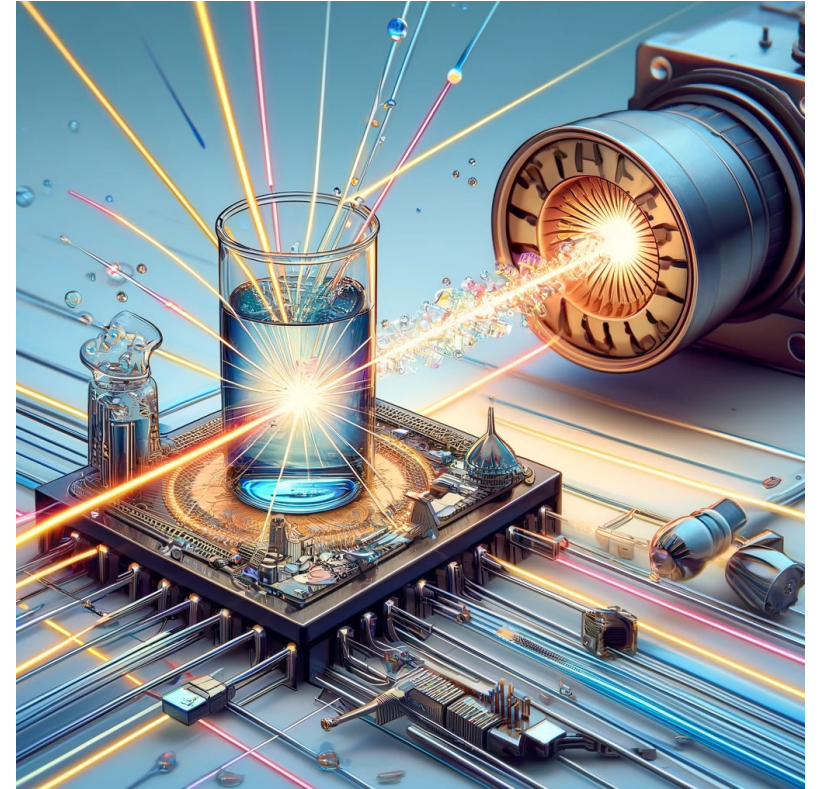


May 2024

Machine Learning Modalities for Materials Science



# Unlocking the Potential of EXAFS: Machine Learning Approaches for Spectroscopic Data

Javier Heras-Domingo

Andrea Ruiz-Ferrando, Stephan Pollit, Adam Clark, Sharon Mitchell,  
Olga Safonova, Prof. Javier Pérez-Ramírez and Prof. Núria López



NCCR  
Catalysis



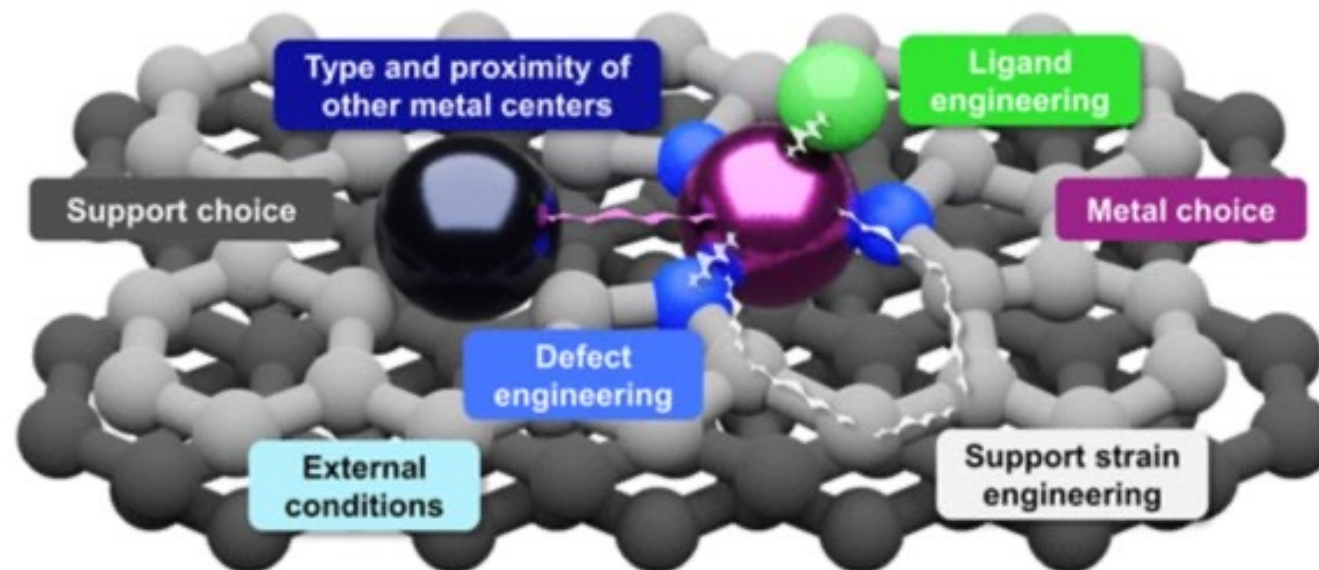
ETH zürich





# Introduction: Single Atom Catalysts (SACs)

SACs **revolutionize catalysis**, excelling in performance, atom utilization, properties, and stability.



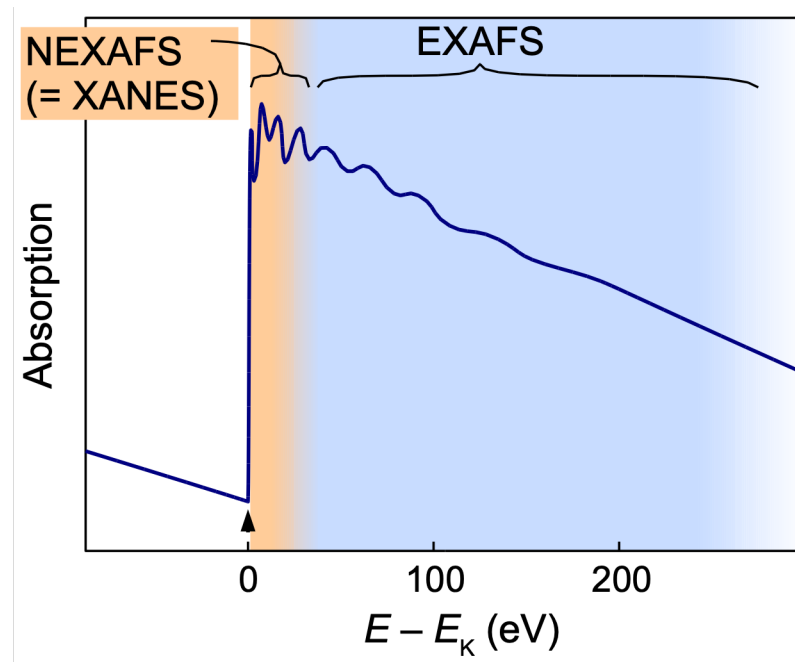


# Introduction: X-ray Absorption Spectroscopy (XAS)

XAS utilizes X-ray radiation to **provide insights** into composition, structure, and bonding.



PAUL SCHERRER INSTITUT  
**PSI**

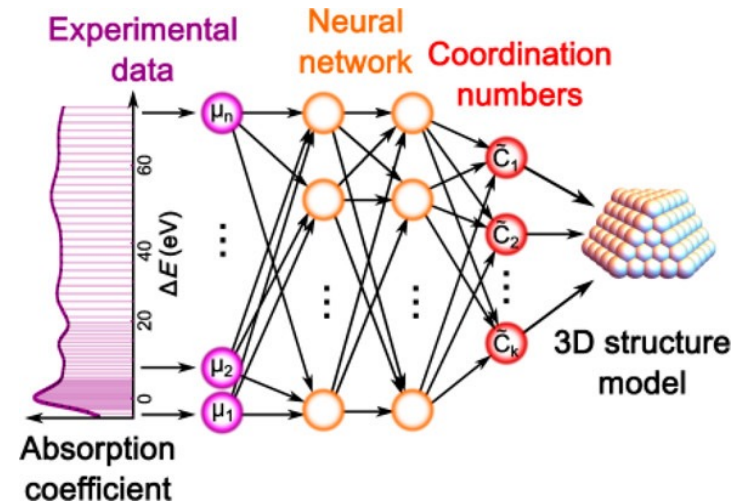




# Introduction: Deep Learning for XAS

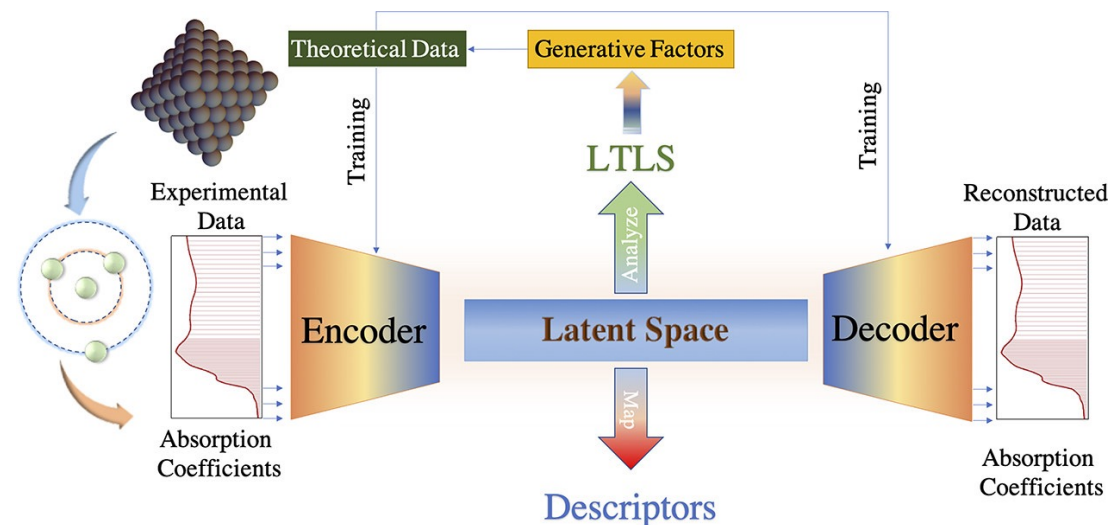
## Supervised Learning:

This kind of ML models can be trained on **specific structural parameters** such as coordination number, distances or **the Pair Distribution functions (PDF)**.

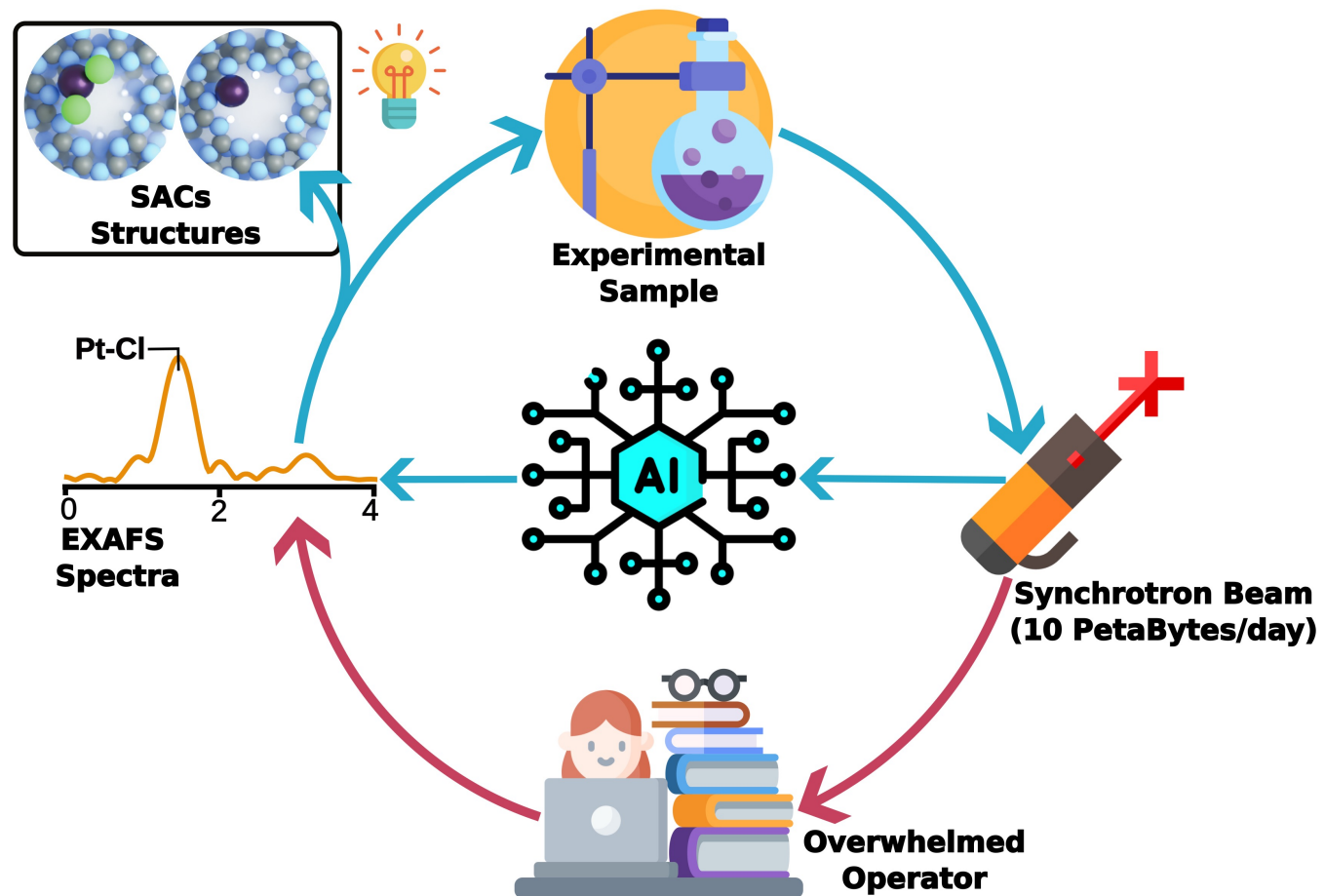


## Unsupervised Learning:

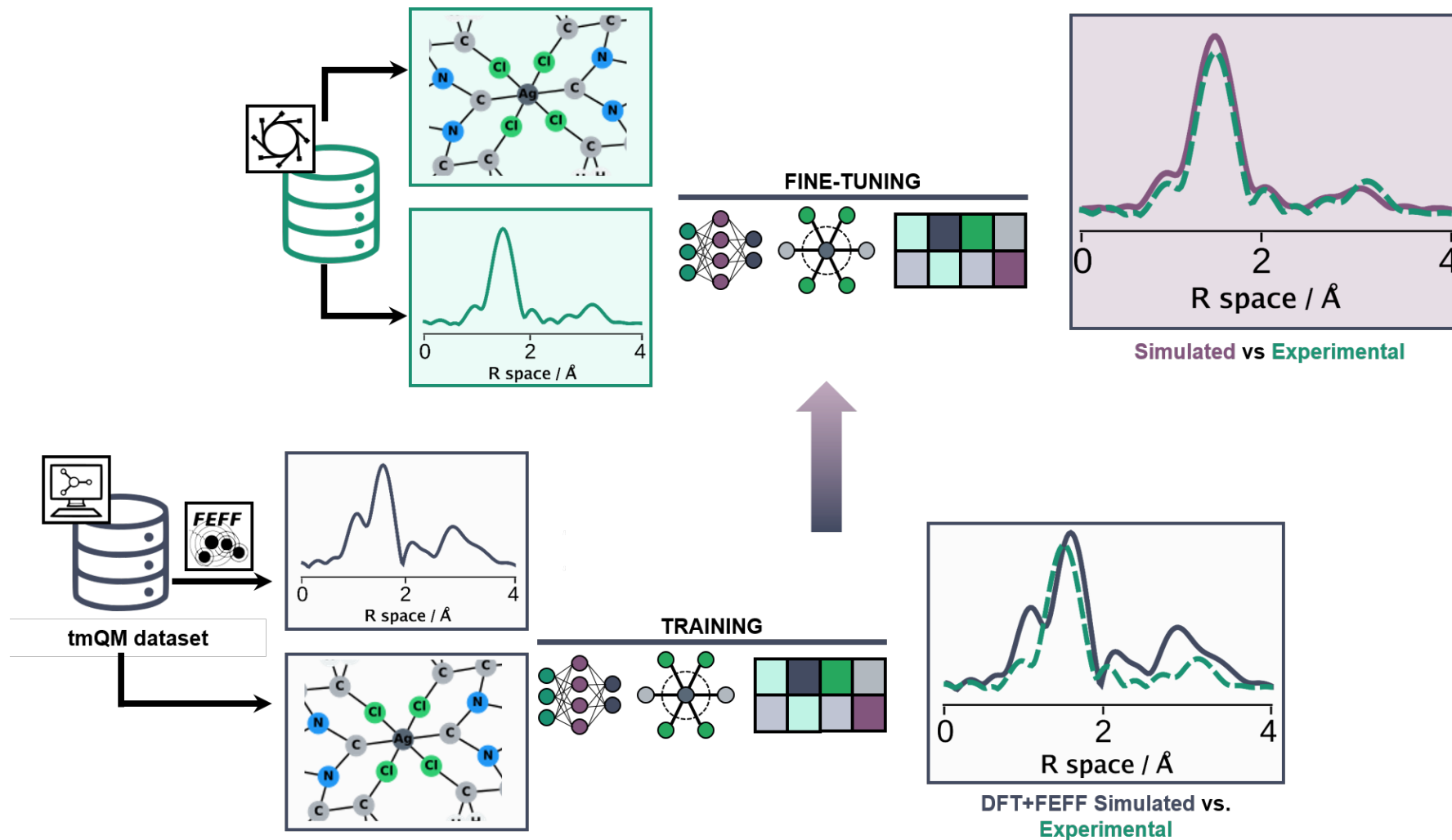
An alternative is to "learn" from a **compressed but still meaningful representation** through unsupervised learning. This unique representation was linked to structural parameters.



# Main Goal



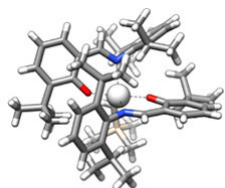
# Results: Training Strategy



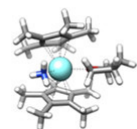
# Results: Theoretical Dataset



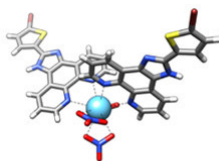
## TmQM\_wB97MV



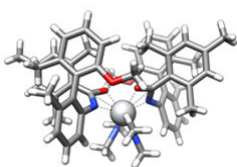
Sc (EGUBIR)



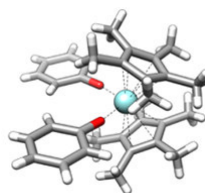
Y (ZERWEB)



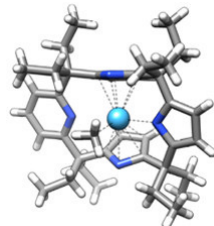
La (SEGSOO)



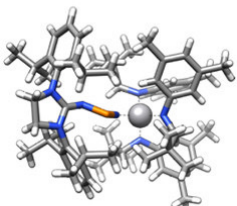
Ti (VUVWOZ)



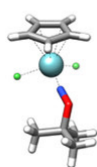
Zr (ZIMYAV)



Hf (YOXTAG)



V (APUPAD)



Nb (TASGAV)

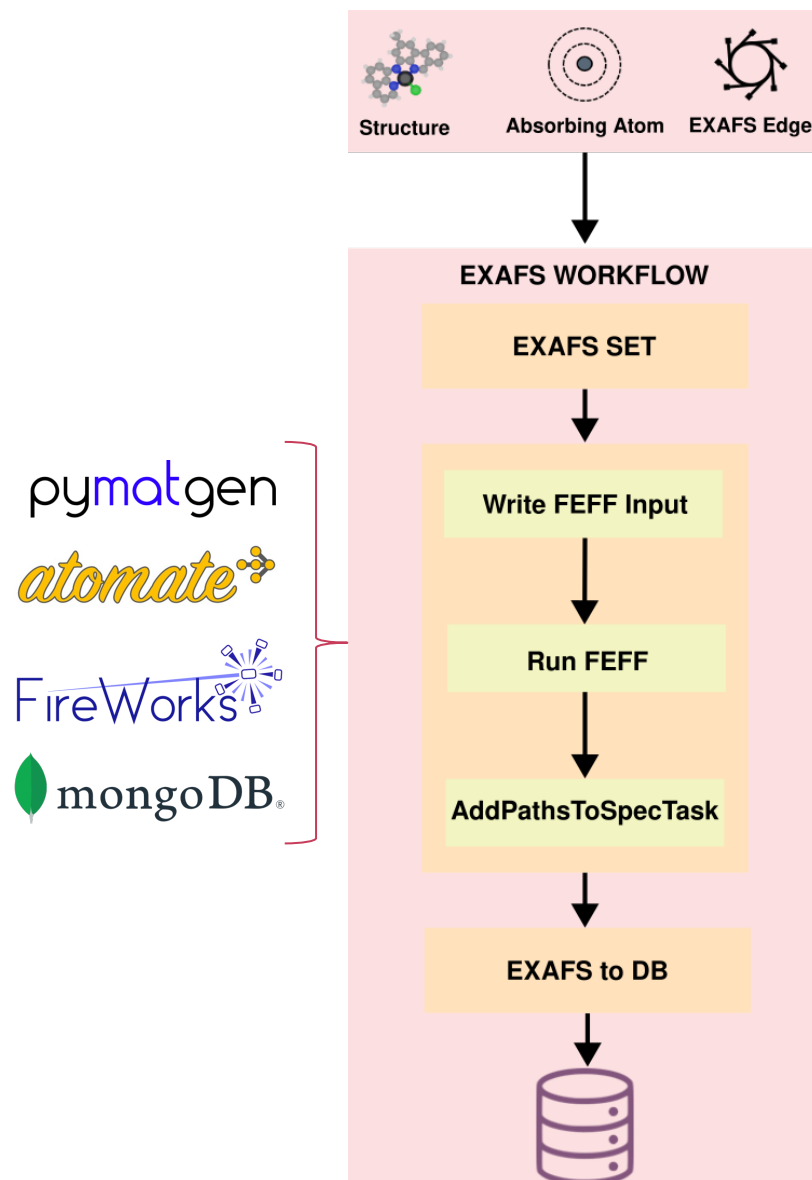


Ta (VIPQEQ)

- **Number of structure: 86k**
- **Molecular**
- **Diverse in terms of structure**
- **Type of data:**
  - **Structure: Experimental**
  - **EXAFS: Theoretical**



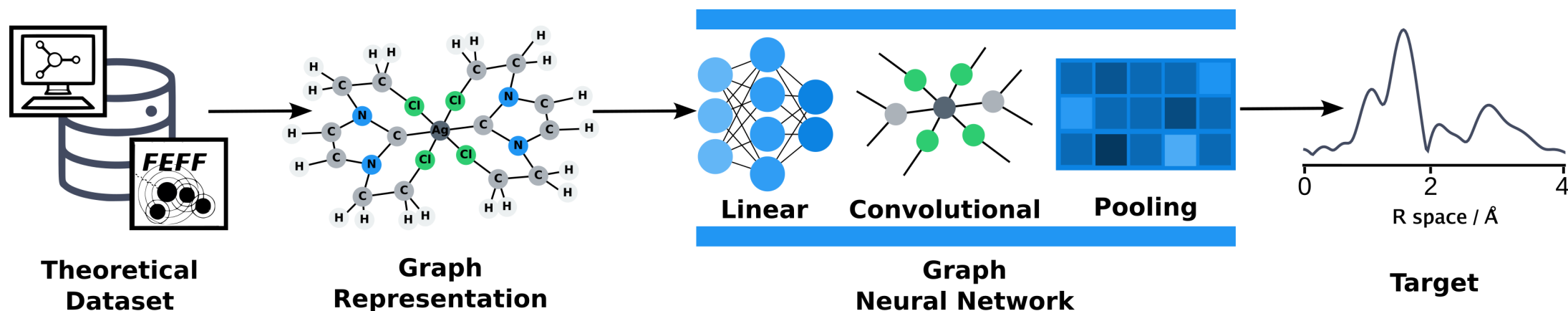
# Results: EXAFS Workflow



- Automatic **decision making**.
- **Reproducibility and traceability** for science.
- The user can **focus on science** instead of tedious tasks.



# Results: GNN Model

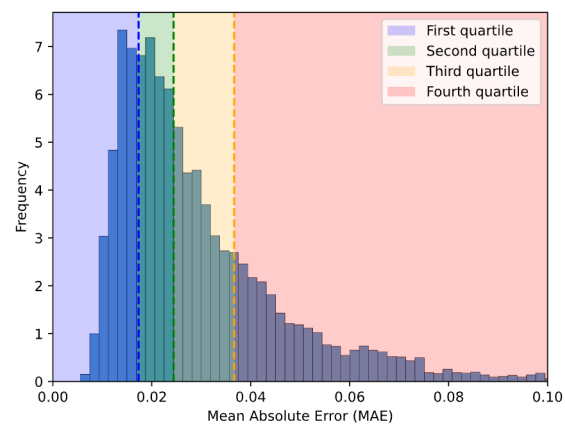


- **Graph Representation:** Nodes are initialized with unique one-hot encoded arrays and edge resembling chemical bonds in the structure.
- **Graph Neural Network (GNN):** The model is constructed with 3 main building blocks, linear transformations, graph convolutions and pooling operations.

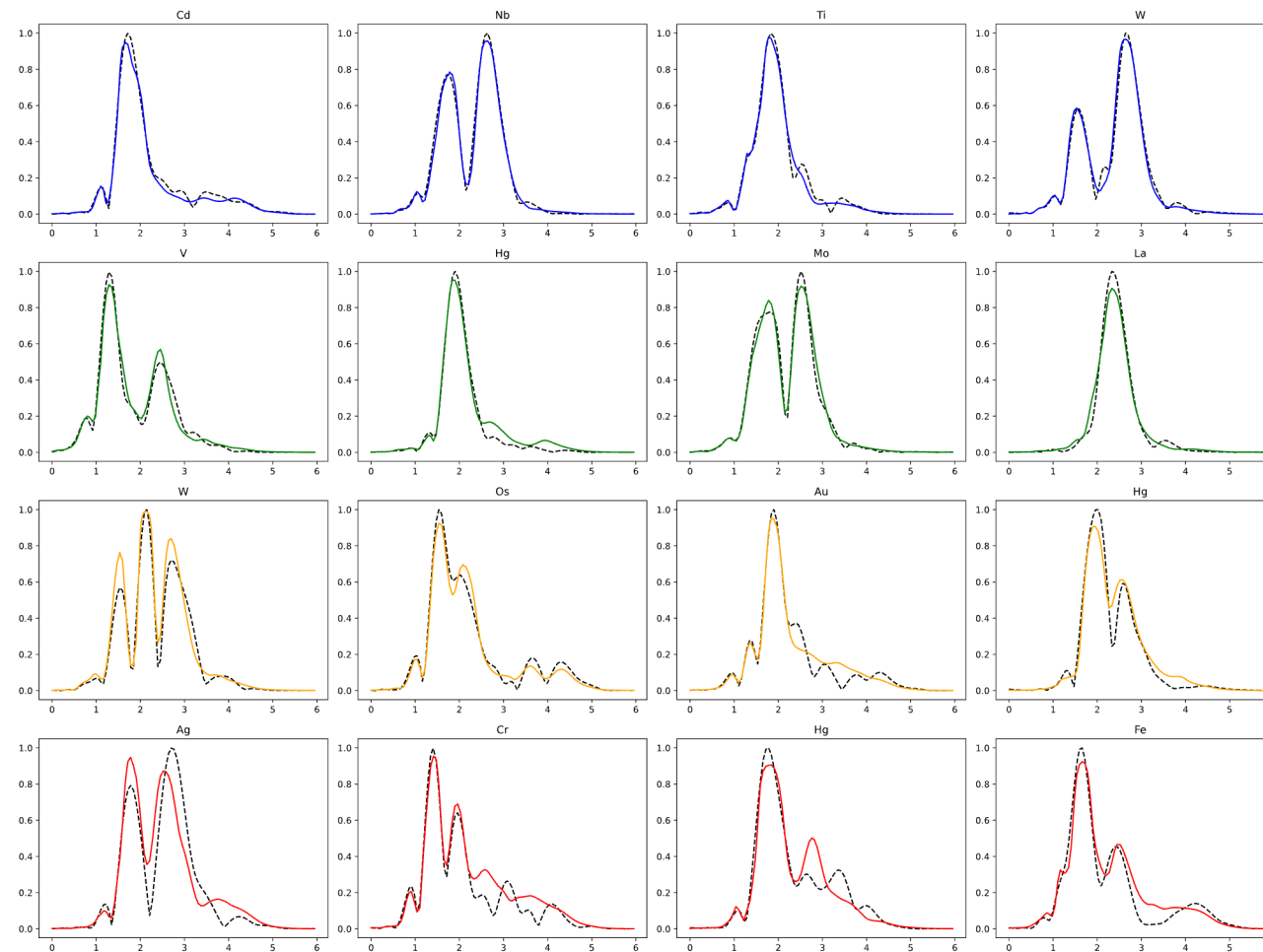


# Results: Model Performance (Theoretical vs Theoretical)

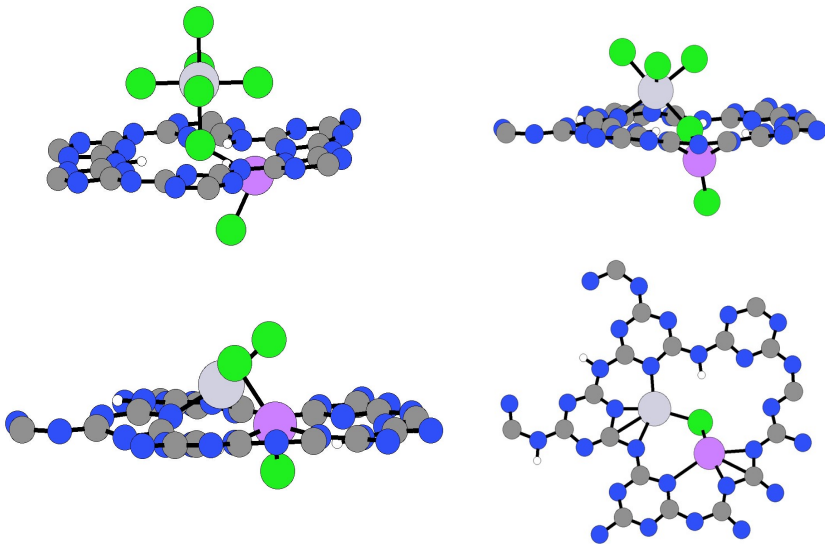
## Training on TmQM\_wB97MV



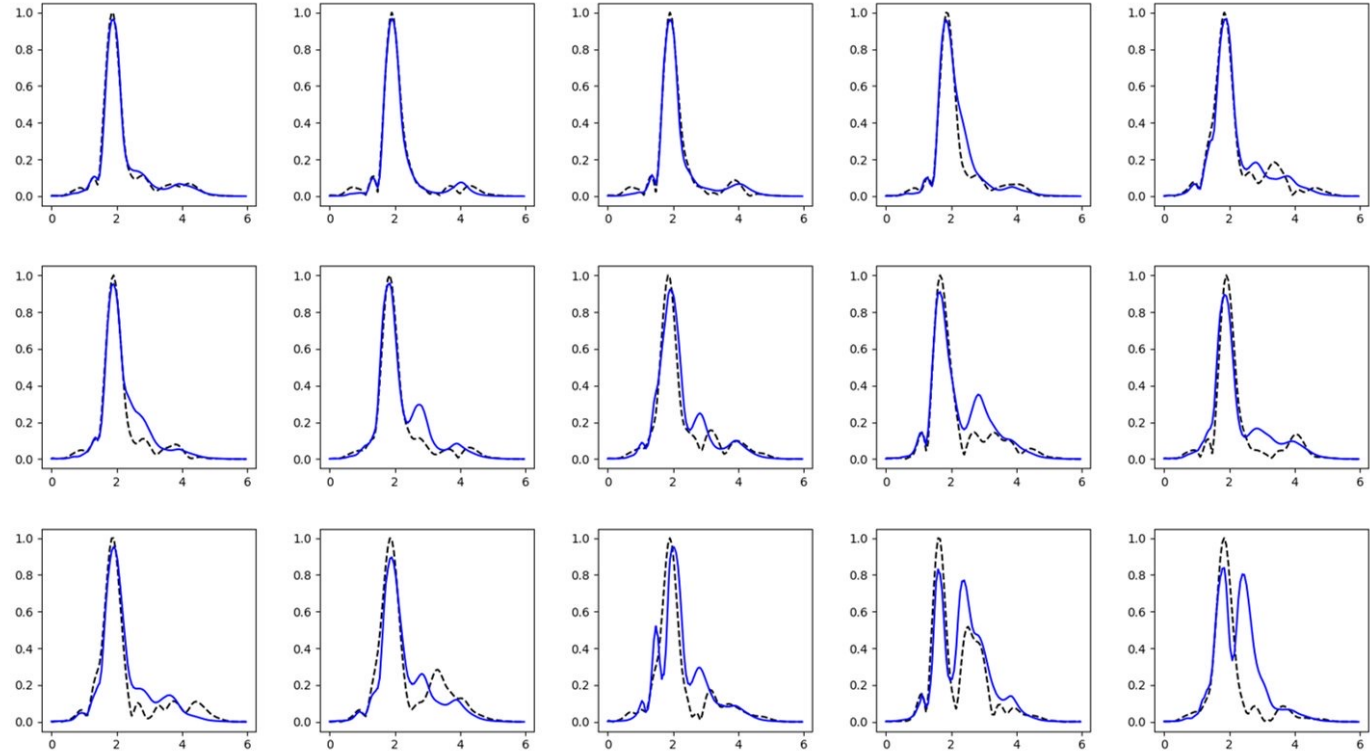
**Performance on 10% Testing  
MAE: 0.026**



# Results: Model Performance (Theoretical vs Theoretical)



Performance on Pt-PTI:  
MAE: 0.044





# Future Steps

## Data Collection:

- Collect experimental data to fine-tune the GNN model

## Deep Learning:

- Implementing different fine-tuning strategies (e.g., Gradual Unfreezing)

## Application:

- Use our method for real world problems

# Collaborations and acknowledgements



**A. Ruiz-Ferrando**



**S. Pollit**



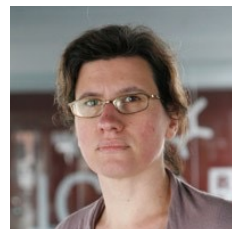
**A. Clark**



**O.V. Safonova**



**S. Mitchell**



**N. López**





**Thank you!**



The National Centres of Competence in Research (NCCRs)  
are a funding scheme of the Swiss National Science Foundation