MOBILE HEALTH AND GENERATIVE AI

MOBILE HEALTH COURSE

TONG XIA

TX229@CAM.AC.UK

UNIVERSITY OF CAMBRIDGE

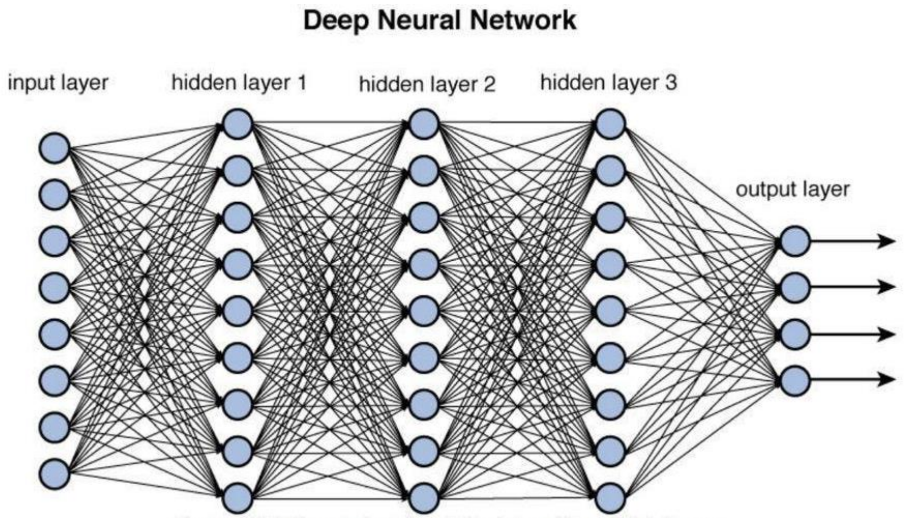


Figure 12.2 Deep network architecture with multiple layers.

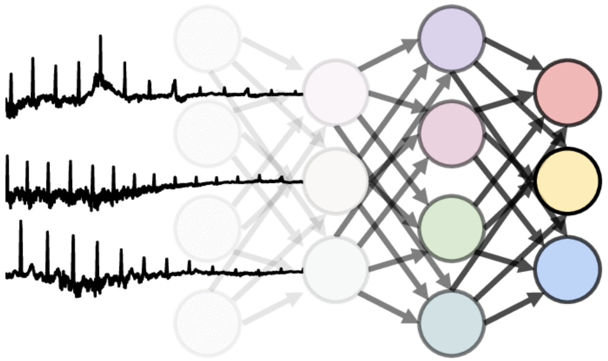
The widespread of mobile and wearable devices

The power of machine learning and deep learning models

Automated health monitoring and diagnostics

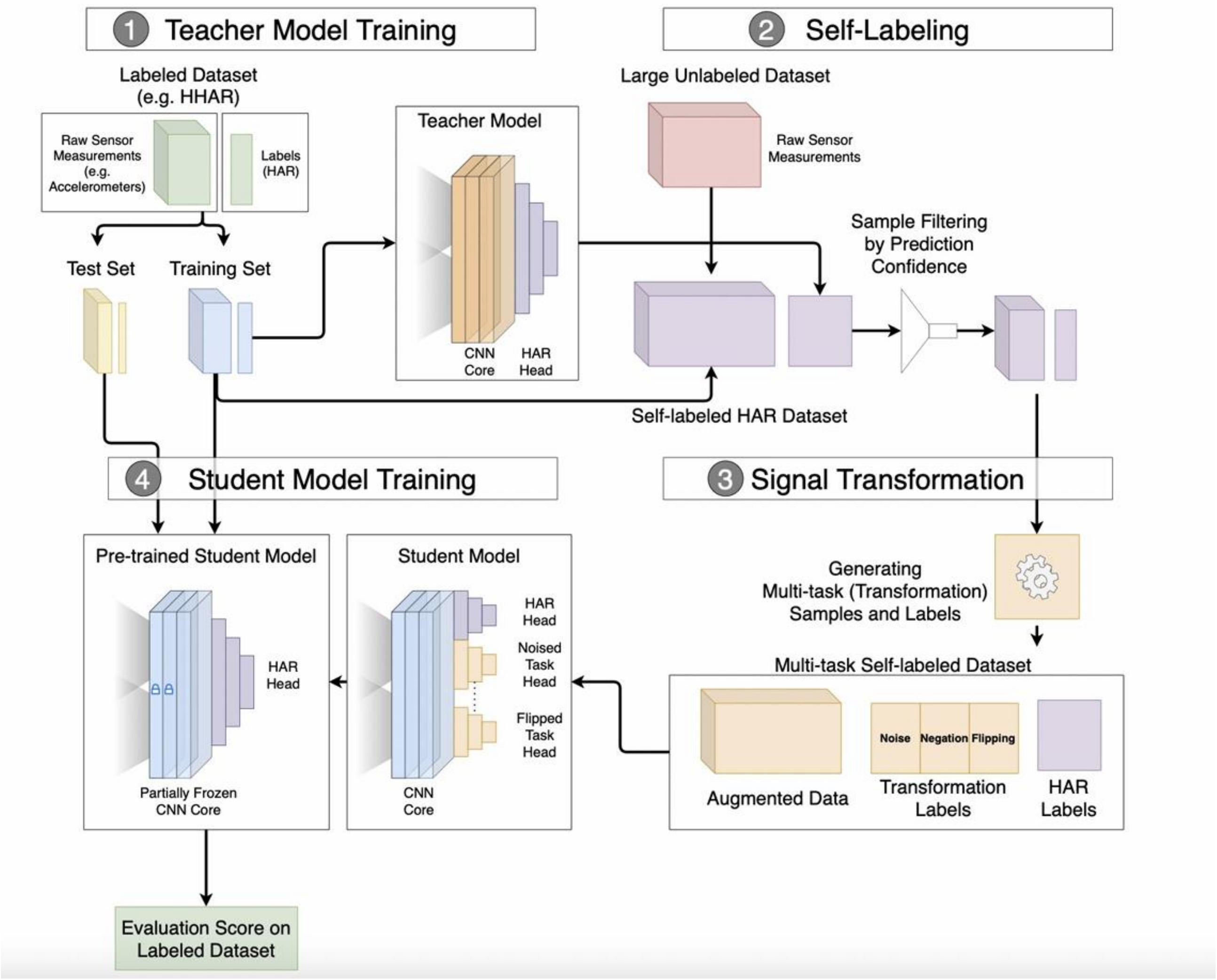
Image from: https://medium.com/@manasim.letsnurture/rise-of-wearables-and-future-of-wearable-technology-1a4e38a2fbb6,
https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964

CHALLENGES



DL models are data hungry

- **Transfer learning**
  - Reduce the need of training data
- **Semi-supervised and self-supervised learning**
  - Reduce the need of annotation



GENERATIVE AI

Generative artificial intelligence (generative AI, GenAI or GAI) is artificial intelligence capable of generating text, images or other data using generative models. Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.

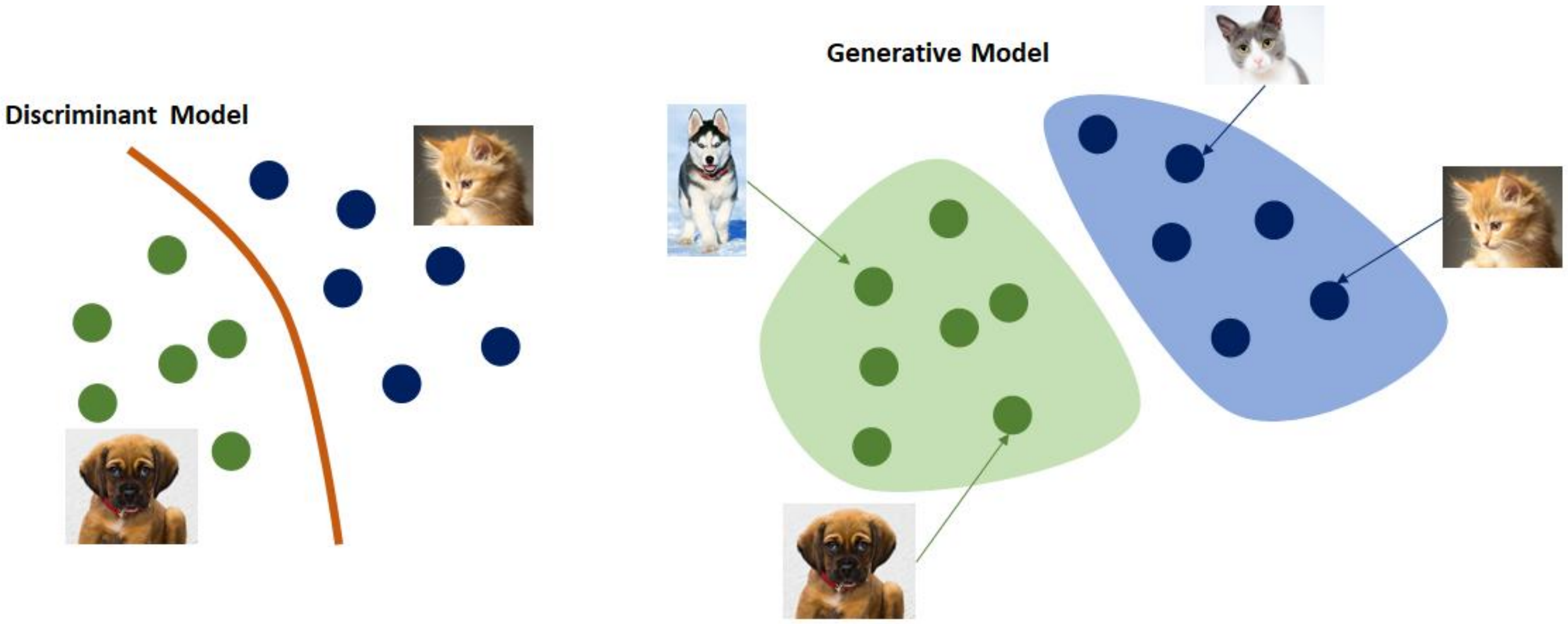


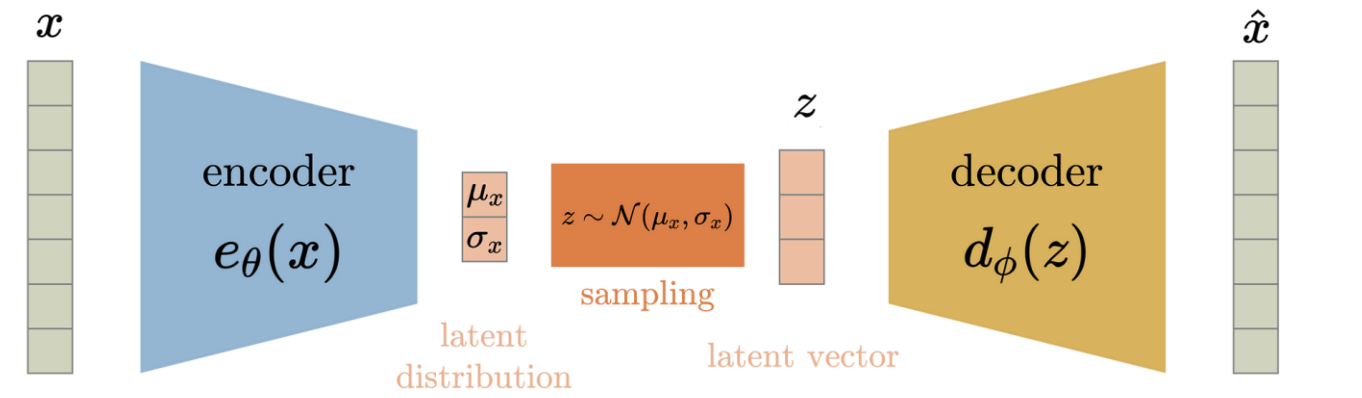
Image from: https://medium.com/@jordi299/about-generative-and-discriminative-models-d8958b67ad32

- **Data generation model**
  - VAE, GAN, Diffusion
  - Examples

- Transformer based generative model
  - Framework
  - Examples

- Foundation model for bio-signals
  - Examples

Generative model - VAE



$$\text{reconstruction loss} = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

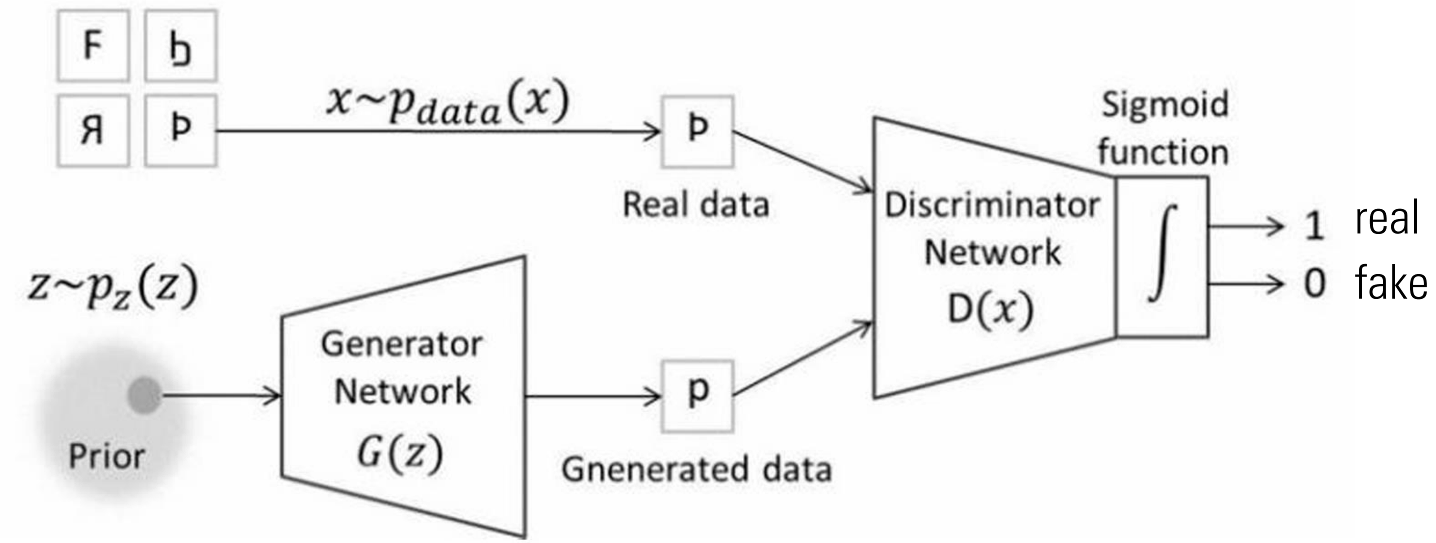
$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\textit{similarity loss} = \textit{KL Divergence} = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$\textit{loss} = \textit{reconstruction loss} + \textit{similarity loss}$$

Pinheiro Cinelli, Lucas; et al. (2021). "Variational Autoencoder". Variational Methods for Machine Learning with Applications to Deep Networks

Generative model - GAN



Train iteratively:

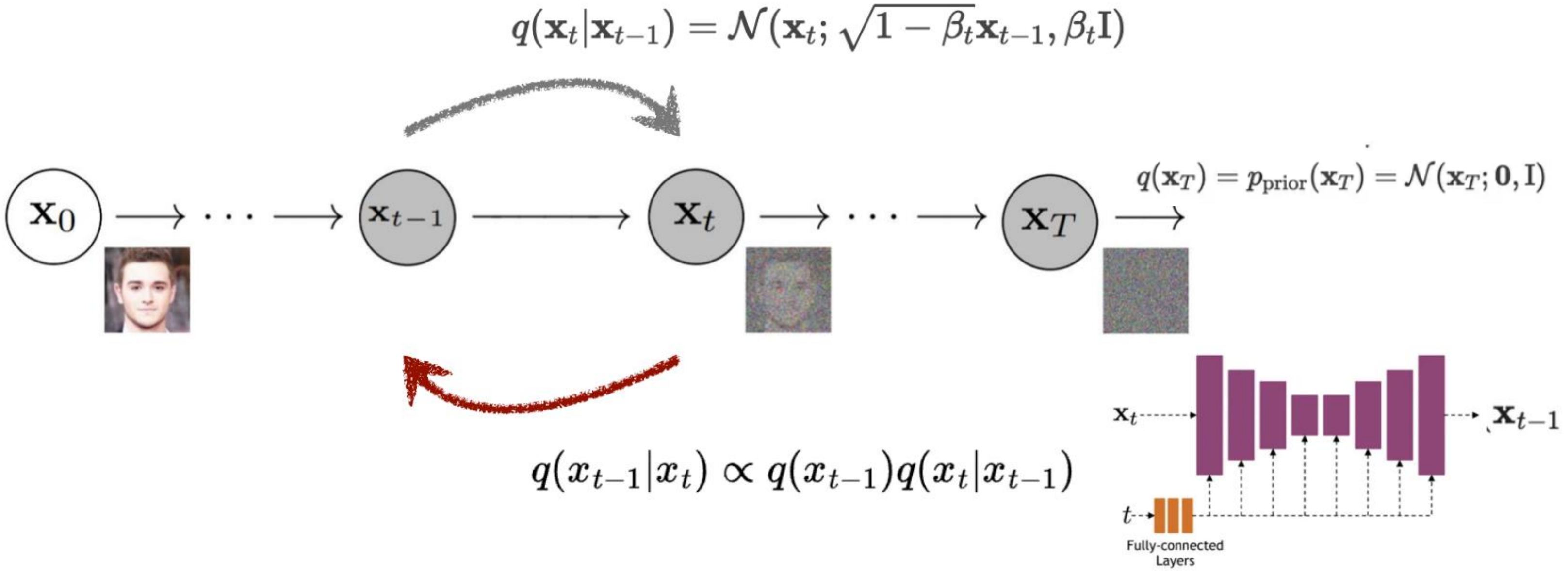
- Step 1: Freeze D(x) to update G(z)
- Step 2: Freeze G(x) to update D(x)

Generation:

- Step 1: Sample z
- Step 2: Use G(z) to generate p

Xu, Dongdong, et al. "Infrared and visible image fusion with a generative adversarial network and a residual network." Applied Sciences 10.2 (2020): 554.

Generative model – Diffusion Models



- Forward diffusion process: Iteratively inject given noise to the data
- Reverse diffusion process: Intractable but can be approximated by a UNet

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851.

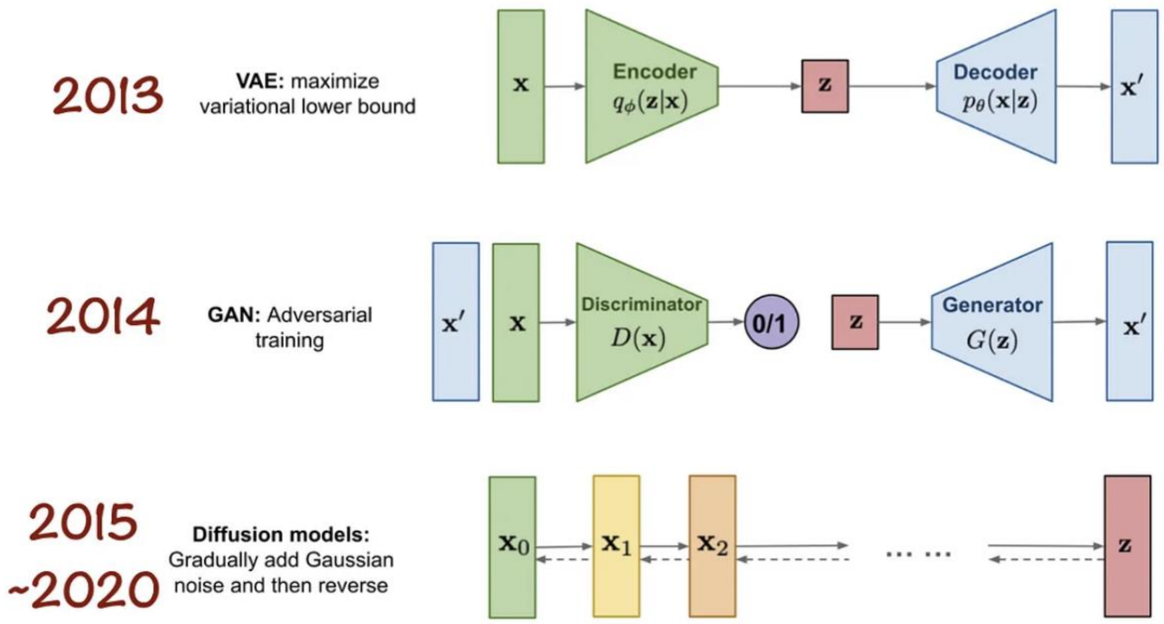
Generative model – Diffusion Models

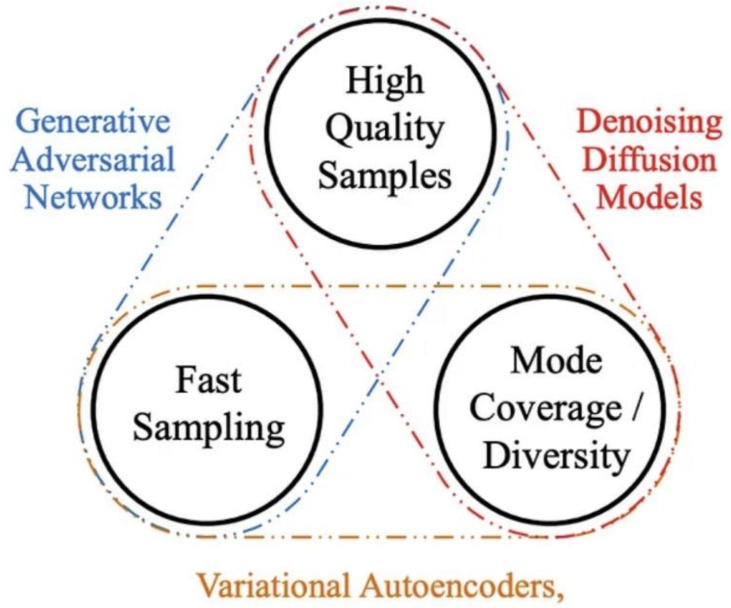
Scaling transformers for video generation

Sora is a diffusion model[21,22,23,24,25]; given input noisy patches (and conditioning information like text prompts), it's trained to predict the original "clean" patches. Importantly, Sora is a diffusion *transformer*.[26] Transformers have demonstrated remarkable scaling properties across a variety of domains, including language modeling,[13,14] computer vision,[15,16,17,18] and image generation.[27,28,29]

Sora (openai.com)

A comparison





- Data quantity augmentation: enabling more data samples for downstream tasks
- Data quality enhancement:
  - Removing noise/artefects
  - Imputing the missenses in the data
  - Privacy-preserving data sharing

Recommend reading: Cao, Hanqun, et al. "A survey on generative diffusion models." *IEEE Transactions on Knowledge and Data Engineering* (2024).

Image from: https://medium.com/@marija.jegorova/a-quick-dive-into-diffusion-models-part-1-fundamentals-of-the-ddpms-4de1304d7b6f

Example 1 : Diffusion model-based EEG generation

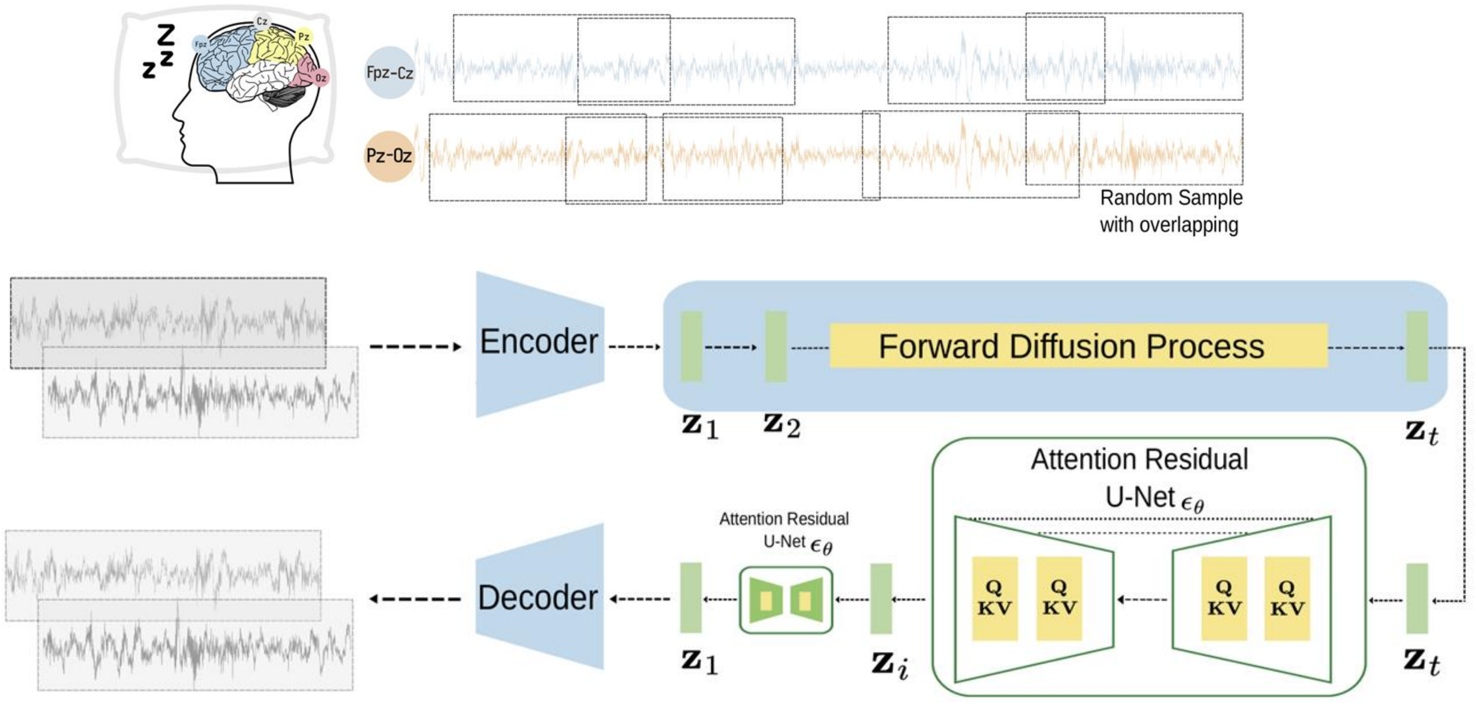


Table 1: Quantitative evaluation

Dataset	FID ↓		
	LDM	LDM_{spec}	Real
Sleep EDFx	11.933	**0.308**	0.015
SHHS_h	0.936	**0.168**	0.086

FID: Fréchet Inception Distance

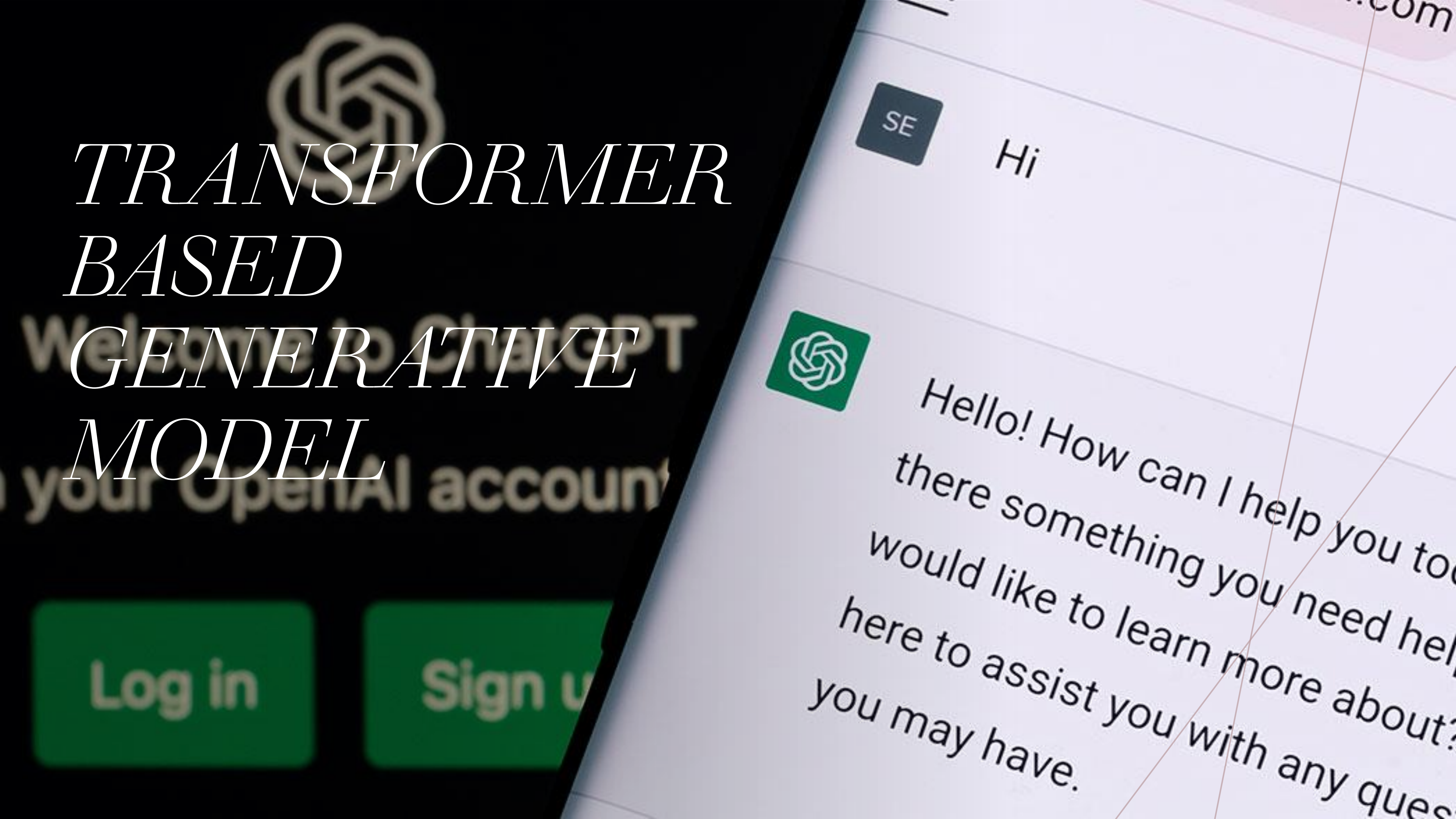
$$\min \sum \boxed{\ell_{\text{recons}}(\mathbf{x}, \widehat{\mathbf{x}})} + \ell_{\text{adv}}(\mathbf{x}, \widehat{\mathbf{x}}) + \ell_{\text{kl}}(\mathbf{z}_\mu, \mathbf{z}_\sigma) + \boxed{\ell_{\text{spec}}(\mathbf{x}_i, \widehat{\mathbf{x}})},$$

Temporal dynamics reconstruction

Spectral feature similarity

Aristimunha, Bruno, et al. "Synthetic Sleep EEG Signal Generation using Latent Diffusion Models." NeurIPS 2023 Workshop DGM4H. 2023.

TRANSFORMER BASED GENERATIVE MODEL



Transformer

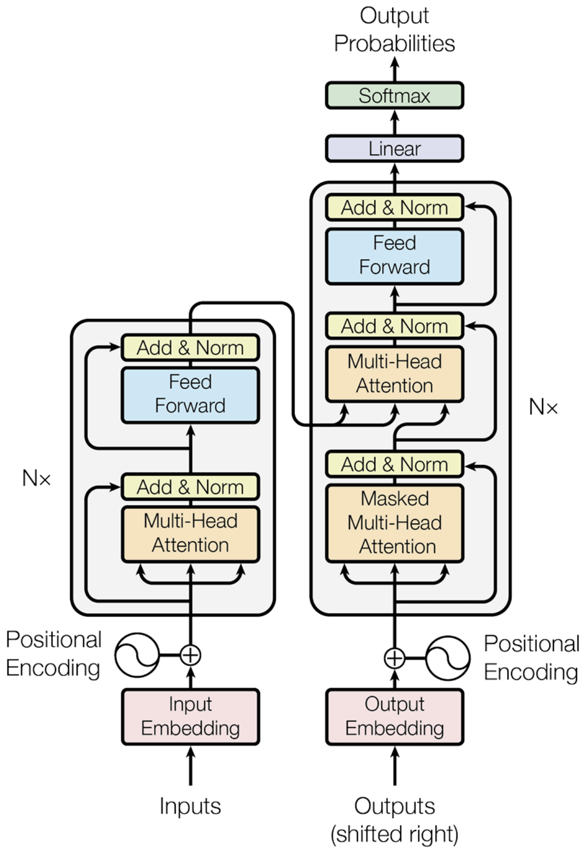
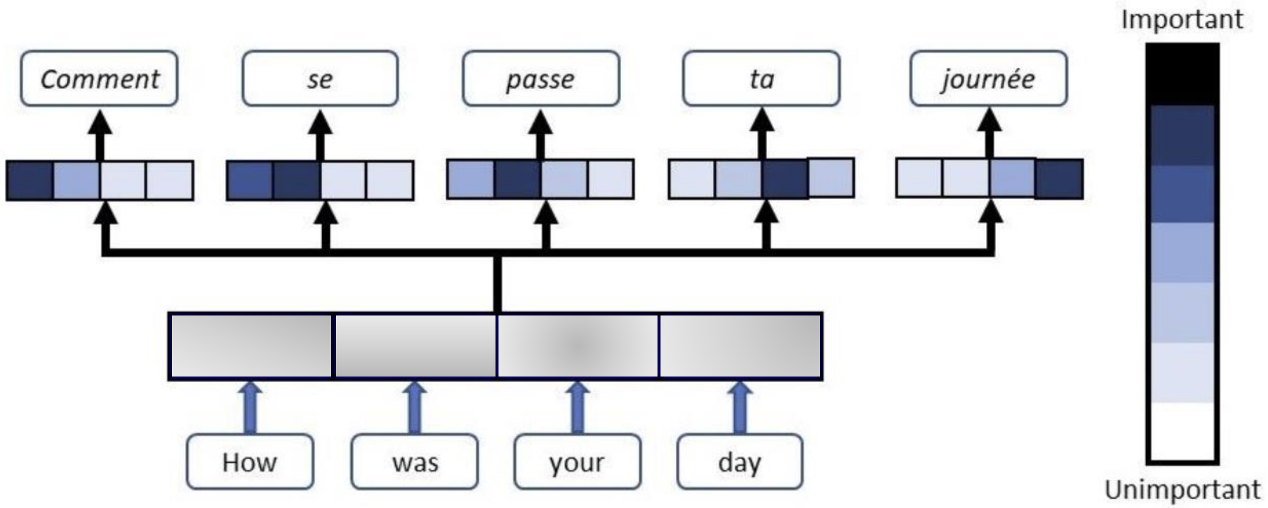


Figure 1: The Transformer - model architecture.

What is Attention?



Weights are assigned to input words at each step of the translation

Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

Generative pre-trained Transformer (GPT)

Stage I: Unsupervised pre-training

Large-scale unlabelled data

Given an unsupervised corpus of tokens $\mathcal{U} = \{u_1, \ldots, u_n\}$, we use a standard language modeling objective to maximize the following likelihood:

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta) \quad (1)$$

Model size is important!

Stage II: Supervised training

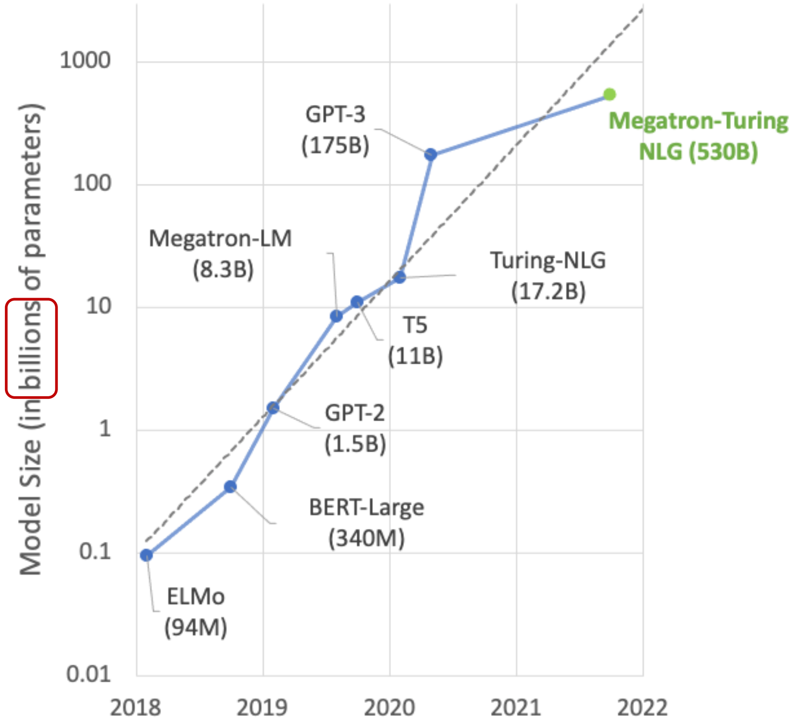
Small-scale labelled data

We assume a labeled dataset $\mathcal{C}$, where each instance consists of a sequence of input tokens, $x^1, \ldots, x^m$, along with a label y. The inputs are passed through our pre-trained model to obtain the final transformer block's activation h_l^m, which is then fed into an added linear output layer with parameters W_y to predict y:

$$L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y | x^1, \ldots, x^m). \quad (4)$$

Radford, Alec, et al. "Improving language understanding by generative pre-training." (2018).

Emergent abilities of large language models (LLMs)



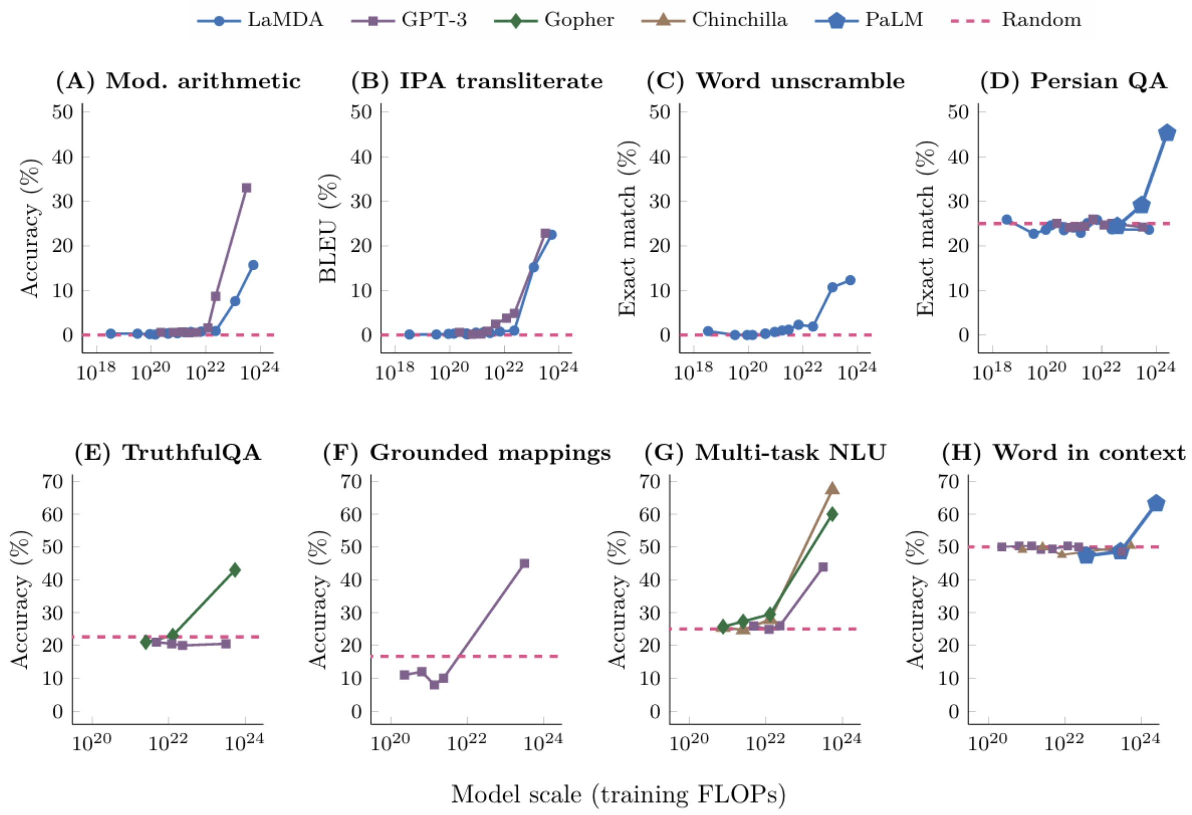


Figure 2: Eight examples of emergence in the few-shot prompting setting. Each point is a separate model.

Wei, Jason, et al. "Emergent abilities of large language models." arXiv preprint arXiv:2206.07682 (2022).

Fine-tuning LLMs

A key to success of ChatGPT

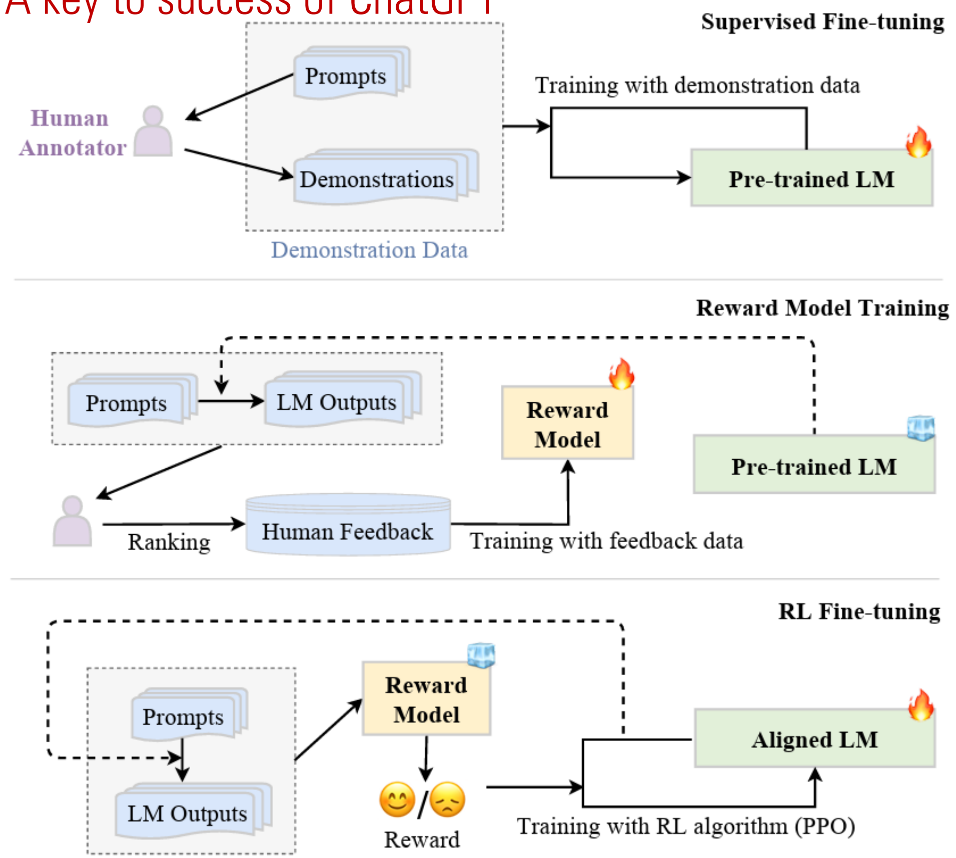


Fig. 12: The workflow of the RLHF algorithm.

MHA: multi-head attention
FFN: feed-forward network

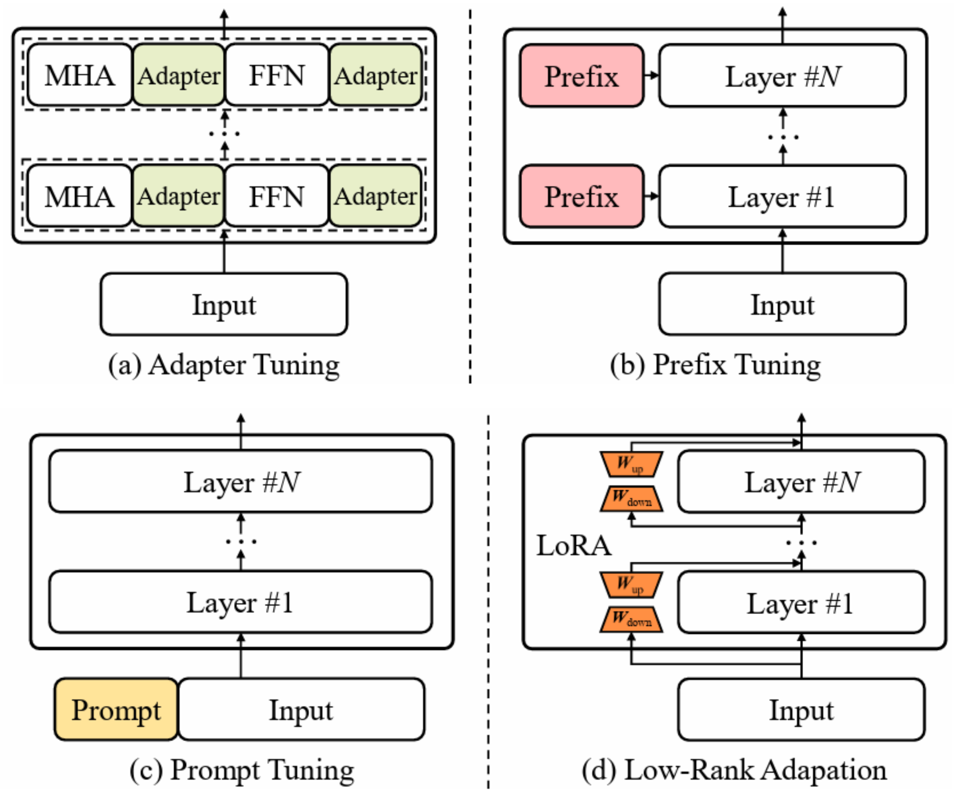
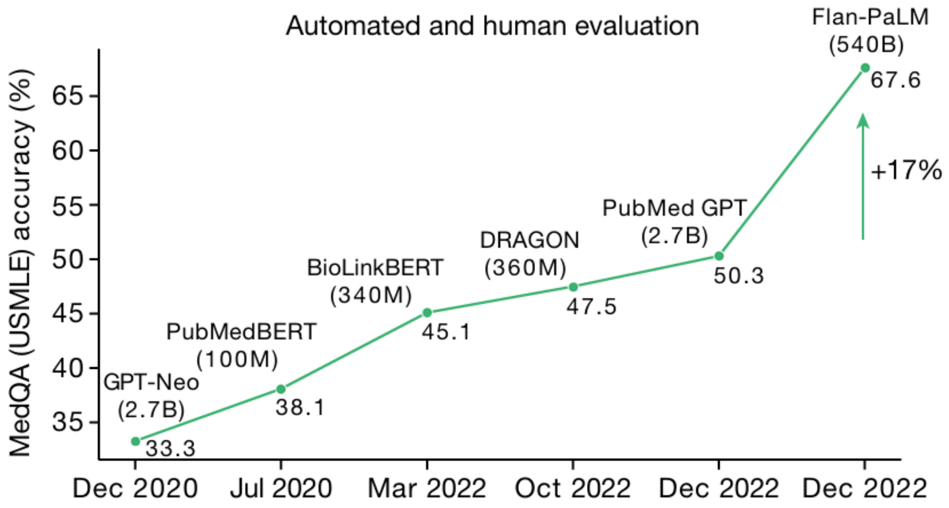


Fig. 13: An illustration of four different parameter-efficient fine-tuning methods.

Fine-tuning the entire model is not practical for most applications

Zhao, Wayne Xin, et al. "A survey of large language models." arXiv preprint arXiv:2303.18223 (2023).

Example 2: Medical large language models



Q: How long does it take for newborn jaundice to go away?

Med-PaLM: Newborn jaundice is when a newborn baby's skin and eyes look yellow. It is caused by a buildup of a substance called bilirubin in the baby's blood. Bilirubin is a byproduct of the normal process of red blood cells breaking down. Newborn jaundice is common and typically harmless, but it can be a sign of a more serious condition. The jaundice typically goes away on its own within a few weeks. In some cases, it may last longer or require treatment. The time it takes for the jaundice to go away can vary depending on the cause and the severity of the jaundice. If the jaundice is severe or lasts longer than a few weeks, the doctor may recommend testing or treatment to determine the cause and prevent complications.

Med-PaLM performs encouragingly
on consumer medical question answering

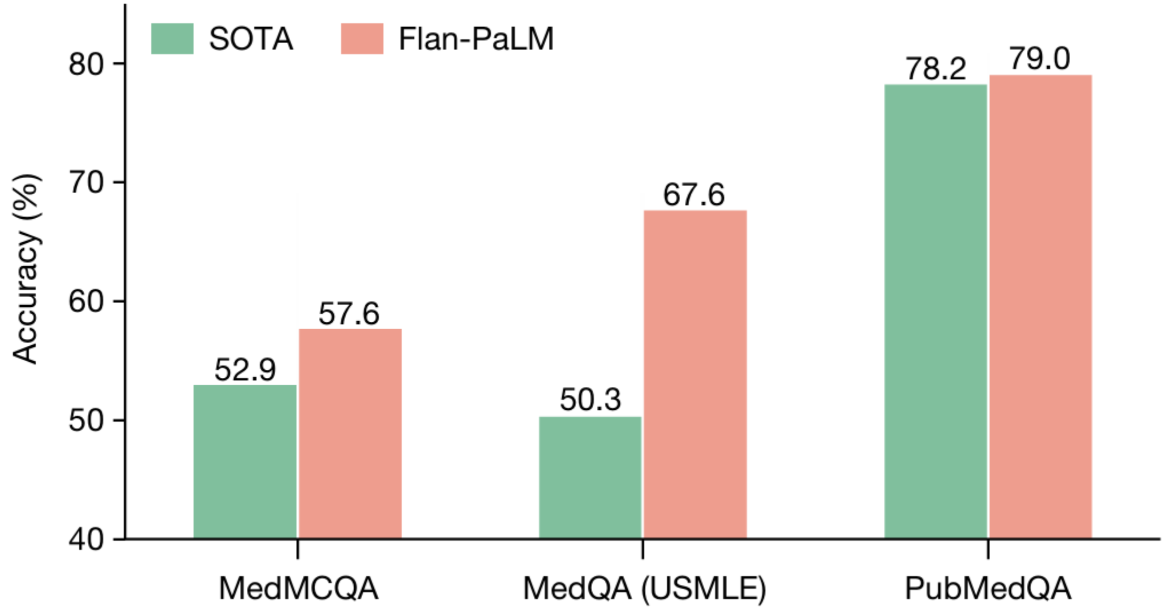
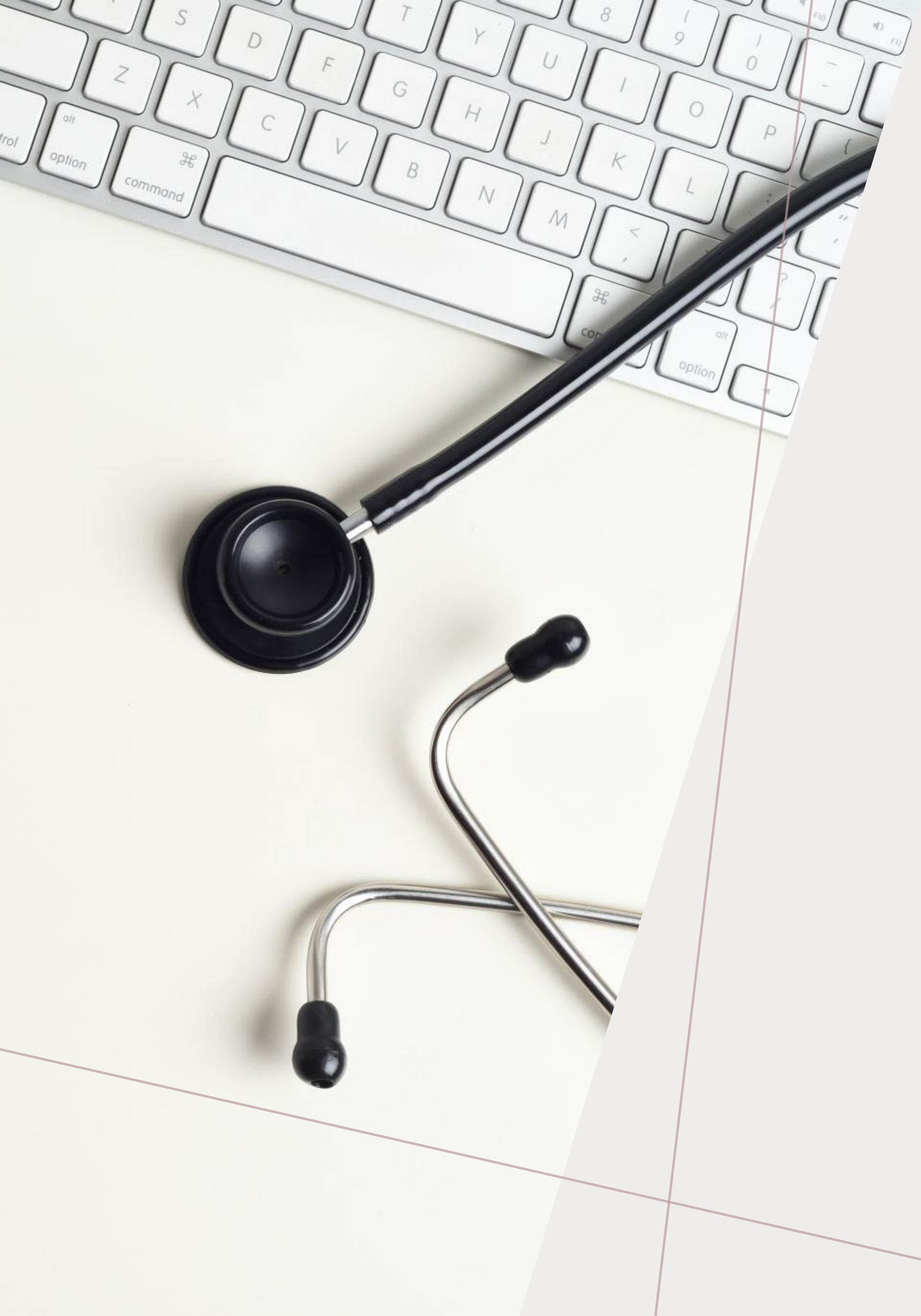


Fig. 2 | Comparison of our method and prior state of the art. Our Flan-PaLM 540B model exceeds the previous state-of-the-art performance (SOTA) on MedQA (four options), MedMCQA and PubMedQA datasets. The previous state-of-the-art results are from Galactica[20] (MedMCQA), PubMedGPT[19] (MedQA) and BioGPT[21] (PubMedQA). The percentage accuracy is shown above each column.

Singhal, Karan, et al. "Large language models encode clinical knowledge." Nature 620.7972 (2023): 172-180.



HOW DO LLMS PERFORM ON MOBILE HEALTH TASKS?

Example 3: Large language models are few-shot learners

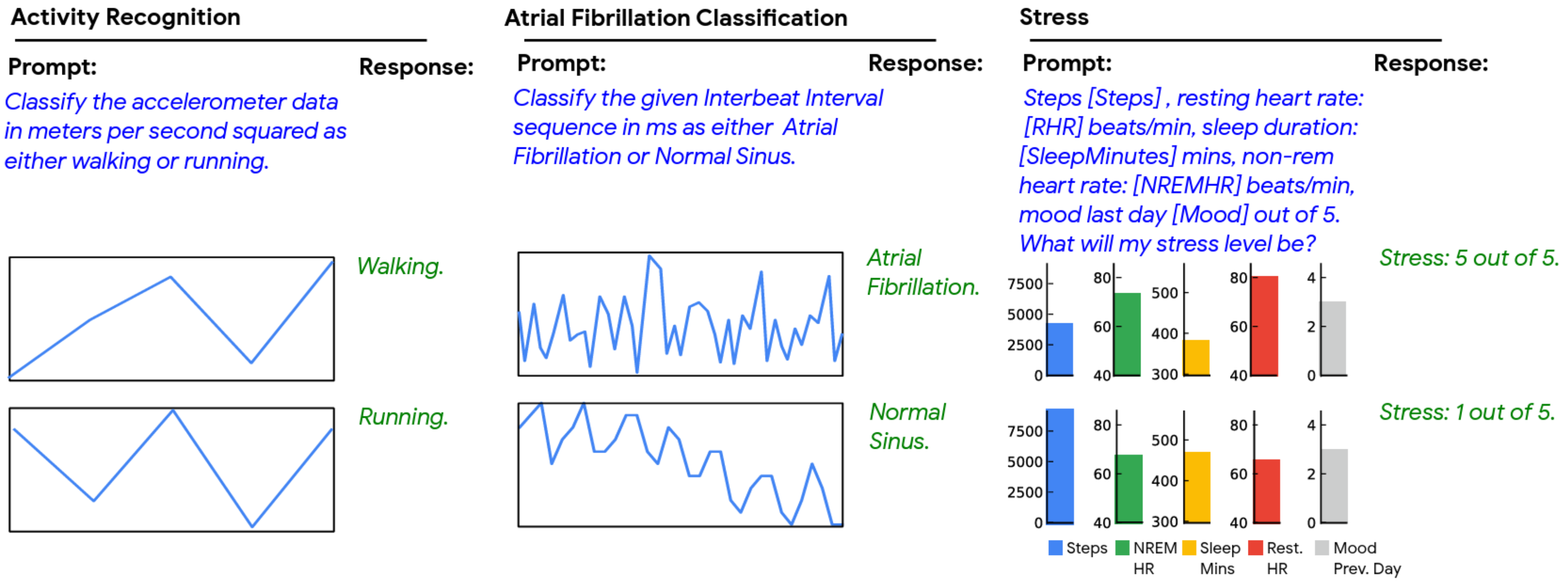


Figure 1: **Examples of question-answer pairs for our health tasks.** In the prompts, data were represented numerically rather than graphically.

Liu, Xin, et al. "Large Language Models are Few-Shot Health Learners." arXiv preprint arXiv:2305.15525 (2023).

Example 4: Large language models are few-shot learners

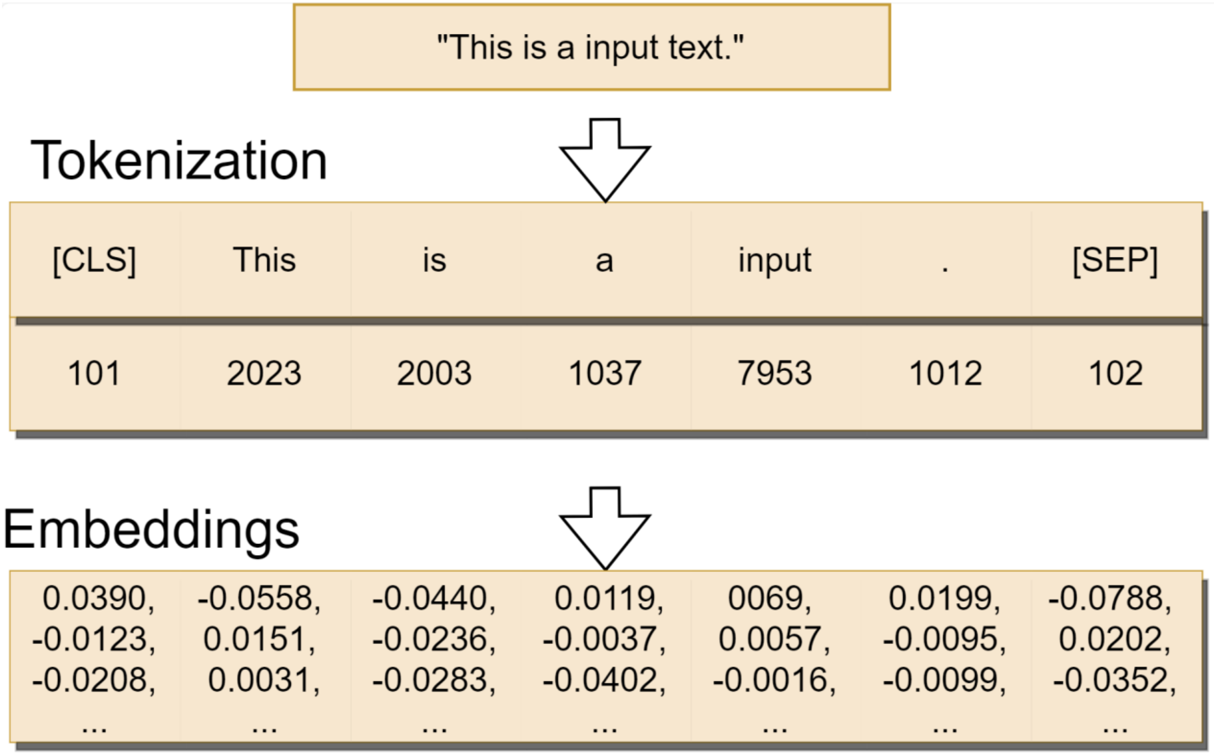
Input: *"Classify the following accelerometer data in meters per second squared as either walking or running: 0.052,0.052,0.052,0.051,0.052,0.055,0.051,0.056,0.06,0.064"*
Label: "Running"

Table 2: **Results.** Comparison of performance between prompt-tuned LLMs (w/ Context-Inclusive Prompts) and supervised neural network training across all consumer health tasks.

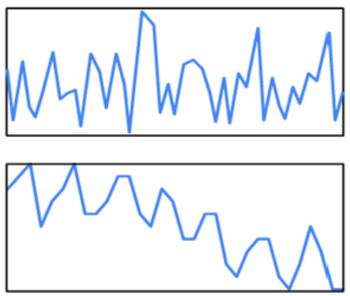
			Supervised Baseline			**LLM with Context**			
Topic	**Task**	**Metric**	3-Shot	10-Shot	25-Shot	3-Shot	10-Shot	25-Shot	% Improvement
Cardio	HRs to Average HR	MAE ↓ (beats/min)	3.41	1.37	1.08	6.00	2.49	**1.06**	**+1.90%**
	IBIs to HR	MAE ↓ (beats/min)	34.0	20.0	19.8	12.3	5.87	**5.01**	**+74.7%**
	IBIs to A.Fib.	Accuracy ↑ (%)	52.5	72.5	75.0	85.0	75.0	**89.0**	**+19.7%**
	IBIs to Sinus B.	Accuracy ↑ (%)	88.0	86.0	86.0	81.0	79.0	**92.0**	**+7.00%**
	IBIs to Sinus T.	Accuracy ↑ (%)	56.0	53.0	61.0	65.0	82.0	**88.0**	**+44.3%**
Activity	IMU Activity	Accuracy ↑ (%)	56.0	60.0	64.0	62.0	80.0	**85.0**	**+32.8%**
Metabolic	Calories	MAE ↓ (calories)	185	97	89	106	77	**48**	**+46.1%**
MHealth	Fitbit to Stress	Accuracy ↑ (%)	37.5	70.5	80.0	72.5	71.5	**82.5**	**+3.10%**
	Fitbit to PHQ	Accuracy ↑ (%)	51.0	52.0	53.0	49.0	59.0	**69.0**	**+30.2%**

Liu, Xin, et al. "Large Language Models are Few-Shot Health Learners." arXiv preprint arXiv:2305.15525 (2023).

❑Tokenizer for natural language

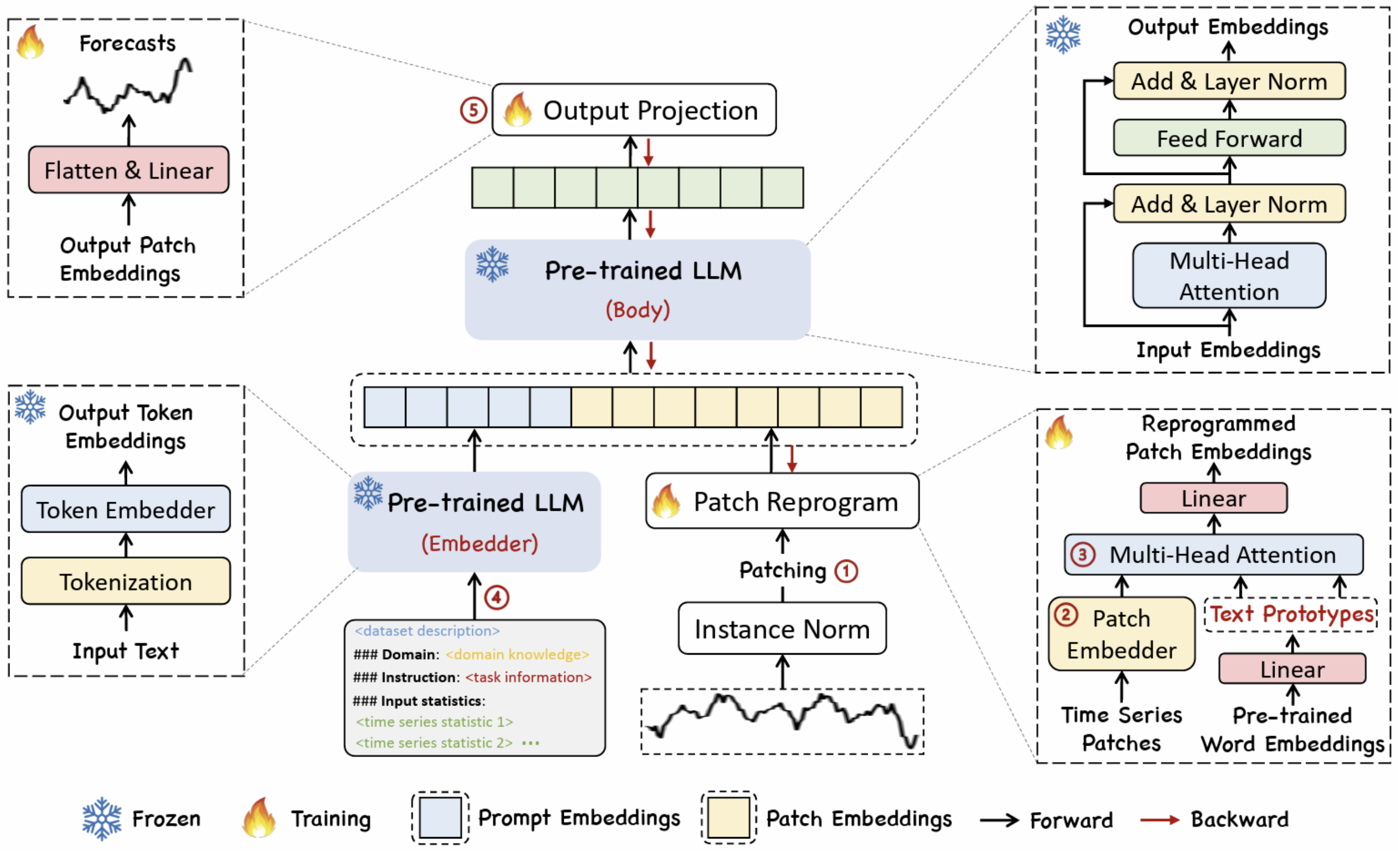


Q: HOW TO REPRESENT TEMPORAL DATA?



Spathis, Dimitris, and Fahim Kawsar. "The first step is the hardest: Pitfalls of representing and tokenizing temporal data for large language models." arXiv preprint arXiv:2309.06236 (2023).

Example 4: Time-LLM: Time series forecasting by reprogramming large language models



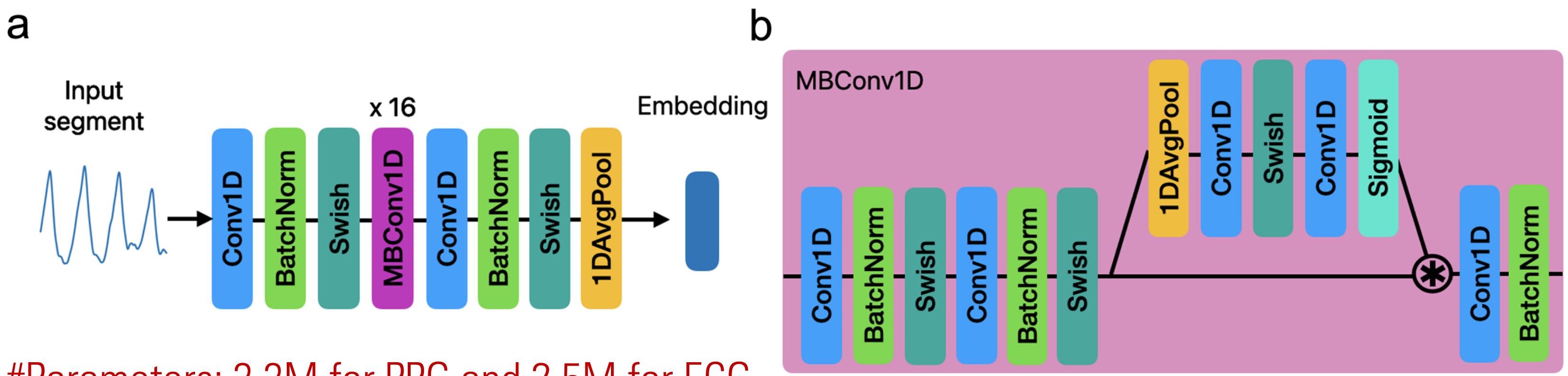
Jin, Ming, et al."Time-llm: Time series forecasting by reprogramming large language models. ICLR 2024

IS THERE ANY PHYSIOLOGICAL DATA SPECIFIC FOUNDATION MODEL?

Example 5: Large-scale training of foundation models for wearable bio-signals

Data used to develop this foundation model

	PPG	ECG
Number of participants	141,207	106,643
Number of segments	19,854,101	3,743,679
Average number of calendar days per participant	92.54	23.27
Total dataset time span (days)	890	1,240



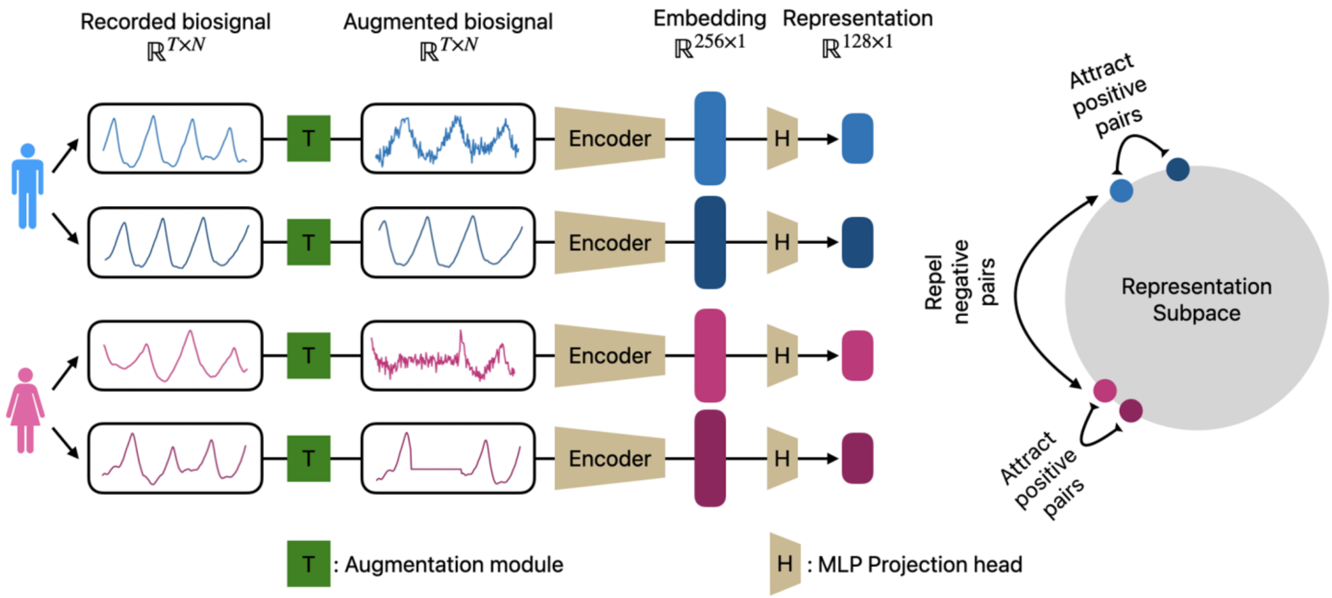
#Parameters: 3.3M for PPG and 2.5M for ECG

Figure 4: Our EfficientNet-style encoder architecture, adapted from (Tan & Le, 2020) for time-series

Salar Abbaspourazad, Oussama Elachqar, Andrew Miller, Saba Emrani, Udhyakumar Nallasamy, Ian Shapiro. "Large-scale training of foundation models for wearable biosignals."ICLR 2024

Example 5: Large-scale training of foundation models for wearable bio-signals

❑ SSL training:



❑ Results on downstream tasks:

Prediction task	PPG	
	AUC (pAUC) ↑	MAE ↓
Age classification	**0.976** (**0.907**)	-
Age regression	-	**3.19**
BMI classification	**0.918** (**0.750**)	-
BMI regression	-	**2.54**
Sex classification	**0.993** (**0.967**)	-

Prediction task	ECG	
	AUC (pAUC) ↑	MAE ↓
Age classification	**0.916** (**0.763**)	-
Age regression	-	**6.33**
BMI classification	**0.797** (**0.612**)	-
BMI regression	-	**3.72**
Sex classification	**0.951** (**0.841**)	-

Salar Abbaspourazad, Oussama Elachqar, Andrew Miller, Saba Emrani, Udhyakumar Nallasamy, Ian Shapiro. "Large-scale training of foundation models for wearable biosignals."ICLR 2024

SUMMARY

Limited labelled data is an obstacle for high-performing DL

- Now we have:
  - Data generation models for data augmentation
  - Pre-trained large (language) models for downstream tasks
  - SSL-empowered foundation models for bio-signals
- Open questions:
  - Evaluation of fine-tuning methods and the foundation models and on mobile health applications
  - Multi-modality foundation models…

FUTURE

Digital health twin

LLMs for health reasoning

THANK YOU!

Tong Xia
tx229@cam.ac.uk