IAMAT 2024

Navigating compound space with physics based machine learning

Anatole von Lilienfeld

[1] *University of Toronto*

[2] *Vector Institute*

Speaker e-mail : Anatole.vonlilienfeld@utoronto.ca

Many of the most relevant observables depend explicitly on atomistic and electronic structure, rendering physics based approaches to chemistry and materials necessary. Unfortunately, due to the curse of dimensionality (combinatorial scaling of possible number of chemicals and reactions) a rigorous first principles based understanding remains computationally prohibitive – even for high-performance computing. Accounting for explicit and implicit dependencies and correlations through machine learning, however, deepens not only our fundamental understanding but also accelerates computational and experimental exploration campaigns by orders of magnitude. I will review insights gained over the last decade, detailing how physically motivated representations, selection of training instances, definition of loss-functions, and label choices systematically impact data efficiency, as monitored by learning curves (scaling) [1]. Given the central role of density functional theory, these findings will be contextualized within the framework of EAST: Efficiency, accuracy, scalability, and transferability [2].

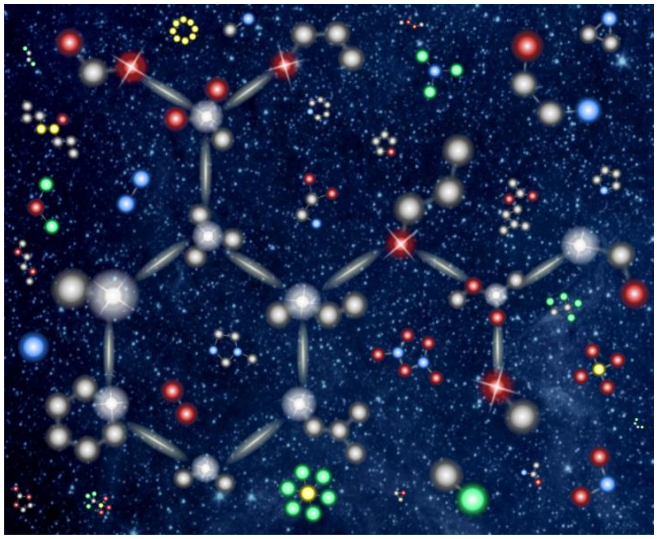

Figure 1: Illustration of redundancy in chemical compound space [3]

REFERENCES :

[1] Huang, B. and Von Lilienfeld, O.A., 2021. Ab initio machine learning in chemical compound space. *Chemical reviews*, *121*(16), pp.10001-10036.

[2] Huang, B., von Rudorff, G.F. and von Lilienfeld, O.A., 2023. The central role of density functional theory in the AI age. *Science*, *381*(6654), pp.170-175.

[3] von Lilienfeld, O A, Müller, K.-R., Tkatchenko, A., Exploring chemical compound space with quantum-based machine learning, Nature Chemistry Reviews (2020)