
University of
Nottingham
UK | CHINA | MALAYSIA
RSE for Data-Driven AI Research
Grazziela Figueredo
University of Nottingham

In times of Software Dev/Engineering living an “existential crisis”…

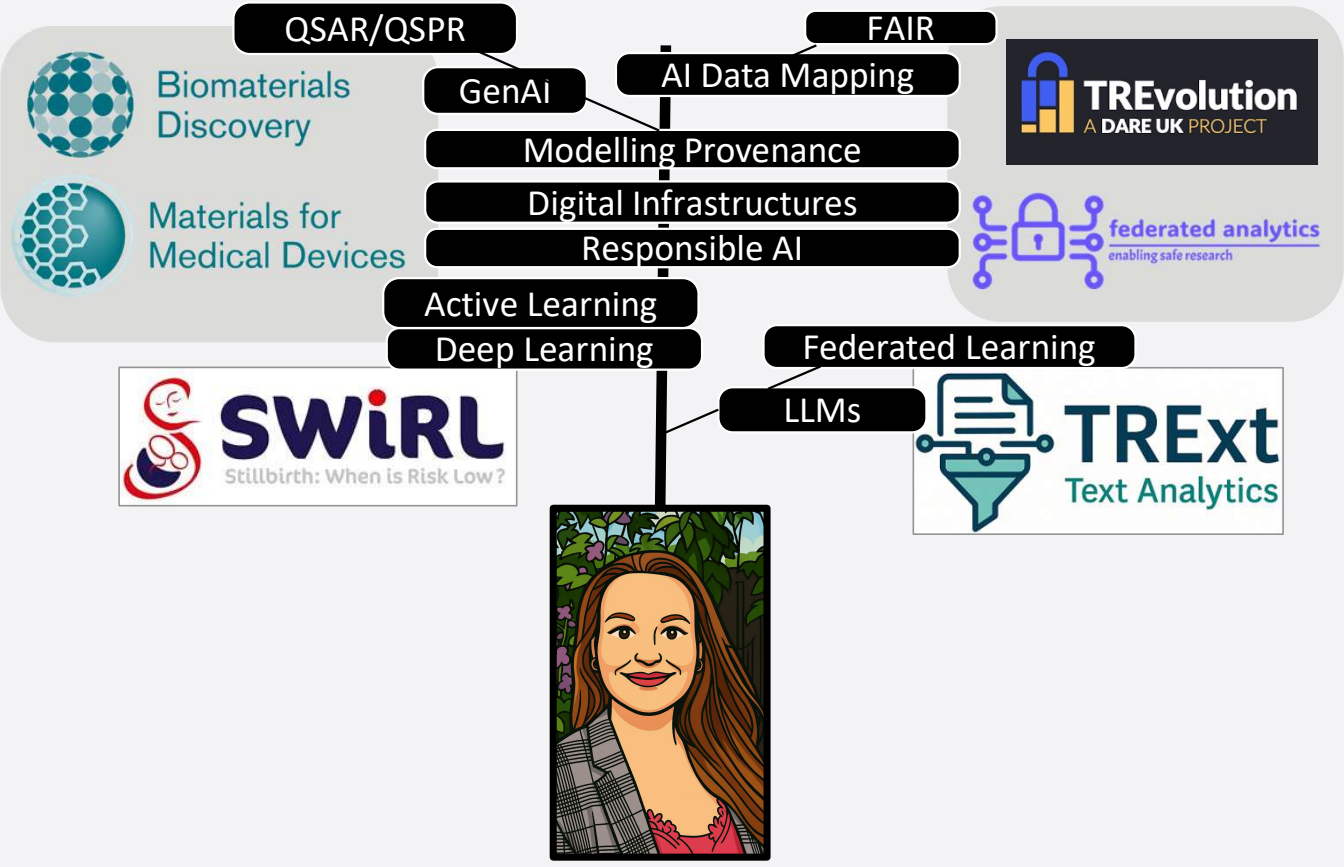
how have we advanced Research and Outputs through the collective power of the University of Nottingham RSE community

About Me

Grazziela Figueredo

Associate Professor in Health Data Science, The University of Nottingham

- Expertise
  - Data science and ML for medical and (bio)materials domains
  - Data science as a service:
    - Research into intelligent products
    - Knowledge transfer
- Worked at/with the DRS in Nottingham
- All projects benefitted from RSE input



Acknowledgements to the Team

Eduardo Aguilar-Bejarano

Rongjun Dong

Jimiama M Mase

James Mitchel-White

Reza Omidvar

Daniel Lea

Troy Kettle

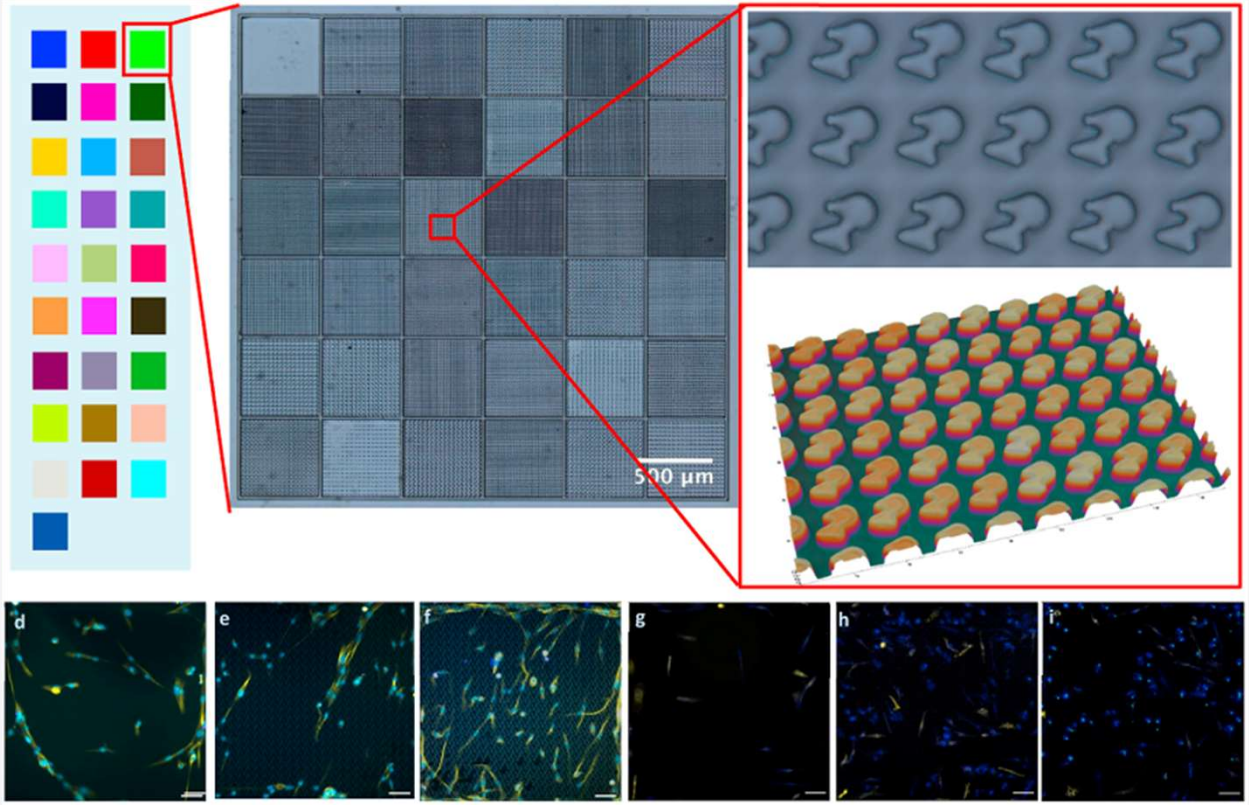
Karthik Sivakumar

and thanks to the Digital Research Service at the University of Nottingham

Biomaterials Discovery

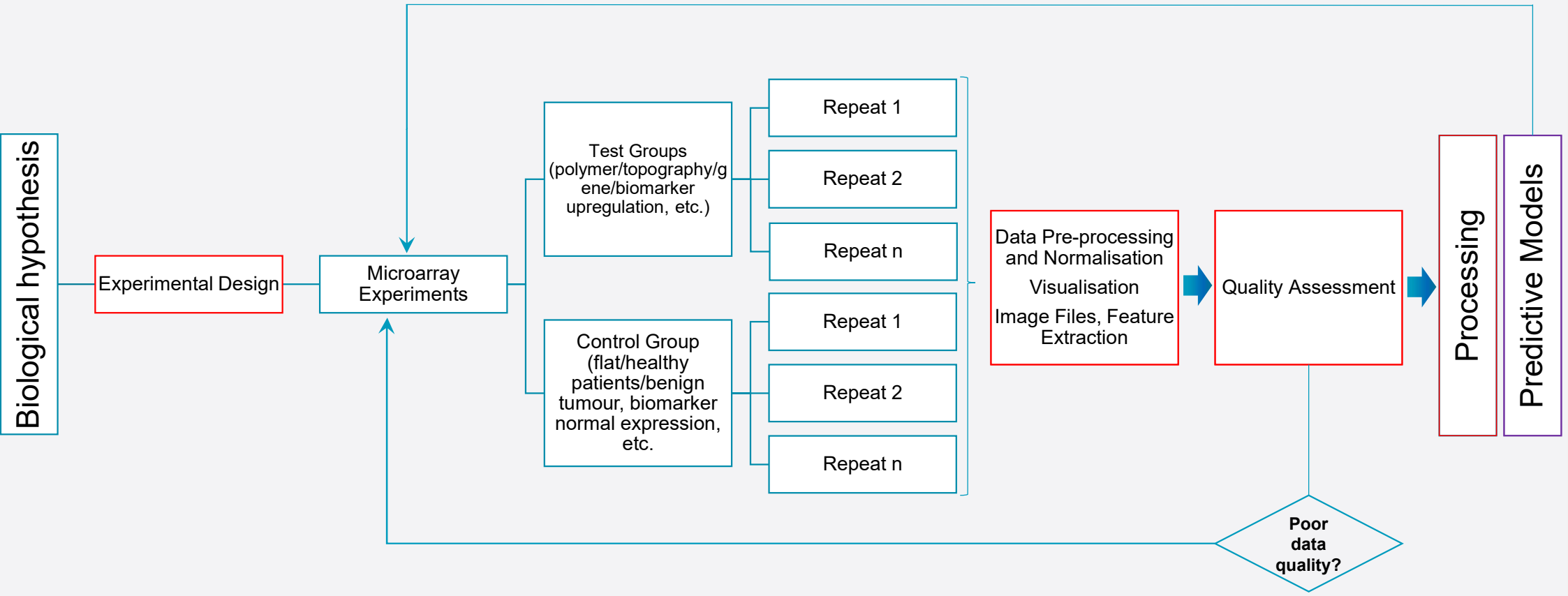
High-throughput Screening

- Chemistries
- Topographies
- Various cellular responses
- Rich data



L. Burroughs et al. 2021. Discovery of synergistic material-topography combinations to achieve immunomodulatory osteoinductive biomaterials using a novel in vitro screening method: The ChemoTopoChip, Biomaterials, 271.

A Generic Analytics Pipeline



Overall Challenges in Data-Driven Discovery of Cell-Instructive Materials

Understanding interplay between chemistry, topography and biology

Understand synergy between chemistry and topographies

Mechanistic interpretation

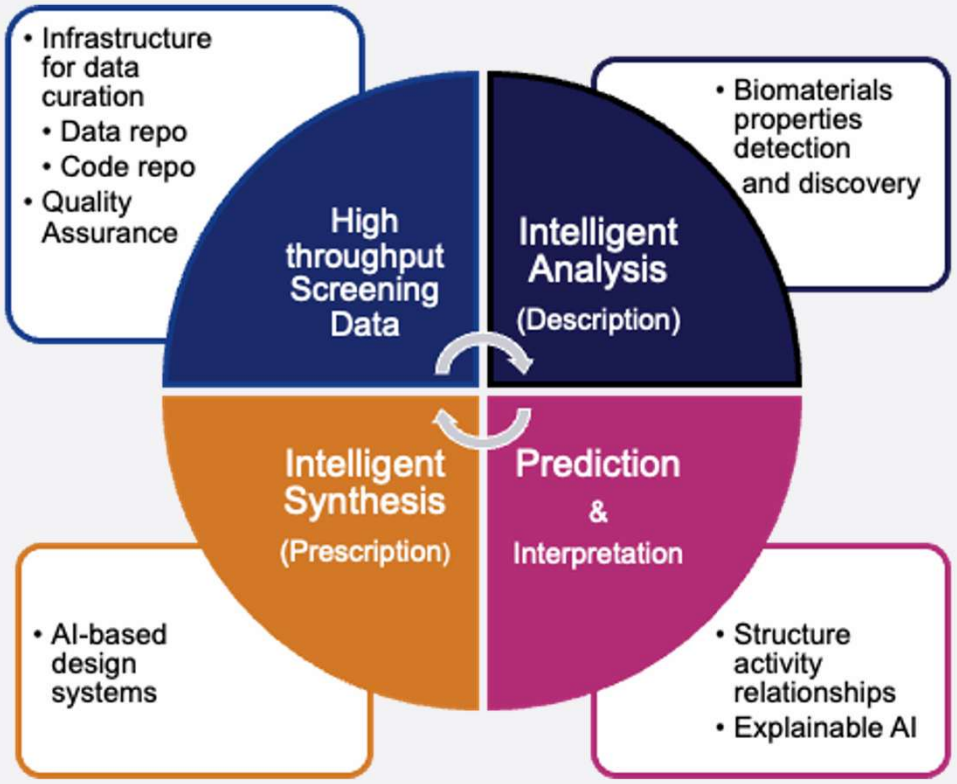
Reliable FAIR (findable, accessible, interoperable, reusable) data

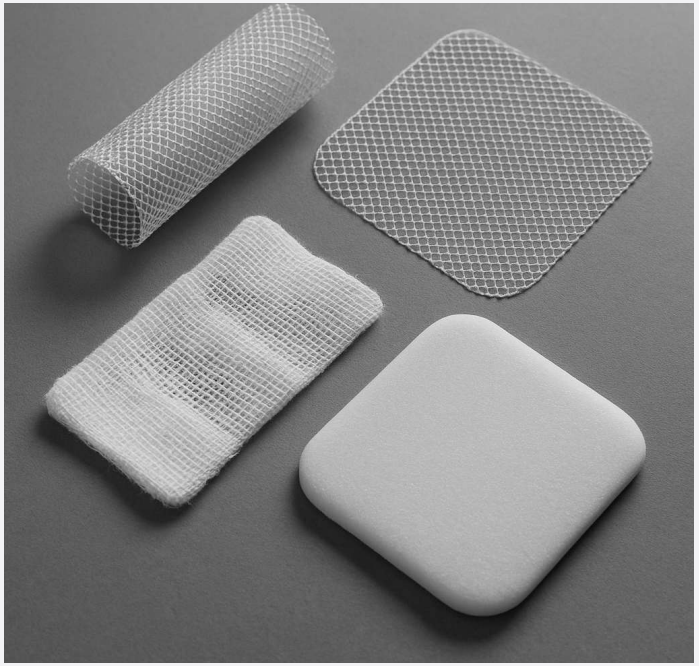
AI (the black box) as a bridge between materials and biology?

- AI Explainability
- AI Reliability
- FAIR models

Materials for Medical Devices

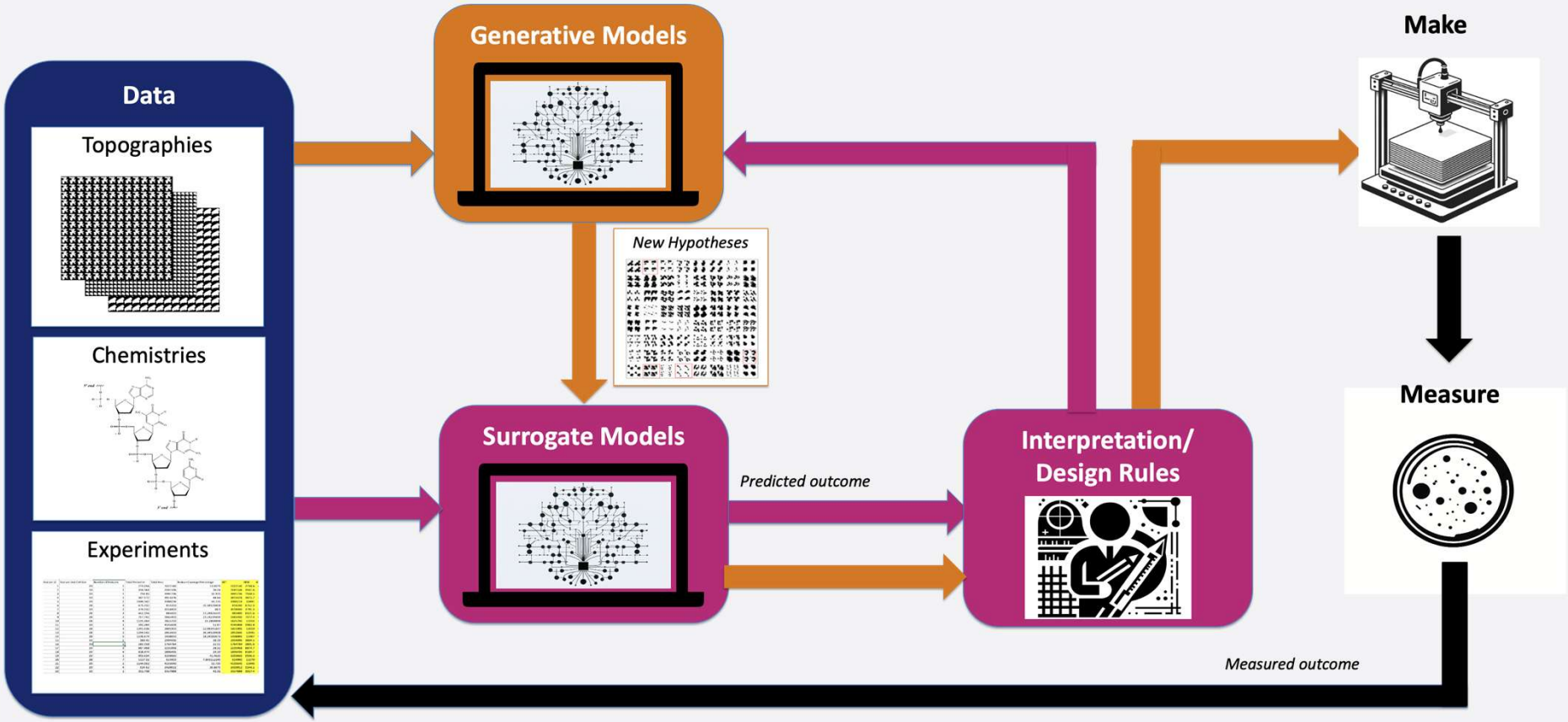
Digital infrastructure for intelligent analysis, interpretation and closed-loop synthesis of Biomaterials chemistries and topographies



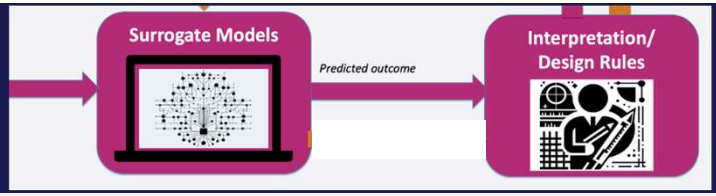


Materials for Medical Devices

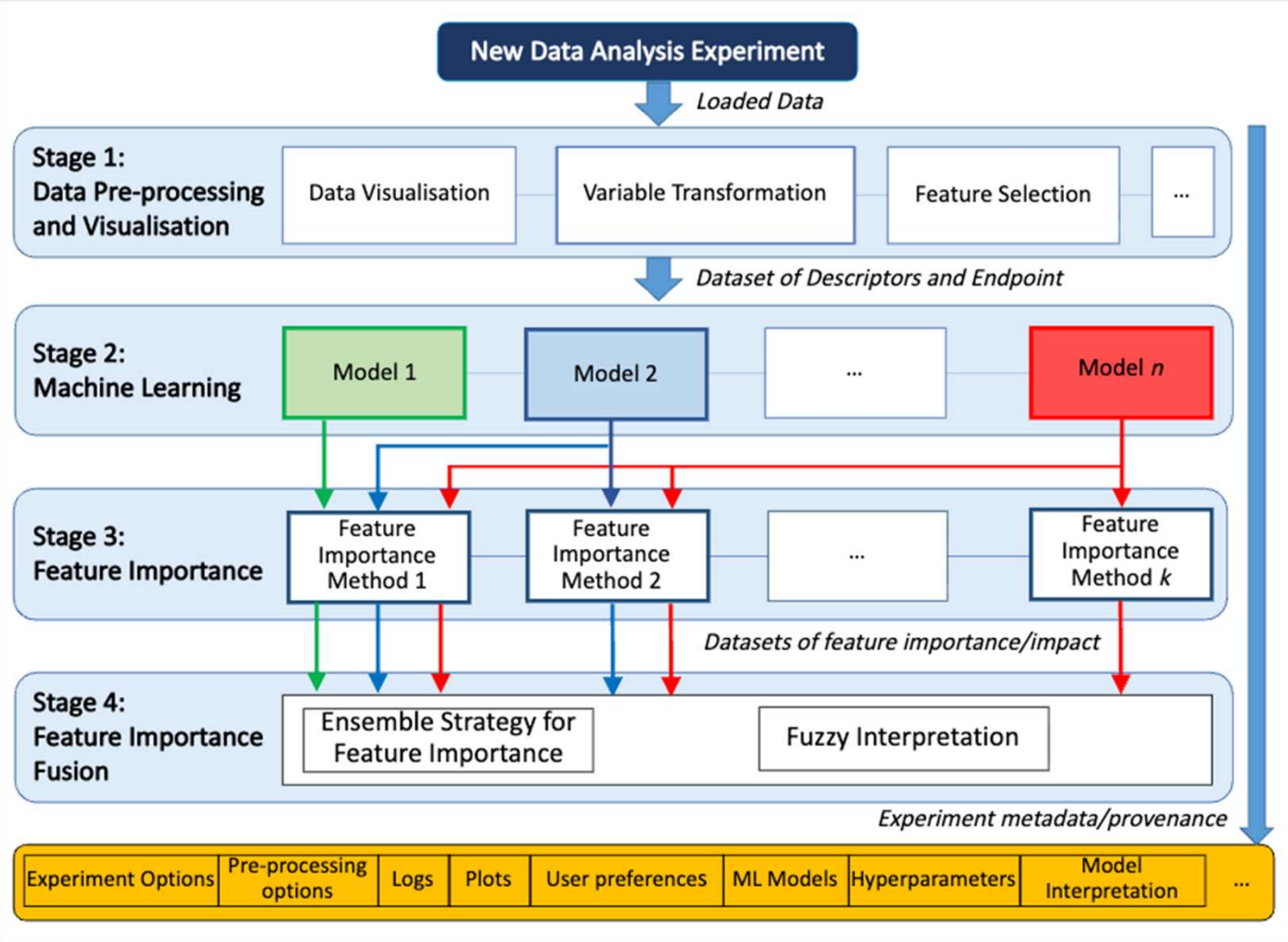
My Research within the Project Ecosystem



Tabular Data - Traditional Models



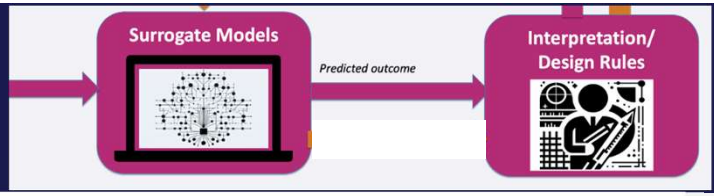
- Robust explainability
- Comprehensive design rules
- Analysis and modelling provenance



Rengasamy, D et al. Feature importance in machine learning models: A fuzzy information fusion approach, Neurocomputing, 2022, 511.

Rengasamy, D.; Rothwell, B.C.; Figueredo, G.P. Towards a More Reliable Interpretation of Machine Learning Outputs for Safety-Critical Systems Using Feature Importance Fusion. Appl. Sci. 2021, 11, 11854.

Tabular Data Solution - Helix



HELIX

README MIT license

Helix: Python Toolkit for Machine Learning, Feature Importance, and Fuzzy Interpretation

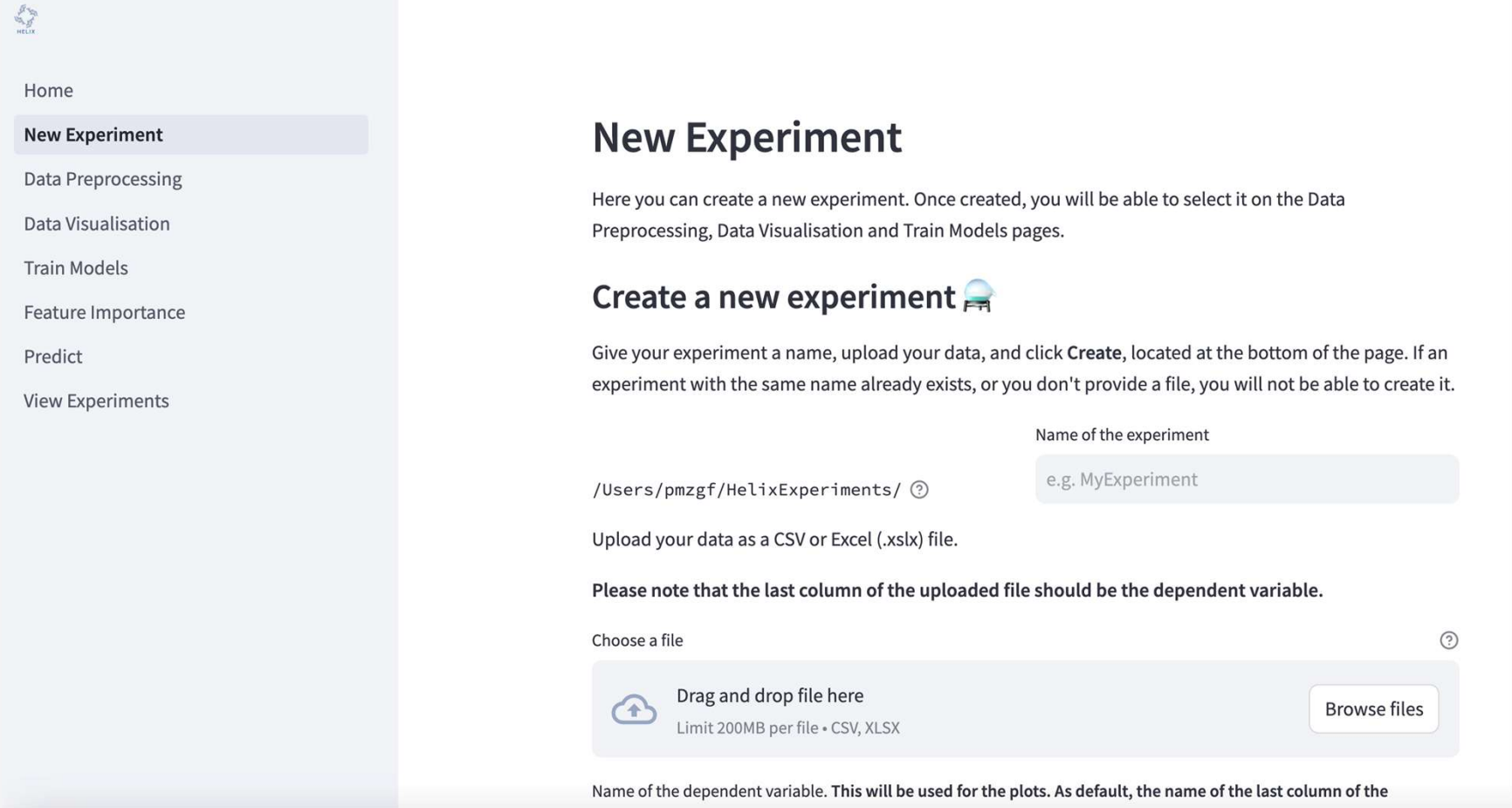
LICENSE MIT | PYTHON | UV | SCIKIT-LEARN | MATPLOTLIB | LINUX | MAC OS | WINDOWS

ISSUES 7 OPEN | BUILD DOCS PASSING | PUBLISH DOCS PASSING | CODE QUALITY PASSING

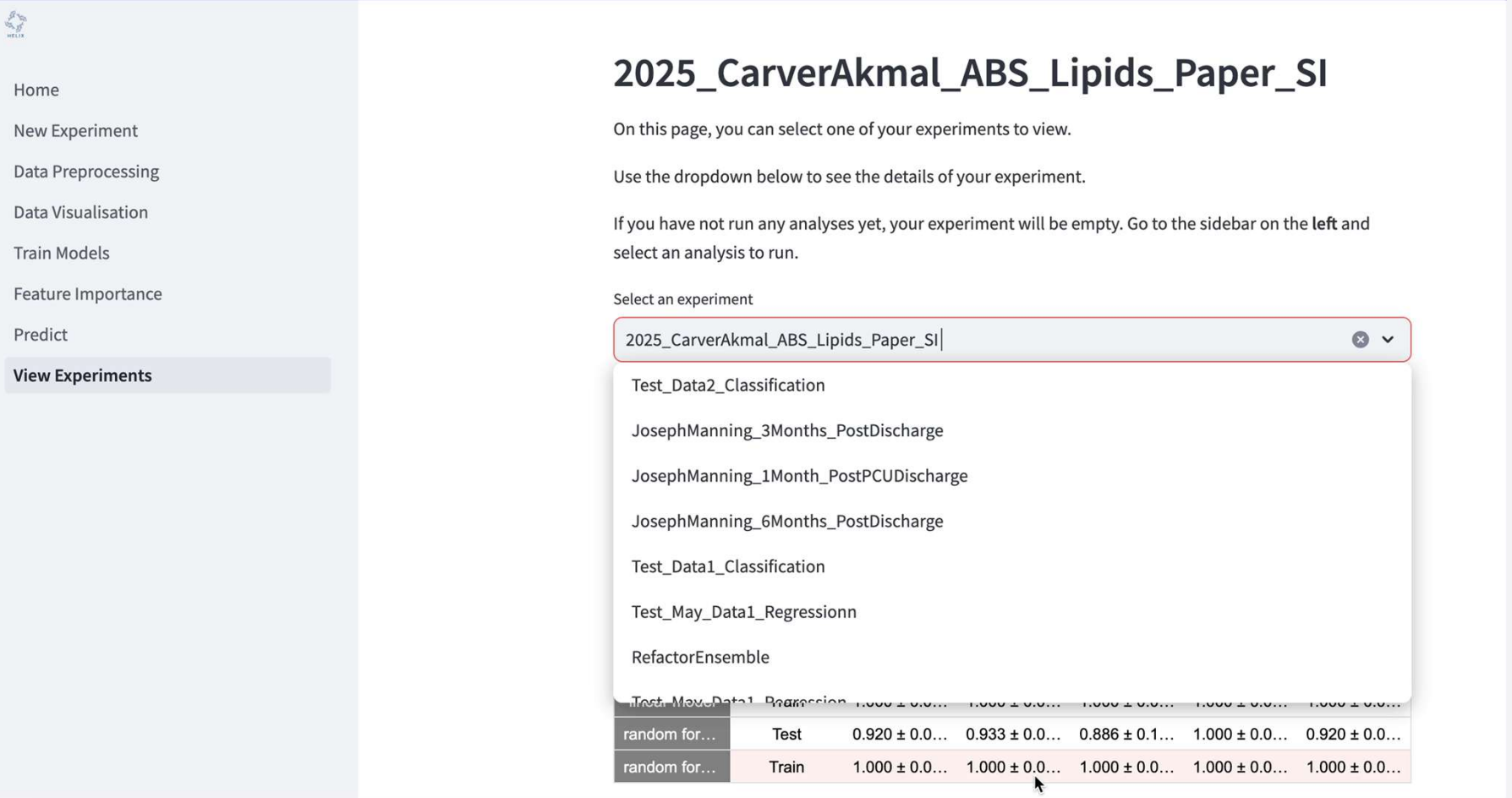
TESTS PASSING | DOWNLOADS 25/MONTH

Aguilar-Bejarano, E., Lea, D., Sivakumar, K., Mase, J.M. et al. (2026) 'Helix 1.0: An open-source framework for reproducible and interpretable machine learning on tabular scientific data', Patterns. Available at: https://doi.org/10.1016/j.patter.2026.101536

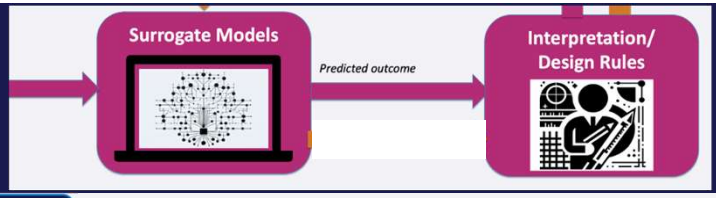
Helix – Analysis and Modelling



Helix Provenance

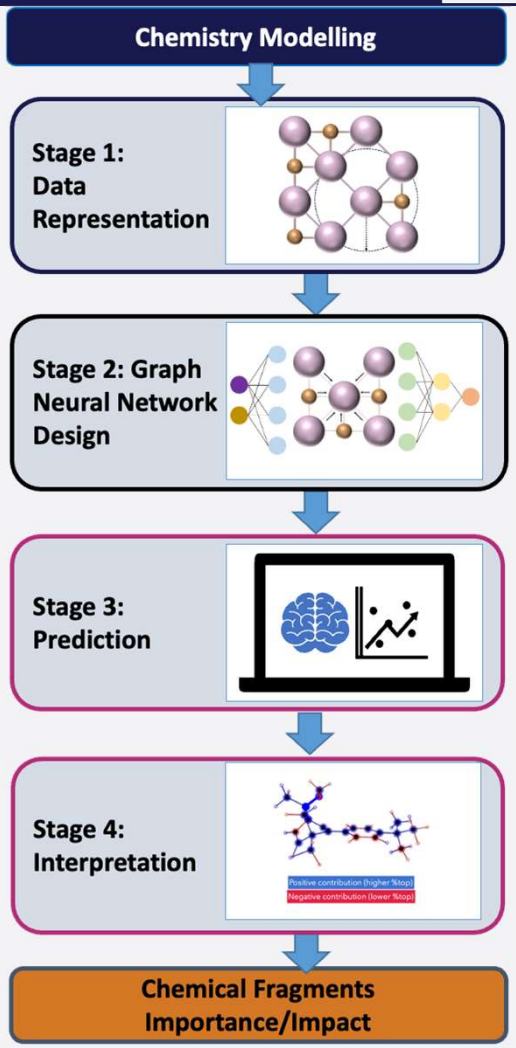


Graph Representation of Polymers



Polymer Modelling

- Multiple representations
- Graph-based
- Transformer-based
- Etc
- Means for interpretation



Graph Representation of Polymers - PolyNet

Eduardo Aguilar-Bejarano

PolyNet

python 3.9+ | license MIT | PyTorch 2.x | PyTorch Geometric enabled | RDKit supported | Tests passing | code style black | dependency management poetry | status active development

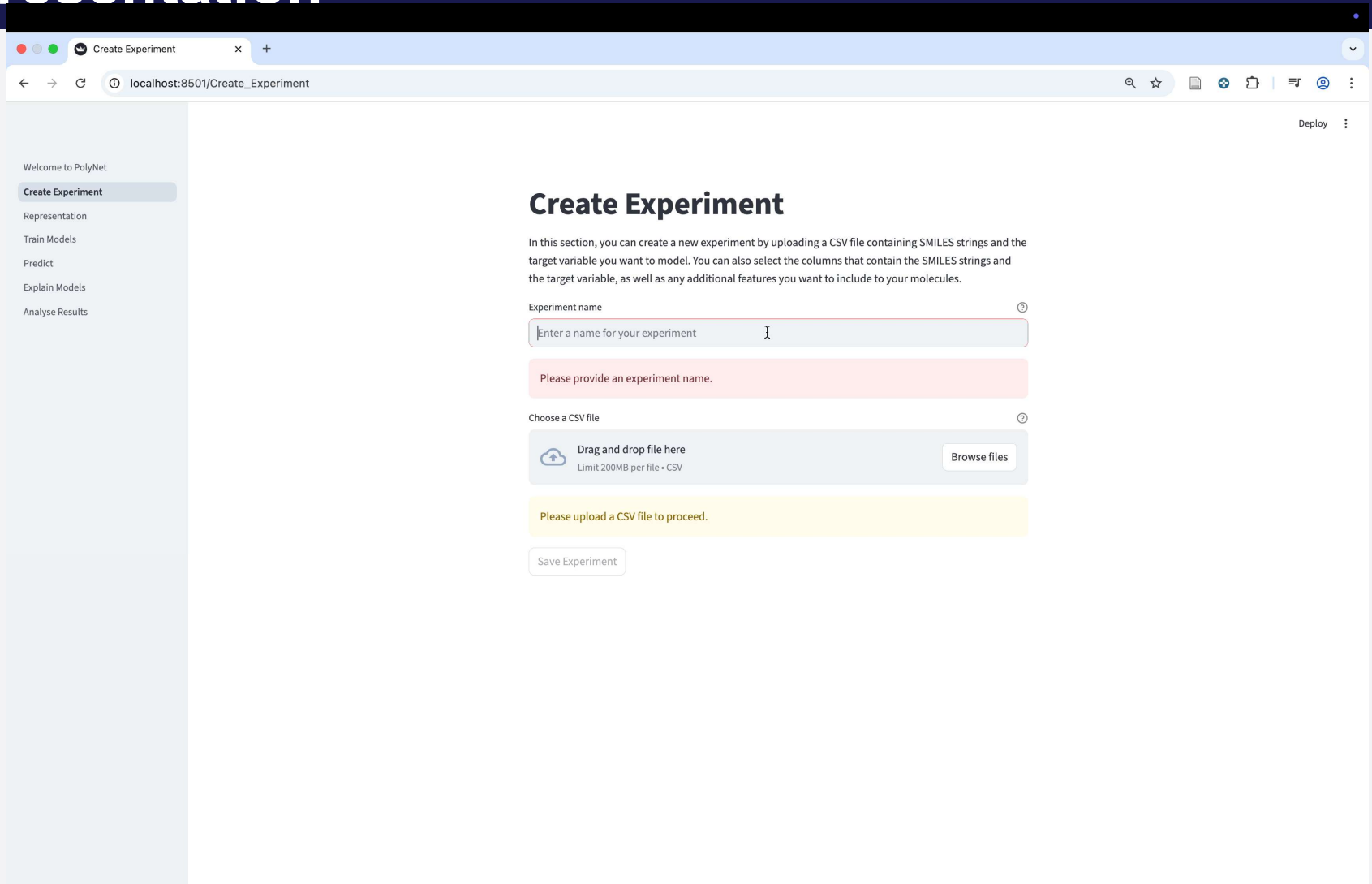
PolyNet
POLYMER INFORMATICS

PolyNet is a Python library for polymer property prediction using graph neural networks (GNNs) and traditional machine learning (TML). It provides a complete, configurable pipeline — from raw SMILES strings to trained models, evaluation metrics, result plots, and atom-level explainability — designed for polymer informatics research.

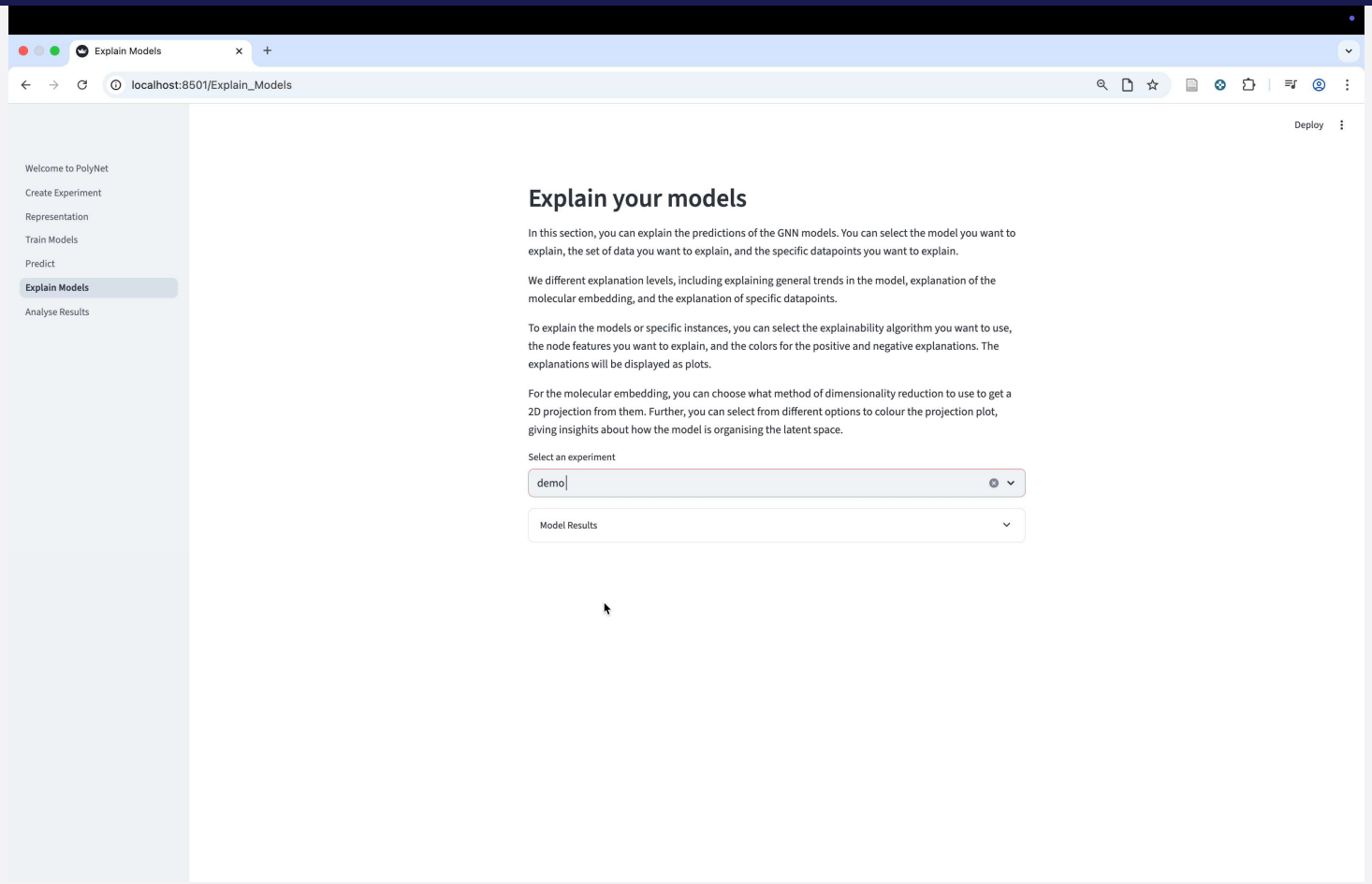
The library is split into two layers:

- **Core package (`polynet`)** — pipeline stages, model training/inference, featurisers, config schemas, and path/IO helpers. No Streamlit dependency; importable from any Python script or notebook.
- **Optional GUI (`polynet.app`)** — a Streamlit web application that wraps the same core stages. Install only when you want an interactive interface.

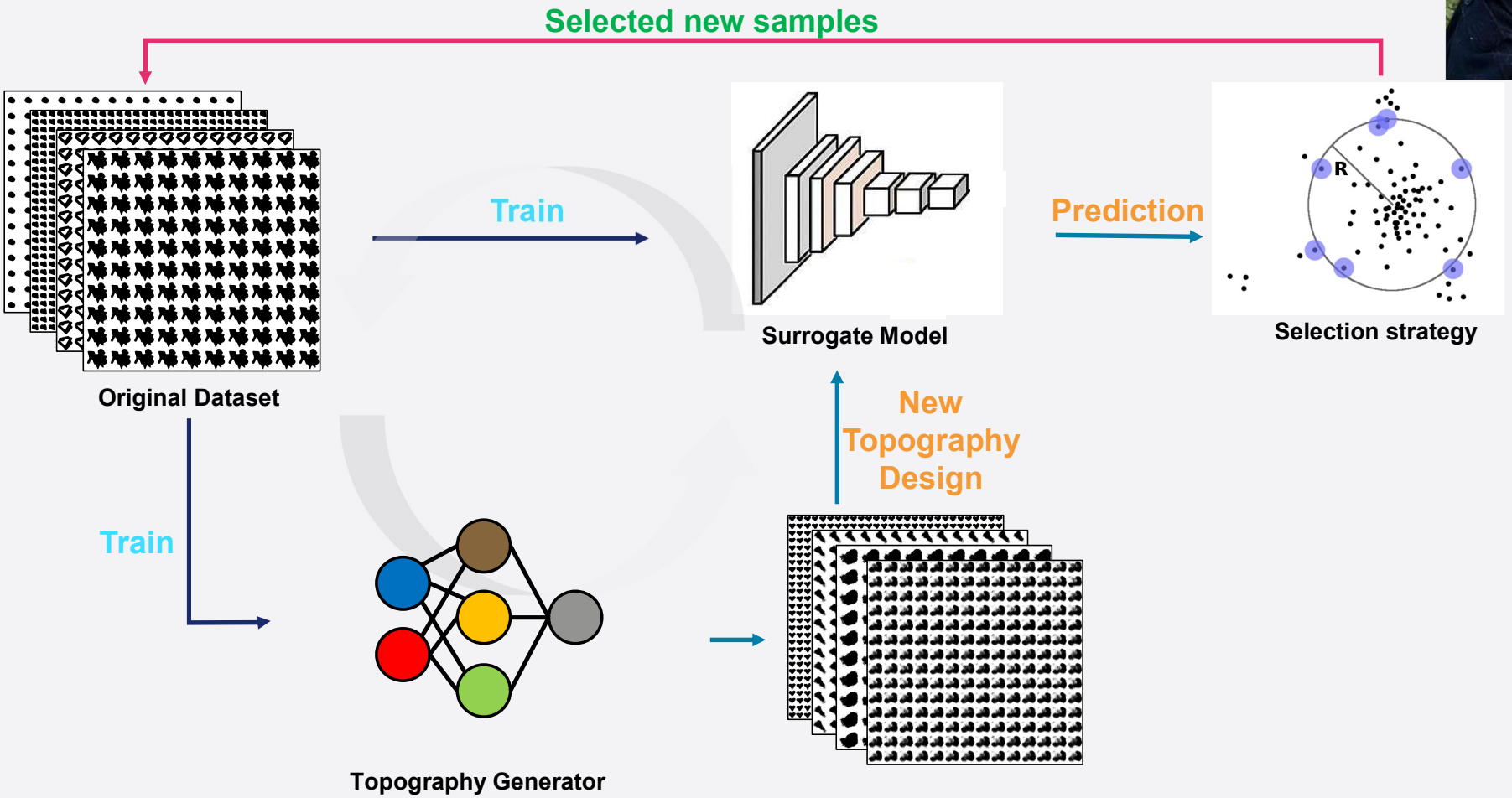
Unstructured Chemistry Data – PolyNet Data Representation



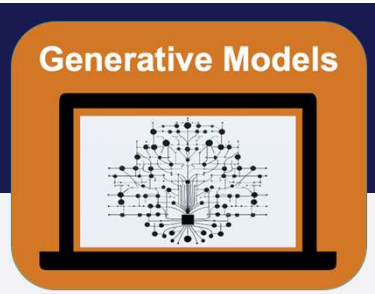
Unstructured Chemistry Data – PolyNet Model Interpretation

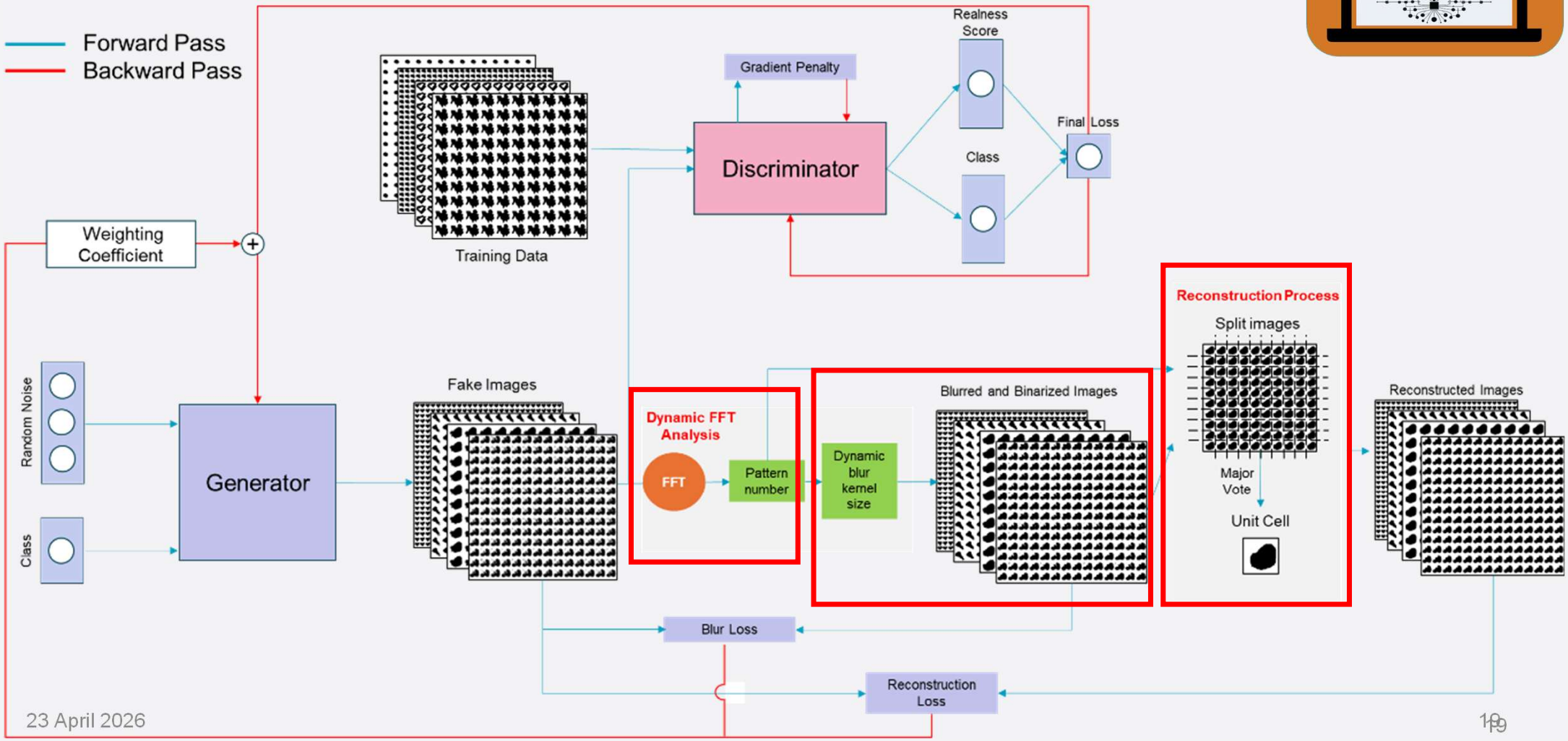


Proposed Solution – Design of Topographies

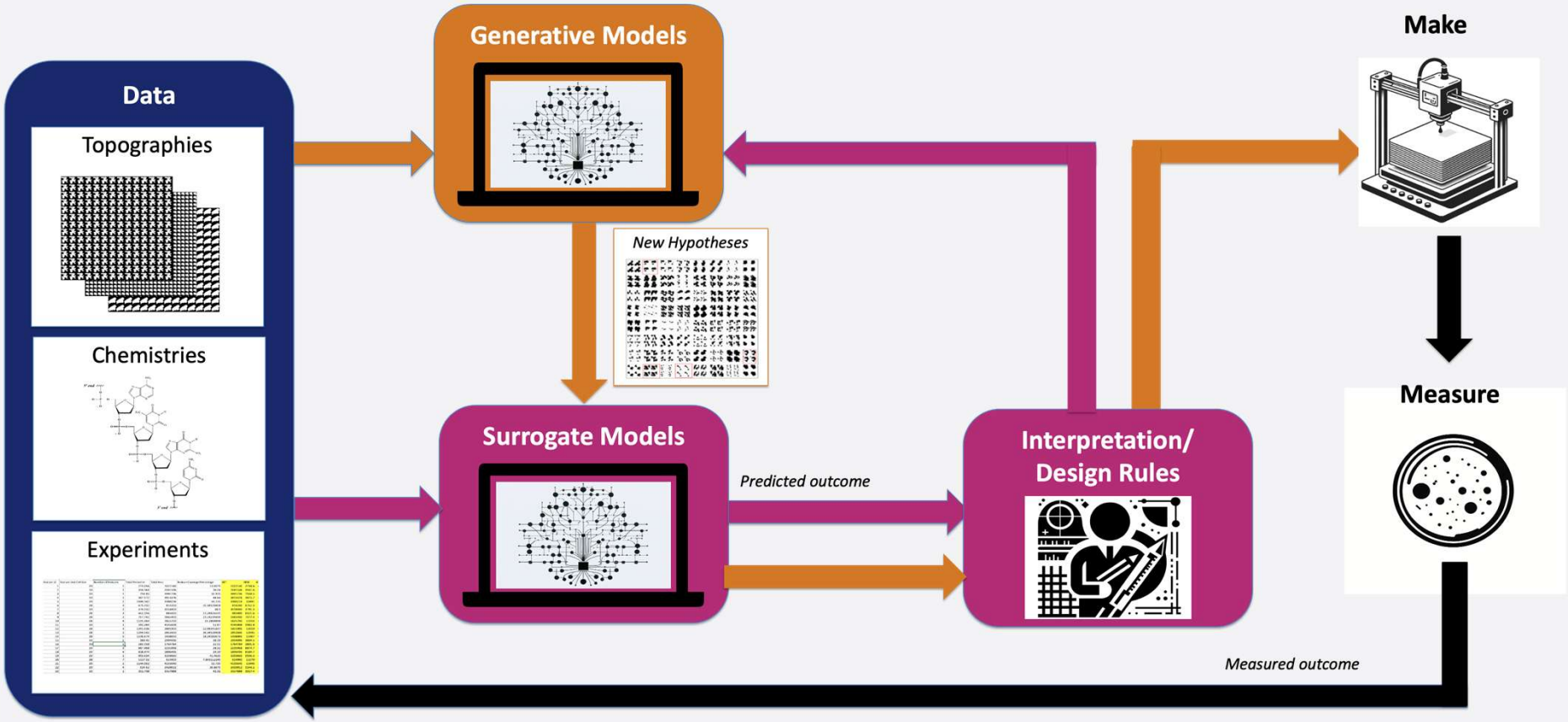


The Dynamic FFT Conditional Blurring Generative Adversarial Network (DF-ACBlurGAN)





Next steps: engineer the closed loop



Final Remarks

- Intelligent collaborative tools focused on
  - FAIR data and models
  - Explainable
  - Human-cooperation
  - Community driven (open source, with open backlog)
  - The value of RSE for FAIR close-loop research to accelerate discovery

Example of communities of practice in Nottingham fostering collaboration, peer-learning, software better fit for purpose, and sharing ideas and best practice

University of Nottingham
UK | CHINA | MALAYSIA