

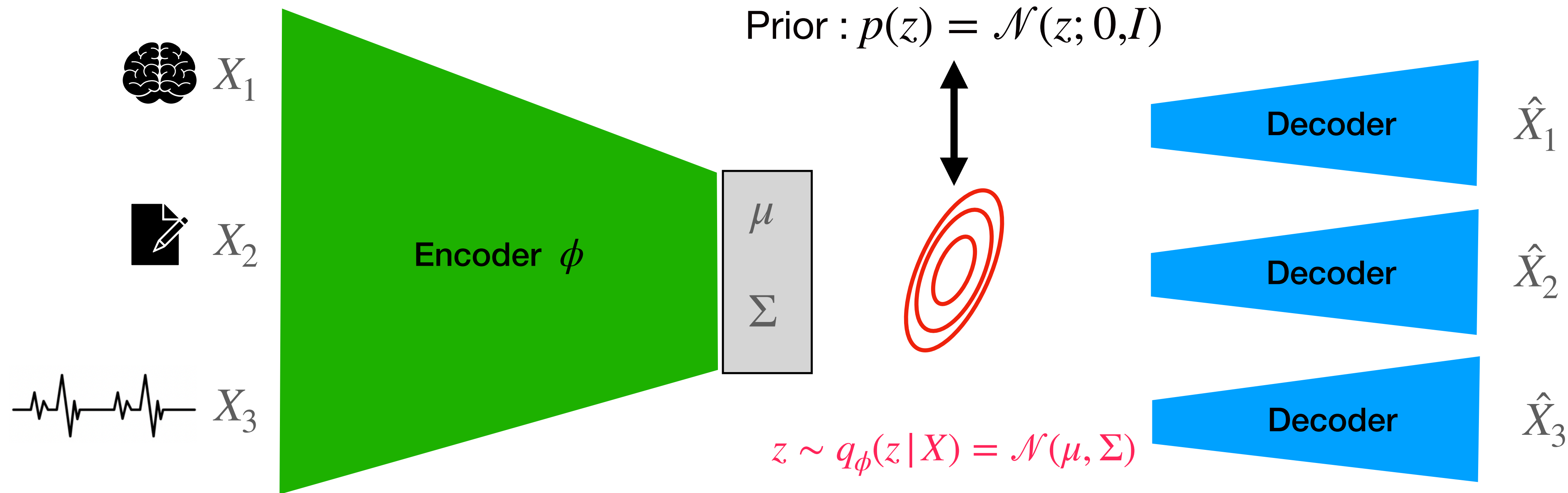
Multimodal Variational Auto-encoders

Interactive tutorial

Agathe Senellart — PhD student with Stéphanie Allasonnière

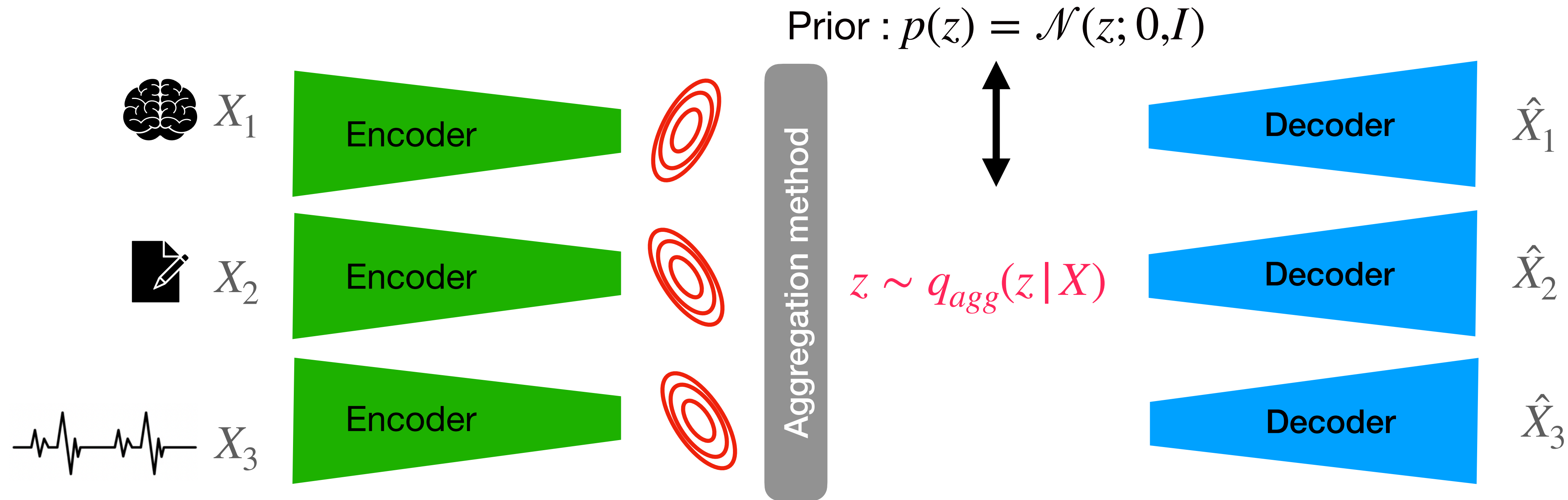
AI+Science Summer School 2025

Multimodal Variational Auto-encoders



$$\mathcal{L} = \underbrace{\mathbb{E}_{q_\phi(z|X)}[\ln p_\theta(X|z)]}_{\text{Reconstruction}} - \underbrace{KL(q_\phi(z|x) || p(z))}_{\text{Regularization}}$$

Multimodal Variational Auto-encoders



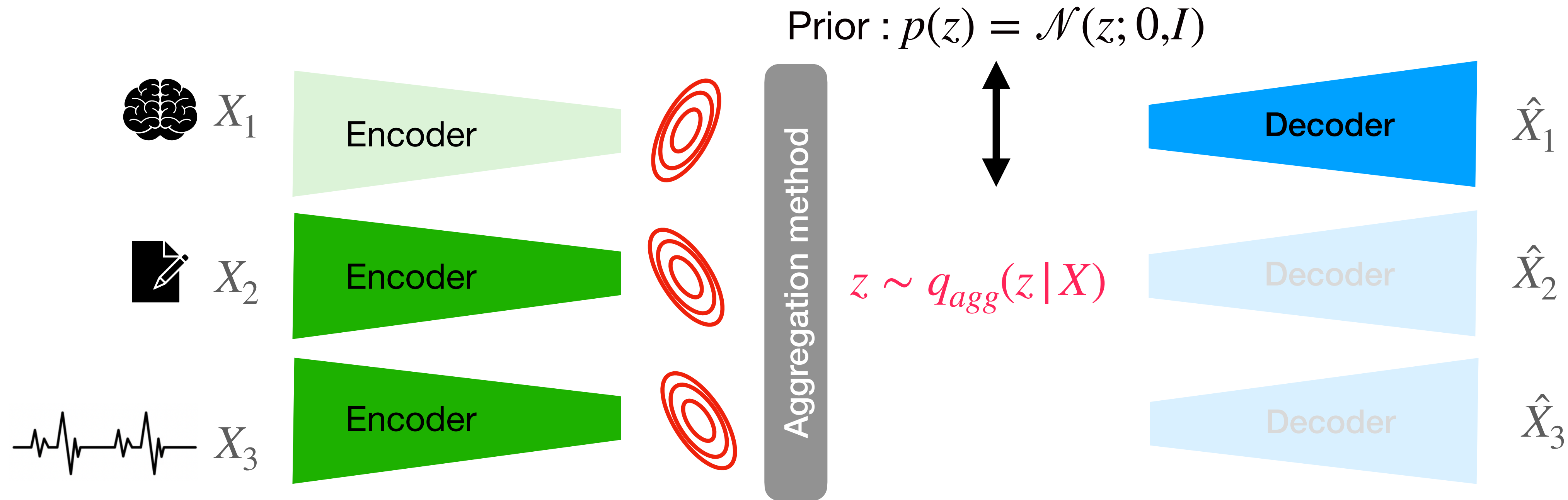
At inference, we can use
observed modalities to
generate a missing one

$$q_{agg}(z | X) \propto \sum_i q_i(z | X_i) \quad \text{Mixture-of-Experts}$$

$$q_{agg}(z | X) \propto \prod_i q_i(z | X_i) \quad \text{Product of Experts}$$

$$q_{agg}(z | X) \propto \sum_S \prod_{i \in S} q_i(z | X_i) \quad \text{Mixture-of-Product-of Experts}$$

Multimodal Variational Auto-encoders



At inference, we can use
observed modalities to
generate a missing one

$$q_{agg}(z | X) \propto \sum_i q_i(z | X_i) \quad \text{Mixture-of-Experts}$$

$$q_{agg}(z | X) \propto \prod_i q_i(z | X_i) \quad \text{Product of Experts}$$

$$q_{agg}(z | X) \propto \sum_S \prod_{i \in S} q_i(z | X_i) \quad \text{Mixture-of-Product-of Experts}$$

With these models, we can do:

- Missing data imputation
- Multimodal Data Augmentation
- Use the joint representation for clustering, classification ...

A (non-exhaustive) list of methods

Used in this
Tutorial

Modèle	Aggregation	Comments
MVAE	Product-of-Experts	
MMVAE	Mixture-of-Experts	
MoPoE	Mixture-of-Product of Experts	
MVTCAE	Product-of-Experts	Additional terms to the loss
CRMVAE	Product-of-Experts	
DMVAE	Product-of-Experts	Additional latent spaces
MMVAE+	Mixture-of-Experts	
Nexus	Joint Network	Hierarchical models
MHVAE	Product-of-Experts	
CMVAE	Mixture-of-Experts	
JMVAE, JNF, TELBO	Joint Network	Uses a joint and unimodal encoders

- We will use the python library :

<https://multivae.readthedocs.io>

- Get the tutorial notebook here :

https://github.com/AgatheSenellart/summer_school_AI_Science_2025

- Download the notebook and run it locally
- Or run it on Google Colab

- Don't hesitate to ask any questions !



A few words for the end

Take home messages:

- Multimodal Variational Autoencoders are models that can be useful for **missing data imputation**, **data augmentation** and **learning multimodal representations**
- For generating fully synthetic samples, it can be very powerful to model the latent space of the variational auto encoder a posteriori.
- Useful libraries for VAE : Pythae and MultiVae

A few examples of medical applications:

Generating synthetic MRI from ultrasounds [Dorent et al 2023] / Unsupervised anomaly detection [Aguila et al 2023]

