Introduction

Introduction to Automated Science

SLAS 2023

Schedule

10:30–11:15	Introduction
11:25–12:30	Planning (Part 1) + Q&A
12:30–1:00	Lunch
1:00–1:30	Models
1:30–2:00	Planning (Part 2)
2:00–2:30	Conclusion + Q&A

Logistics

- “Parking lot” questions
- Restrooms
- Lunch
- Slides available at `http://jensenlab.net/automatedscience`

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**.

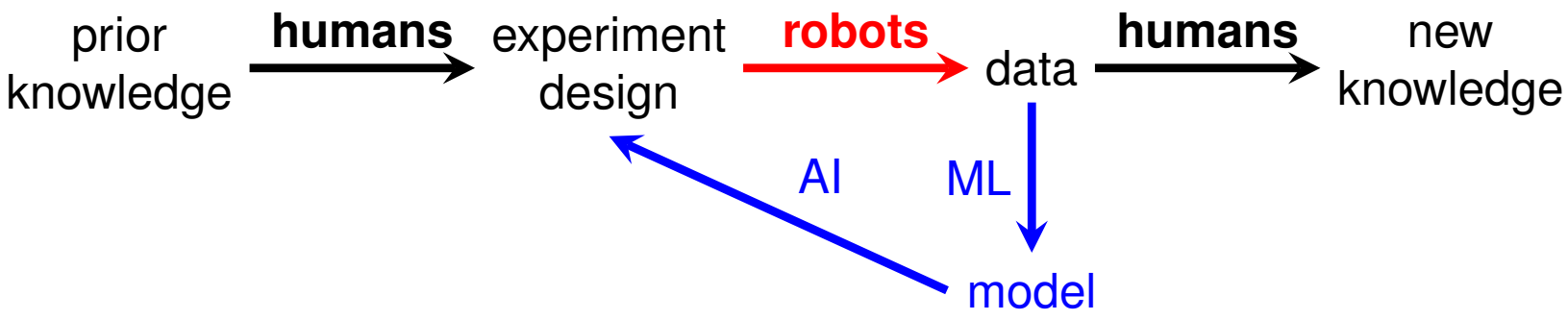
Automated Science platforms combine three technologies:

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

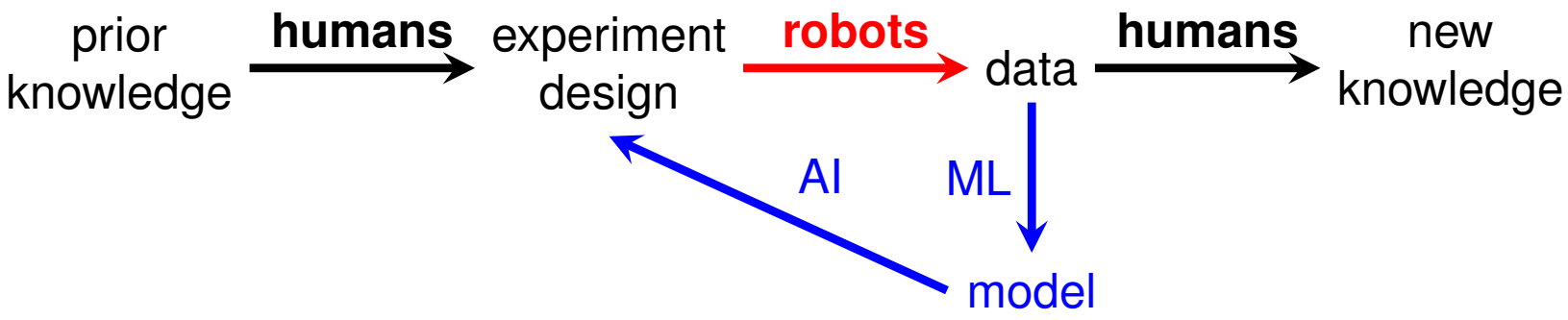
Automated Science can use **laboratory automation** to run the experiments it designs; however, we consider experiments designed by AI and executed by humans to be “automated science”.

Laboratory Automation vs. Automated Science

Laboratory Automation vs. Automated Science



Laboratory Automation vs. Automated Science



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

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.

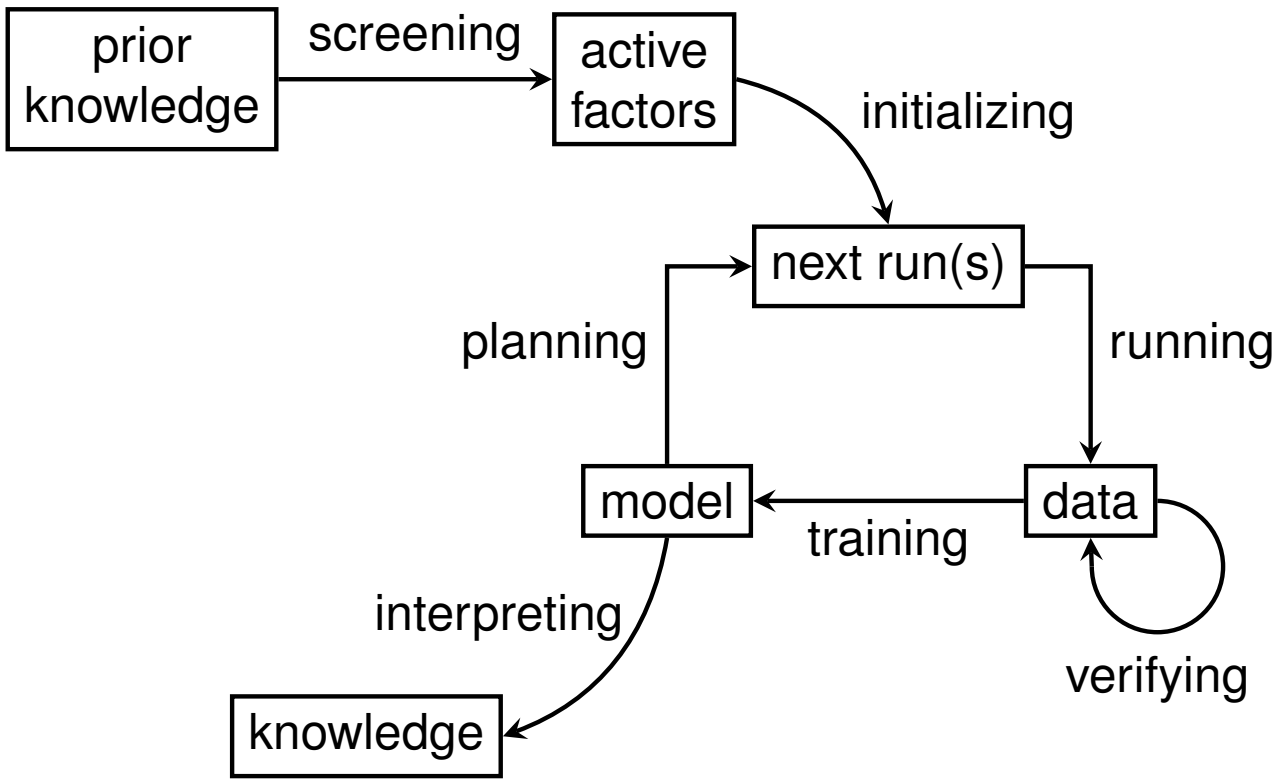
The goal of Automated Science

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.

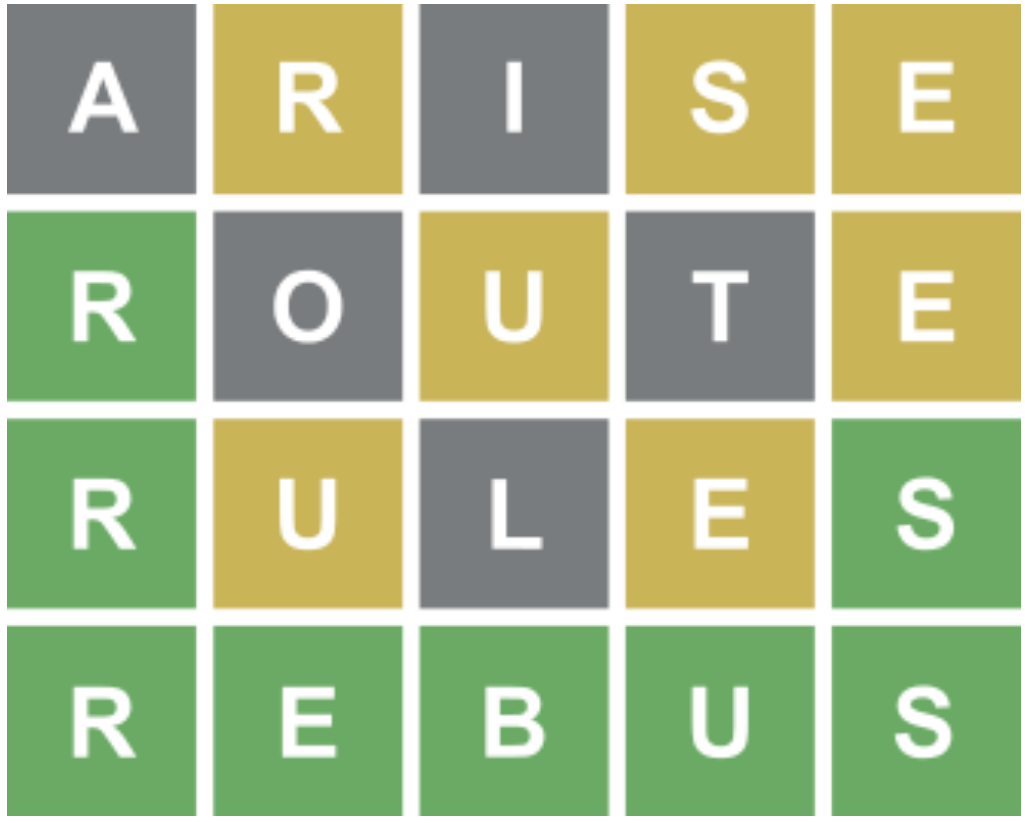
We'll see later that optimization and characterization are related by the *exploration/exploitation tradeoff*.

The Automated Science Cycle

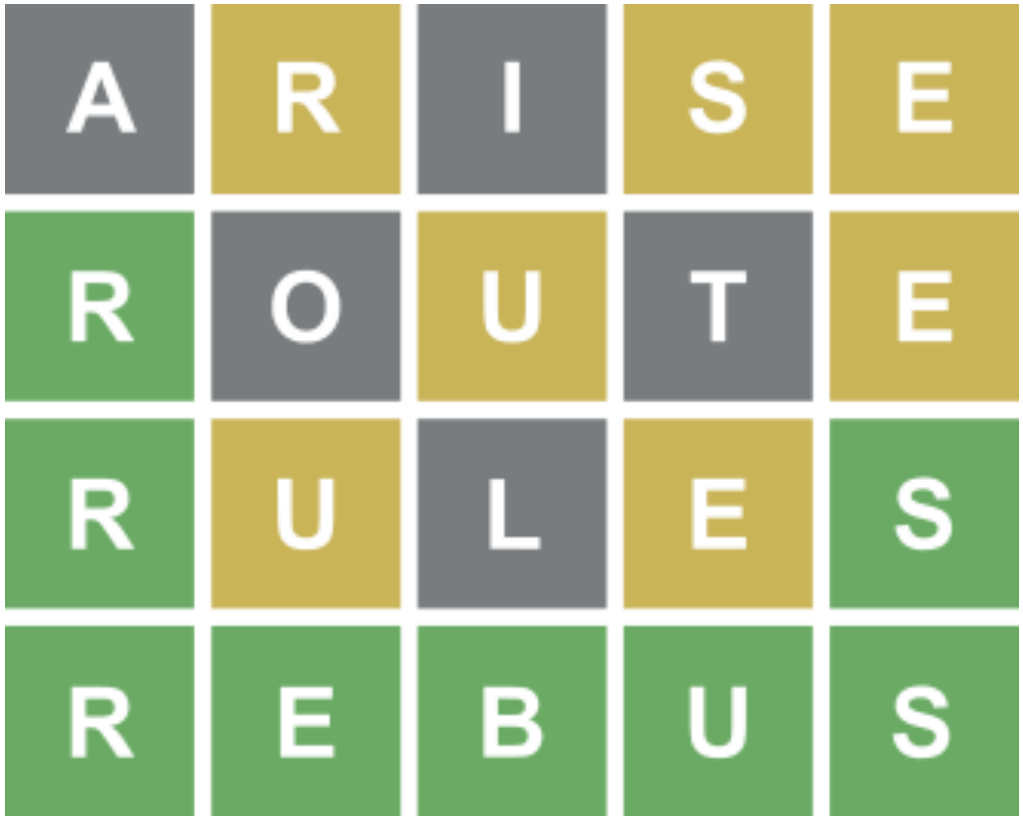


Case Studies

Why sequential experiments?



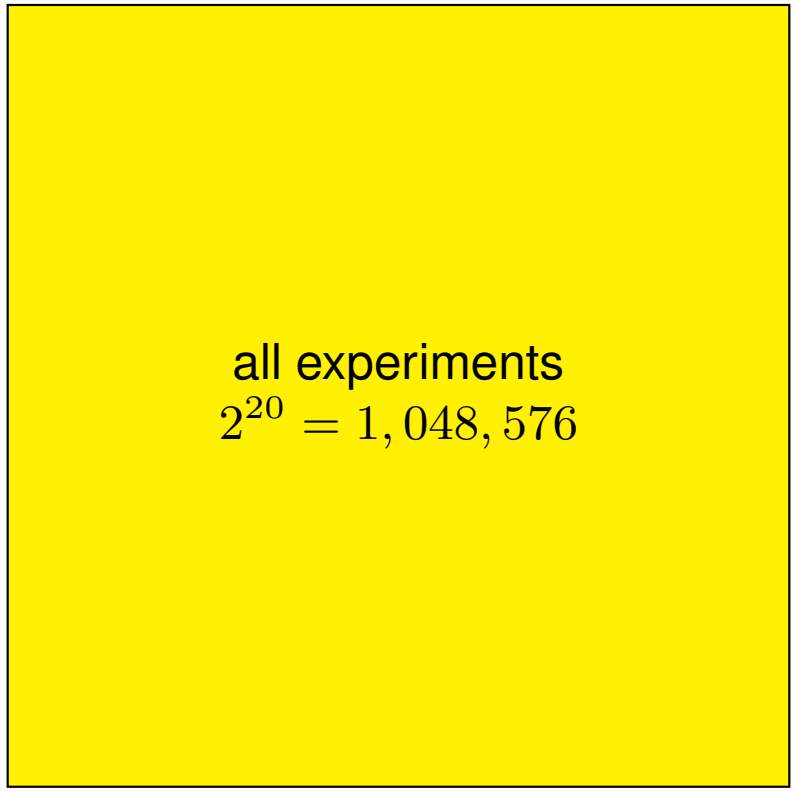
Why sequential experiments?



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

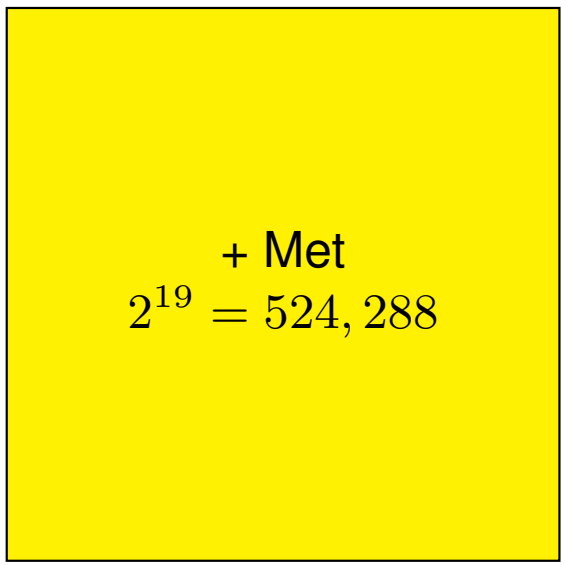
Why sequential experiments? Searching amino acid combinations

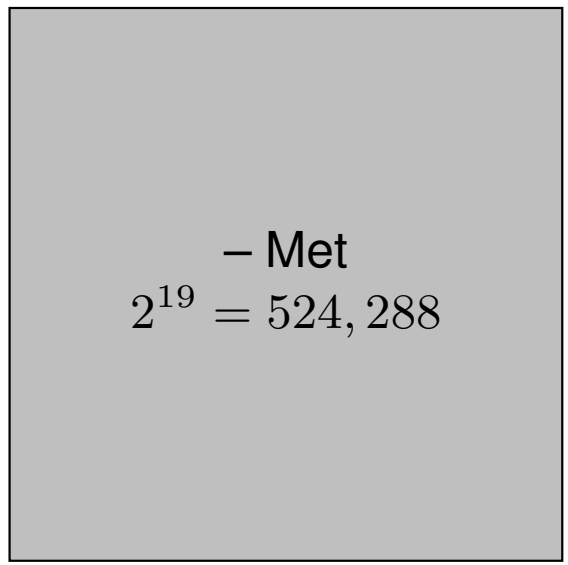
Before any experiments.



Why sequential experiments? Searching amino acid combinations

After learning Met is essential.





Why sequential experiments? Searching amino acid combinations

After learning Met and Cys are essential.

+ Met, + Cys
$2^{18} = 262,144$

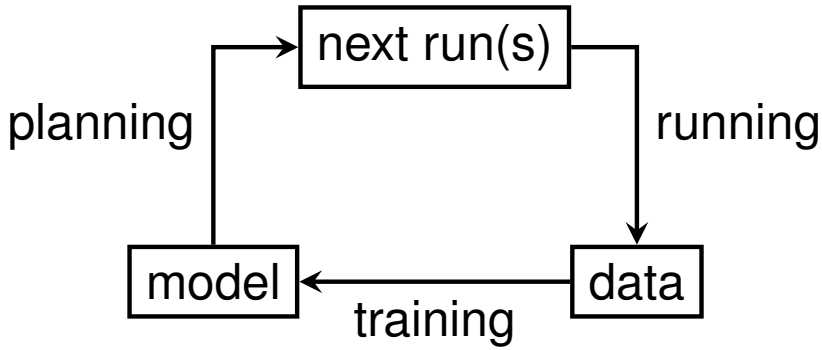
+ Met, – Cys
$2^{18} = 262,144$

– Met, + Cys
$2^{18} = 262,144$

– Met, – Cys
$2^{18} = 262,144$

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.