# Computational materials science with machine learning: from data to insights

**Silvana Botti** and **Miguel A.L. Marques**
Research Center Future Energy Materials and Systems
Faculty of Physics and Astronomy and ICAMS
Ruhr University Bochum, Germany

silvana.botti@rub.de

Machine learning (ML) models for materials science are rapidly evolving, driven by large-scale, high-quality datasets and innovative neural network architectures. This talk explores critical challenges in improving the accuracy and reliability of complex ML models, examining the interplay between quality and quantity of the training data and model performance across material properties.

Recent advances have been marked by the creation of extensive FAIR databases, such as alexandria, which provides over 5 million density-functional theory calculations spanning periodic compounds of various dimensionalities. These comprehensive datasets enable systematic investigation of the relationship between training data volume/quality and model accuracy.

Comparative assessments of machine learning approaches—ranging from composition-based models to crystal-graph neural networks—demonstrate that architectures with detailed geometrical information consistently outperform simpler compositional models. To address representation challenges, we discuss a novel material fingerprinting technique that we have recently proposed and that balances computational efficiency with human interpretability, offering new insights into representation of crystalline materials for machine learning inference.

We finally discuss the ongoing race to develop universal machine learning interatomic potentials. Our recent benchmarking study focuses on predicting phonon properties, which are critical for understanding the vibrational and thermal behavior of materials. By analyzing approximately 10,000 ab initio calculations of phonons, we highlight the nuanced performance of different potentials on the Matbench Discovery leaderboard, revealing both their promising capabilities and existing limitations.

## References

[1] Alexandria Database. https://alexandria.icams.rub.de/
[2] J. Schmidt, T.F.T. Cerqueira, A.H. Romero, A. Loew, F. Jäger, H.-C. Wang, S. Botti, Miguel A.L. Marques, Improving machine-learning models in materials science through large datasets, Mater. Today Phys. 48, 101560 (2024).
[3] A. Loew, D. Sun, H.-C. Wang, S. Botti, M. A. L. Marques, Universal Machine Learning Interatomic Potentials are Ready for Phonons, https://doi.org/10.48550/arXiv.2412.16551
[4] T.F.T. Cerqueira, S. Botti, M.A.L. Marques A non-orthogonal representation of the chemical space, https://doi.org/10.48550/arXiv.2406.19761
[5] J. Schmidt, N. Hoffmann, H.-C. Wang, P. Borlido, PJ.M.A. Carriço, T.F.T. Cerqueira, S. Botti, M.A.L. Marques, Machine-learning-assisted determination of the global zero-temperature phase diagram of materials, Adv. Mater. 35, 2210788 (2023).
[6] H.Kulik et al., Roadmap on machine learning in electronic structure, Electron. Struct. 4, 023004 (2022).
[7] J. Schmidt, L. Pettersson, C. Verdozzi, S. Botti, M.A.L. Marques, Crystal-graph attention networks for the prediction of stable materials, Sci. Adv. 7, eabi7948 (2021).
[8] P. Borlido, J. Schmidt, A.W. Huran, F. Tran, M.A.L. Marques, S. Botti, Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning, NPJ Comput. Mater. 6, 96 (2020).

## Figures

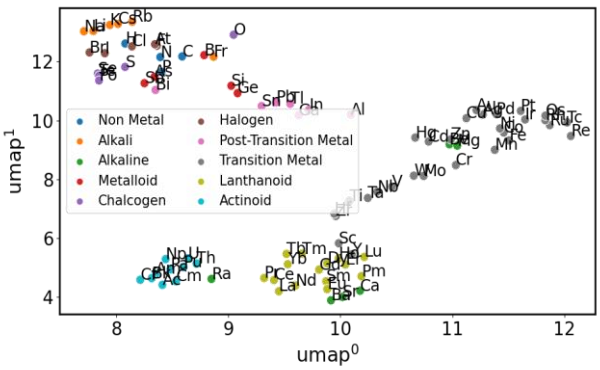


**Figure 1.** Two-dimensional map obtained by reducing the dimensions of the compositional fingerprints of the chemical elements from Ref.[4] that can be thought of as a data-mined and machine-learned periodic table.