study on molecular solubility using Graph Neural Networks and highlight other emerging tools for polymers.

Furthermore, we integrate Open AI GPT4 into a virtual chemistry assistant, facilitating access to our internal chemistry database and deep learning-based chemistry toolsets via a chat interface. This integration embodies our commitment to augmenting human innovation with AI, fostering an ecosystem where data science enriches research and development endeavors in chemistry and promotes a culture of interdisciplinary collaboration.