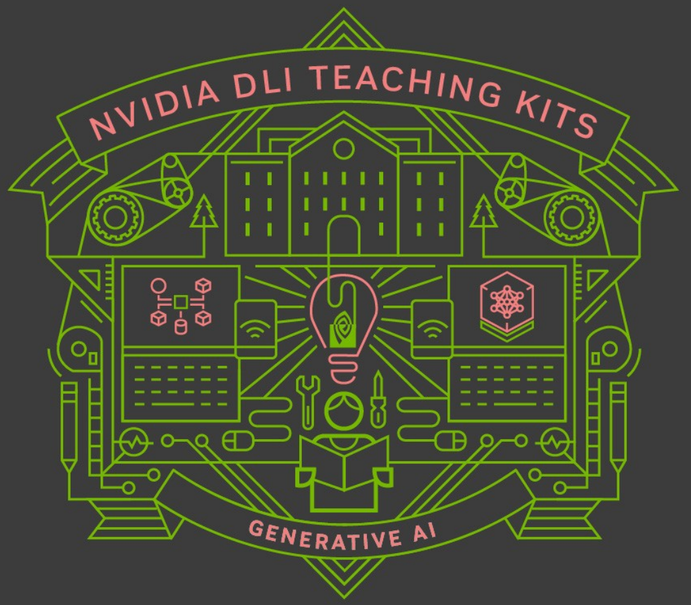
Lecture 1.1 – Introduction to the Generative AI Course

Generative AI Teaching Kit

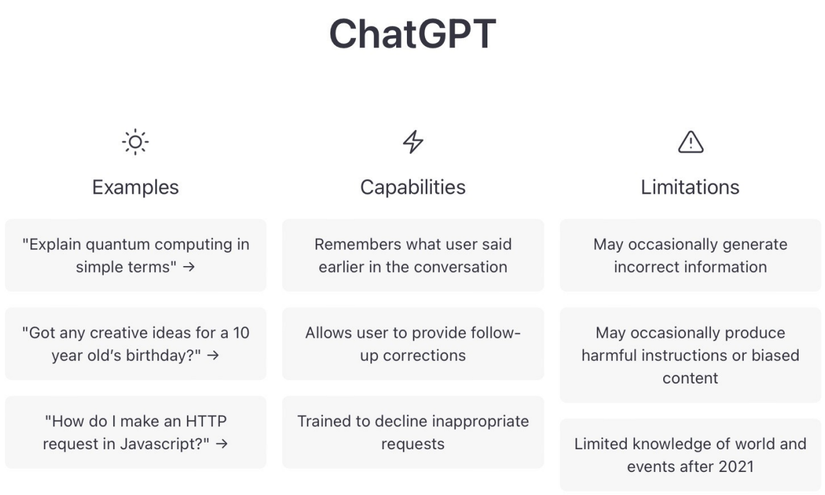


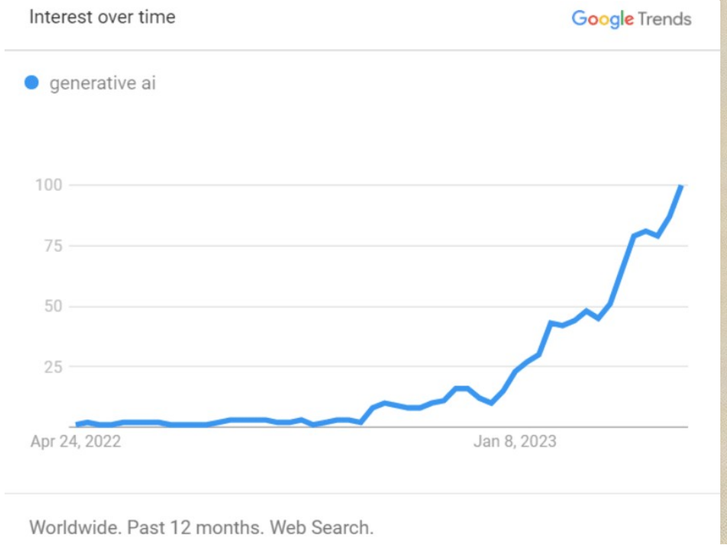
Course Structure

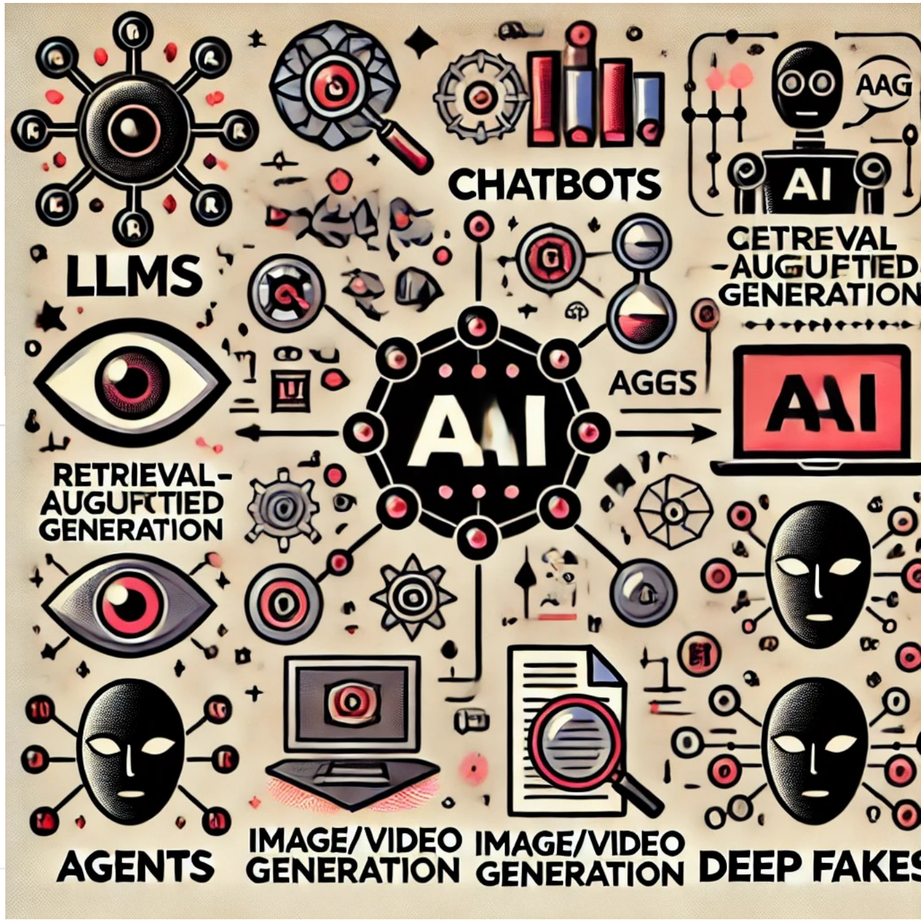
Motivation for this course

Over the last two years you've probably hear all of these GenAI buzzwords:

- LLMs
- Chatbots
- RAG
- Agents
- Image/Video generation
- Deep Fakes





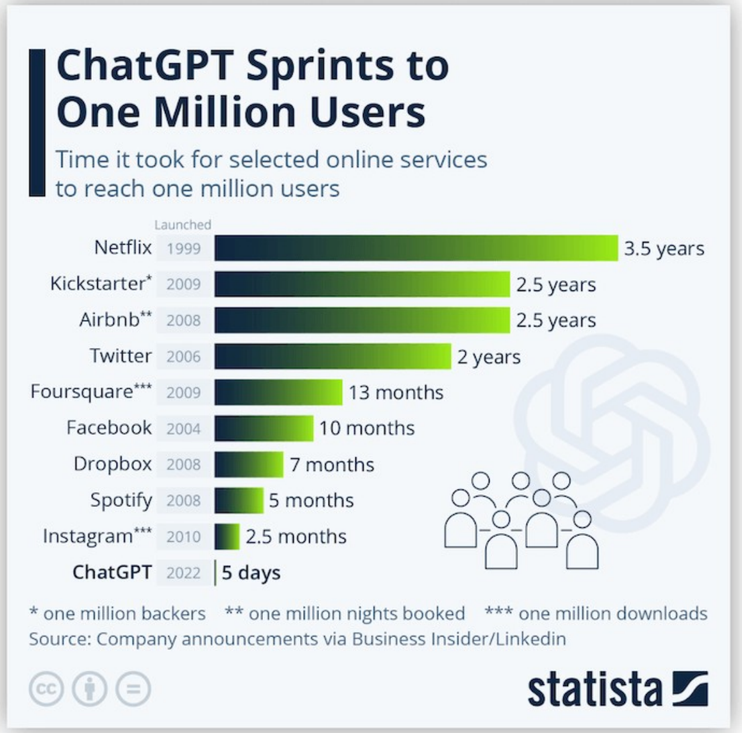


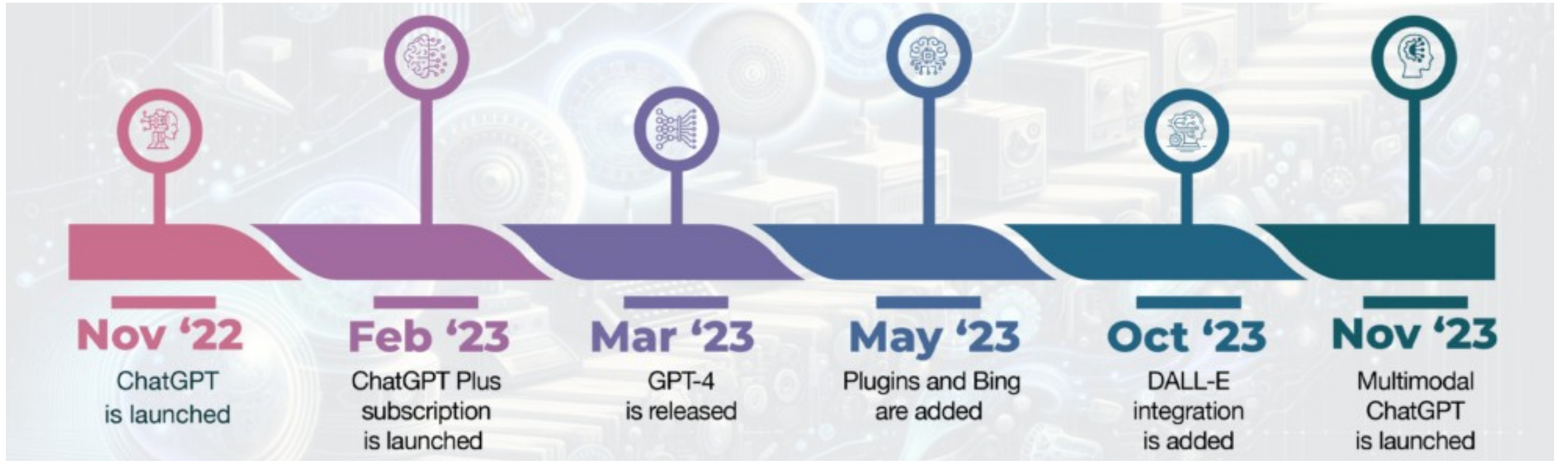
DARTMOUTH ENGINEERING | NVIDIA

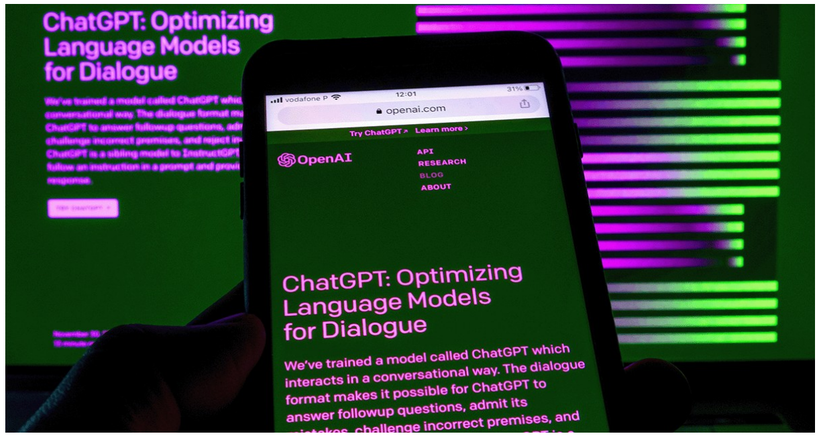
Motivation for this course

ChatGPT Changed the world of knowledge forever

- Released Nov 2022 Took the world by storm
- How does it know so much?
- Is it “thinking” ?
- What is it made of?





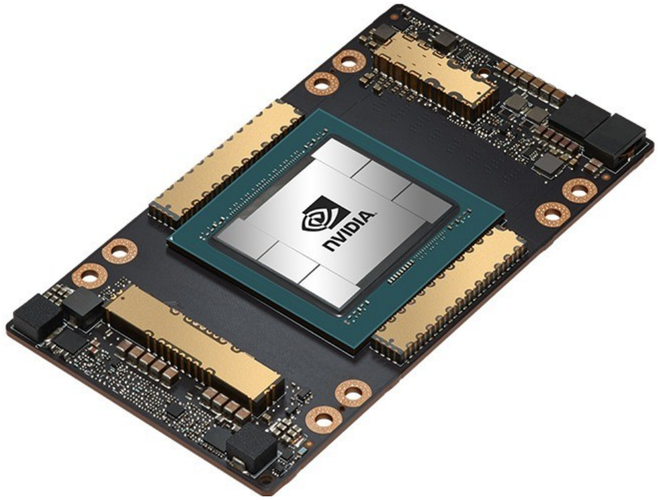


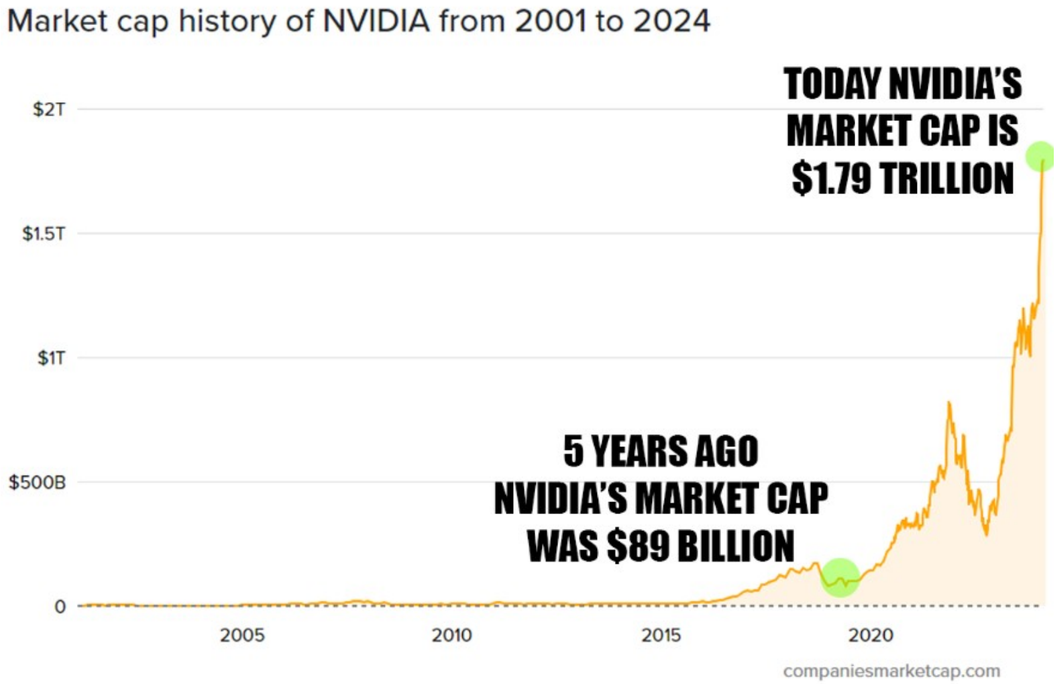
DARTMOUTH ENGINEERING | NVIDIA

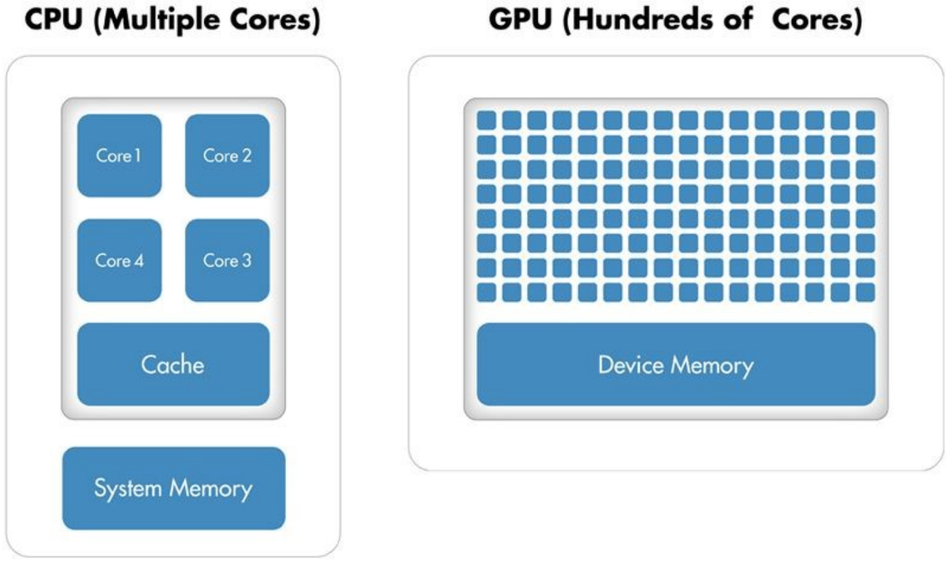
Motivation for this course

In 2024, NVIDIA's market cap reached 3T! Why?

Why are GPUs are now the hottest commodity for GenAI ?

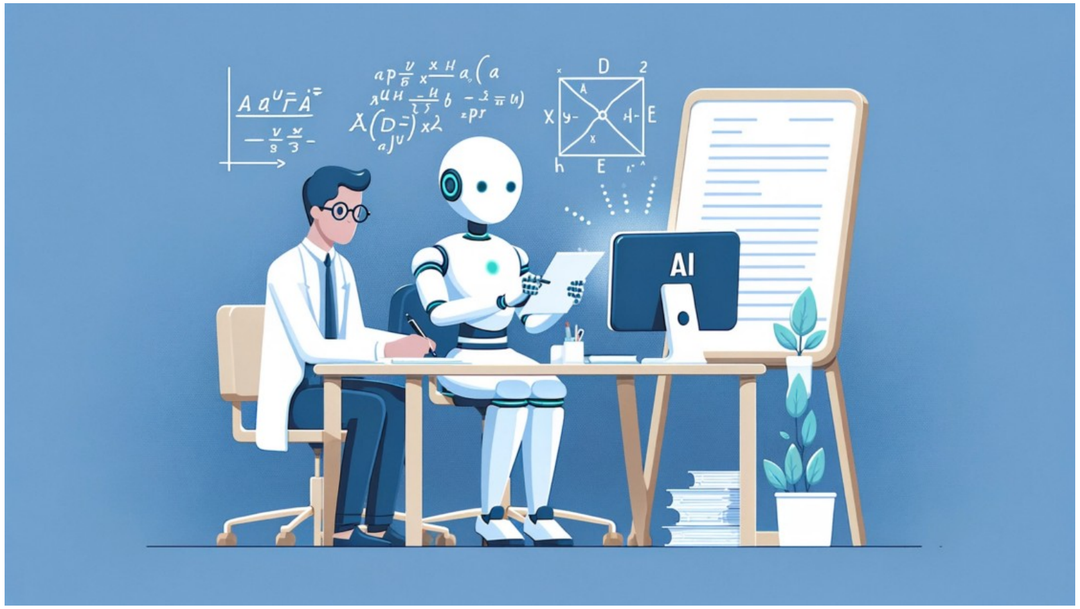






Motivation for this course

- GenAI poses a challenge for Academia
- Pace of development has far outpaced adoption in schooling
- How do we do teaching given these extremely knowledgeable, highly accessible tools?
- Familiarity with the abilities and limits of these tools is vital to maintain competitiveness for both teachers and students

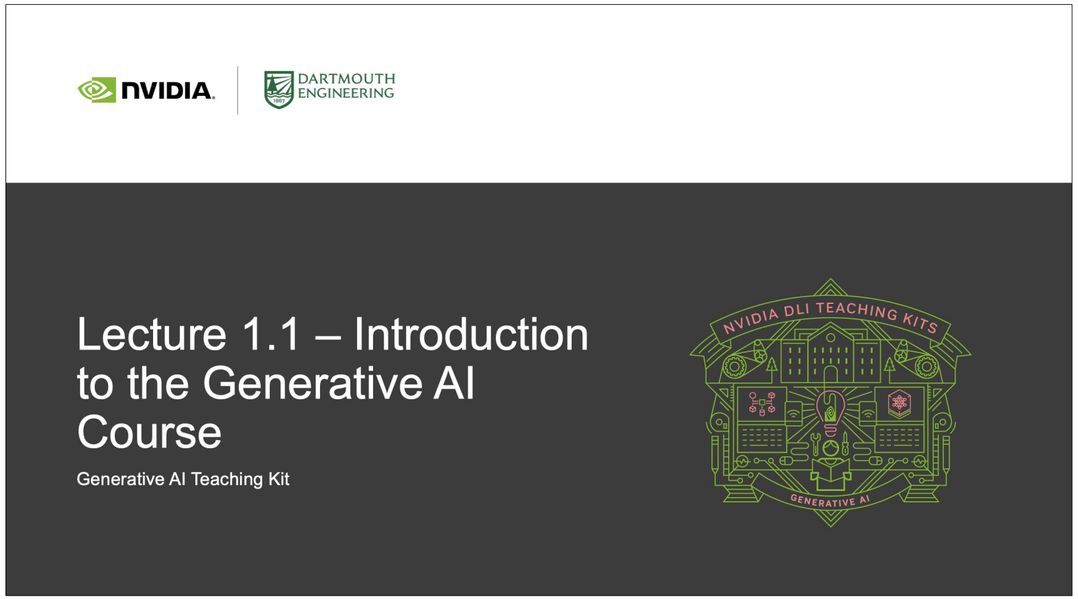


DARTMOUTH ENGINEERING | NVIDIA

Structure of the course

How each module of the course is designed

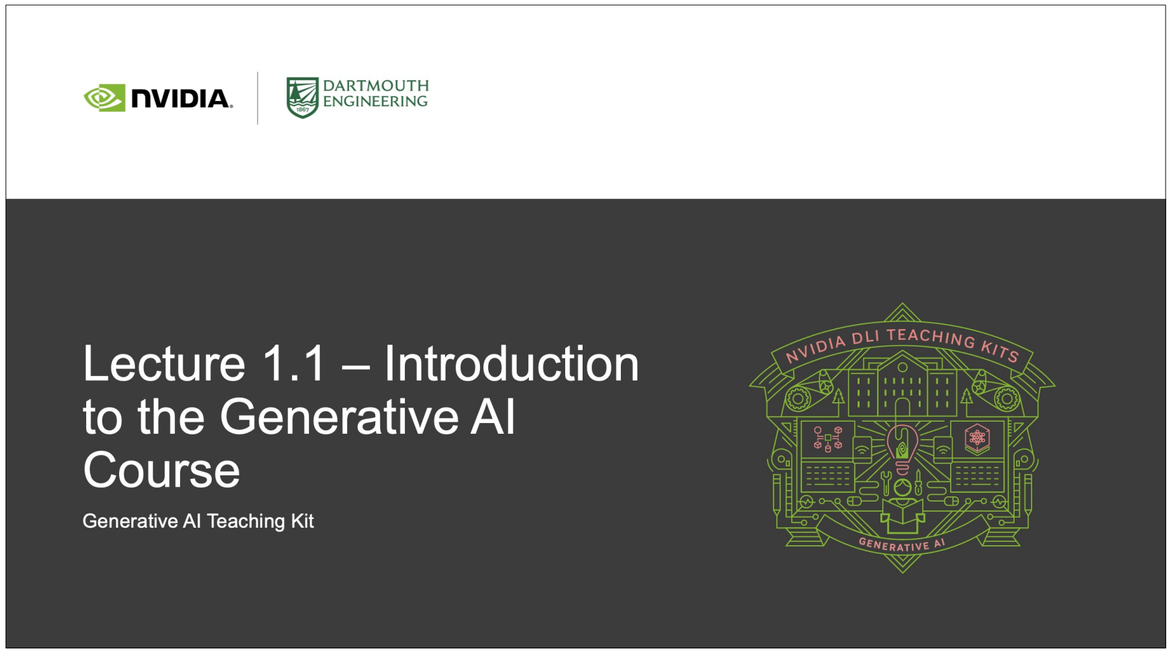
- Online Syllabus
  - Lays out entire course including planned future content
- Lecture Slides
  - Introduce content
- Demo Notebooks
  - Show the tools in practice
- Knowledge Checks
  - Keeping track of the content
- Lab Notebooks
  - Your turn to build with the tools
- DLI Online Courses
  - Online, self-paced courses offering certificate



Structure of the course

Lecture Slides

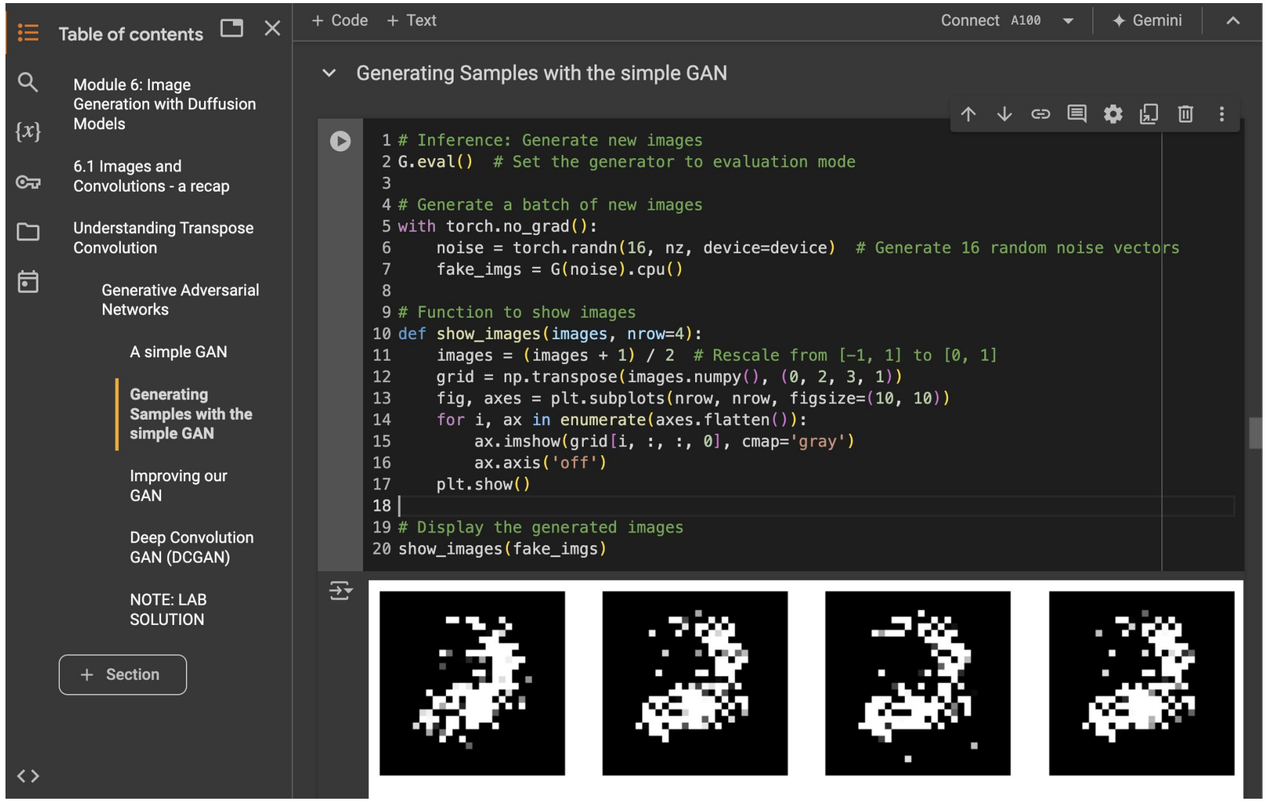
- 2 or three lectures per module
- Designed to introduce the necessary information needed to understand what/how things work
- Not deep mathematical proofs, more applied science/engineering focus



Structure of the course

Demo Notebooks

- Designed to be used during class time for the instructor to show how these tools/topics can work
- Built with an interactive Python notebook, namely Google Colab



DARTMOUTH ENGINEERING | NVIDIA

Structure of the course

Knowledge Checks

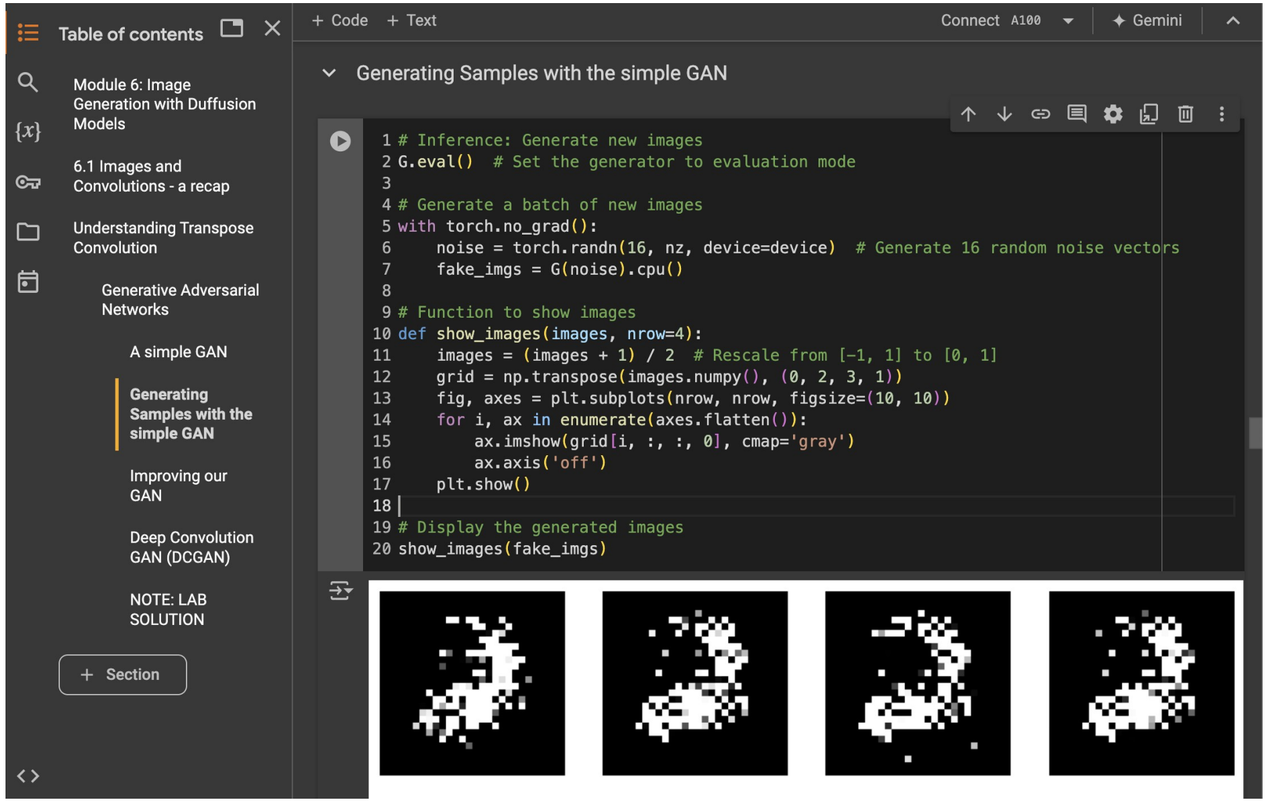
- Confirmation of the development of a working knowledge of GenAI topics

Question	Option A	Option B	Option C	Option D
What is the main advantage of Few-Shot Learning in LLMs?	Requires a large labeled dataset	Reduces the need for labeled data	Improves the model's interpretability	Requires extensive fine-tuning
How does self-supervised learning work?	It uses labeled data for predictions	It masks parts of the data as input/output pairs	It uses external labels for classification	It combines labeled and unlabeled data
What problem does self-supervised learning address?	It reduces the need for labeled data	It addresses data scarcity in supervised learning	It simplifies the model training process	It ensures consistency in model predic
What is the role of autoregressive learning in LLMs?	Predicts the next token based on prior tokens	Enhances the model's output accuracy	Identifies relationships between labeled data	Increases the model's parameter coun
What is Few-Shot Learning in the context of GPT-3?	Training the model with many labeled examples	Providing a few examples before the main input	Building a single model for multiple tasks	Training the model on multiple dataset
How does in-context learning differ from fine-tuning?	In-context learning is task-specific	In-context learning uses the model's parameters flexibly	In-context learning is only used during training	In-context learning replaces fine-tunin
What is the purpose of prompt templating in LLMs?	To create unique responses	To format inputs consistently	To standardize outputs	To introduce randomness in outputs
Why is temperature an important factor in LLM outputs?	It controls the sequence length	It affects the randomness of outputs	It sets the learning rate	It determines the output length

Structure of the course

Lab Notebooks

- Based on the demo notebooks
- These will give you the chance to implement the tools and explore the topics covered in each module

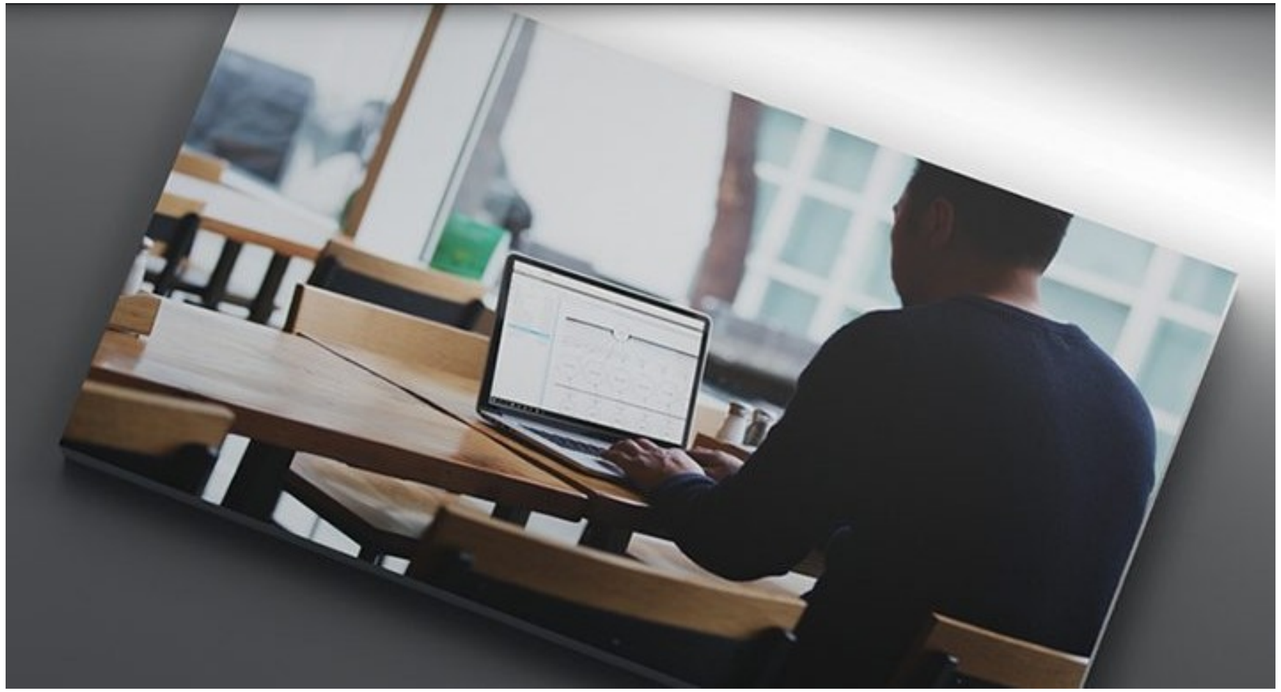


DARTMOUTH ENGINEERING | NVIDIA

Structure of the course

DLI Online Courses

- Related self-contained, self-paced DLI courses
- Free for students using promo codes provided
- Runs on DLI Platform instead of Colab
- Offers student certificates of competency based on built-in assessments



Tools used in GenAI

GenAI Tools in this Course

Python is the most widely used language in the AI/ML community, making it the go-to language for developing Generative AI models.
Python's simple syntax and readability allow developers and researchers to quickly prototype and implement complex algorithms.

- TensorFlow & PyTorch: Essential for building and training deep learning models, which are at the core of GenAI.
- Hugging Face Transformers: A popular library for working with large language models (LLMs), crucial for tasks like text generation and RAG.
- OpenCV & PIL: Important for image and video manipulation in GenAI applications like deep fakes and image generation

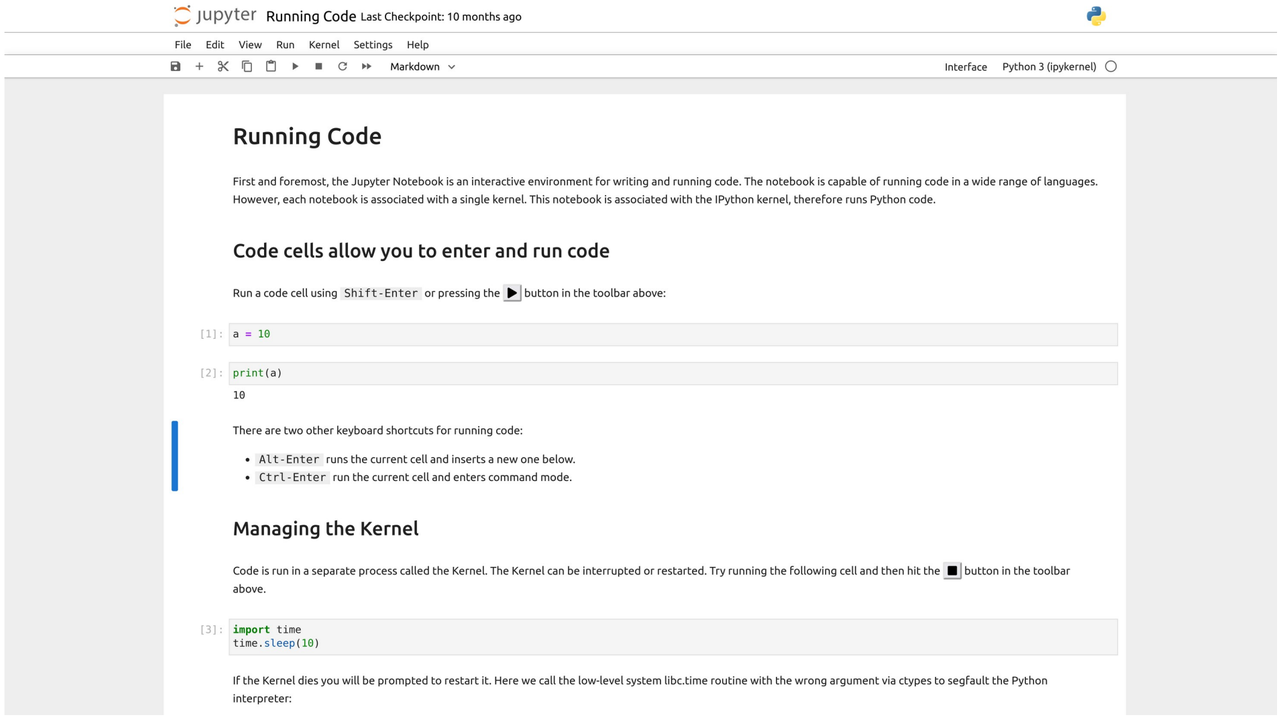
GenAI Tools in this Course

Coding Notebooks

The primary platform to work with these AI tools and code libraries will be through the Google Colab notebooks

These notebooks allow for easy interaction with the data and models in a visual manner

Platforms like Google Colab and other cloud notebook providers also supply access to cloud-based GPUs and compute infrastructure.



DARTMOUTH ENGINEERING | NVIDIA

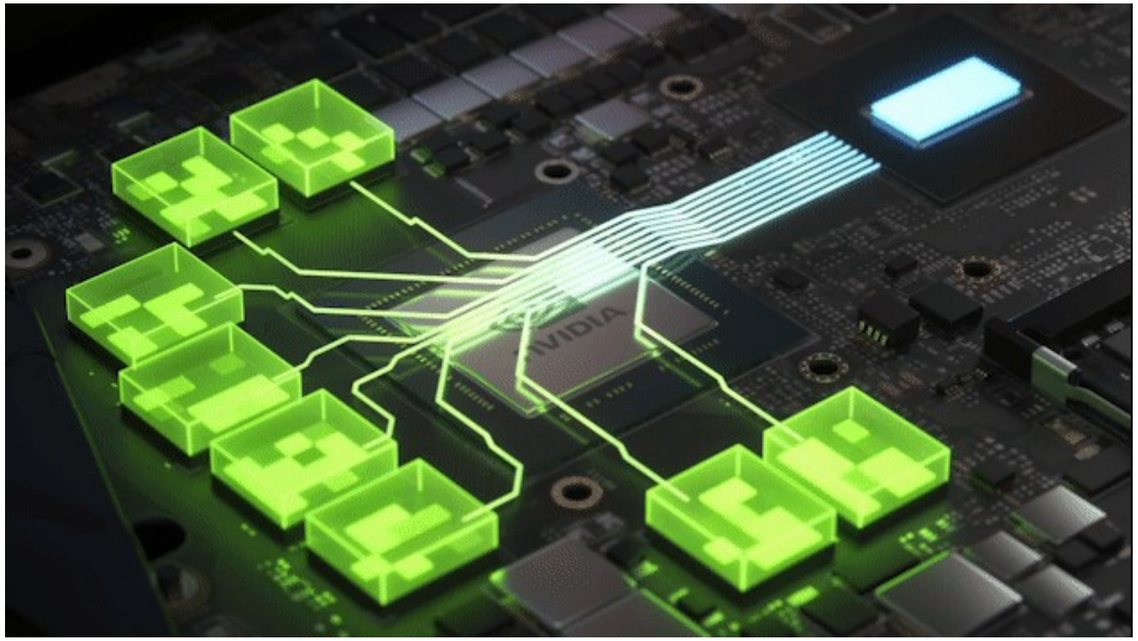
GenAI Tools in this Course

The Importance of GPUs in GenAI

Parallelism: GPUs have thousands of cores optimized for parallel processing, allowing them to handle many operations at once. This is essential for the matrix multiplications and other computations involved in AI inference.

Throughput: GPUs can process large batches of data simultaneously, which is critical for applications like image generation or video synthesis where high throughput is needed.

Latency: GPUs reduce latency by executing multiple operations in parallel, which speeds up the time it takes to get from input to output in GenAI models.



Course Content

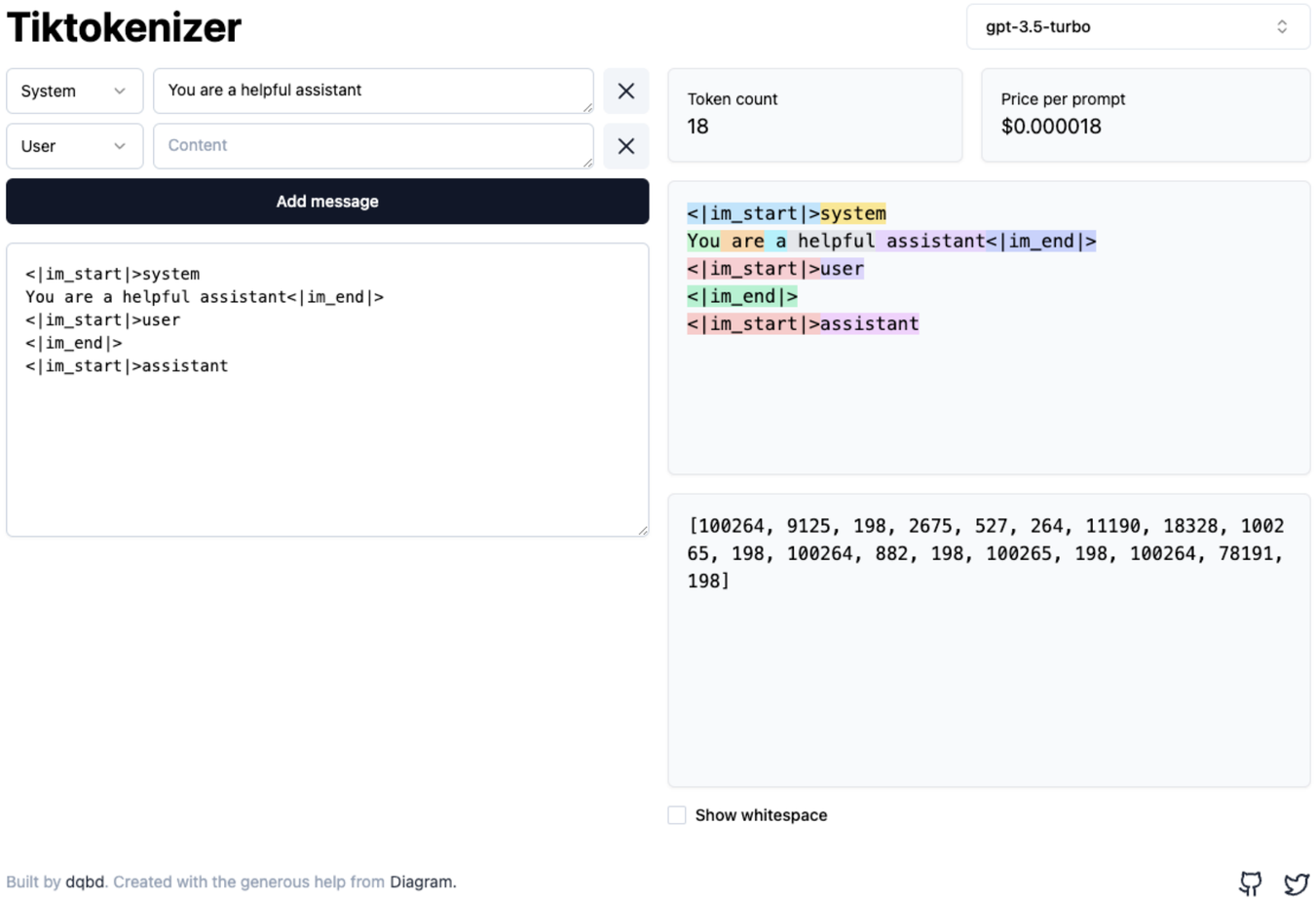
What this course will cover

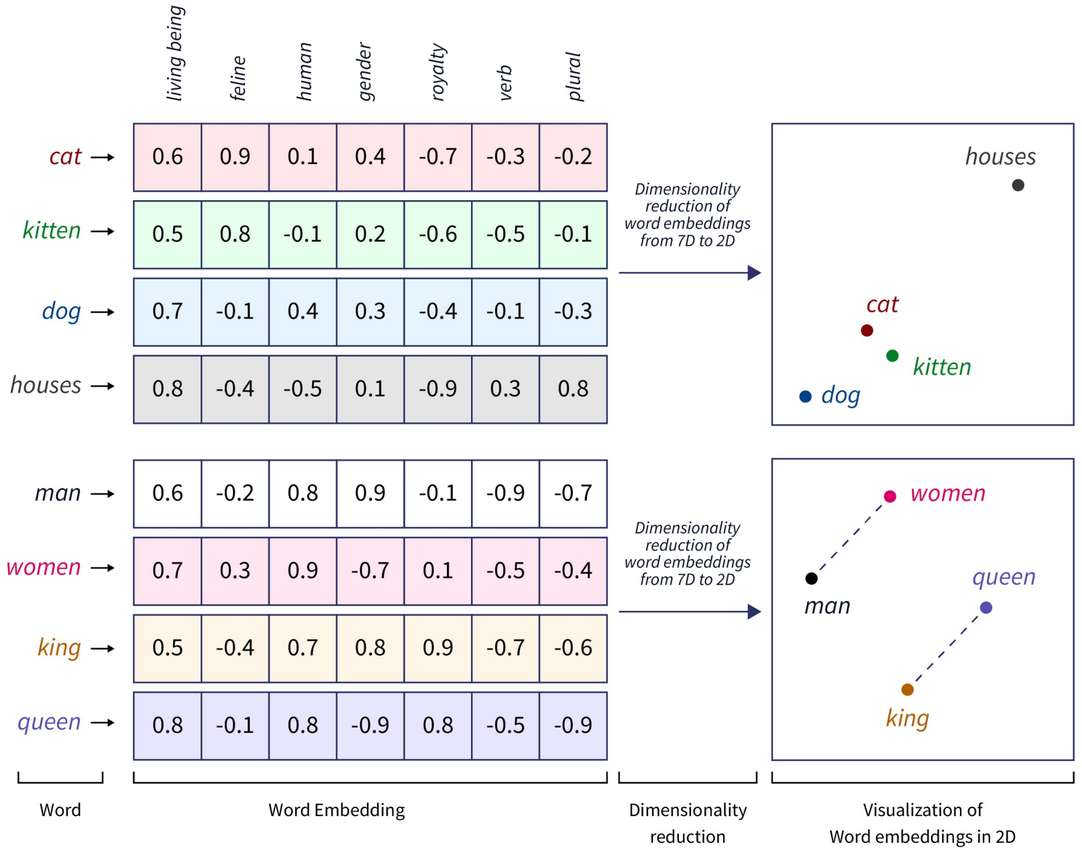
Modules

1. Introduction to Generative AI
2. Word Embeddings, Tokens, and NLP
3. Large Language Models and the Transformer
4. LLM Scaling Laws and LLM Families
5. Multimodal Learning and its Applocations
6. Diffusion Models in Generative AI
7. Model Training (Pre-Training, Instruction Following, and PEFT)
8. LLM Orchestration
9. Scaling Model Training to Distributed Workloads

Full online syllabus

2. Word Embeddings, Tokens, and NLP

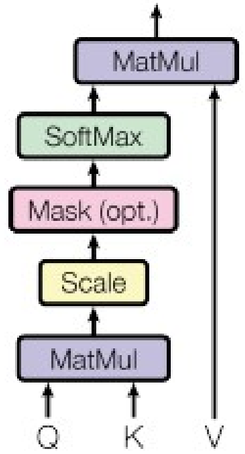




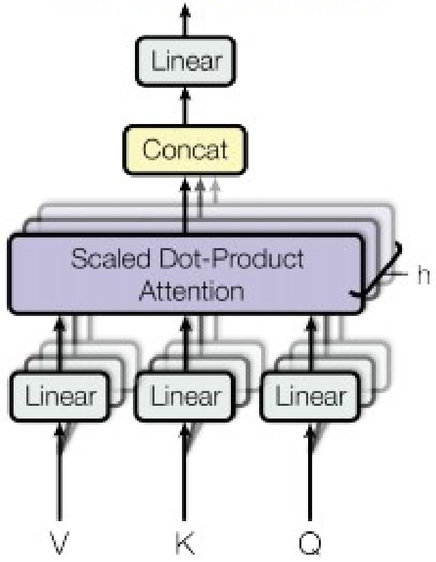
DARTMOUTH ENGINEERING | NVIDIA

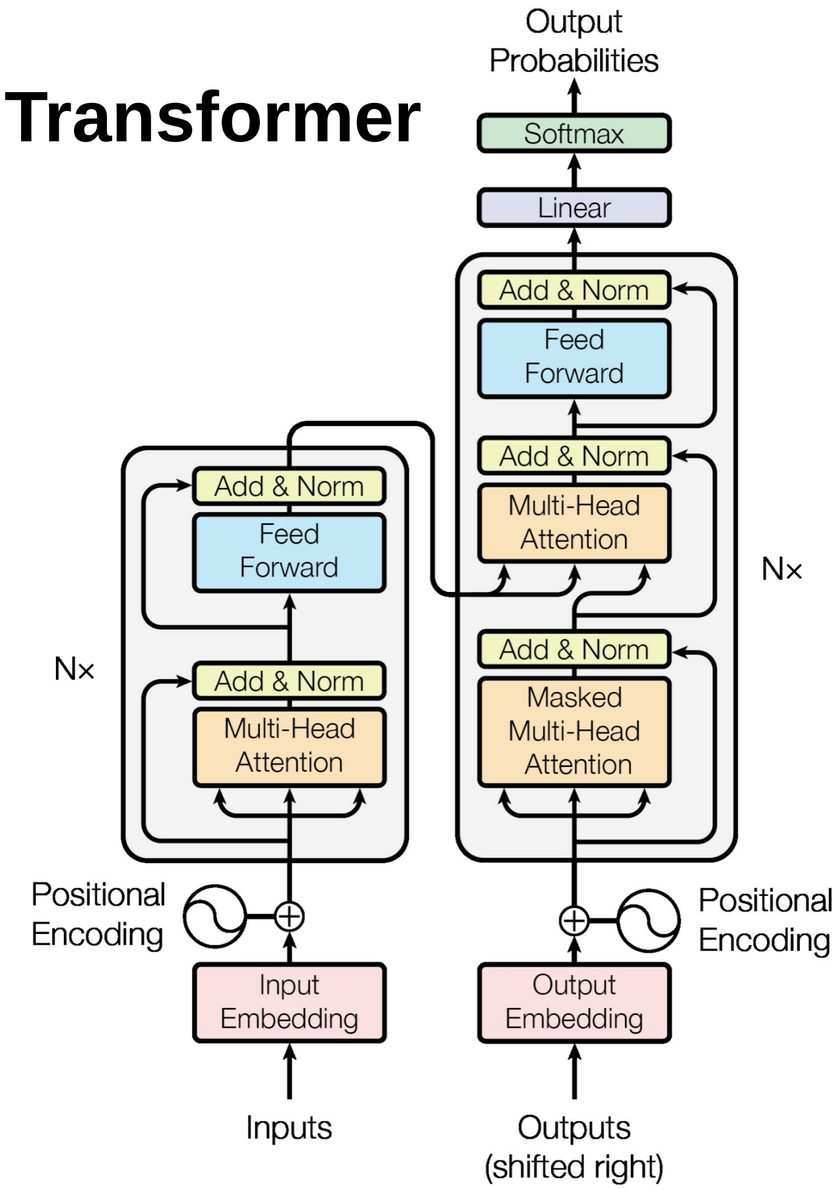
3. Large Language Models and the Transformer

Scaled Dot-Product Attention

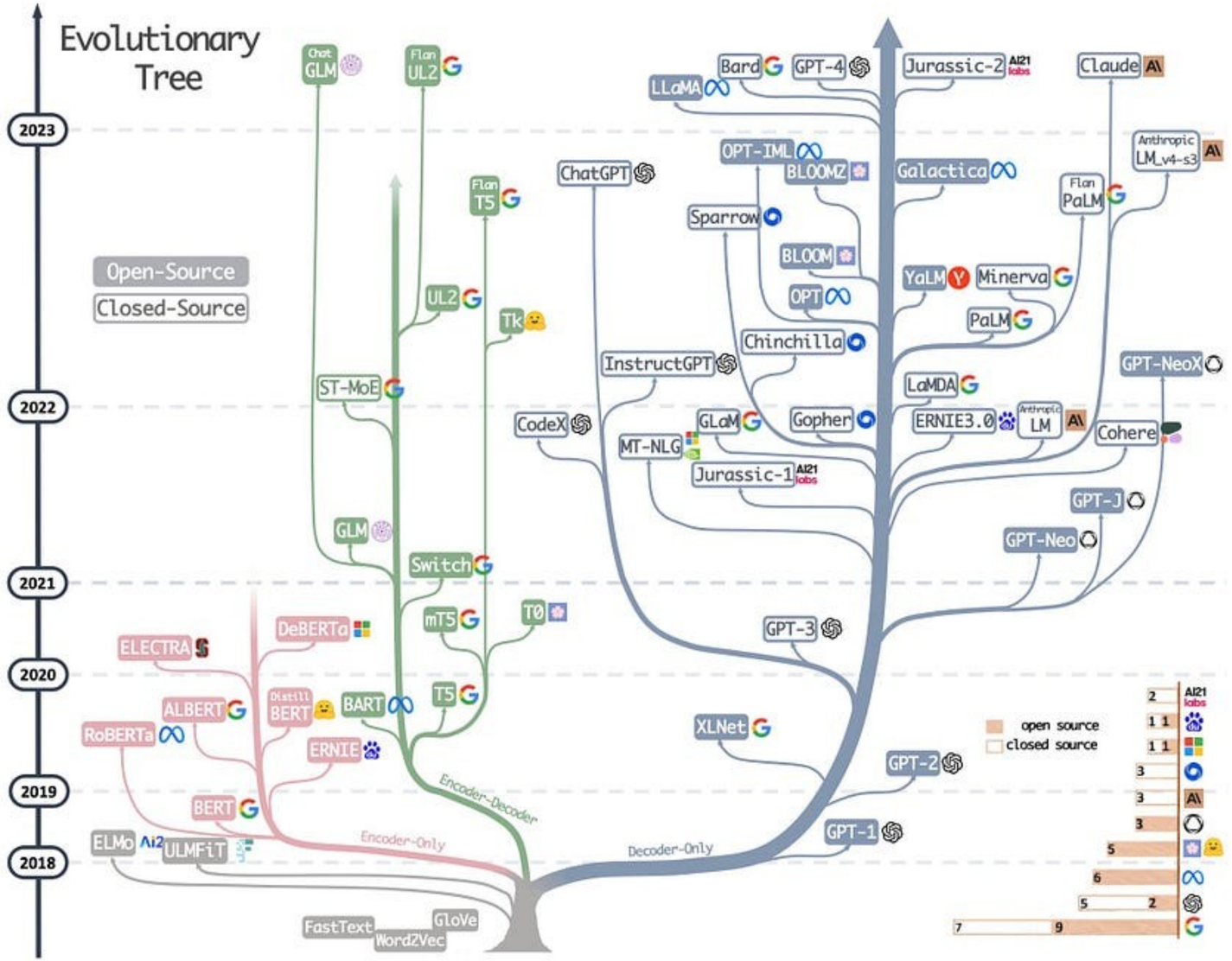


Multi-Head Attention

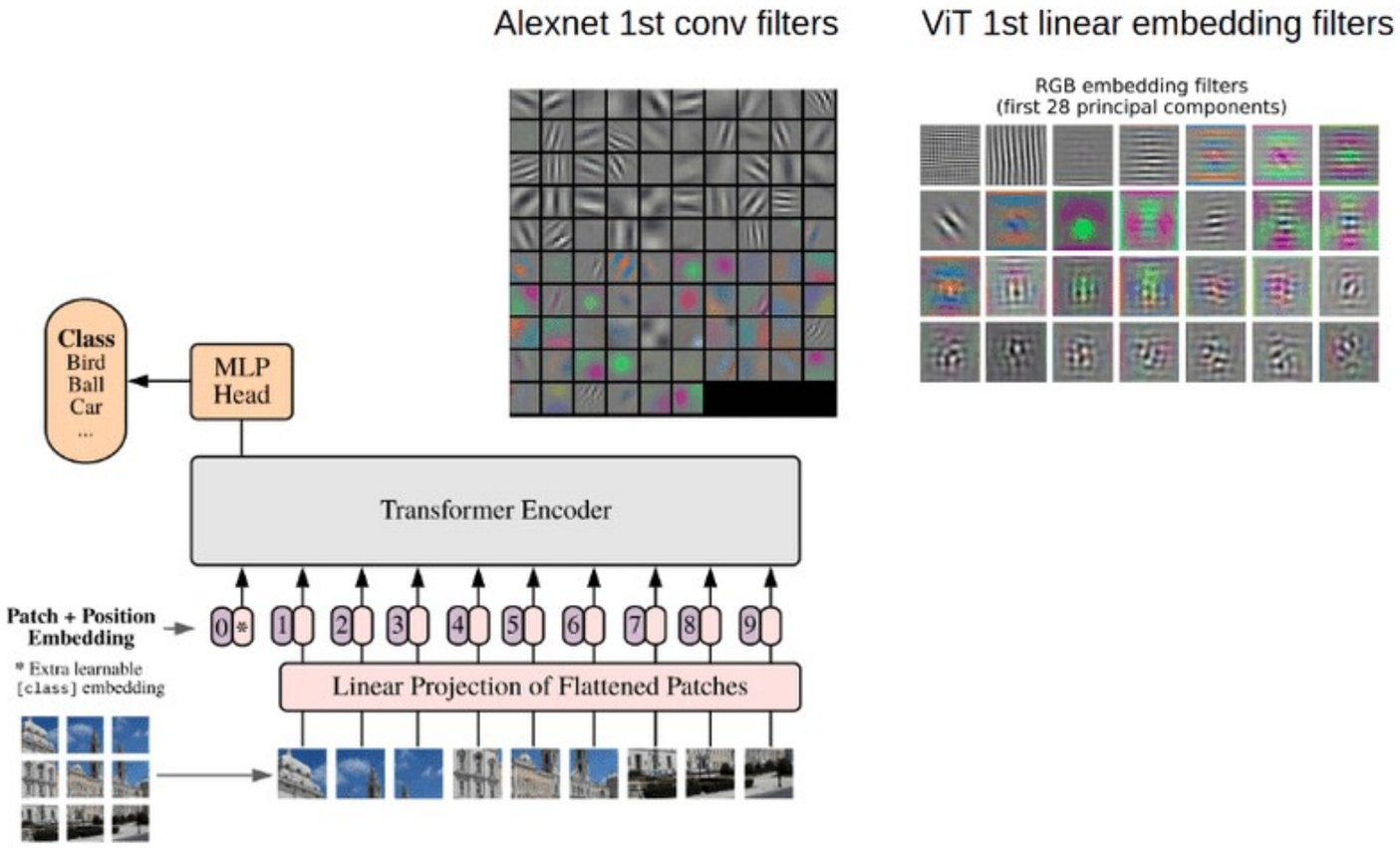


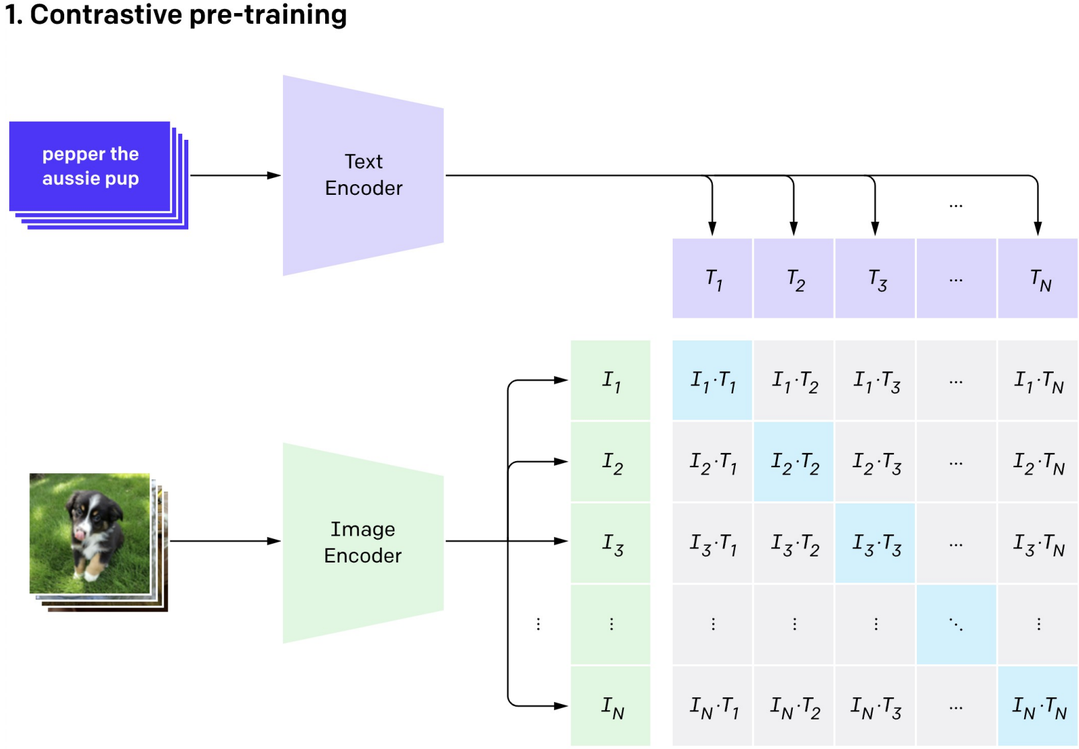


4. LLM Scaling Laws and LLM Families

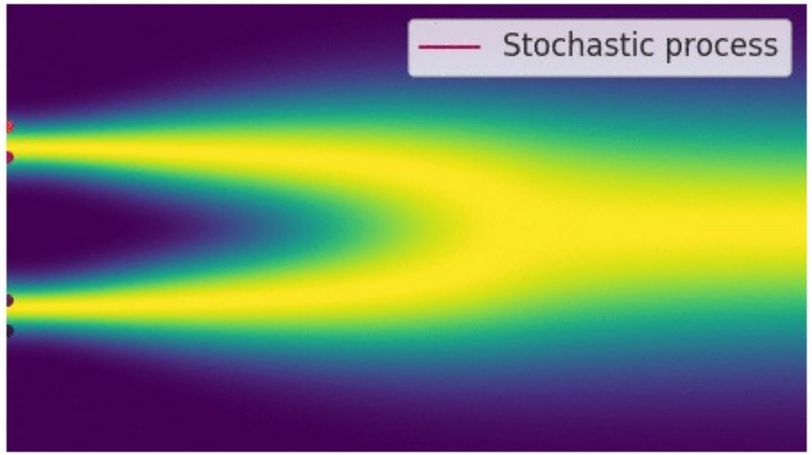


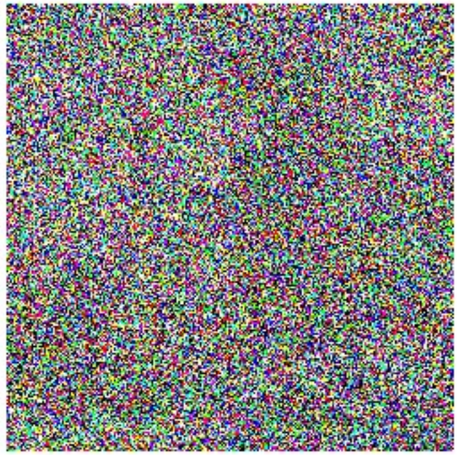
5. Multimodal Learning and its Applications

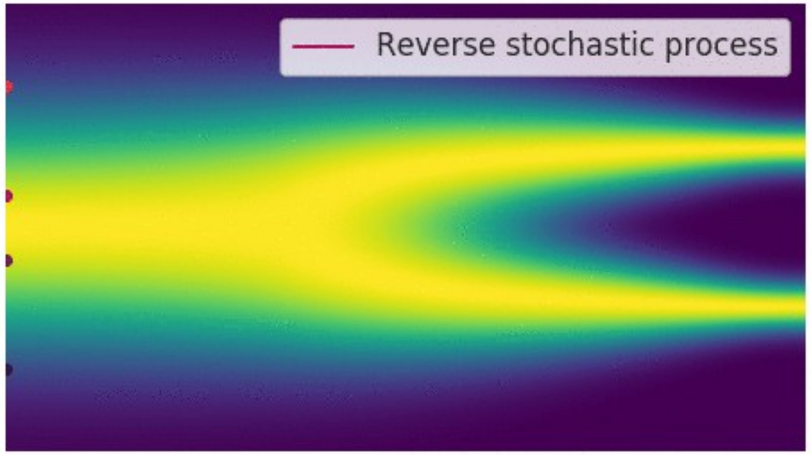




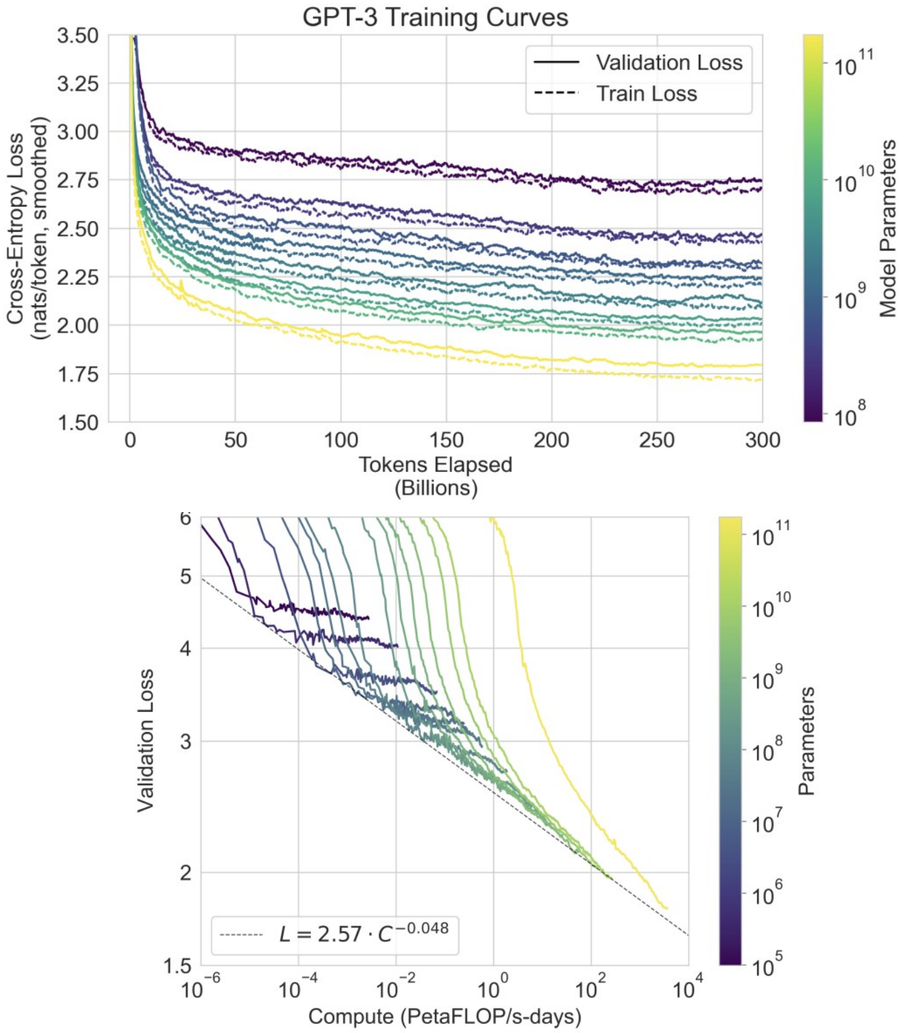
6. Diffusion Models in Generative AI

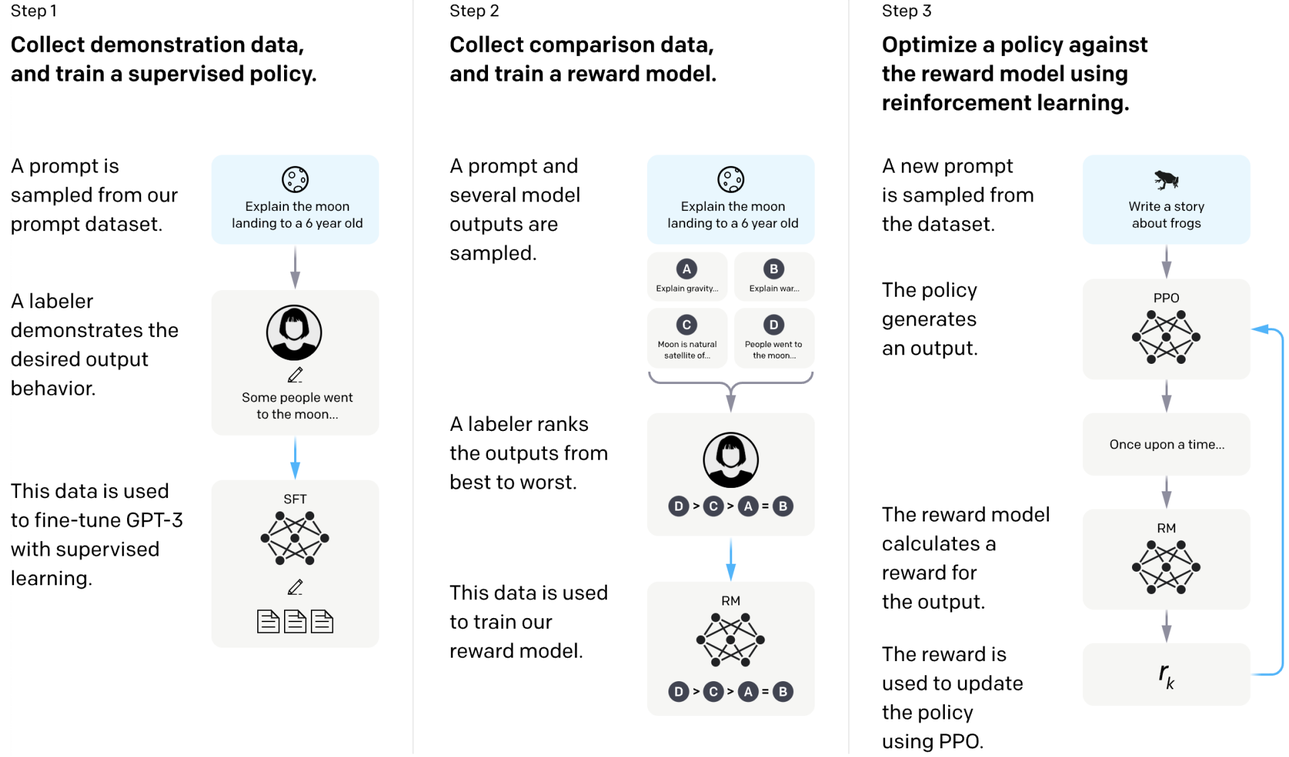






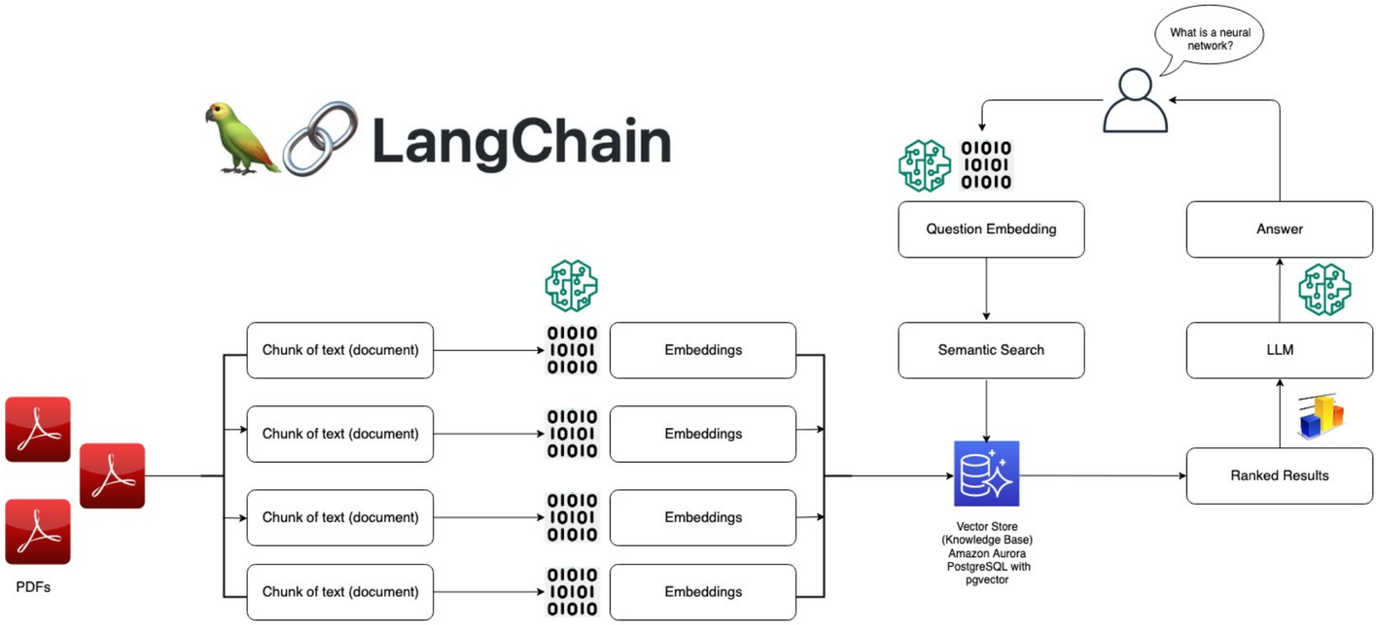
7. Model Training (Pre-Training, Instruction Following, and PEFT)

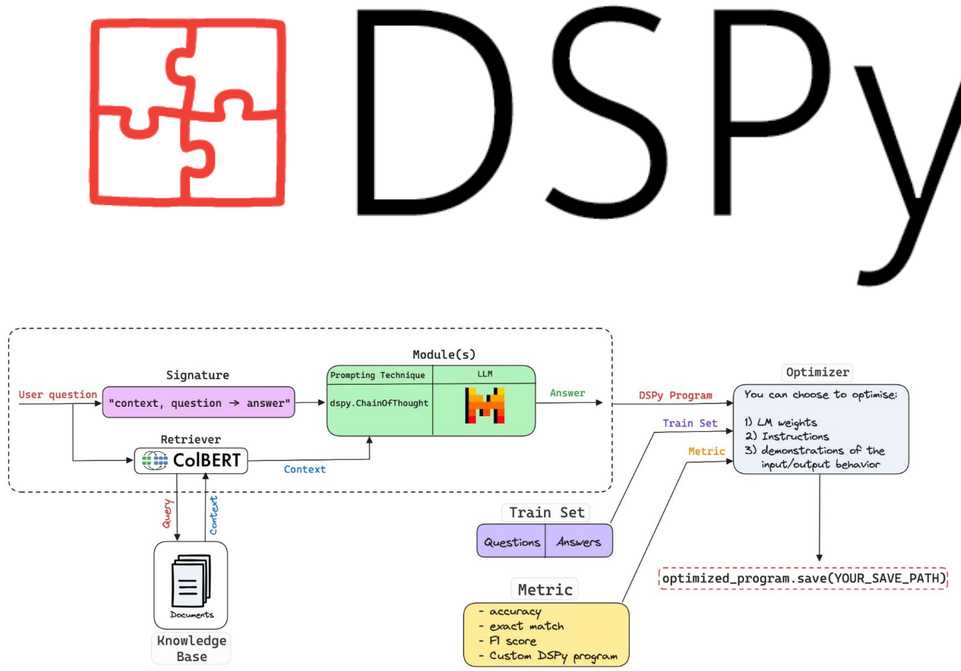




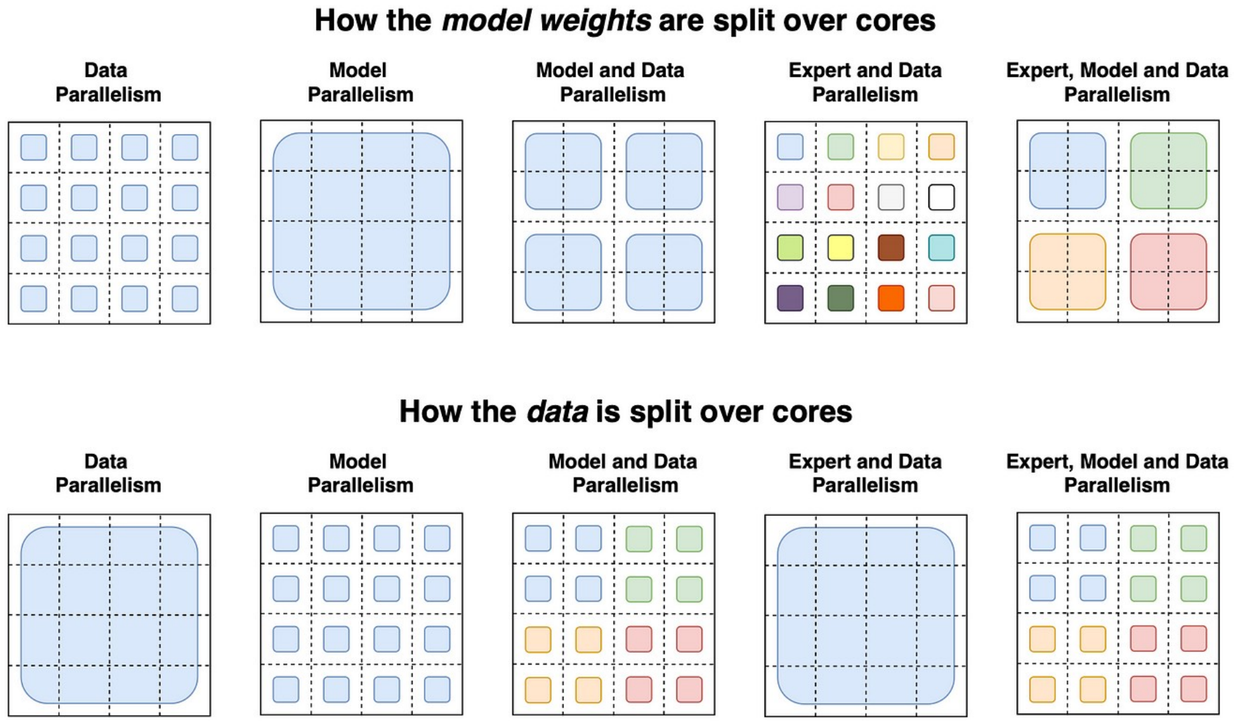
DARTMOUTH ENGINEERING | NVIDIA

8. LLM Orchestration





9. Scaling Model Training to Distributed Workloads



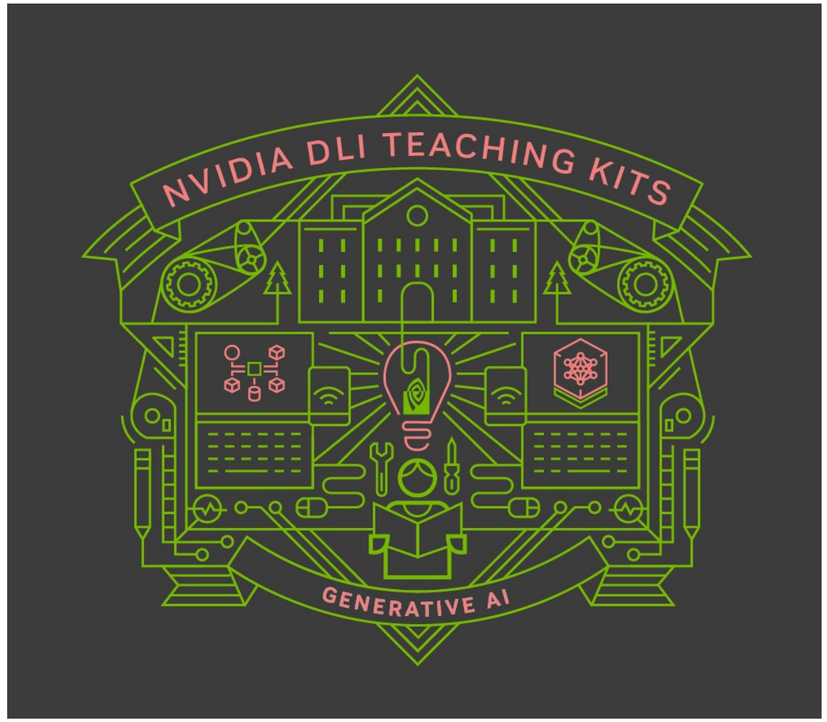
Wrap Up

Introduction to GenAI

- Today we introduced the GenAI Course
- We covered the different materials used in the class
- Provided an overview of the modules to come
- Highlighted how GenAI has exploded onto the scene of science, engineering, and the economy at large

---

In the next lesson we will start our discussion into how we got to this point and what advances led us from the digital computer, to ChatGPT



Thank you!