# Algorithmic decision-making in neuroscience: how can we improve algorithmic interpretability and reduce bias?


Florence Aellen

Athina Tzovara

University of Bern

florence.Aellen@inf.unibe.ch

athina.Tzovara@inf.unibe.ch

@AthinaTzovara

# Flo Aellen: About

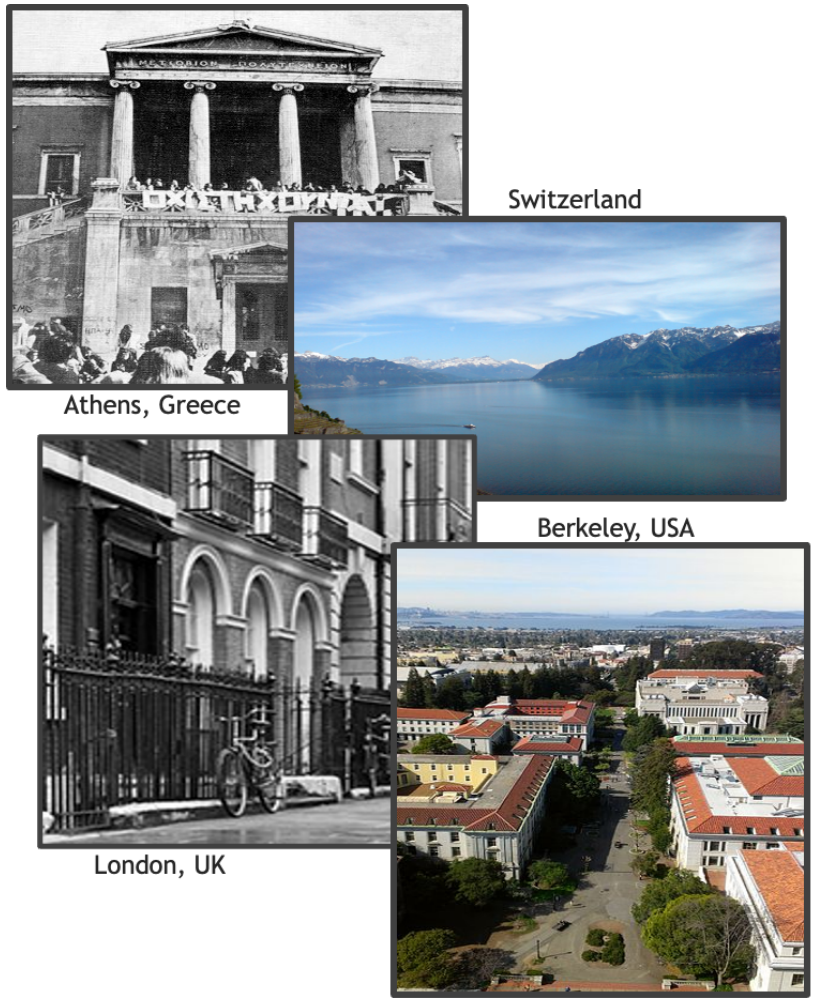
**Today:**
PhD Candidate; Institute of Computer Science;
University of Bern, Switzerland

**Past:**
Master of Science in Theoretical Physics;
University of Bern, Switzerland

Bachelor of Science in Mathematics;
University of Bern, Switzerland

b
UNIVERSITÄT
BERN

# Athina Tzovara: About




**Today:**
Assistant Professor
University of Bern, Switzerland

Visiting scholar at Helen Wills Neuroscience Institute, UC Berkeley, USA.

**Past:**
Postdoc at University of Zurich, Switzerland & Wellcome Centre for Human Neuroimaging, UCL, UK.

PhD in Neuroscience, University of Lausanne, Switzerland.

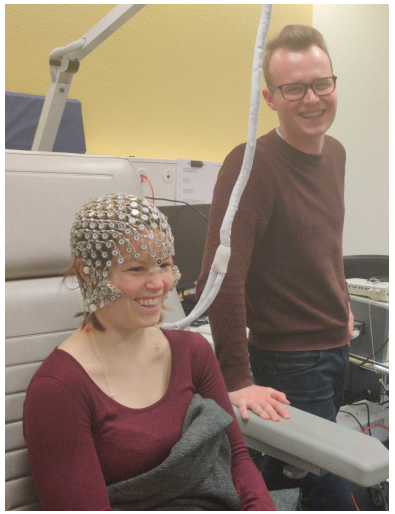
Electrical & computer engineering, National Technical University of Athens, Greece.

# Who are we?

Cognitive Computational Neuroscience Group
Institute of Computer Science University of Bern
Sleep Wake Epilepsy Centre, Inselspital, University Hospital Bern
https://neuro.inf.unibe.ch/

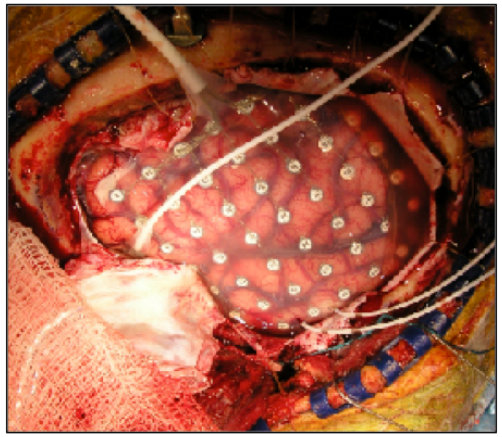
UNIVERSITÄT BERN

## Scalp EEG



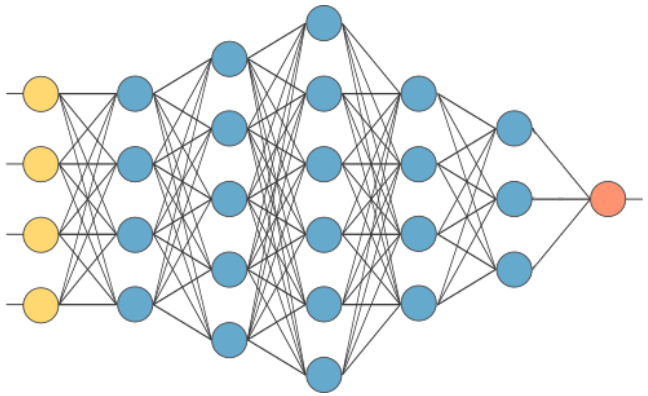
## MEG

## Intracranial EEG

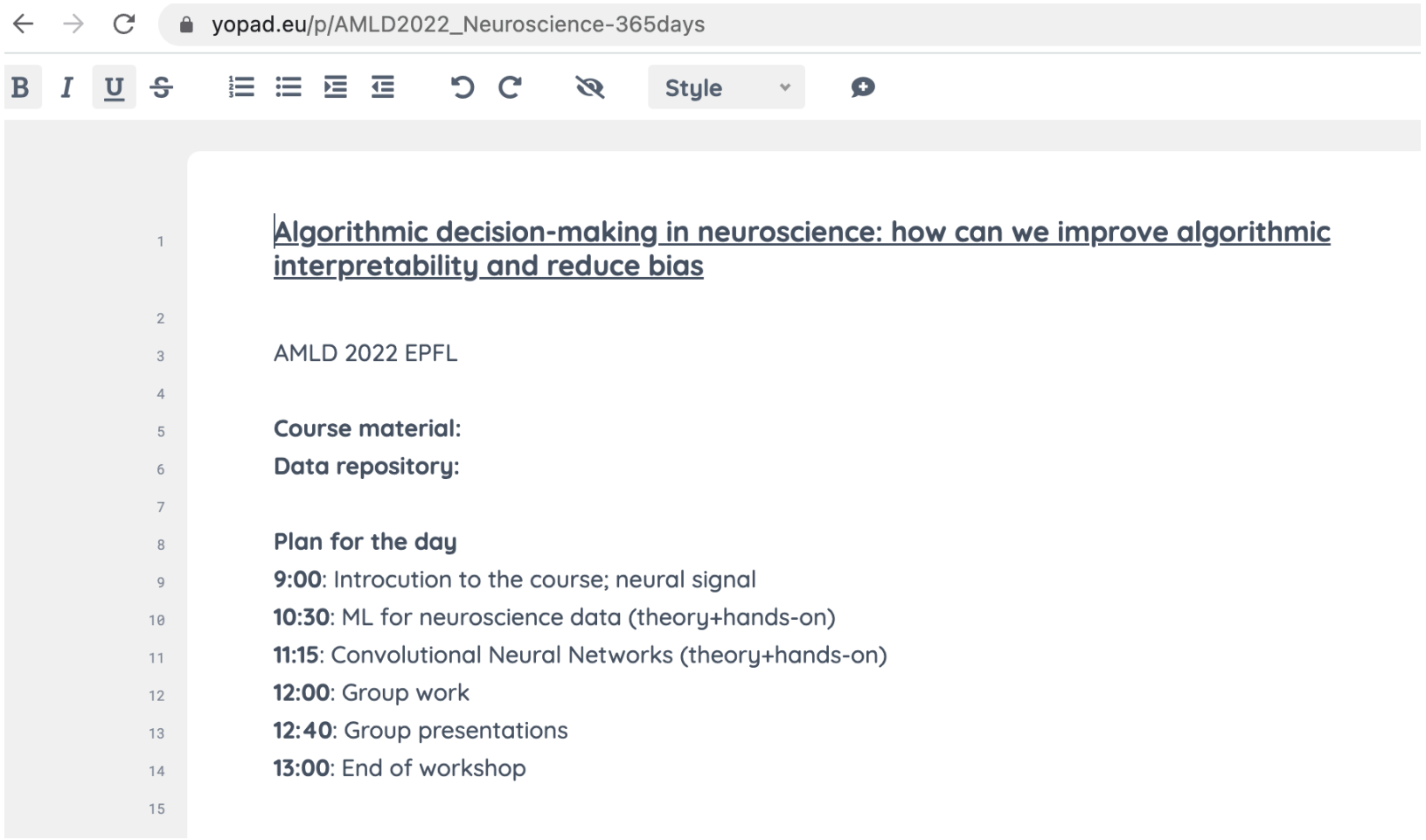


## Machine learning Computational modeling

+



# Who are you?




Experience with ML?

Experience with Python?

Experience with Neuroscience/EEG?

Why are you here?

https://yopad.eu/p/AMLD2022_Neuroscience-365days

# Overview for today

## Introduction to AI in neuroscience :

- Electroencephalography (EEG) signals
- Hands-on: working with EEG

## Machine Learning in neuroscience

- Supervised learning: training classifiers
- Measuring performance
- Hands-on: Classifying EEG data

## Convolutional Neural networks for EEG signals

- Training networks & measuring performance
- Hands-on: working with neural networks

## Group work & presentations:

- Mini projects: try out what we learned in short projects & your own ideas

# Neuroscience and AI?

**In a first, brain implant lets man with complete paralysis spell out thoughts: 'I love my cool son.'**

Surgically placed electrodes enable person with late-stage ALS to communicate via neural signals

22 MAR 2022 · 12:00 PM · BY KELLY SERVICK

**An algorithm is learning to detect whether patients will wake from a coma**

**Artificial Intelligence Detects Epileptic Seizures in Real Time**

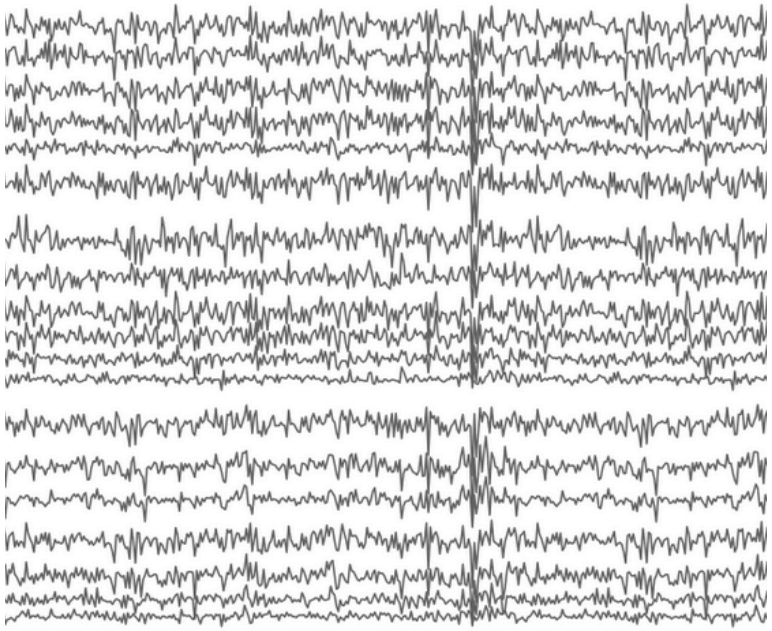
An artificial intelligence system was able to efficiently and accurately identify epileptic seizures in real time.

**Artificial Intelligence May Boost Sleep Disorder Treatment, Diagnosis**

Artificial intelligence has the potential to enhance sleep studies, improving the diagnosis and treatment of sleep disorders.

# Neurology today: advanced Neuro-monitoring techniques

**Electroencephalography -EEG-**



Sleep disorders
Epilepsy
Coma

# Computer Science: Artificial Intelligence algorithms

**Powerful algorithms**

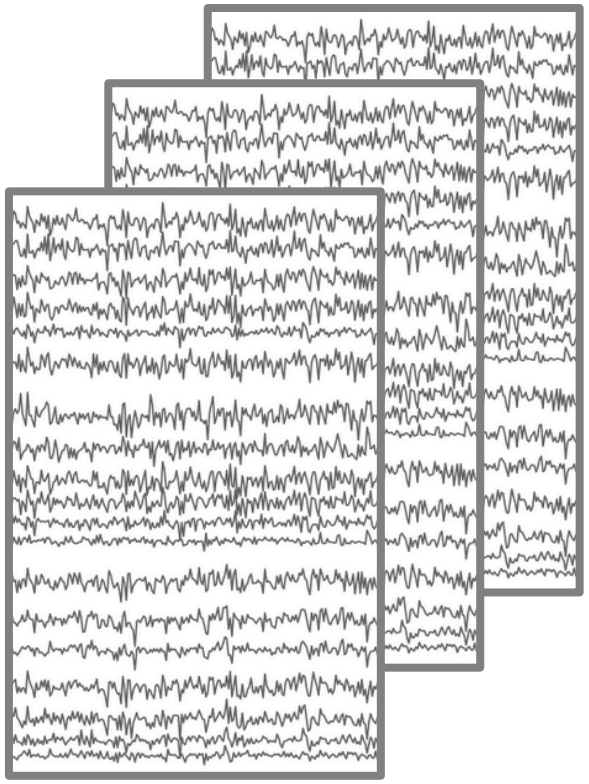
**Labeling images; videos; speech**

**Neurological data?**
**Healthcare?**

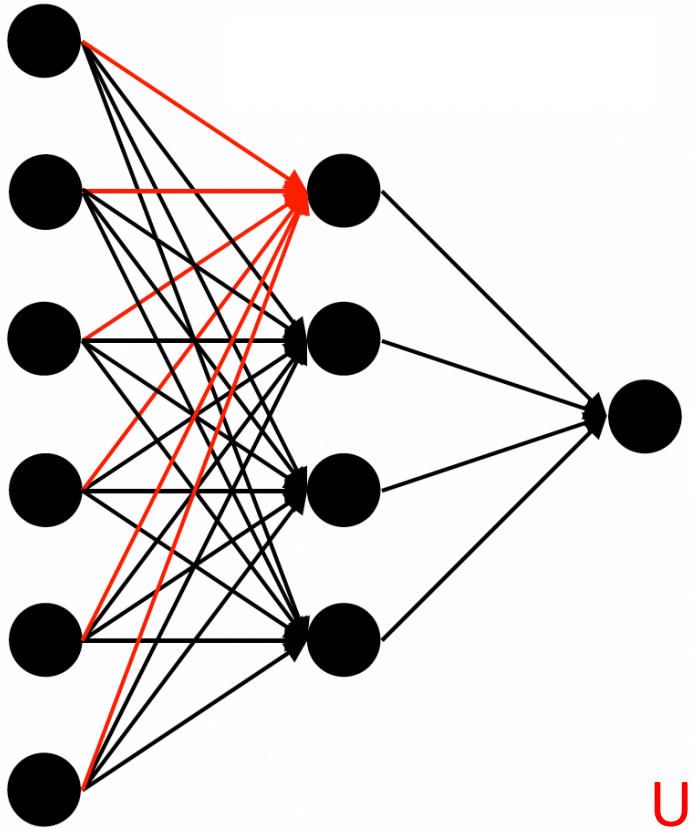
# Our approach: Artificial Intelligence in Neurology

F. Aellen

## EEG data



## Machine Learning



Patterns of neural activity

Requirements:
Understandable by a human
Interpretable
Bias-free

Aellen et al., 2021

# Predicting outcome from coma after cardiac arrest

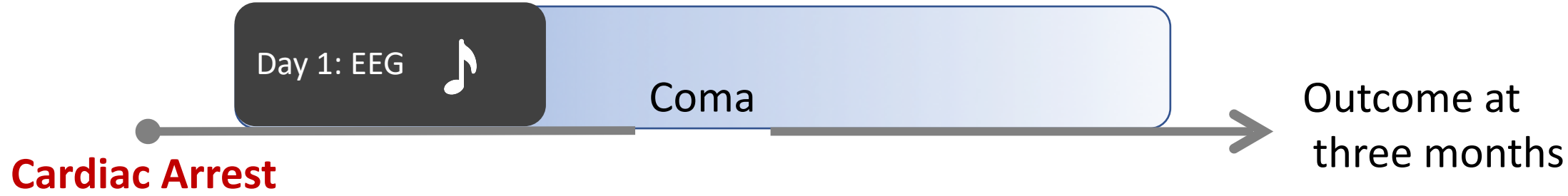
19 million cases per year worldwide

Cardiac arrest is the leading cause of coma in developed countries

Early medical intervention has a strong impact on patients' chance of survival

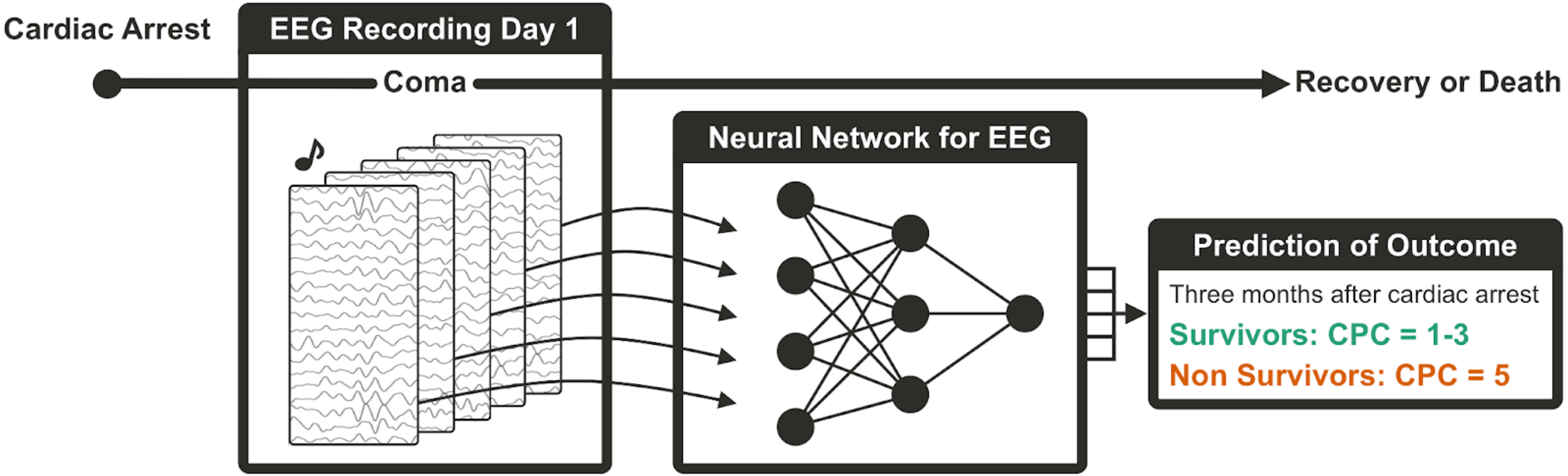
# Artificial Intelligence at patients' bedside

Predictors for patients' outcome?



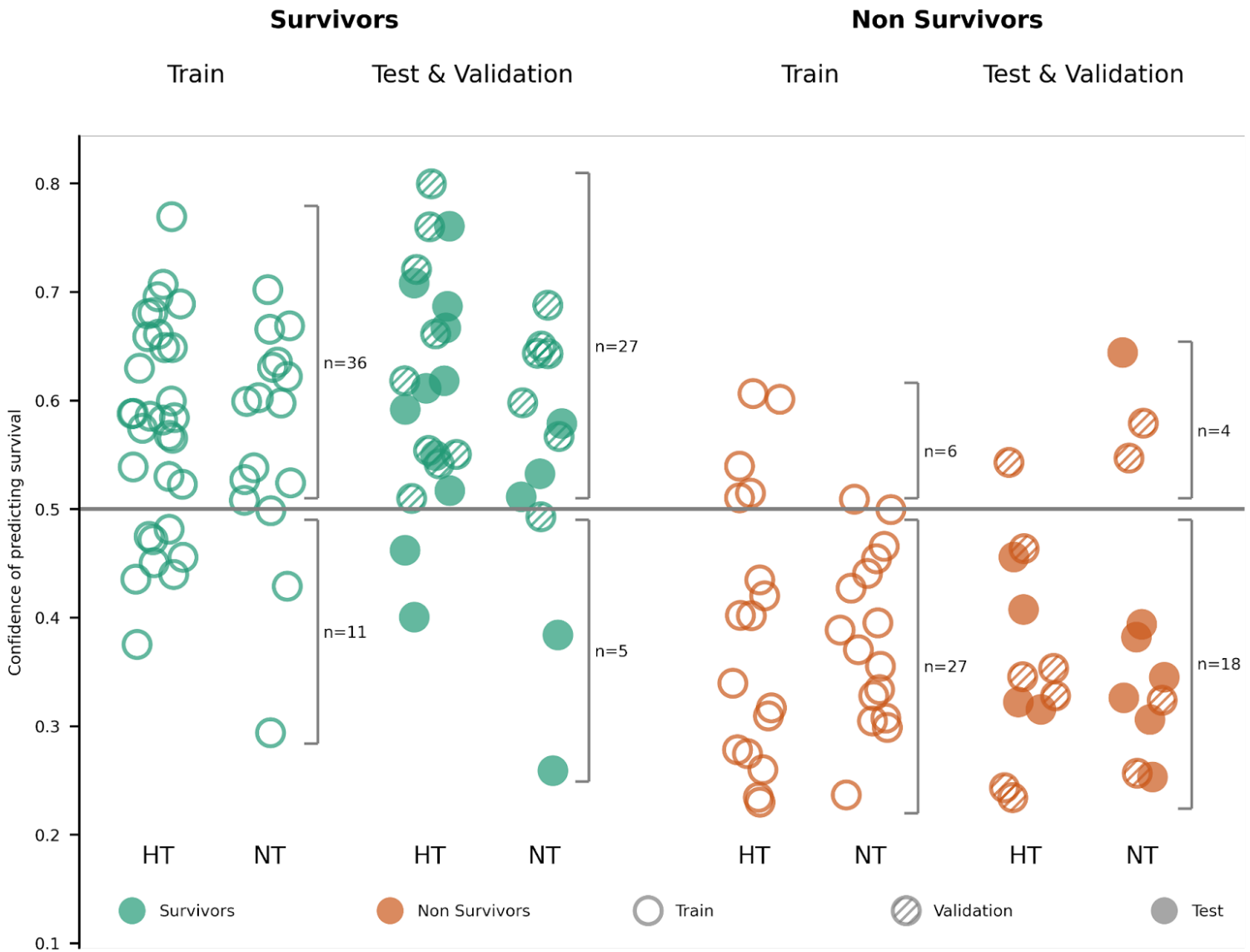

Tzovara et al., Brain 2013; Rossetti et al., J of Clin Neuroph 2014
Tzovara et al., Brain 2015; Tzovara et al., Annals of neurol. 2016;
Alnes et al., Neuroimage 2021; Aellen et al., under review
International patent application: PCT/EP2013/055036

# Artificial Intelligence at patients' bedside




Aellen et al., Under review

# Artificial Intelligence at patients' bedside




Convolutional Neural Networks:

Trained on day 1 of coma

Predict chances of awakening 3 months later

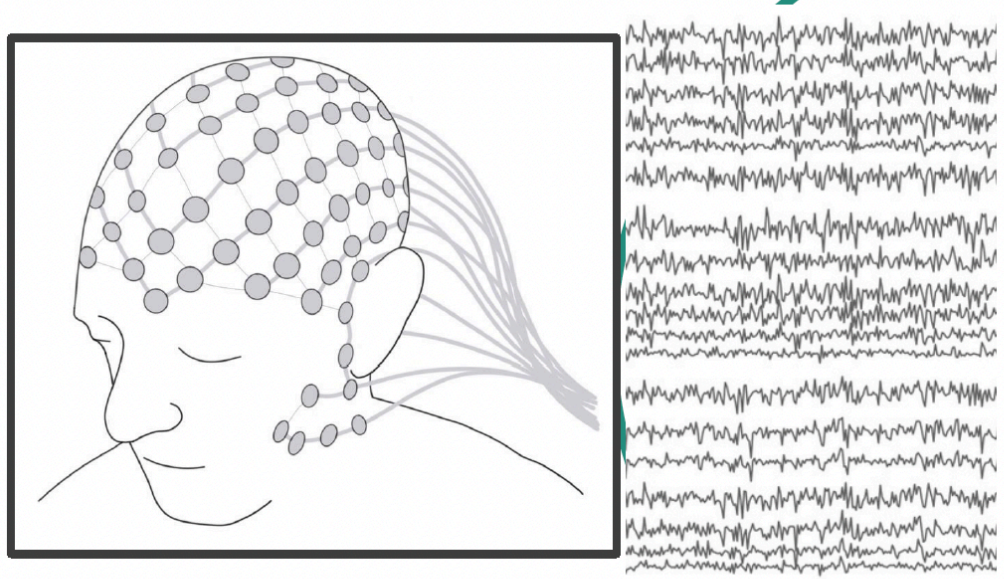
Aellen et al., under review

# Vision for the future

Excellent performance

Interpretable?

**Neurology**

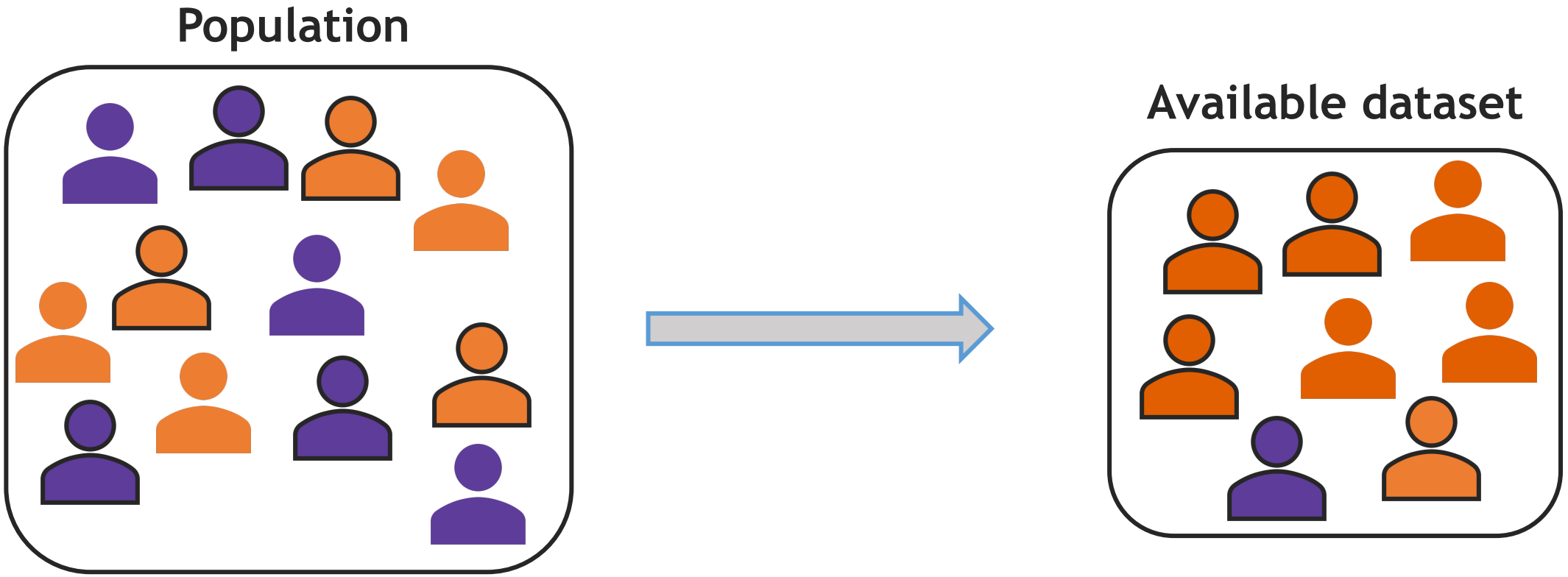


**Artificial Intelligence**

Bias-free?

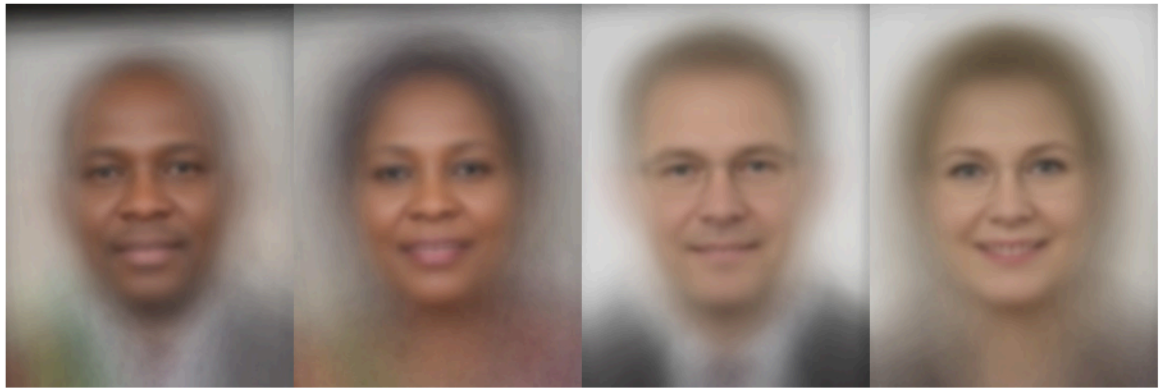
Fair?

# An algorithm can be as good as the data we used to train it



# Data might reproduce human bias at scale

| Gender Classifier | Darker Male | Darker Female | Lighter Male | Lighter Female | Largest Gap |
|---|---|---|---|---|---|
| Microsoft | 94.0% | 79.2% | 100% | 98.3% | 20.8% |
| FACE++ | 99.3% | 65.5% | 99.2% | 94.0% | 33.8% |
| IBM | 88.0% | 65.3% | 99.7% | 92.9% | 34.4% |



Source: MIT Media Lab, Joy Buolamwini
BY-NC-ND license

- Bias is present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups
- This bias reflects the composition of datasets on which algorithms are trained

# WEIRD Individuals

**Behavioral and Brain Sciences, Volume 33, Issue 2-3**

June 2010, pp. 61-83

## The weirdest people in the world?

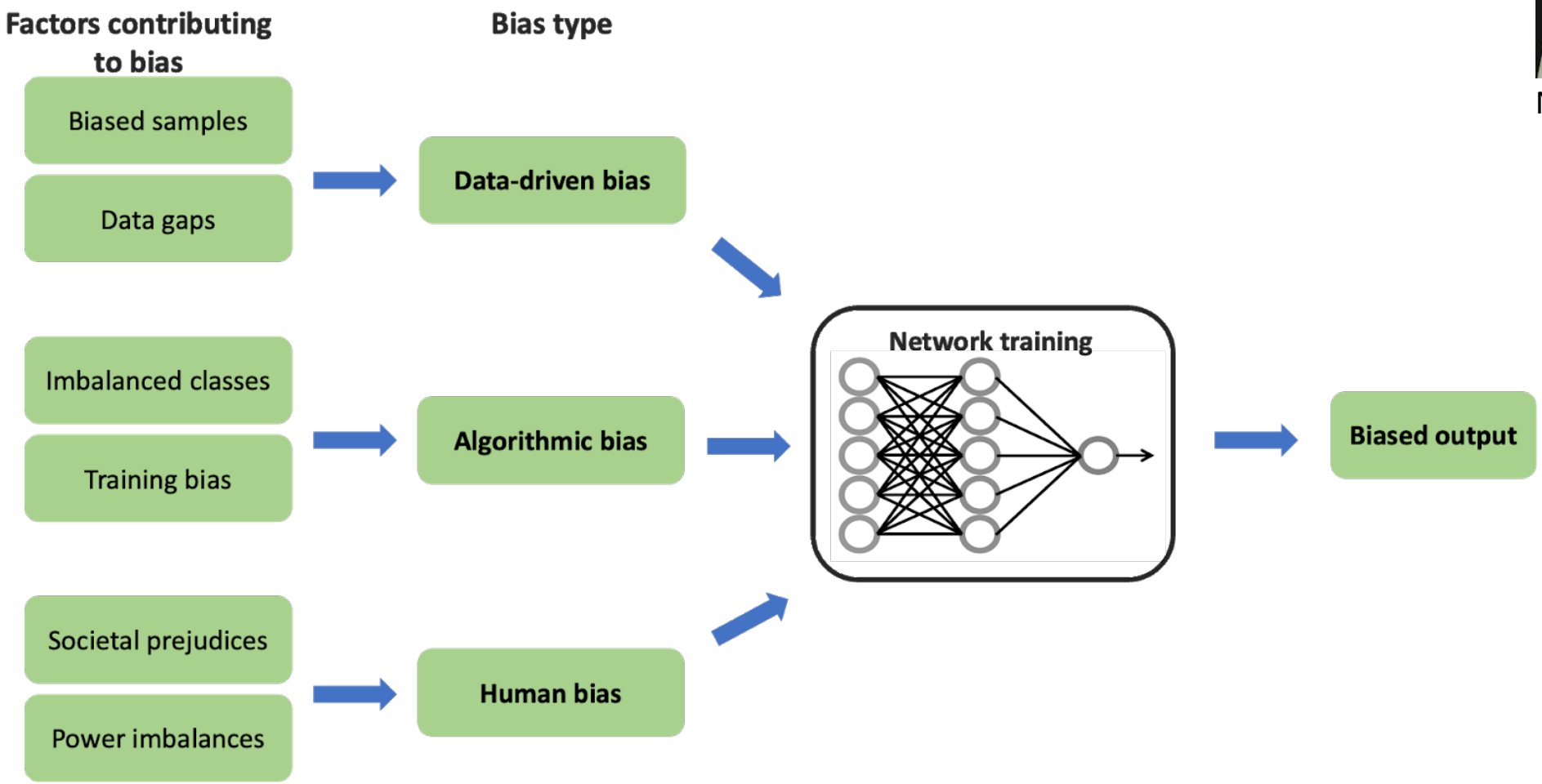
Joseph Henrich [a1], Steven J. Heine [a2] and Ara Norenzayan [a3]

**W**estern
**E**ducated
**I**ndustrialized
**R**ich
**D**emocratic

# Caution:Bias in available datasets & by extension in AI


N. Norori


Q. Hu




Norori et al., 2021

# Learning outcomes for today

1. Obtain an understanding of EEG signals: what we measure, how we can use as a diagnostic tool

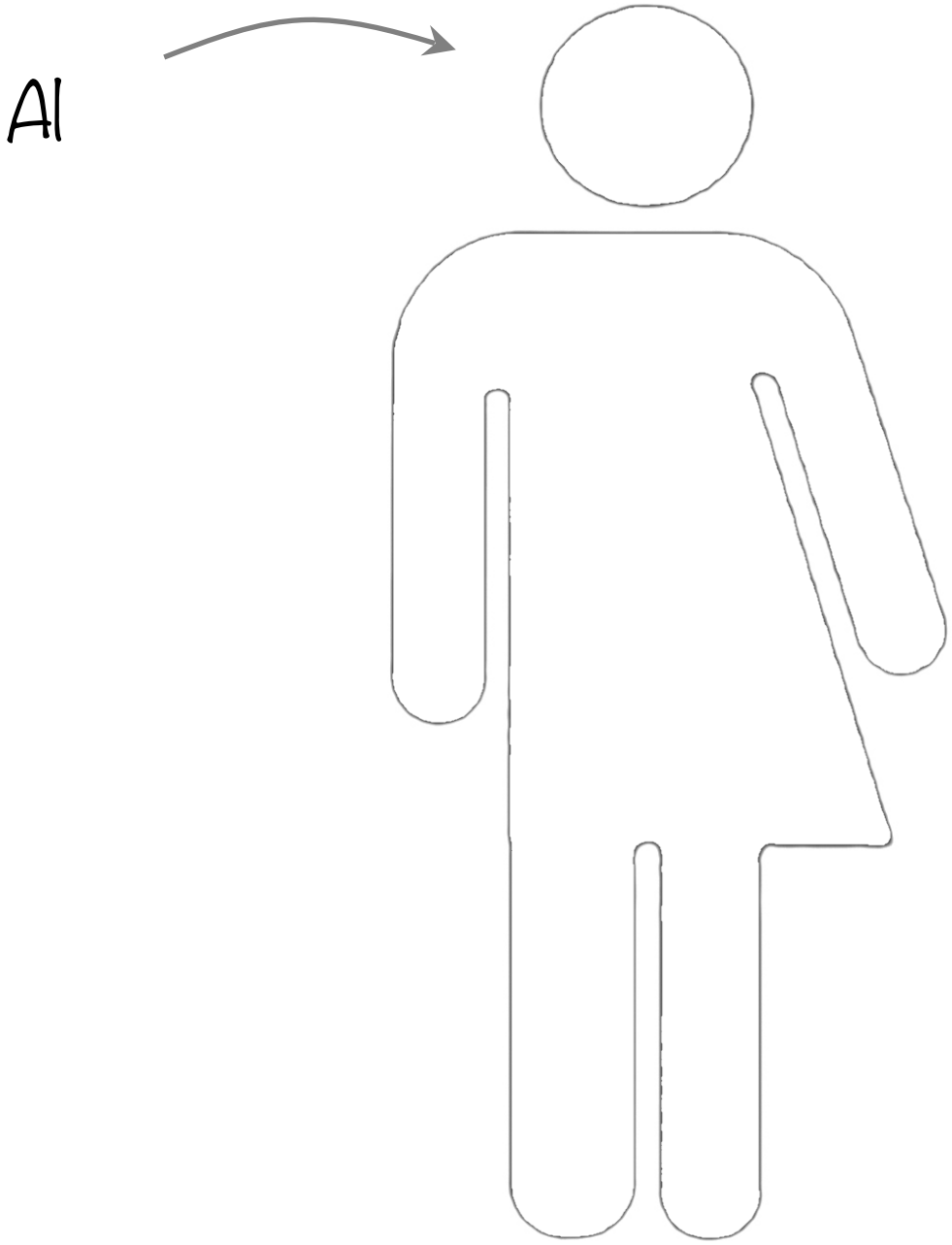
2. Gain independence in applying machine learning techniques (time-domain classifiers; CNNs)

3. Understand how we can interpret features of time-domain classifiers & CNNs, commonalities and differences; and understand how we can deal with bias

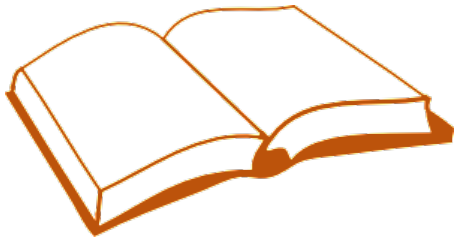
4. Come up with your own questions and implement in mini-projects to discuss with the group

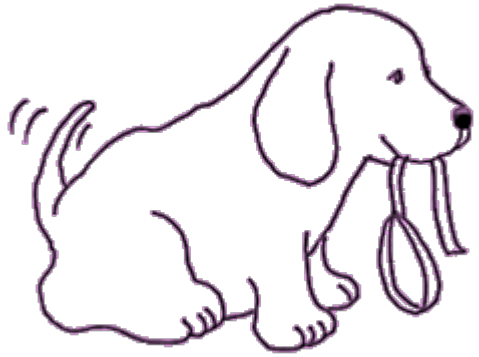
# What is neuroscience?

# What is neuroscience?

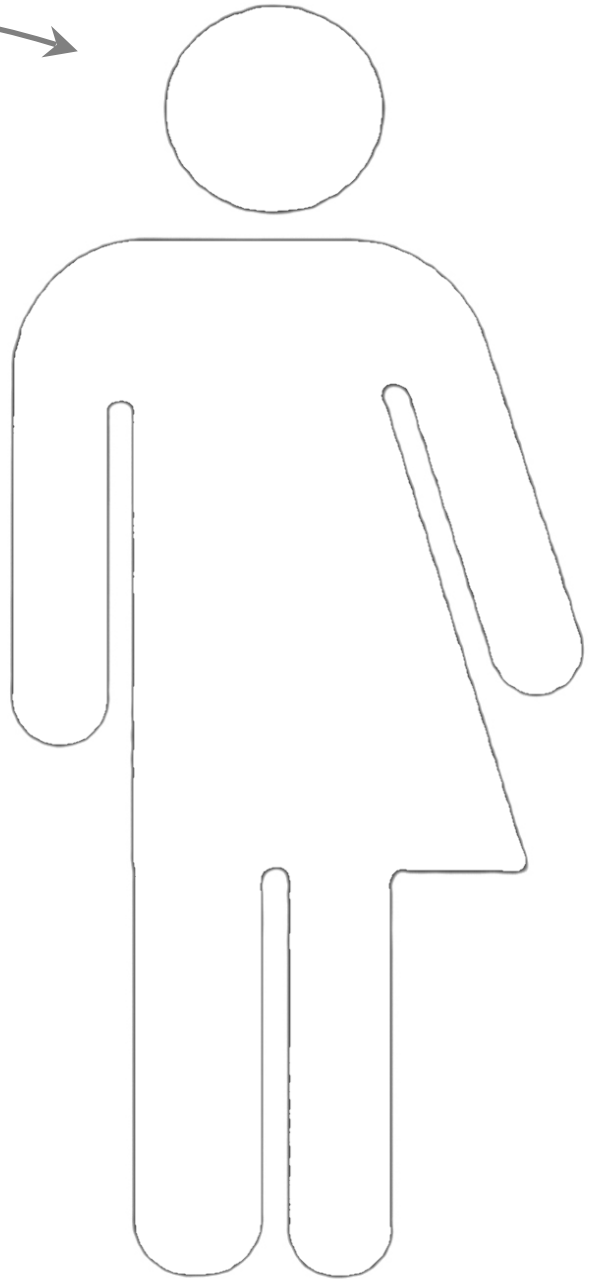


# What is neuroscience?

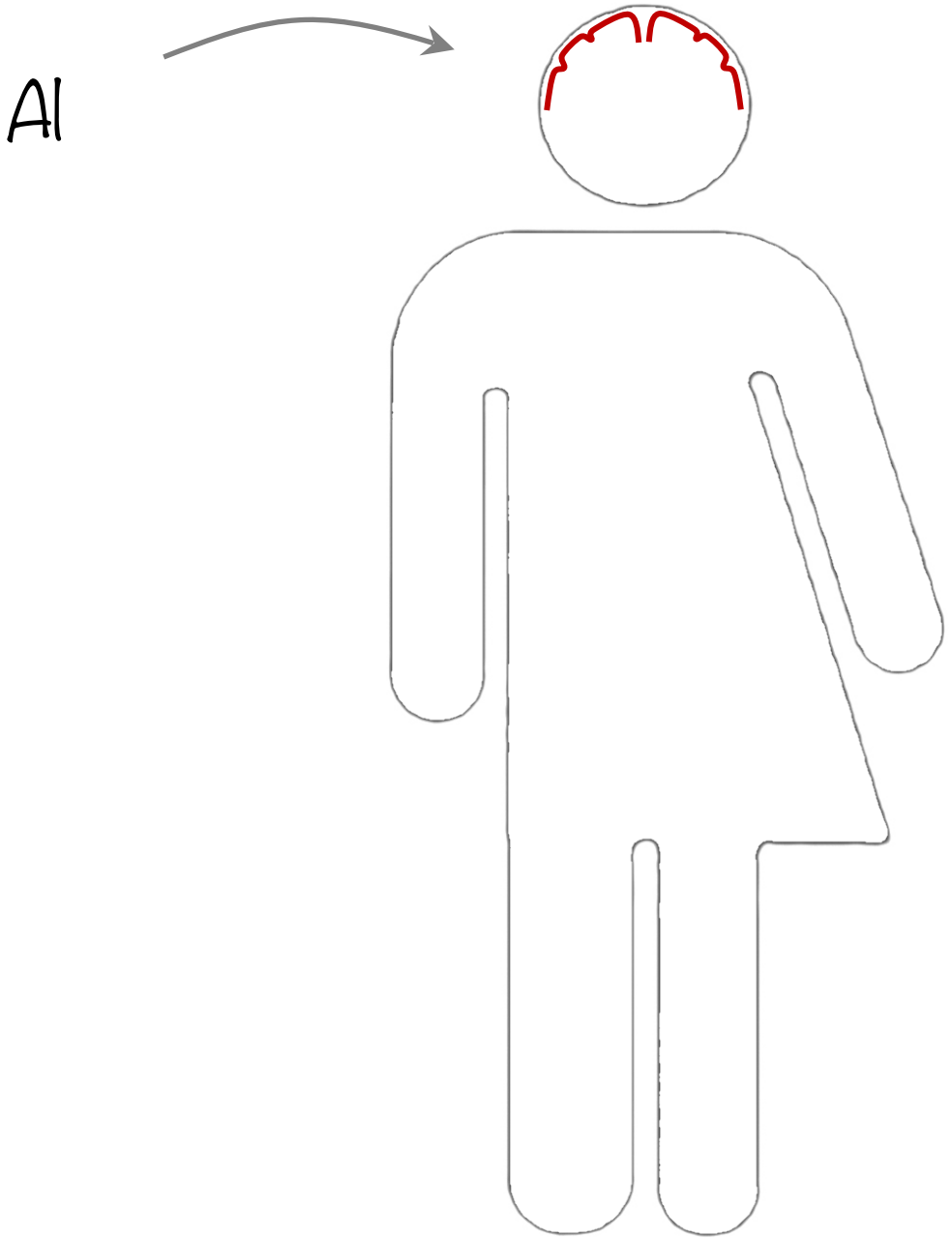




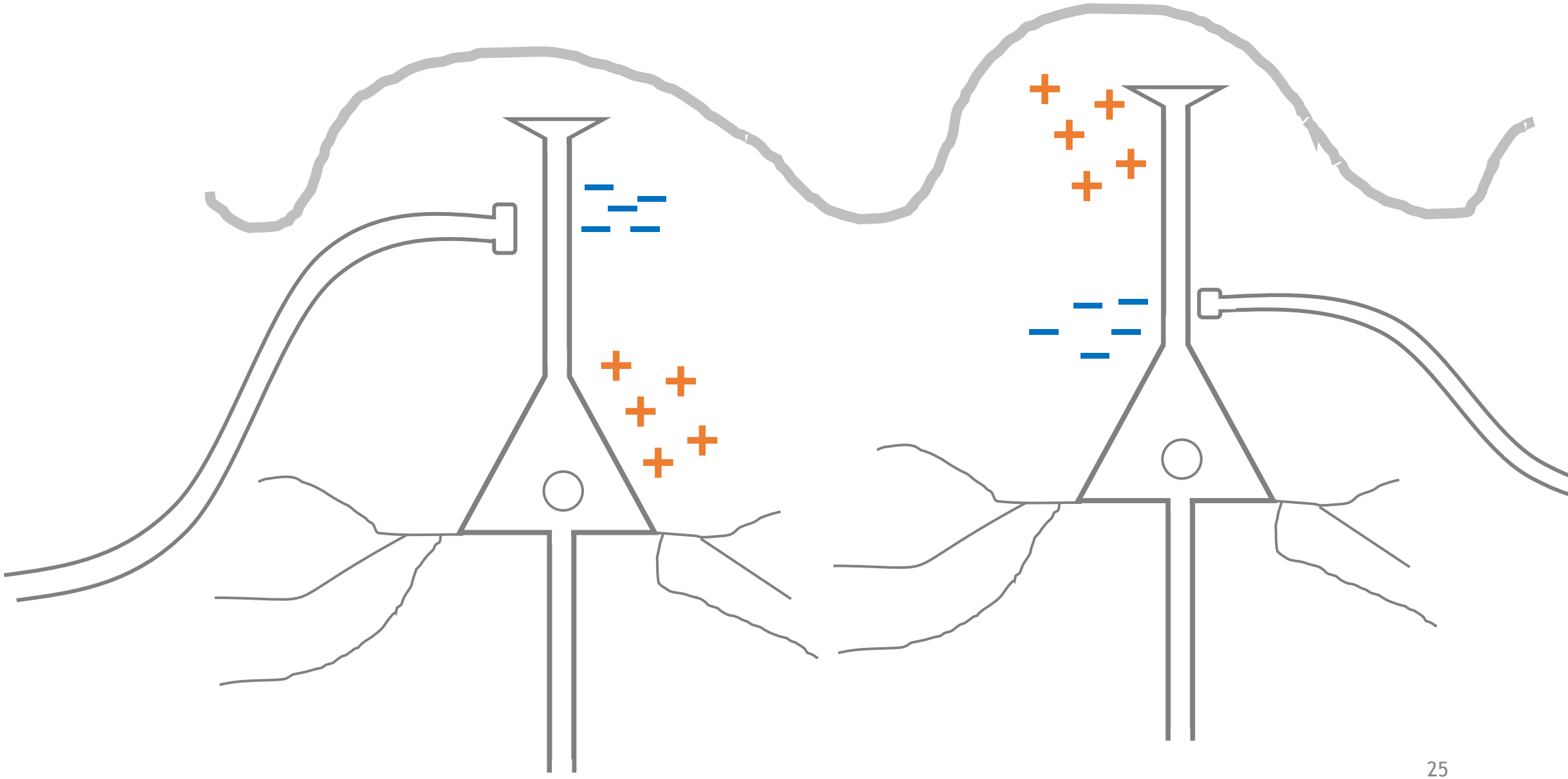
AI



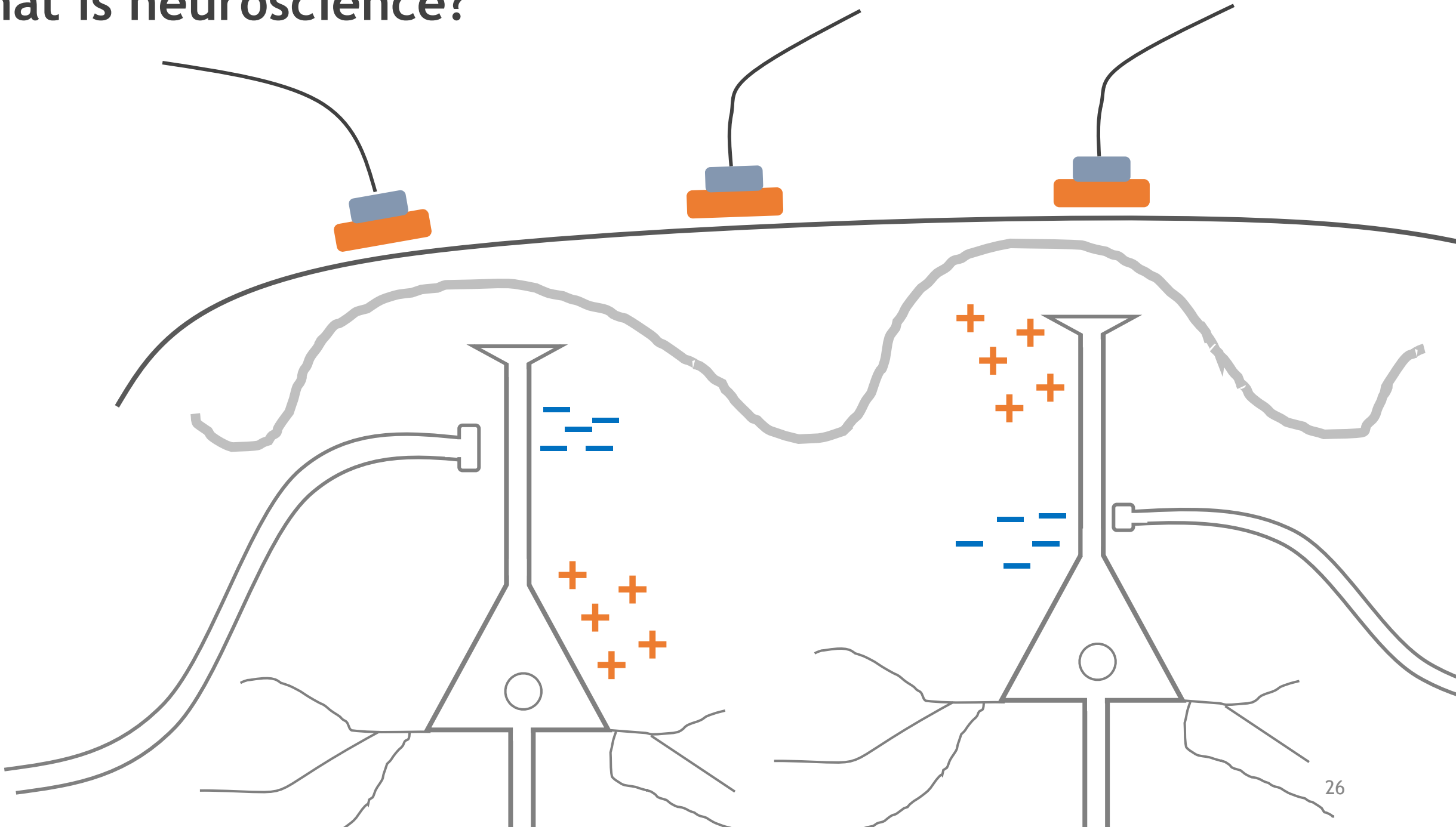
# What is neuroscience?



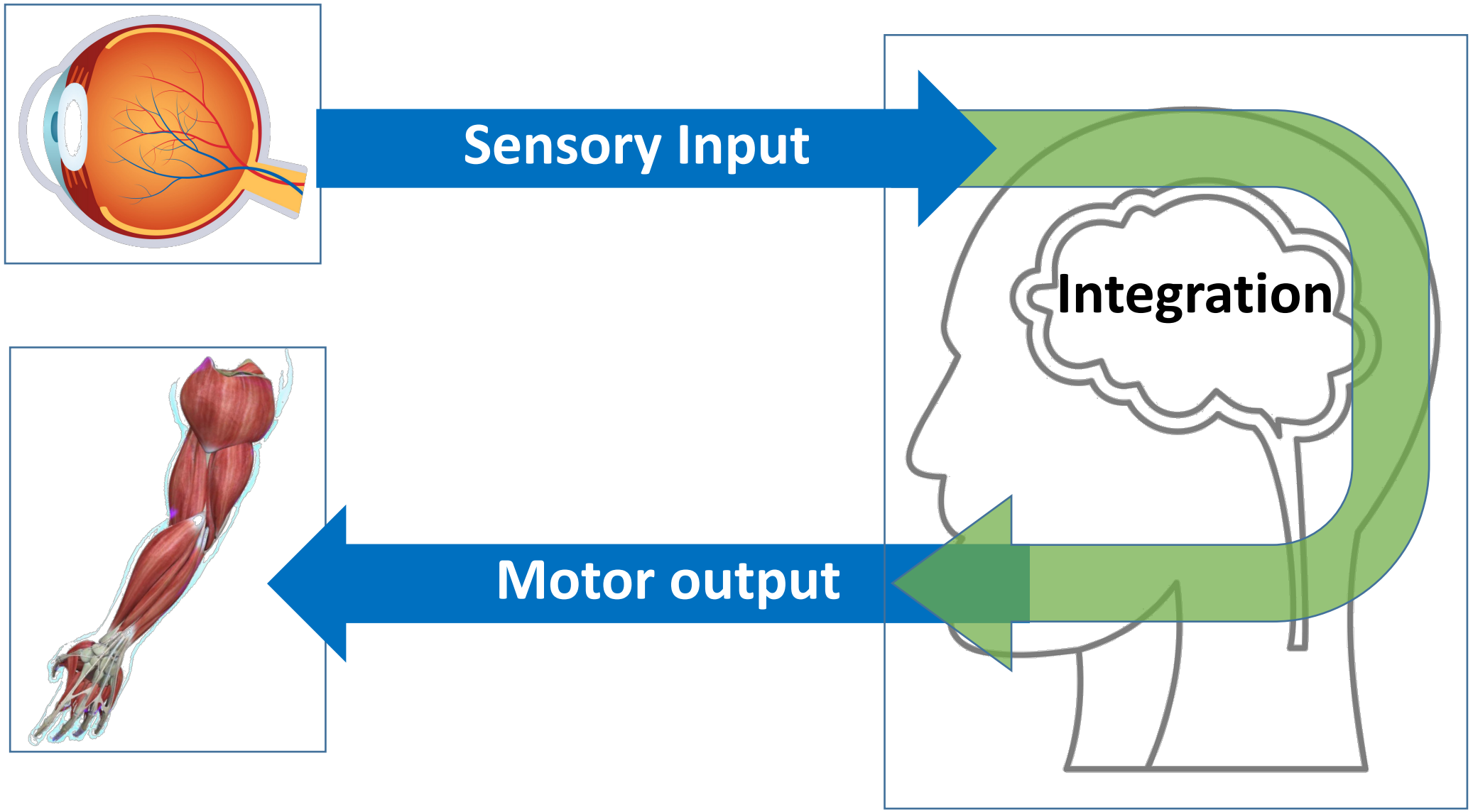
# What is neuroscience?



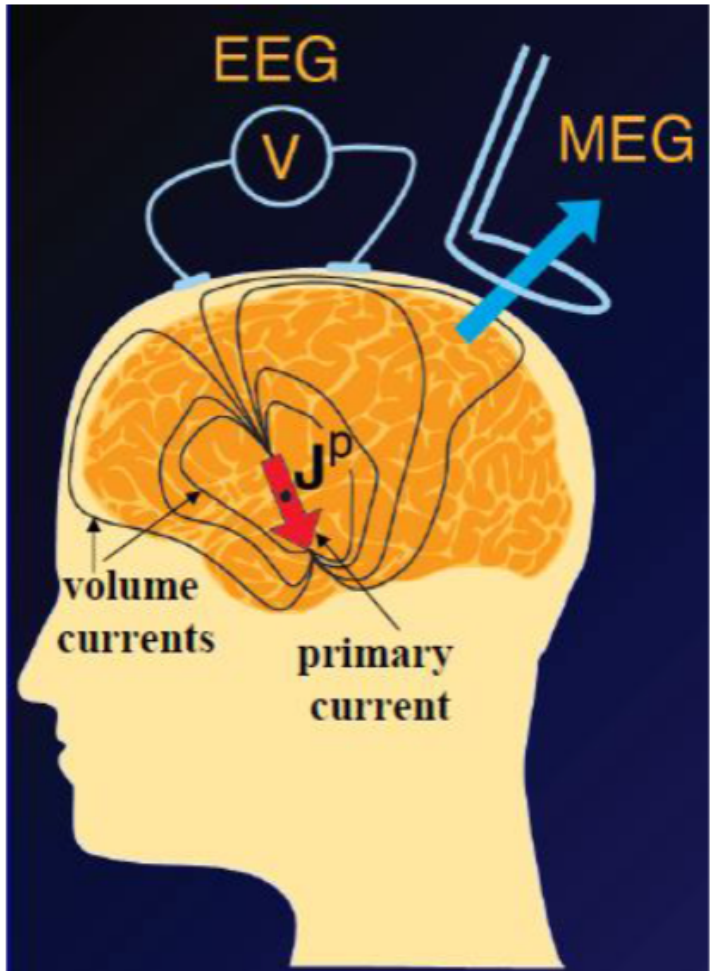
# What is neuroscience?



# The brain as a black box



# Tool to investigate neural activity of the brain: electroencephalography -EEG-




A dipole generates current which is propagated in the conducting medium of the head.

This reaches the head surface and can be seen as:

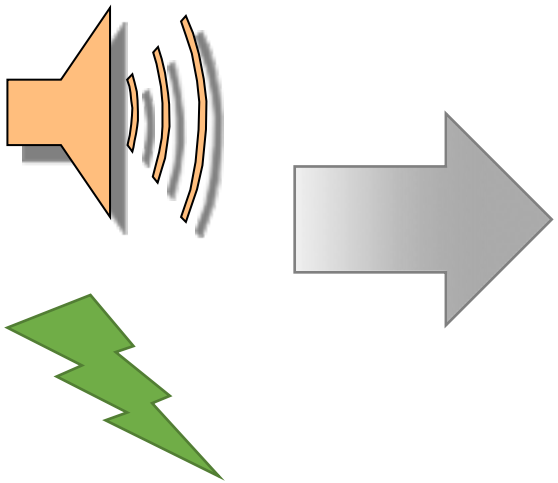
A. Electric current

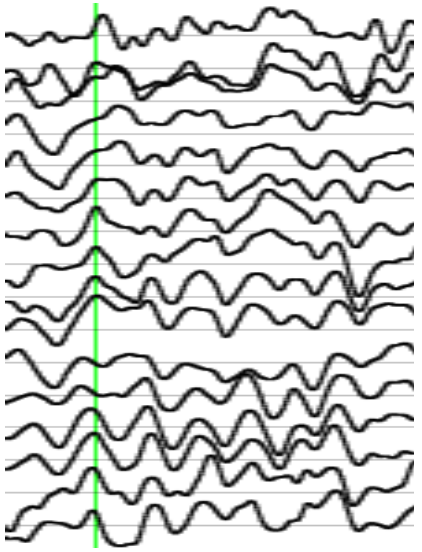
B. Magnetic field

# The recipe for measuring EEG

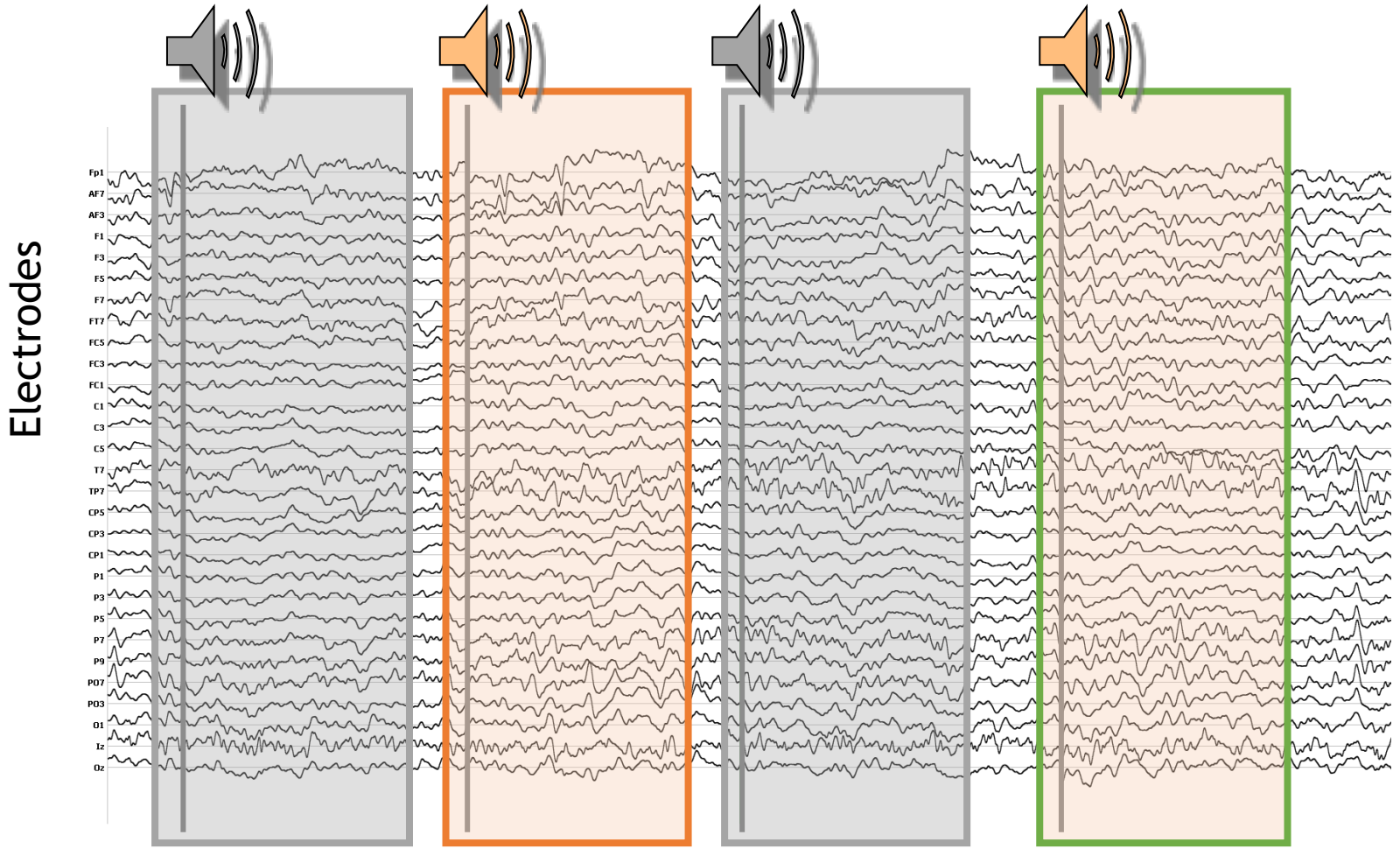
Good friend

# The recipe for measuring EEG

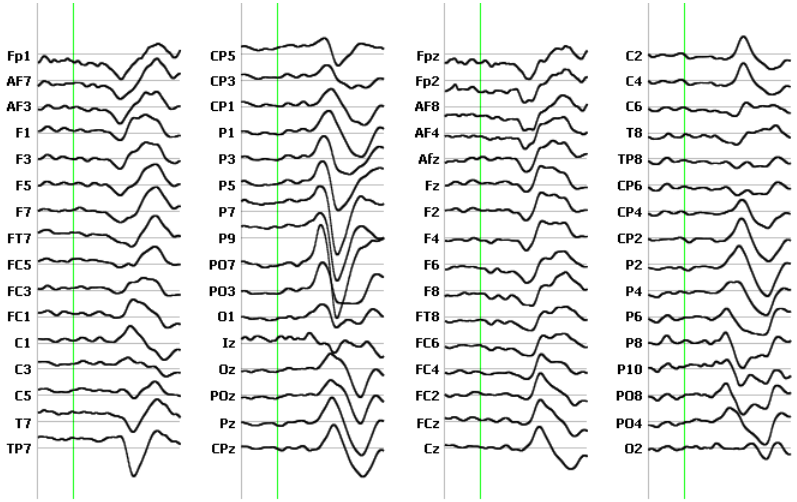




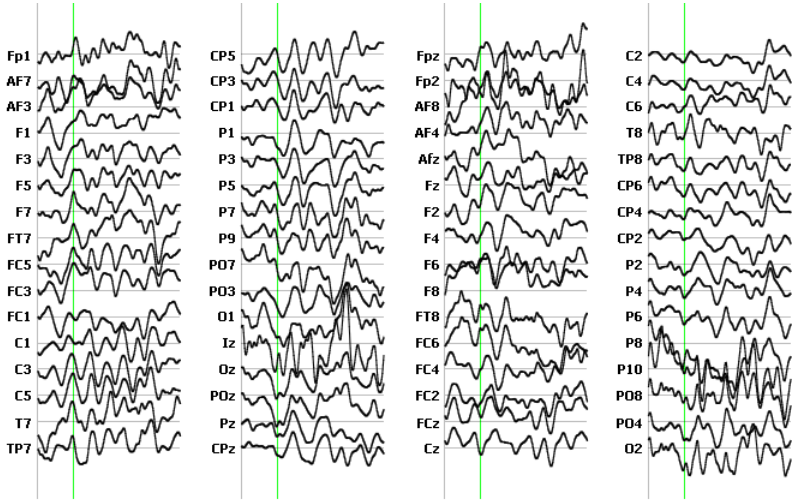
# Event-related potentials



# Event-related potentials

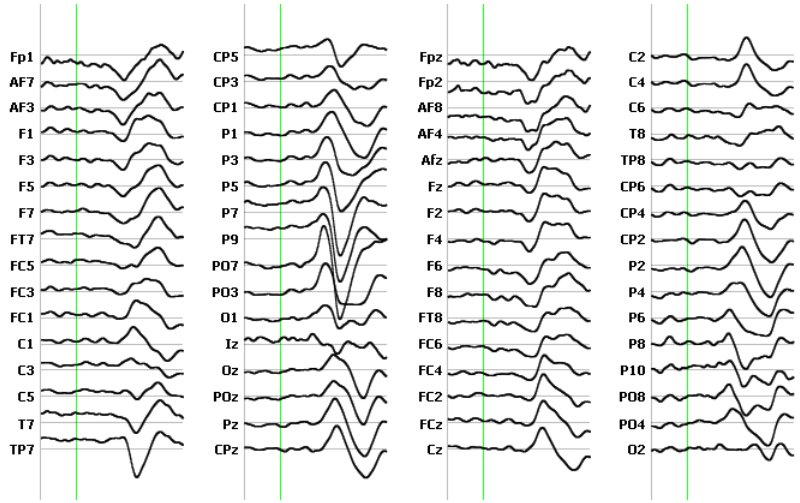


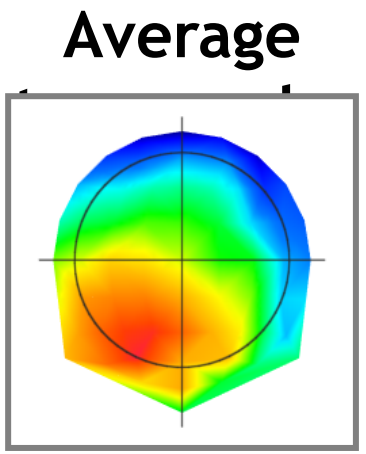

Average ERP

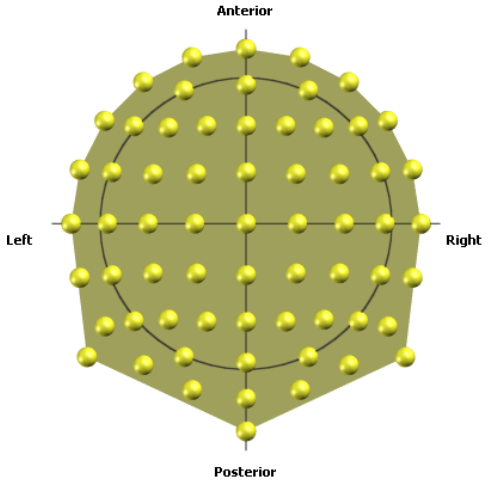



Single-trial ERP

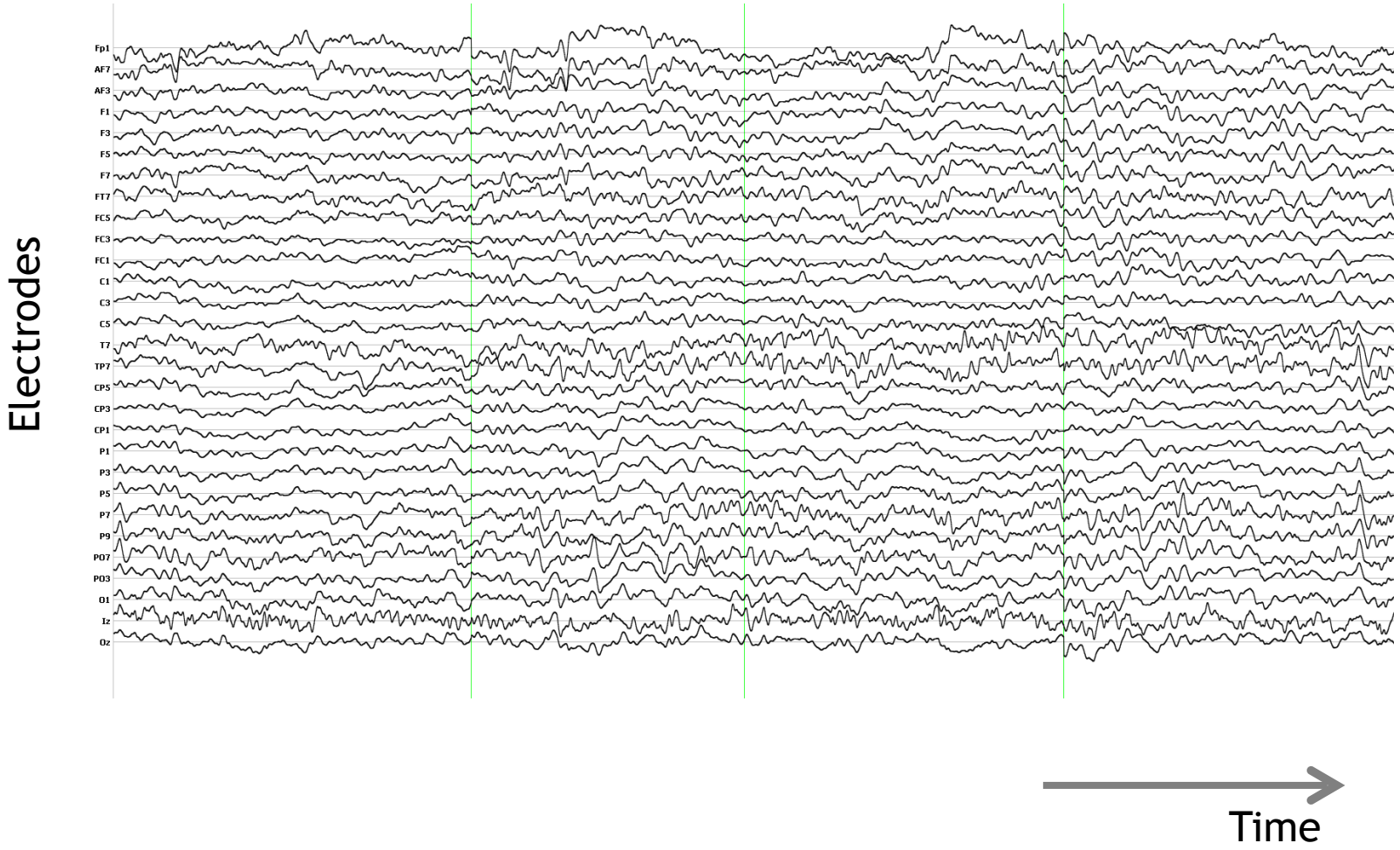
# Event-related potentials



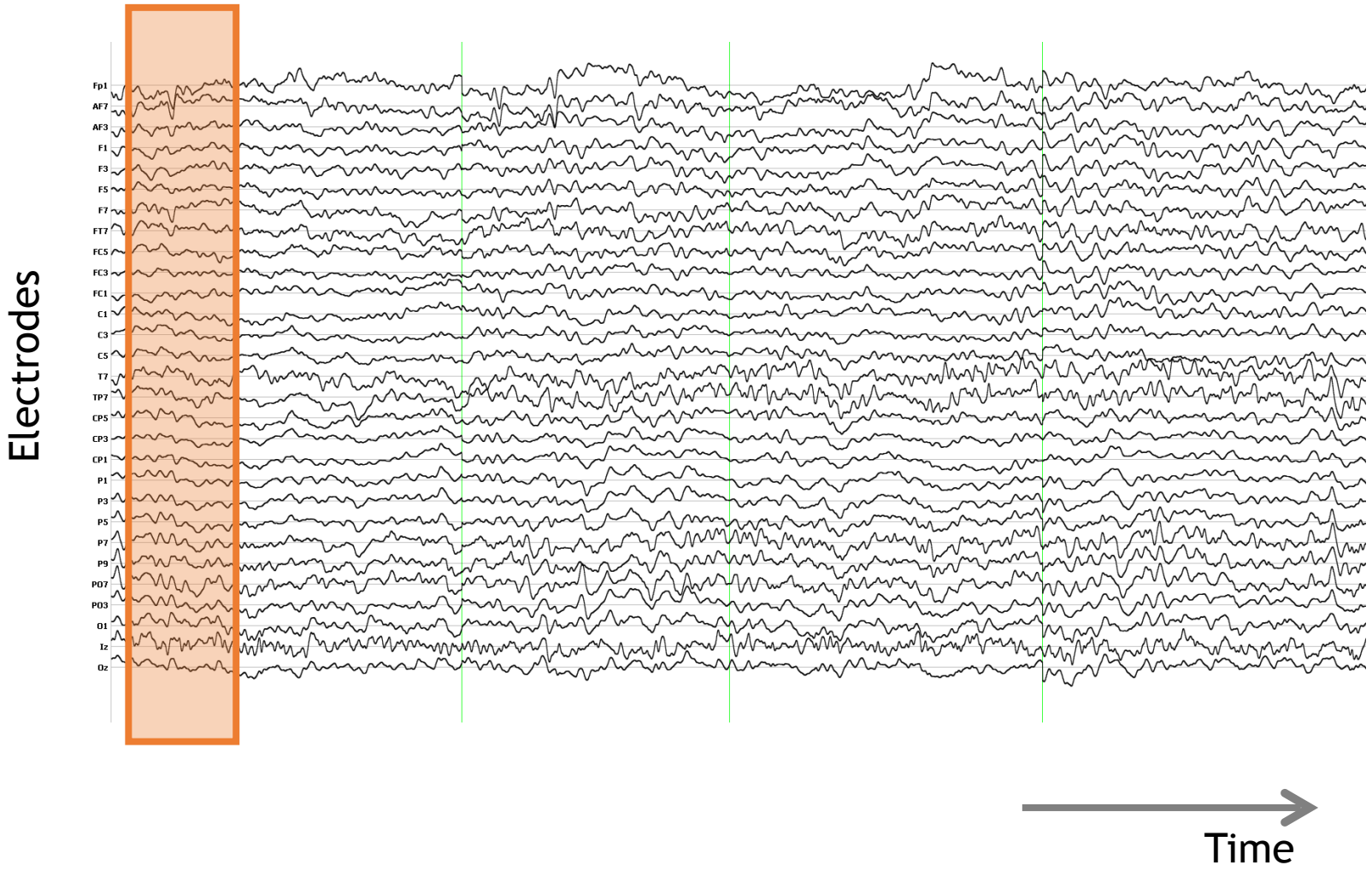



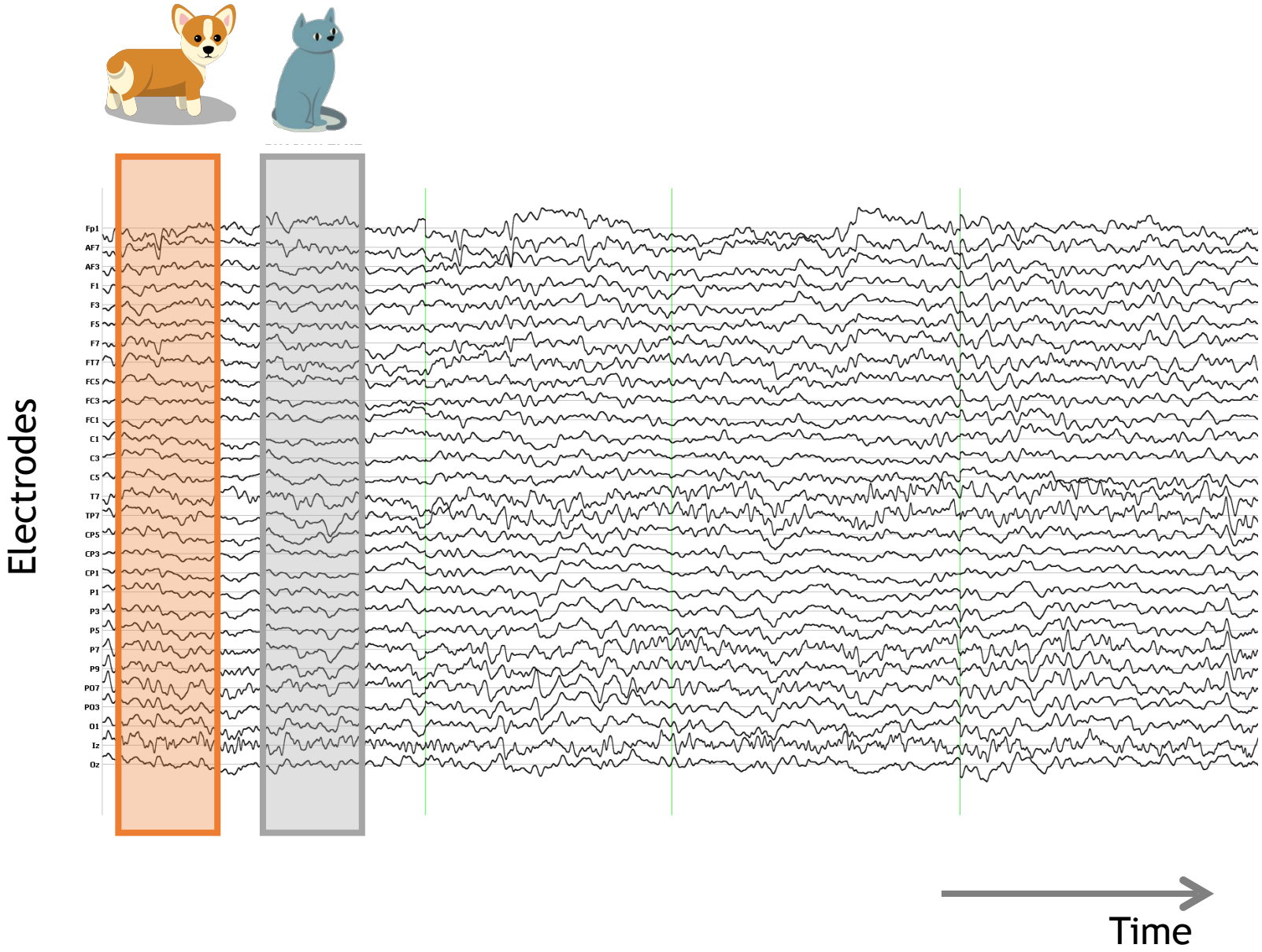
# Analyzing EEG responses with Machine Learning -ML-



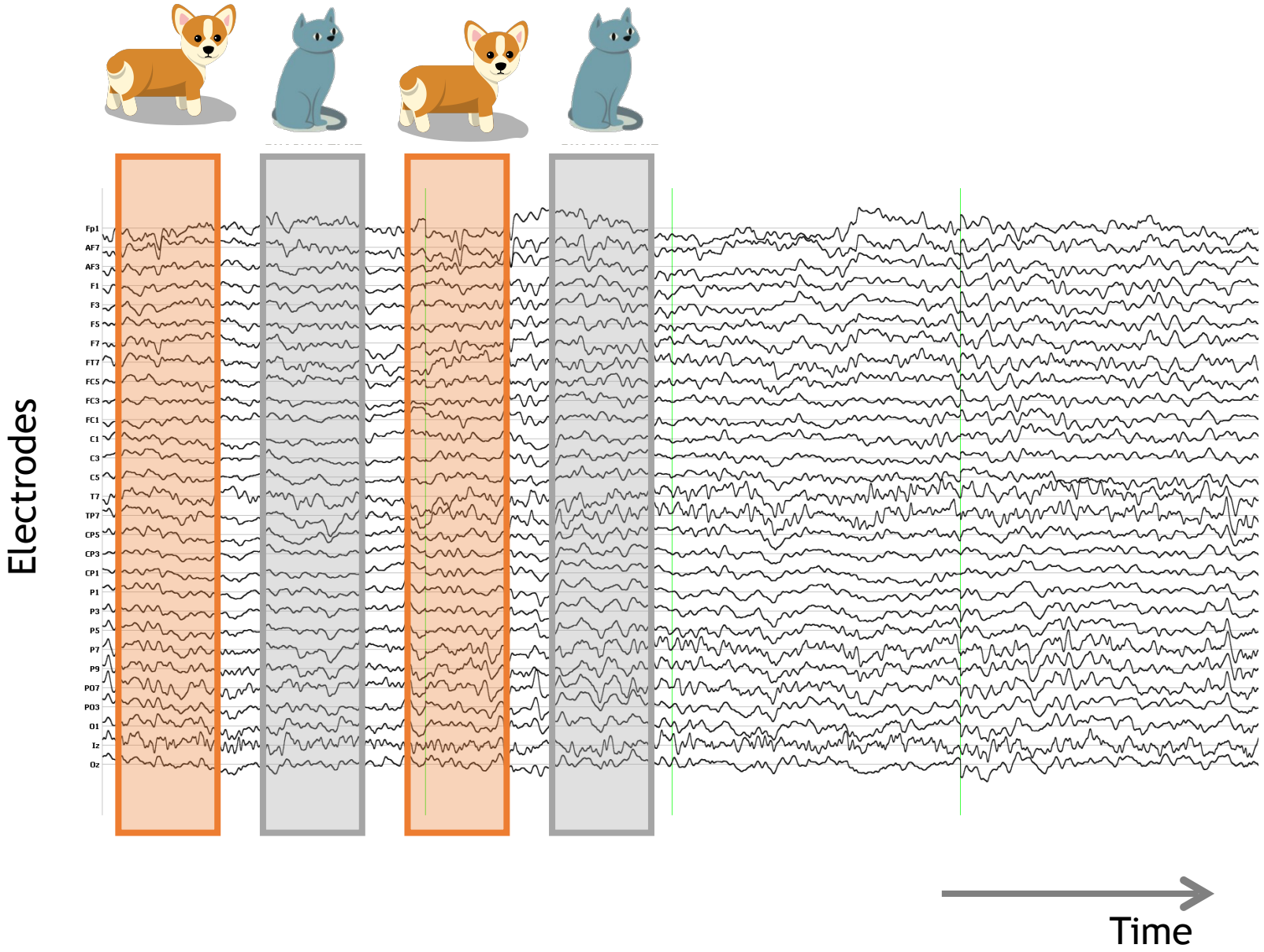
# Analyzing EEG responses with Machine Learning -ML-



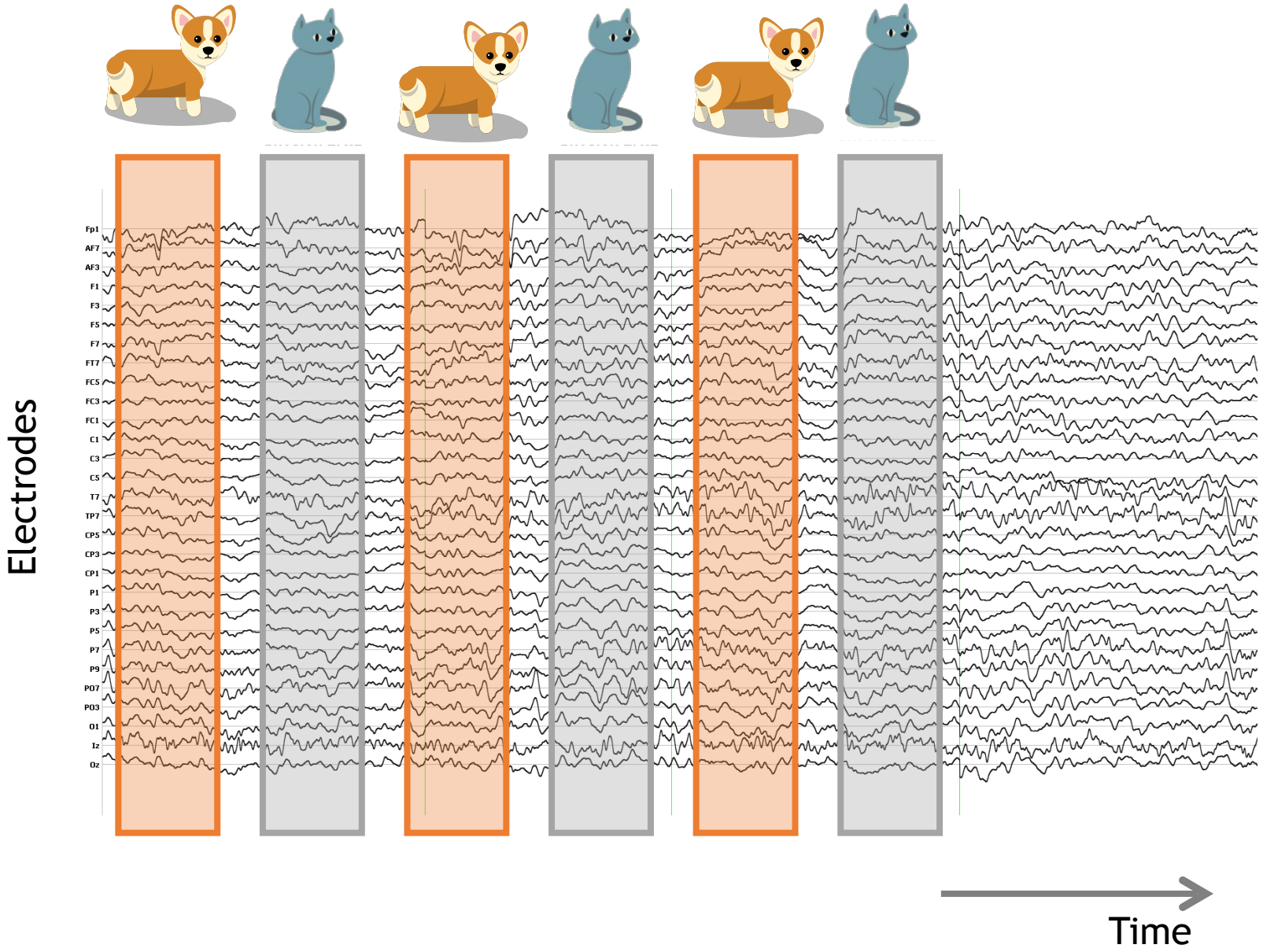
# Analyzing EEG responses with Machine Learning -ML-



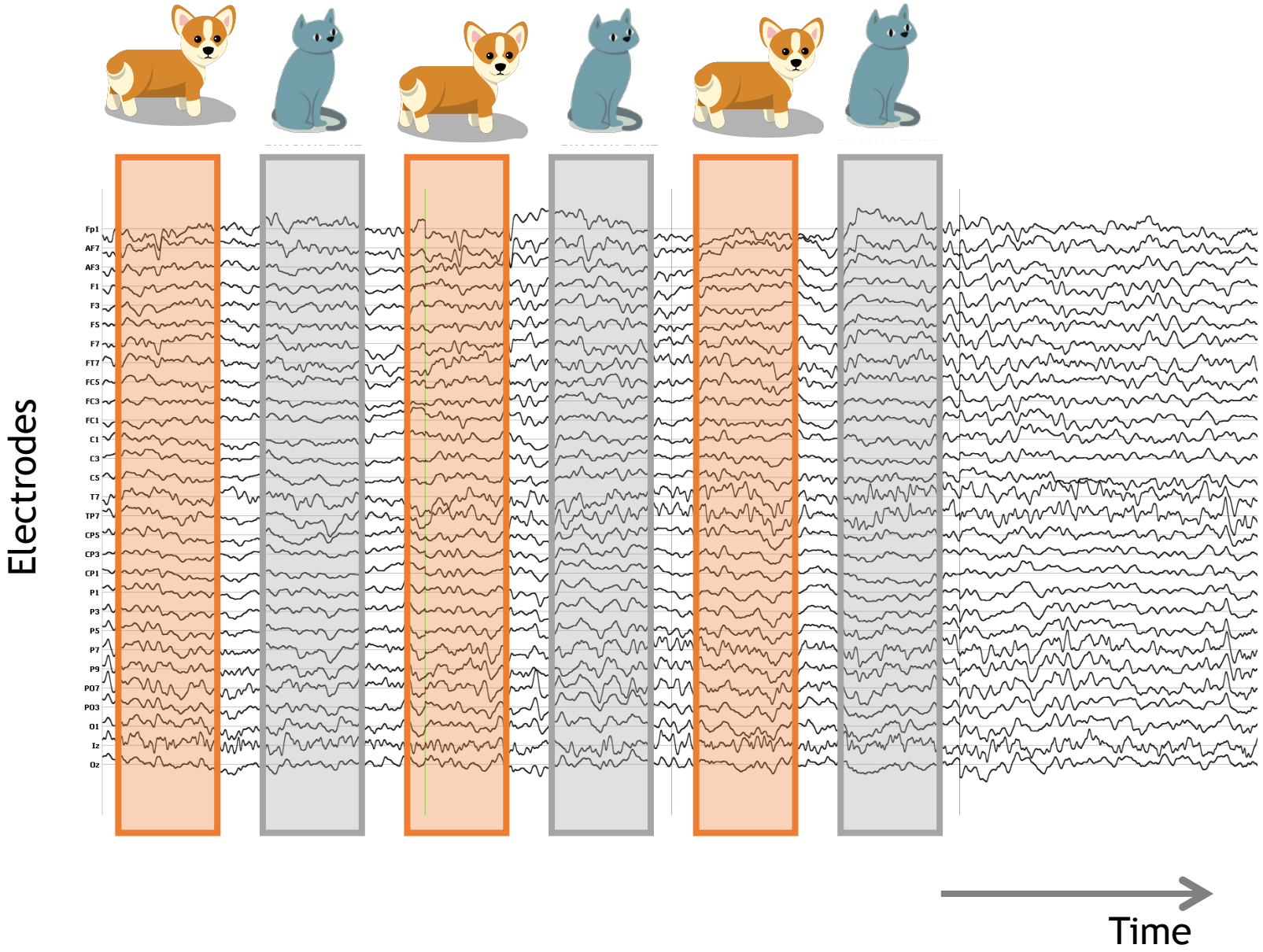
# Analyzing EEG responses with Machine Learning -ML-



# Analyzing EEG responses with Machine Learning -ML-



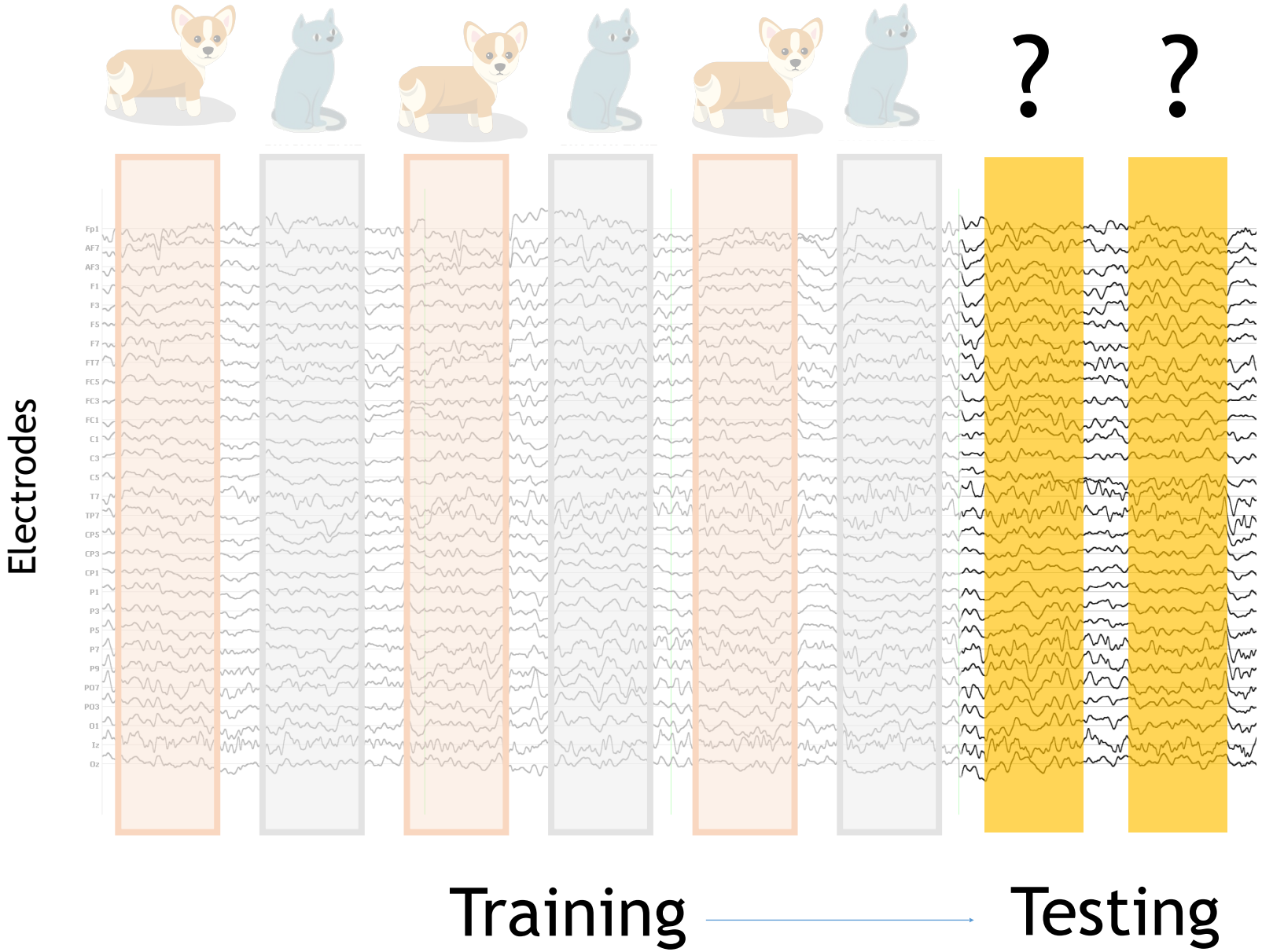
# In other words…




**We focus on the EEG responses we have already observed.**

**We build a model that learns patterns of responses to corgi vs. cats**

# In other words…

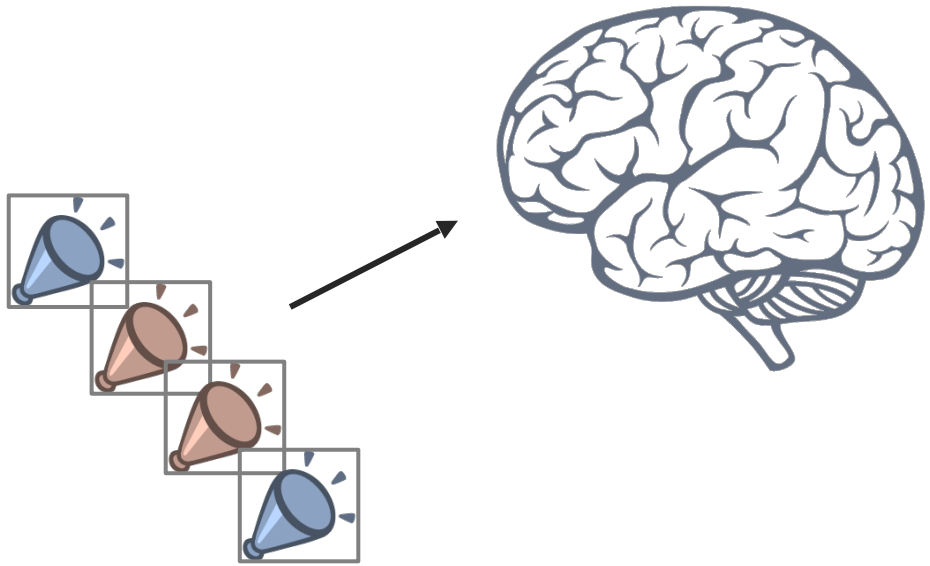



**Can we now guess, based on the model we built, whether participants were looking at corgi or cats?**

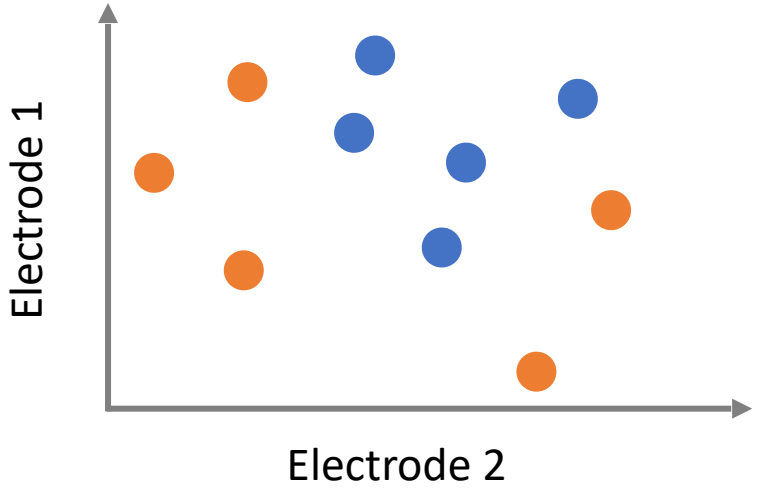
**If yes: The neural patterns of responses to corgi vs. cats were different!**

# The recipe for applying ML on EEG signals

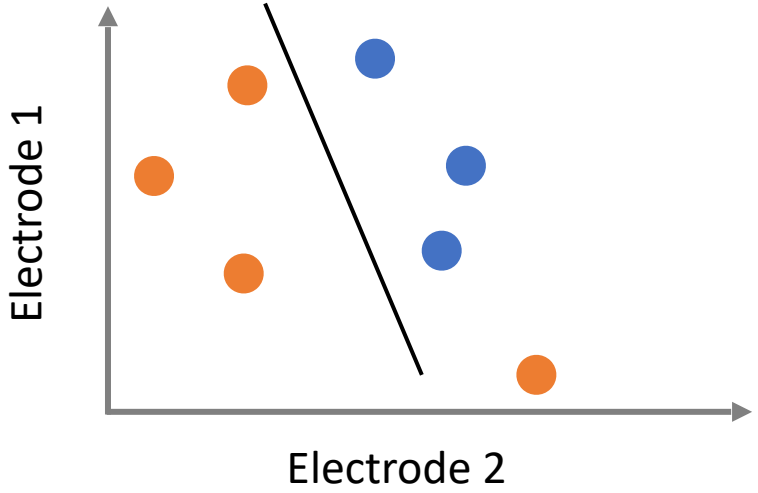
## 1. Record data



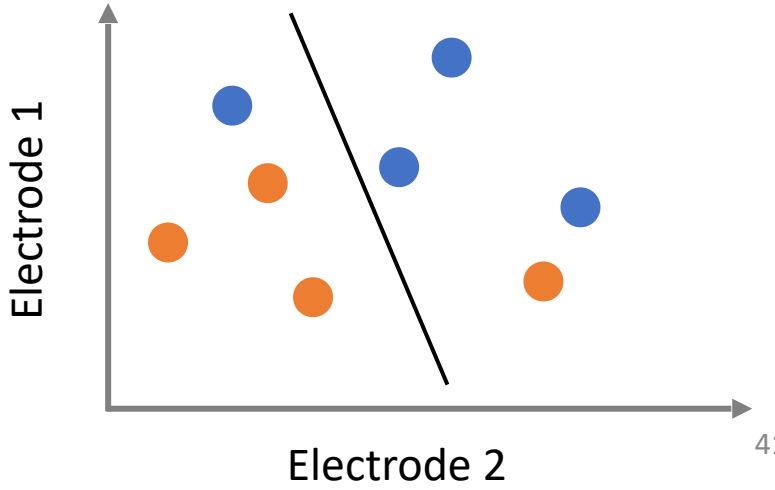
## 2. Extract patterns of activity




## 3. Train classifier on a subset




## 4. Test classifier on new data

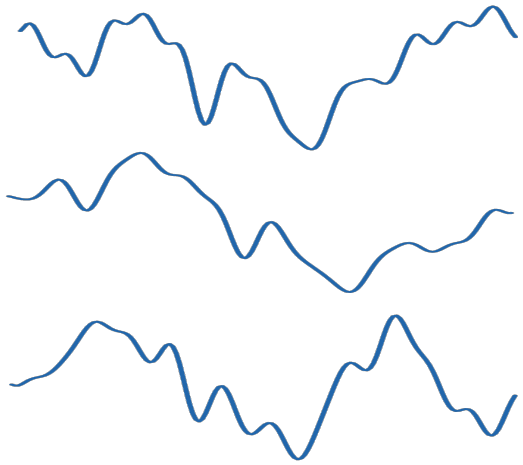


# The recipe for applying ML on EEG signals

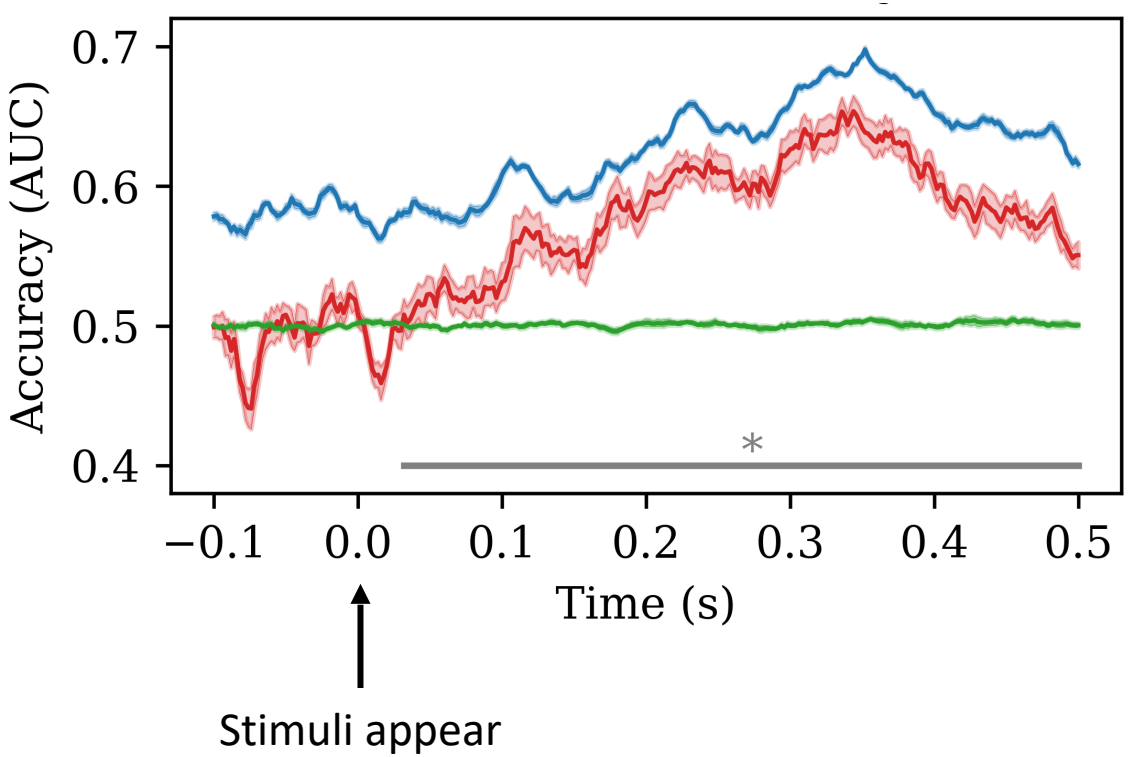
**External stimuli**

**ML Models**

**EEG responses**



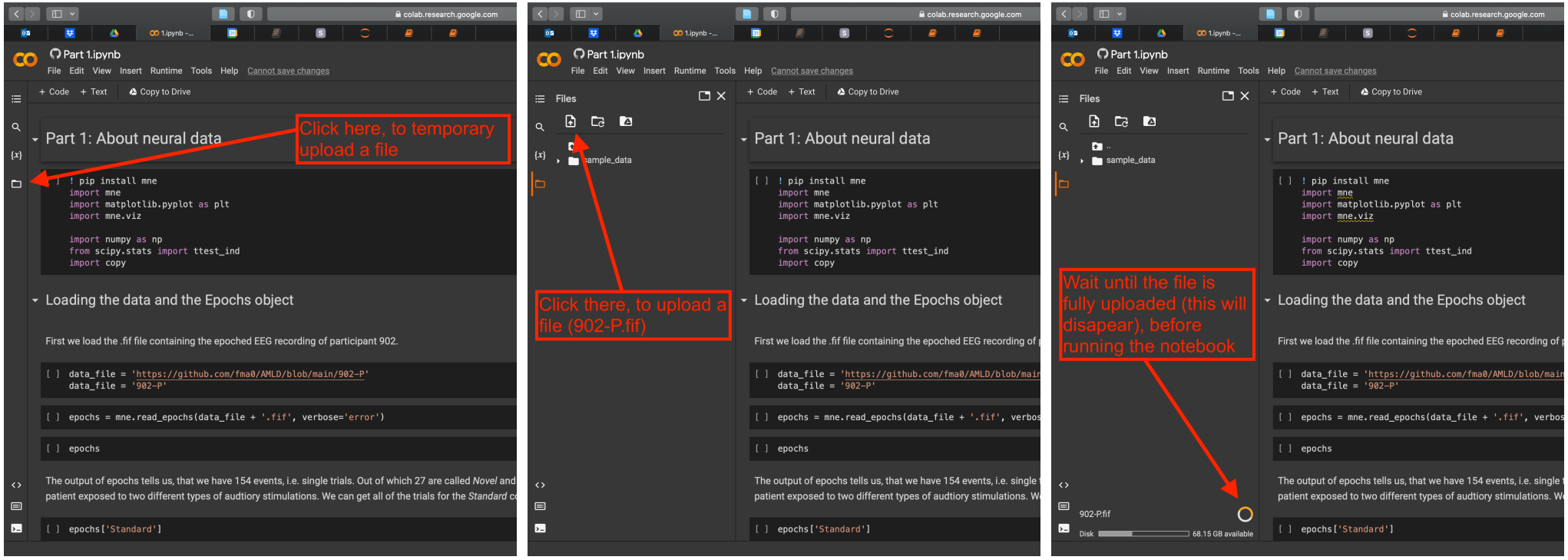
**Classification performance**



# Hands ON Tutorial 1

GitHub repository, with exercises: https://github.com/fma0/AMLD/

Run in Google Colab: https://colab.research.google.com
(Make sure to save the notebooks in your google drive)



# Overview for today

Introduction to AI in neuroscience :

- Electroencephalography (EEG) signals
- Hands-on: working with EEG

Machine Learning in neuroscience

- Supervised learning: training classifiers
- Measuring performance
- Hands-on: Classifying EEG data

Convolutional Neural networks for EEG signals

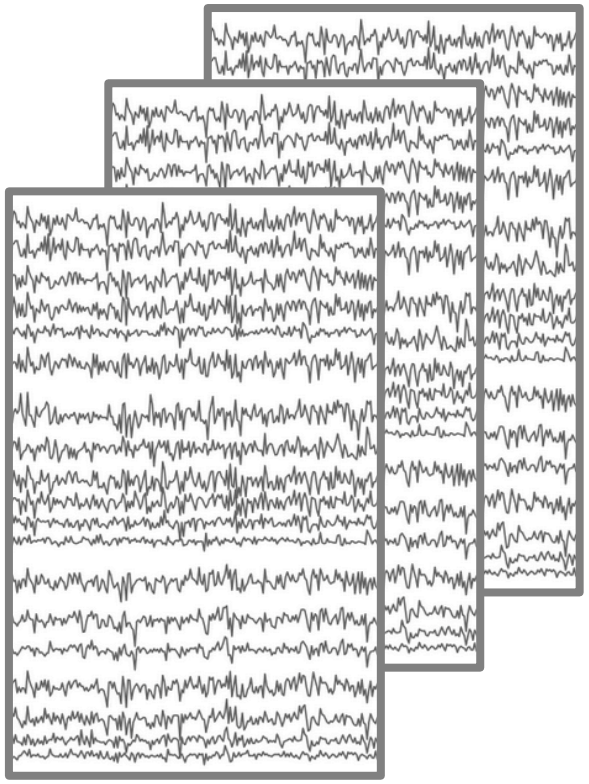
- Training networks & measuring performance
- Hands-on: working with neural networks

Group work & presentations:

- Mini projects: try out what we learned in short projects & your own ideas

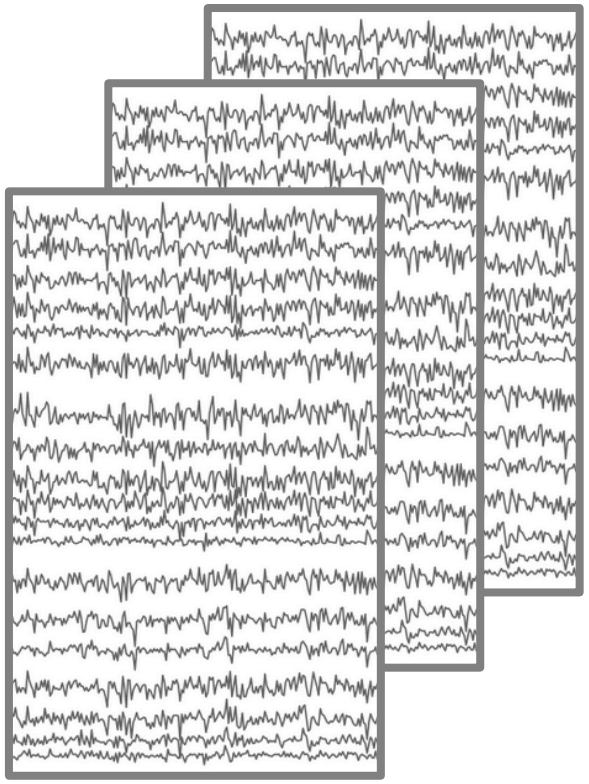
# Now that we have the data:

**EEG data**

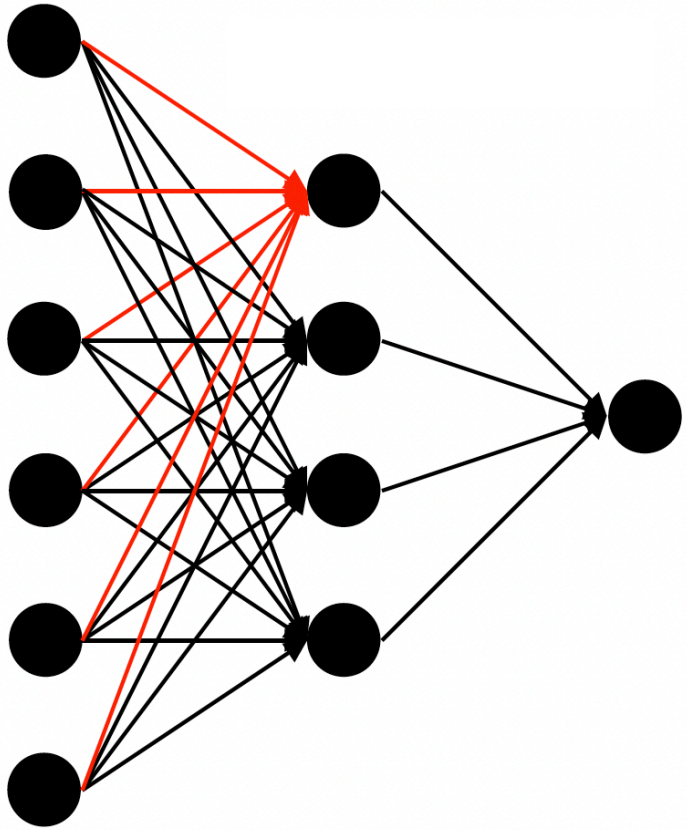


# Now that we have the data

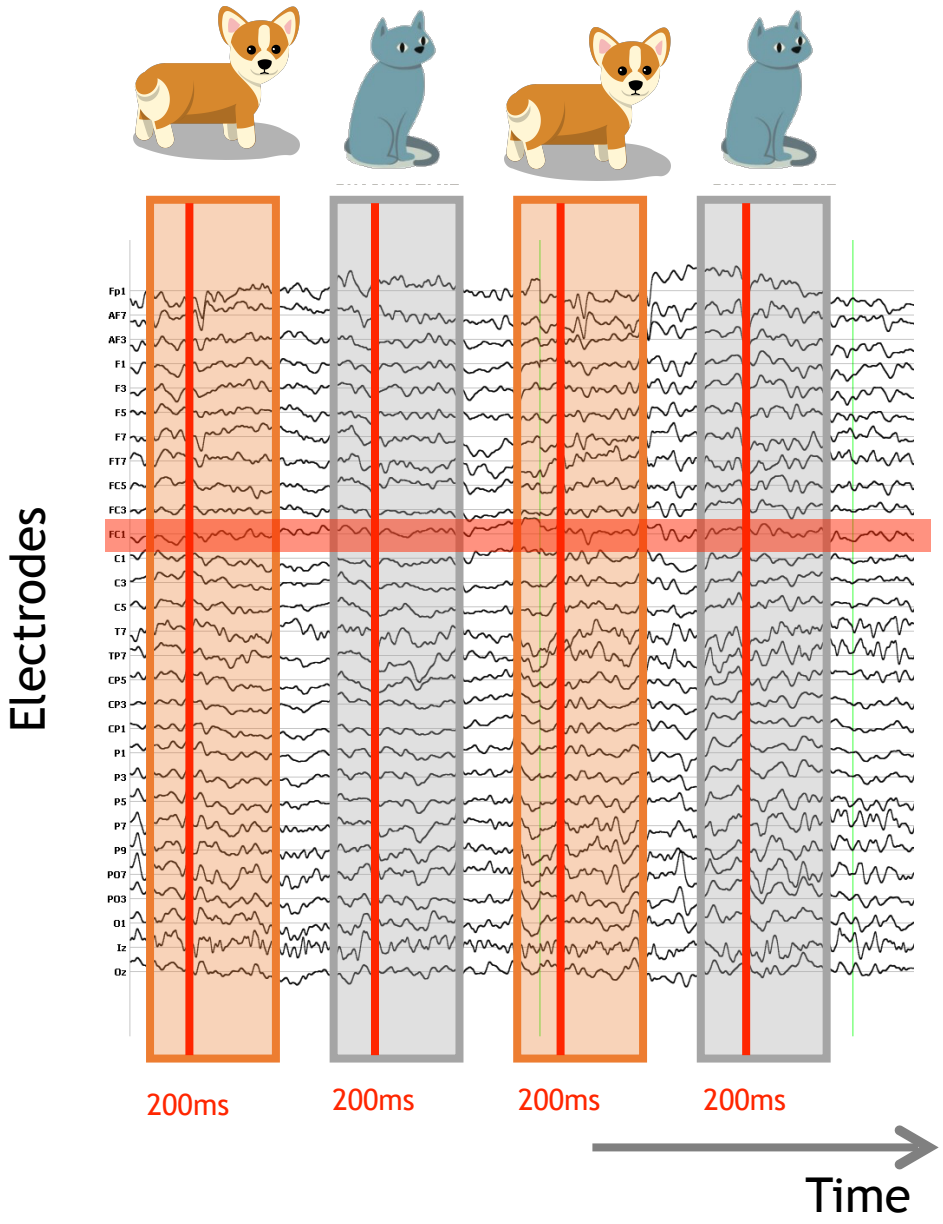
## EEG data



## Machine Learning

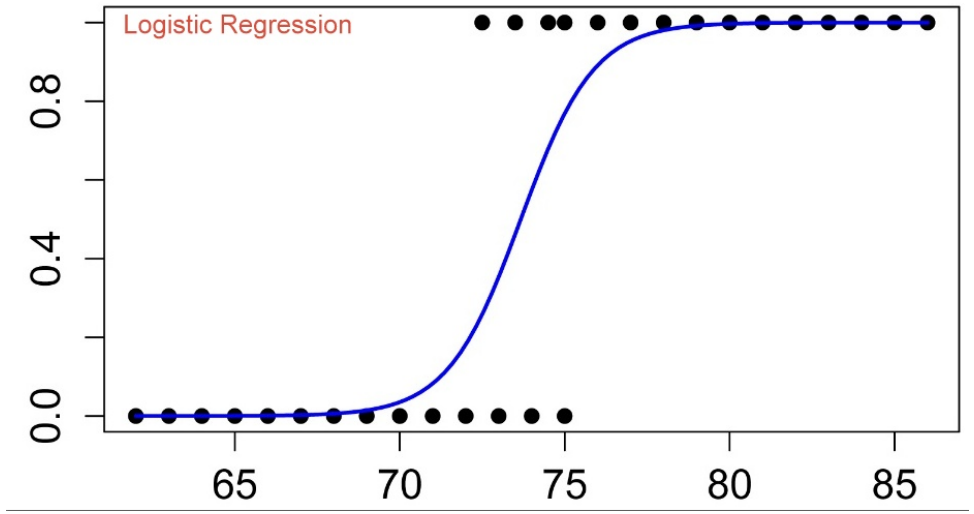


# Logistic Regression 1 dimensional



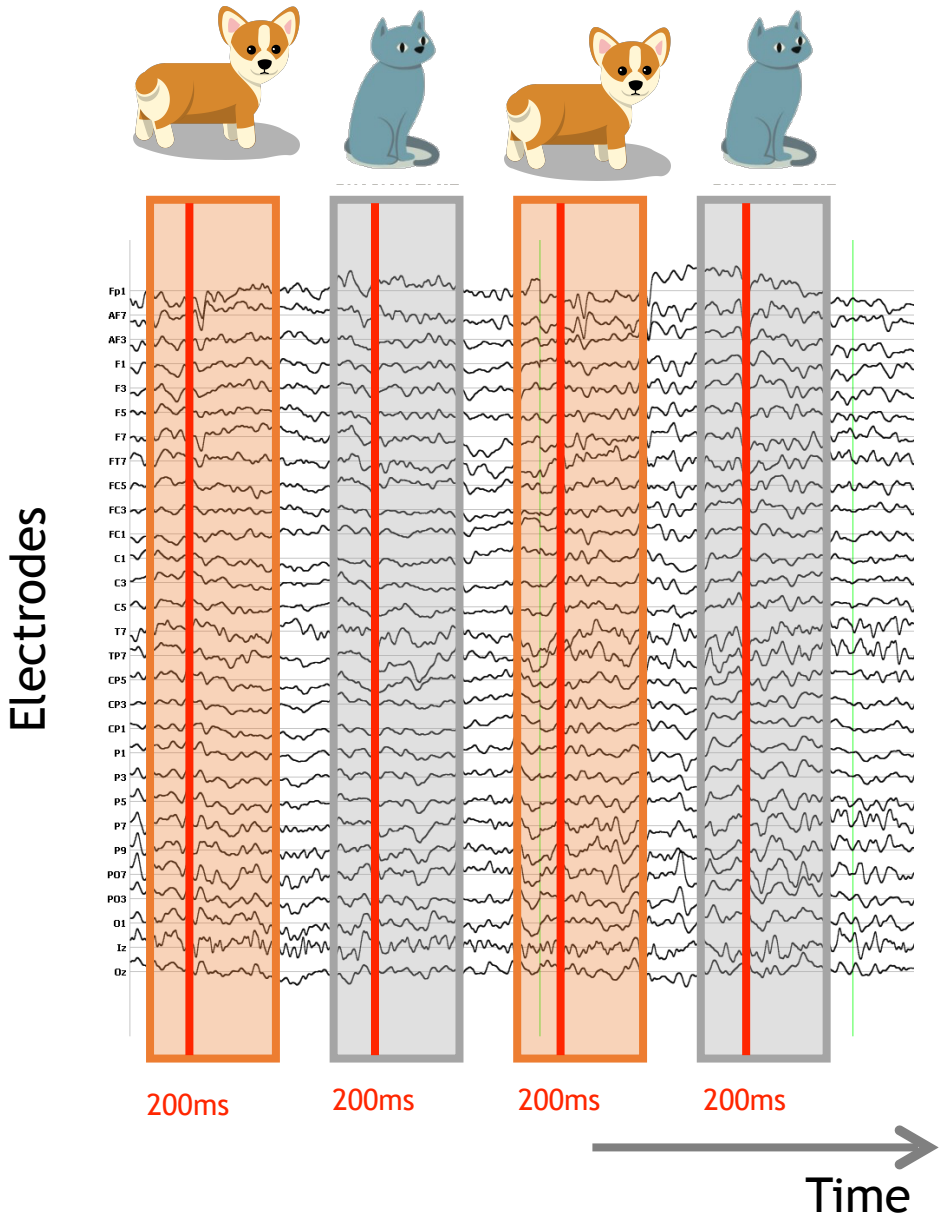

Example: Predicting from activation of FC1 at 200ms for Corgi vs. Cat

Mapping from continuous feature to binary values




https://medium.com/@cmukesh8688/logistic-regression-sigmoid-function-and-threshold-b37b82a4cd79

# Logistic Regression N dimensional



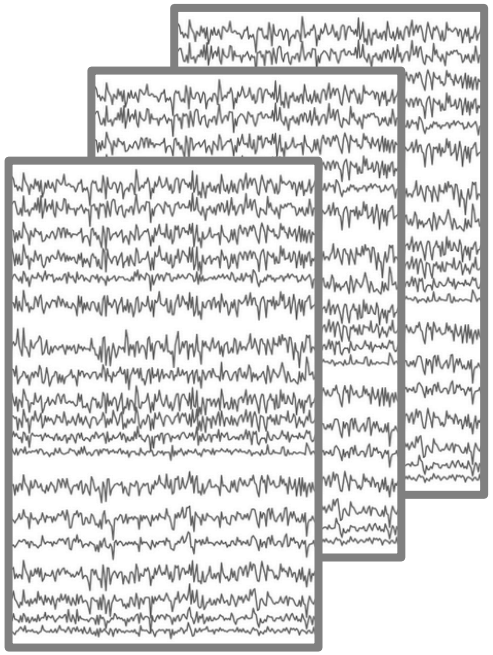

Example: Predicting from activation of all electrodes at 200ms for Corgi vs. Cat

| Input | | Weight | | | | |
|---|---|---|---|---|---|---|
| 0.5 | x | 0.1 | = | -1.23 | → Sigmoid → | 0.23 |
| -0.3 | | -0.5 | | | | |
| -1.5 | | 0.8 | | | | |
| -0.5 | | 0.4 | | | | |
| 0.1 | | -0.3 | | | | |
| -0.4 | | 0.6 | | | | |
| 1.2 | | 0.2 | | | | |

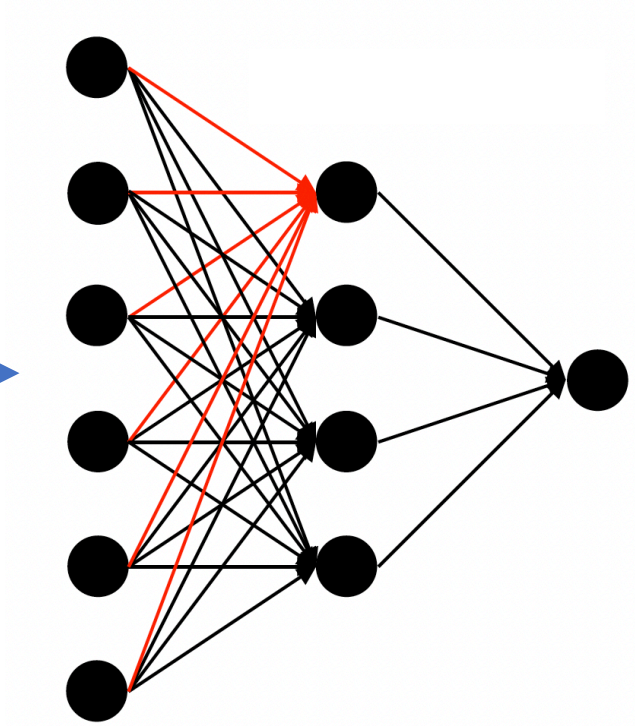
Prediction

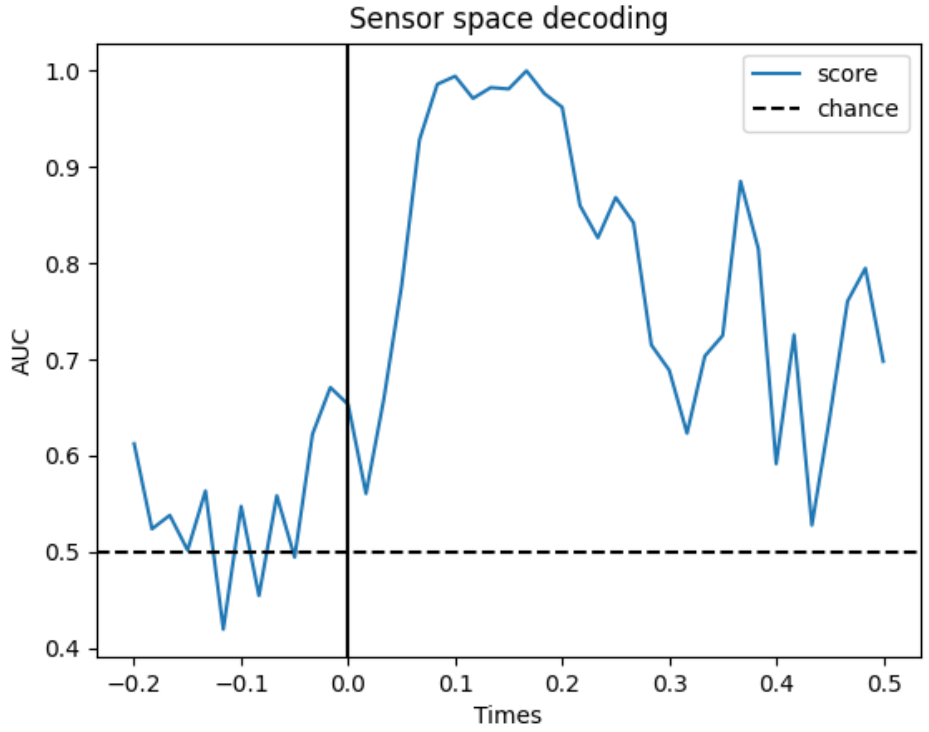
# The Performance Results

**EEG data**



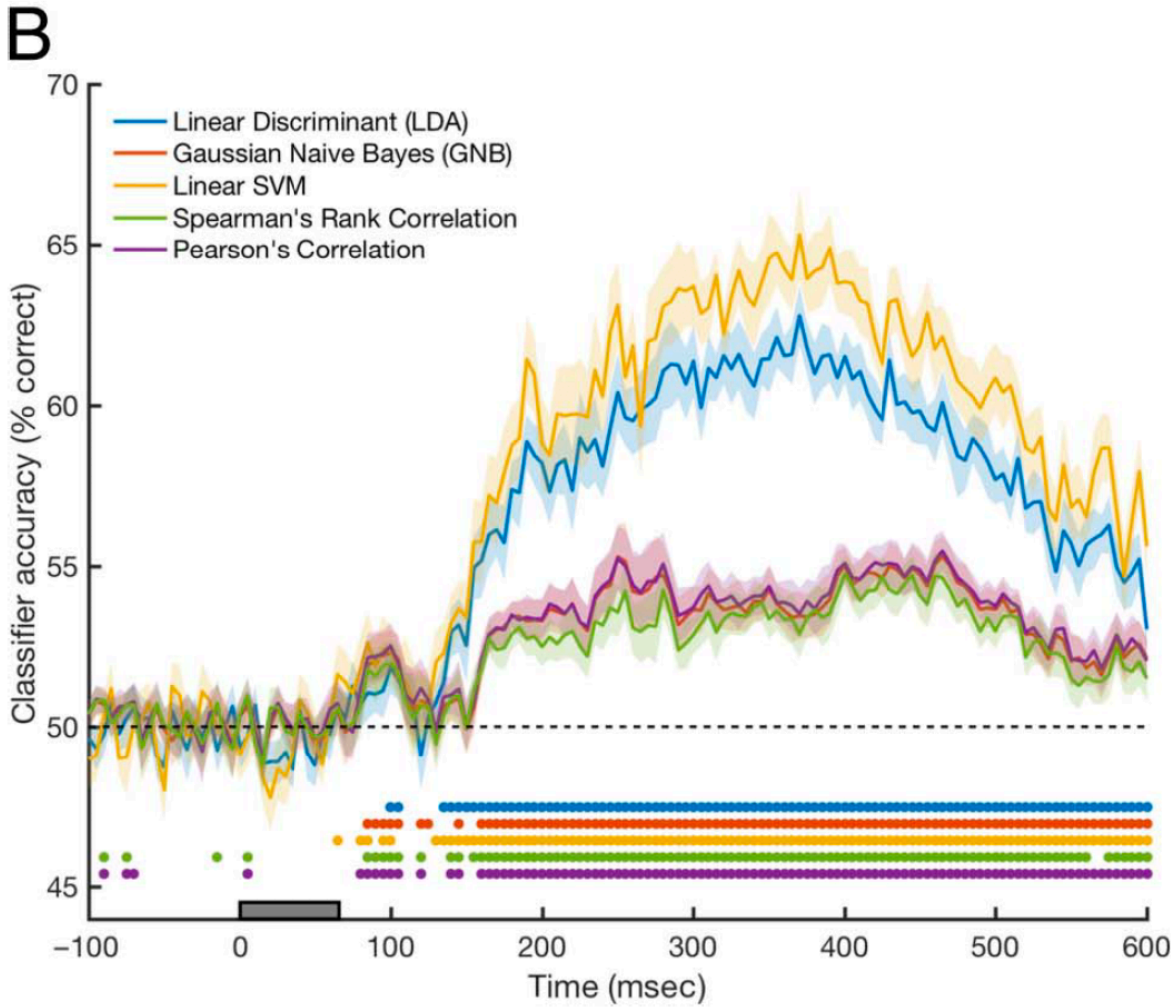
**Machine Learning**






https://mne.tools/stable/auto_tutorials/machine-learning/50_decoding.html

# Choosing a classifier




Grootswagers et al., 2017, J Cogn. Neuroscience

Classification performance is affected by the classifier that we choose

However, in most applications we are interested in evaluating if our results are above chance

**The values of classification per se do not matter as much, as the comparison to chance**

# Moment of truth: how well are we doing?

**Our goal:** How robust are our findings?

**The ultimate test**: how well can we classify a new –but similar to our old-dataset?

**We need**: metrics of performance

# Metrics of performance

| | | **True values** | |
|---|---|---|---|
| | | Positives | Negatives |
| **Predicted values** | Positives | True positive | False positive |
| | Negatives | False negative | True negative |

# Metrics of performance

| | | **True values** | |
|---|---|---|---|
| | | Positives | Negatives |
| **Predicted values** | Positives | **True positive** | False positive |
| | Negatives | False negative | **True negative** |

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

# Metrics of performance

**True values**

| **Predicted values** | Positives | Negatives |
|---|---|---|
| Positives | True positive | False positive |
| Negatives | False negative | True negative |

**Is this enough?**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

# Metrics of performance: Example

| | | True values: Positives | True values: Negatives |
|---|---|---|---|
| Predicted values | Positives | 0 | 0 |
| | Negatives | 20 | 80 |

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{0 + 80}{0 + 80 + 0 + 20} = 0.8$$

# Metrics of performance

**Is this enough?** **NO**

| | | True values: Positives | True values: Negatives |
|---|---|---|---|
| **Predicted values** | Positives | 0 | 0 |
| | Negatives | 20 | 80 |

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{0 + 80}{0 + 80 + 0 + 20} = 0.8$$

# Metrics of performance

| | | True values: Positives | True values: Negatives |
|---|---|---|---|
| **Predicted values** | Positives | True positive | False positive |
| | Negatives | False negative | True negative |

$$F1\text{-}score = \frac{TP}{TP + \frac{1}{2}(F, P, +, F, N)}$$

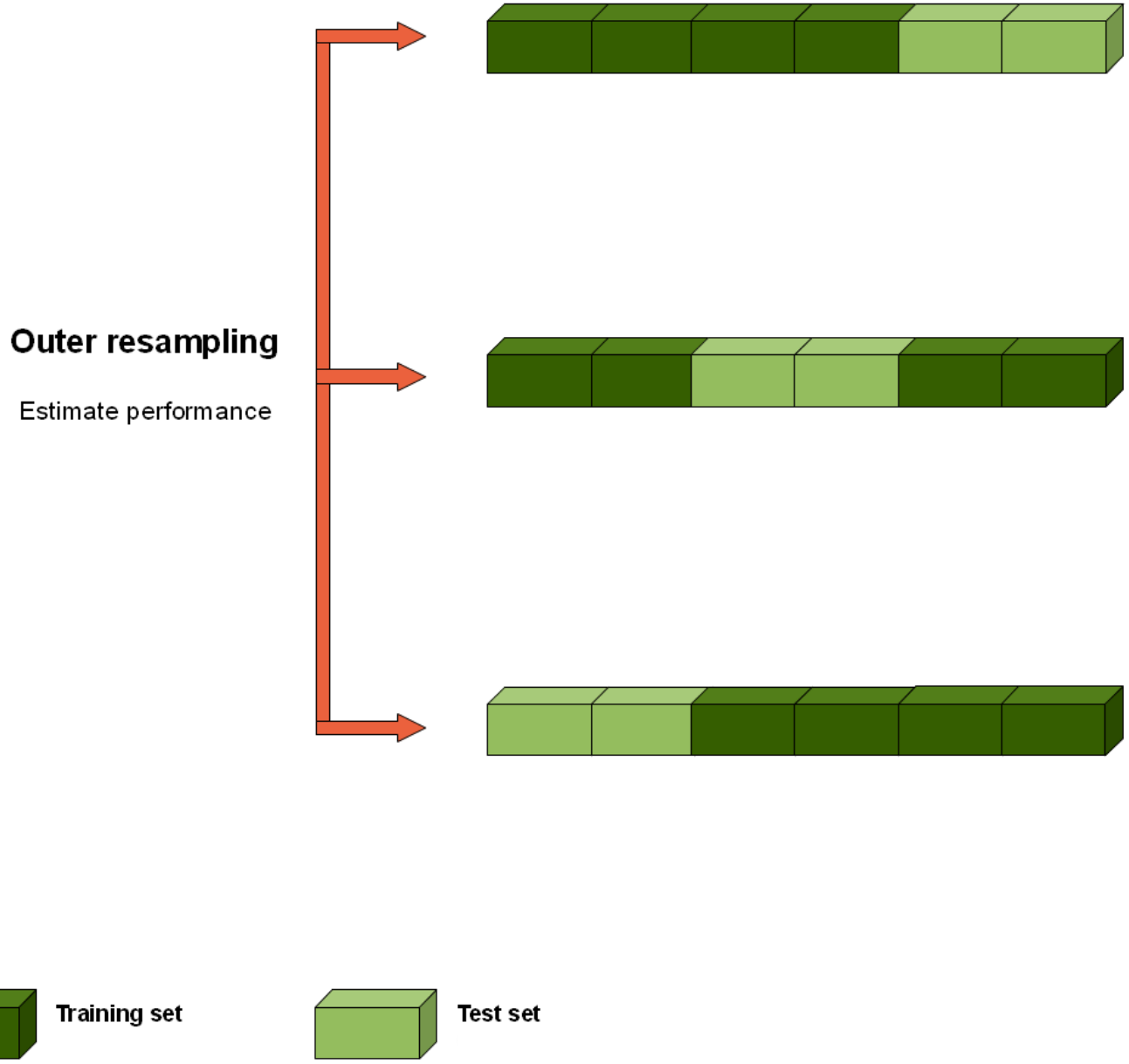
# Metrics of performance

| | | True values: Positives | True values: Negatives |
|---|---|---|---|
| Predicted values | Positives | 0 | 0 |
| | Negatives | 20 | 80 |

$$F1\text{-}score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

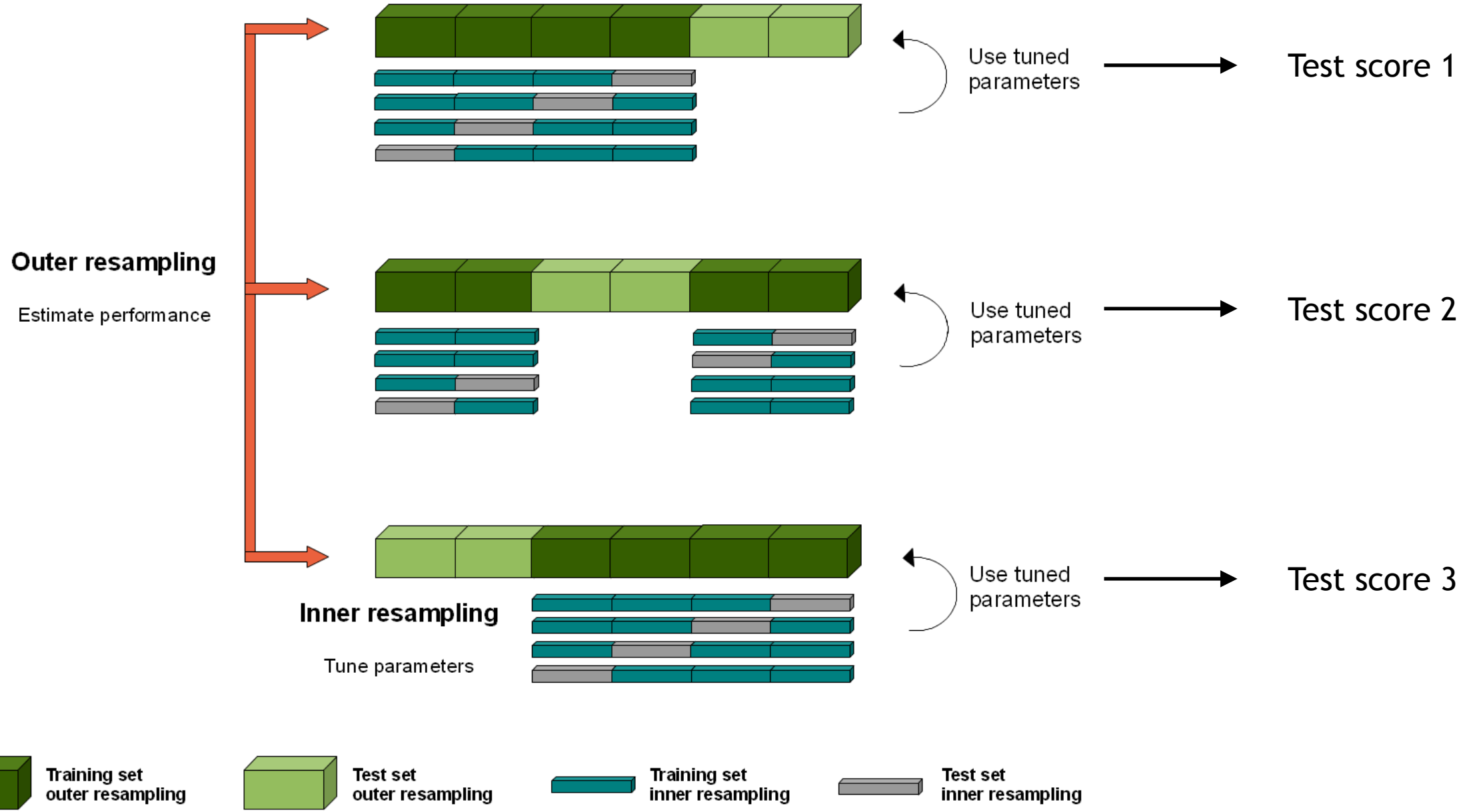
$$F1\text{-}score = \frac{0}{0 + \frac{1}{2}(0 + 20)} = 0$$

# Measuring performance: Train - Test split




https://mlr.mlr-org.com/articles/pdf/img/nested_resampling.png

# Measuring performance: Cross Validation

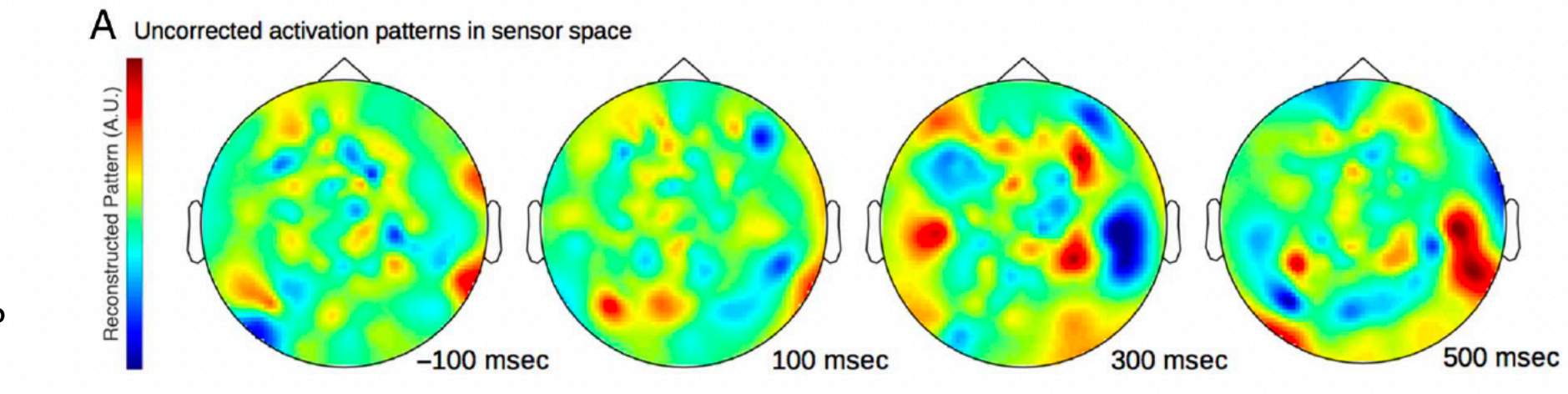



https://mlr.mlr-org.com/articles/pdf/img/nested_resampling.png

# What features has our classifier learned?

# What features has the classifier learned?

**Classifier weights**
How much does each electrode contribute to the classifier's decision?

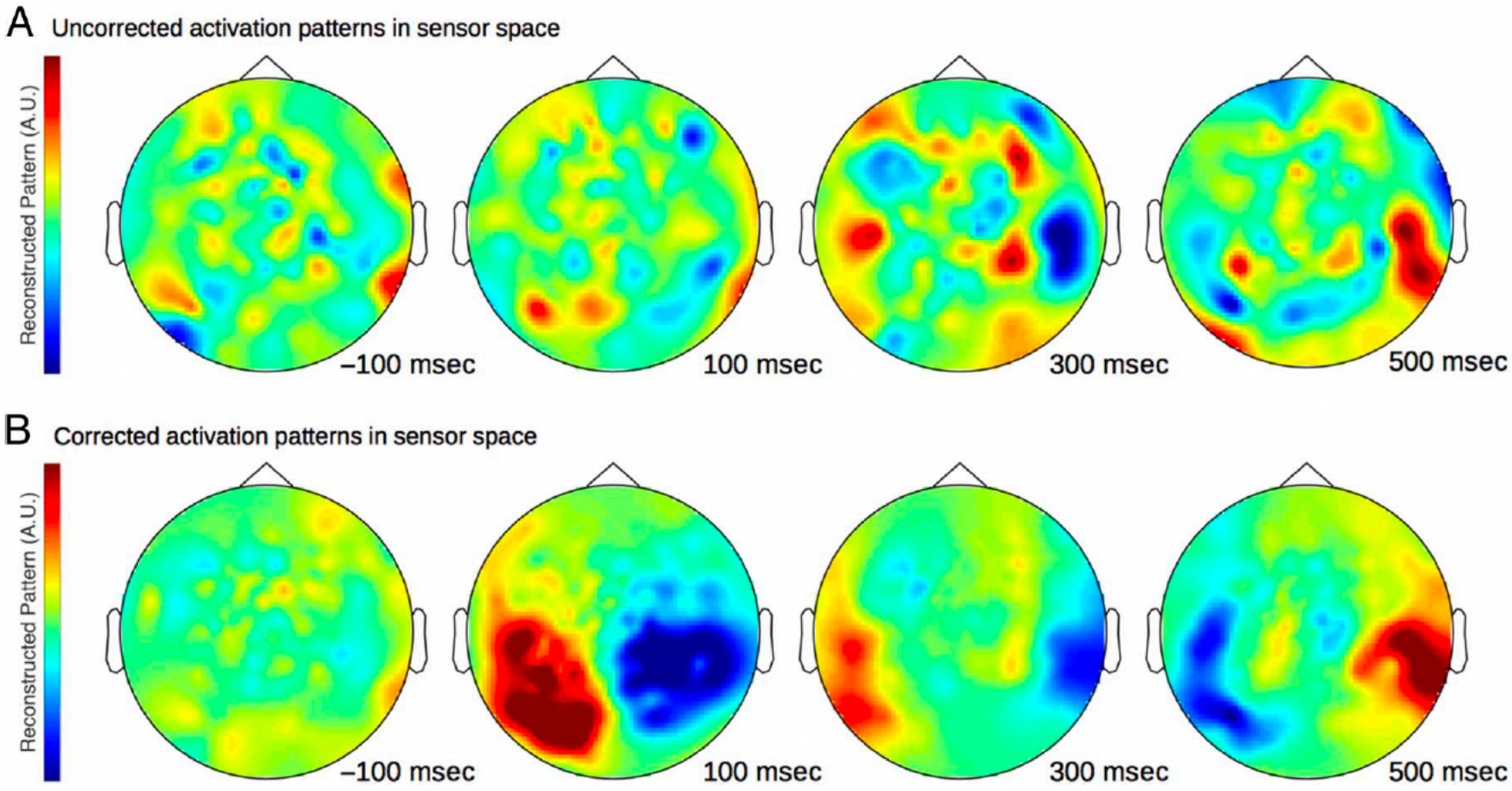



Grootswagers et al., 2017, J Cogn. Neuroscience

# What features has the classifier learned?

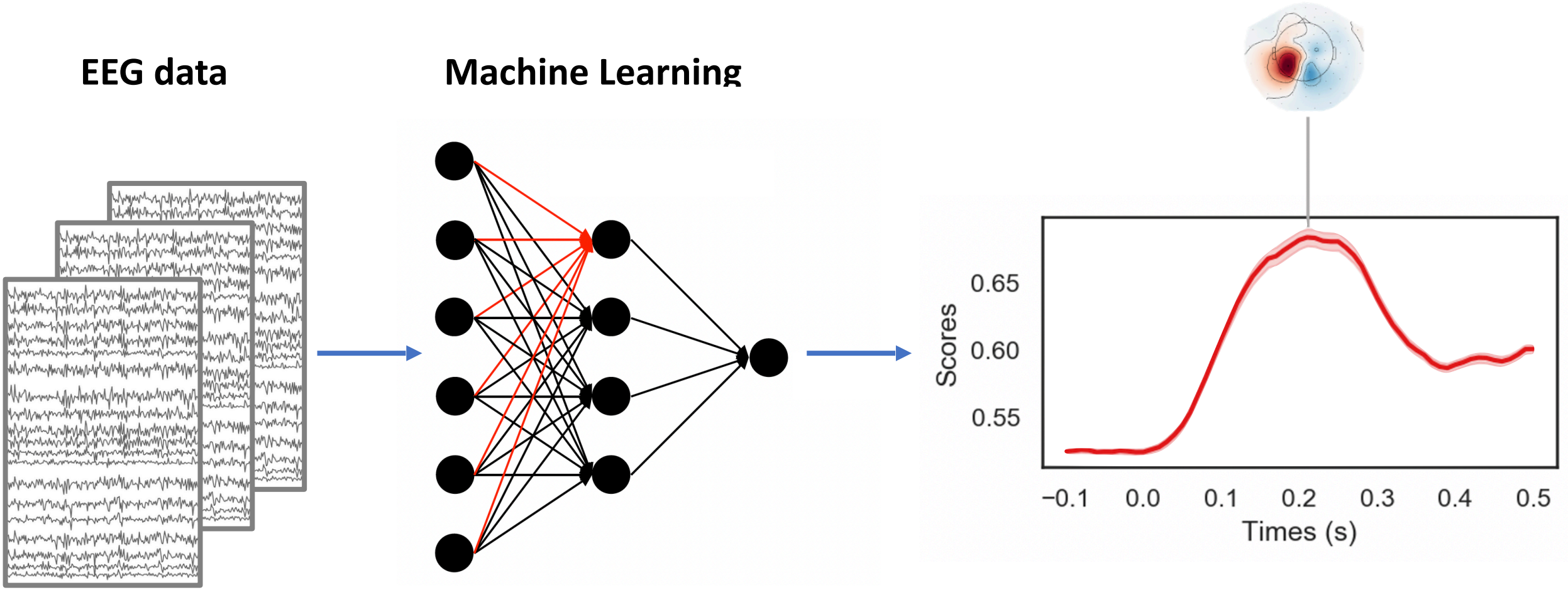
**Classifier weights**
How much does each electrode contribute to the classifier's decision?

**Activation maps**
Weights * Covariance in data
Interpretability
Less prone to noise




Grootswagers et al., 2017, J Cogn. Neuroscience

# What do we retain for “traditional”ML applications on EEG data?



# What do we retain for “traditional”ML applications on EEG data?

Features: They are not class-specific!

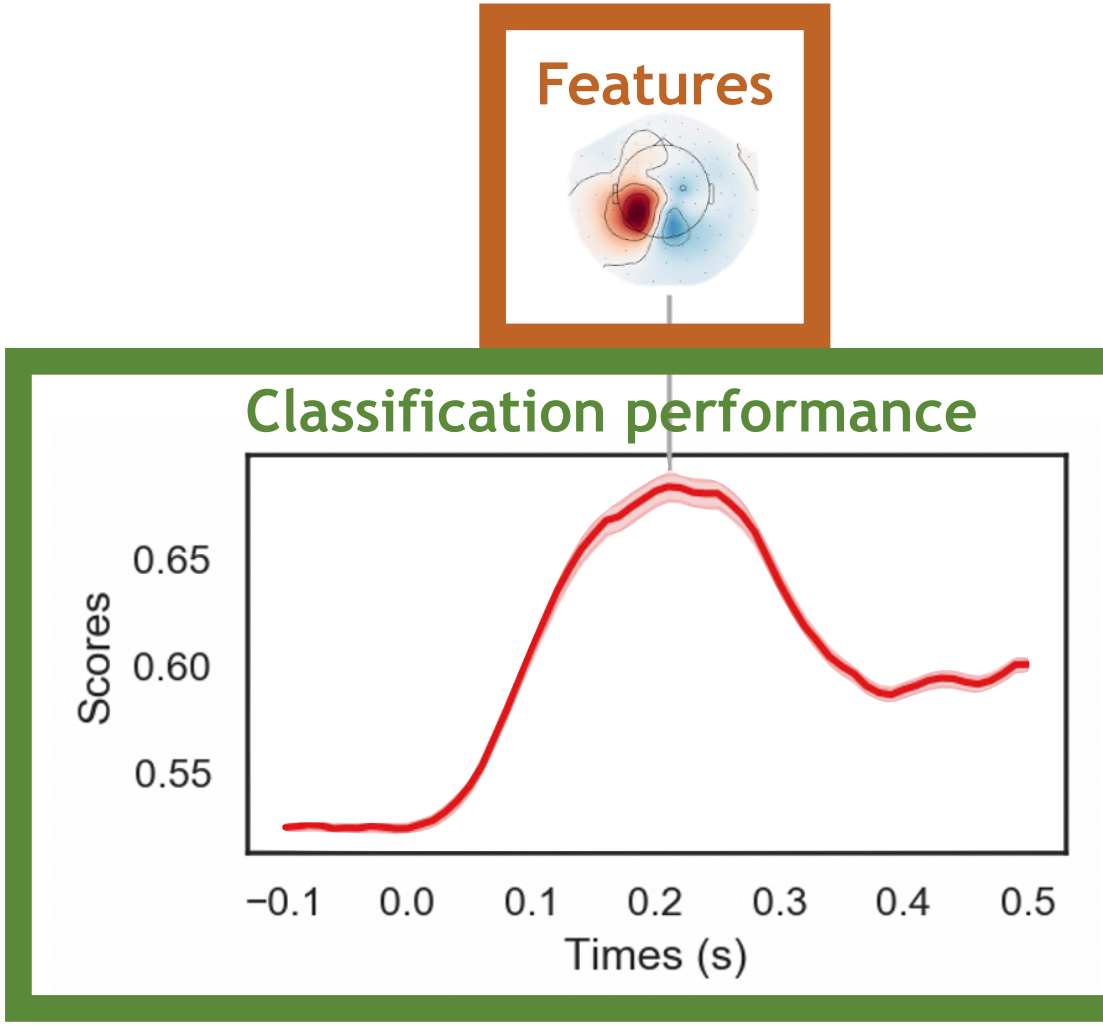
i.e. we don’t know which of our classes has mainly contributed to them!

1. **Which electrodes contribute to accurate classification?**

Classification performance is typically a time-course

It is not class specific either, but it evolves over time:

2. **When is it above chance?**



# Tutorial II : HANDS ON

# Overview for today

Introduction to AI in neuroscience :

- Electroencephalography (EEG) signals
- Hands-on: working with EEG

Machine Learning in neuroscience

- Supervised learning: training classifiers
- Measuring performance
- Hands-on: Classifying EEG data

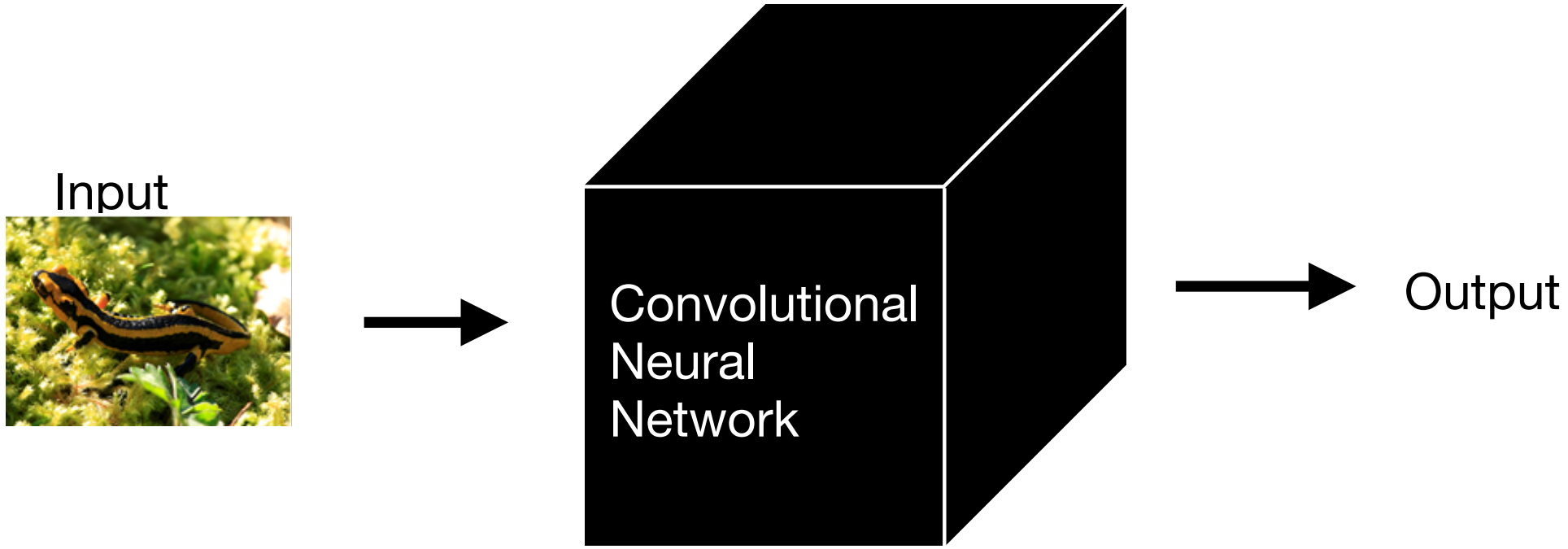
Convolutional Neural networks for EEG signals

- Training networks & measuring performance
- Hands-on: working with neural networks

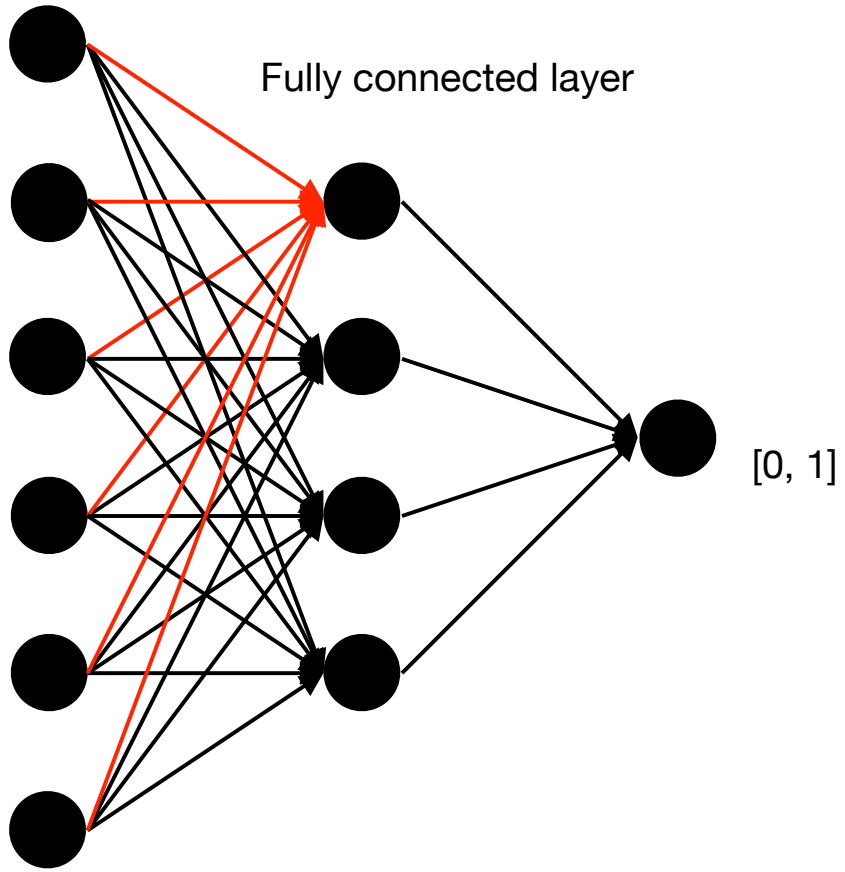
Group work & presentations:

- Mini projects: try out what we learned in short projects & your own ideas

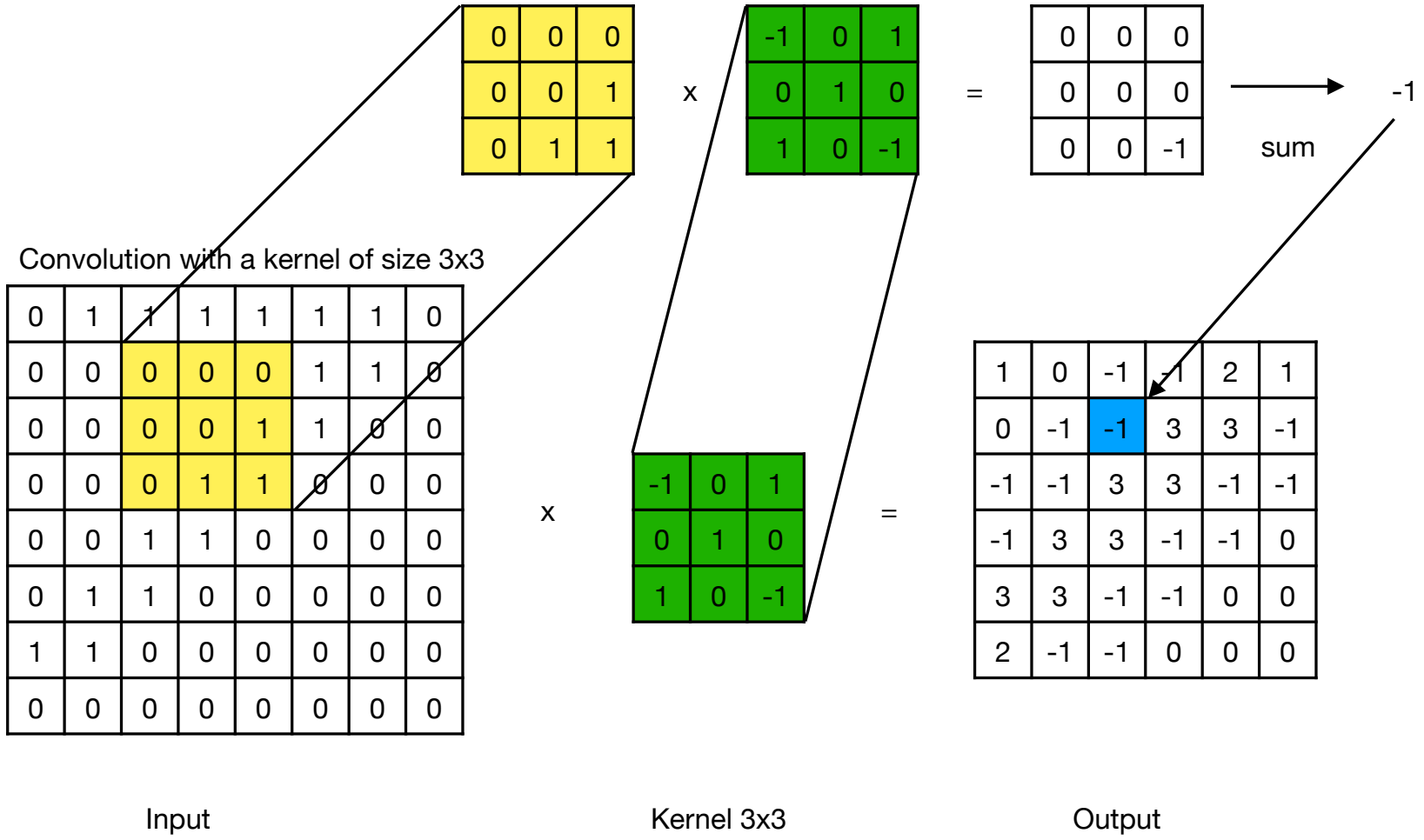
# Convolutional Neural Networks



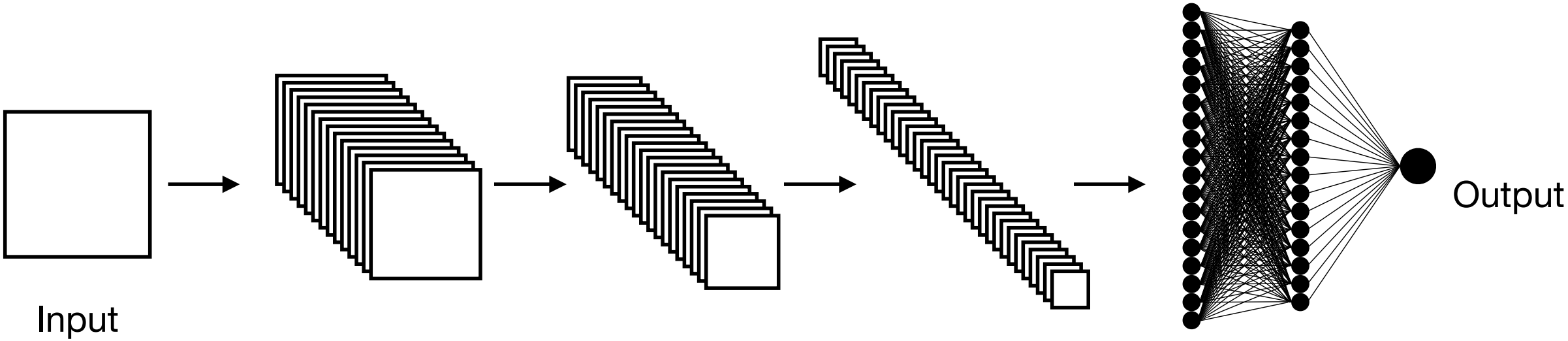
# Convolutional Neural Networks: Fully Connected Layers



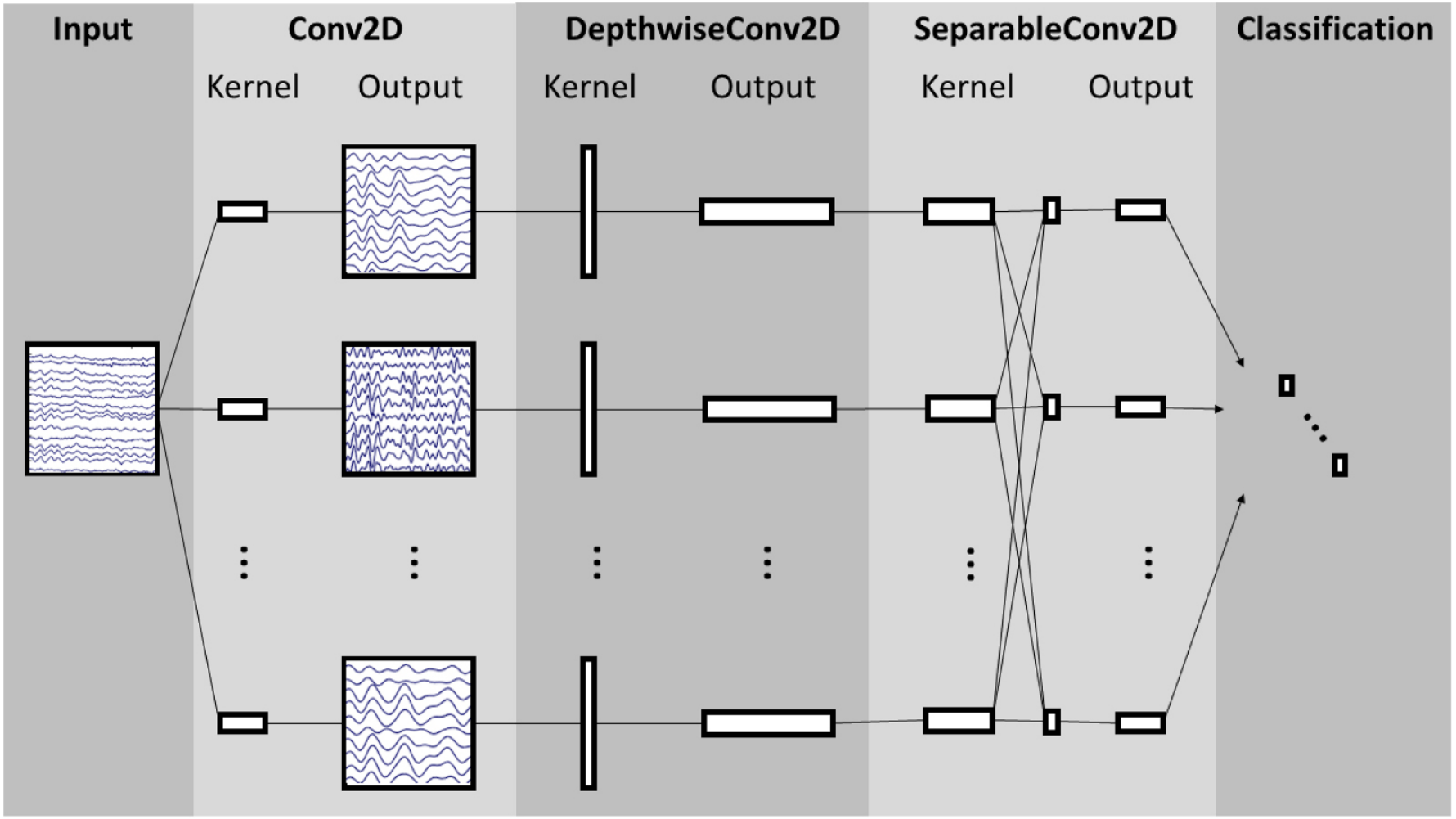
# Convolutional Neural Networks: Convolutions



# Convolutional Neural Networks: Full Network

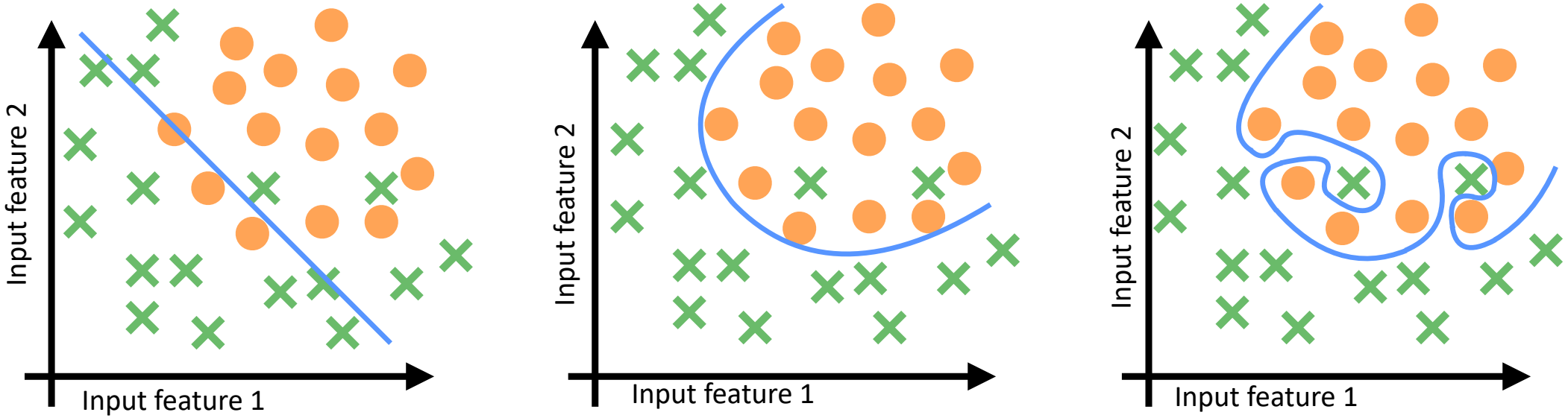


# We are going to use: EEGNet




Lawhern et al., 2018

# Overfitting



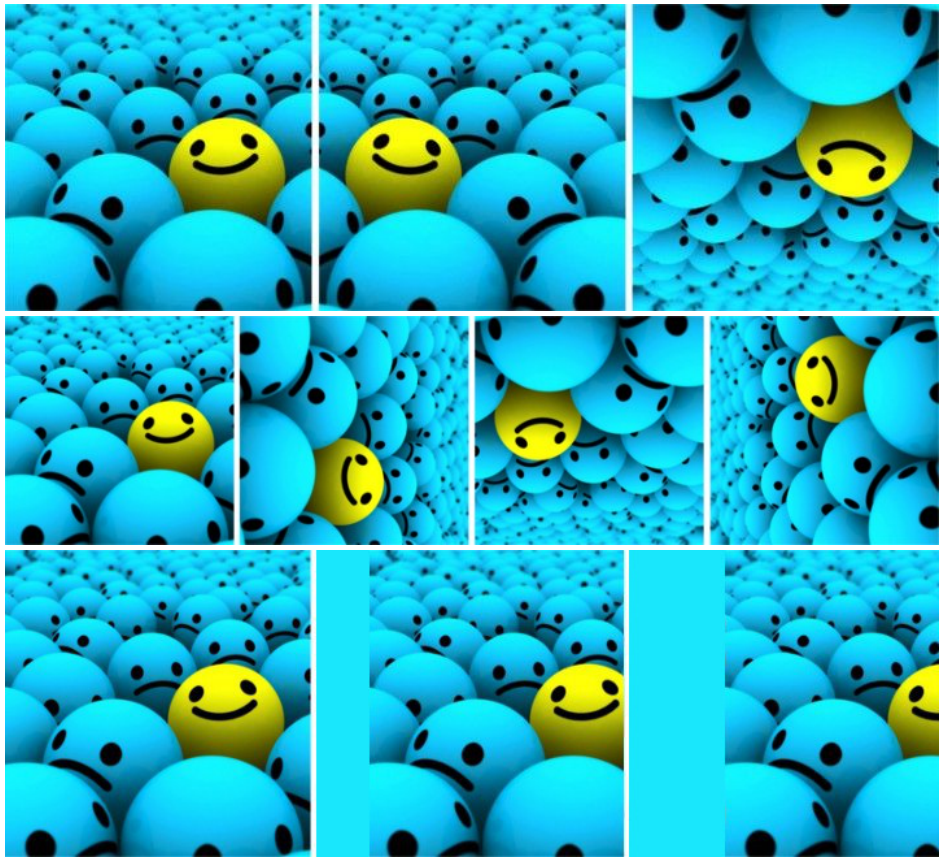

Some Problems:

- Limited amount of data
- A lot of parameters

Some Solutions:

- Dropout
- Batch Normalisation
- **Data Augmentation**

# Data augmentation techniques from computer vision



Flipping

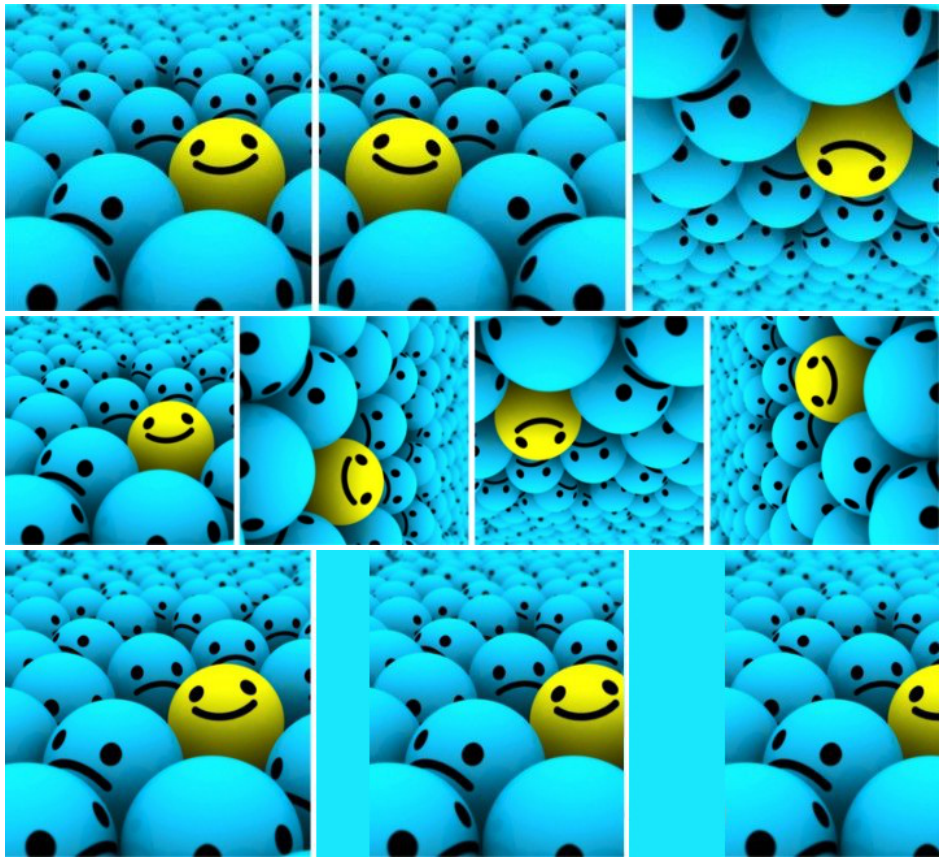
Rotation

Translation

https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/

# Data augmentation techniques from computer vision

**Extra advantage of data augmentation techniques:**
**- They can help us deal with bias by generating synthetic data for underrepresented classes**



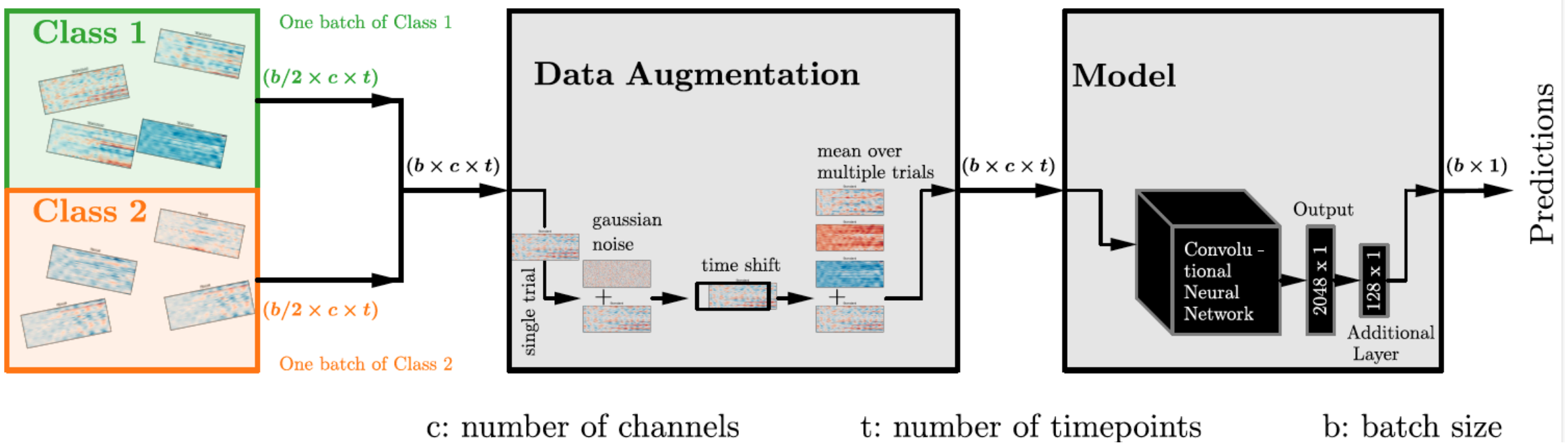
Flipping

Rotation

Translation

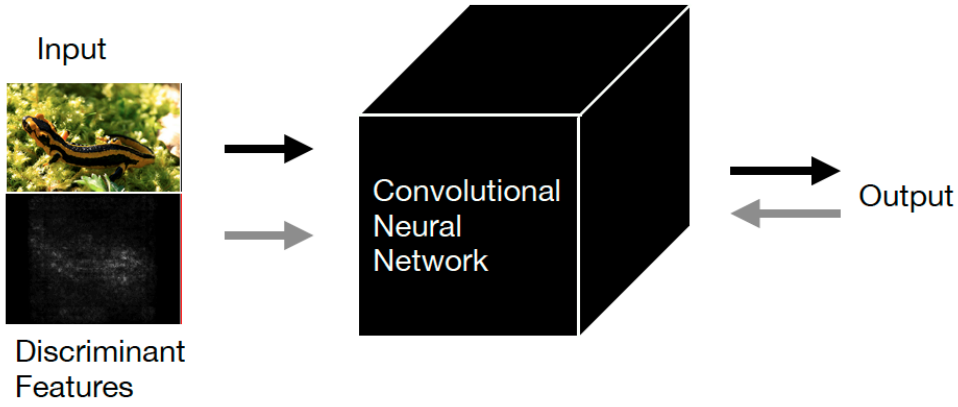
https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/

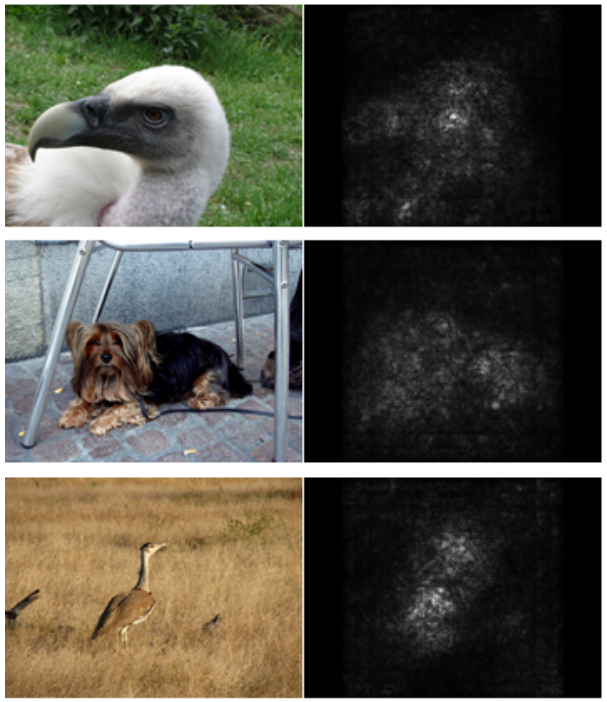
# Pipeline for CNNs for classifying EEG data




Aellen et al., 2021

# Feature Extraction

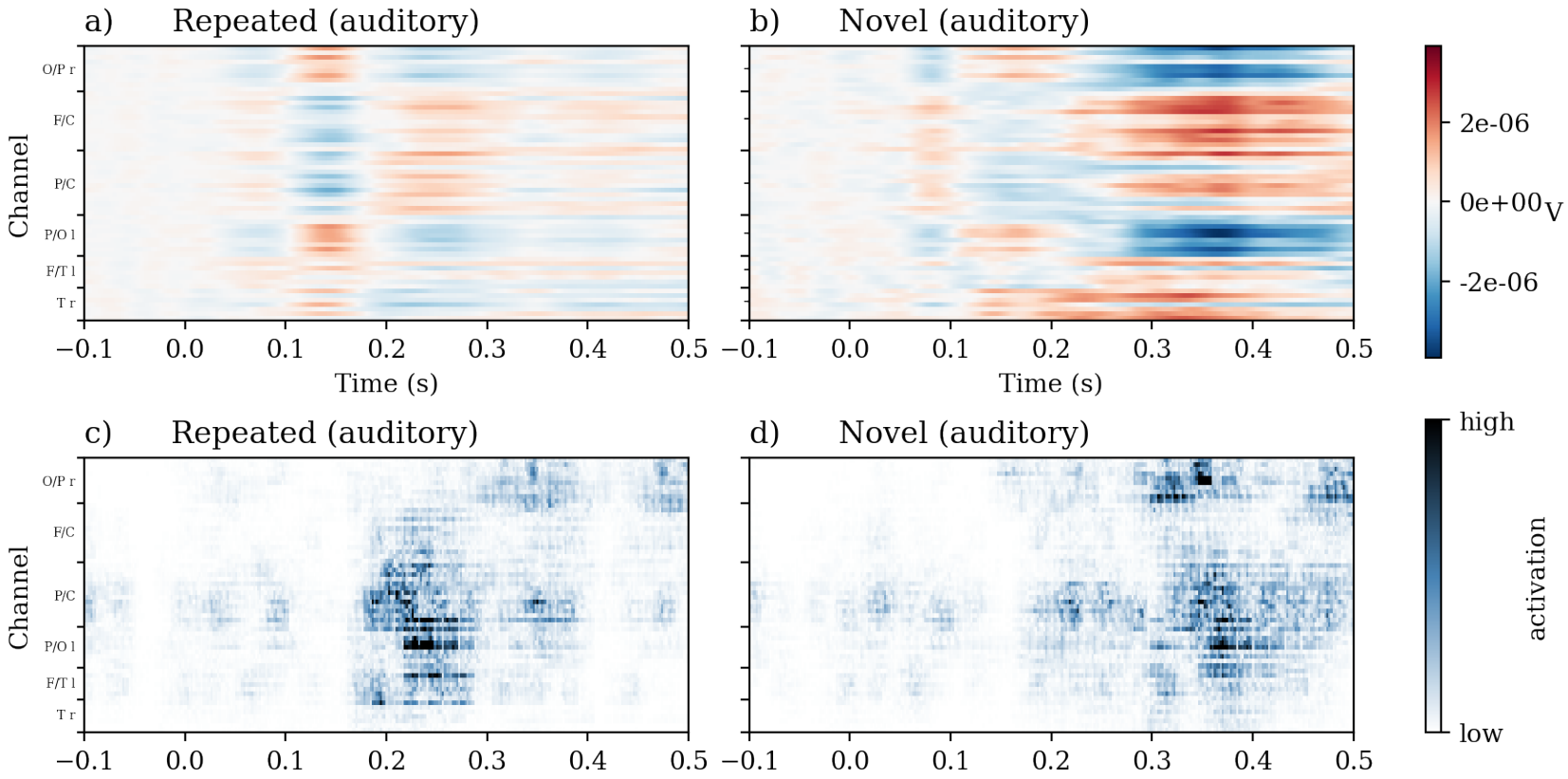


Simonyan et. al (2013)

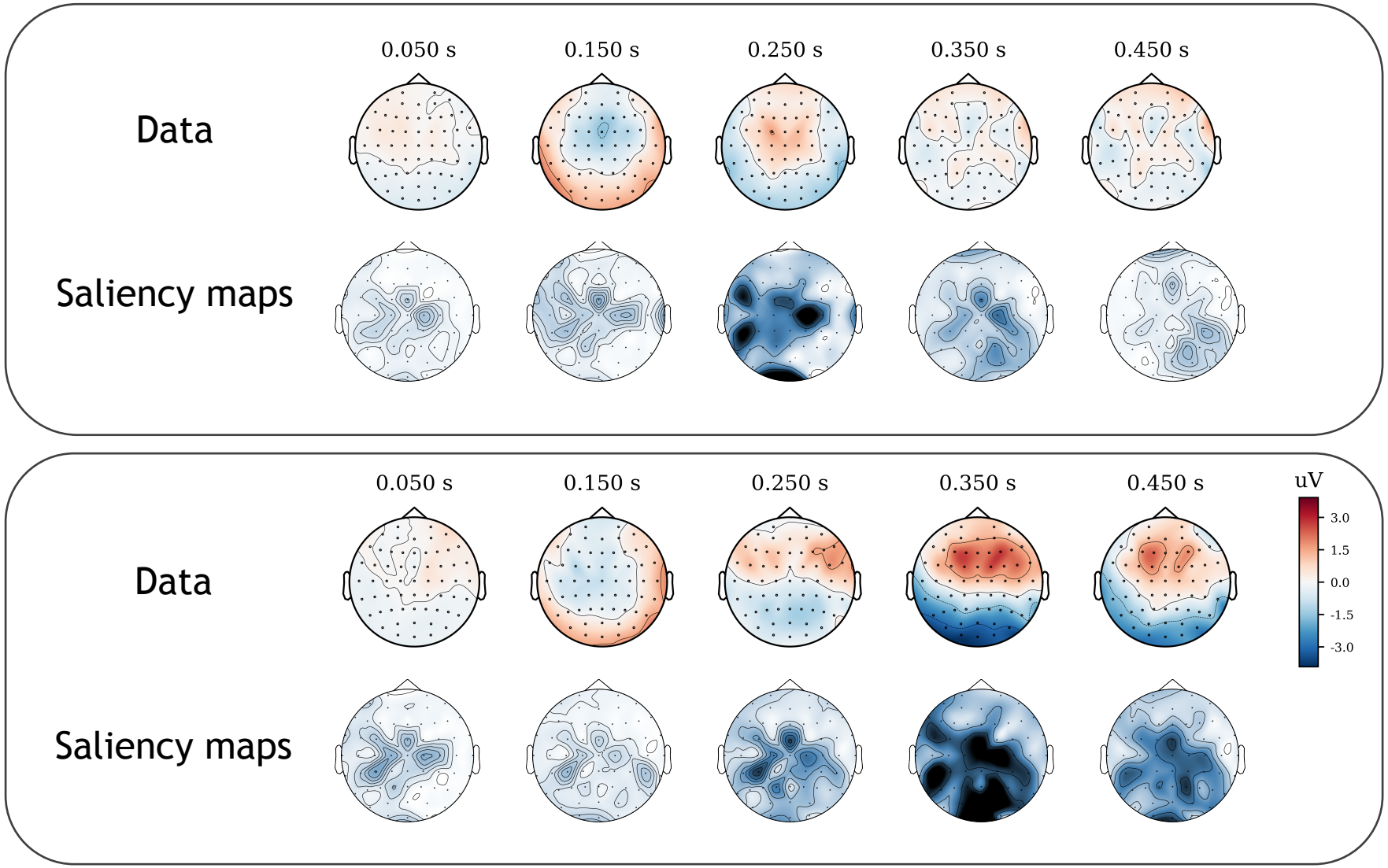
Saliency maps show the pixels most relevant for the networks decision.

# Additional advantage: feature extraction



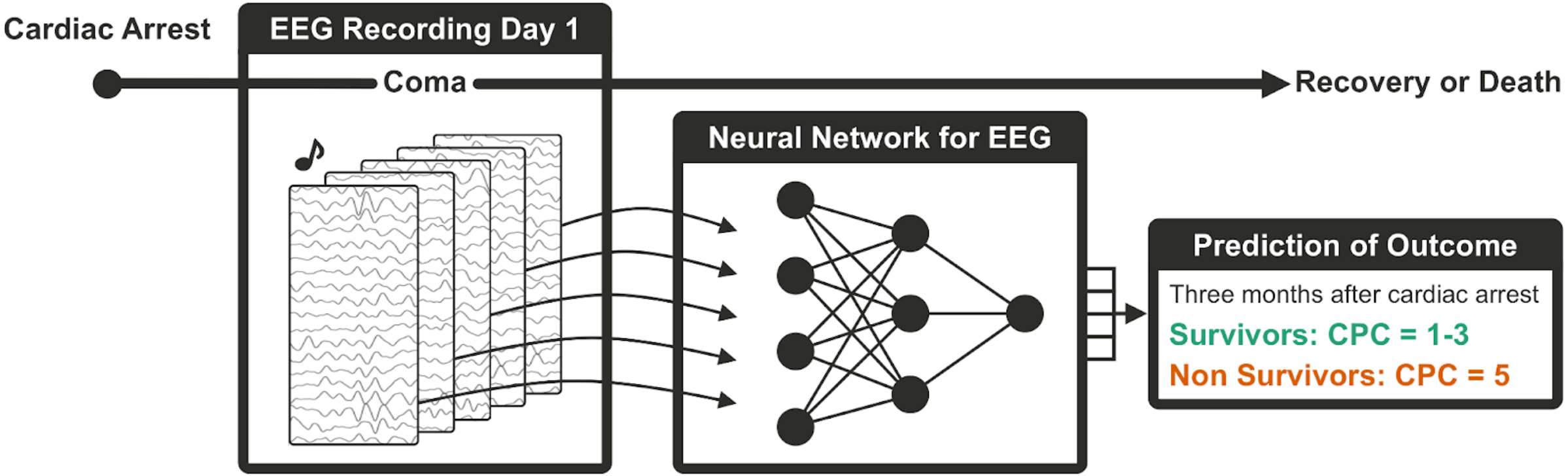

Aellen et al., 2021

# Class - specific features



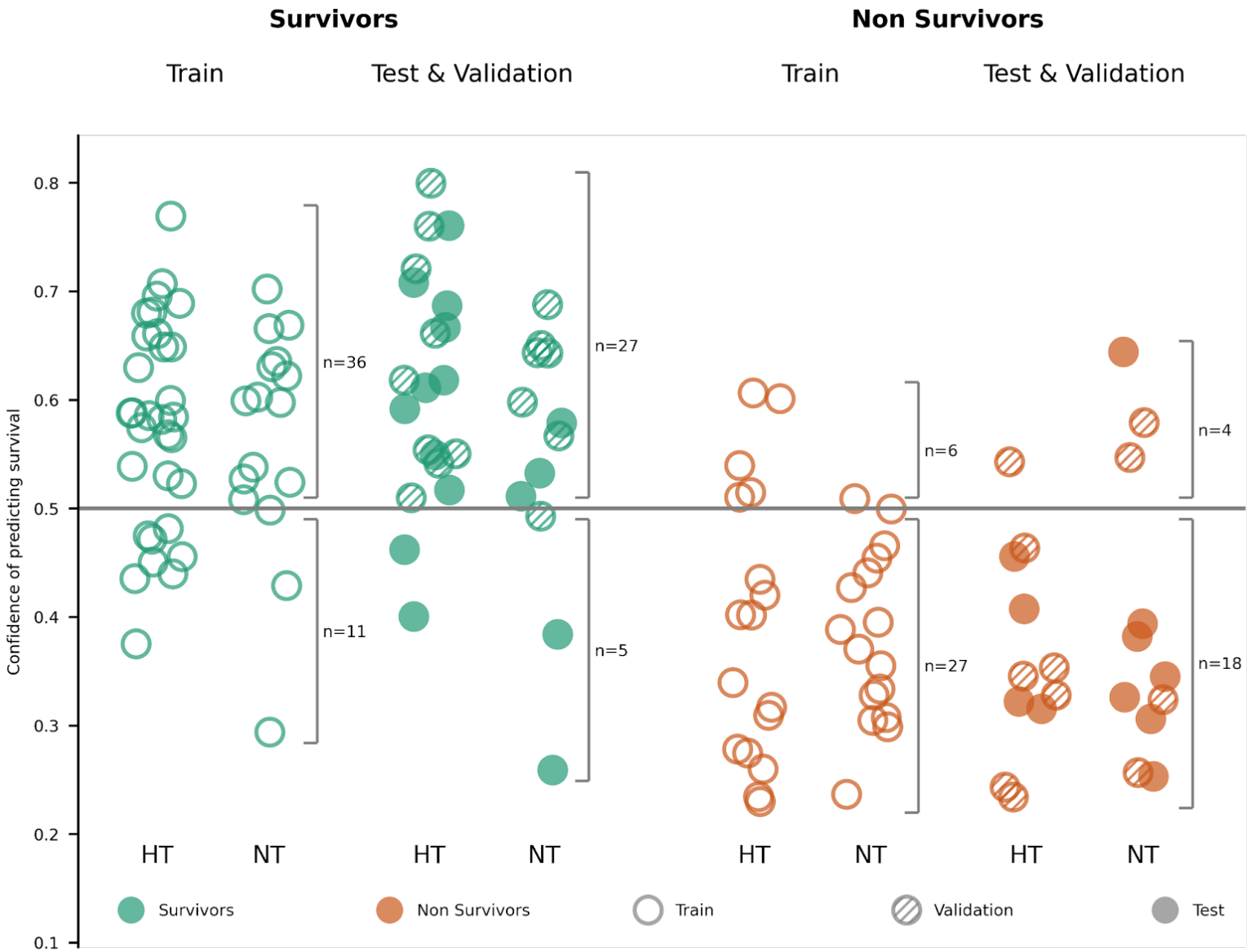

Aellen et al., 2021

# Application: Predicting outcome from coma




Aellen et al., Under review

# CNNs are as good as clinicians in predicting a patients' chances of awakening

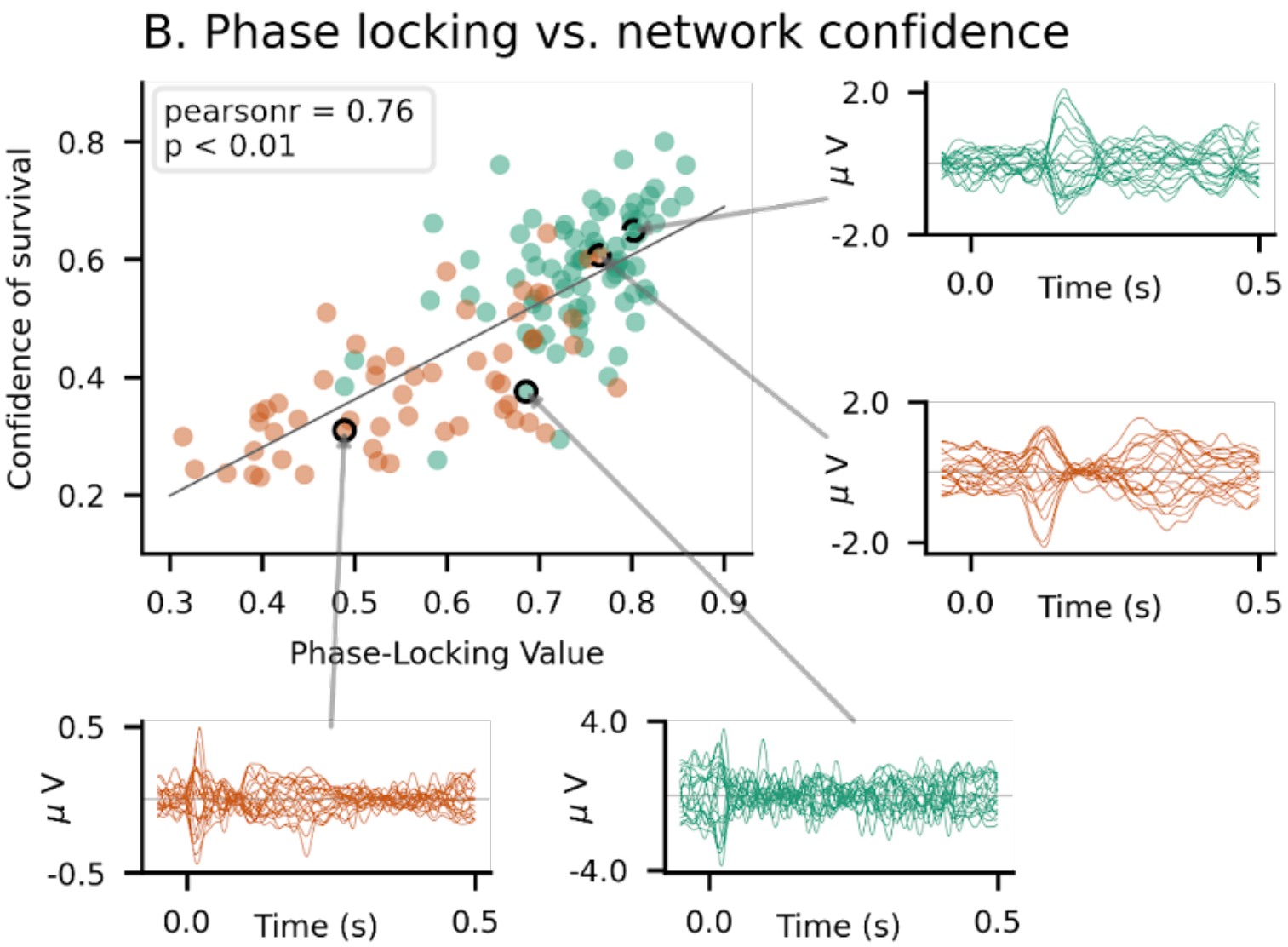



CNNs applied on EEG data of first day of coma can predict chances of awakening 3 months later, with
83% predictive value

**This is as good as clinical tests that rely on medical expertise, with the advantage of objective and automatic procedure !!**

Aellen et al., Under review

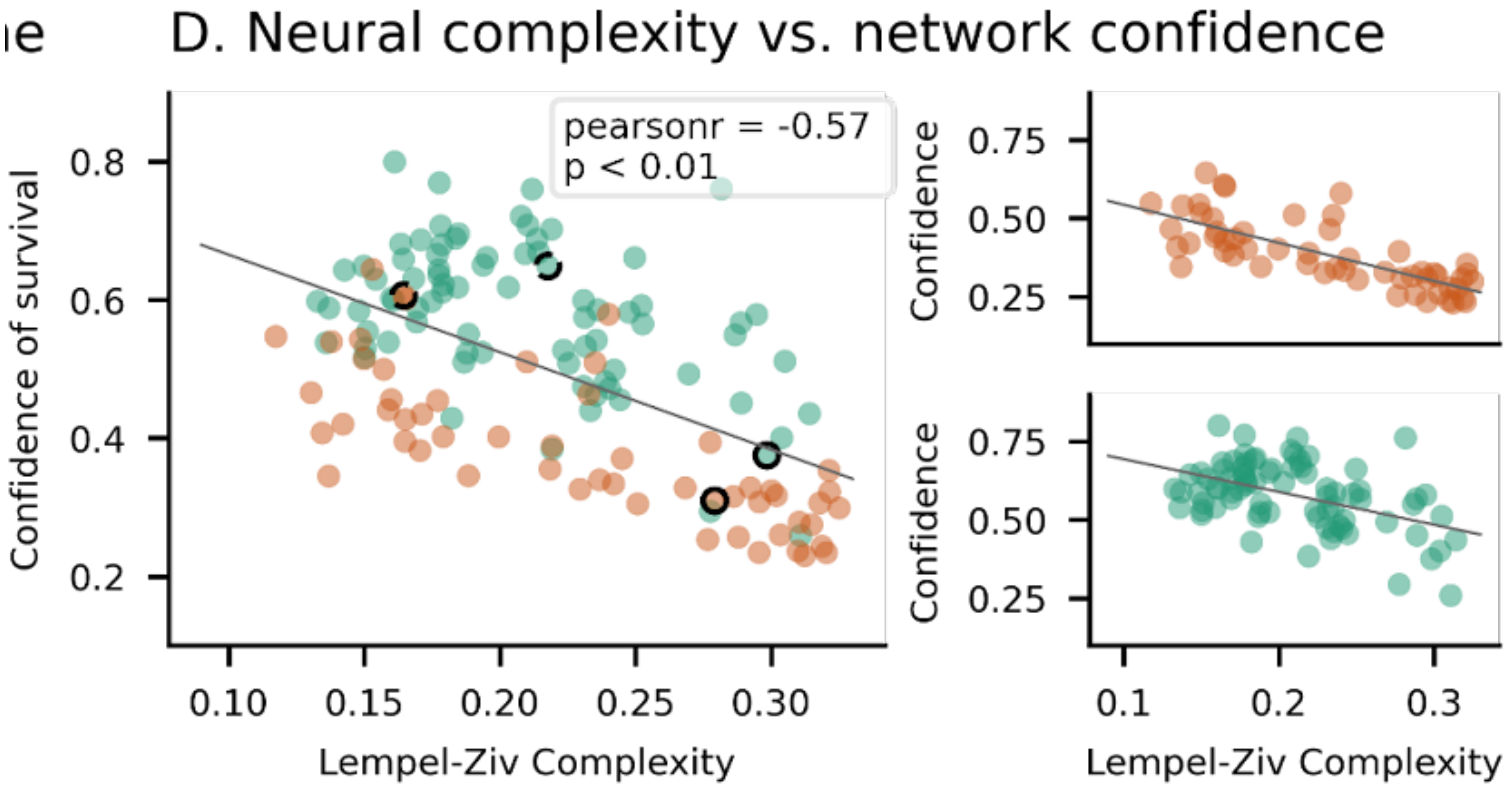
# Interpreting features of CNNs during prediction of coma outcome




The confidence of CNNs in predicting outcome reflects the neural synchrony of EEG responses

Aellen et al., Under review

# Interpreting features of CNNs during prediction of coma outcome




The confidence of CNNs in predicting outcome reflects the neural synchrony of EEG responses

And also neural complexity

High confidence of CNN:

- Low complexity in EEG signals
- High synchrony

Indicative of survival

Aellen et al., Under review

# Tutorial III : HANDS ON

# Overview for today

Introduction to AI in neuroscience :

- Electroencephalography (EEG) signals
- Hands-on: working with EEG

Machine Learning in neuroscience

- Supervised learning: training classifiers
- Measuring performance
- Hands-on: Classifying EEG data

Convolutional Neural networks for EEG signals

- Training networks & measuring performance
- Hands-on: working with neural networks

Group work & presentations:

- Mini projects: try out what we learned in short projects & your own ideas

# Mini projects

- Now it is time for **you** to come up with your own ideas of mini-projects where you can explore the notions that we saw today. For inspiration we give you some ideas:

1. Compare different time-domain classifiers (Part II)
2. Compare how CV influences results, experiment with  train test ; validation splits (Part II)
3. Optimize EEGNet parameters and compare output (Part III)
4. Build a classifier to decoding the identity of different participants (Part II-III)
5. Train a classifier with different references (Part II) or different filters
6. Bias: artificially reduce or augment your dataset, classify and observe the effects on the output (Part II - III)

# Part IV : Group presentations

# Summary

1. AI for EEG signals is a powerful tool that can assist clinical decision-making (e.g. prognosis of coma outcome; diagnosis of sleep disorders)

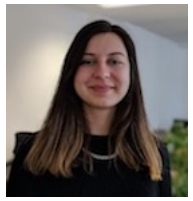
2. ‘Traditional’ machine learning algorithms provide limited performance & interpretability; require some a priori hypotheses and data curation

3. CNNs increase performance; rely on minimally processed data

4. Extraction of features needs to complement AI algorithms

5. Dealing with bias: Generating synthetic data; caution in selecting and evaluating algorithms

# Online resource for classification of neural signals



P. Göktepe

**EEG Data Processing**

1. Background

2. Introduction to Signal Processing
   - Complex Numbers
   - Temporal Vs. Spectral Space
   - Acquiring Data

3. Preprocessing
   - Data Loading
   - Data Visualization
   - Baseline Correction
   - Choosing a Reference

4. Machine Learning
   - Formatting a dataset
   - Single-Participant Analysis
   - Group-Level Analysis
   - Classication over time

5. About Us

6. Outline

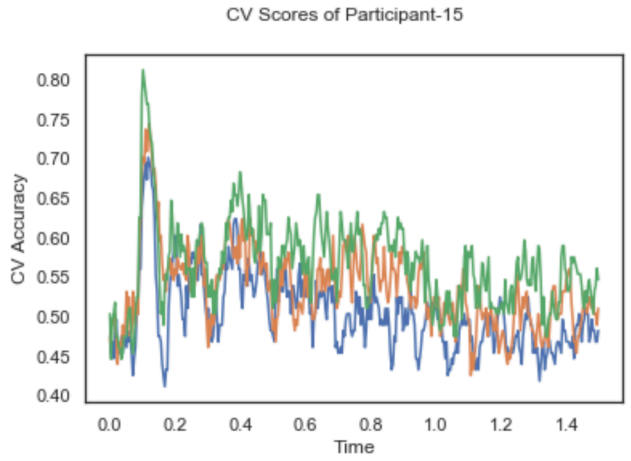
```
import matplotlib.pyplot as plt
%matplotlib inline

def plotCVScores(times, CV_score_time, id=None):
    fig, ax = plt.subplots()
    if id != None:
        fig.suptitle('CV Scores of Participant-'+str(id))
    else:
        fig.suptitle('CV Scores')
    ax.plot(times, CV_score_time.T)
    plt.xlabel('Time')
    plt.ylabel('CV Accuracy')
    plt.show()
```

**Classification Between Unplesant and Neutral Events**

```
CV_score_time_UN = []
for i in range(len(data_UN)):
    print('Participant id: '+ str(ids[i]))
    clf = make_pipeline(Vectorizer(), StandardScaler(), LinearDiscrim
    CV_score_time_UN.append(applyCrossValidation(data_UN[i], labels_U
```

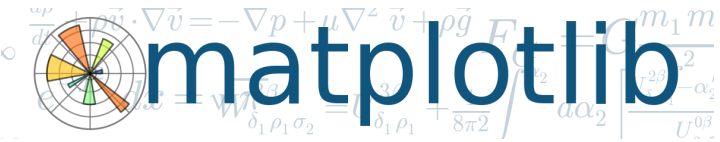
```
Participant id: 15
[.....................................] 100.00% Fitting SlidingEstimator |
[.....................................] 100.00% Fitting SlidingEstimator |
[.....................................] 100.00% Fitting SlidingEstimator |
```




https://neuro.inf.unibe.ch/

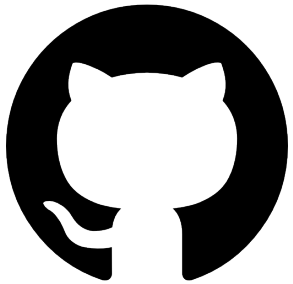
# Online resource: the key ingredients

## 1. Open source libraries for signal processing, visualization and machine learning




## 2. Open source infrastructure

## 3. Open access data


# Thank you!

**Cognitive & Computational Neuroscience Group**
**University of Bern**

Dr. Ruxandra Tivadar
Florence Aellen
Sigurd Lurked Alnes
Pinar Göktepe
Riccardo Cusinato
Lea Zoe Meret Bächlin
Anna Morf
Fabian Loosli

**AI & Bias**

Qiyang Hu

Natalia Norori

florence.Aellen@inf.unibe.ch

athina.Tzovara@inf.unibe.ch

@AthinaTzovara

https://neuro.inf.unibe.ch/

b
UNIVERSITÄT
BERN

Interfaculty Research Cooperation:
Decoding Sleep

moz://a

FNSNF
SWISS NATIONAL SCIENCE FOUNDATION