The Alan Turing Institute

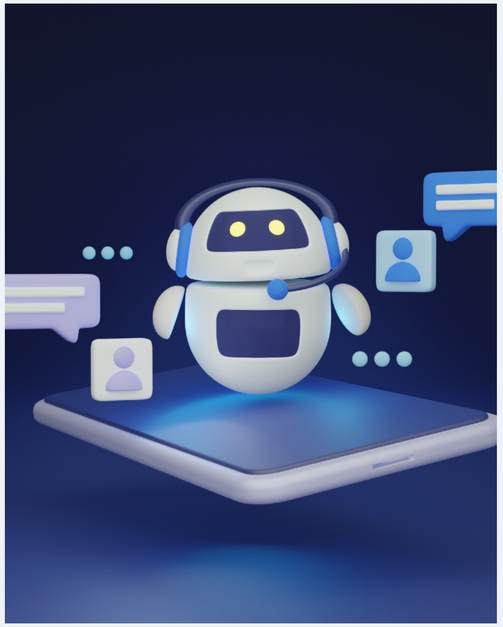
Turing Commons

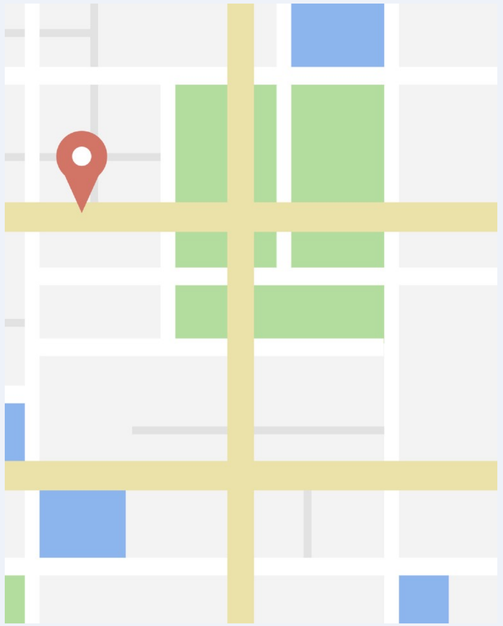
Understanding Data

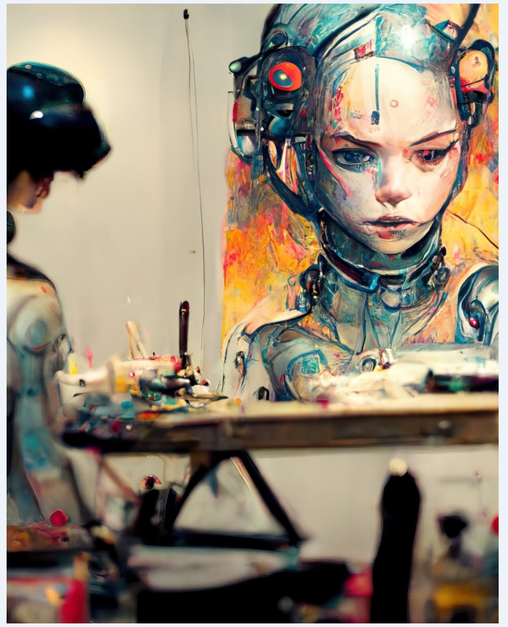
What is Data? Why is it important?

14/01/2023

Dr Christopher Burr
Claudia Fischer







OVERVIEW

A Case Study in Healthcare

01

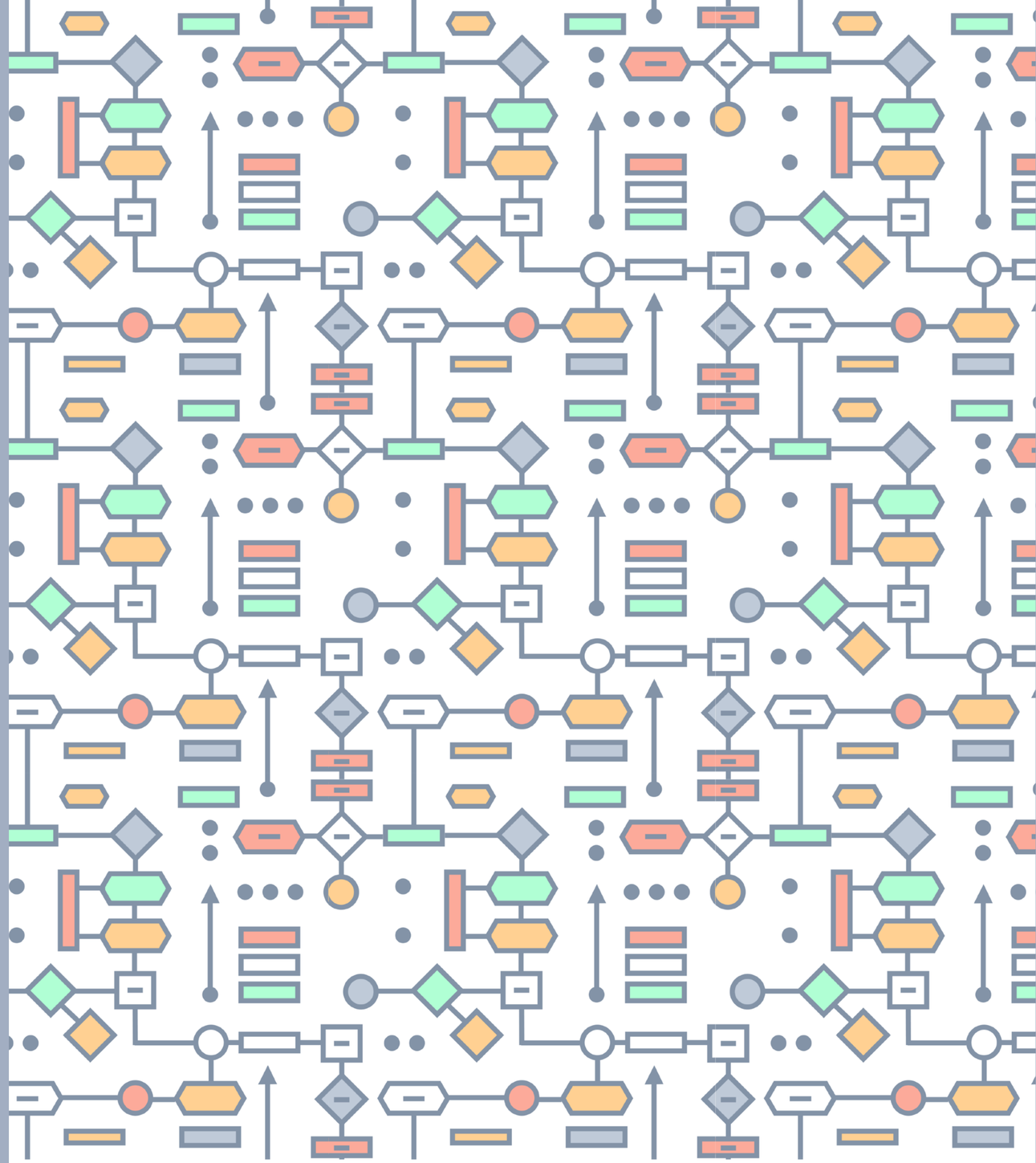
Activity 1: Data Types

02

Activity 2: Stakeholder Analysis

03

01 Case Study in Healthcare



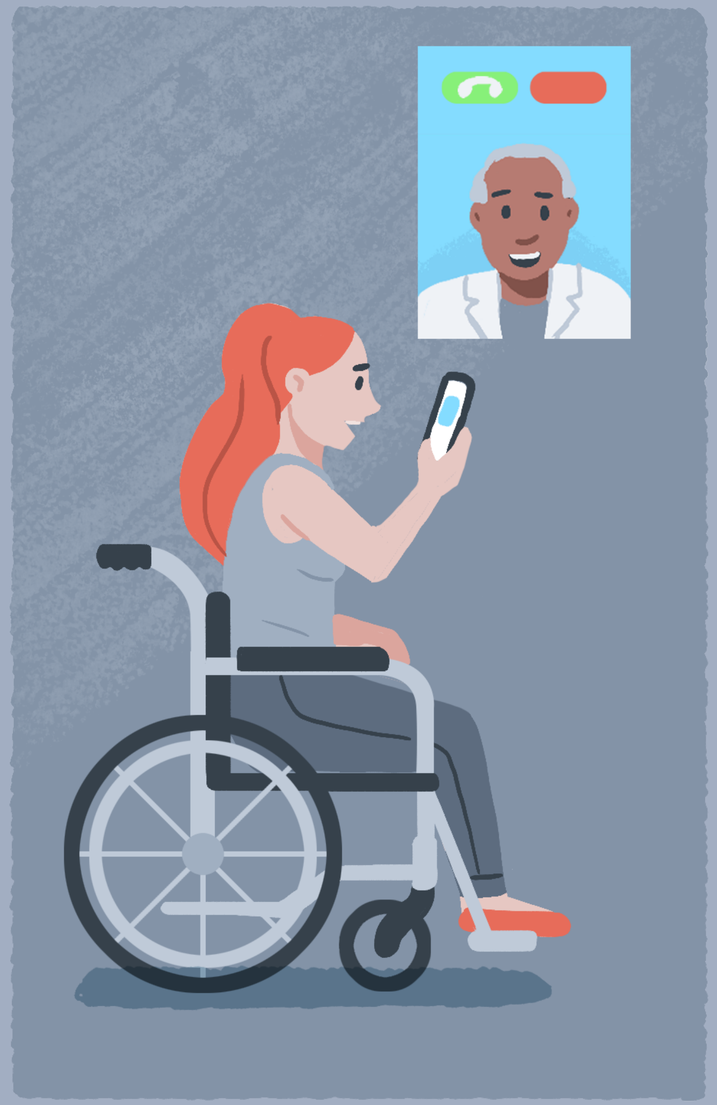
NHS Triaging Algorithm

An **algorithmic system** that can be used by frontline healthcare professionals (e.g. nurses, doctors) to assess the risk of patients entering a hospital.

NHS Triaging Algorithm

The system will use the data to **“sort” (or, classify)** patients into three categories:

- Low Risk
- Medium Risk
- High Risk



NHS Triaging Algorithm

The algorithm will use **data from previous patients** to identify and learn patterns that can help **predict health outcomes** (e.g. increased risk of requiring emergency treatment).


?
75%

The Lifecycle of Data

Data Collection

Data Analysis

Machine Learning Algorithm

System Use and Monitoring

The Lifecycle of Data

Data Collection

Data Analysis

Machine Learning Algorithm

System Use and Monitoring

DATA COLLECTION

1/ **Which** data are needed? How will the data **help** design and develop the system?

2/ **Can** the data be collected? **Should** the data be collected?

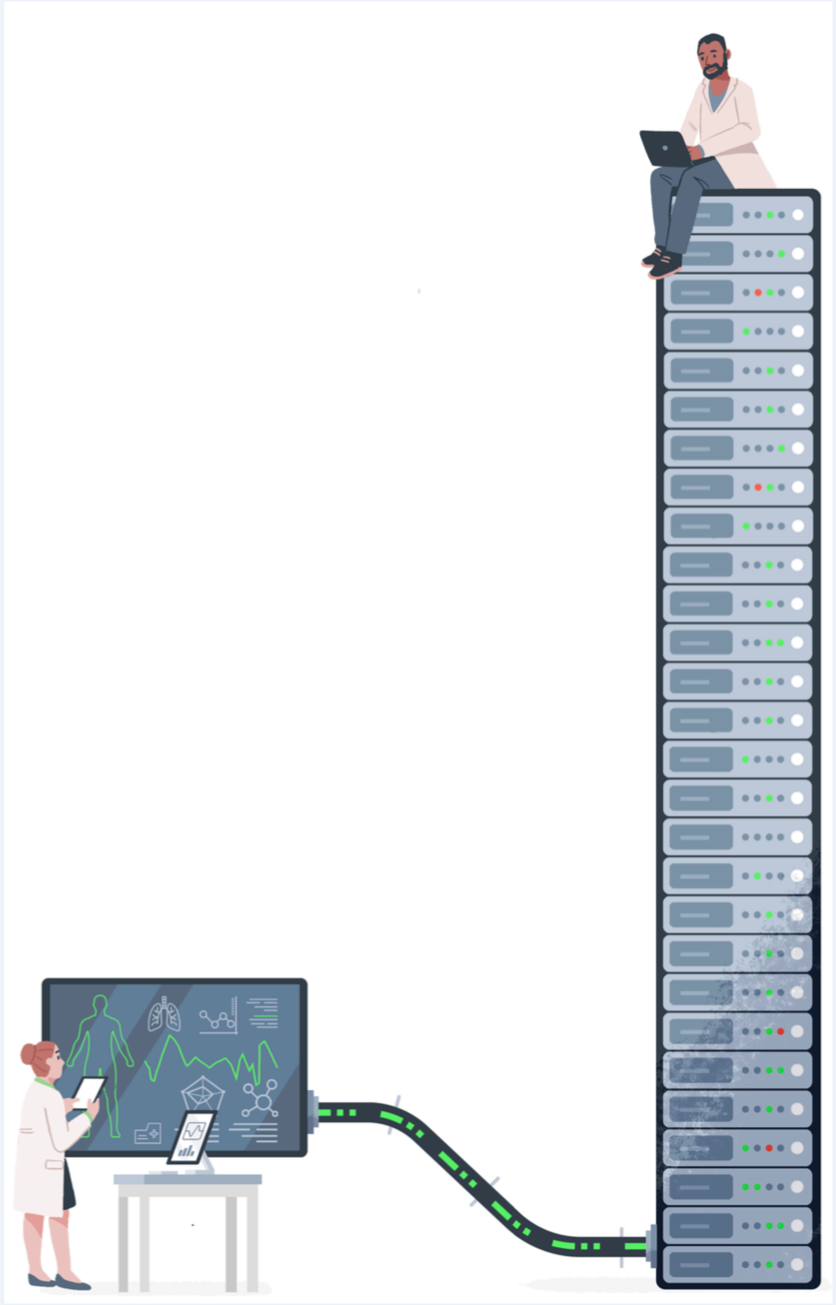
3/ Does the team have **sufficient resources** to use and analyse the data? Can they keep the data **secure**?

DATA ANALYSIS

1/ Which data types **are strongly associated** with the risk categories?

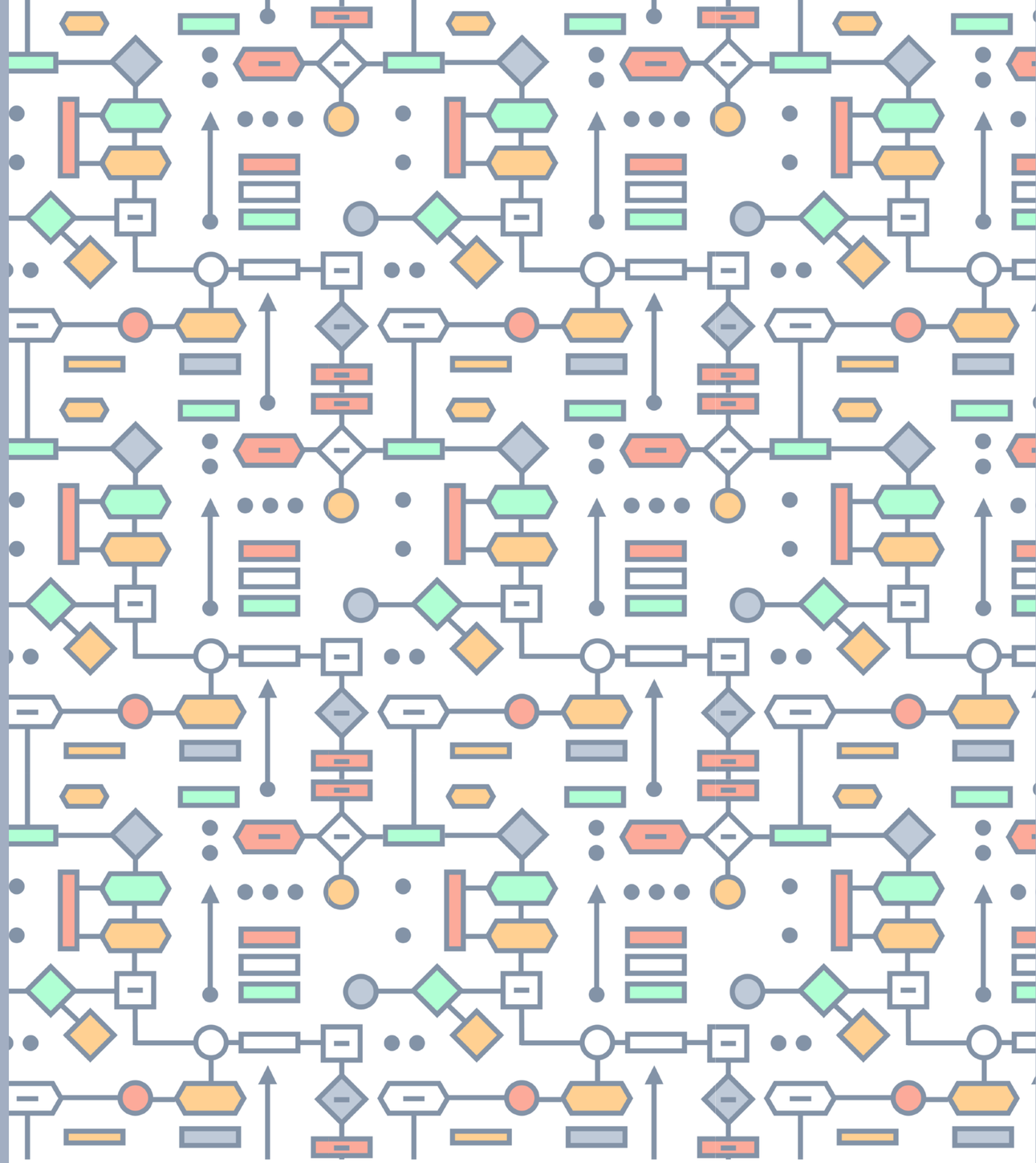
2/ Are there patterns in the existing data that could **help predict future risk**?

3/ Do these patterns make sense? Are they **acceptable** to all patients and staff?



02

Activity 1: Data Types



Which data types do you think are appropriate to use?

Why?

Which data types do you think are problematic?

Why?

The Lifecycle of Data

Data Collection

Data Analysis

Machine Learning Algorithm

System Use and Monitoring

The Lifecycle of Data

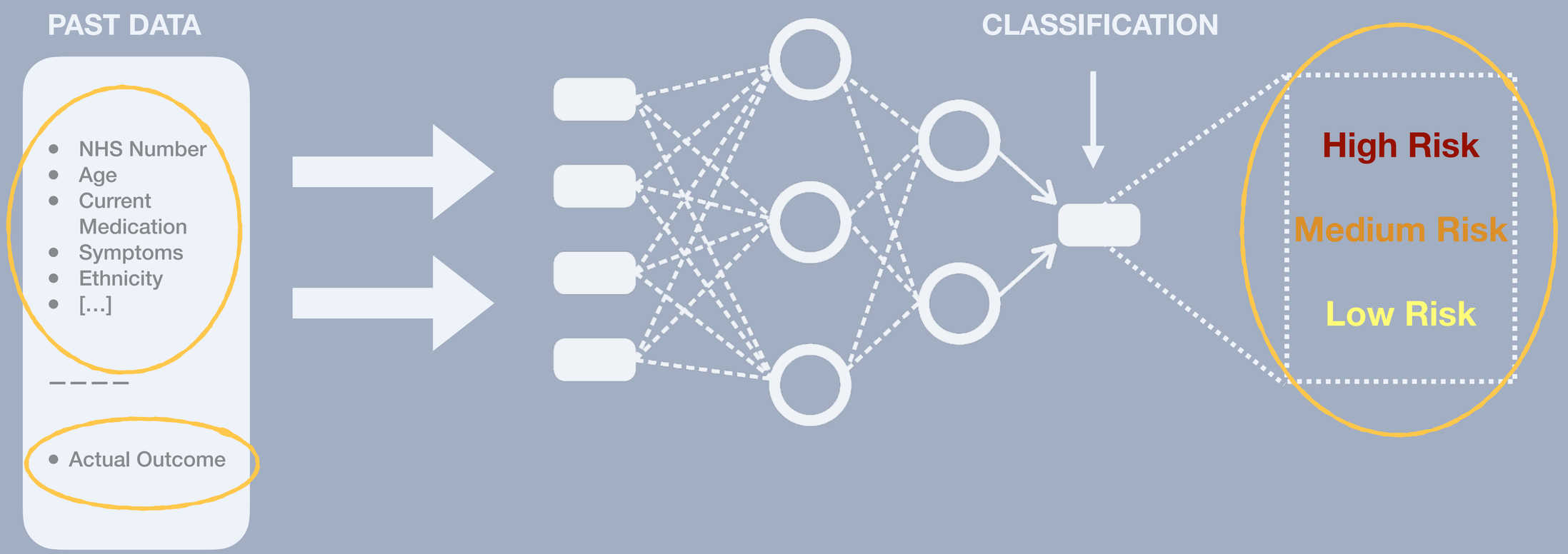
Data Collection

Data Analysis

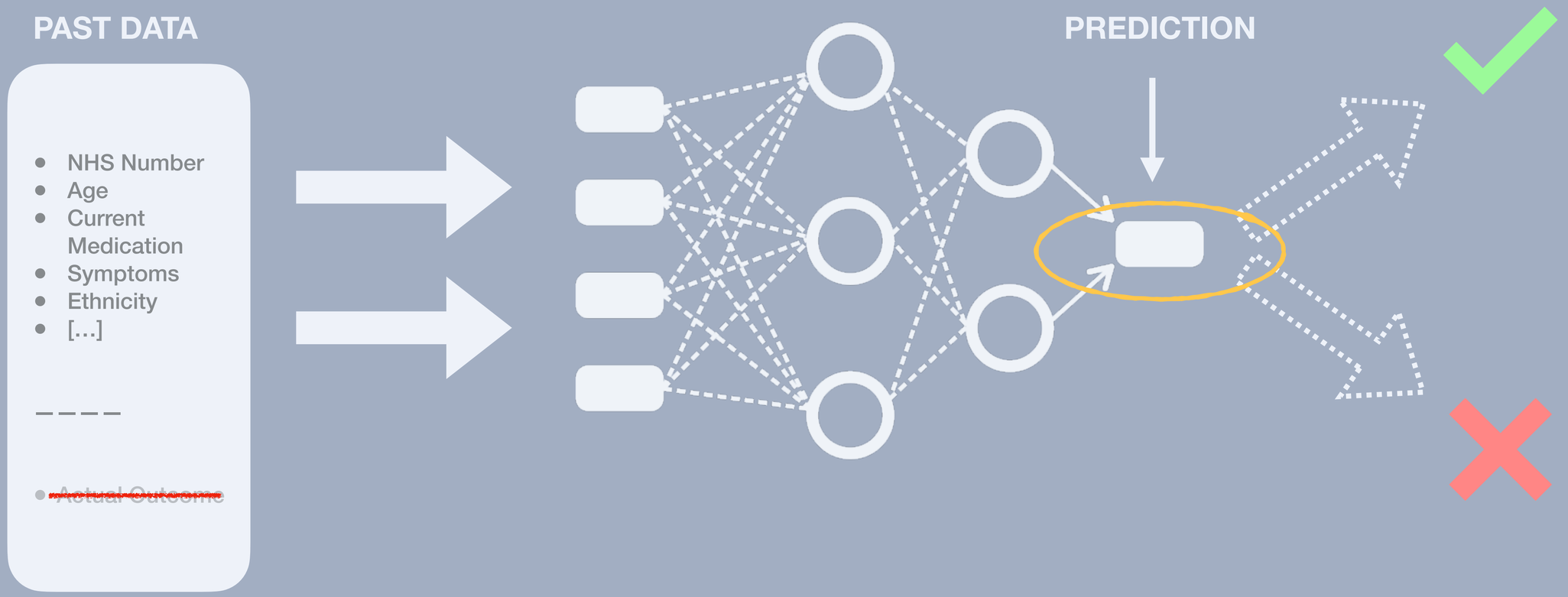
Machine Learning Algorithm

System Use and Monitoring

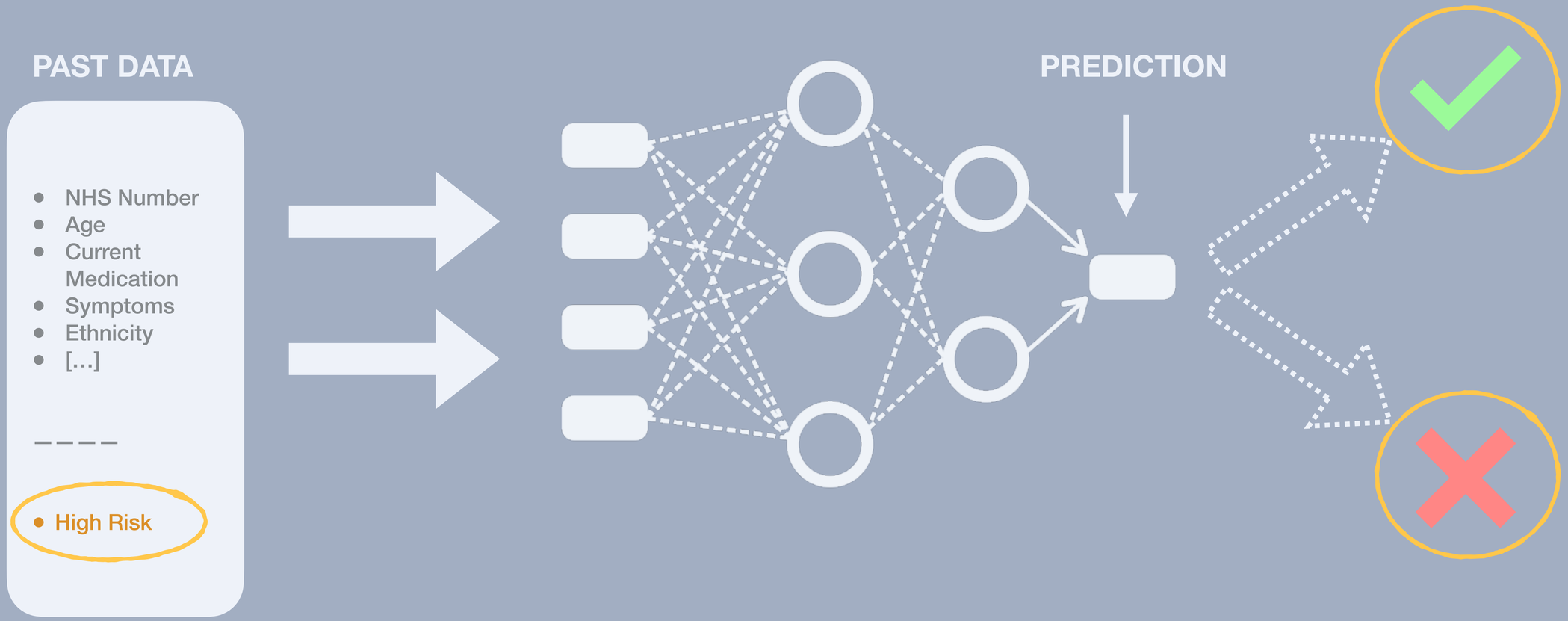
MACHINE LEARNING ALGORITHM



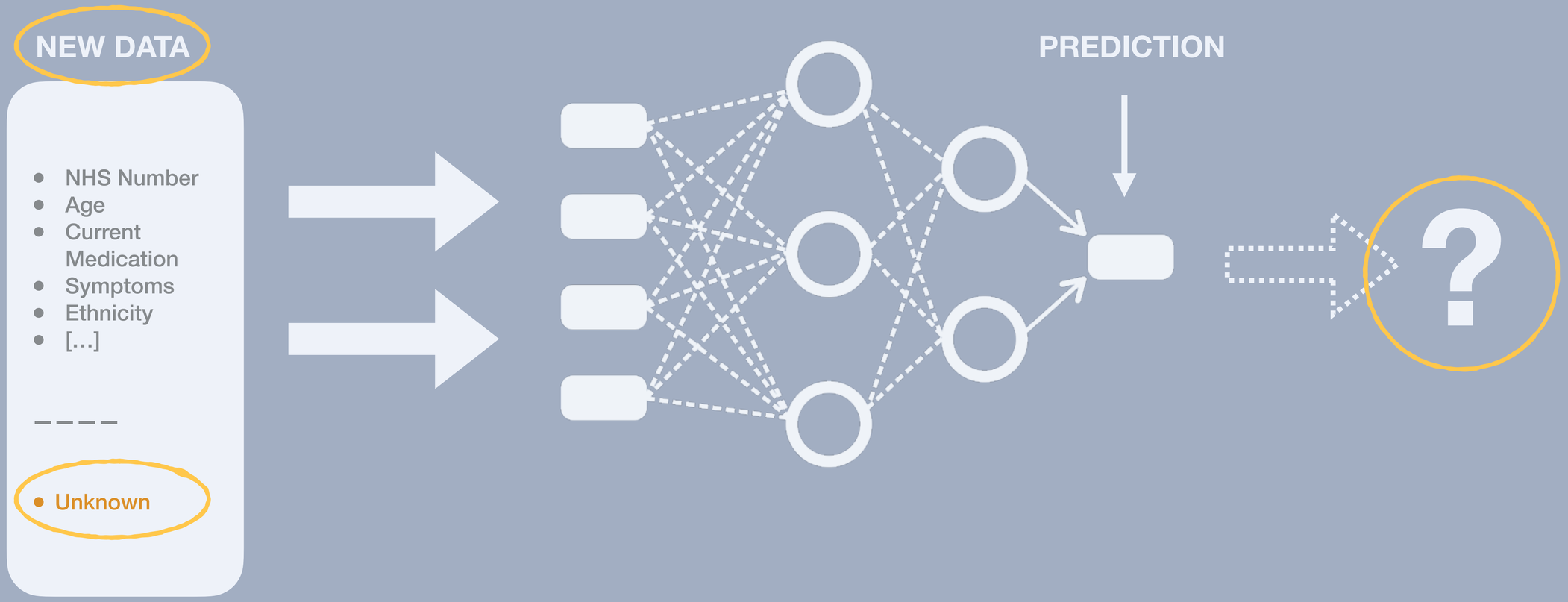
MACHINE LEARNING ALGORITHM



MACHINE LEARNING ALGORITHM



MACHINE LEARNING ALGORITHM



SYSTEM USE AND MONITORING

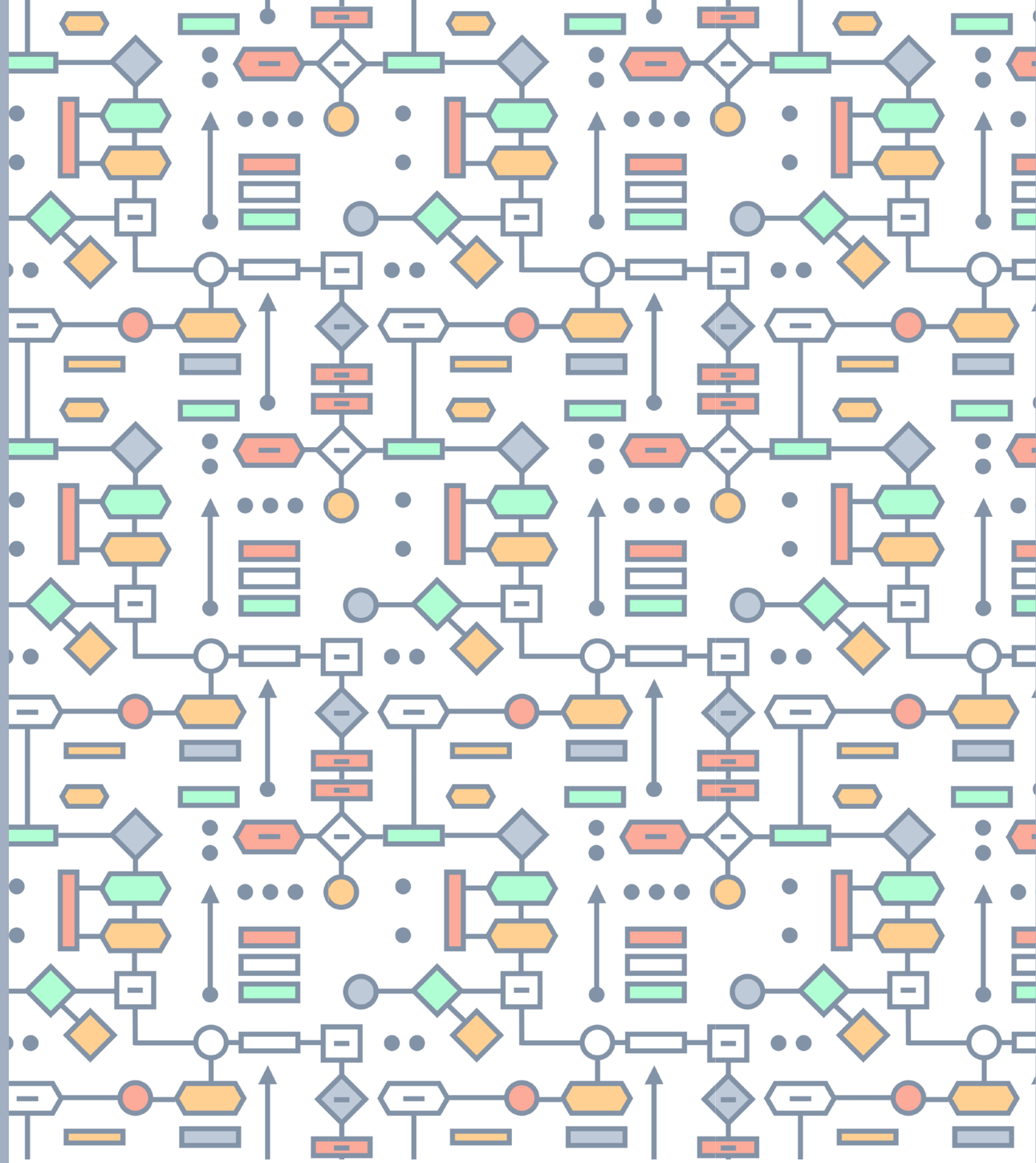
1/ Is the algorithmic system working as expected? Is it **accurate** when used in real situations?

2/ Are there **specific patients or conditions** where the algorithmic system fails to perform well?

3/ What **new possibilities** does this system open up (e.g. remote diagnosis)?

03

Activity 2: Stakeholder Analysis



The data types used in this project are about people and affect people.

But people are not represented or affected in equal ways.

PERSONAS

1/ Set of profiles and data for hypothetical patients.

2/ Helps you consider a range of situations and circumstances.

3/ How could the people be differentially affected by the algorithmic system?

George

- Gender: Male
- Age Range: 51-60
- Ethnicity: Black
- BMI: 27.1
- Smoker: True
- Alcohol consumption: 12
- Postcode: SE15 2DF
- Number of Dependants: 1

George is a 60 year old black British man. He owns a popular restaurant at a high street within an economically deprived area in London.
He usually works late on weekdays and at least one day on weekends.

George's oldest daughter is suffering from severe depression. She moved into George's flat a few months ago and he is financially supporting her.

Hayley

- Gender: Female
- Age Range: 31-40
- Ethnicity: White
- BMI: 25.4
- Smoker: True
- Alcohol consumption: 1
- Postcode: E7 9PE
- Number of Dependants: 2

Hayley is a 40 year old white British woman and office assistant.

Hayley and her husband's living situation has been overcrowded since the birth of their third child a year ago, and greatly worsened when they started working from home.

The lack of space has been extremely challenging for their eldest son in particular, who has Autism Spectrum Disorder (ASD).

- How could this person benefit from the use of the triaging system?
- Are there any risks or harms that could arise from using this system, which would disproportionately affect this individual (or people like them)?
- Is it likely that this person would be concerned about the collection and use of specific data types (see previous activity)?
- Any other thoughts or comments?

Thank You!

The Alan Turing Institute

Turing Commons

Understanding Data

Dr Christopher Burr

Claudia Fischer