ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Ο.Π.Μ.Σ. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

DEMOKRITOS
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Επιβλέπων:
Δρ. Γιώργος Μπουρίτσας,
Μονάδα Αρχιμήδης, ΕΚ Αθηνά / Τμήμα Πληροφορικής & Τηλεπικοινωνιών, ΕΚΠΑ

A. De novo design of environmentally friendly molecules with deep generative models

The use of deep learning in the physical sciences (AI for science) has been steadily gaining traction, showing the potential to revolutionize scientific discovery through data-driven methodologies. In the field of chemistry, significant progress has been made in tasks such as molecular property prediction, chemical reaction prediction, and de novo drug design.

This thesis will focus on the computational design of molecular structures with desired properties. Recent advancements at the intersection of *generative modeling* and *geometric deep learning* have shown promising results for molecular generation tasks, through architectures such as Variational Autoencoders (VAEs) [1] and Generative Adversarial Networks (GANs) [2]. A novel paradigm, known as GFlowNets [3], which combines generative modeling with reinforcement learning, has demonstrated impressive capabilities in generating graph-structured objects, particularly drug-like small molecules.

The aim of this thesis is to leverage these techniques, with a specific focus on GFlowNets, to design *environmentally friendly molecules*. To achieve this, the first step will be the collection of a dataset containing information about molecular properties that directly impact environmental safety, such as biodegradability and toxicity. This data will be used to train a proxy model (reward model) for environmental friendliness, which will be later employed to guide the decisions of the generative model. The workflow will incorporate single or multi-objective optimization to ensure that key factors are optimized simultaneously.

[1] W. Jin, R. Barzilay, and T. Jaakkola, “Junction Tree Variational Autoencoder for Molecular Graph Generation,” ICML 2018.
[2] N. De Cao and T. Kipf, “MolGAN: An implicit generative model for small molecular graphs,” ICLR workshops 2018.
[3] Y. Bengio, S. Lahlou, T. Deleu, E. J. Hu, M. Tiwari, and E. Bengio, “GFlowNet Foundations,” JMLR 2021.