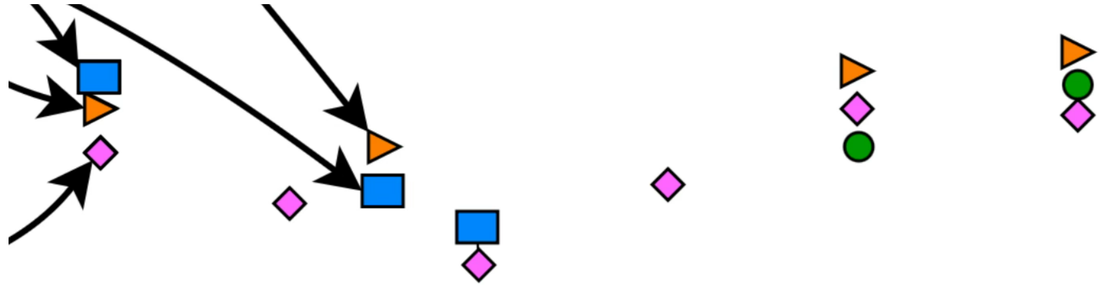
small data
MORE FROM LESS
CRC 1597

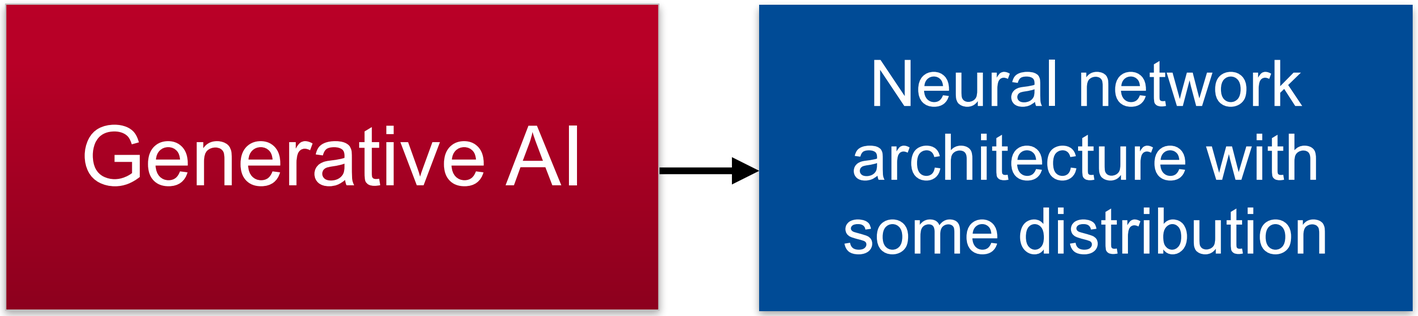
UNIVERSITÄTS KLINIKUM FREIBURG

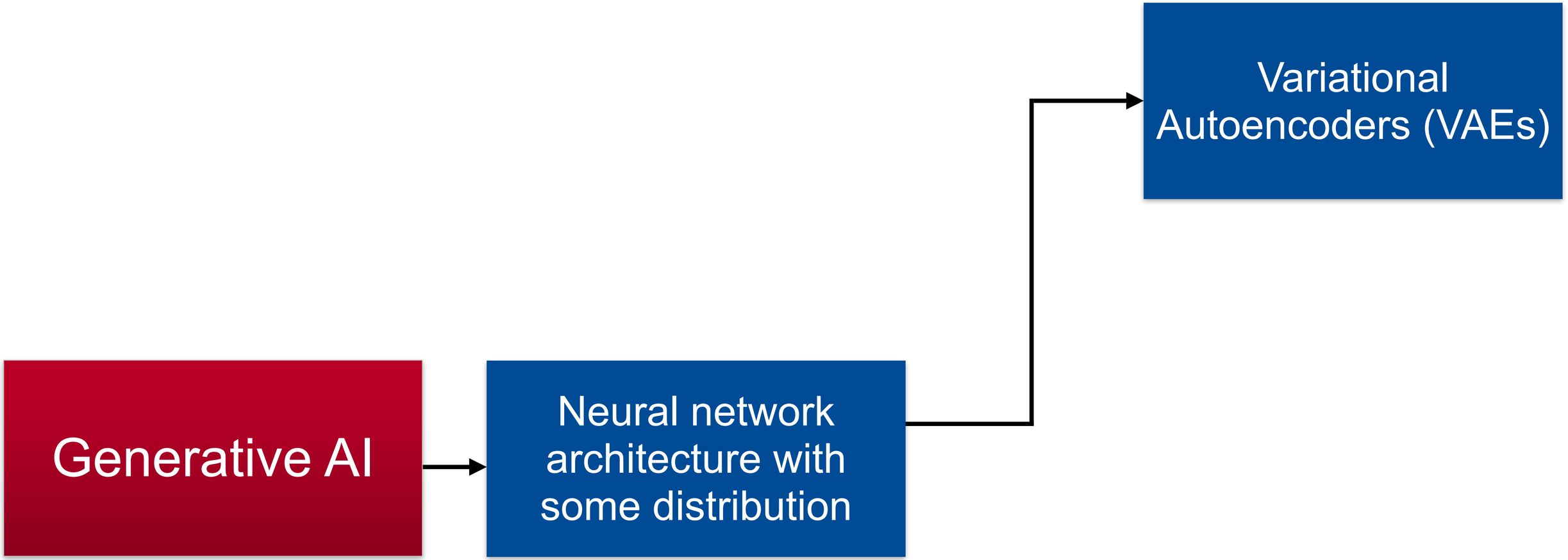


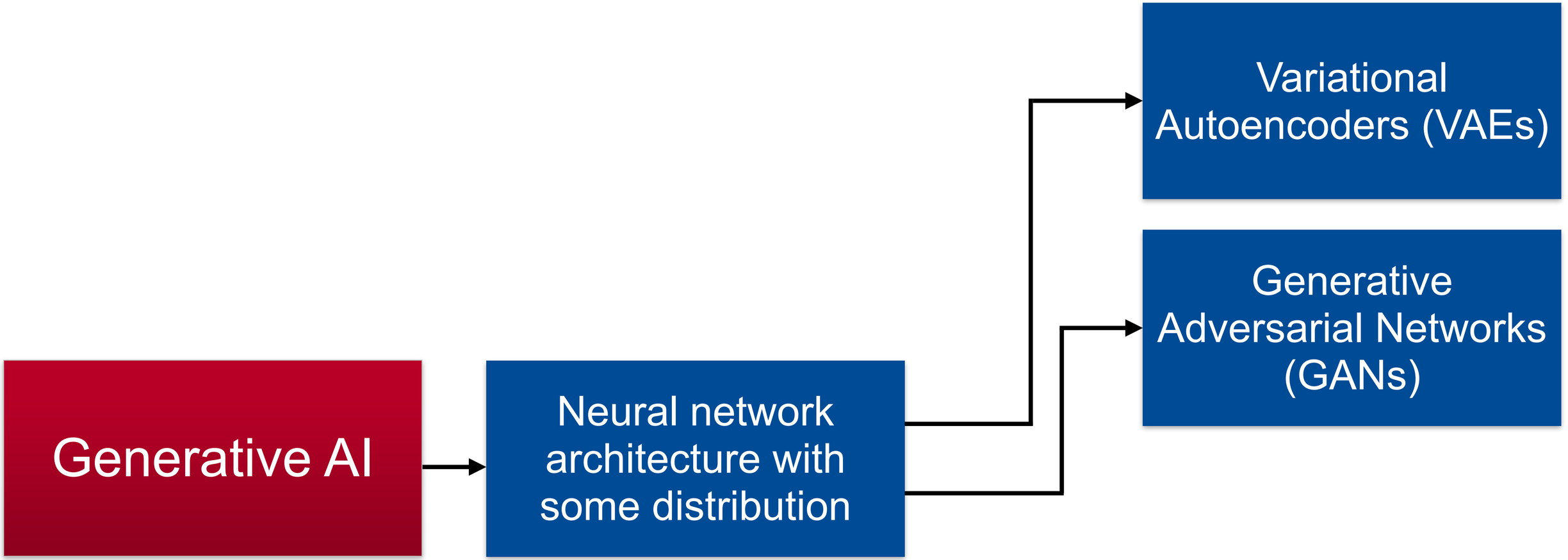
Leveraging generative AI approaches for small data settings in clinical research

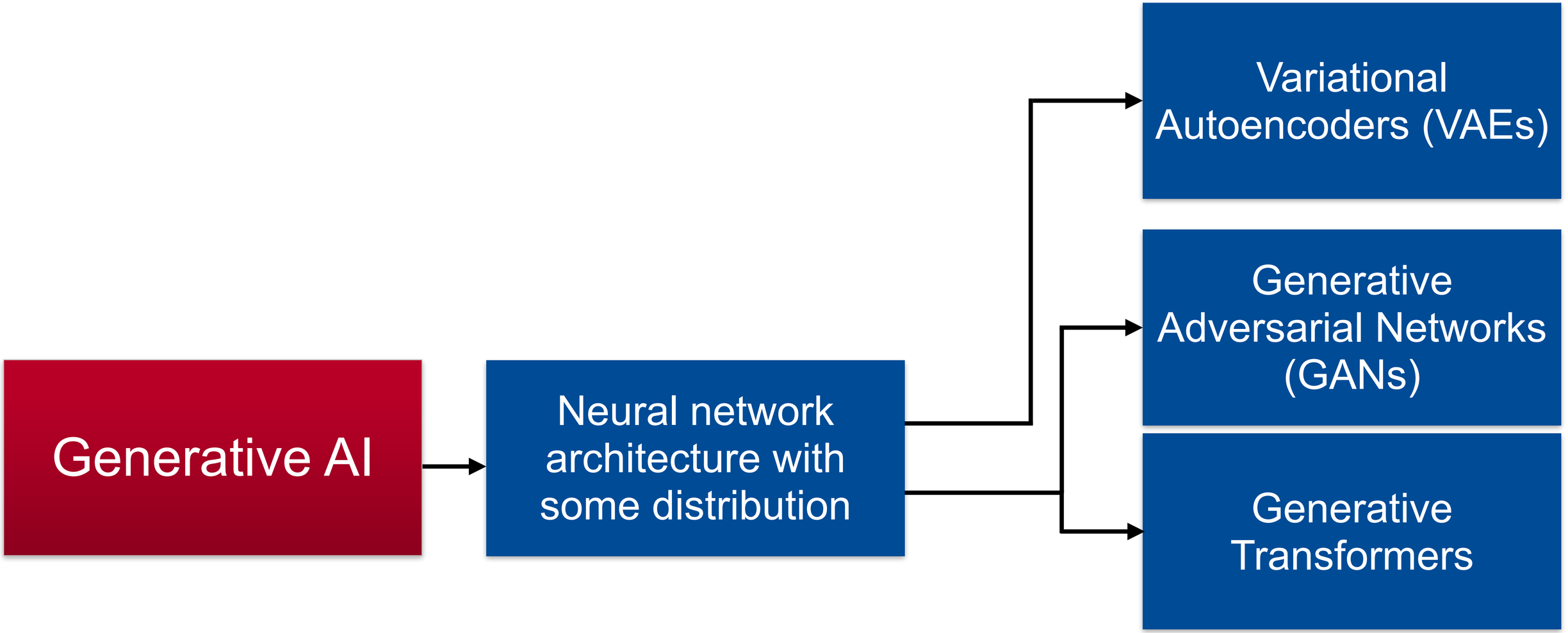
Harald Binder, Institute of Medical Biometry and Statistics (IMBI)

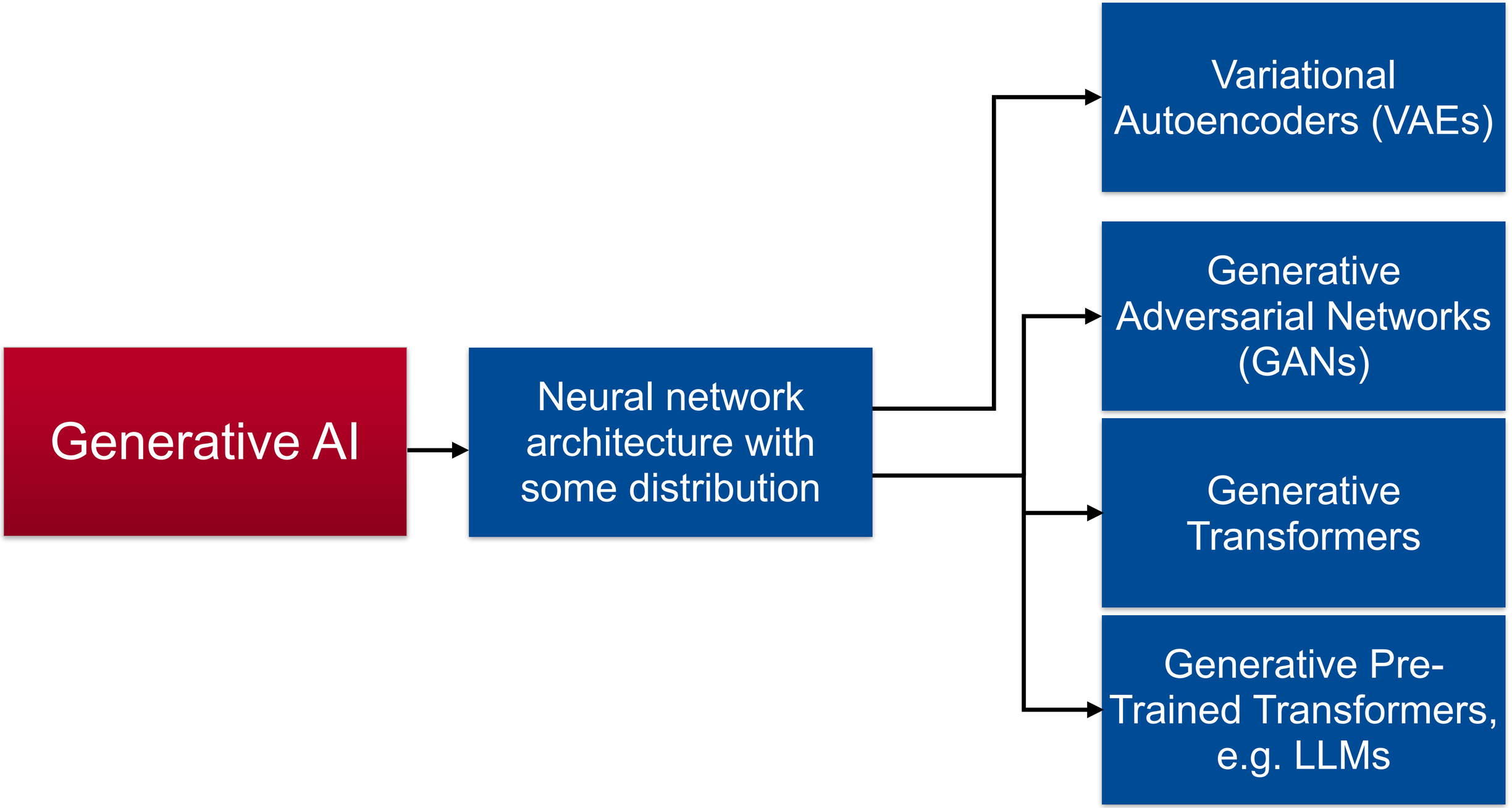
Generative AI


Generative AI
Neural network architecture with some distribution









Longitudinal rare disease registries

Pechmann *et al. Orphanet Journal of Rare Diseases* (2019) 14:18
https://doi.org/10.1186/s13023-019-0998-4

Orphanet Journal of Rare Diseases

RESEARCH Open Access

SMArtCARE - A platform to collect real-life outcome data of patients with spinal muscular atrophy

Astrid Pechmann[1], Kirsten König[2], Günther Bernert[3], Kristina Schachtrup[2], Ulrike Schara[4], David Schorling[1], Inge Schwersenz[5], Sabine Stein[1,6], Adrian Tassoni[2], Sibylle Vogt[1,6], Maggie C. Walter[7], Hanns Lochmüller[1,8] and Janbernd Kirschner[1*]

Abstract

Background: Survival and quality of life for patients affected by spinal muscular atrophy (SMA) are thought to have improved over the last decade due to changes in care. In addition, targeted treatments for SMA have been developed based on a better understanding of the molecular pathology. In 2016 and 2017, nusinersen was the first drug to be approved for treatment of all types of SMA in the United States and in Europe based on well-controlled clinical trials in a small subgroup of pediatric SMA patients. Systems are required to monitor treated and untreated SMA patients in a real-life environment to optimize treatment and care, and to provide outcome data to regulators, payers, and the SMA community.

Methods: Within SMArtCARE, we conduct a prospective, multicenter non-randomized registration and outcome study. SMArtCARE collects longitudinal data on all available SMA patients independent of their actual treatment regime as disease-specific SMA registry. For this purpose, we provide an online platform for SMA patients

UNIVERSITÄTS KLINIKUM FREIBURG

Dimension reduction with VAEs

Items from motor function assessment in SMA

Test items

Encoder

Q1: Can the patient stand?

Unable: True

With support: False

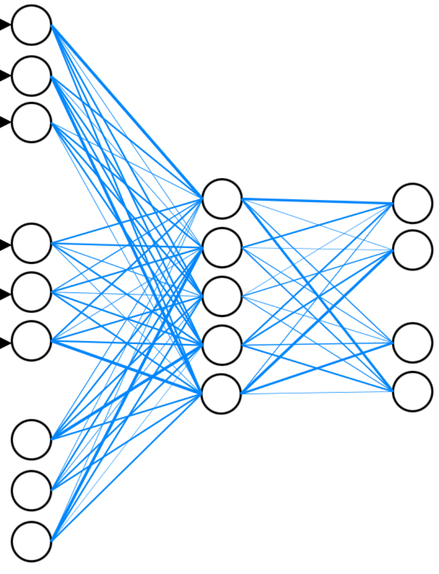
Without support: False

Q2: Can the patient grab a ball?

Unable: False

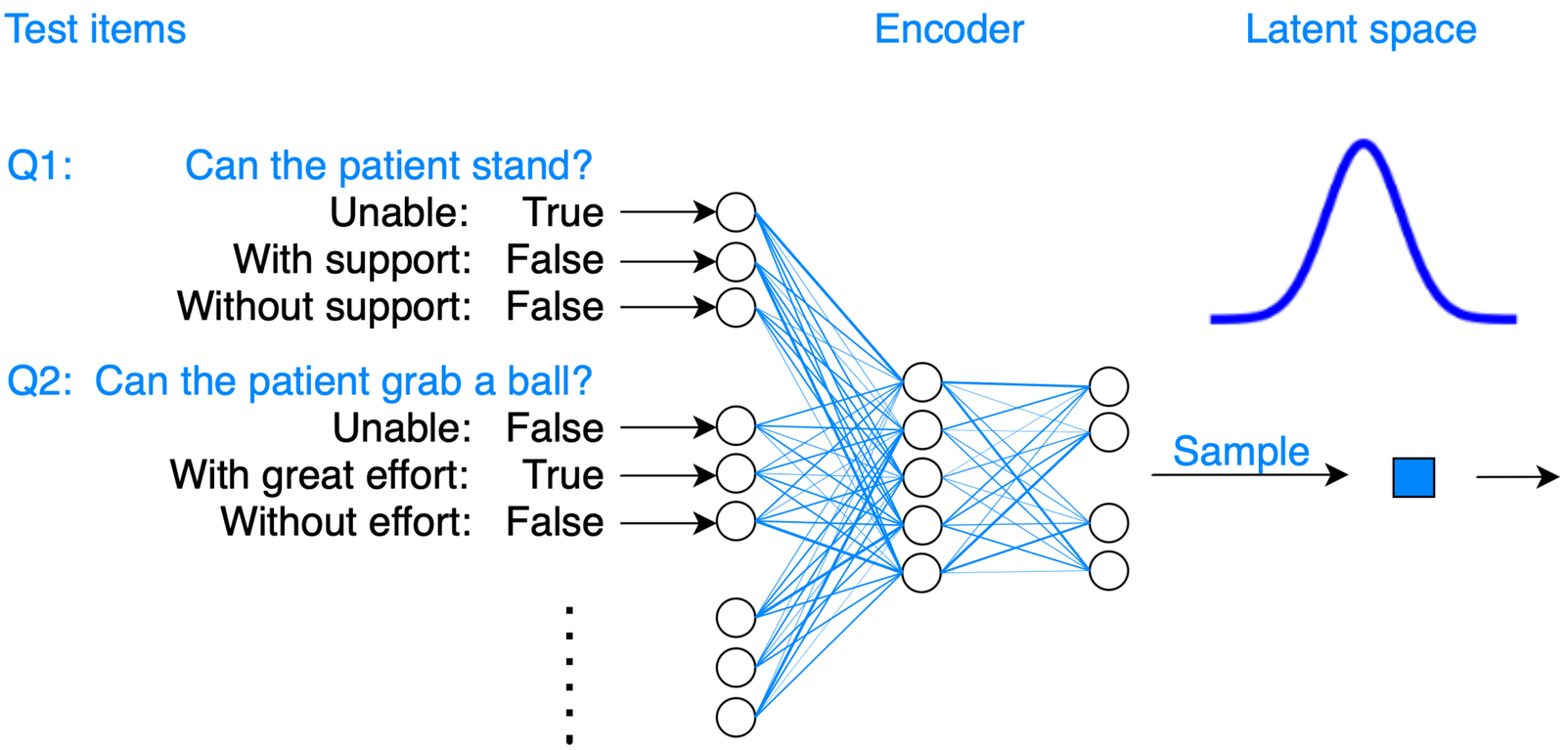
With great effort: True

Without effort: False



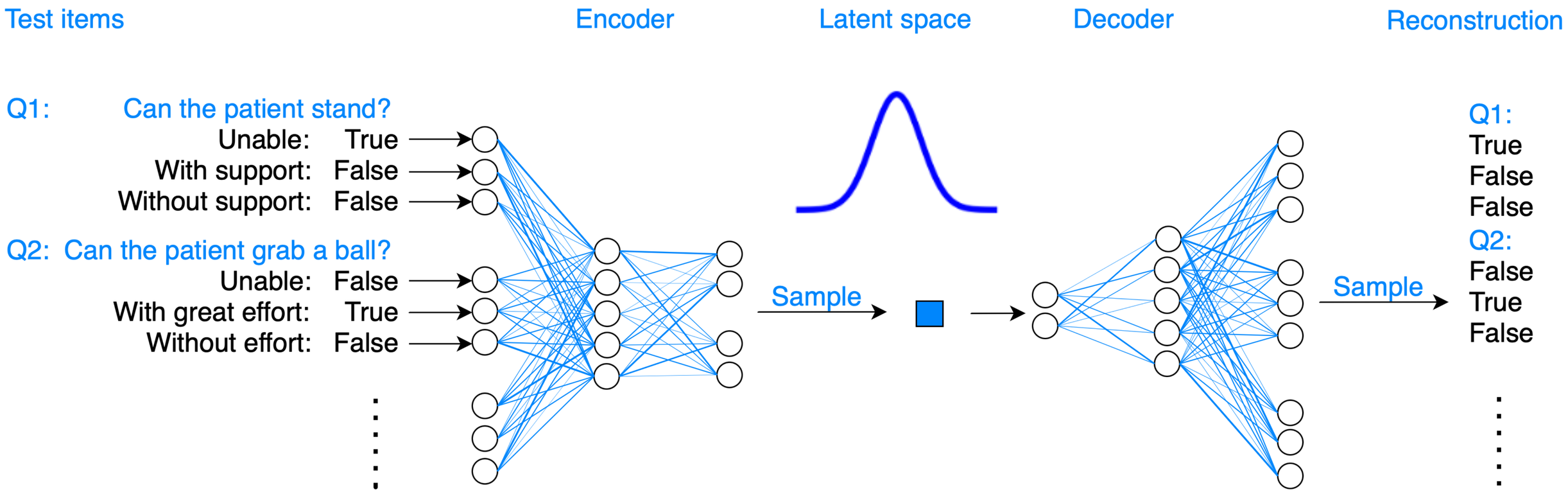
Dimension reduction with VAEs

Items from motor function assessment in SMA



Dimension reduction with VAEs

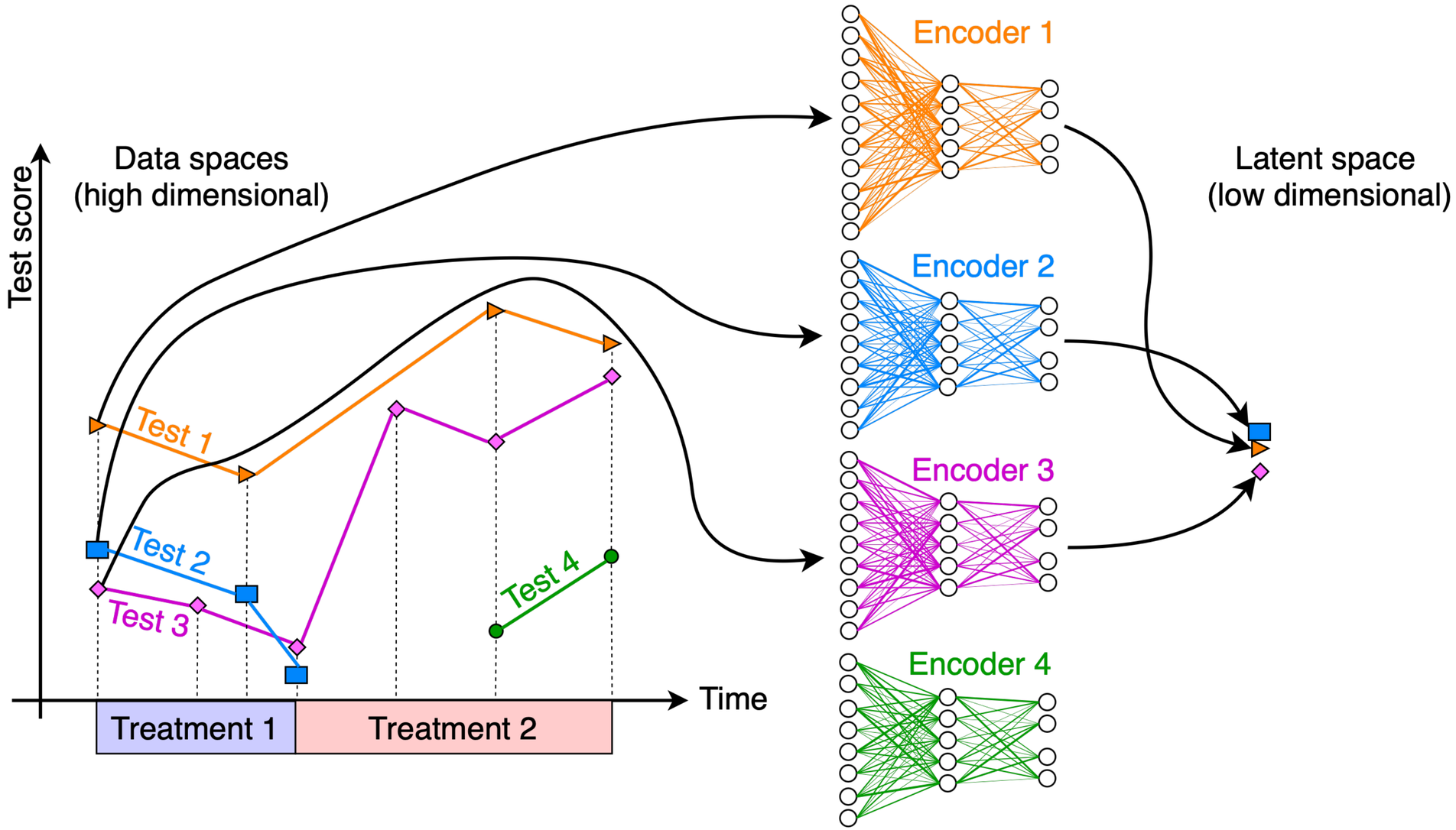
Items from motor function assessment in SMA



UNIVERSITÄTS KLINIKUM FREIBURG

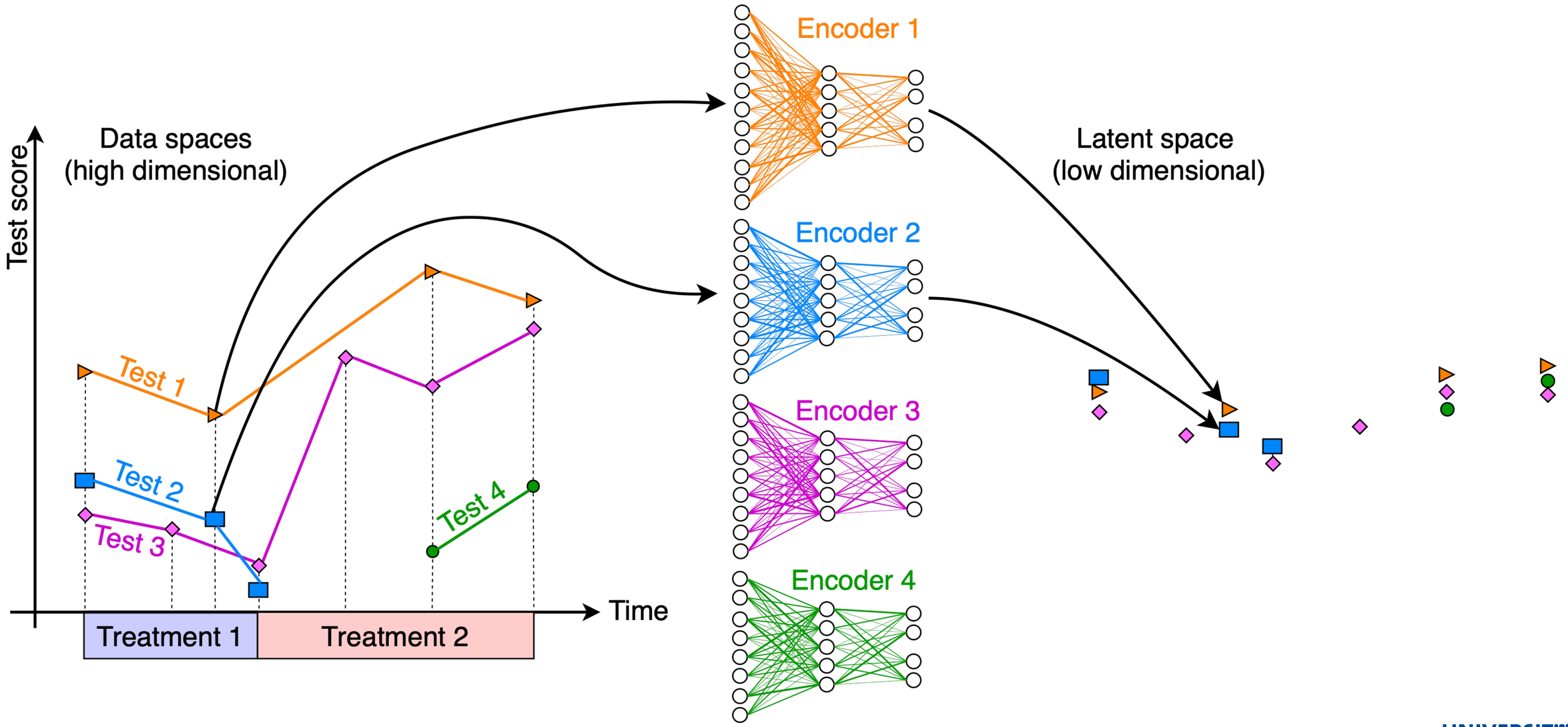
Incorporating several measurement instruments

Mapping to a joint latent space



Incorporating several measurement instruments

Mapping to a joint latent space

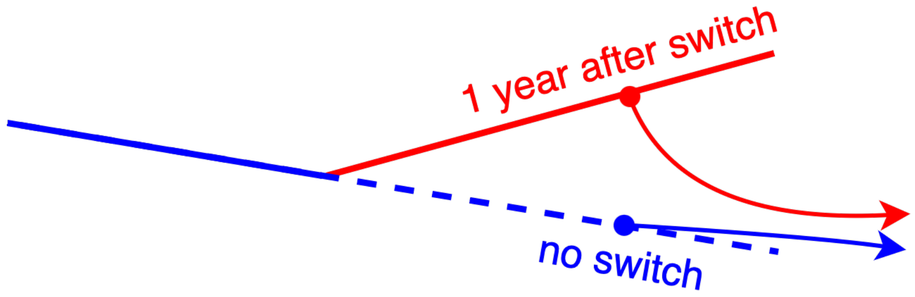


UNIVERSITÄTS KLINIKUM FREIBURG

Incorporating several measurement instruments

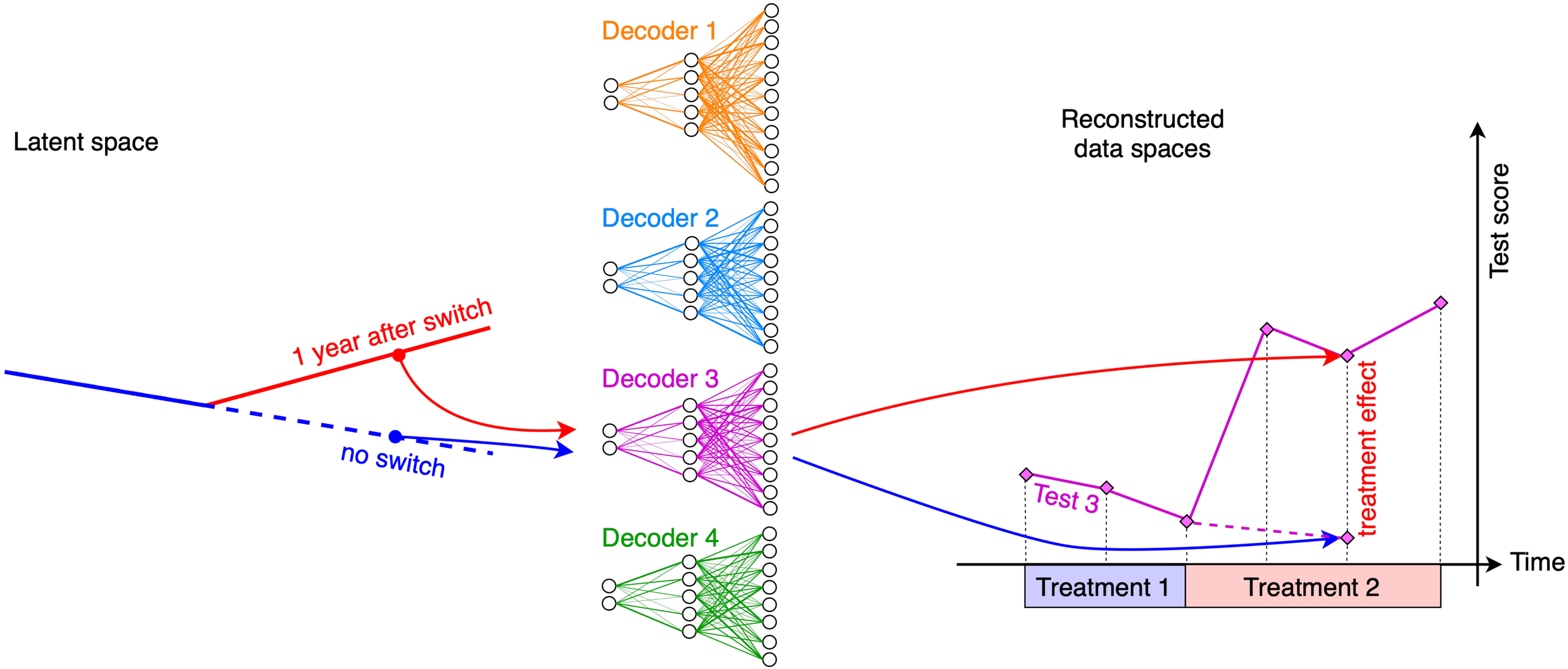
Modeling treatment switches

Latent space



Incorporating several measurement instruments

Modeling treatment switches



Calibrating ODEs with VAEs for synthetic data

Rare disease example: Epodermolysis bullosa (EB)

ODE system:

CRP: $$\frac{dC}{dt} = (r_C + \delta) \bullet \mathrm{C}\,(1 - \frac{C}{K_C}) - \alpha_{CB} \bullet \frac{B}{K_B}$$

Haemoglobin: $$\frac{dH}{dt} = r_H \bullet \mathrm{H}\,(1 - \frac{H}{K_H}) + \alpha_{HB} \bullet \frac{B}{K_B}$$

BMI: $$\frac{dB}{dt} = r_B \bullet \mathrm{B}\,(1 - \frac{B}{K_B}) - \alpha_{CB} \bullet \frac{B}{K_B}$$

Albumin: $$\frac{dA}{dt} = r_A \bullet \mathrm{A}\,(1 - \frac{A}{K_A}) + \alpha_{AB} \bullet \frac{B}{K_B}$$

Iron: $$\frac{dI}{dt} = r_I \bullet \mathrm{I}\,(1 - \frac{I}{K_I}) + \alpha_{IB} \bullet \frac{B}{K_B}$$

Calibrating ODEs with VAEs for synthetic data

Rare disease example: Epodermolysis bullosa (EB)

ODE system:

CRP: $$\frac{dC}{dt} = (r_C + \delta) \bullet \mathrm{C}\,(1 - \frac{C}{K_C}) - \alpha_{CB} \bullet \frac{B}{K_B}$$

Haemoglobin: $$\frac{dH}{dt} = r_H \bullet \mathrm{H}\,(1 - \frac{H}{K_H}) + \alpha_{HB} \bullet \frac{B}{K_B}$$

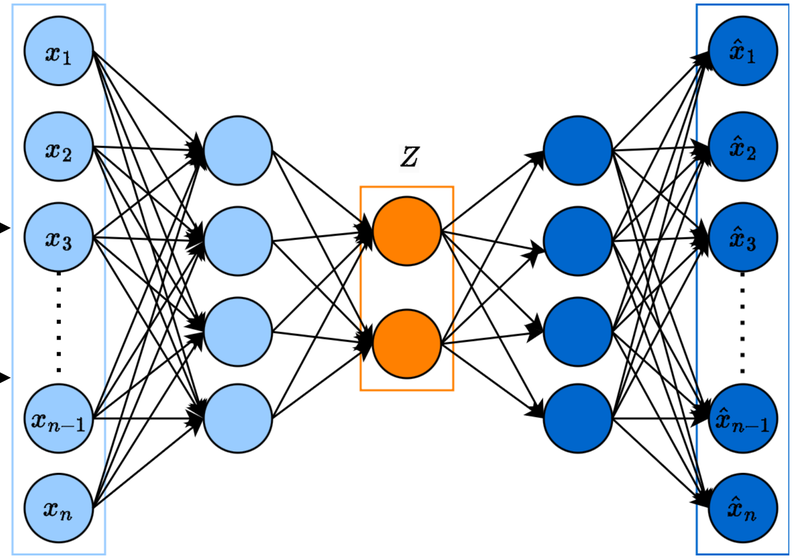
BMI: $$\frac{dB}{dt} = r_B \bullet \mathrm{B}\,(1 - \frac{B}{K_B}) - \alpha_{CB} \bullet \frac{B}{K_B}$$

Albumin: $$\frac{dA}{dt} = r_A \bullet \mathrm{A}\,(1 - \frac{A}{K_A}) + \alpha_{AB} \bullet \frac{B}{K_B}$$

Iron: $$\frac{dI}{dt} = r_I \bullet \mathrm{I}\,(1 - \frac{I}{K_I}) + \alpha_{IB} \bullet \frac{B}{K_B}$$

x_{syn}

x_{real}



UNIVERSITÄTS KLINIKUM FREIBURG

Calibrating ODEs with VAEs for synthetic data

Rare disease example: Epodermolysis bullosa (EB)

ODE system:

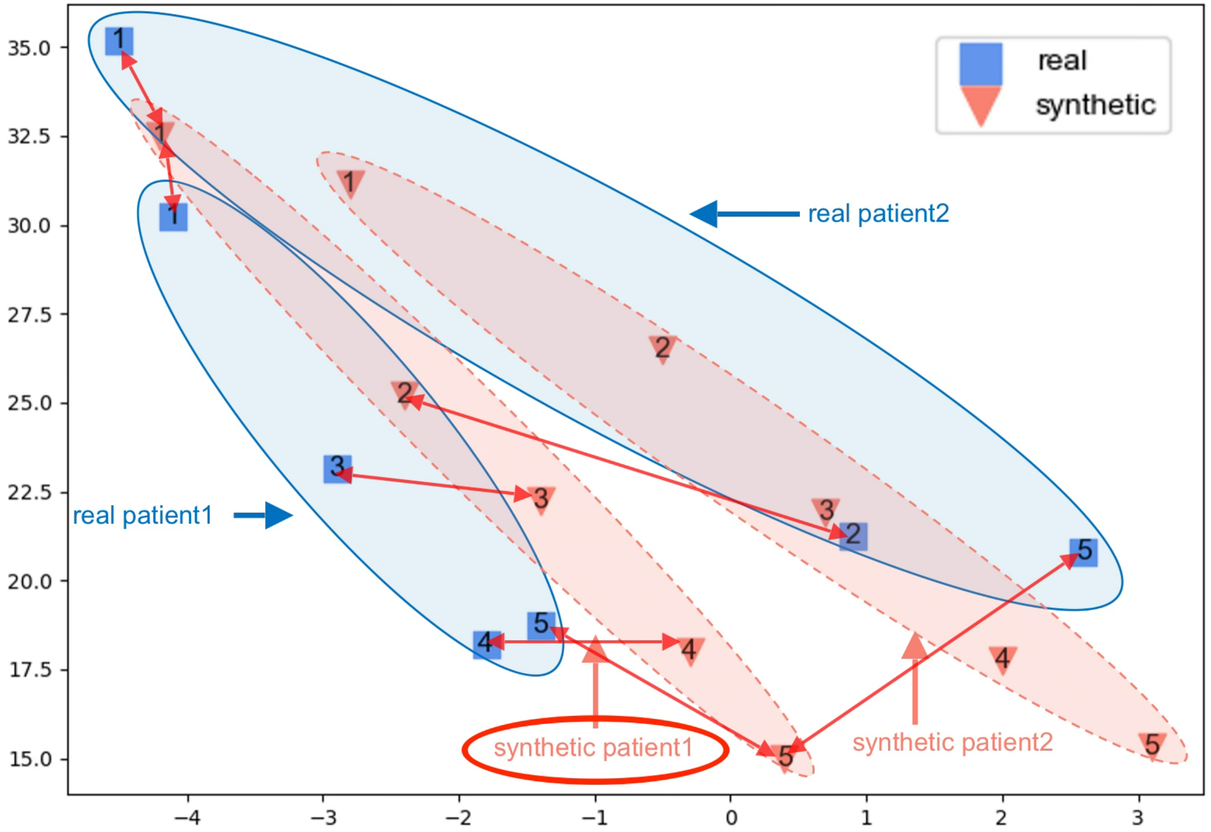
CRP:
$$\frac{dC}{dt} = (r_C + \delta) \bullet \mathrm{C}\,(1 - \frac{C}{K_C}) - \alpha_{CB} \bullet \frac{B}{K_B}$$

Haemoglobin:
$$\frac{dH}{dt} = r_H \bullet \mathrm{H}\,(1 - \frac{H}{K_H}) + \alpha_{HB} \bullet \frac{B}{K_B}$$

BMI:
$$\frac{dB}{dt} = r_B \bullet \mathrm{B}\,(1 - \frac{B}{K_B}) - \alpha_{CB} \bullet \frac{B}{K_B}$$

Albumin:
$$\frac{dA}{dt} = r_A \bullet \mathrm{A}\,(1 - \frac{A}{K_A}) + \alpha_{AB} \bullet \frac{B}{K_B}$$

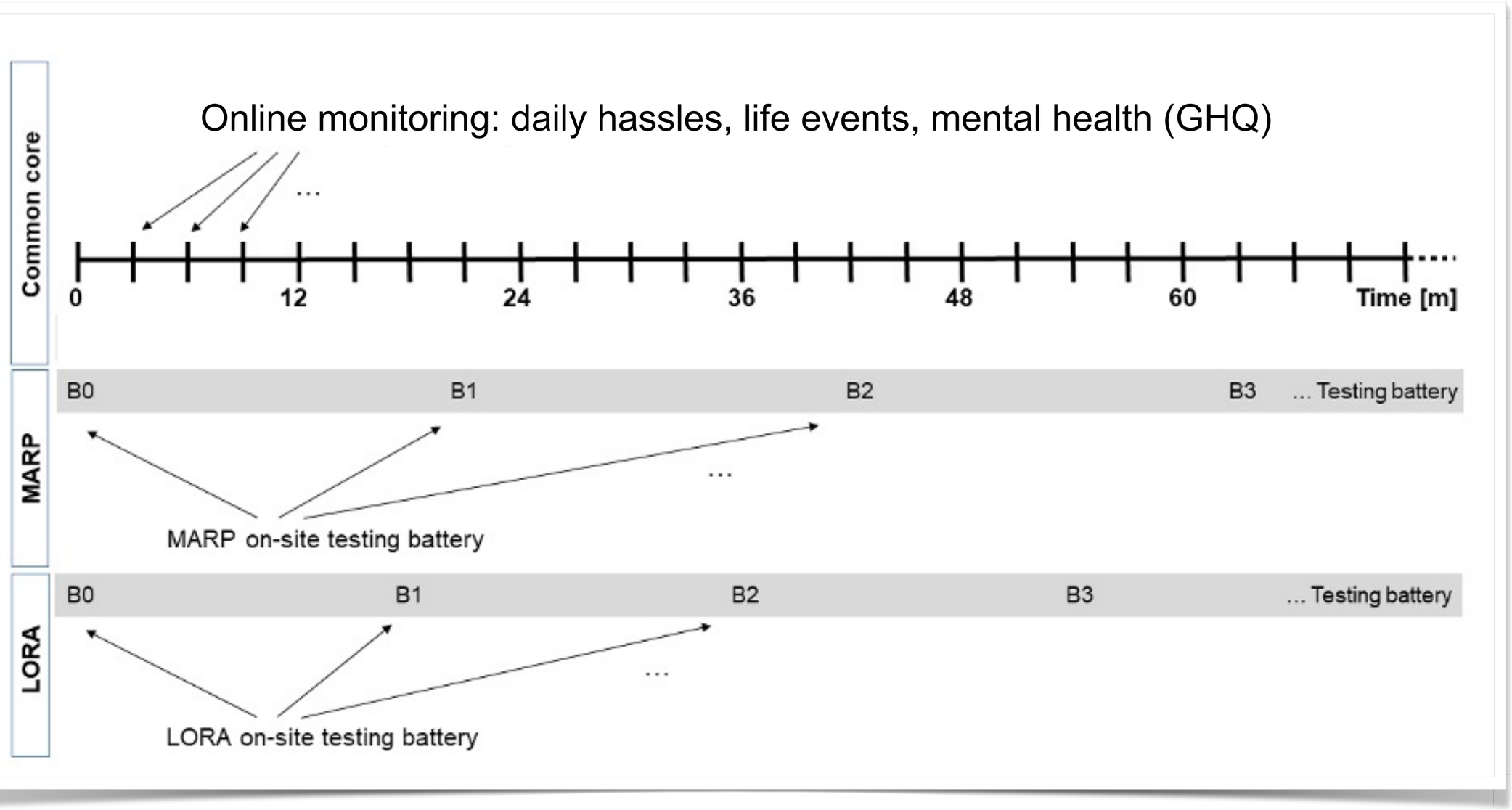
Iron:
$$\frac{dI}{dt} = r_I \bullet \mathrm{I}\,(1 - \frac{I}{K_I}) + \alpha_{IB} \bullet \frac{B}{K_B}$$



UNIVERSITÄTS KLINIKUM FREIBURG

Application: Psychological resilience

DynaMORE project: The MARP and LORA studies



UNIVERSITÄTS KLINIKUM FREIBURG

Turning the data into a sequence of tokens

ID 1

Noise	Noise	Nightmares	Noise	Lack of sleep or sleeping problems	Noise	Severe anxiety and sleep disturbances

ID 2

Noise	Lack of sleep or sleeping problems	Severe anxiety and sleep disturbances

ID 3

Noise	Noise	Noise	Noise	Lack of sleep or sleeping problems	Noise	Noise

ID 4

Noise	Conflict with close person	Conflict with close person	Job loss	Lack of sleep or sleeping problems	Financial problems	Severe anxiety and sleep disturbances

The attention mechanism

- Starting point: sequence of embedding vectors $x_i, i = 1,\ldots n$ (maybe with position information added)

The attention mechanism

- Starting point: sequence of embedding vectors $x_i, i = 1,\ldots n$ (maybe with position information added)
- Key-value representation:

$$k_i = W_k x_i \quad \text{key tokens}$$
$$v_i = W_v x_i \quad \text{value tokens}$$
$$q_i = W_q x_i \quad \text{query tokens}$$

The attention mechanism

- Starting point: sequence of embedding vectors $x_i, i = 1,\ldots n$ (maybe with position information added)
- Key-value representation:

$$k_i = W_k x_i \quad \text{key tokens}$$
$$v_i = W_v x_i \quad \text{value tokens}$$
$$q_i = W_q x_i \quad \text{query tokens}$$

- Updating the embedding vectors:

$$h_i = \sum_{j=1}^{n} \text{softmax}_j \left(g(q_i, k_j) \right) \cdot v_j$$

Generating synthetic data by continuing sequences

In the first visit, the participant reported, [conflict with a close person], [severe argument with partner,] [job loss,] [end.] In the next visit, the participant reported [looking for a new job,] [conflict with a close person,] [meeting,] [end.] In the next visit the participant reported [conflict with a close person,] [severe argument with partner,] [lack of sleep,] [end.]

Generating synthetic data by continuing sequences

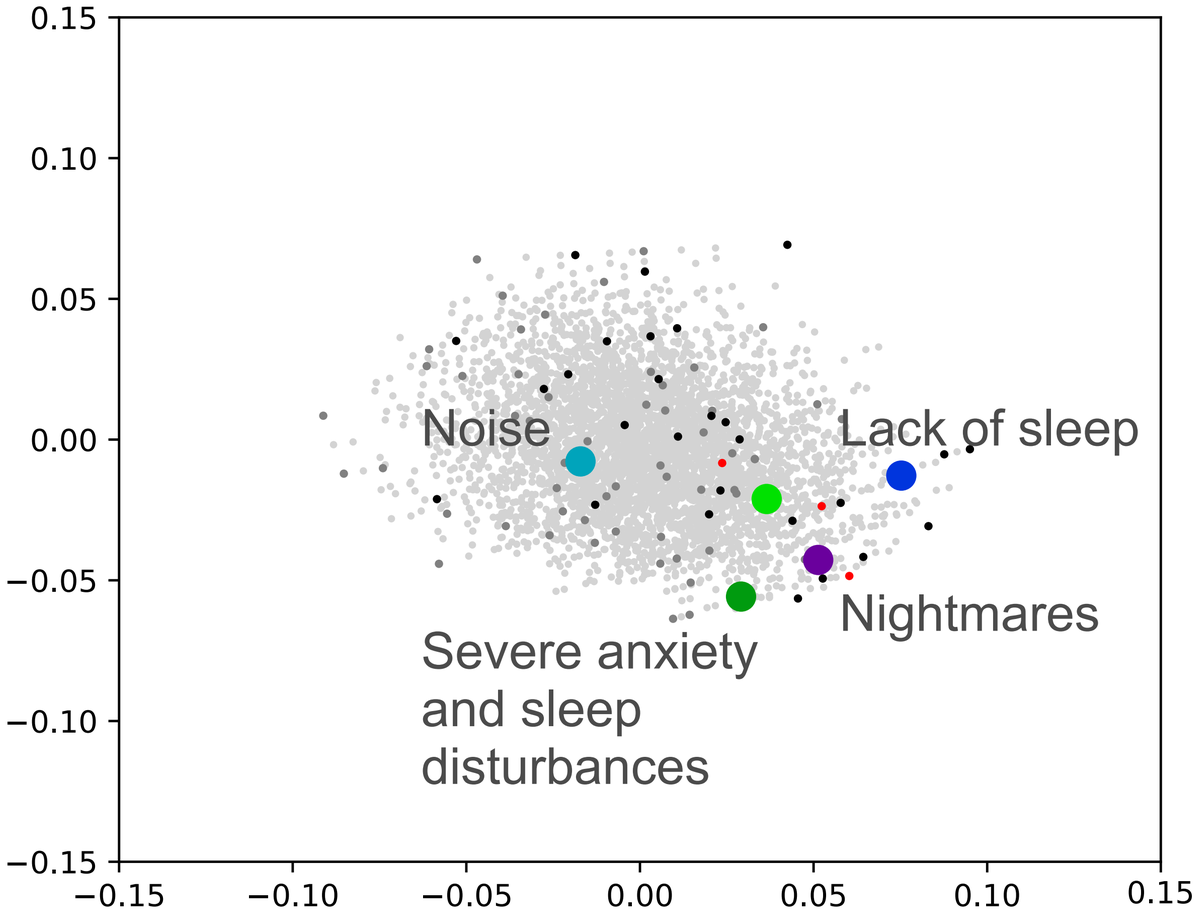
In the first visit, the participant reported, [conflict with a close person], [severe argument with partner,] [job loss,] [end.] In the next visit, the participant reported [looking for a new job,] [conflict with a close person,] [meeting,] [end.] In the next visit the participant reported [conflict with a close person,] [severe argument with partner,] [lack of sleep,] [end.]

In the next visit the participant reported [looking for a new job,][financial problems,] [conflict with a close person,][time pressure,][end.]

Borrowing label embeddings from a large language model (LLM)

Dimension reduction using a variational autoencoder

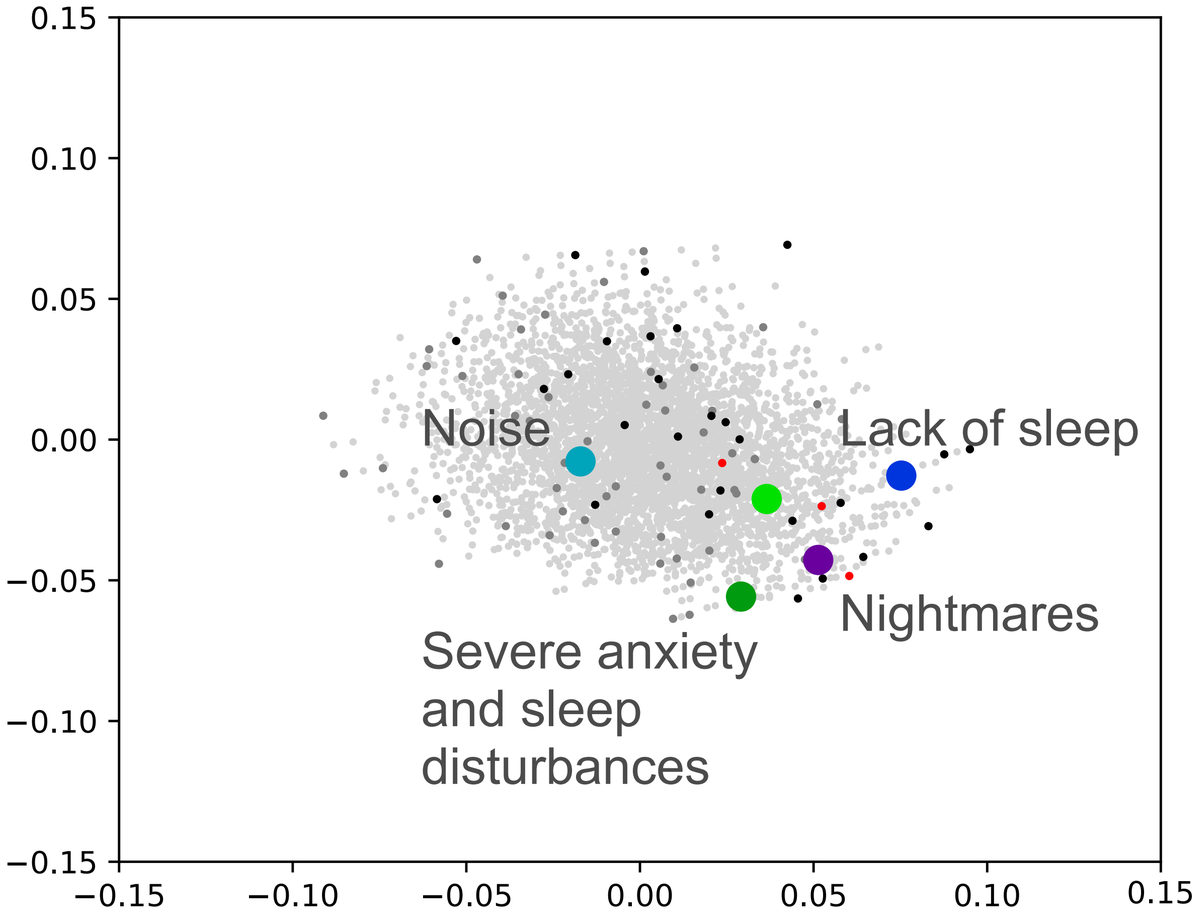
Averaging the embedding of “Nightmares”, “lack of sleep”, and “Noise”



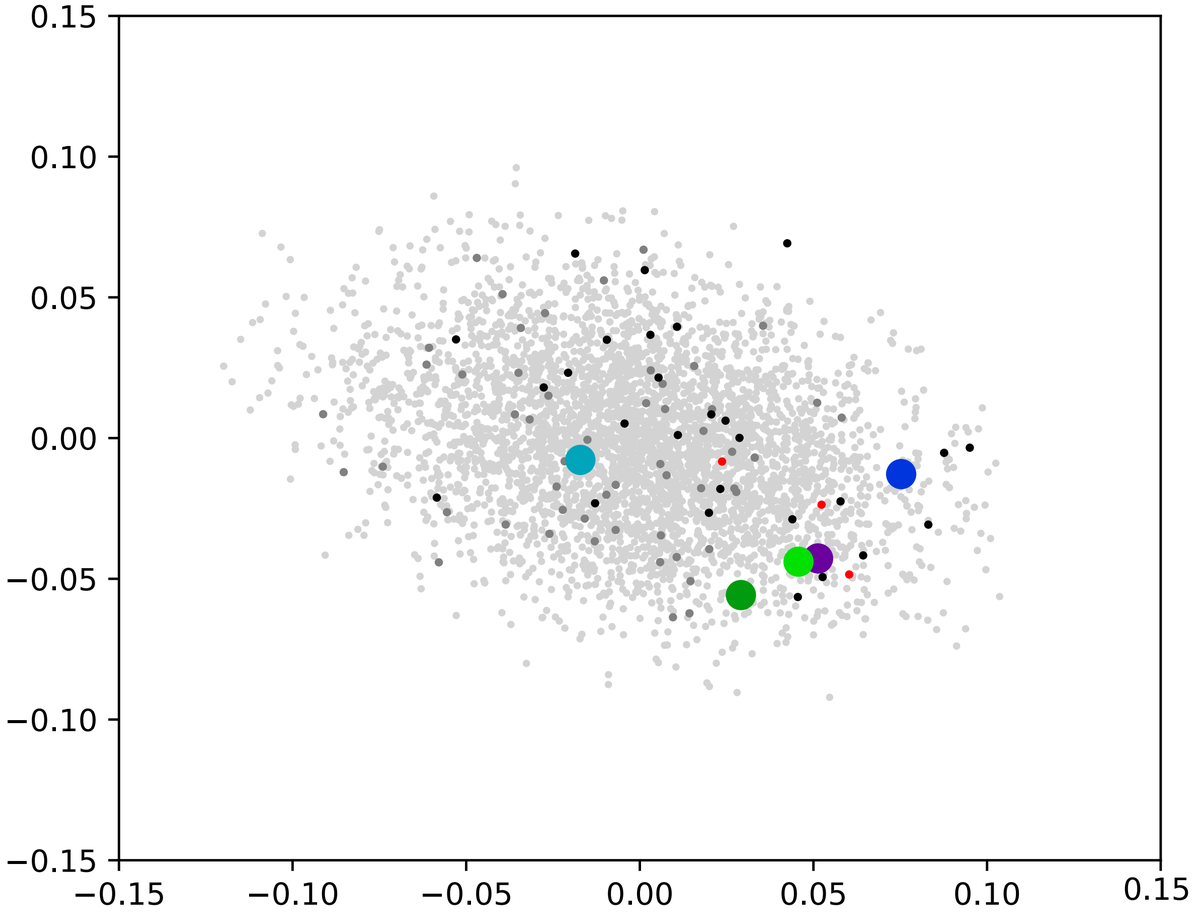
Borrowing label embeddings from a large language model (LLM)

Dimension reduction using a variational autoencoder

Averaging the embedding of “Nightmares”, “lack of sleep”, and “Noise”



Using “Nightmares” and “Lack of sleep” as a context for “Noise”



UNIVERSITÄTS KLINIKUM FREIBURG

Thanks to …

Members of my group who did the actual work:
Max Behrens, Kiana Farhadyar, Maren Hackenberg, Michelle Pfaffenlehner, Clemens Schächter, Hanning Yang

DFG Deutsche Forschungsgemeinschaft

small data
MORE FROM LESS
CRC 1597

Contact:
Harald Binder
Institute of Medical Biometry and Statistics (IMBI)
Faculty of Medicine and Medical Center — University of Freiburg, Germany
harald.binder@uniklinik-freiburg.de

UNIVERSITÄTS KLINIKUM FREIBURG