NMR Spectroscopy as Molecular Language: A Deep Learning Approach to Bioactivity Prediction

Artificial intelligence (AI) is revolutionizing the life sciences – not just by automating routine tasks, but by challenging our very assumptions about how we explore and understand the molecular world. In chemistry and drug discovery, computer models are often trained on simplified digital fingerprints of molecules – strings of binary numbers that describe a compound's structure in two dimensions. While efficient, these representations overlook crucial details like 3D shape, electron distribution, or how a molecule might behave in a real, messy biological environment.

This project sets out to change that by giving machines a more human-like way of "seeing" molecules – through the lens of spectroscopy. Instead of working with abstracted descriptors, we generate detailed virtual nuclear magnetic resonance (NMR) spectra for thousands of compounds. These spectra, commonly used by chemists to study molecular structure, encode rich information about how atoms are arranged, what their chemical environments look like, and how they might interact with biological targets.

Using powerful algorithms inspired by the human brain – such as *graph neural networks* (GNNs), which mimic how neurons process connections, and *convolutional neural networks* (CNNs), which specialize in detecting patterns – the project combines these spectral fingerprints with information about molecular structure. The result? A new generation of hybrid AI models that don't just guess a molecule's properties, but interpret them in context.

These models are not just academic exercises. They're trained to predict biological activity across three clinically significant targets in the central nervous system:

- the **serotonin 5-HT$_6$ receptor**, linked to memory and cognition,
- the **μ-opioid receptor**, a key player in pain regulation and addiction,
- and the **histamine H$_3$ receptor**, involved in alertness, sleep, and neurodegeneration.

Each of these receptors presents unique challenges for drug development – such as activity cliffs, where small changes in molecular structure cause dramatic shifts in effect, or biased signaling, where a drug activates only selected pathways. Traditional models struggle with this level of nuance. The models developed here, by contrast, aim to capture those subtleties by integrating structural and spectroscopic information at a depth previously unattainable.

What makes this project truly exciting is not just its predictive power, but its transparency. By using explainable AI techniques, we can trace back the predictions to specific parts of a molecule's spectrum or structure. That means chemists won't just get a "yes" or "no" – they'll see *why* a compound is expected to be active, and where to tweak it for better results.

Ultimately, this project offers a bridge between human intuition and machine precision. If successful, it could redefine how we explore chemical space, accelerate the search for new therapeutics, and shift the paradigm of drug discovery toward models that are not only accurate, but also interpretable, scalable, and deeply grounded in the chemistry they aim to describe.