A deep learning approach to private data sharing of medical images using conditional generative adversarial networks

Hanxi Sun, Jason Plawinski, **Sajanth Subramaniam**, Amir Jamaludin, Timor Kadir, Aimee Readie, Gregory Ligozio, David Ohlssen, Mark Baillie, Thibaud Coroller

NOVARTIS | Reimagining Medicine

BIG DATA INSTITUTE | UNIVERSITY OF OXFORD

A look at modern foundation models

BERT

Text classification

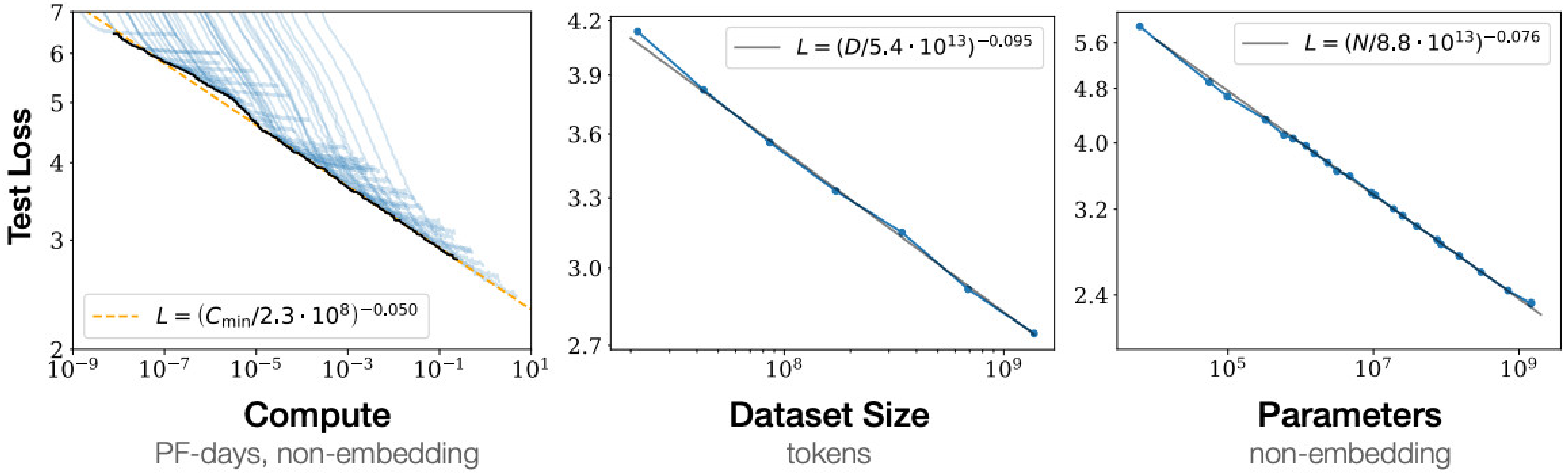
GPT

Text generation

DALL-E

Image generation

A look at modern foundation models



Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, & Dario Amodei. (2020). Scaling Laws for Neural Language Models.

What about the *medical* domain?

Problem

Use of medical data often restricted due to **privacy concerns**

Potential solutions

Anonymization

Remove or modify potentially identifying features from the data

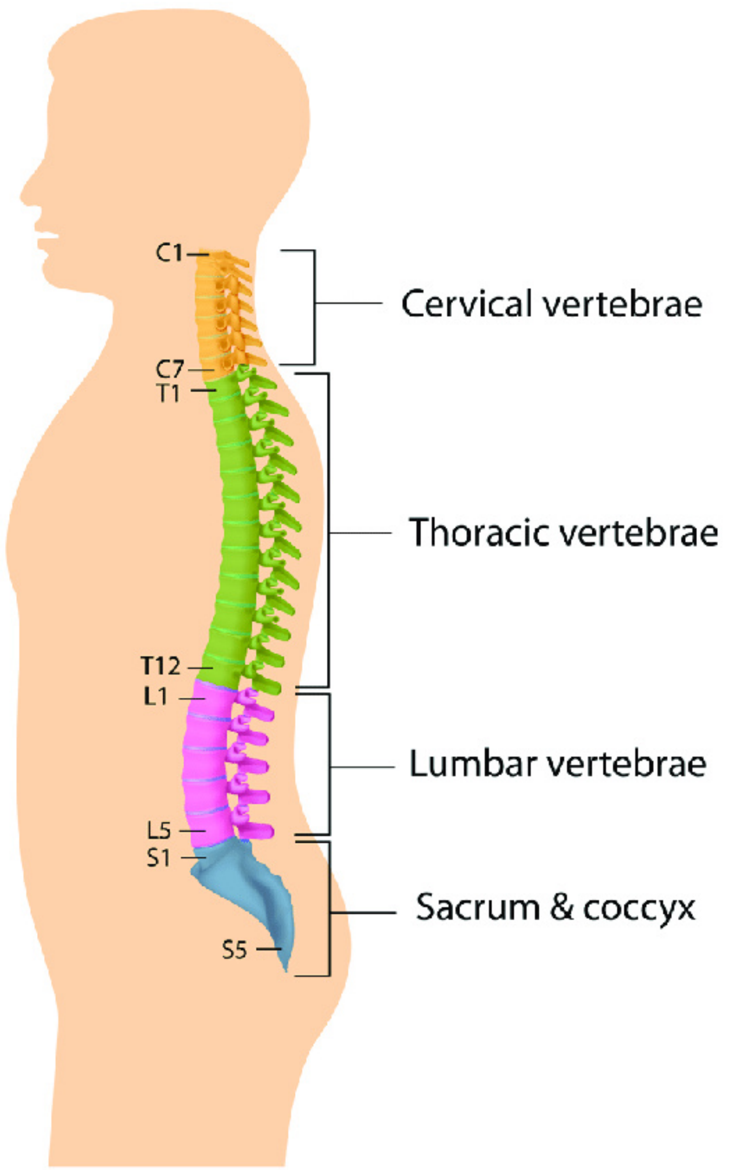
Federated learning

Training of a centralized model by multiple parties without sharing data

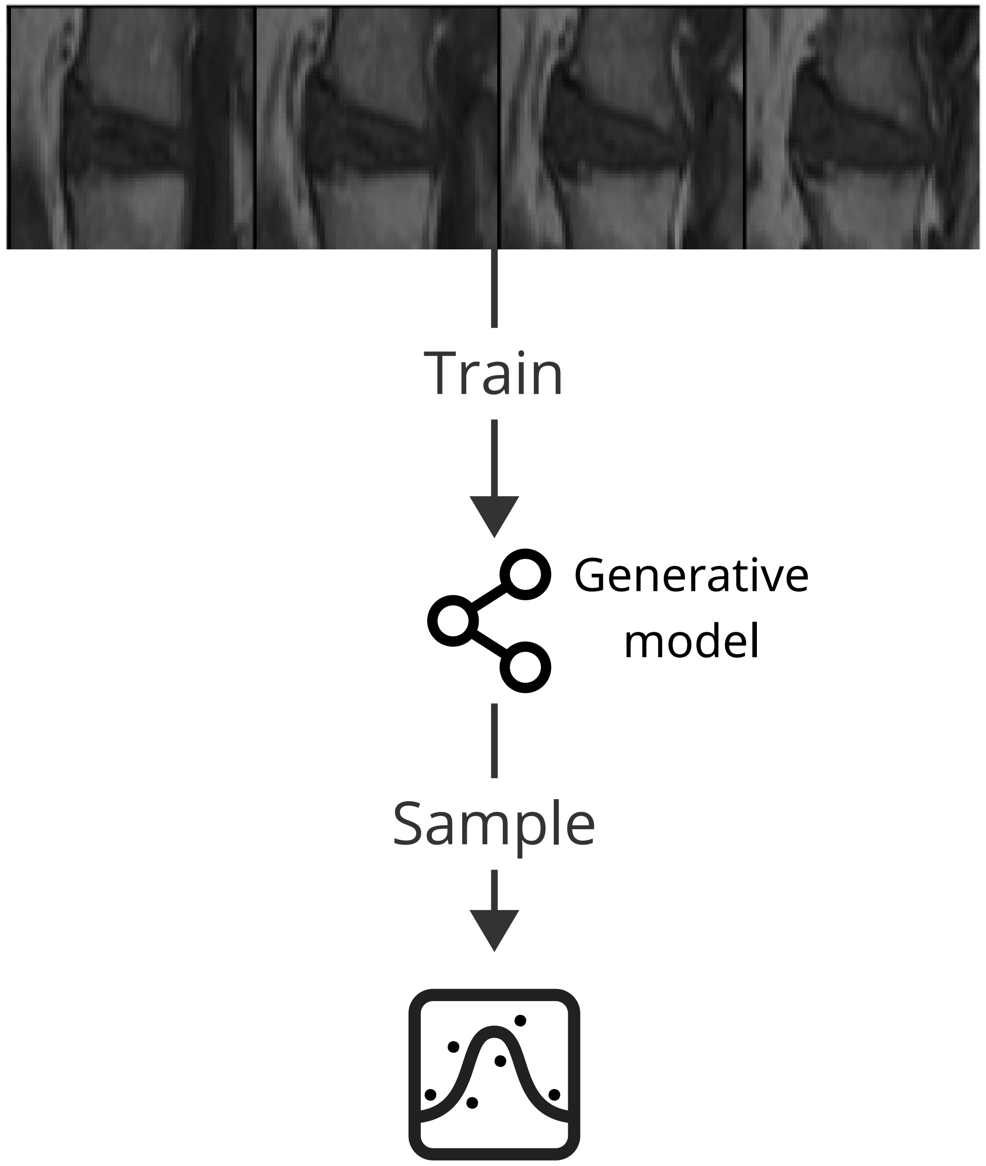
Synthetic data

Generate synthetic data that closely models the original dataset without revealing patient information

Case study: MRI images of vertebral units (VUs)



MRIs of VUs from clinical COSENTYX® study on Ankylosing spondylitis (AS)
, ~10'000 samples

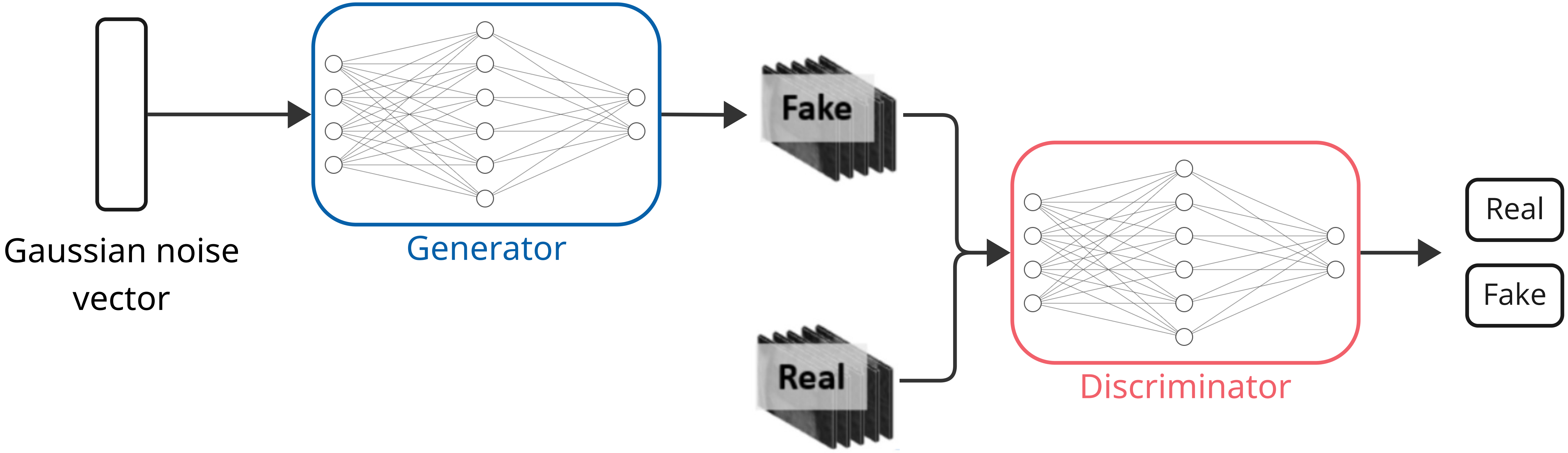


Synthetic MRI images of VUs

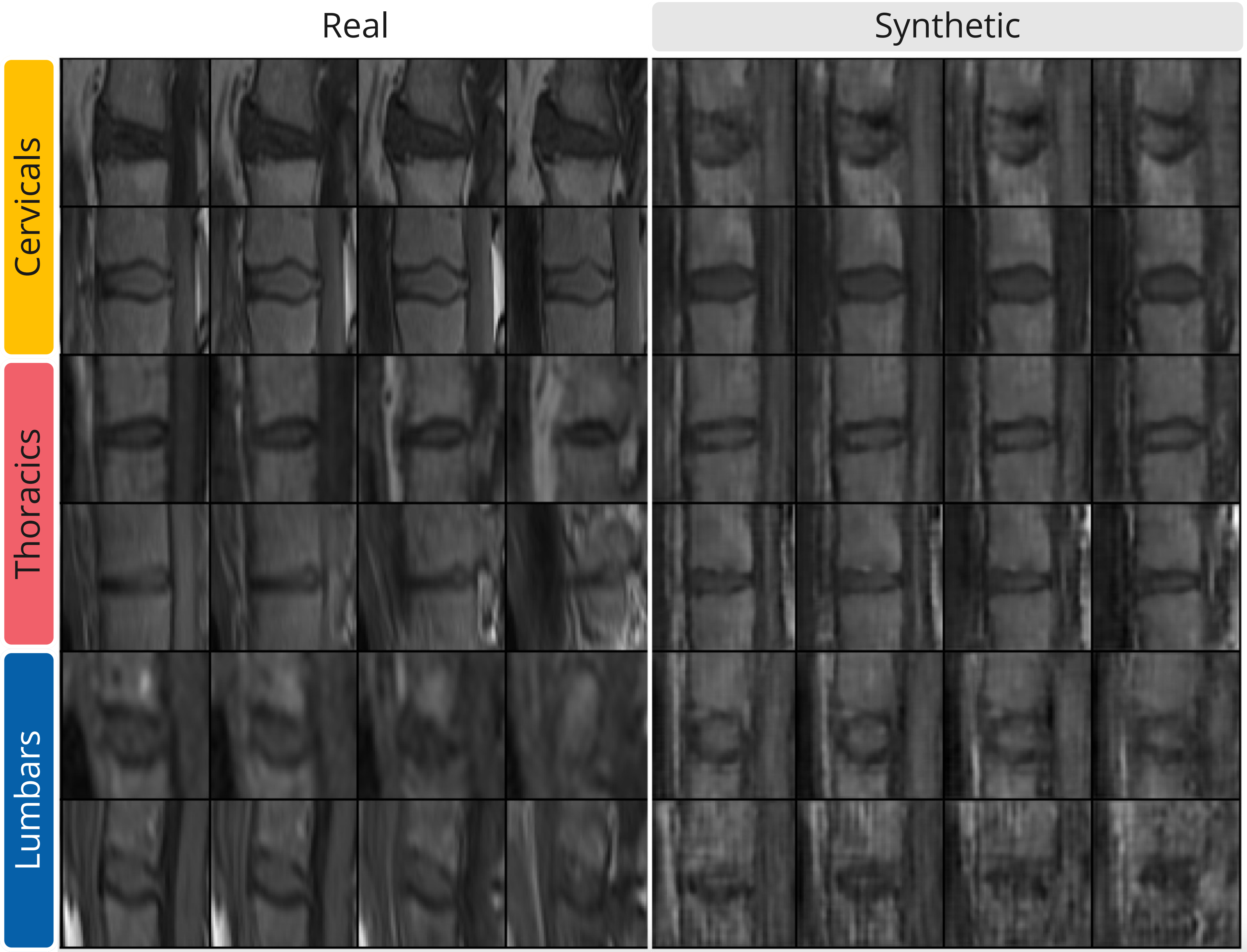
Berlin, C., & Adams, C. (2017). *Production Ergonomics: Designing Work Systems to Support Optimal Human Performance*. Ubiquity Press. http://www.jstor.org/stable/j.ctv3t5qtf

Generative adversarial network (GAN)

Idea: Let two neural networks compete with each other in creating and identifying fake images.



Results

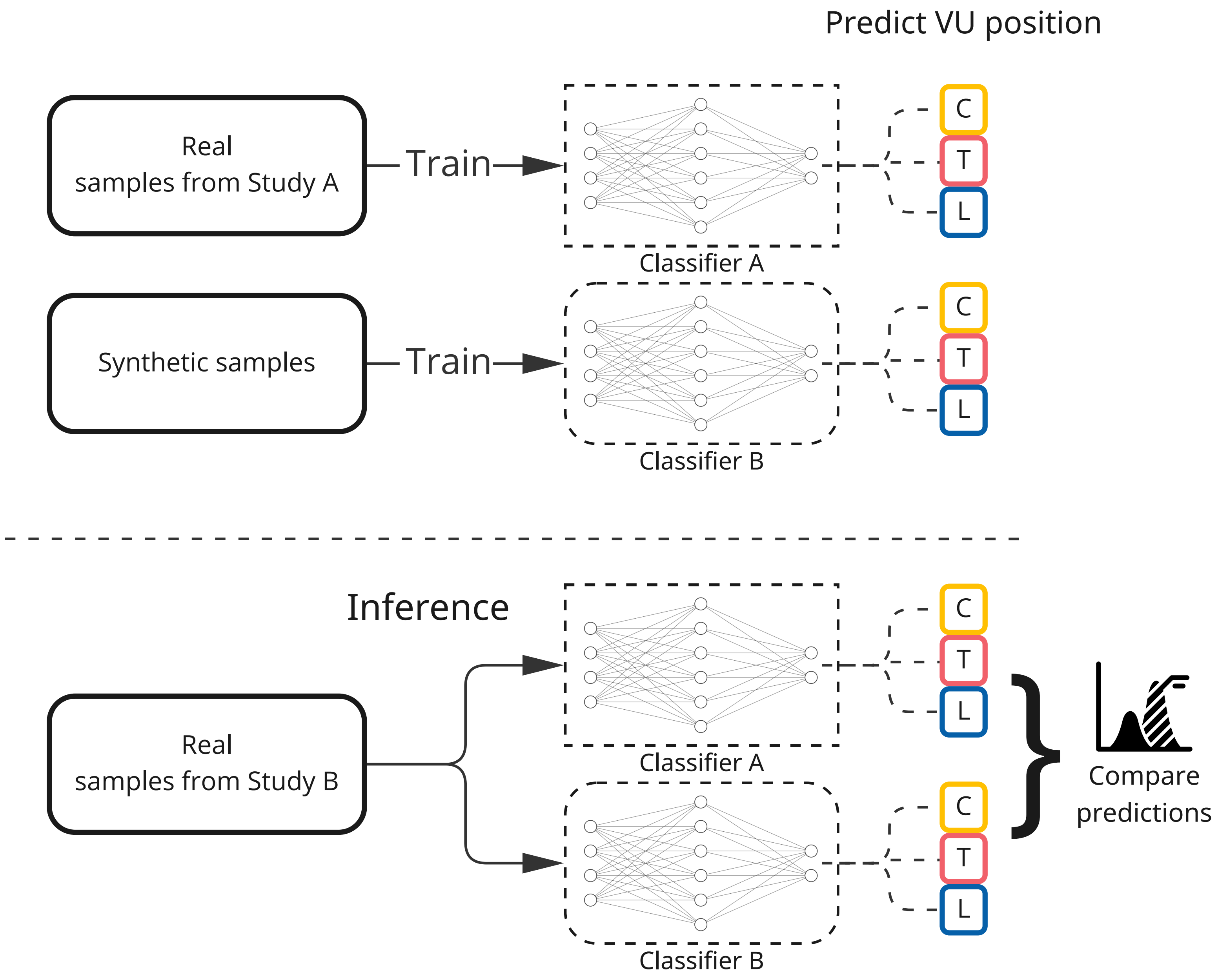


Evaluation

Compare UMAP embeddings
between real and synthetic samples

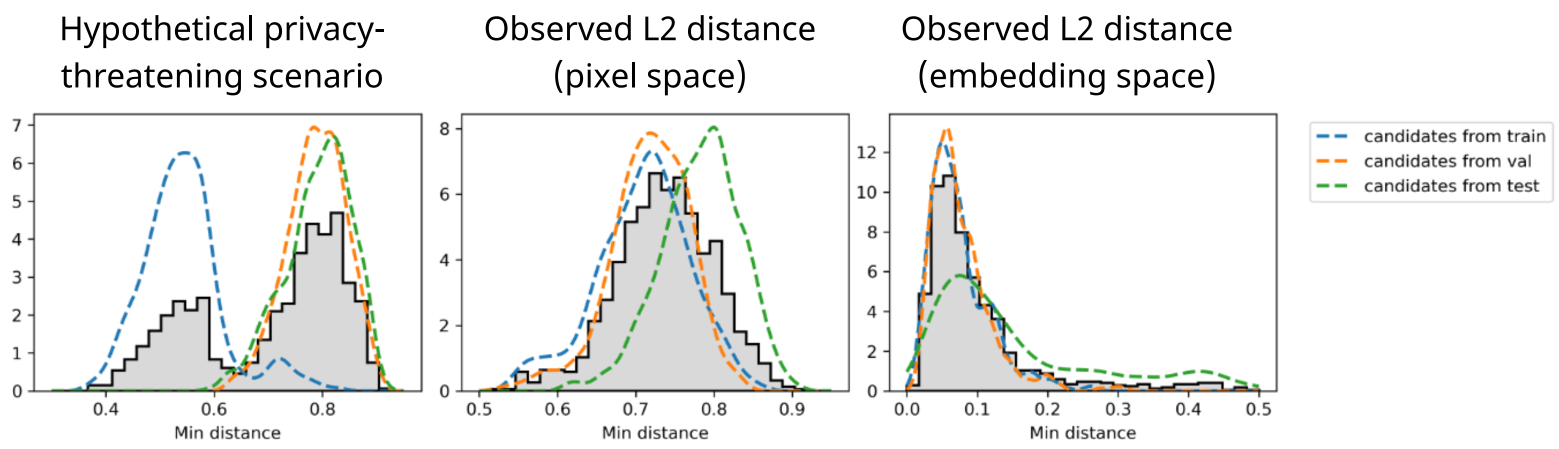
Evaluation

Compare UMAP embeddings between real and synthetic samples



Train two separate classifiers, one on real and one on synthetic data, and compare performance

Privacy evaluation

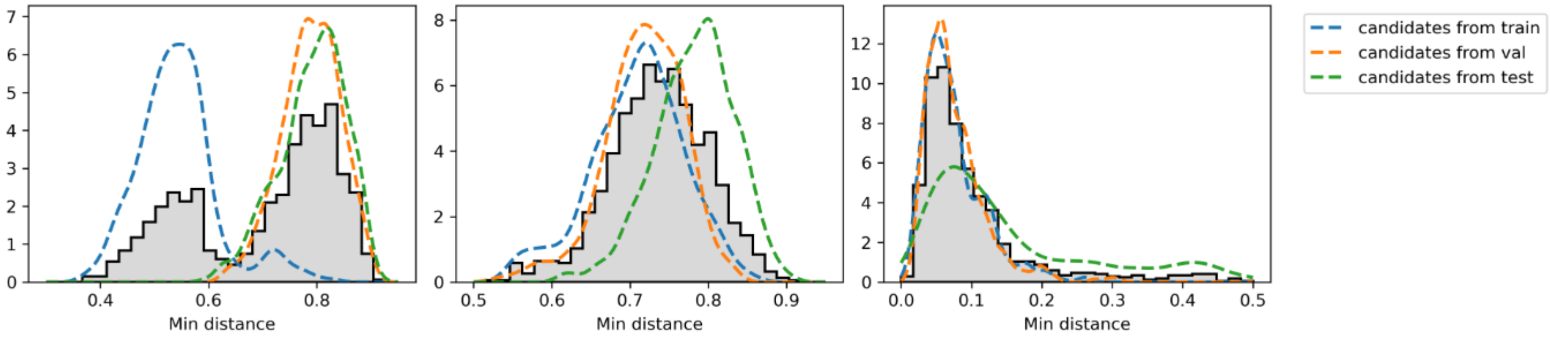


Pairwise attacks

Can we tell whether a given sample was used during training of the model?

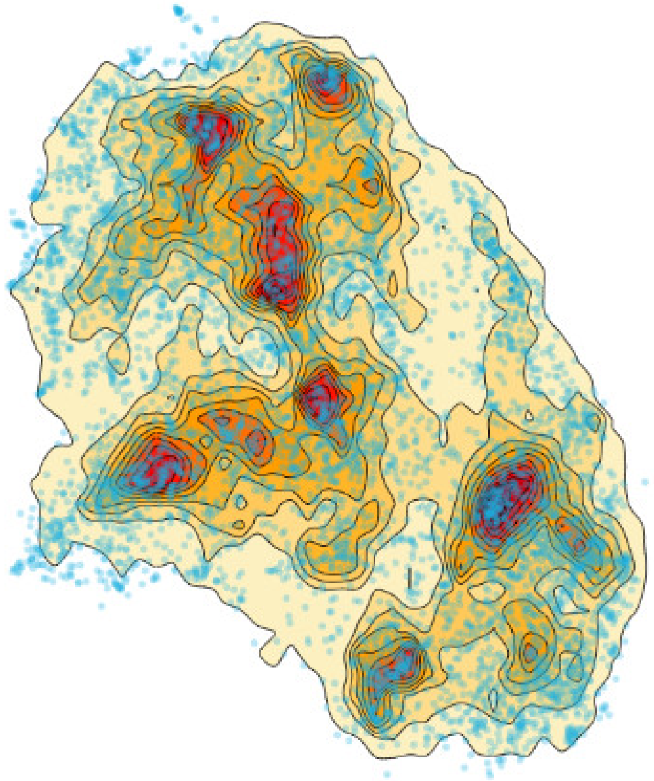
Privacy evaluation

Hypothetical privacy-threatening scenario | Observed L2 distance (pixel space) | Observed L2 distance (embedding space)



Pairwise attacks

Can we tell whether a given sample was used during training of the model?



Synthetic (red/orange) and train (blue dots) samples in embedding space.

Distribution attacks

Can we identify clusters of synthetic images around real images?

Limitations

No privacy guarantee

Incorporate differential privacy (DP) methods, see e.g. [1]

GAN training unstable

Explore diffusion models for diverse, high-quality images, see e.g. [1]

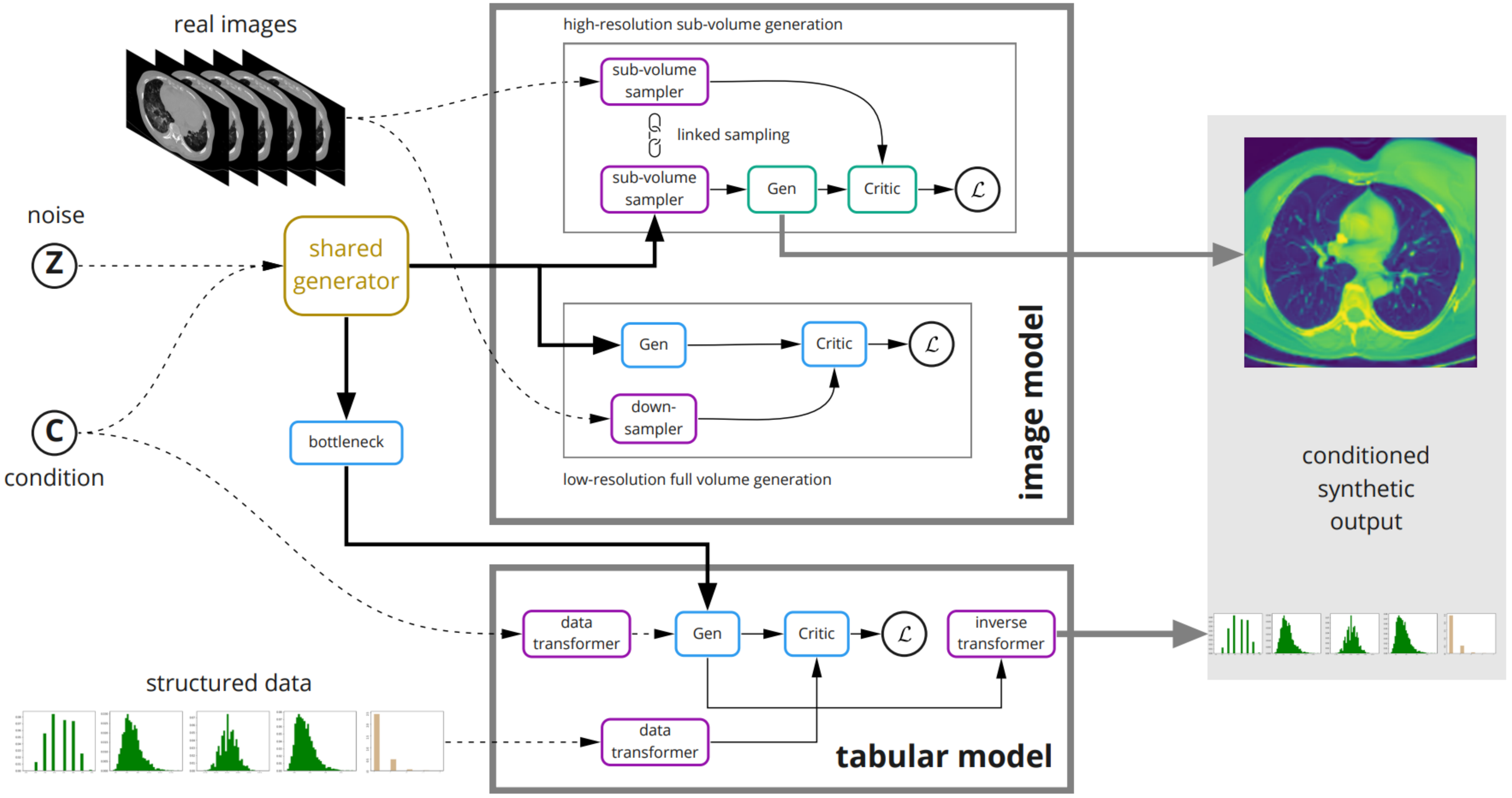
[1] Deniz Daum, Richard Osuala, Anneliese Riess, Georgios Kaissis, Julia A. Schnabel, & Maxime Di Folco. (2024). On Differentially Private 3D Medical Image Synthesis with Controllable Latent Diffusion Models.

Conclusion

- We explored **synthetic data** as a potential remedy for **privacy concerns** in the medical domain
- Deep learning methods like **GANs** can create realistic synthetic images without replicating patients from the original distribution
- Avenues for improvement include better privacy guarantees and more scalable model training

Backup: Follow-up work

Idea: Extend approach to multi-modal data (clinical and imaging)



Jonathan David Ziegler, Sajanth Subramaniam, Michela Azzarito, Orla Doyle, Peter Krusche, & Thibaud Coroller (2022). Multi-Modal Conditional GAN: Data Synthesis in the Medical Domain. In *NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research*.

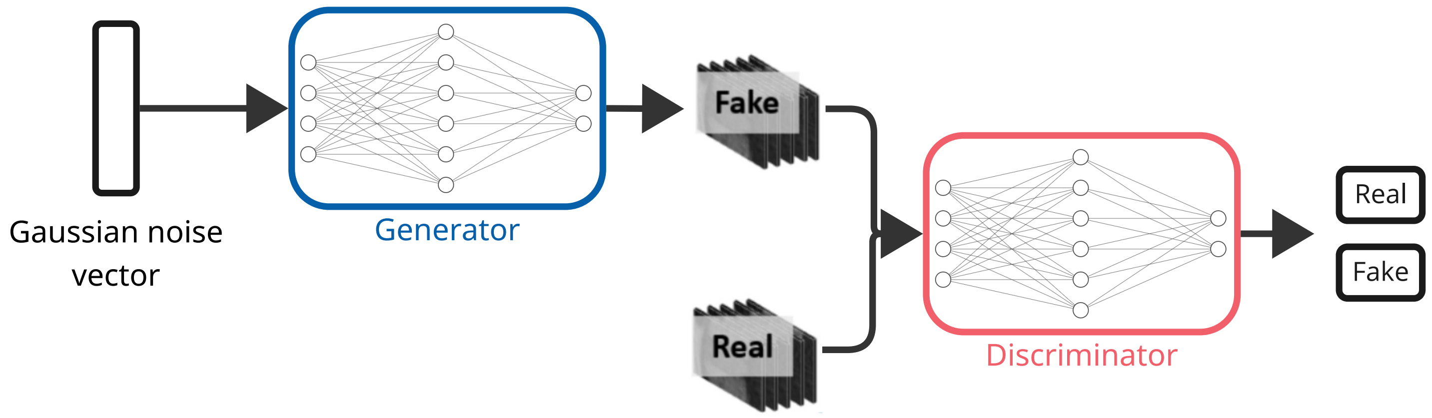
Backup: Optimizing a GAN

Discriminator loss

$$E_x[log(D(x))] + E_z[log(1 - D(G(z)))]$$

where D(x) is the probability that x (real image) is classified as real. The variable z denotes the Gaussian noise vector.

The Generator is trained to *minimize* this loss.



Backup: Auxiliary Classifier GAN

Idea: Introduce conditioning on class variable