Introduction

Introduction to Automated Science

SLAS 2024

What is Automated Science?

Automated Science uses AI to plan, execute and interpret experiments **without human intervention**.

What is Automated Science?

Automated Science uses AI to plan, execute and interpret experiments **without human intervention**.

Automated Science platforms combine three technologies:

1. laboratory robots that perform physical experiments
2. a machine learning model that predicts results
3. an AI agent that plans future experiments

What is Automated Science?

Automated Science uses AI to plan, execute and interpret experiments **without human intervention**.

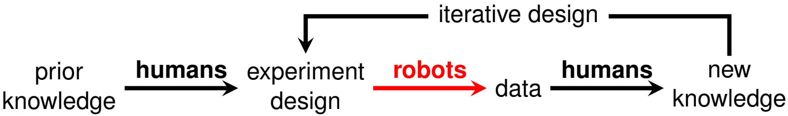
Automated Science platforms combine three technologies:

1. laboratory robots that perform physical experiments
2. a machine learning model that predicts results
3. an AI agent that plans future experiments

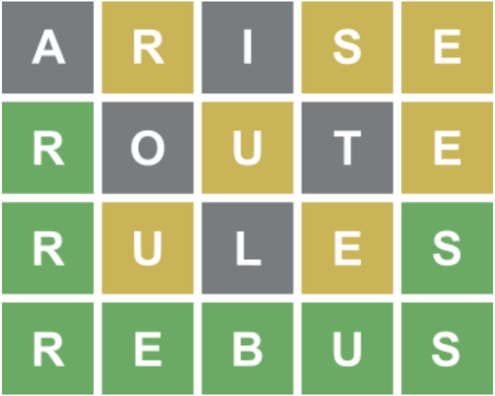
Automated Science can use **laboratory automation** to run the experiments it designs; however, we consider experiments designed by AI and executed by humans to be "automated science".

Laboratory Automation vs. Automated Science

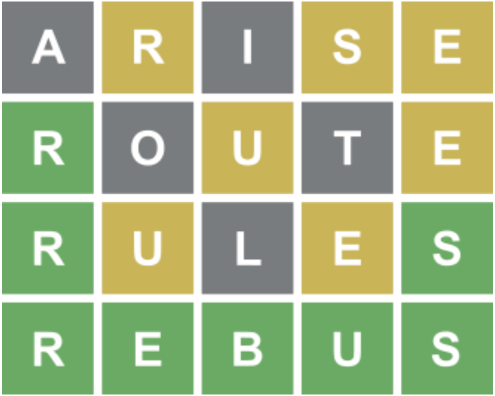
Laboratory Automation vs. Automated Science



Why sequential experiments?



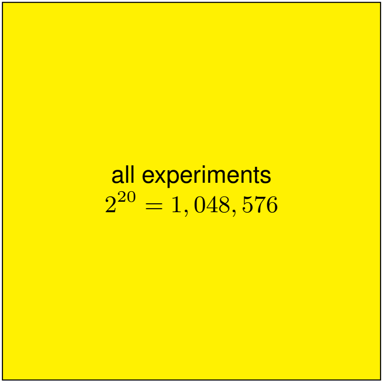
Why sequential experiments?



How many guesses would you need if all the guesses were checked in a single batch?

Why sequential experiments? Searching amino acid combinations

Before any experiments.



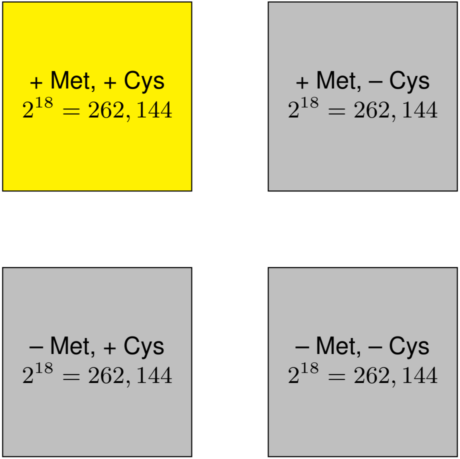
Why sequential experiments? Searching amino acid combinations

After learning Met is essential.

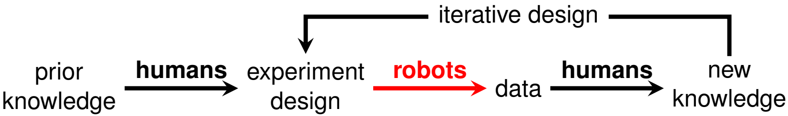
+ Met	– Met
$2^{19} = 524,288$	$2^{19} = 524,288$

Why sequential experiments? Searching amino acid combinations

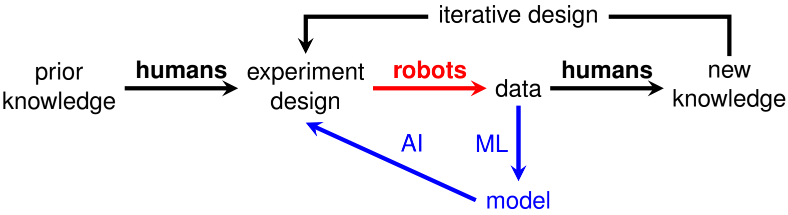
After learning Met and Cys are essential.



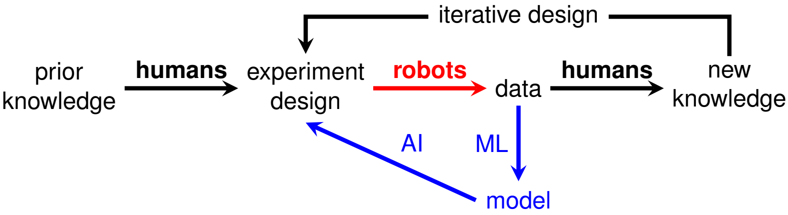
Laboratory Automation vs. Automated Science



Laboratory Automation vs. Automated Science

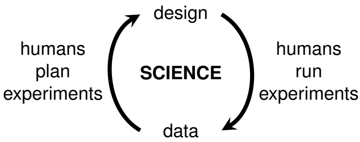


Laboratory Automation vs. Automated Science

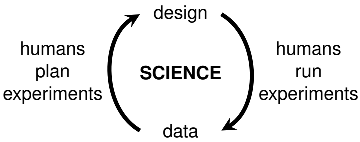


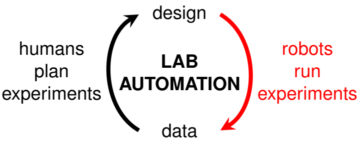
Designing experiments to improve a model
is easier than
designing experiments to create knowledge.

Automated Science does not require robots

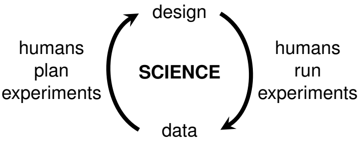


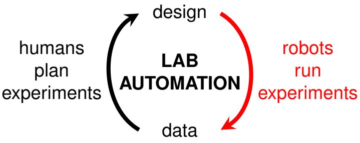
Automated Science does not require robots

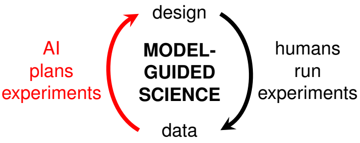




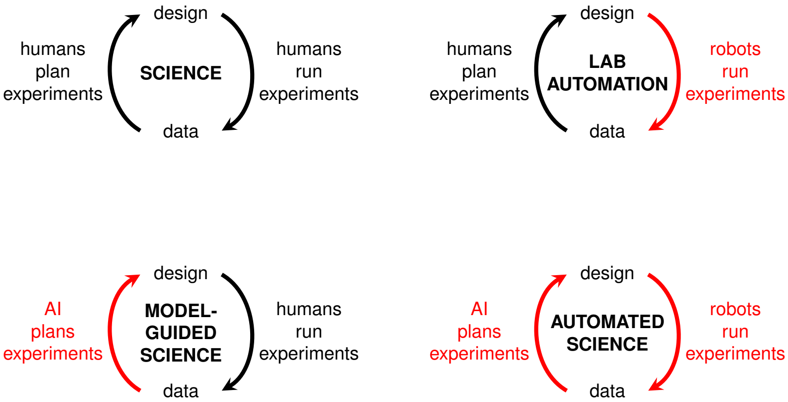
Automated Science does not require robots







Automated Science does not require robots



The language of Automated Science

AI agents design experiments by varying **factors**.

- **Discrete factors** belong to a fixed set of values (e.g. cell lines, reagent supplier, drug A or drug B).
- **Continuous factors** can take any value inside a range (e.g. concentration, time, temperature).

The language of Automated Science

AI agents design experiments by varying **factors**.

- **Discrete factors** belong to a fixed set of values (e.g. cell lines, reagent supplier, drug A or drug B).
- **Continuous factors** can take any value inside a range (e.g. concentration, time, temperature).

A **treatment** defines a unique combination of levels for each factor.

cell line 1 + 0.1 μM drug A + 45 minute incubation

The language of Automated Science

AI agents design experiments by varying **factors**.

- **Discrete factors** belong to a fixed set of values (e.g. cell lines, reagent supplier, drug A or drug B).
- **Continuous factors** can take any value inside a range (e.g. concentration, time, temperature).

A **treatment** defines a unique combination of levels for each factor.

cell line 1 + 0.1 μM drug A + 45 minute incubation

A **run** is an experiment performed to test a treatment. Runs can be

- **replicated** to capture the variation in the experimental system (i.e. biological replicates).
- **duplicated** to capture the variation in the measurement system (i.e. technical replicates).

The language of Automated Science

AI agents design experiments by varying **factors**.

- **Discrete factors** belong to a fixed set of values (e.g. cell lines, reagent supplier, drug A or drug B).
- **Continuous factors** can take any value inside a range (e.g. concentration, time, temperature).

A **treatment** defines a unique combination of levels for each factor.

cell line 1 + 0.1 μM drug A + 45 minute incubation

A **run** is an experiment performed to test a treatment. Runs can be

- **replicated** to capture the variation in the experimental system (i.e. biological replicates).
- **duplicated** to capture the variation in the measurement system (i.e. technical replicates).

The **response** is a quantitative output of the systems that will be learned or optimized.

The goals of Automated Science

Automated Science can use either of two objectives.

- **Optimization** seeks the treatment with the maximum (or minimum) response.
- **Characterization** seeks the most informative treatments for learning the response.

We'll see later that optimization and characterization are related by the *exploration/exploitation tradeoff*.

Optimization ("engineering")

Optimization seeks a treatment that maximizes, minimizes, or moves the response closest to a goal.

Optimization can be *constrained* or *unconstrained*.

Optimization (“engineering”)

Optimization seeks a treatment that maximizes, minimizes, or moves the response closest to a goal.

Optimization can be *constrained* or *unconstrained*.

Examples

- Find enzymes expression levels in a pathway that maximize lipid titers in a fermenter.
- Find reactions conditions that maximize yield of a chemical synthesis.
- Find drug combinations that return phosphorylation to normal levels.
- Find gene deletions that minimize production of a byproduct *while maintaining a threshold level of primary production.*

Characterization (“science”)

Characterization learns the effects of each factor on the response.

Complete characterization learns both the main effects of the factors and their interactions. The complexity of the interactions depends on both the model and the experiments.

Characterization ("science")

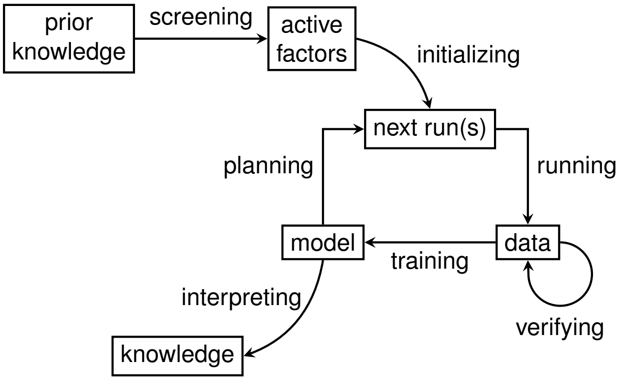
Characterization learns the effects of each factor on the response.

Complete characterization learns both the main effects of the factors and their interactions. The complexity of the interactions depends on both the model and the experiments.

Examples

- Learn how media factors affect expression of a target gene.
- Learn how reagent concentrations affect absorbance of a nanocrystal.
- Learn how drug combinations affect protein phosphorylation.

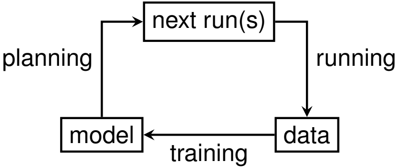
The Automated Science Cycle



Case Studies

Outline

This course focuses on the core of the Automated Science cycle:



We will discuss

- Planning experiments with a trained model.
- Selecting an appropriate Bayesian model.
- Challenges when performing AI-planned experiments.

The course emphasizes your role as an *integrator* between modelers, automation engineers, and scientists.