

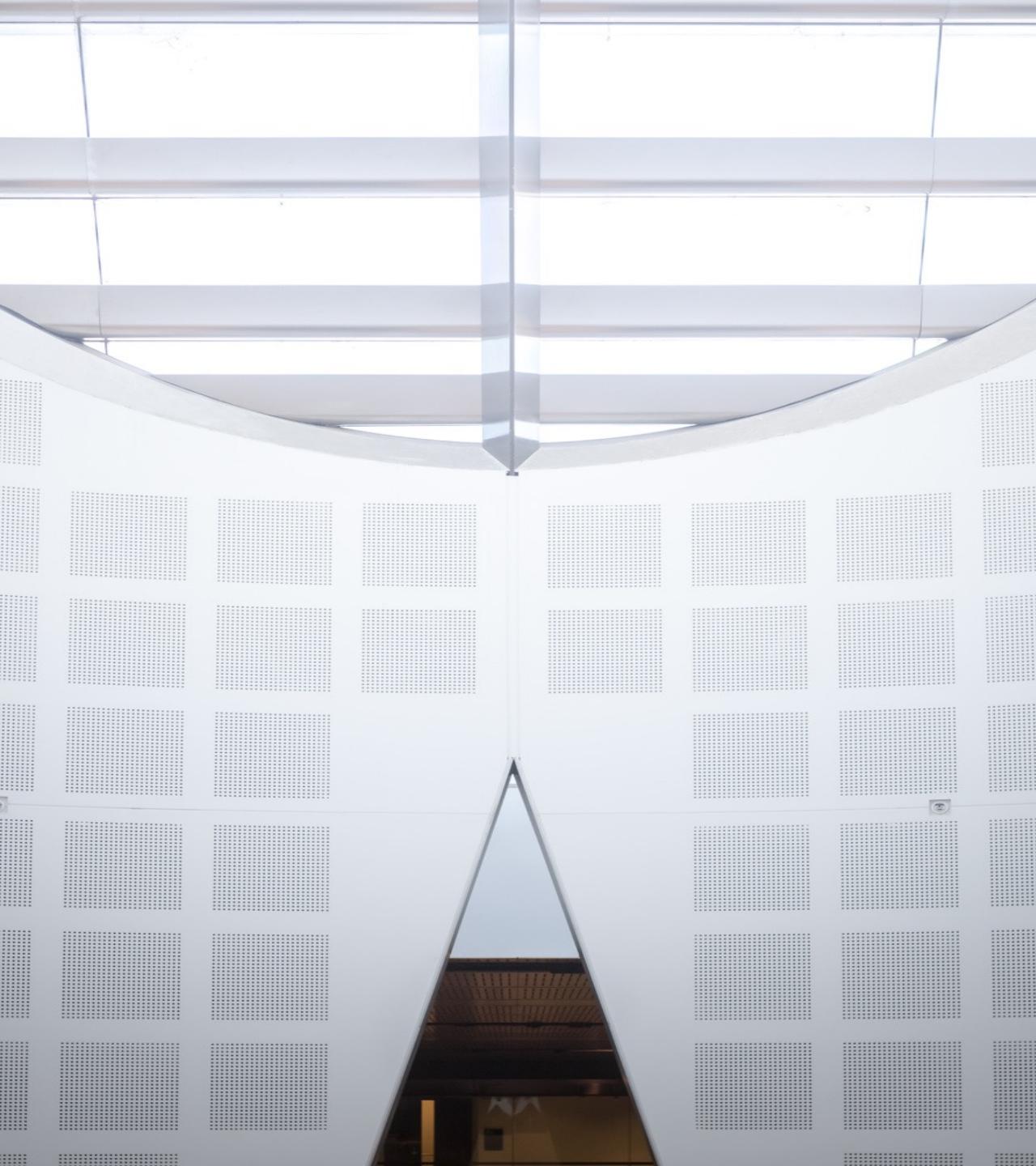
Seminar 1: Overview of GenAI

Yunke Wang

yunke.wang@sydney.edu.au
School of Computer Science



THE UNIVERSITY OF
SYDNEY



GenAI Applications: ChatGPT by OpenAI

ChatGPT

Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →

Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests

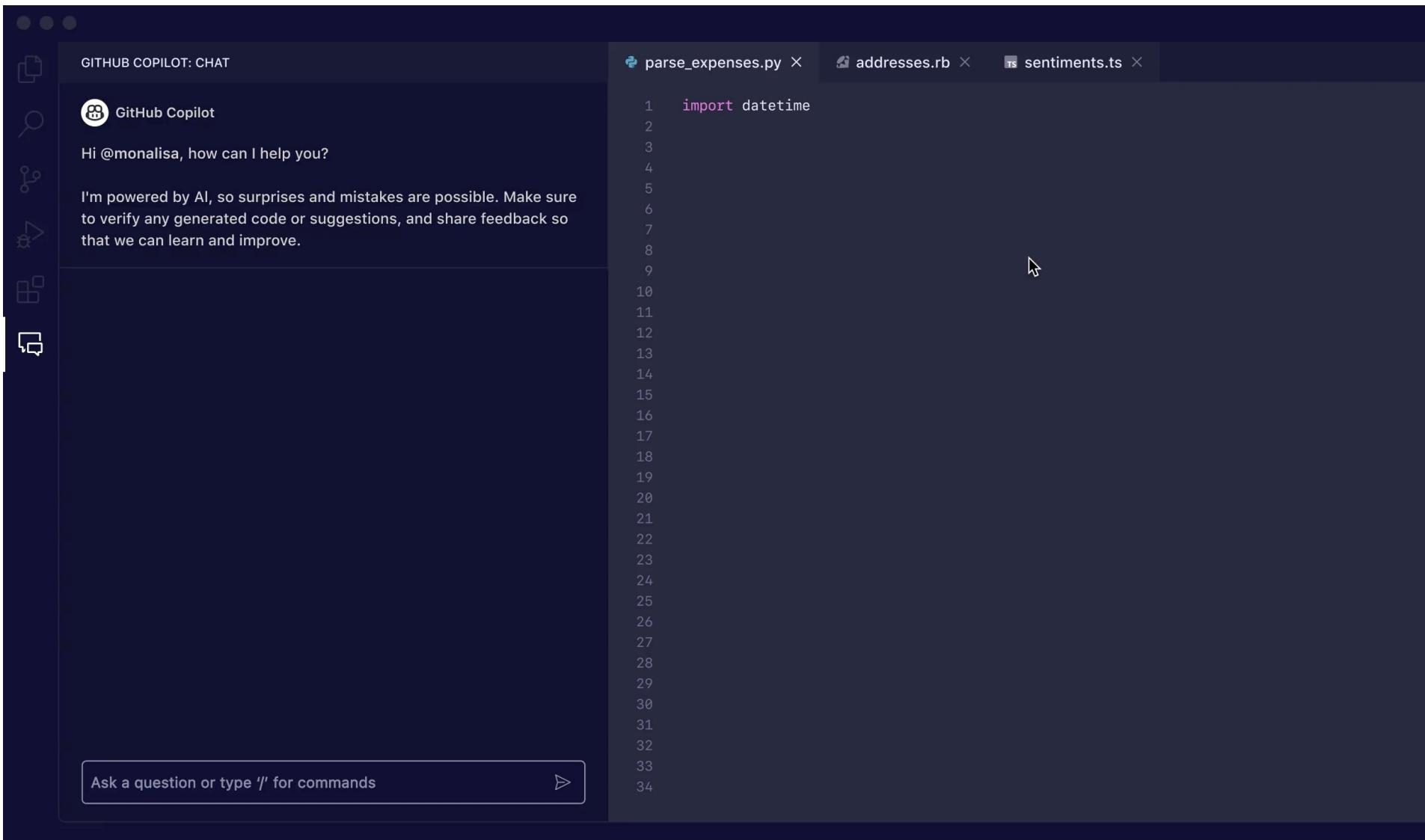
Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

GenAI Applications: Github Copilot

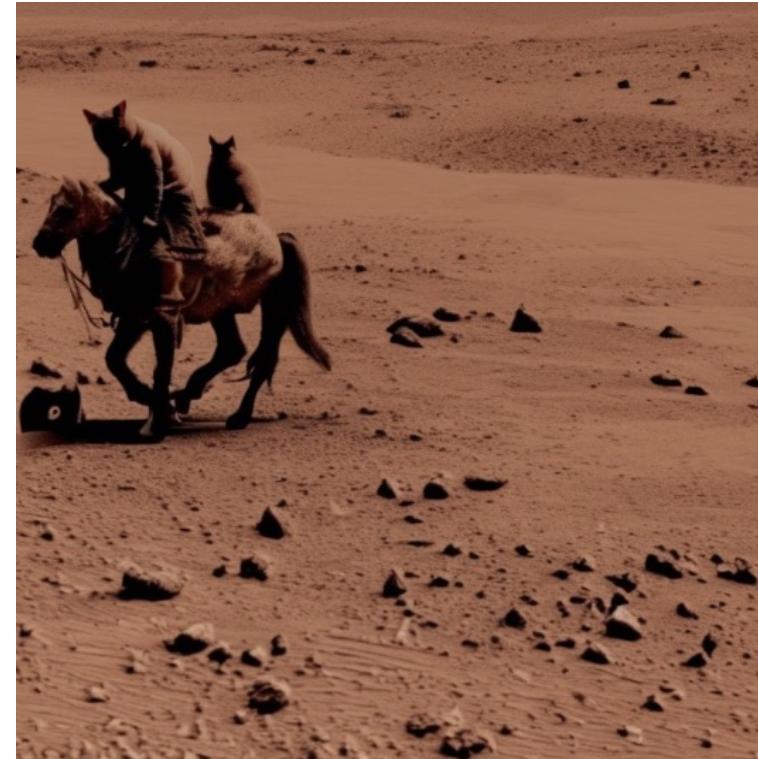


GenAI Applications: Stable Diffusion

Stable Diffusion is a latent **text-to-image** diffusion model capable of generating photo-realistic images given any text input, cultivates autonomous freedom to produce incredible imagery, empowers billions of people to create stunning art within seconds.



a photo of an astronaut riding a horse on mars



a photo of a cat riding a horse on mars

GenAI Applications: Sora by OpenAI

Sora is a latent **text-to-video** diffusion model. Sora can generate videos up to **one-minute** long while maintaining visual quality and adherence to the user's prompt.

Prompt:

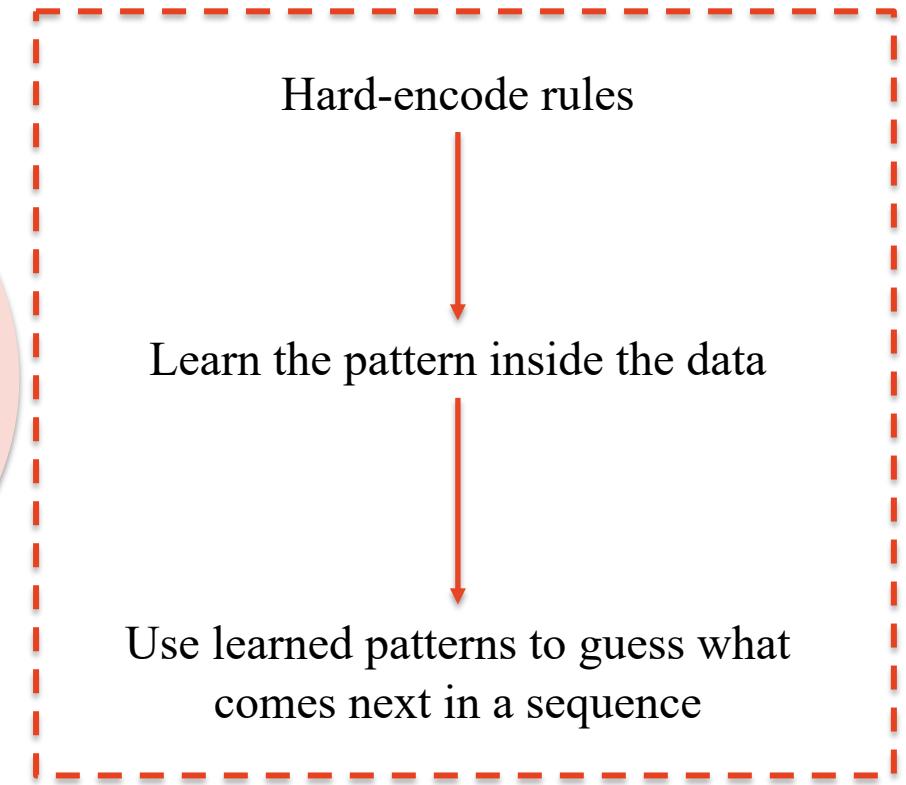
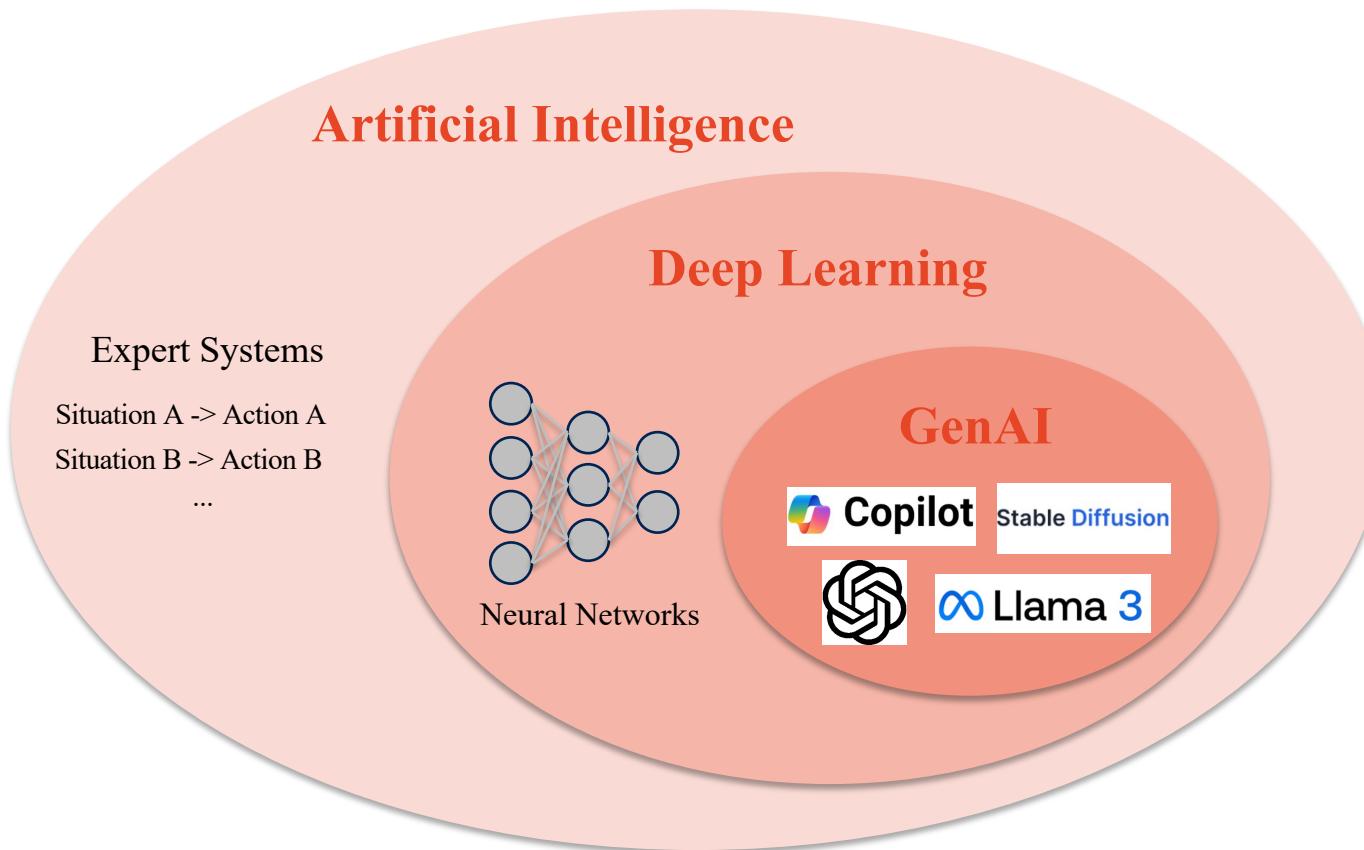
A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.



What is Generative AI?



Generative AI is a type of artificial intelligence that creates new content—like text, images, music, or code—based on patterns it has learned from existing data. -- *Generative by ChatGPT-4o*



What is Large Language Model (LLM)?

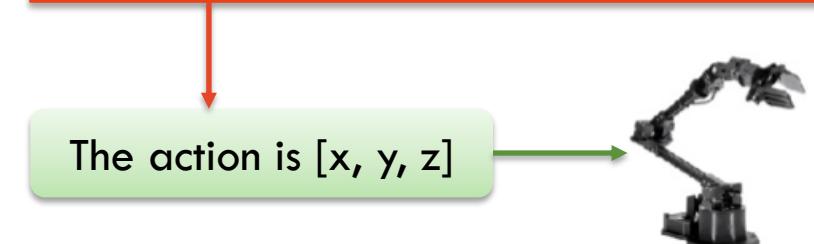


Large Language Model (LLM) is the most representative model of GenAI designed to process and generate human-like text. It uses vast amounts of data and sophisticated algorithms to understand, predict, and create natural language. -- *Generative by ChatGPT-4o*

Text Prompt:
What is Large Language model?



Put Eggplant
in bowl



[1] Kim et al., OpenVLA: An Open-Source Vision-Language-Action Model.

What to Learn in this GenAI Summer School?

Seminar 1: Development of GenAI (Yunke Wang, Jinxu Lin)

- *Part 1:* Overview of GenAI architectures including Transformer, Autoregressive models, and Diffusion Models.
- *Part 2:* A look at various existing types of LLMs, including Multi-Modal LLMs.

Seminar 2: Training and Inference of GenAI (Anh-Dung Dinh, Xiaohuan Pei)

- *Part 1:* Efficient training methods for LLMs, such as LoRA.
- *Part 2:* LLM with Mixture-of-Expert, Retrieval-augmented Language models (RGA).

Seminar 3: Applications of GenAI in Robotics (Zunzhi You, Xiyu Wang)

- *Part 1:* LLM applications in generating robotics data.
- *Part 2:* Use cases of Vision-Language Models (VLMs) in robotics.

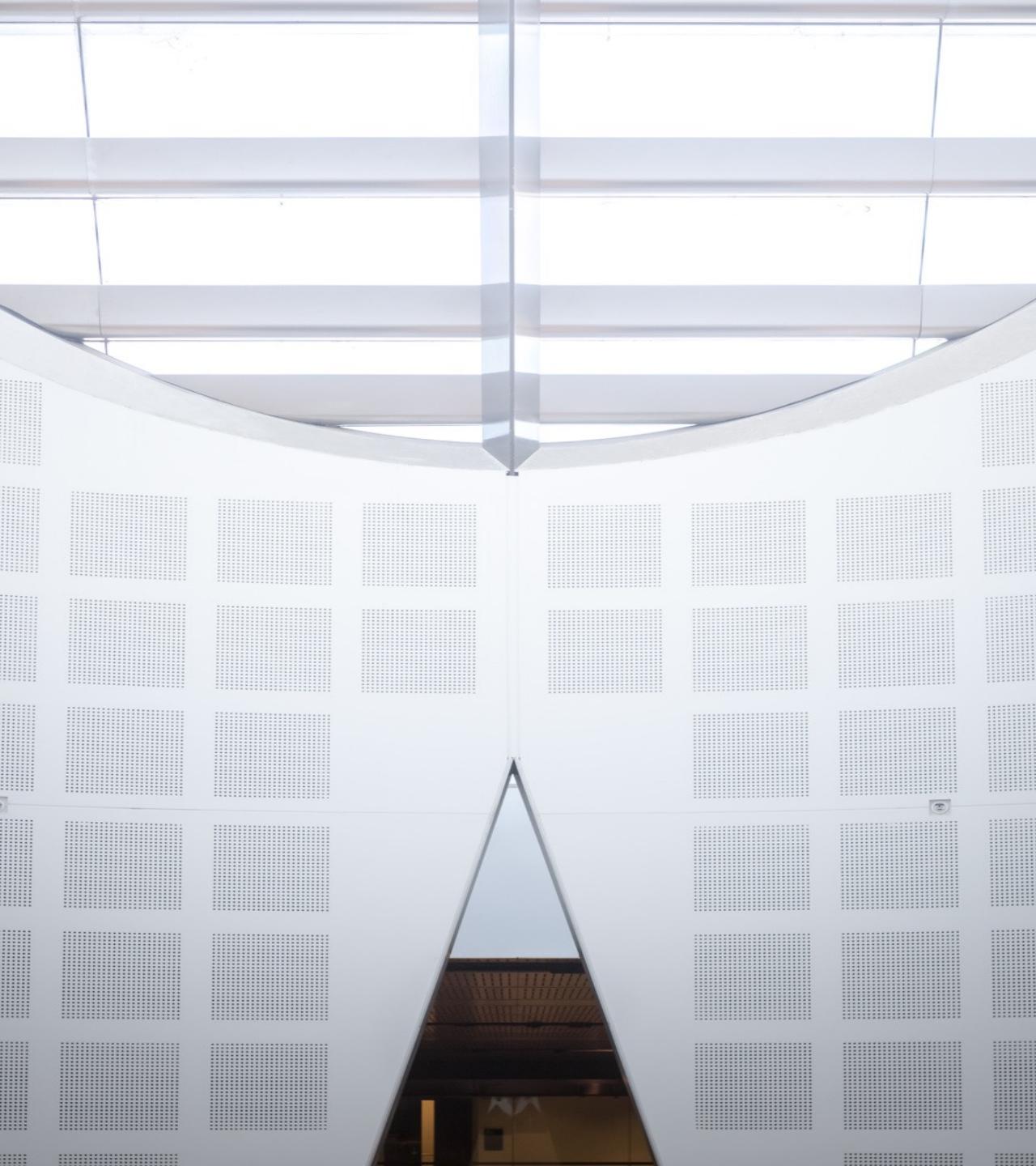
Seminar 4: Robustness of GenAI (Linwei Tao, Younan Zhu)

- Discussions on robustness and phenomena like hallucination and jailbreak in LLMs and GenAI.

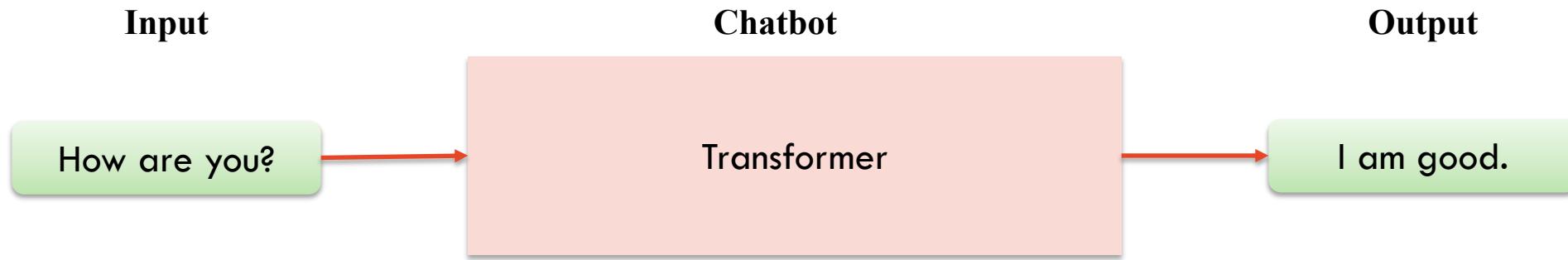
Seminar 1, Part 1.1: Architecture of LLM



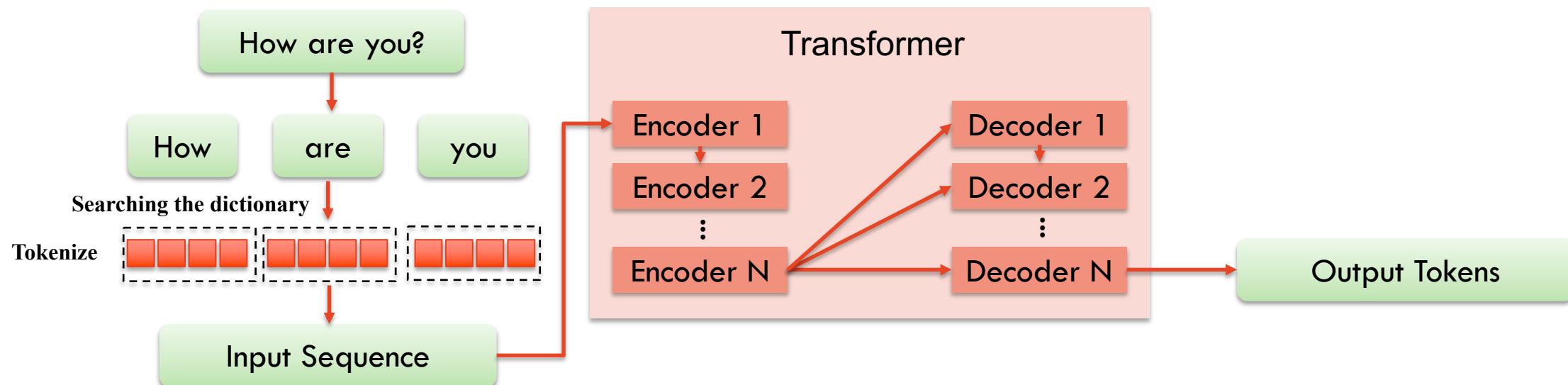
THE UNIVERSITY OF
SYDNEY



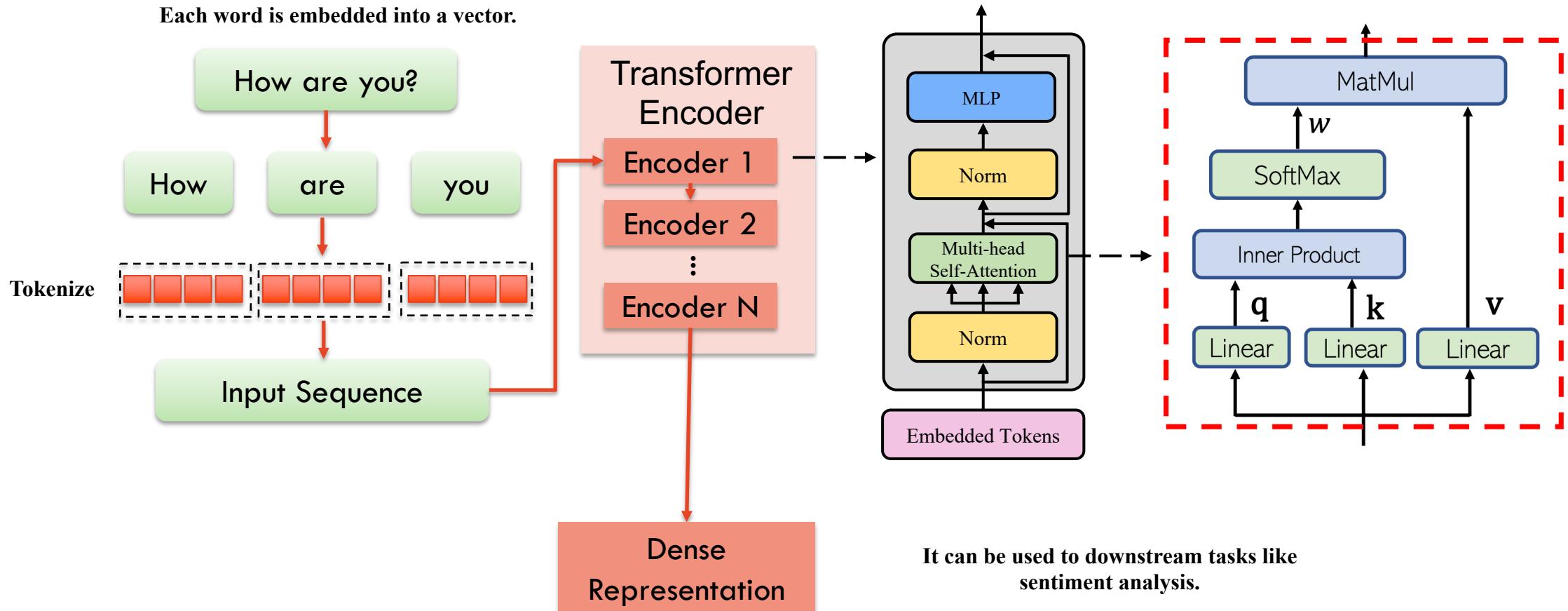
Transformer Architecture



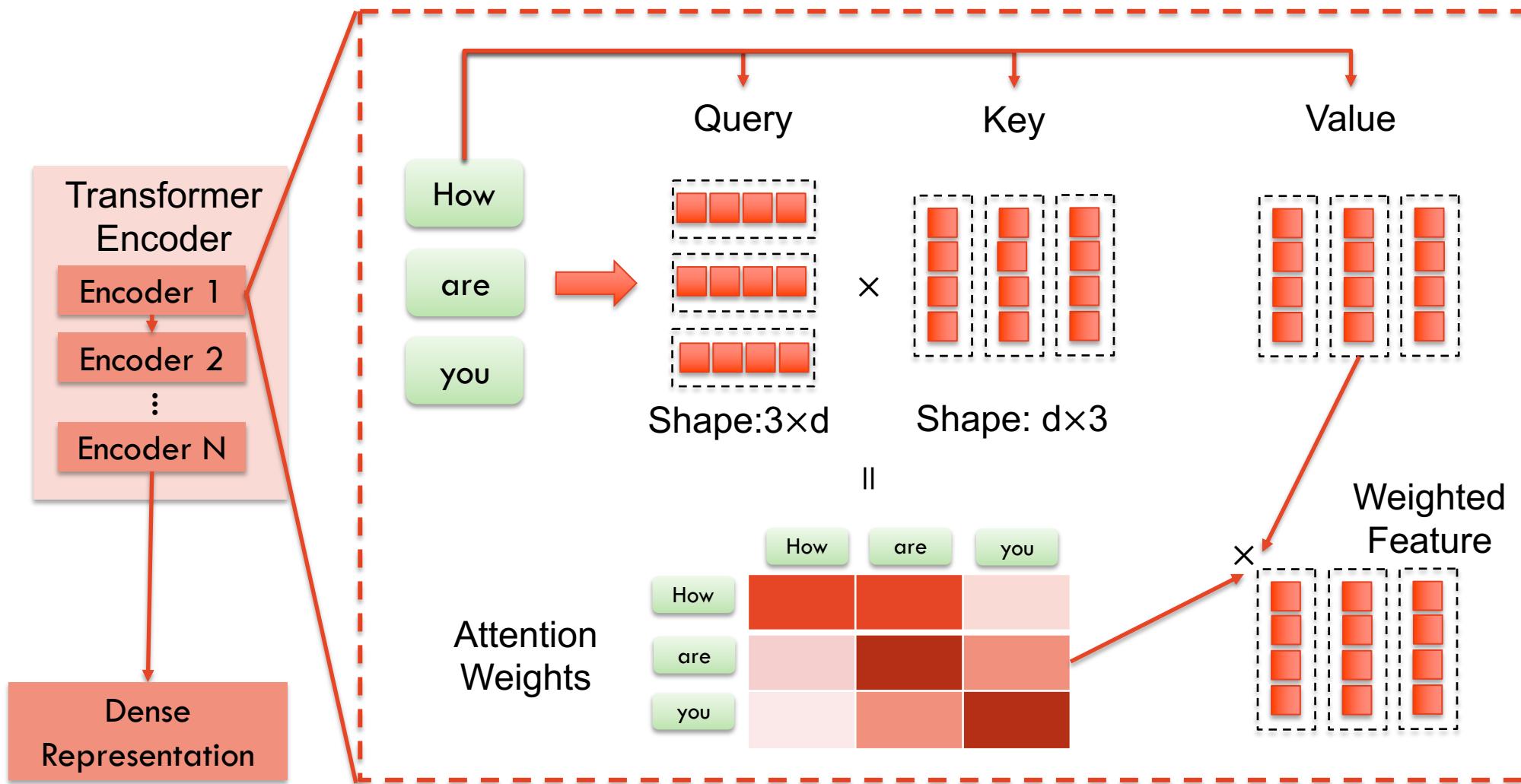
Each word is embedded into a vector.



Transformer Encoder

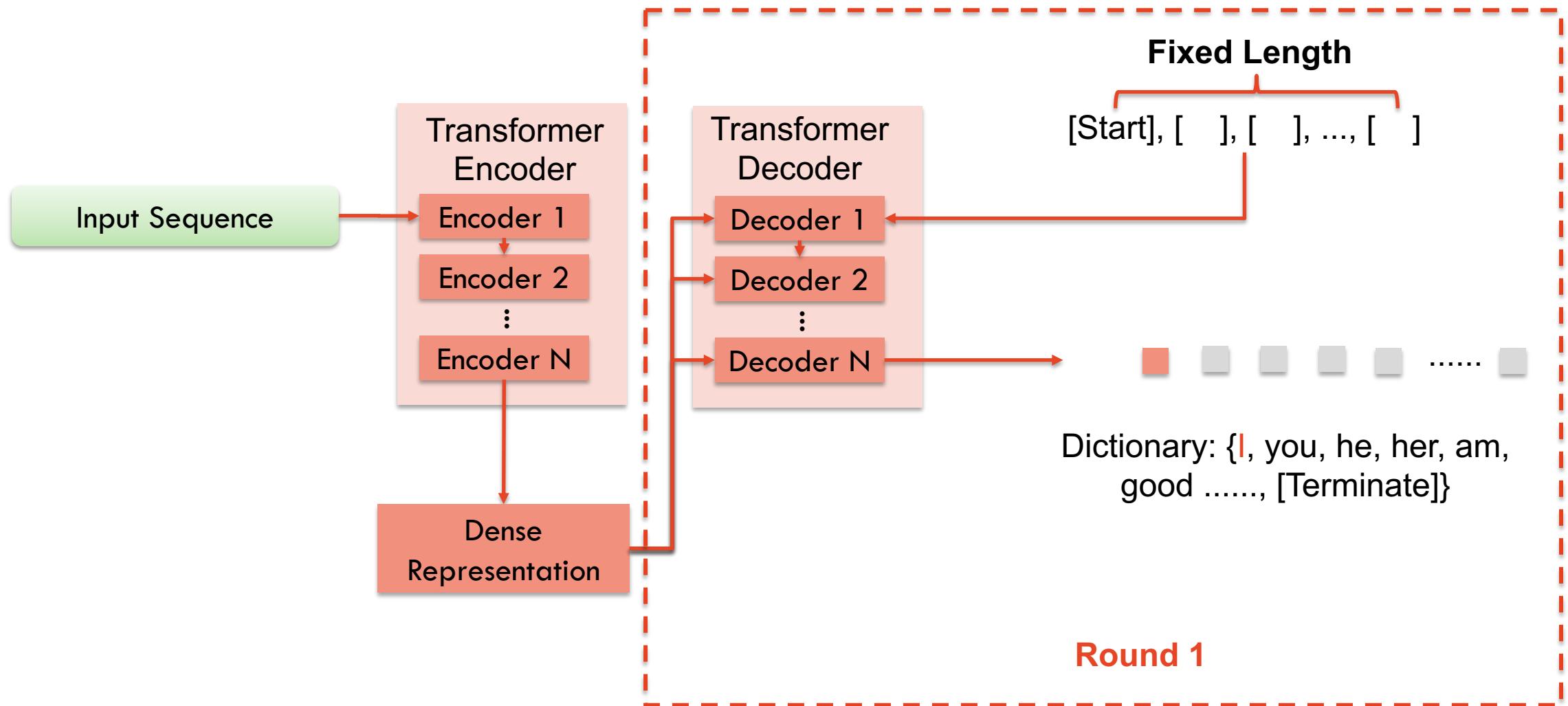


Attention

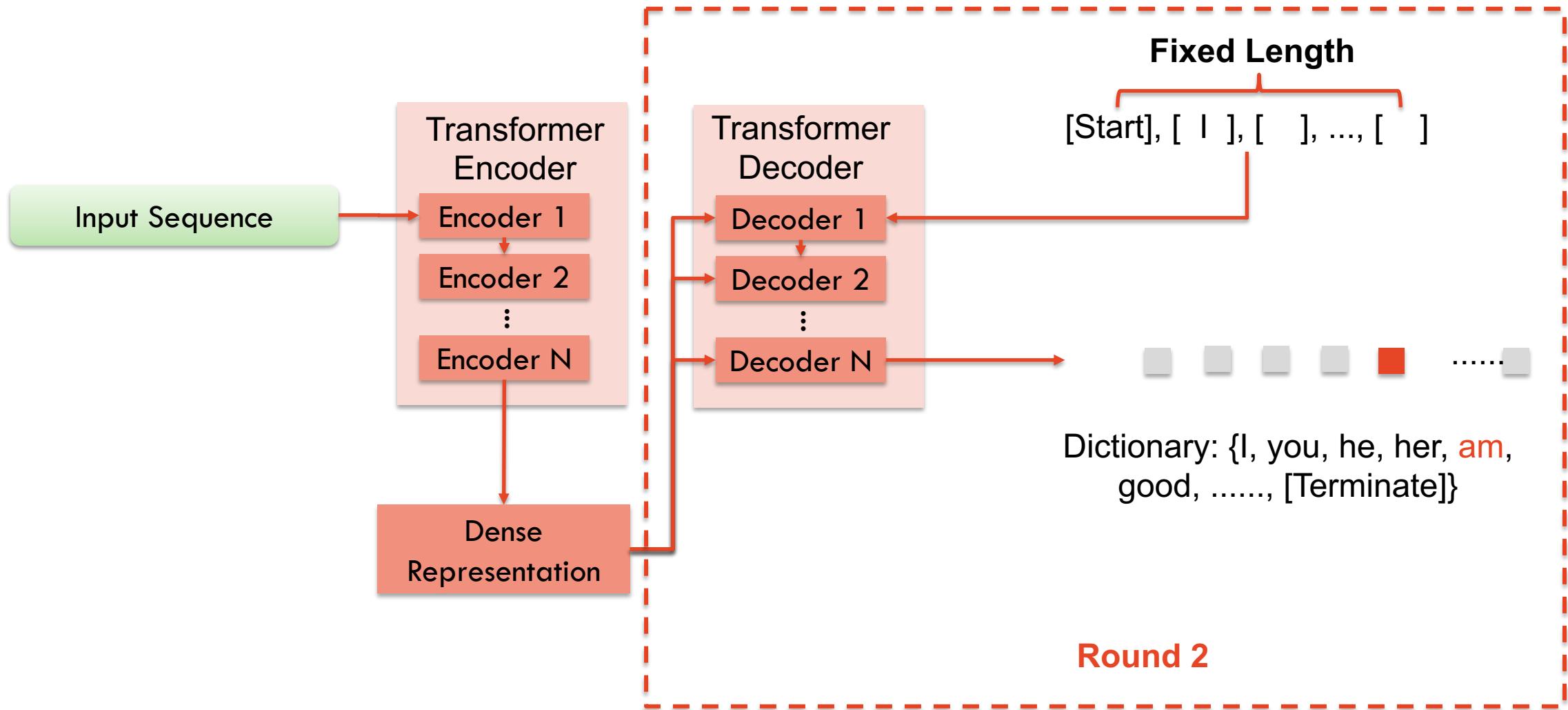


[2] Vaswani et al., Attention is All You Need, NeurIPS 2017.

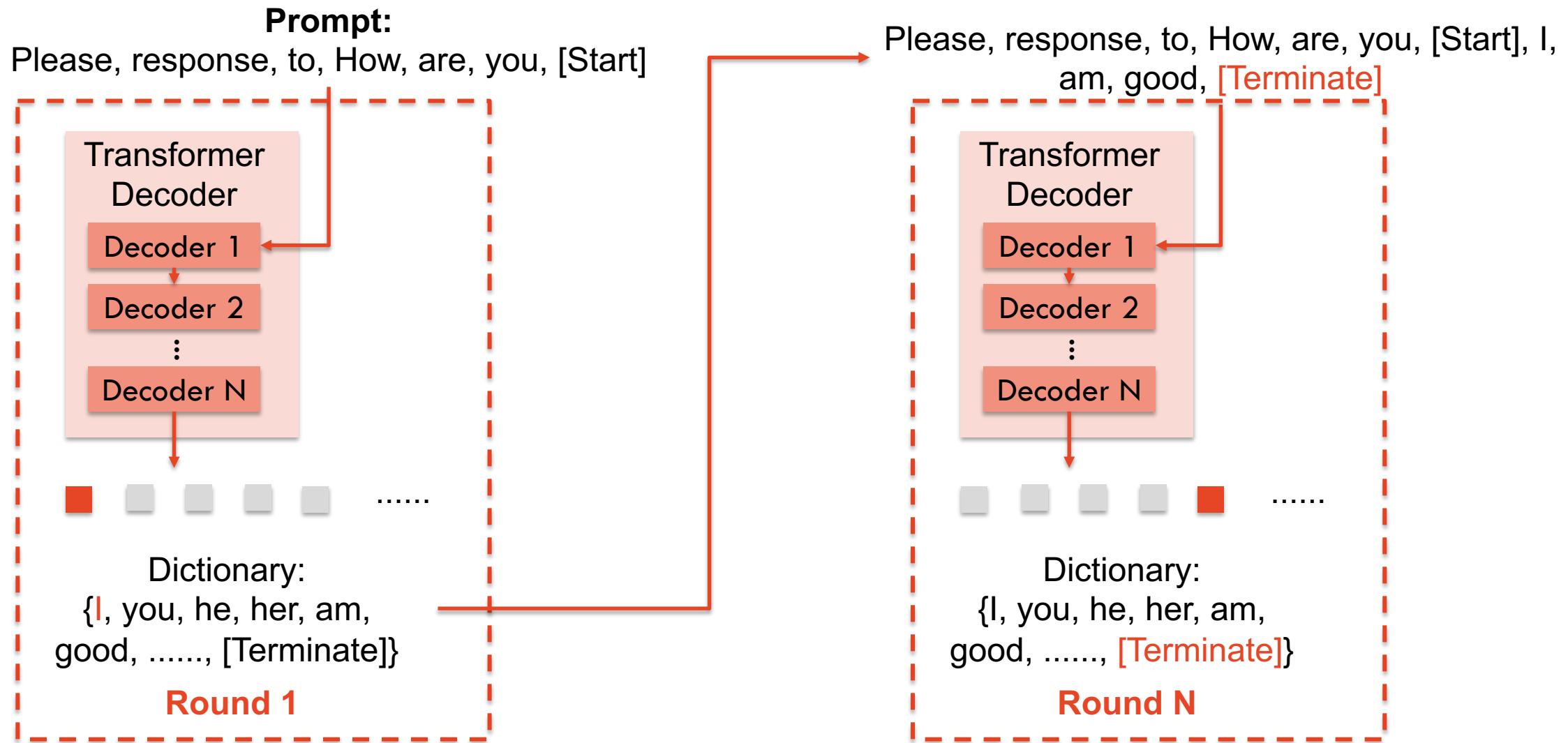
Transformer Decoder – Autoregressive Generation



Transformer Decoder – Autoregressive Generation



Transformer Decoder Only – How GPT works



Decoder Only v.s. Encoder-Decoder

Decoder-only Architecture

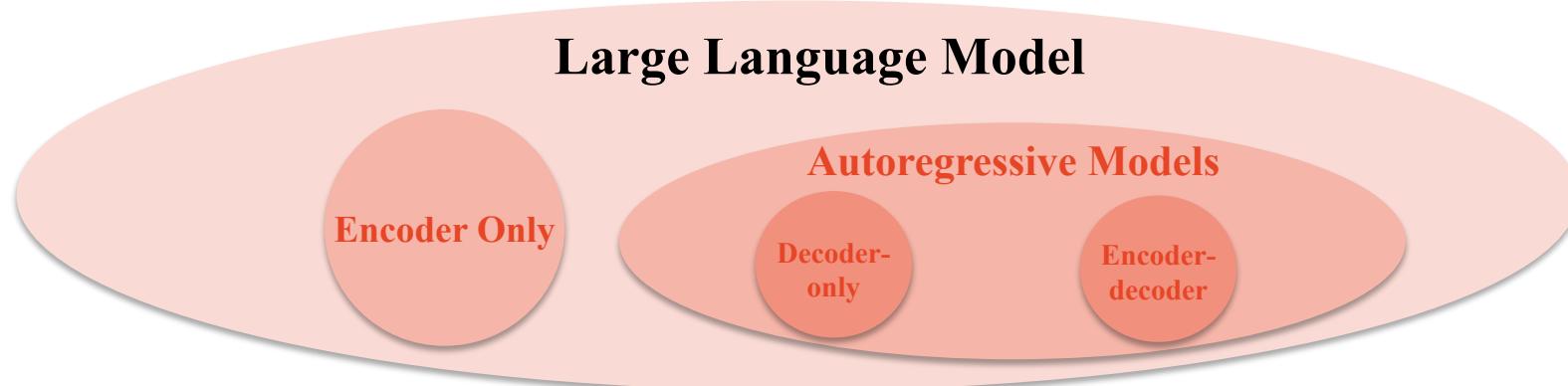
- *Like writing a story one word at a time, only considering what you've already written*

Encoder-Decoder Architecture

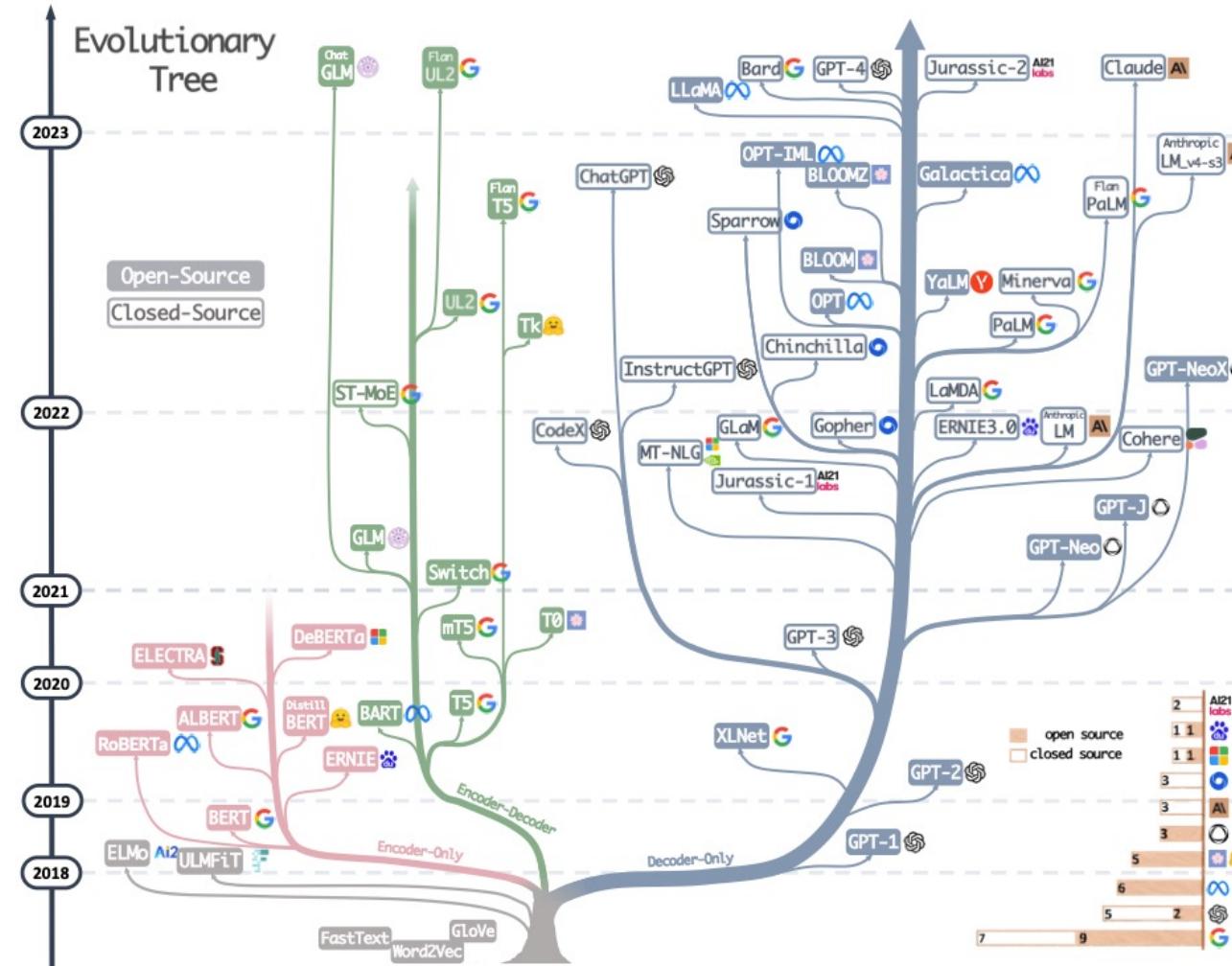
- *Like translating a sentence from one language to another—understand the full meaning first (encoder) and then write the translation (decoder).*

Autoregressive Models

- *Like predictive texting, where each next word is based on what you've written so far.*



Development of LLMs

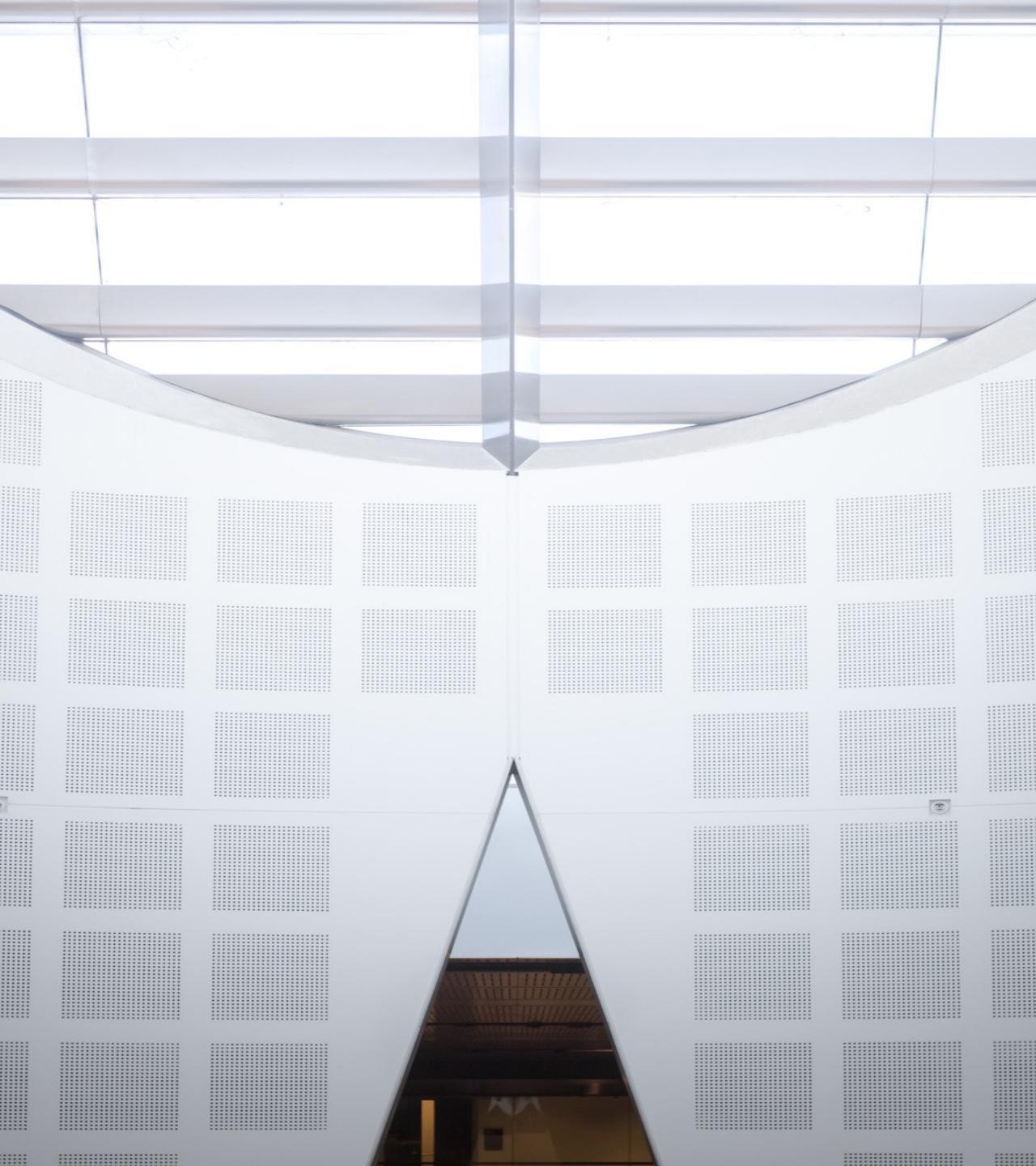


[2] Yang et al., Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond

Seminar 1, Part 1.2: Diffusion Model for Image Generation



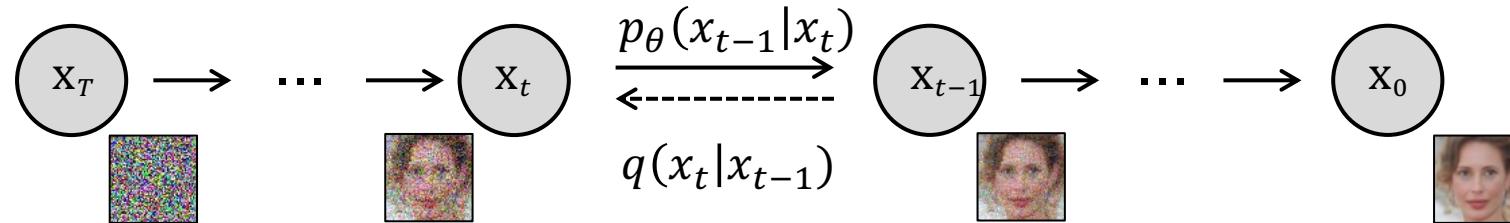
THE UNIVERSITY OF
SYDNEY



Diffusion Model

- Framework

Reverse Process: Noise to Data



Gaussian Noise

Forward Process: Data to Noise

-
- Forward step: (Iteratively) Add noise to the original sample

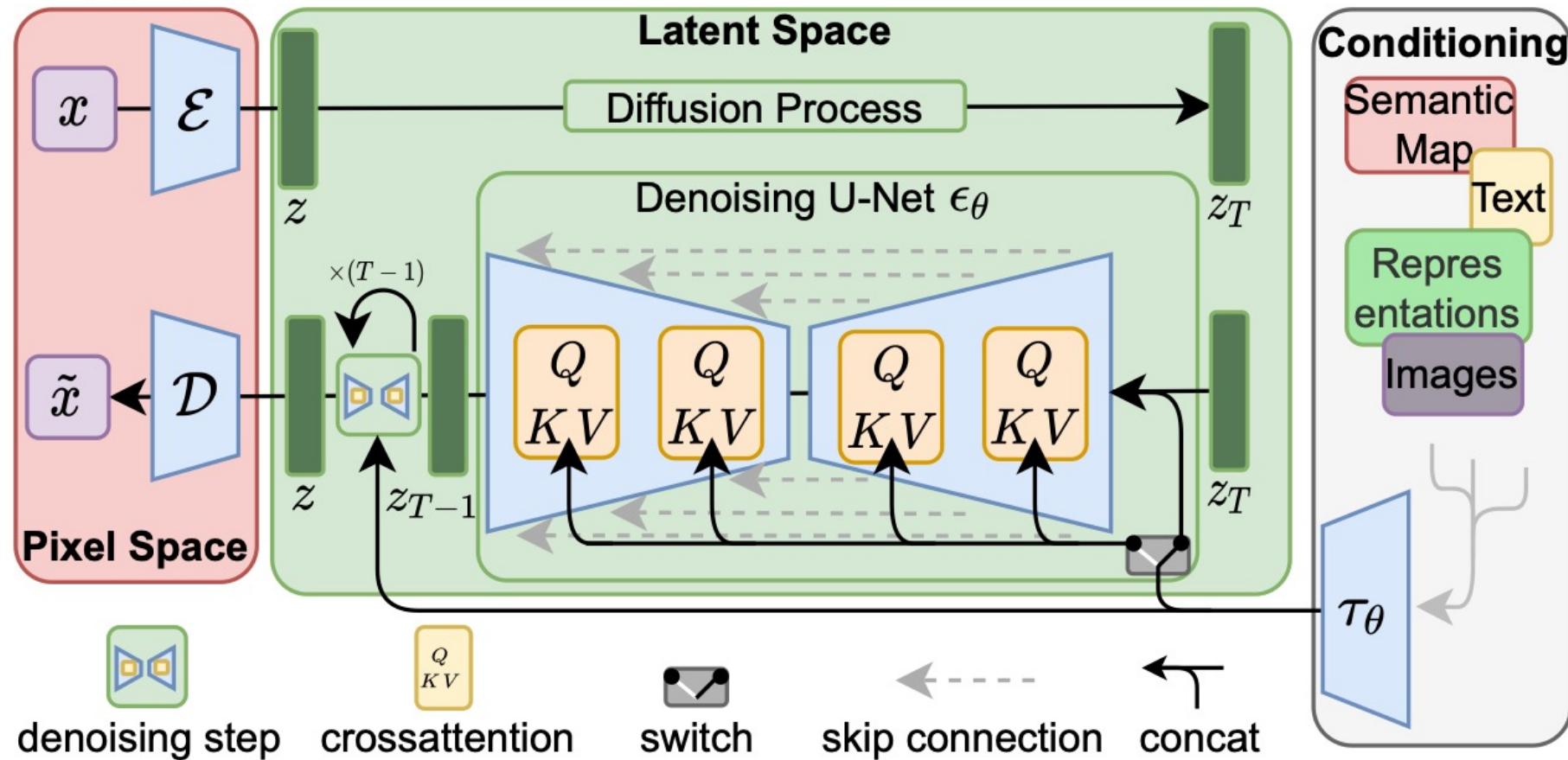
$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I) \quad q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1})$$

- Reverse step: Recover the original sample from the noise

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \rightarrow \text{Represent by a Neural Network}$$

Stable Diffusion

- Framework



[4] High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.

Stable Diffusion

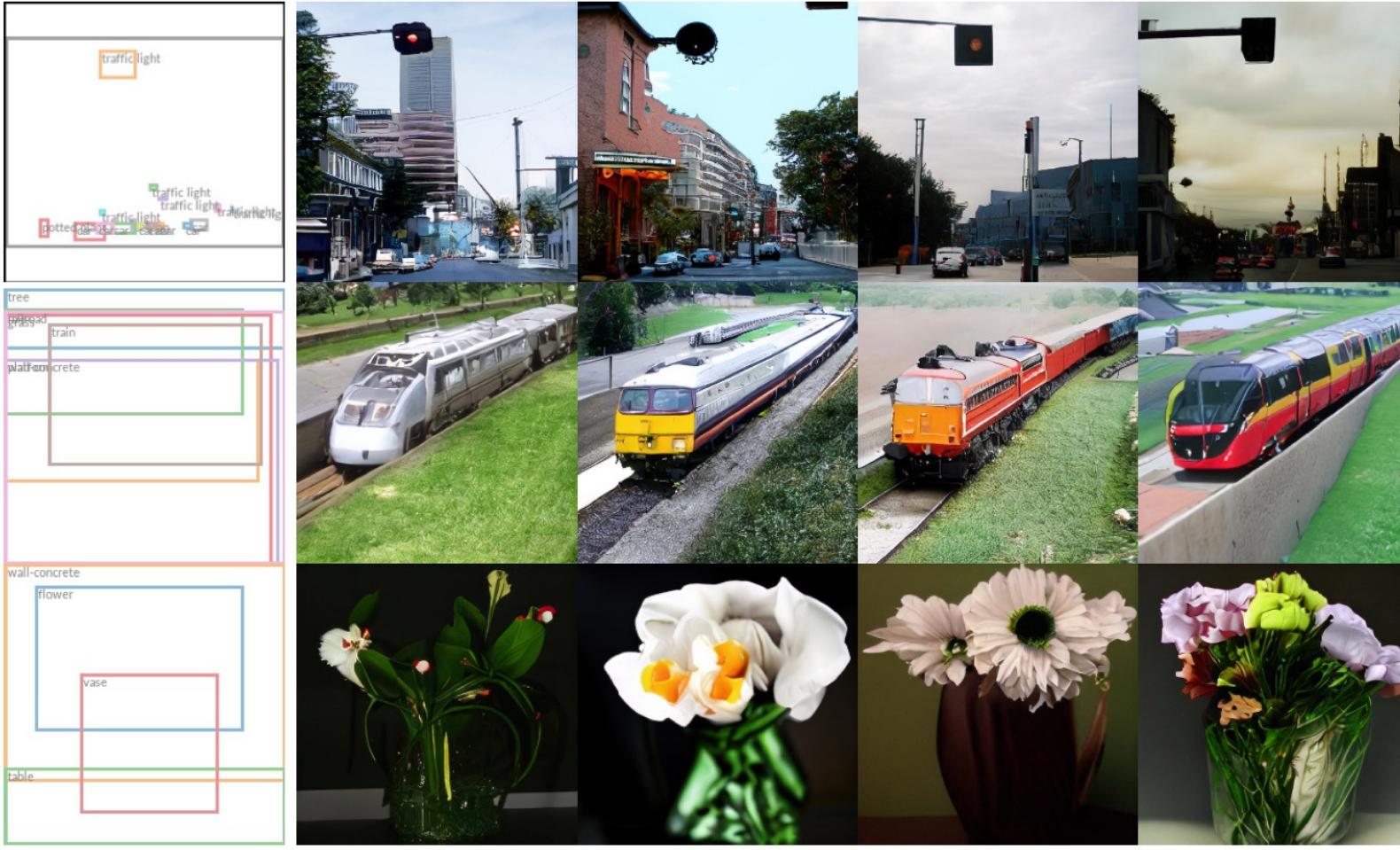
- Text-to-image generation

Text-to-Image Synthesis on LAION. 1.45B Model.

'A street sign that reads "Latent Diffusion"'	'A zombie in the style of Picasso'	'An image of an animal half mouse half octopus'	'An illustration of a slightly conscious neural network'	'A painting of a squirrel eating a burger'	'A watercolor painting of a chair that looks like an octopus'	'A shirt with the inscription: "I love generative models!"'

Stable Diffusion

- Layout-to-image generation



Stable Diffusion

- Image Inpainting



Summary

- We give a brief introduction GenAI architectures including Transformer, Autoregressive models, and Diffusion Models.
- We introduce some popular applications of GenAI like Stable Diffusion, Sora, etc.
- **Future trend** is Multi-modal Large Language Model, which has stronger capability in dealing with multi-modal data, and has better ability in generalization.

Thank you!
Q&A