Automated Generative Experimental Design via Relational Learning-Guided Large Language Models

Daniel Brunnsåker[1], Alexander H. Gower[1], Erik Y. Bjurström[1], Prajakta Naval[1], Filip Kronström[1], Ievgeniia A. Tiukova[1,2] & Ross D. King[1,3,4]

[1] Chalmers University of Technology, Gothenburg, Sweden
[2] KTH Royal University of Technology, Stockholm, Sweden
[3] Cambridge University, Cambridge, United Kingdom
[4] Alan Turing Institute, London, United Kingdom

CHALMERS

Introduction

Large Language Models (LLMs) have revolutionized AI, achieving impressive results on tasks traditionally requiring human intelligence. At the forefront of applying AI in science is the automation of research, offering faster discoveries and new insights. By using traditional models and logic to generate hypotheses, LLMs can then be given a logical scaffold to reason around, reducing hallucinations and enhancing the reliability of its outputs. Here we demonstrate a way to leverage the use of LLMs to automate experimental design, given relational learning derived hypotheses and physical laboratory constraints. We integrate this methodology with an automated laboratory cell, demonstrating a way of flexibly doing automated scientific discovery in a predefined hypothesis space.

1 Establishing hypothesis space

The **hypothesis space** is defined through a Datalog database with **applicable phenotypes,** combined with ontological descriptors of involved treatments.

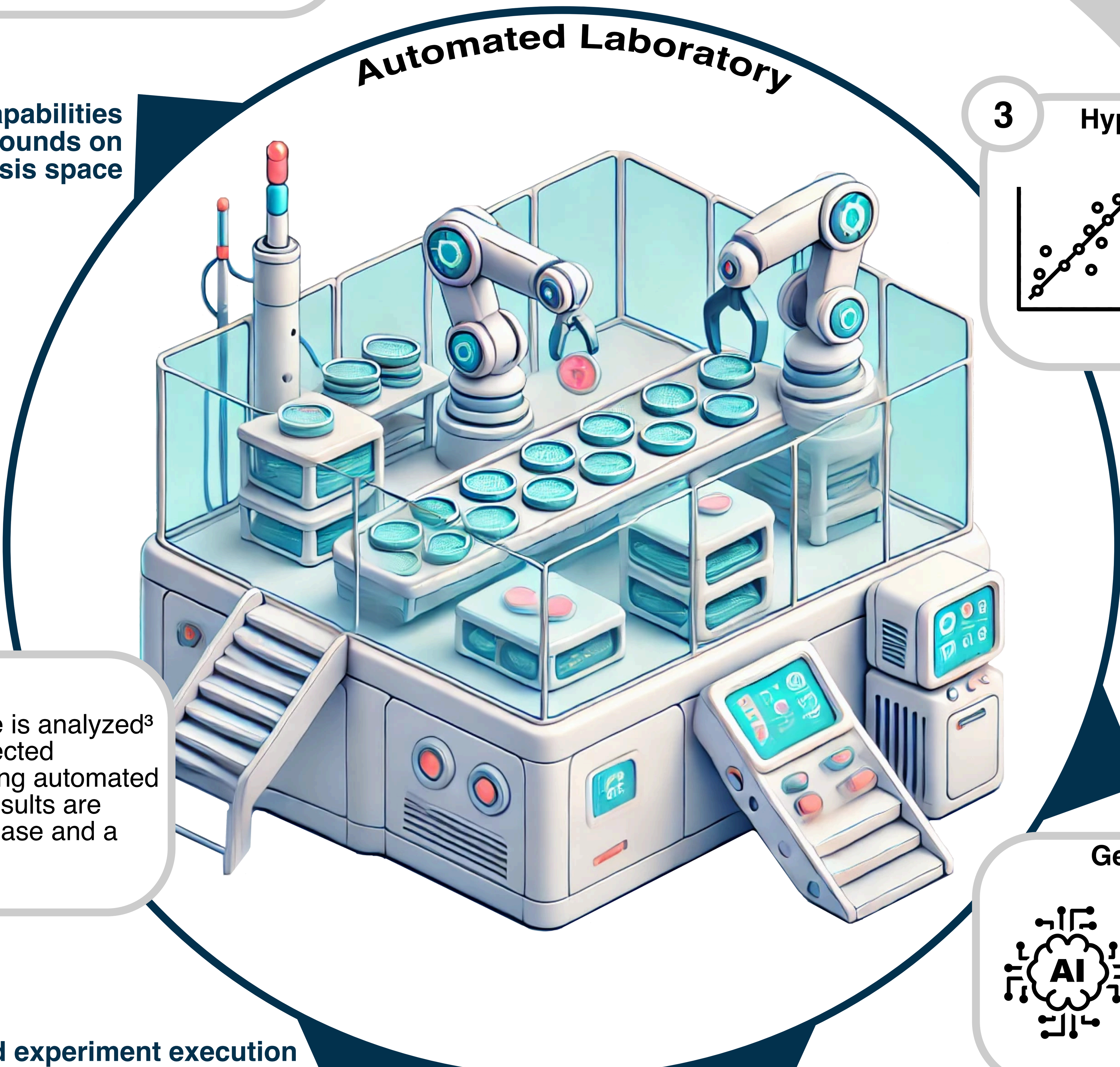
2 Pattern mining & Hypothesis Generation

Human readable Logic Programs are generated by mining the database for frequent sub-patterns using the inductive logic programming (ILP) based algorithm **WARMR**[1].

Example(A) :- Phenotype(A, Increased Resistance to drug, B, C), Drug(B, Aspirin), Condition(C, Nitrogen rich media)

3 Hypothesis Filtering & Weighting

Learn an association between logic programs and a biological observable through regularized linear models, creating a **descriptive hypothesis** concerning a measurable biological observable (e.g. metabolite)[2].



Laboratory capabilities provides bounds on hypothesis space

Hardware availability constrains design-space

Control software determines formalization

Automated experiment execution through standardized lab protocols

4 Generative Experimental Design

The hypothesis is passed to an **LLM** along with hardware and feasibility constraints. From that, automatically **generate an experimental design** to prove or disprove generated hypothesis

5 Experiment Formalization

Experimental design is **automatically formalized** in adherence to the interface of the automated laboratory, producing a standardized, machine readable, experimental protocol in JSON

6 Running the experiment

Experiment is **automatically performed** in an **automated laboratory cell** along with integrated phenotyping, sampling and metabolomics readouts based on ion mobility mass spectrometry[3]

7 Automated analysis

Experimental outcome is analyzed[3] and compared to expected outcomes from (4) using automated hypothesis testing. Results are added to knowledge-base and a report is generated.

Conclusion

Generative experimental design is a promising avenue for enabling the complete automation of science. By integrating traditional modeling and relational learning alongside Large Language Models, we can guide this process to not only generate valid scientific hypotheses and creative experimental designs without excessive hallucinations, but also automatically incorporate insights from complex interconnected datasets into the scientific process.

Image generated using GPT4o with the prompt: Isometric view of a stylized automated laboratory cell performing biology experiments

1. Ross D. King, Ashwin Srinivisan, Luc Dehaspe. *WARMR: A Data Mining Tool for Chemical Data* (2001). Journal of Computer-Aided Molecular Design.
2. Daniel Brunnsåker, Filip Kronström, Ievgeniia A. Tiukova, Ross D. King. *Interpreting protein abundance in Saccharomyces cerevisiae through relational learning* (2024). Bioinformatics.
3. Gabriel K. Reder, Erik Y. Bjurström, Daniel Brunnsåker et al.. *AutonoMS: Automated Ion Mobility Metabolomic Fingerprinting* (2024). Journal of the American Society for Mass Spectrometry.

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM