SciLifeLab

Introduction to Data-driven Life Sciences

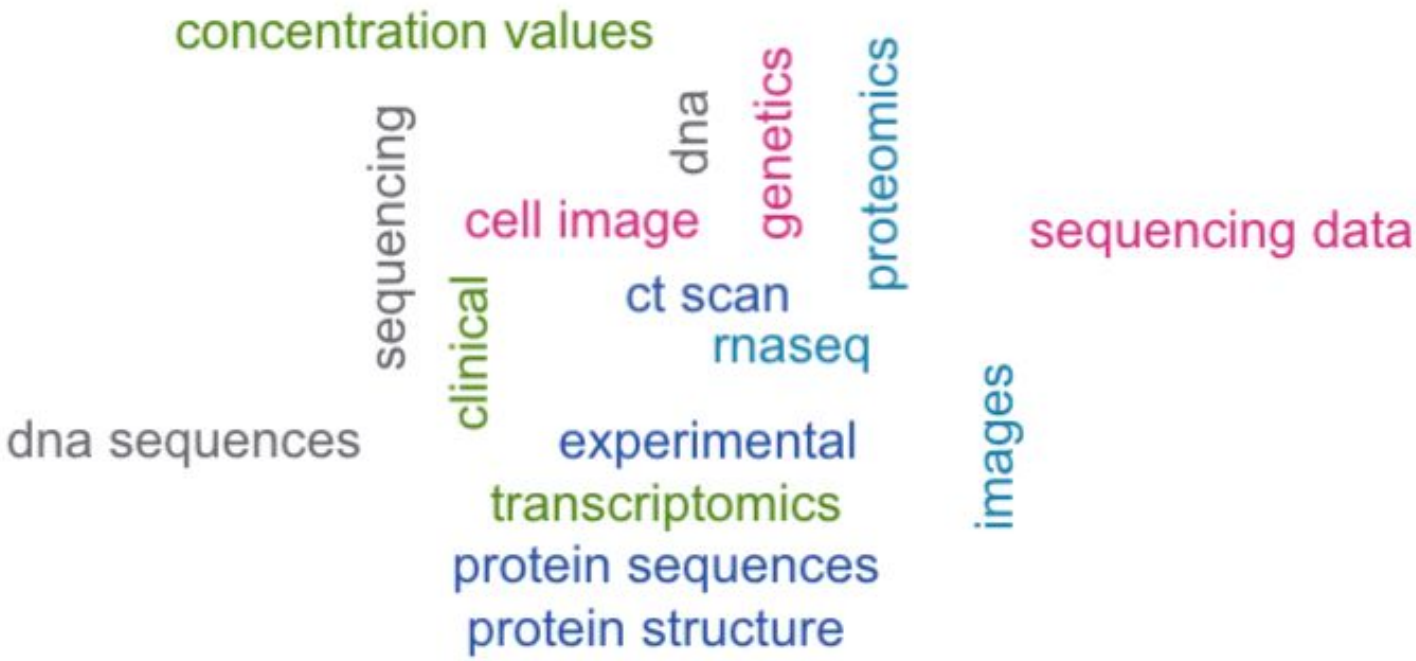
Wei Ouyang - August 29th 2023

What are the data types in life science?

Join at menti.com use code 8676 5947

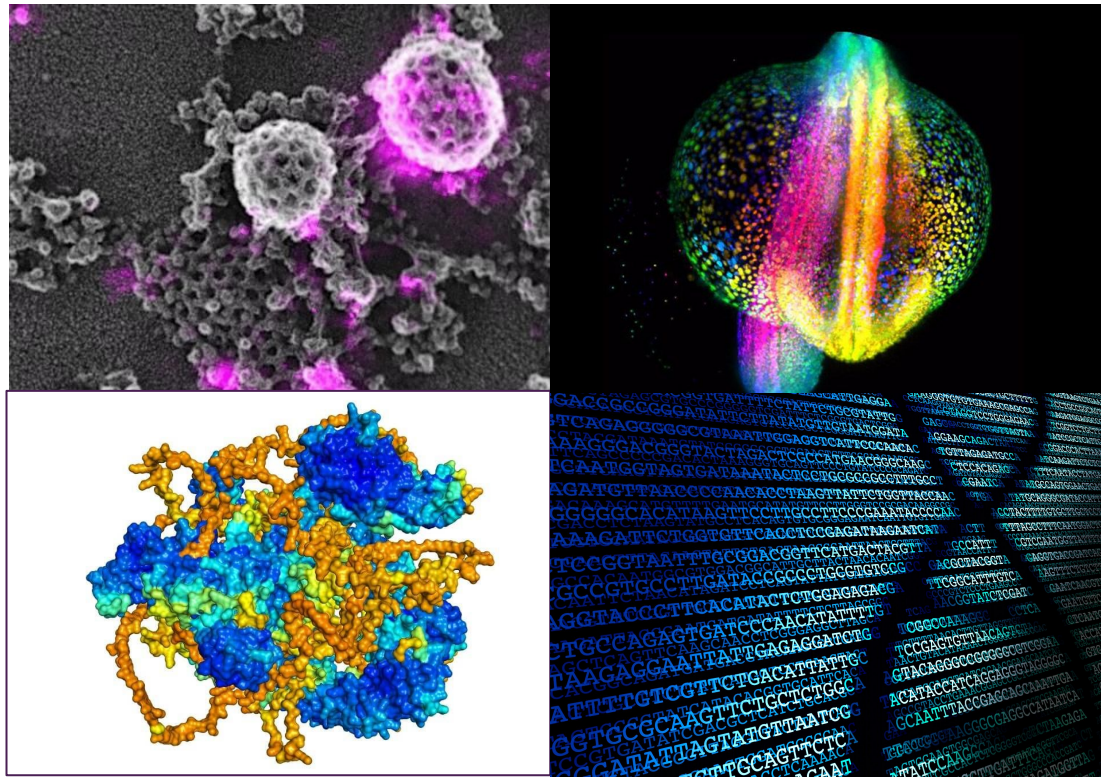
The data types in life science?

16 Responses

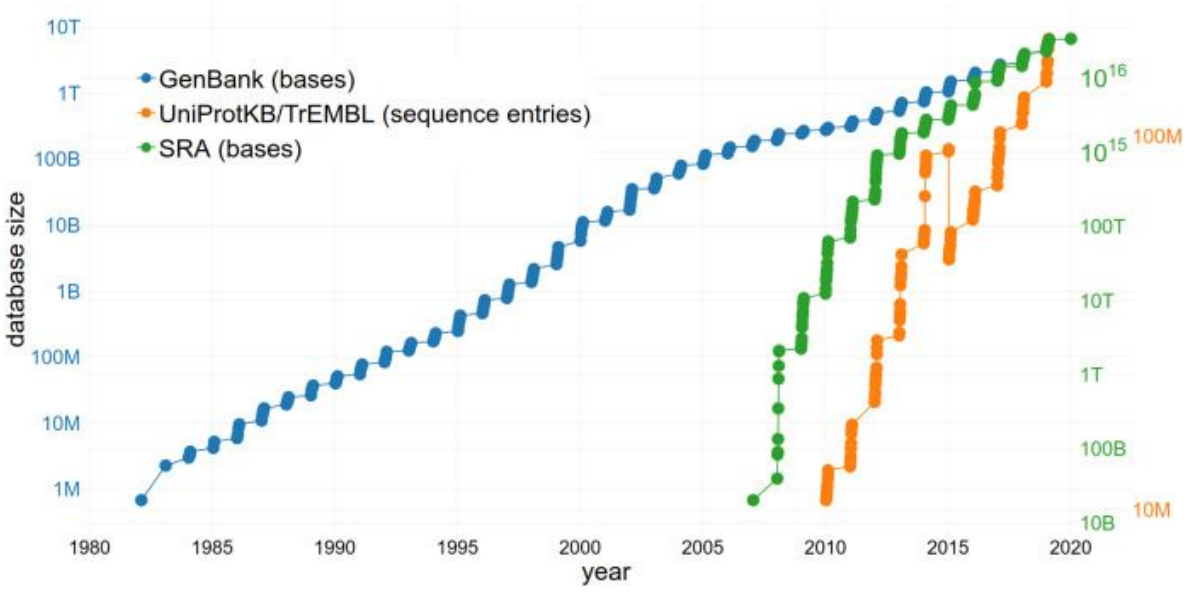


7

Paradigm Shift: Big Data in Life Science

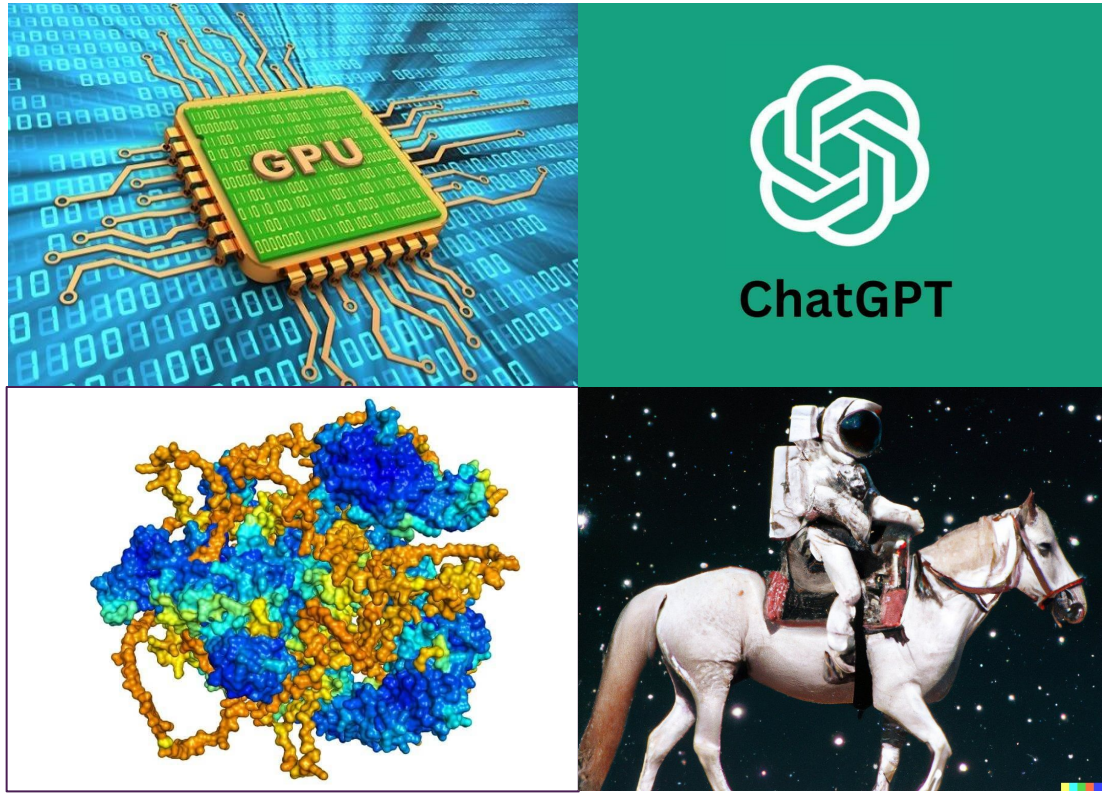


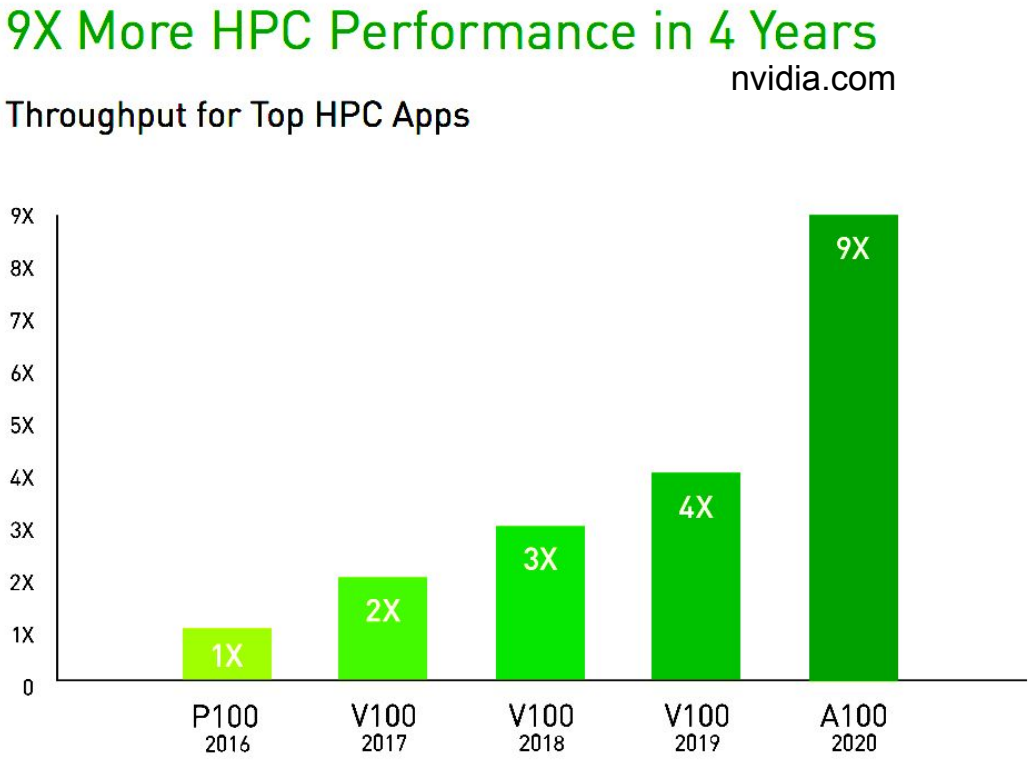
The rapidly increasing databases



Sielemann et al, 2020

Paradigm Shift: Compute Power & AI





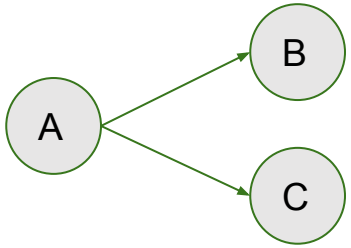
Paradigm shift: Data-driven Life Science

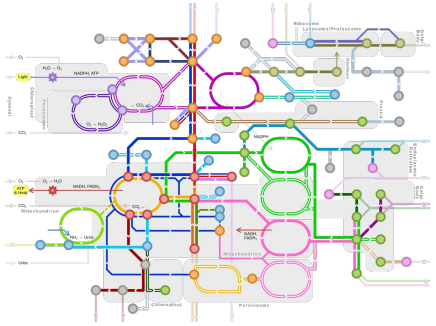
Hypothesis driven research → Data driven research → AI-powered data-driven research

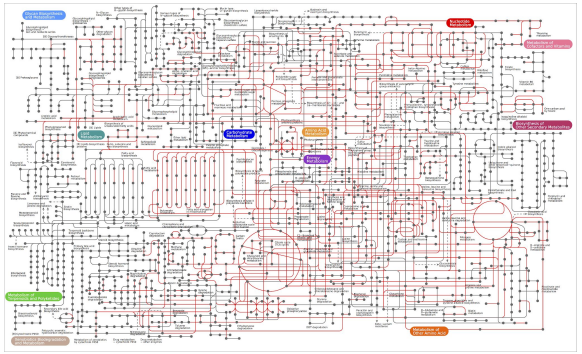
1. Hypothesis
2. Experiment
3. Accept or reject

1. Acquire data
2. Find pattern
3. Evaluate pattern

1. Acquire big data
2. Fit AI model
3. Evaluate AI model



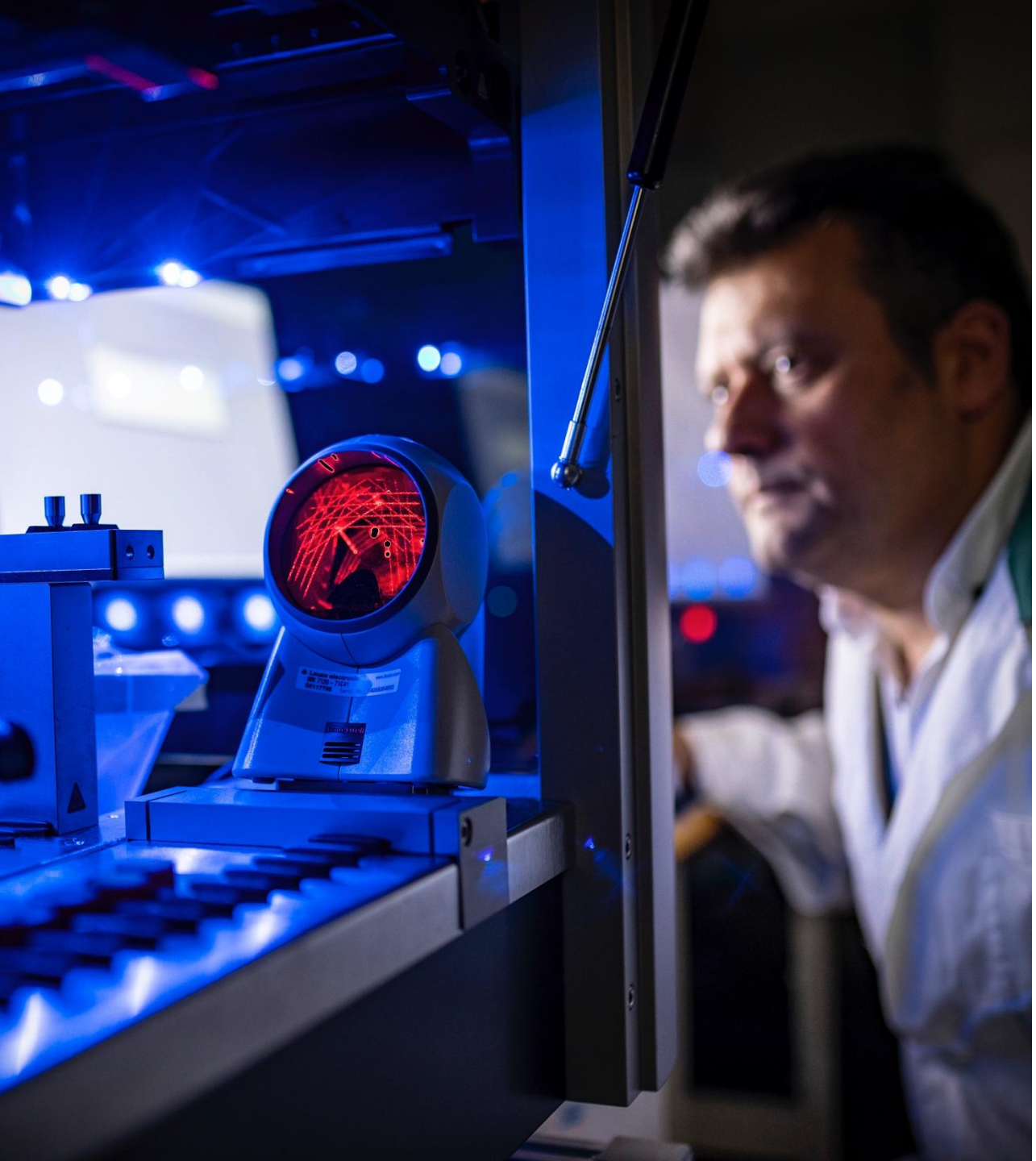




SciLifeLab

About SciLifeLab

Our vision, mission and strategic objectives.



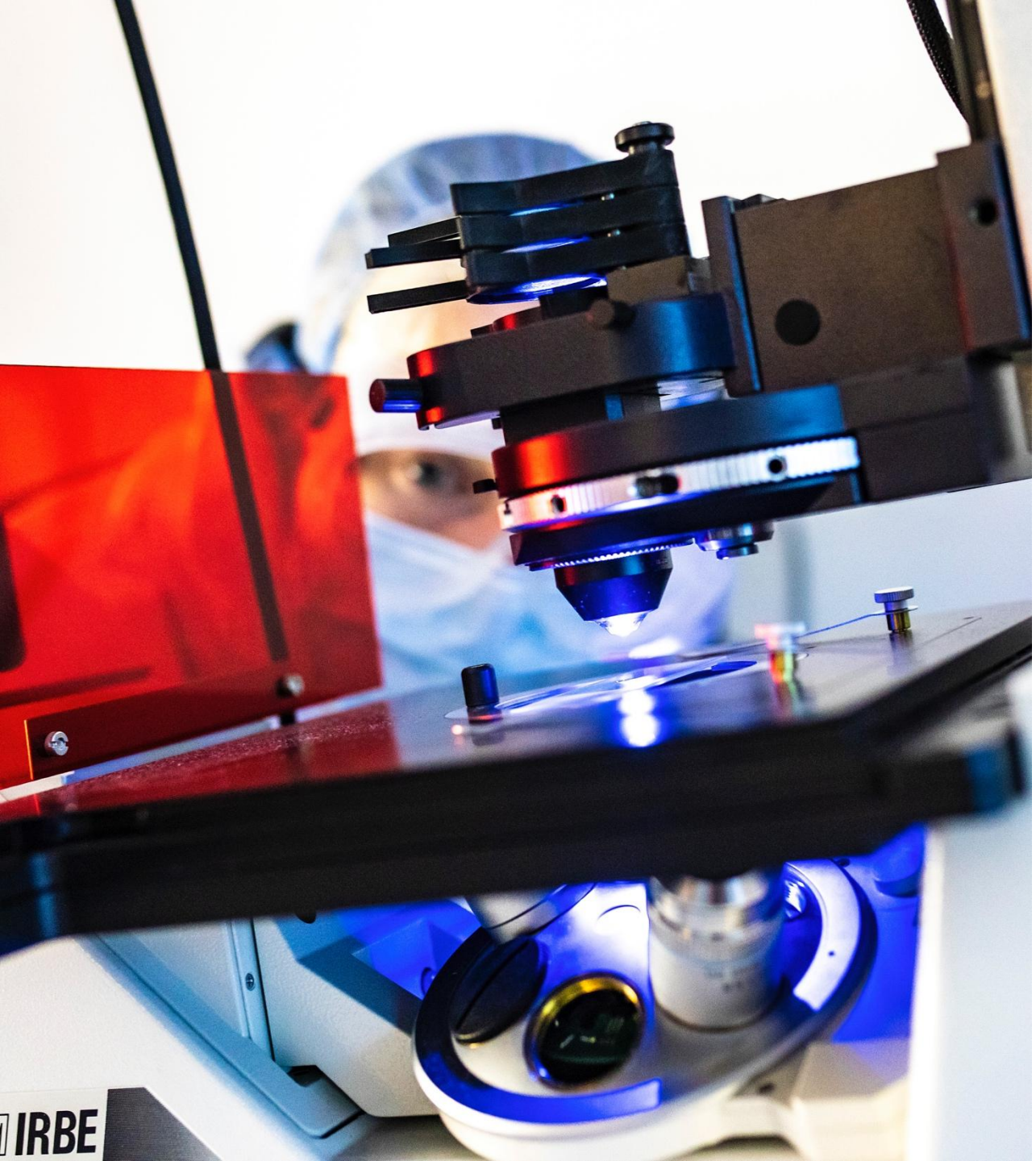
What is SciLifeLab?

National hub **enabling** life science research that would otherwise not be possible

Government appointed mission as a **national research infrastructure.**

Started in 2010 by Karolinska Institutet, KTH Royal Institute of Technology, Stockholm University and Uppsala University.

Today, activities at **all major Swedish universities.**



Vision and mission

Vision: for Sweden to be a world-leading nation in life science

Mission: Enable life science research that would otherwise not be possible

Three dimensions of SciLifeLab

Research environment

Approx. 190 affiliated research groups

- Environment and climate change
- Farming and forestry
- Evolution and biodiversity
- Gene editing
- Biofuels and biomaterials
- Microbiology and microbiome
- Drugs and biomedicine
- Healthcare and aging

Infrastructure

Service to ~ 1400 Swedish researchers annually (2020)

- Bioinformatics
- Cellular and molecular imaging
- Clinical diagnostics
- Single cell biology
- Genomics
- Chemical biology and gene editing
- Drug development
- Proteomics and metabolomics

Data-driven life science

3.1 billion SEK, 12-year-program

Putting Sweden at the forefront of data-driven life science research and fostering the next generation of life scientists

- Four strategic research areas
- Recruiting talent from across the globe
- Academic and industry PhD and postdoc programs
- Sparking collaborations, innovation and interdisciplinary team science
- Building a strong computational and data science base for open, real-time data

SciLifeLab

SciLifeLab and Wallenberg National Program for Data-Driven Life Science

Changing the way life science is carried out

SciLifeLab and Wallenberg National Program for Data-Driven Life Science

Knut och Alice Wallenbergs Stiftelse

SciLifeLab

WALLENBERG CENTRES FOR MOLECULAR MEDICINE

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

WACQT | Wallenberg Centre for Quantum Technology

CHALMERS

GÖTEBORGS UNIVERSITET

Karolinska Institutet

KTH VETENSKAP OCH KONST

LINKÖPINGS UNIVERSITET

LUNDS UNIVERSITET

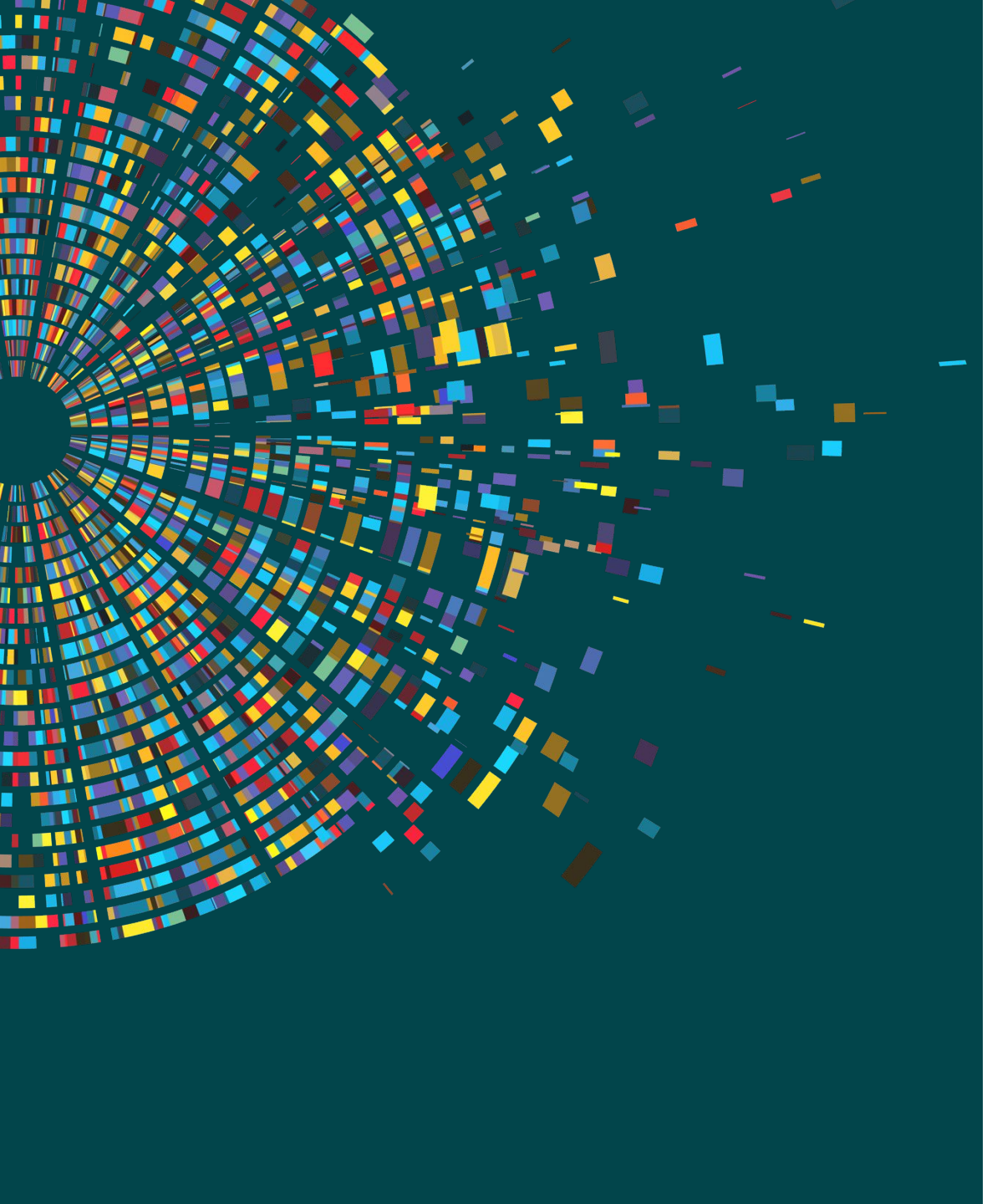
SLU

Stockholms universitet

UMEÅ UNIVERSITET

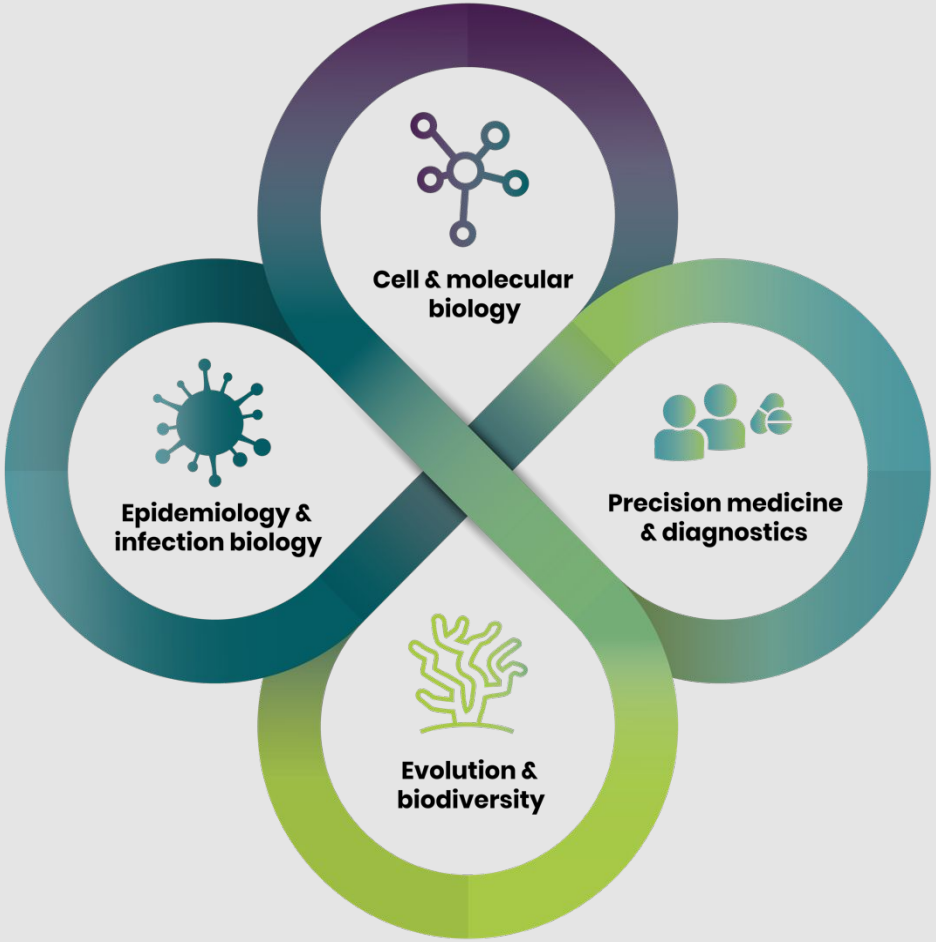
Naturhistoriska riksmuseet

UPPSALA UNIVERSITET



Data has a central place in life science

- Practice of life science is more and more data-dependent
- The amount of data grows exponentially
- Data becomes more complex, continuous, and needs to be openly available accessible and reusable in real-time to all



The four strategic research areas of the DDLS program

Promoting a paradigm shift in life sciences

Launching a research program to:

- Foster the next generation of life scientists
- Enable every biologist to better analyze and interpret data patterns and integrate their own data seamlessly with the global life science data streams
- Create a strong computational and data science base
- Enable deep understanding of life through data

Advancing competence and innovation in **four research areas**

Using new **technologies** and **strategies** to utilize open and real-time data

Engaging in education, training, recruiting new talent, sparking collaborations, and in innovation activities

The data life cycle

Data Centre

Created to **maximize impact** of SciLifeLab generated data

Assists in communication between platforms, users, and research community

Acts as a point of contact for data management questions relating to SciLifeLab generated data

Assists platforms with data tracking and statistics

Facilitates providing SciLifeLab generated data with SciLifeLab funded bioinformatics and data management support

Assists with planning the handling of SciLifeLab generated data throughout projects

FAIR principles

Findable – assigning a globally unique and eternally persistent identifier (like a DOI or Handle), describing the data with rich metadata, and making sure it is findable through disciplinary discovery portals.

Accessible – data and metadata should be retrievable in a variety of formats that are sensible to humans and machines using persistent identifiers.

Interoperable – the description of metadata elements should follow community guidelines that use an open, well defined vocabulary.

Reusable – the data should maintain its initial richness. The description of essential, recommended, and optional metadata elements should be machine processable and verifiable, use should be easy and data should be citable to sustain data sharing and recognize the value of data.

Group Discussion

Discuss in groups of 3-4 on the following questions (5 minutes):

- What are the obstacles for implementing FAIR in Life Science?
- How can the FAIR principles be implemented more effectively in life sciences?

Reminder: FAIR = Findable, Accessible, Interoperable, Reusable

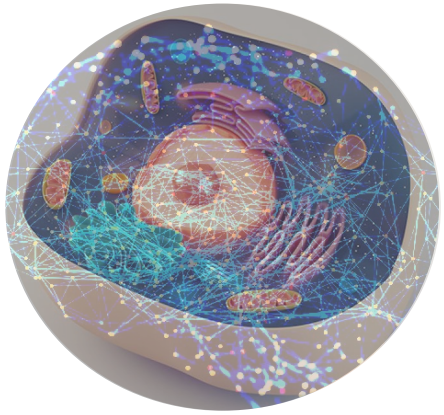
SciLifeLab

Getting prepared for Data-Driven Life Science Efforts in the AICell Lab

15 minutes

AICell Lab (https://aicell.io)

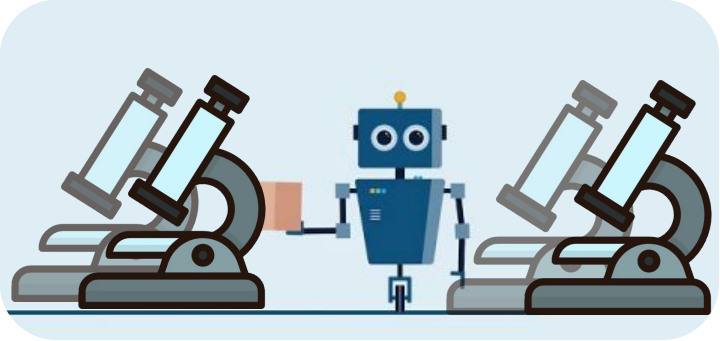
Human Cell Simulator



Data Generation

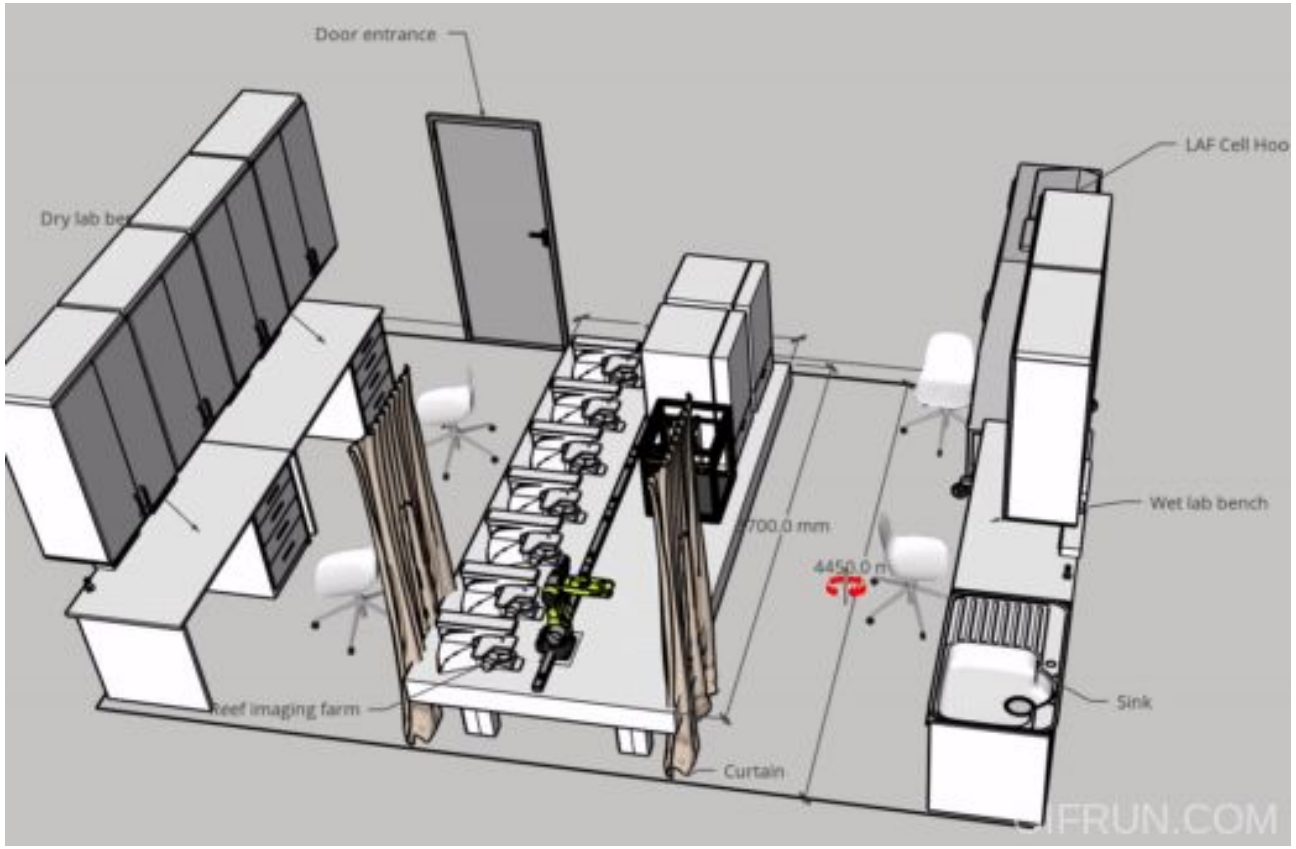
AI Models Execution

Microscopy Imaging Farm

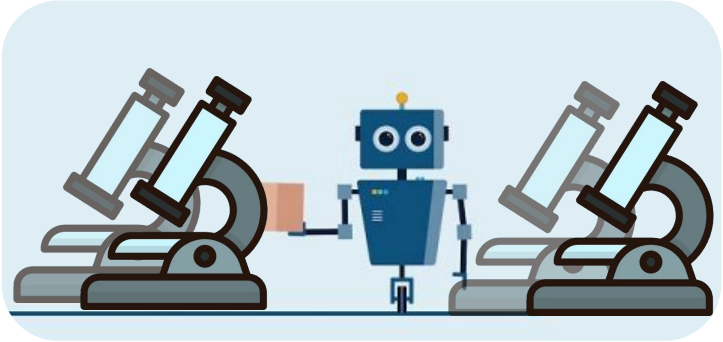


AI Cloud Infrastructure

REEF: Smart Microscopy Imaging Farm

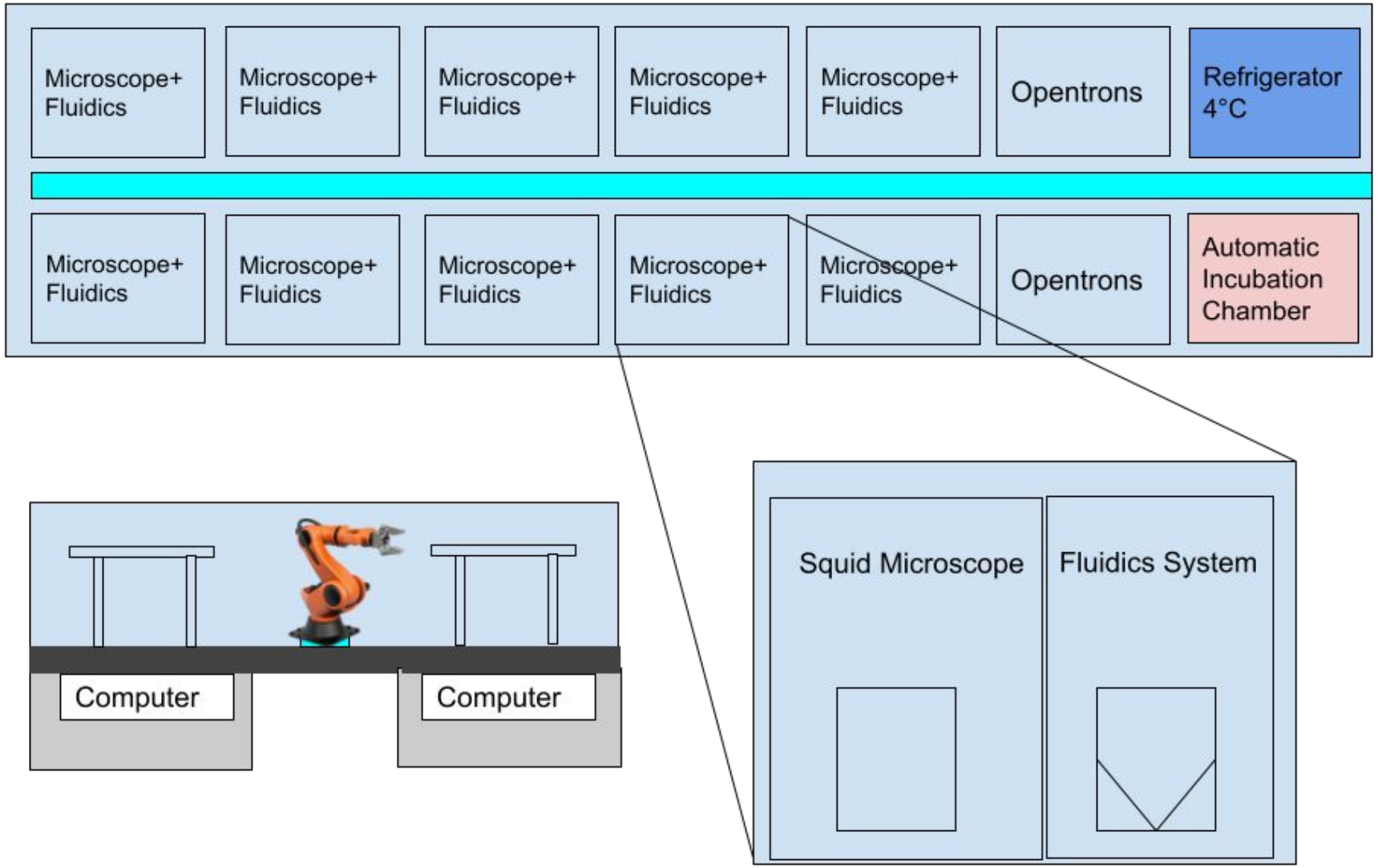


REEF Imaging Farm @ SciLifeLab

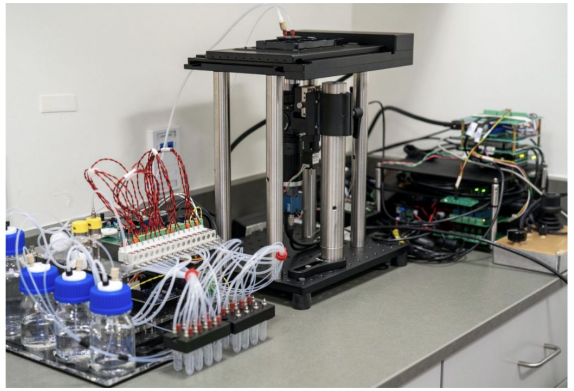


For Massive Dataset Generation

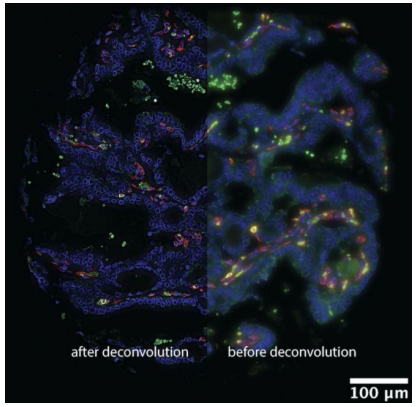
REEF: Smart Microscopy Imaging Farm



Squid Microscope Prototype



CODEX multiplex imaging

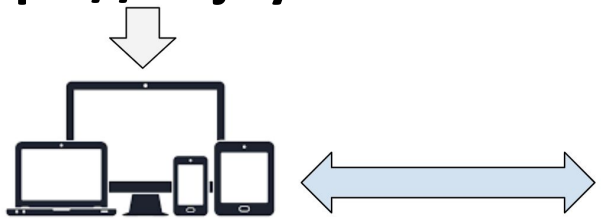


In collaboration with Manu Prakash group at Stanford and Heidstar CO., LTD

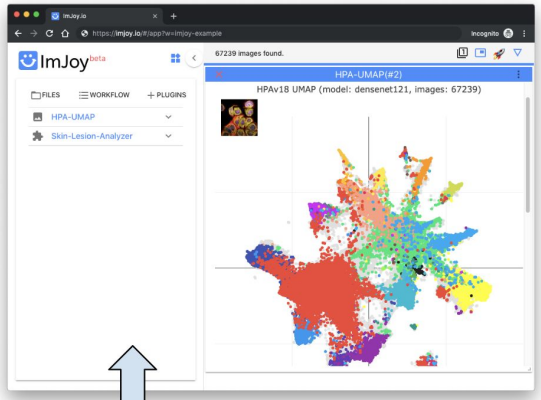
ImJoy AI computing platform

Supercharging interactivity and scalability in AI-powered life science

https://imjoy.io



ImJoy browser app



ImJoy plugins

HTML | JavaScript | Python

Plugin Engine

Cloud servers

Local cluster

Local workstation

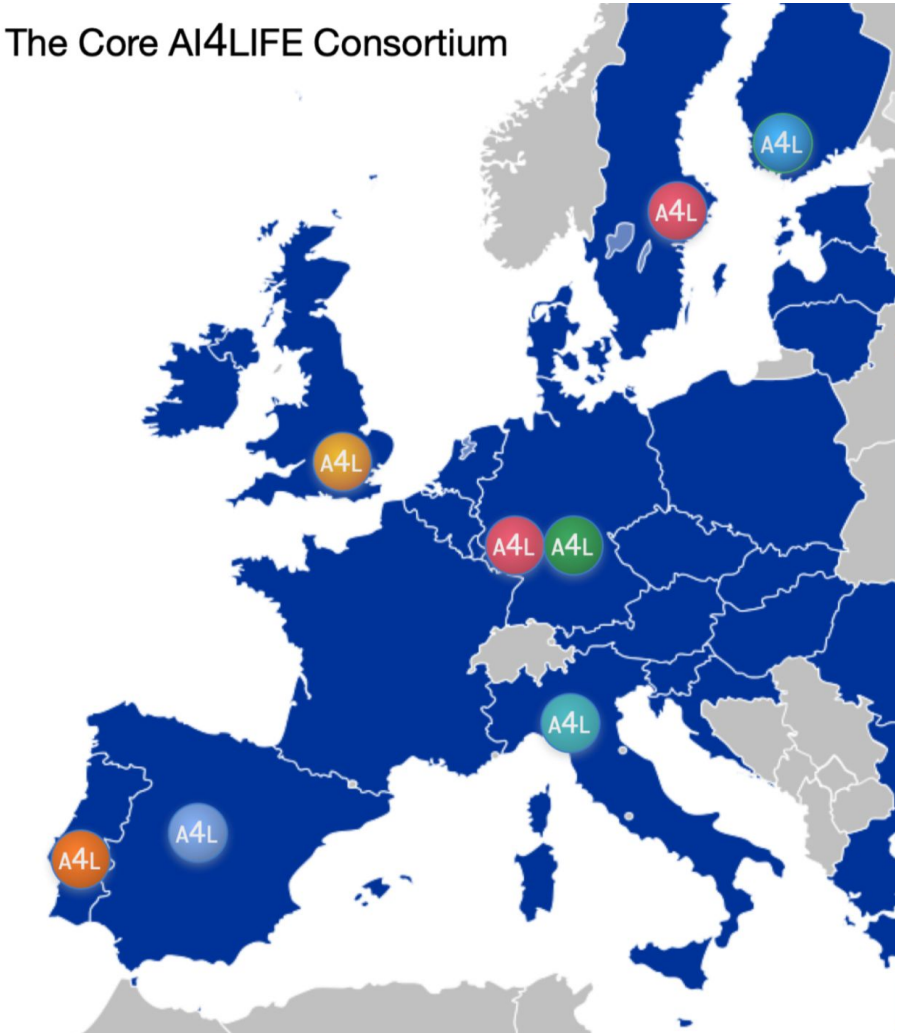
ImageJ.JS: ~1000 users / day

Ouyang* et. al, Nat Methods, 2019 (*co-corresponding)

AI4Life Consortium

- **Making AI models more FAIR**
- **Supported by EU Horizon grant**

- Since 2019, with EMBL, HT, Euro Bioimaging…



https://ai4life.eurobioimaging.eu/

BioImage.IO

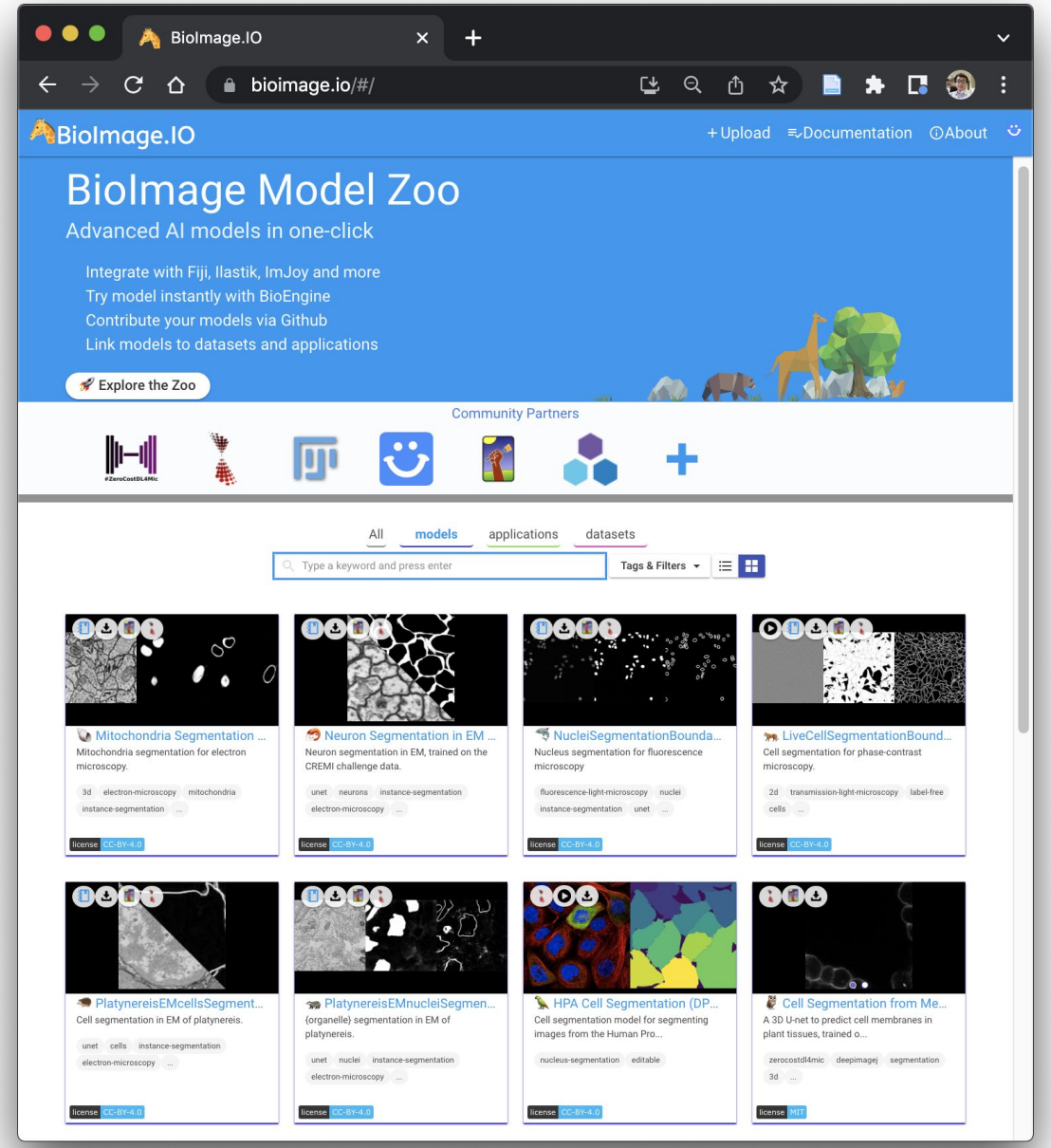
https://bioimage.io

- Model description file standard
- Repository for sharing models
- Continuous integration for model testing
- Cloud-based model serving and test run

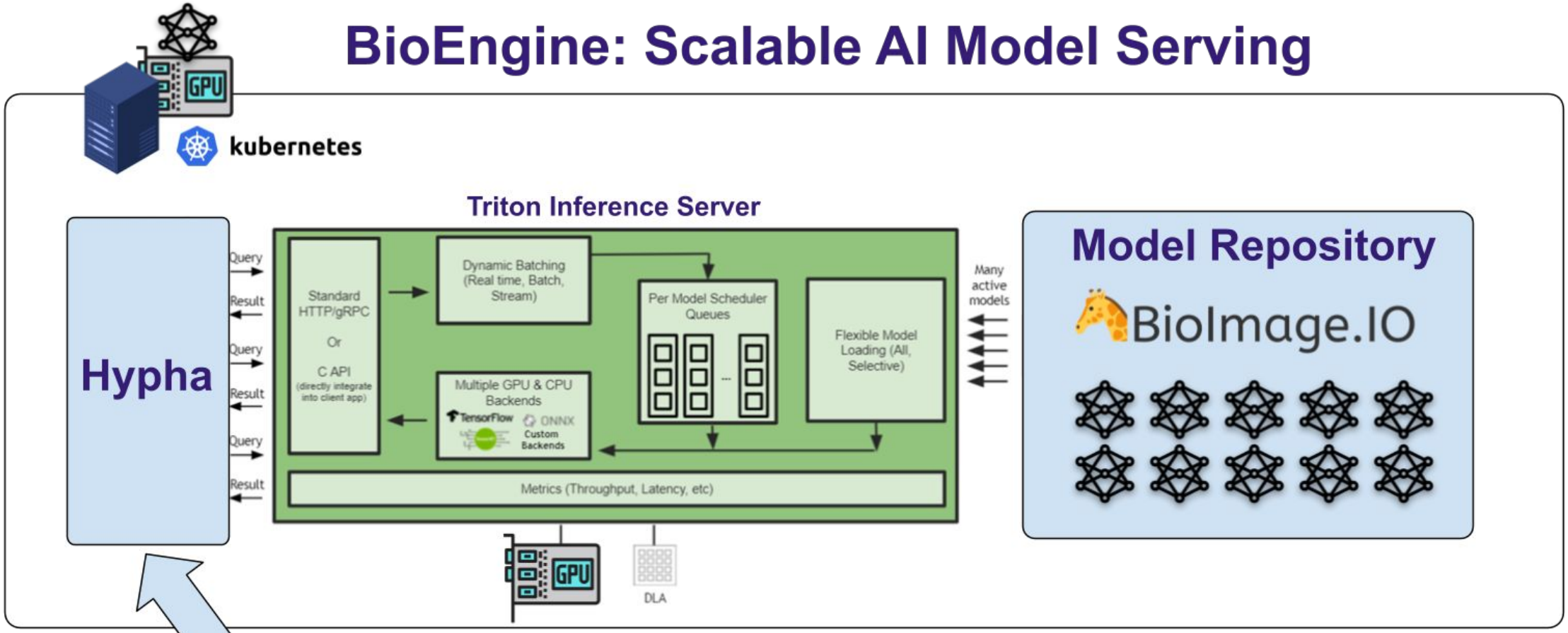
Join us as a community partner!

NEW

Preprint for the BioImage Model Zoo
https://doi.org/10.1101/2022.06.07.495102



BioEngine: Scalable AI Model Serving



Users with thin client computers

Desktop Clients

deepImageJ QuPath ...

Web Clients

ImJoy jupyterlite

No Installation Required!

SciLifeLab

About the course

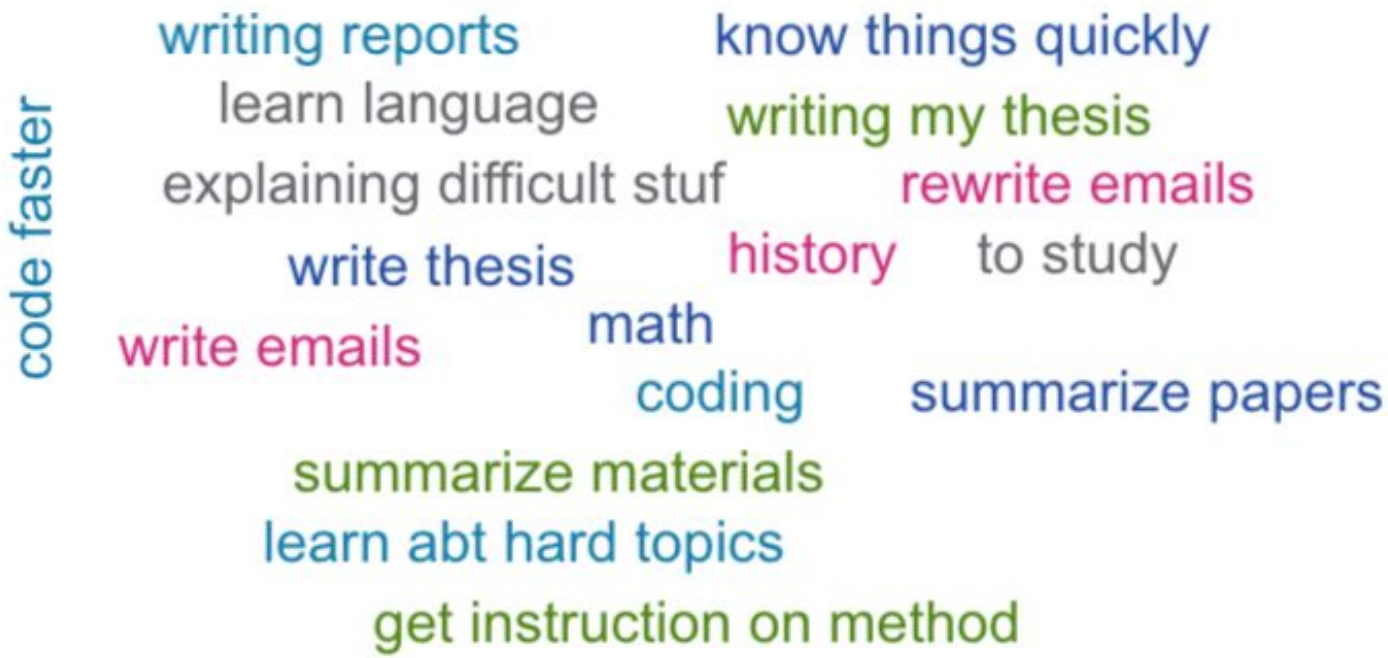
More information at https://ddls.aicell.io

What do you do with ChatGPT?

Join at menti.com use code 8676 5947

What do you do with ChatGPT?

17 Responses



10

Tips for using ChatGPT

- Set a role
- Be specific
- Provide examples
- Use formatting
- Avoid assumptions
- Check limitations
- Try different approaches
- Watch for repetition
- Limit prompt length
- Correct gently
- Give feedback

Assignments

See here: https://ddls.aicell.io/course/ddls-2023/module-1/#lecture-tuesday