FOUNDATIONS OF HEALTH DATA SCIENCE

An Introduction for SUND Researchers & Educators

UCPH HeaDS

PROGRAM

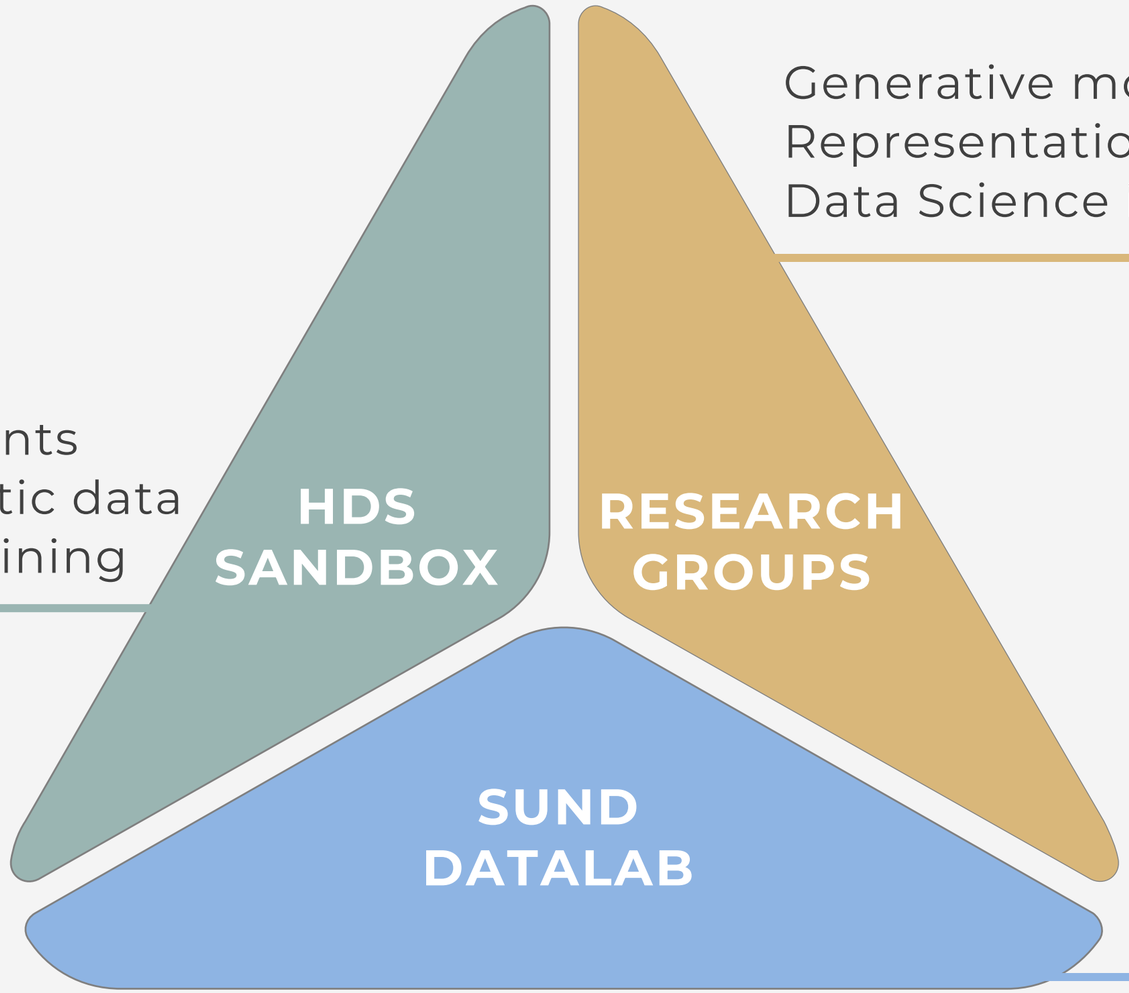
08:15	MORNING COFFEE
08:30	INTRODUCTION TO HDS
09:15	DATA COLLECTION
10:00	COFFEE BREAK
10:15	EXPLORATORY DATA ANALYSIS
11:30	DATA ANALYSIS PART 1
12:00	LUNCH
13:00	DATA ANALYSIS PART 2
13:30	MODEL EVALUATION PART 1
14:15	COFFEE BREAK
14:30	MODEL EVALUATION PART 2
15:00	WRAP-UP & DISCUSSION

COURSE MATERIALS:

https://github.com/Center-for-Health-Data-Science/Foundations-of-Health-Data-Science

CENTER FOR HEALTH DATA SCIENCE (HeaDS)

UCPH HeaDS



HPC environments
Public & synthetic data
Courses and training

Generative modelling
Representation learning
Data Science in Epidemiology


Henrike


Thilde

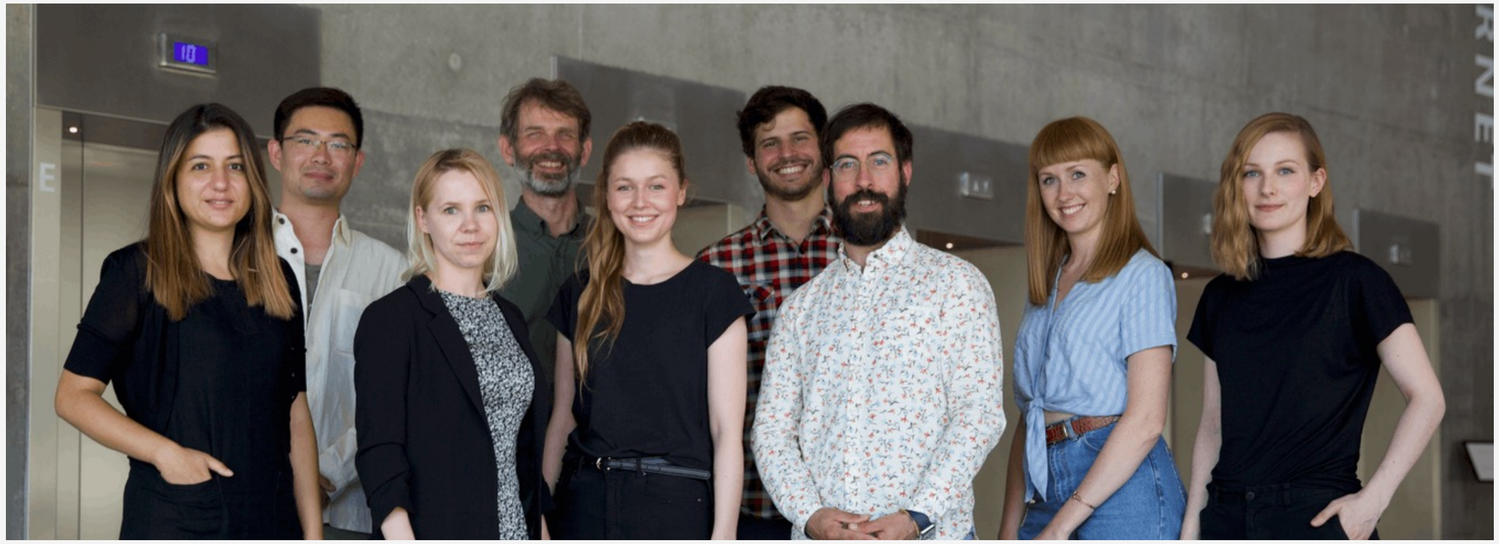
Courses, Consulting,
Commissions, Supervision

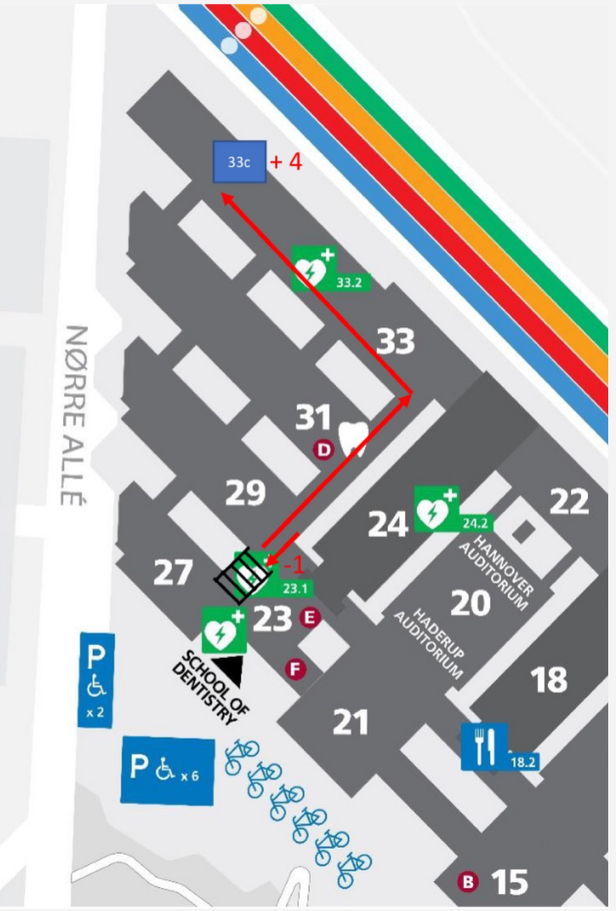
CONTACT US

Email: heads-admin@sund.ku.dk | datalab@sund.ku.dk

Website: https://heads.ku.dk/

Location: Building 33, 4. floor, Section C, Panum

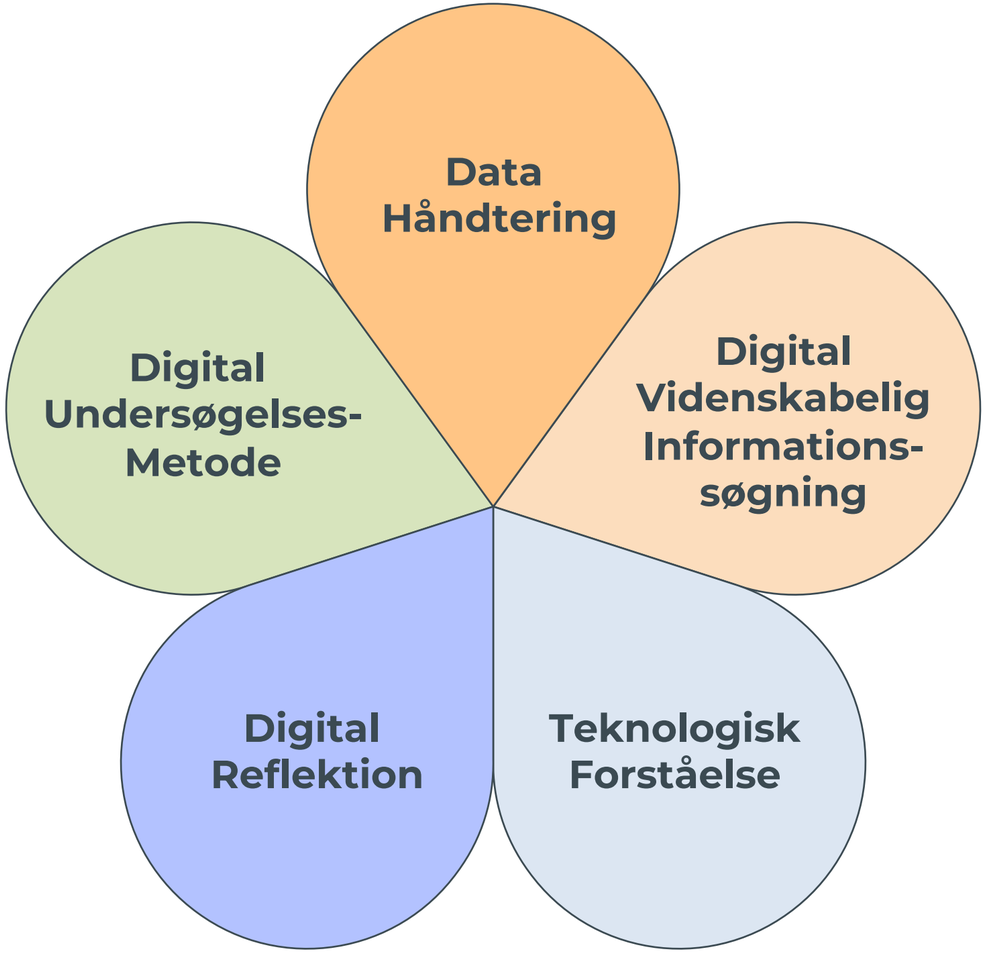




DIGITAL CORE CURRICULUM INITIATIVE

The Digital Core Curriculum (DCC) initiative

- KU-wide initiative, started at SUND
- Update KU educations to include digital literacy skills and data science competences
- Each study board does its own implementation.
- DCC working group supports and advises
- Goal: A digital core curriculum corresponding to 5 – 7,5 ECTs per study line

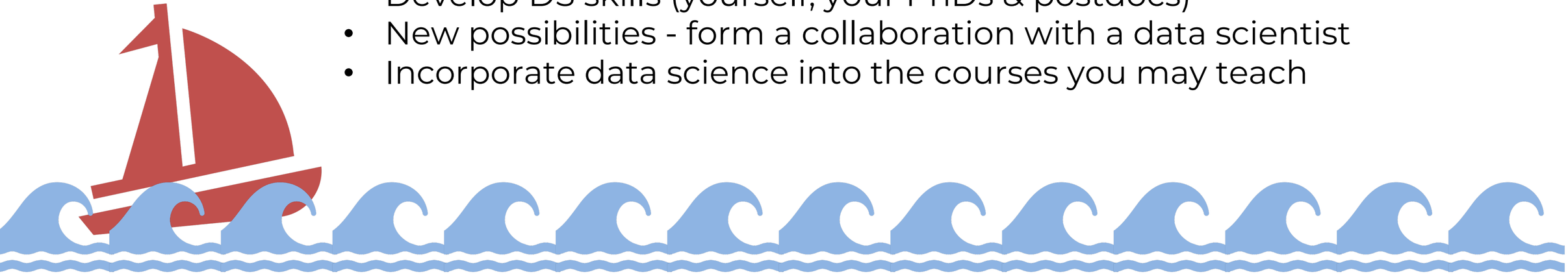


UCPH HeaDS

THE PURPOSE OF THIS WORKSHOP

In this workshop, we are going to begin our journey into Data Science

- **What is (Health) Data Science (DS)?** Roles, definitions, data types
- **A Data's journey**: From data collection to scientific results?
- Data science is not black box or mystical. Now you know the 'fancy words'.
- Inspiration:
  - Develop DS skills (yourself, your PhDs & postdocs)
  - New possibilities - form a collaboration with a data scientist
  - Incorporate data science into the courses you may teach



UCPH HeaDS

In this workshop, we are going to begin our journey into Data Science.

On our way we shall touch upon data collection, exploration, analysis and evaluation.

UCPH HeaDS

Mentimeter

During the course we will use **Mentimeter** for feedback and discussion.

Let’s try it out!

UCPH HeaDS

Mentimeter

Q2 & Q3:

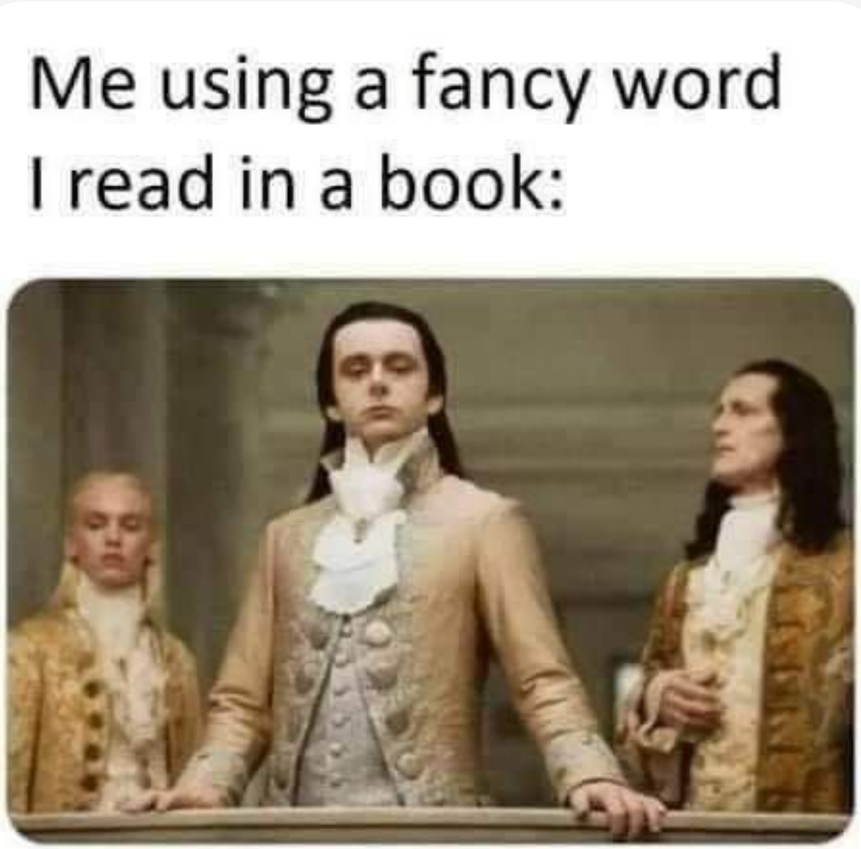
Now that you know about **HeaDS** and this course, we want to hear some things about **you**.

UCPH HeaDS

WHAT IS DATA SCIENCE?

UCPH HeaDS

WHAT DO THE WORDS MEAN?



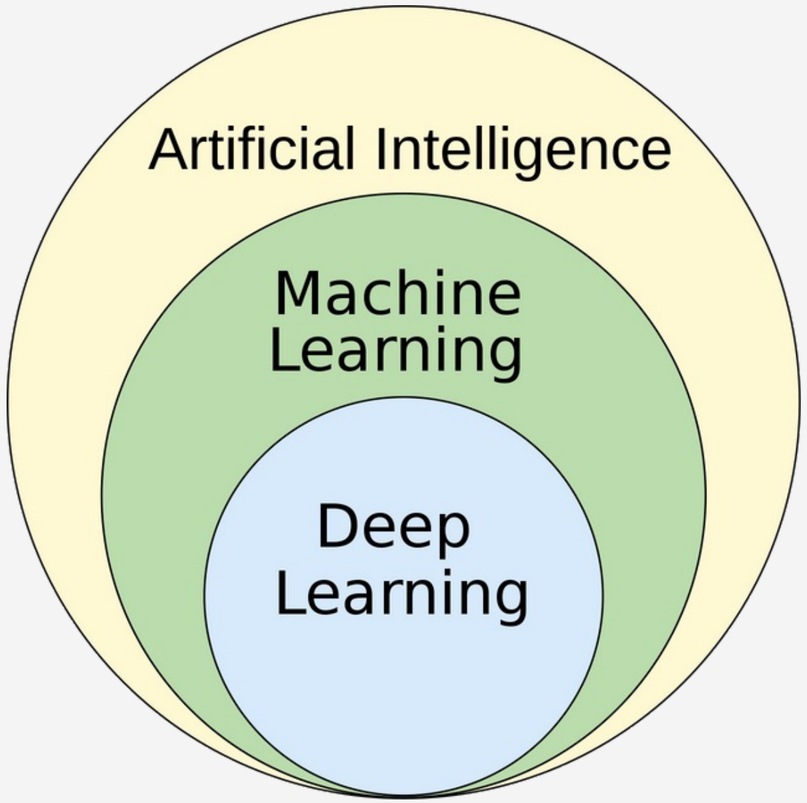
You have likely heard the words **Data Science (DS)**, **Machine Learning (ML)** and **Artificial Intelligence (AI)**.

These are often used as synonyms, **buuuut, they are not :)**

What do they mean? - Let's try some differentiation.

UCPH HeaDS

WHAT DO THE WORDS MEAN?

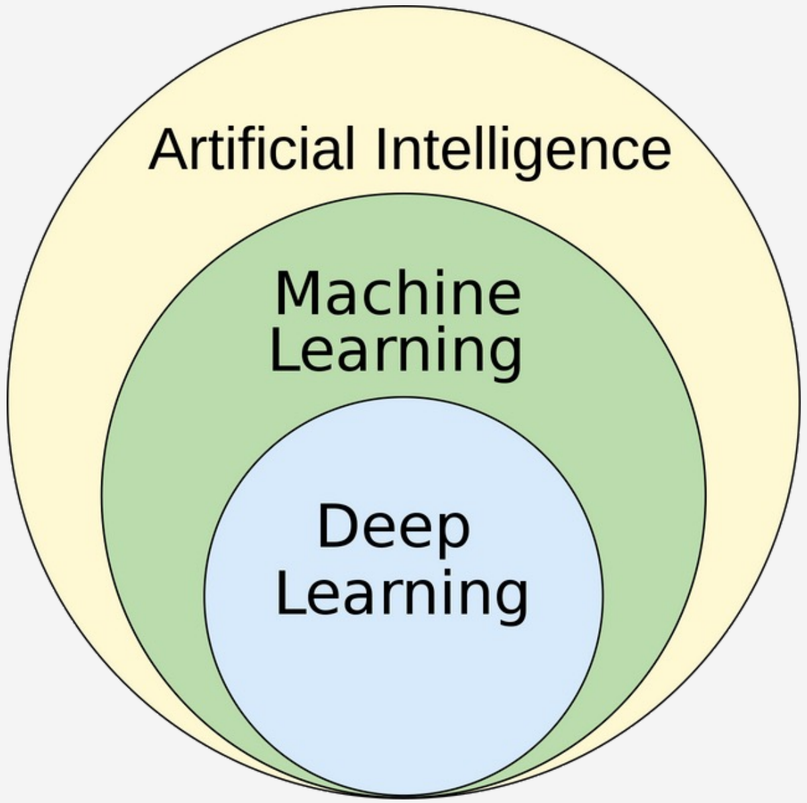


DATA SCIENCE != MACHINE LEARNING

Machine learning - development of algorithms that enable computers to learn from data and make predictions or decisions (with human-like performance or better).

Data Science (DS) - a variety of techniques for extracting knowledge from data. This involves cleaning, normalizing, and analyzing large data to uncover patterns and trends.

WHAT DO THE WORDS MEAN?



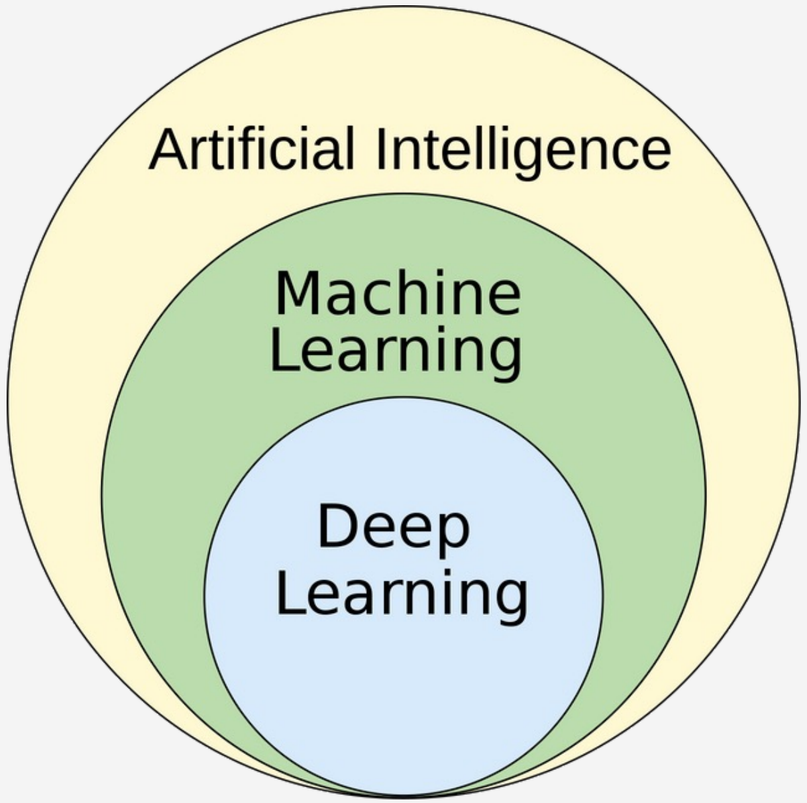
Artificial Intelligence:

- The ability of computing systems to achieve human-like performance on complex tasks
- Conceptual umbrella term

- AI is the **outcome, not the method**

UCPH HeaDS

WHAT DO THE WORDS MEAN?



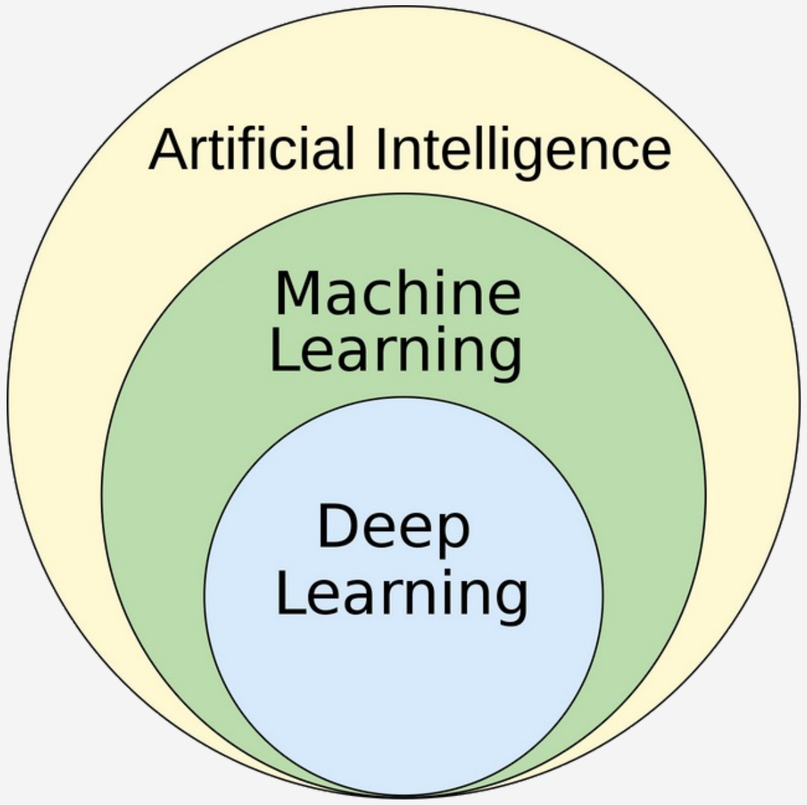
Machine Learning:

- "Technologies and algorithms that enable systems to identify patterns, make decisions, and improve themselves through experience and data" [1]
- Machine learning is methodology.

- Currently our most successful way of achieving AI

1. https://ai.engineering.columbia.edu/ai-vs-machine-learning/

WHAT DO THE WORDS MEAN?



Deep Learning:

- When you do **ML** by means of a **deep neural network**
- This is a type of/sub-class of (ML).

Data science is often taken to mean ML/AI, but the term is much broader than that!

...and this is not a course on ML/AI.

WHAT IS DATA SCIENCE?

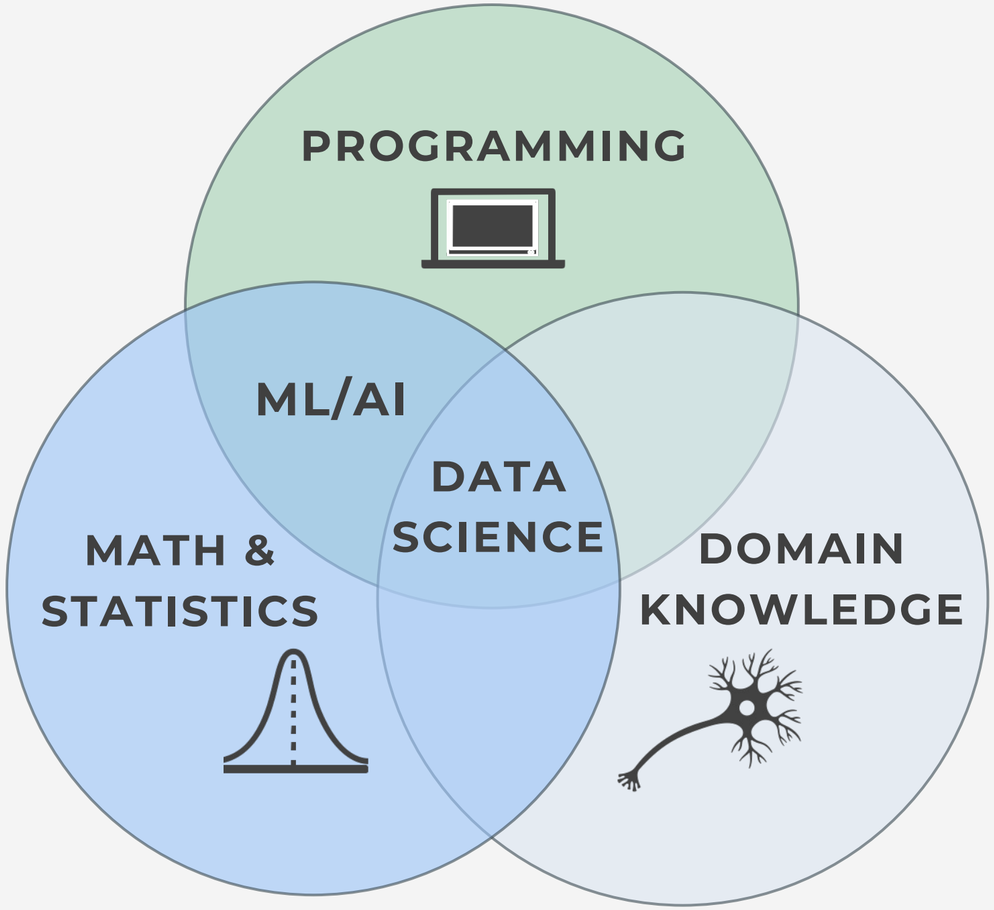
Data Science combines math, statistics, programming and algorithms with **domain expertise** in order to extract insights from data.
- **IBM**

Data Science is the processing and analysis of data with the goal of **learning something** about its characteristics or answer a scientific question.
- **HeaDS**

UCPH HeaDS

WHAT IS DATA SCIENCE?

- A **cross-disciplinary** undertaking that draws on many disciplines and is in turn becoming part of many disciplines.
- Data matters, one size analysis does not fit all.



UCPH HeaDS

WHY DO WE WORK WITH DATA?

Goal: To extract knowledge about how the world works and, if possible, make generalizations and predictions.

Data Science is the more **formalized process** where we make use of tools such as computers and algorithms to help us make sense of **vast amounts data**.

UCPH HeaDS

ROLES

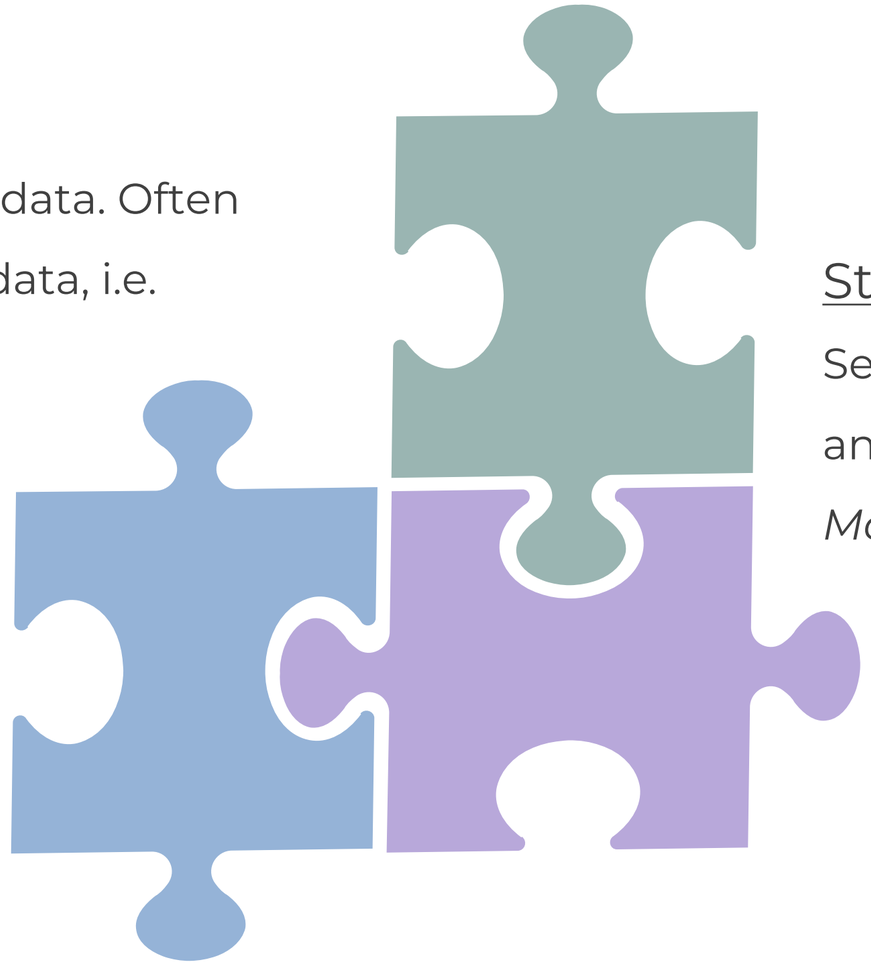
Not everybody is involved in every step.

Data Collector:

Produces or gives access to the data. Often has domain knowledge on the data, i.e. doctors working at the hospital

Principal Investigator:

Introduces the research question



Statistician/Mathematician:

Selects the appropriate tests and/or models

May do the data science analysis

UCPH HeaDS

ROLES

Data Scientist (Computer Scientist / Bioinformatician)

Does the data wrangling, cleaning and pre-processing

Does the data analysis

May standardize and/or implement as software

Visualisations expert:

Makes the plots to illustrate results

Selects proper type of plot

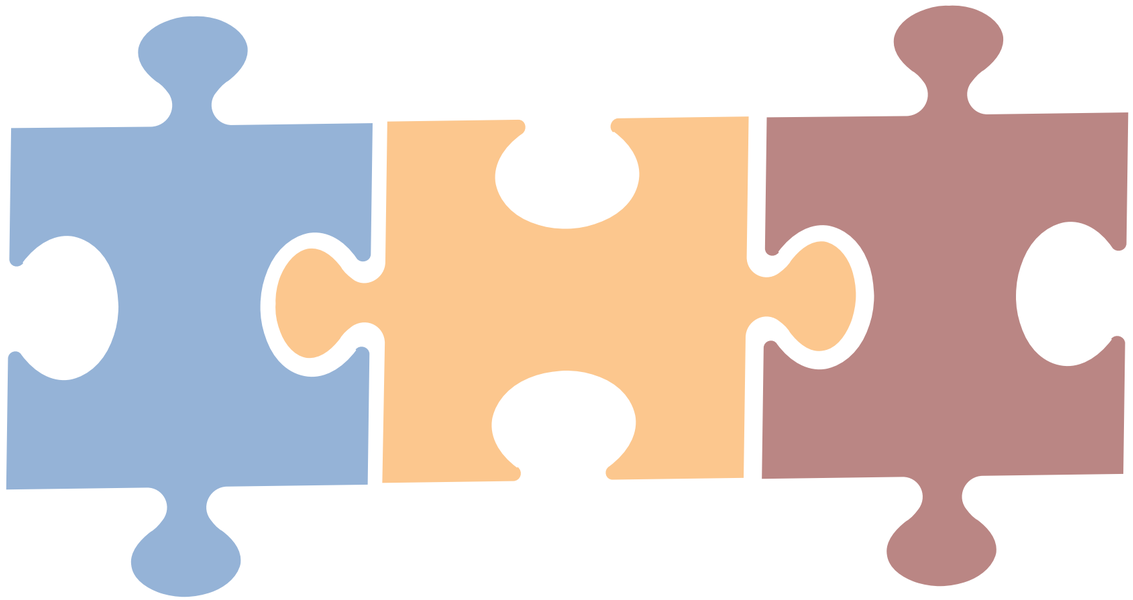
Focus on interpretability and accessibility

Domain expert:

Sparring about results

Biological/clinical relevance

Scientific Publication



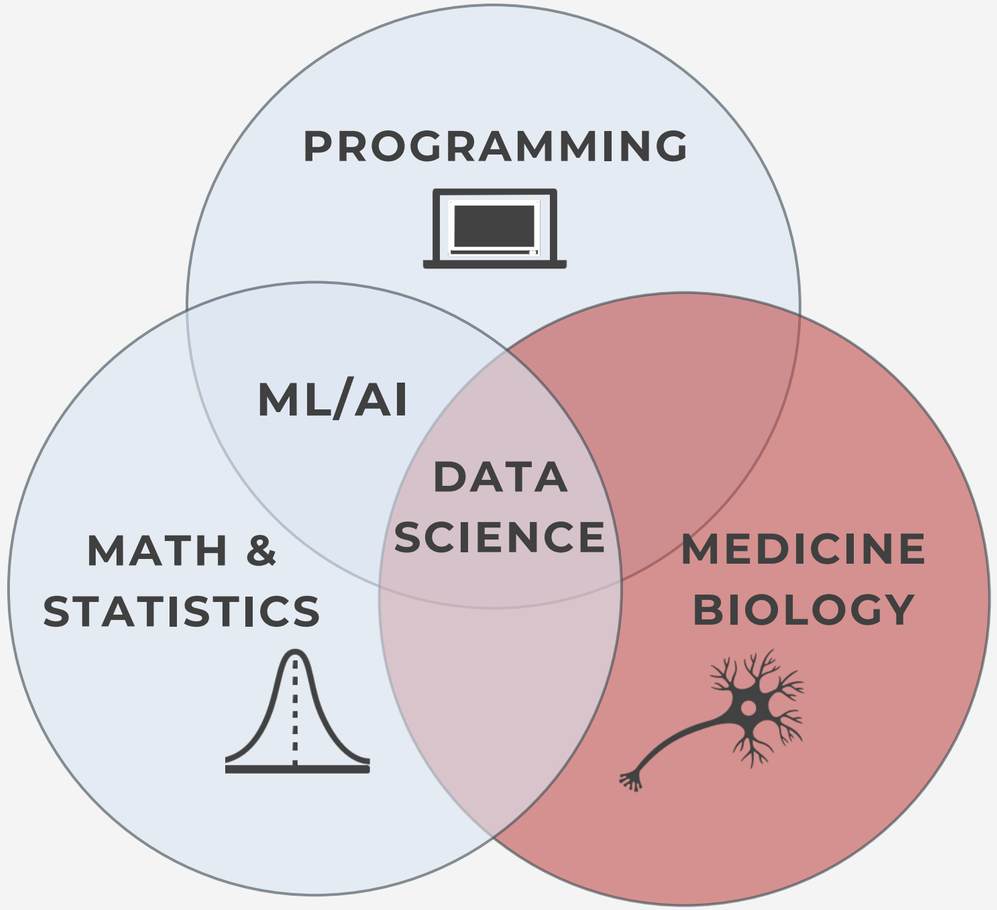
UCPH HeaDS

WHAT IS HEALTH DATA SCIENCE?

In **Health Data Science** the domains of interest are medicine (micro)biology, biochemistry, etc.

Field is concerned with:

- Biological mechanisms central to disease development
- Discovery and assessment of disease specific drug treatment
- Disease progression and patient survival
- **Personalized medicine**

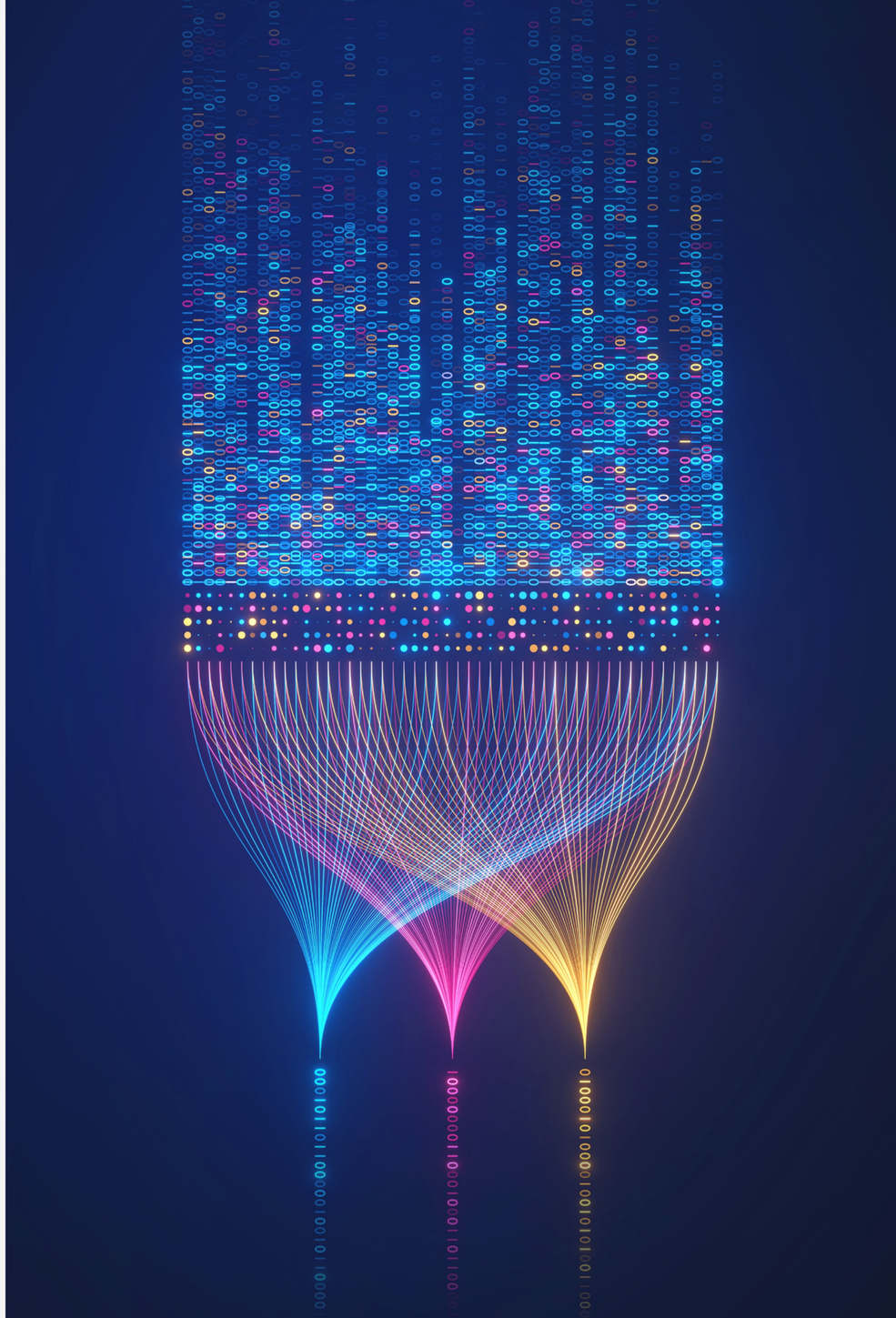


UCPH HeaDS

WE HAVE THE WHAT – NOW THE WHY?

The world is becoming data driven!

- Amount and quality of data is growing every year
- Now data drive research questions & theory
- **Improve *your* science**, as well as your **CV**
- **Teach the next generation** how to utilize big data!



ANALYSIS OF BIG BIO-MEDICAL DATA

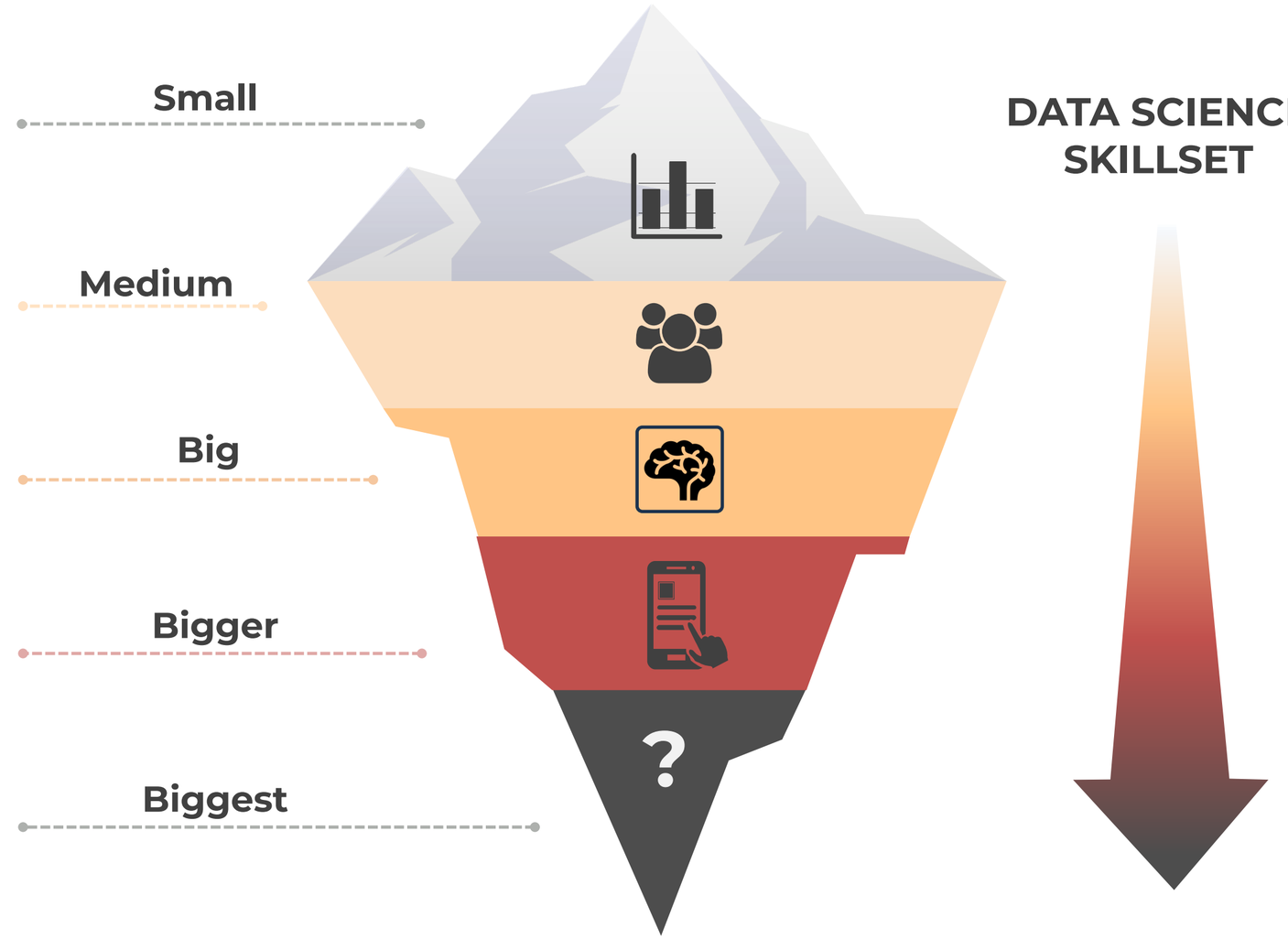
Research /Clinical Studies

Registries & Questionnaires
Pharmacological Databases

Biological (omics) Databases
Medical Image Data

Wearable Devices
Search Engine Data

‘Future’ Data



Mentimeter

Q4:

After this introduction, how do you see the importance of data science to you personally (i.e. in your research, for collaboration, etc.)?

UCPH HeaDS

GROUP DISCUSSION – 1.0

Which of the **roles** we have introduced **do you see yourself in**?

Do you have people in your group or among your collaborators to fill the other roles? If not, **what are the alternatives?**

UCPH HeaDS