Generative AI Innovative Applications

NTPU
國立臺北大學
National Taipei University

NVIDIA Generative AI with Diffusion Models
Part III
AI Agents and Large Multimodal Agents (LMAs)

1132GAIIA09
MBA, IM, NTPU (M6031) (Spring 2025)
Tue 2, 3, 4 (9:10-12:00) (B3F17)

https://meet.google.com/paj-zhhj-mya

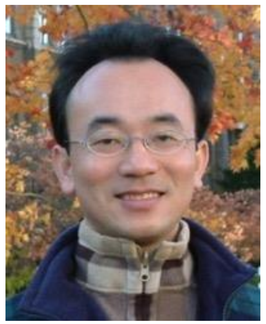
NVIDIA University Ambassador Certified Instructor

aws educate Cloud Ambassador 2020 Cohort

aws academy Accredited Educator

aws certified Cloud Practitioner

aws certified Solutions Architect Associate



Min-Yuh Day, Ph.D,
Professor

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

2025-05-20

Syllabus

Week	Date	Subject/Topics
1	2025/02/18	Introduction to Generative AI Innovative Applications
2	2025/02/25	Transformers for Natural Language Processing and Computer Vision
3	2025/03/04	Large Language Models (LLMs), NVIDIA Building RAG Agents with LLMs Part I
4	2025/03/11	Case Study on Generative AI Innovative Applications I
5	2025/03/18	NVIDIA Building RAG Agents with LLMs Part II
6	2025/03/25	NVIDIA Building RAG Agents with LLMs Part III

Syllabus

Week	Date	Subject/Topics
7	2025/04/01	Self-Learning
8	2025/04/08	Midterm Project Report
9	2025/04/15	Generative AI for Multimodal Information Generation
10	2025/04/22	NVIDIA Generative AI with Diffusion Models Part I
11	2025/04/29	NVIDIA Generative AI with Diffusion Models Part II
12	2025/05/06	Case Study on Generative AI Innovative Applications II

Syllabus

Week	Date	Subject/Topics
13	2025/05/13	NVIDIA DLI and NVIDIA-Certified Associate (NCA) Generative AI LLMs Certification [Invited Talk]
14	2025/05/20	NVIDIA Generative AI with Diffusion Models Part III AI Agents and Large Multimodal Agents (LMAs)
15	2025/05/27	Final Project Report I
16	2025/06/03	Final Project Report II

AI Agents and Large Multimodal Agents (LMAs)

NVIDIA
Generative AI with Diffusion Models
Part III

[0,1,2] [3, 4] [5, 6]

Generative AI with Diffusion Models

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance (CFG)
- Part 5: CLIP (Contrastive Language-Image Pre-Training)
- Part 6: Wrap-up & Assessment

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Diffusion Models

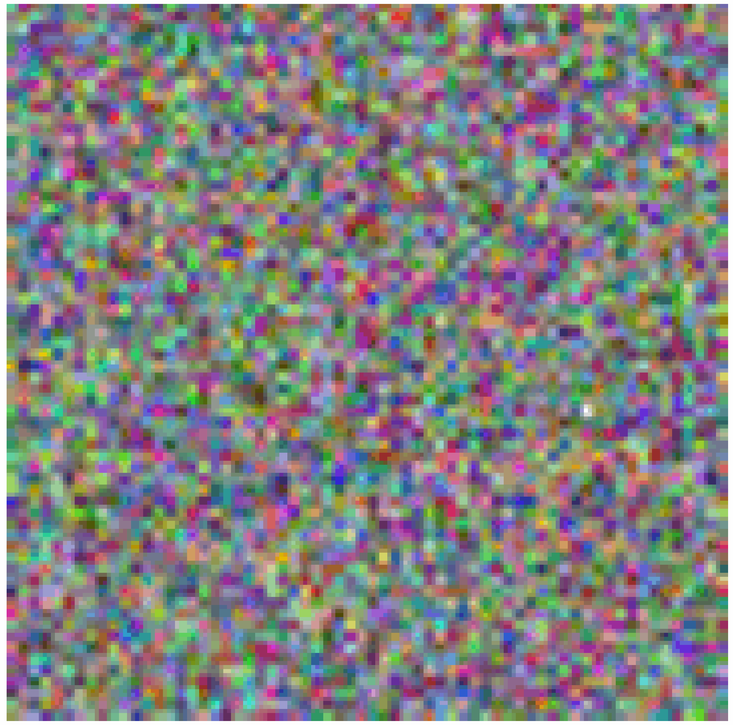
diffusion models are trained to denoise noisy images,
and can generate images by iteratively denoising pure noise.

Forward diffusion

image

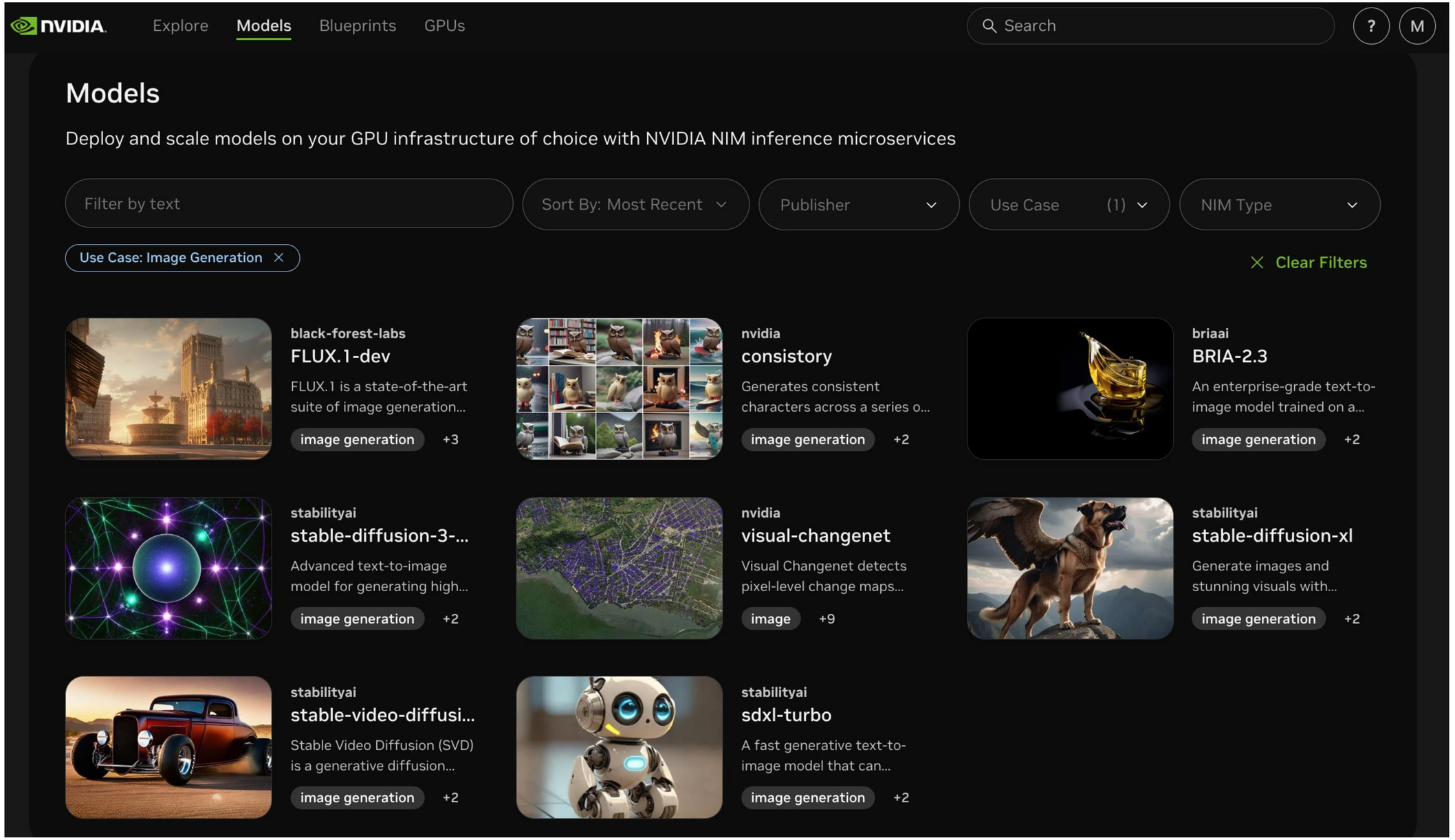
noisy image

noise



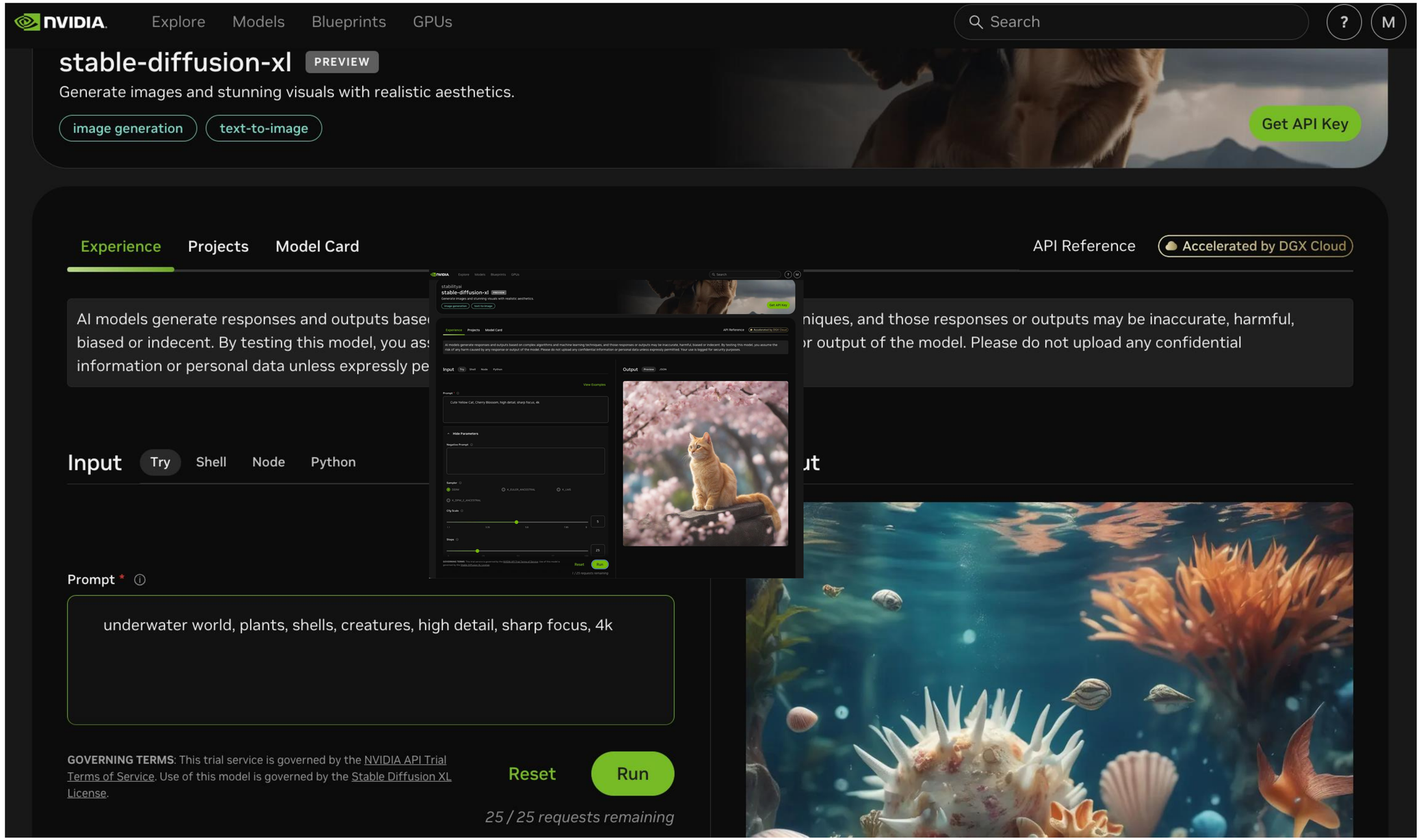
Source: https://keras.io/examples/generative/ddim/

build.nvidia.com **Models: Image Generation**



https://build.nvidia.com/models?filters=usecase%3Ausecase_image_gen

build.nvidia.com **stabilityai/stable-diffusion-xl**

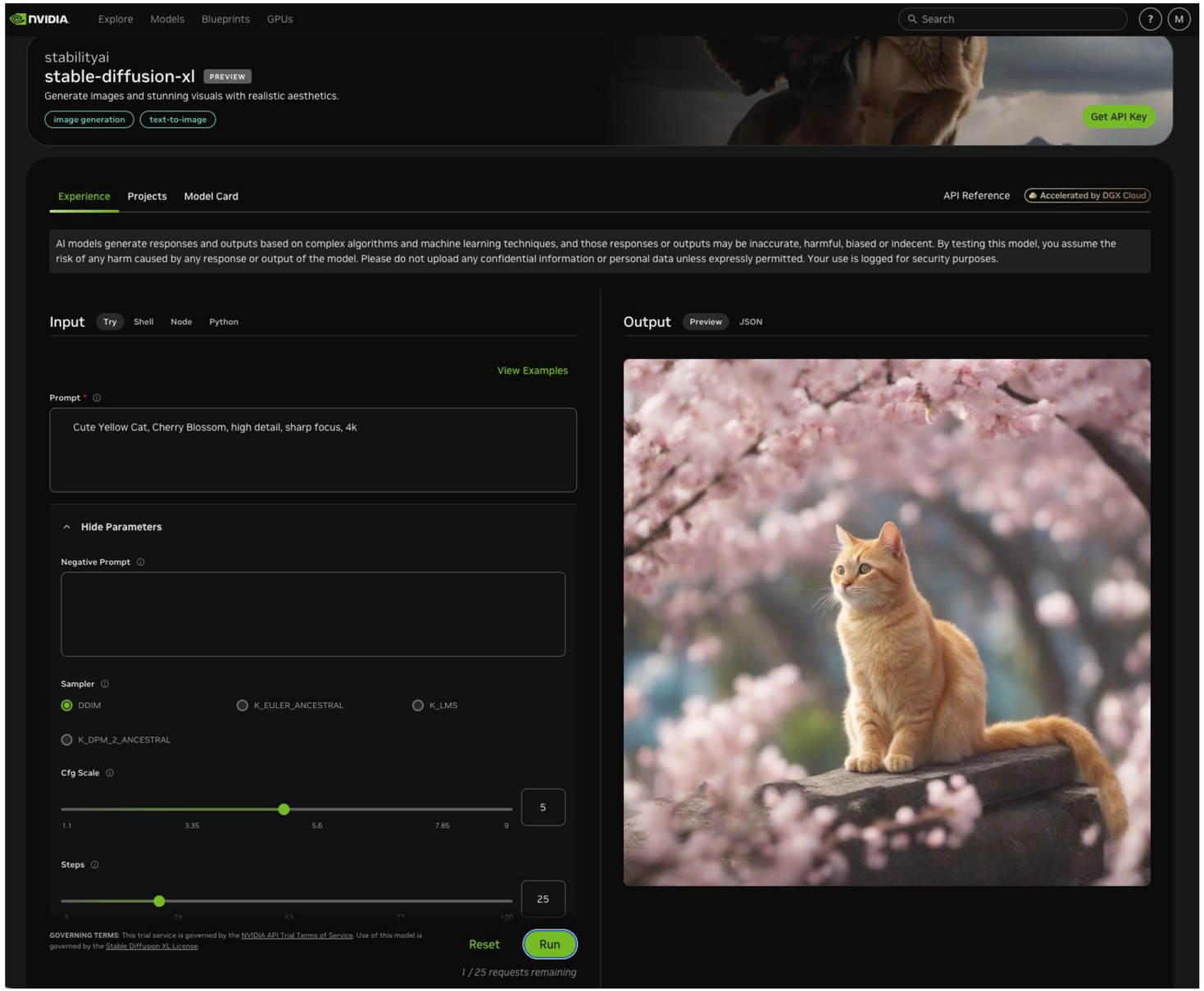


https://build.nvidia.com/stabilityai/stable-diffusion-xl

build.nvidia.com **stabilityai/stable-diffusion-xl**

Input Prompt

Cute Yellow Cat, Cherry Blossom, high detail, sharp focus, 4k



https://build.nvidia.com/stabilityai/stable-diffusion-xl

Spring 2025

Generative AI Innovative Applications

NVIDIA

University Ambassador

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver NVIDIA instructor-led workshop for academia

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Ambassador Certification ID: cCFh1ZWWTvqKTq7dcKkEWw

https://learn.nvidia.com/certificates?id=cCFh1ZWWTvqKTq7dcKkEWw

NVIDIA

Certified Instructor

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver the instructor-led workshop

Building RAG Agents with LLMs

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Certification ID: OVmqY4cSSya0BdMQBWHxzw

https://learn.nvidia.com/certificates?id=OVmqY4cSSya0BdMQBWHxzw

NVIDIA

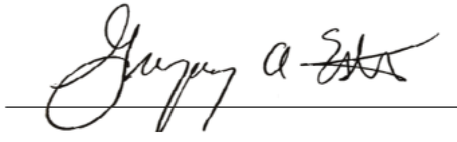
Certificate of Completion

This certificate is awarded to

Min-Yuh Day

for successfully completing

Building RAG Agents with LLMs



Greg Estes

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQaatzU8SNUNxgw |

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw/courses/course?course_id=course-v1:DLI+S-FX-15+V1

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

NVIDIA

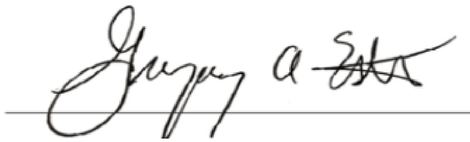
Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Generative AI with Diffusion Models



Greg Estes

Vice President, NVIDIA

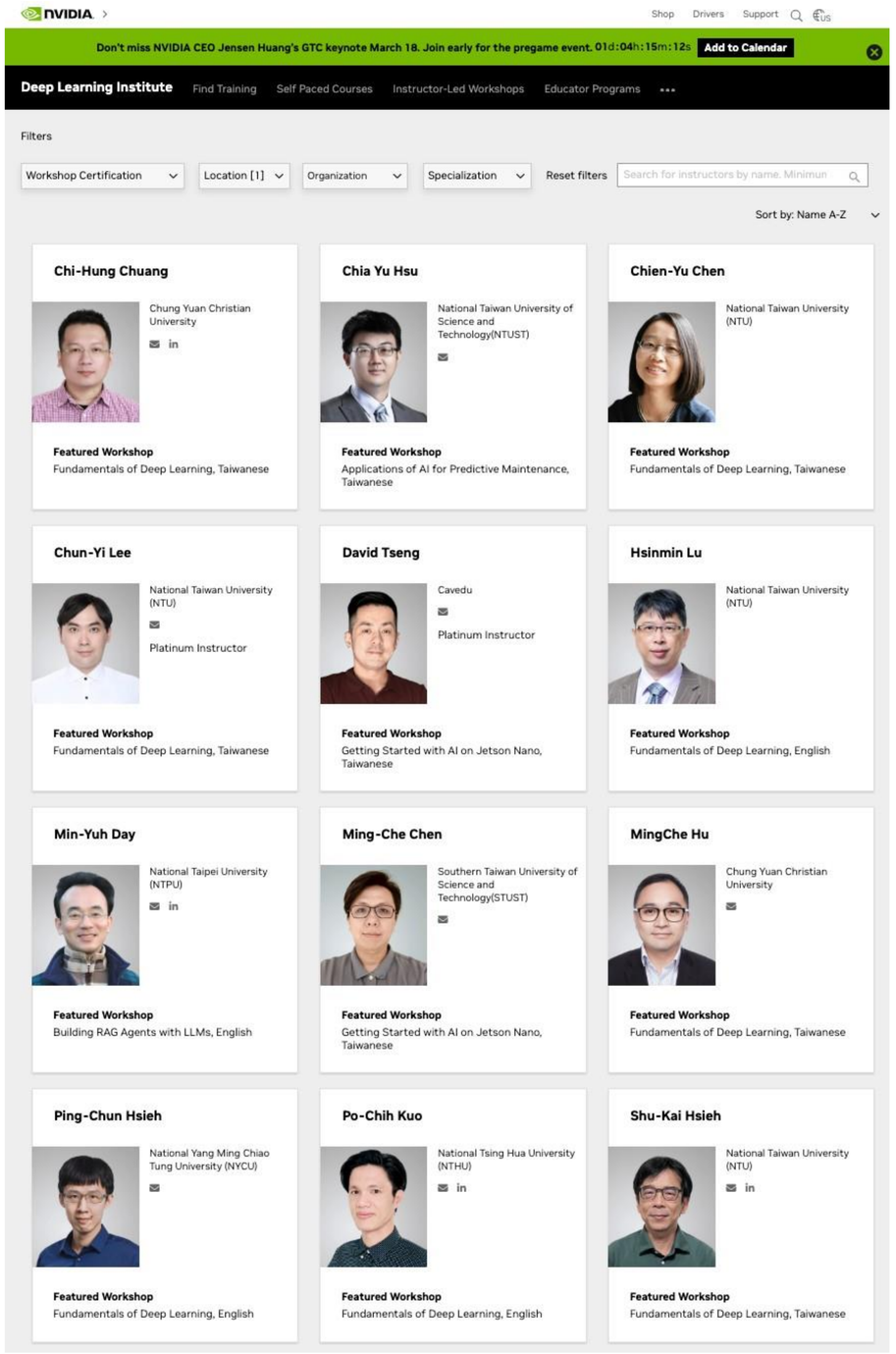
Issue Date: : February 28, 2025

Certification ID: q3oo-oBhTQKtyCCote2E-Q

https://learn.nvidia.com/certificates?id=q3oo-oBhTQKtyCCote2E-Q

NVIDIA Certified Instructors

(12 from Taiwan)

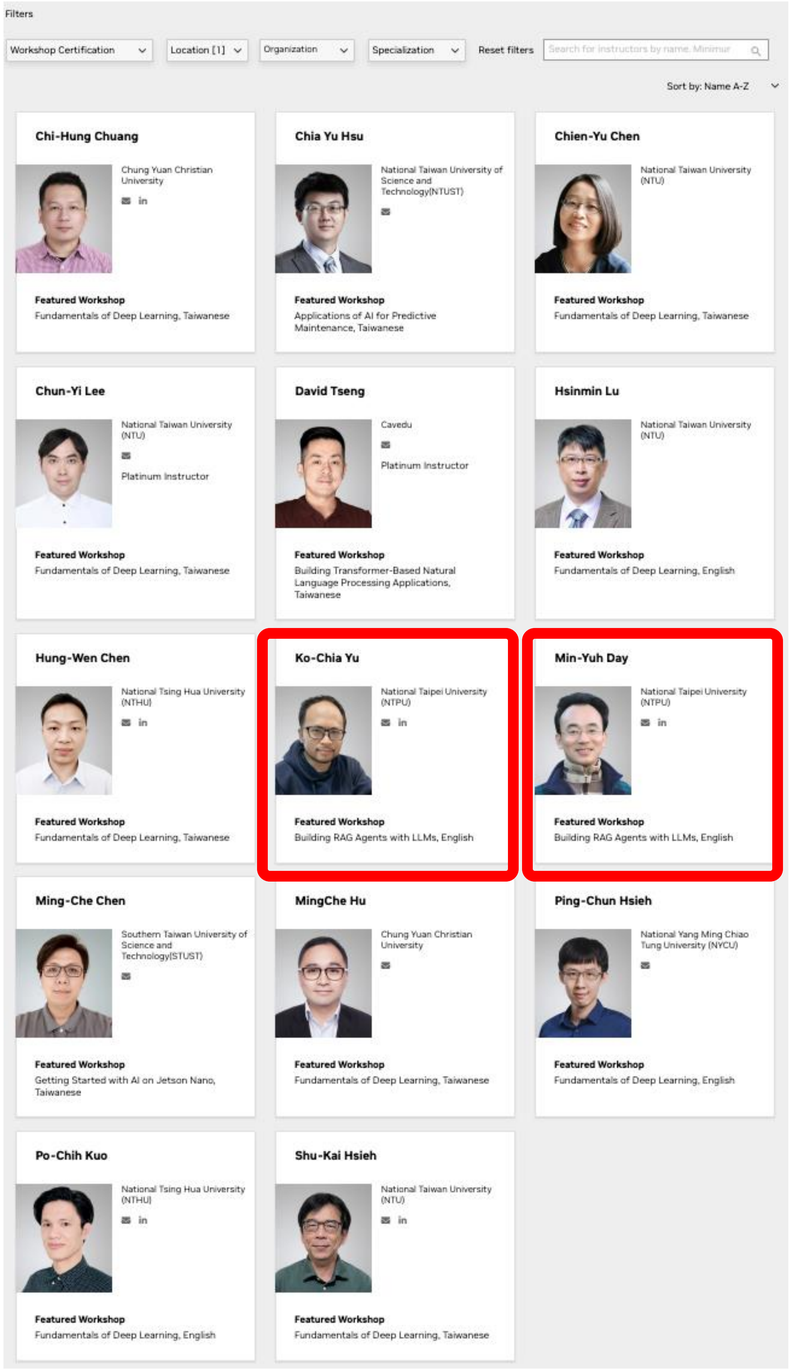


https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Certified Instructors

(14 from Taiwan)

(2 from NTPU)(April 2025)



https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Developer Program

https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA
Deep Learning Institute (DLI)

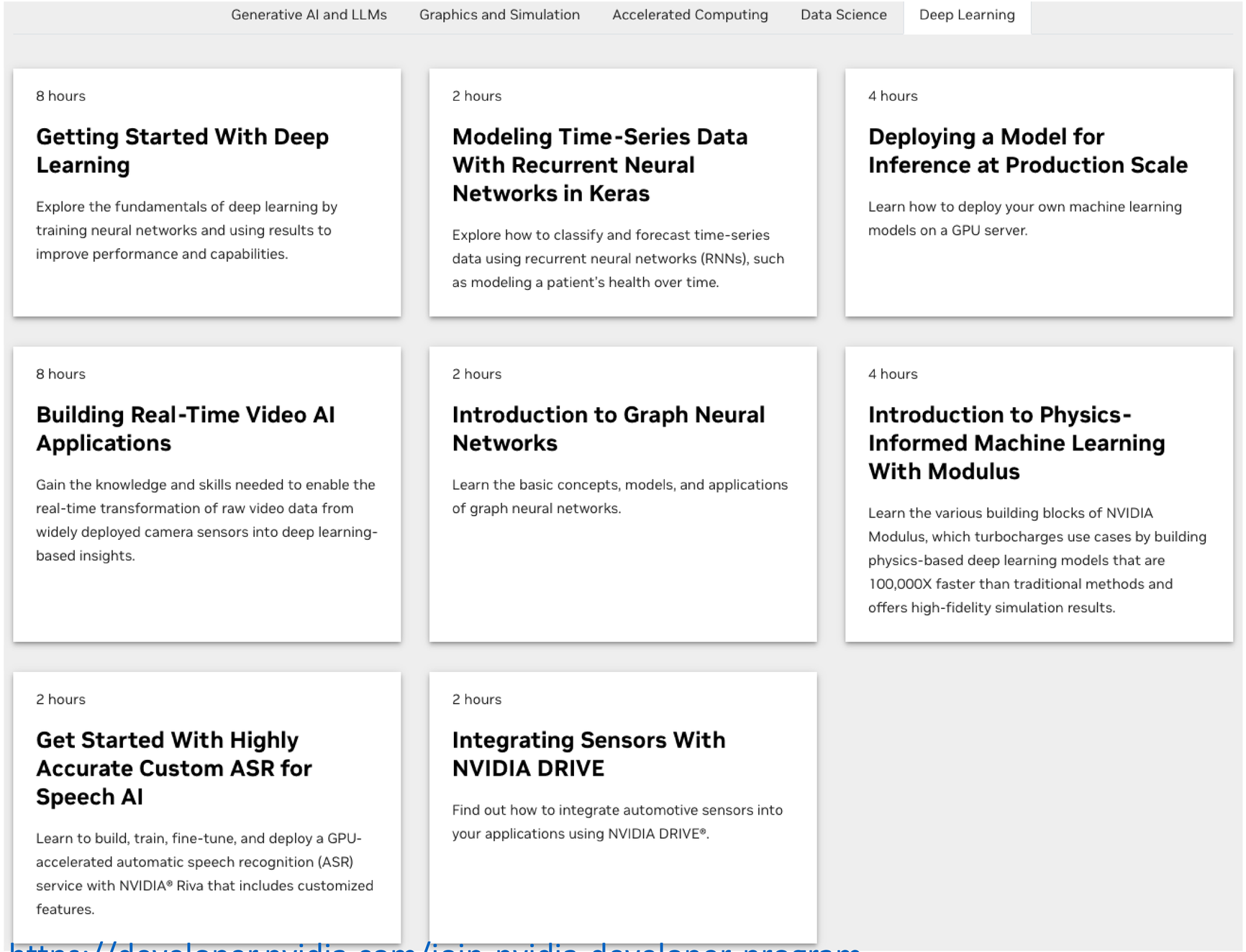
https://learn.nvidia.com/

Get NVIDIA DLI Certificate

- **Step 1. Join NVIDIA Developer Program (Free)**
  https://developer.nvidia.com/join-nvidia-developer-program
- **Step 2. Visit NVIDIA Deep Learning Institute (DLI)**
  https://learn.nvidia.com/
- **Step 3. Enroll "Generative AI with Diffusion Models"**
  Self-Paced Course ($90)
  https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)



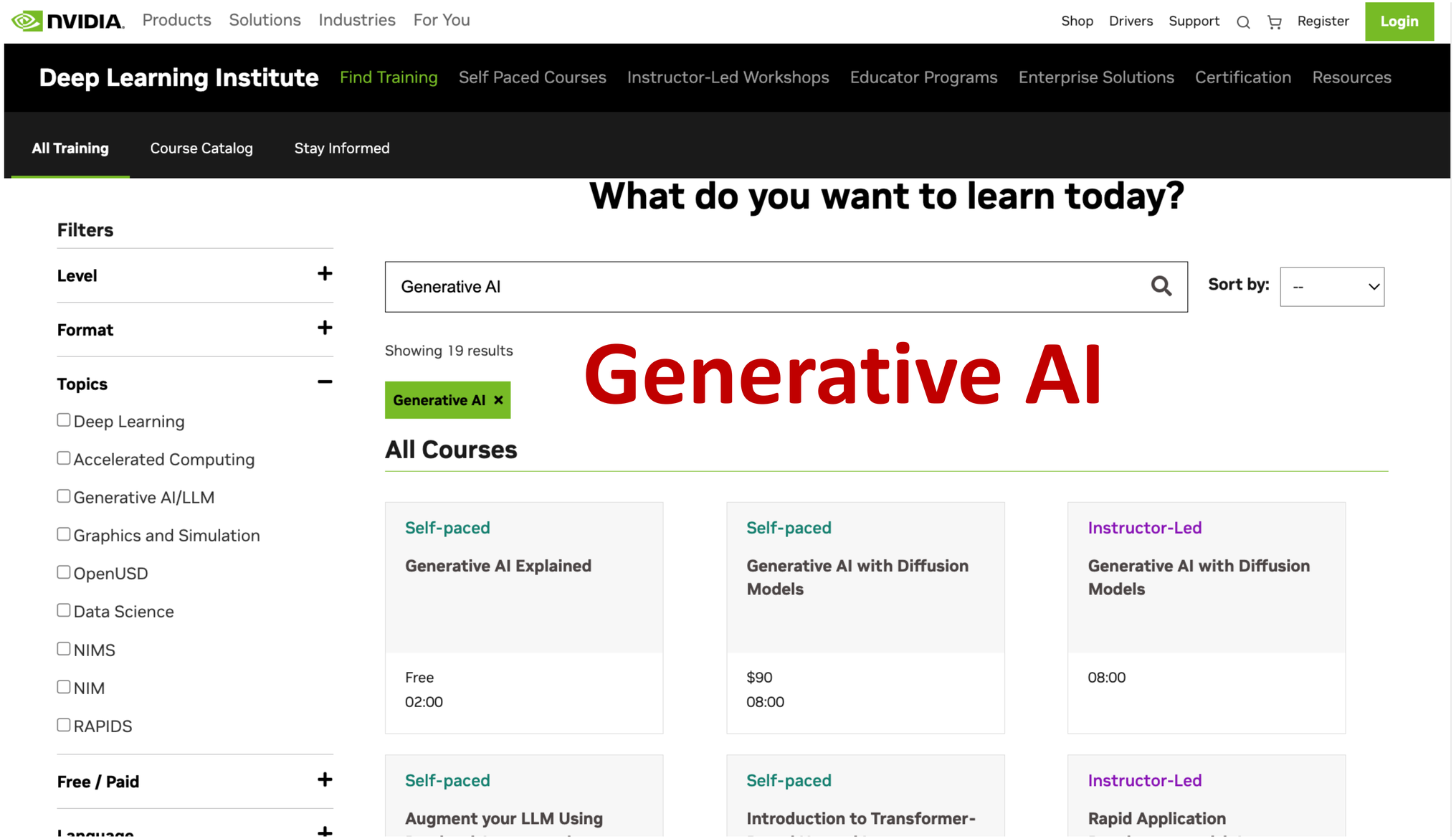
https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA Deep Learning Institute (DLI)

Self-Paced Course **Generative AI Explained** Free 2 hours	Self-Paced Course **Getting Started With Deep Learning** Certificate available $90 8 hours	Instructor-Led Workshop **Fundamentals of Deep Learning** Certificate available $500 8 hours	Self-Paced Course **Introduction to Transformer-Based Natural Language Processing** Certificate available $30 6 hours
Self-Paced Course **Building RAG Agents With LLMs** Certificate available Free 8 hours	Instructor-Led Workshop **Building RAG Agents With LLMs** Certificate available $500 8 hours	Self-Paced Course **Generative AI with Diffusion Models** Certificate available $90 8 hours	Instructor-Led Workshop **Generative AI with Diffusion Models** Certificate available $500 8 hours

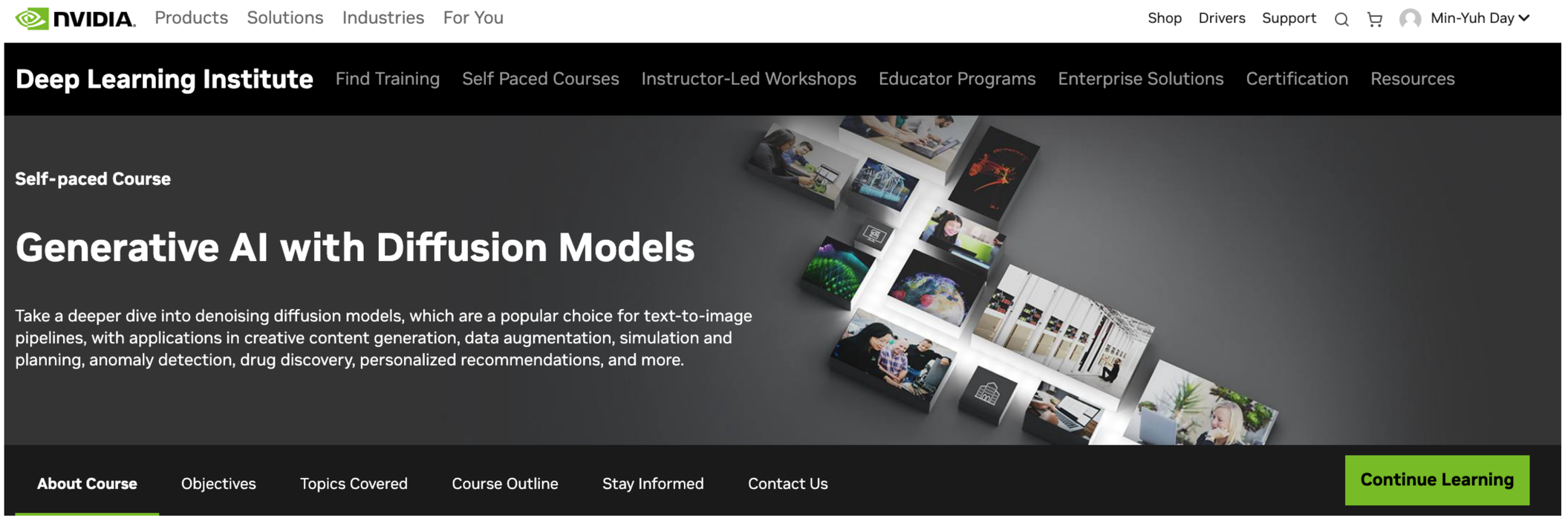
https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

NVIDIA Deep Learning Institute (DLI)



https://learn.nvidia.com/en-us/training/find-training?q=Generative+AI

Generative AI with Diffusion Models



About this Course

Thanks to improvements in computing power and scientific theory, generative AI is more accessible than ever before. Generative AI plays a significant role across industries due to its numerous applications, such as creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more. In this course, learners will take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines.

Course Details

Duration: 08:00

Price: $90

Subject: Generative AI/LLM

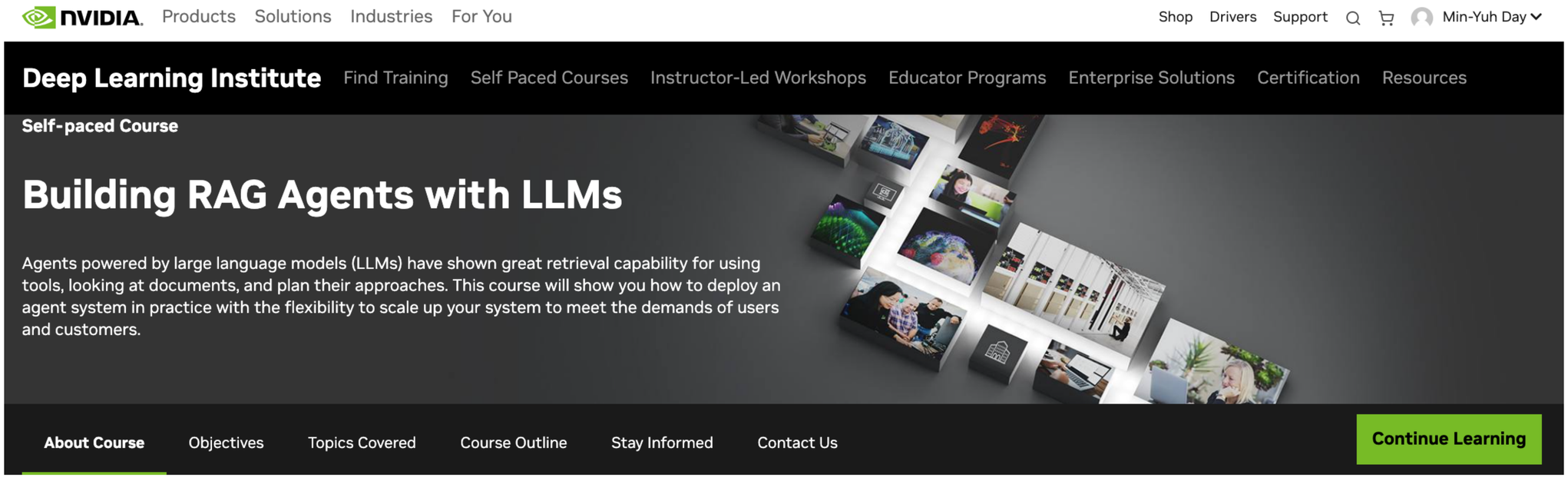
Language: English

Course Prerequisites:
A basic understanding of Deep Learning Concepts.

Learning Objectives

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Building RAG Agents with LLMs



About this Course

This course is free for a limited time.

The evolution and adoption of large language models (LLMs) have been nothing short of revolutionary, with retrieval-based systems at the forefront of this technological leap. These models are not just tools for automation; they are partners in enhancing productivity, capable of holding informed conversations by interacting with a vast array of tools and documents. This course is designed for those eager to explore the potential of these systems, focusing on practical deployment and the efficient implementation required to manage the considerable demands of both users and deep learning models. As we delve into the intricacies of LLMs, participants will gain insights into advanced orchestration techniques that include internal reasoning, dialog management, and effective tooling strategies.

Course Details

Duration: 08:00

Price: Free

Level: Technical - Intermediate

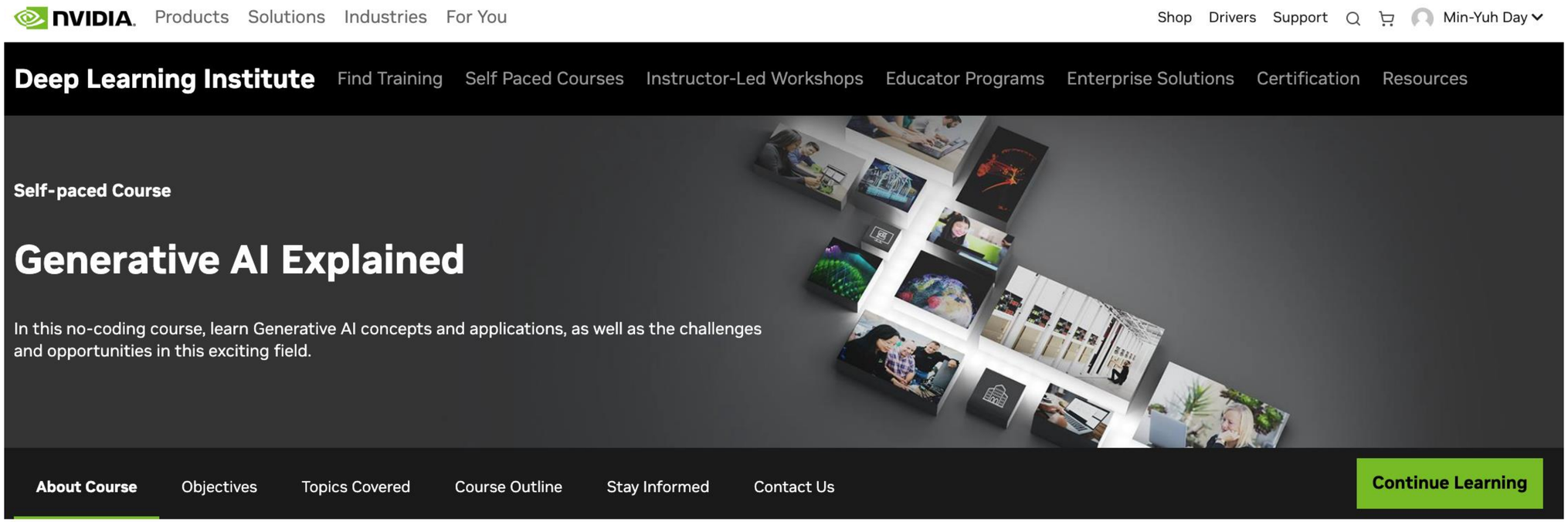
Subject: Generative AI/LLM

Language: English

Course Prerequisites:
Introductory deep learning knowledge, with comfort

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Generative AI Explained



About this Course

Generative AI describes technologies that are used to generate new content based on a variety of inputs. In recent time, Generative AI involves the use of neural networks to identify patterns and structures within existing data to generate new content. In this course, you will learn Generative AI concepts, applications, as well as the challenges and opportunities in this exciting field.

Learning Objectives

Upon completion, you will have a basic understanding of Generative AI and be able to more effectively use the various tools built on this

Course Details

Duration: 02:00

Price: Free

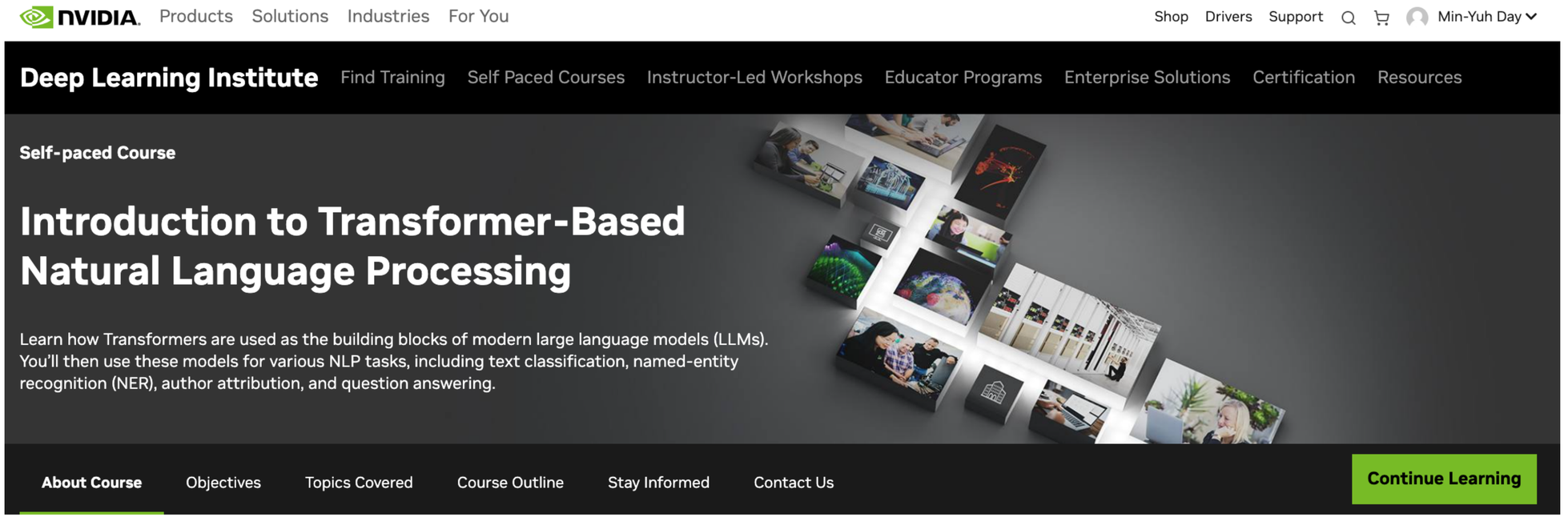
Level: Technical - Beginner

Subject: Generative AI/LLM

Language: English

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Introduction to Transformer-Based Natural Language Processing



About this Course

Large Language Models (LLMs), or Transformers, have revolutionized the field of natural language processing (NLP). Driven by recent advancements, applications of NLP and generative AI have exploded in the past decade. With the proliferation of applications like chatbots and intelligent virtual assistants, organizations are infusing their businesses with more interactive human-machine experiences. Understanding how Transformer-based large language models (LLMs) can be used to manipulate, analyze, and generate text-based data is essential. Modern pre-trained LLMs can encapsulate the nuance, context, and sophistication of language, just as humans do. When fine-tuned and deployed correctly, developers can use these LLMs to build powerful NLP applications that provide natural and seamless human-computer interactions within chatbots, AI voice agents, and more. In this course, you'll learn how Transformers are used as the building blocks of modern large language models (LLMs). You'll then use these models for various NLP

Course Details

Duration: 06:00

Price: $30

Level: Technical - Beginner

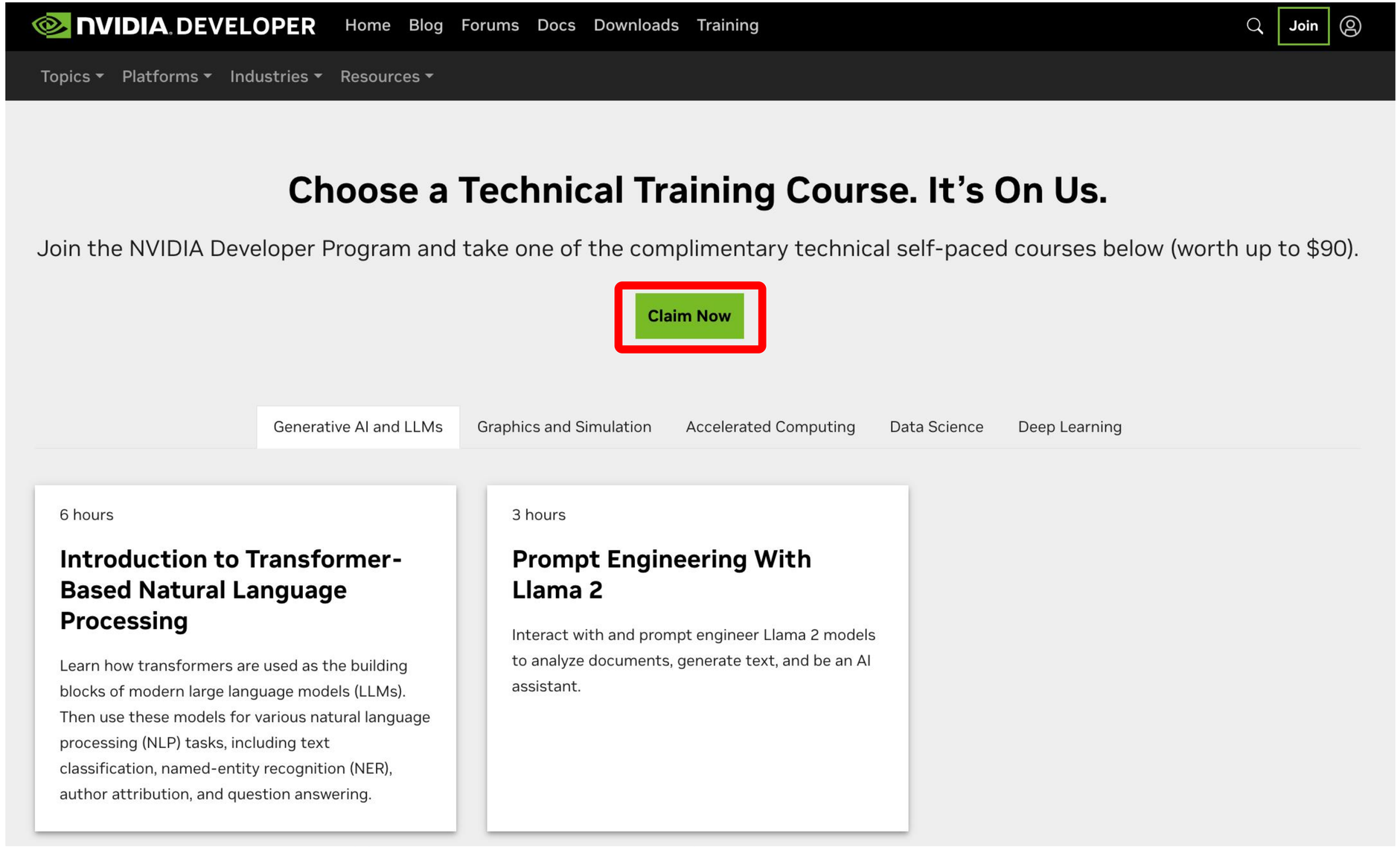
Subject: Generative AI/LLM

Language: English

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-08+V1

Join the NVIDIA Developer Program

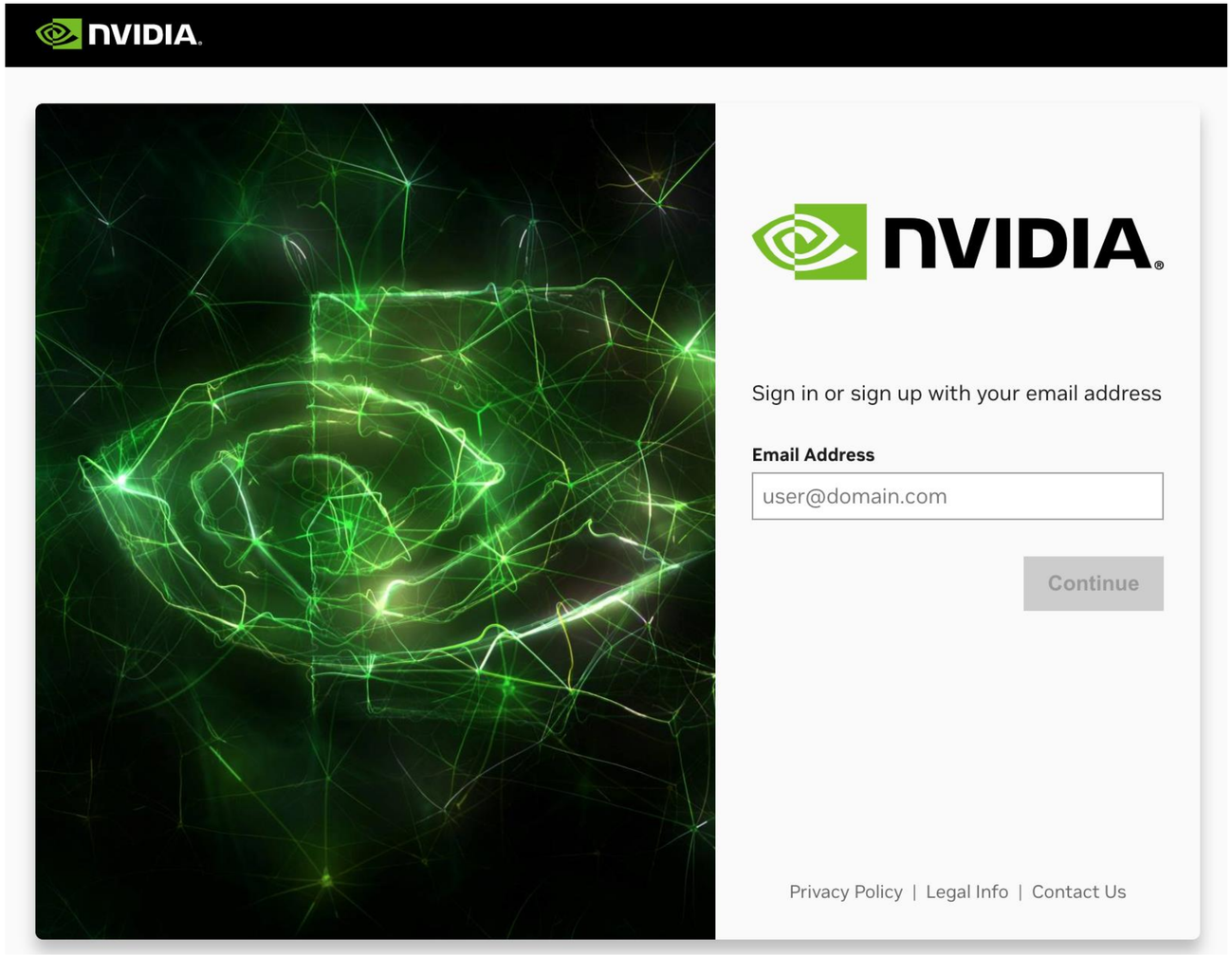
take one of the complimentary technical self-paced courses (worth up to $90)



https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

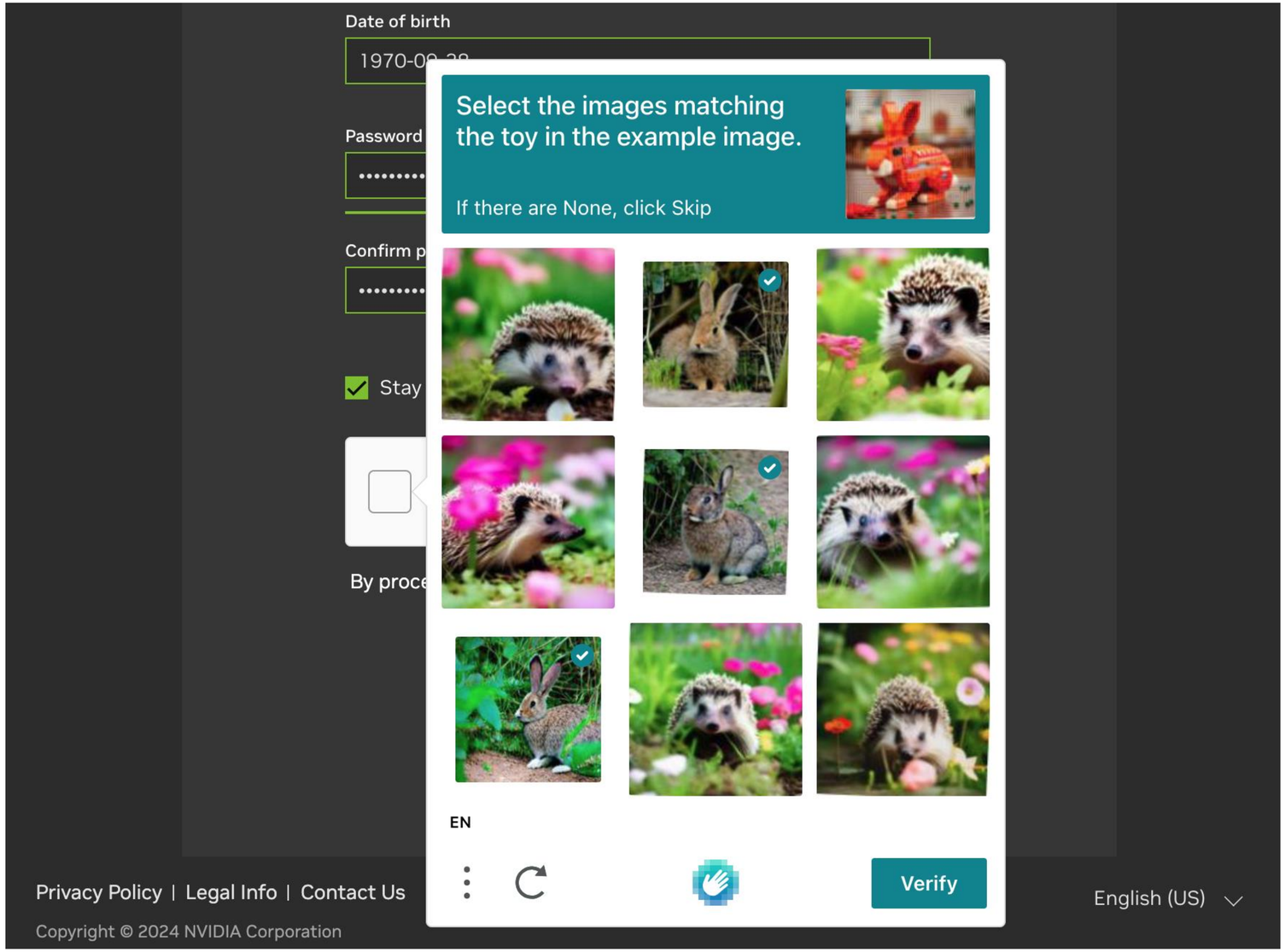
take one of the complimentary technical self-paced courses (worth up to $90)



https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

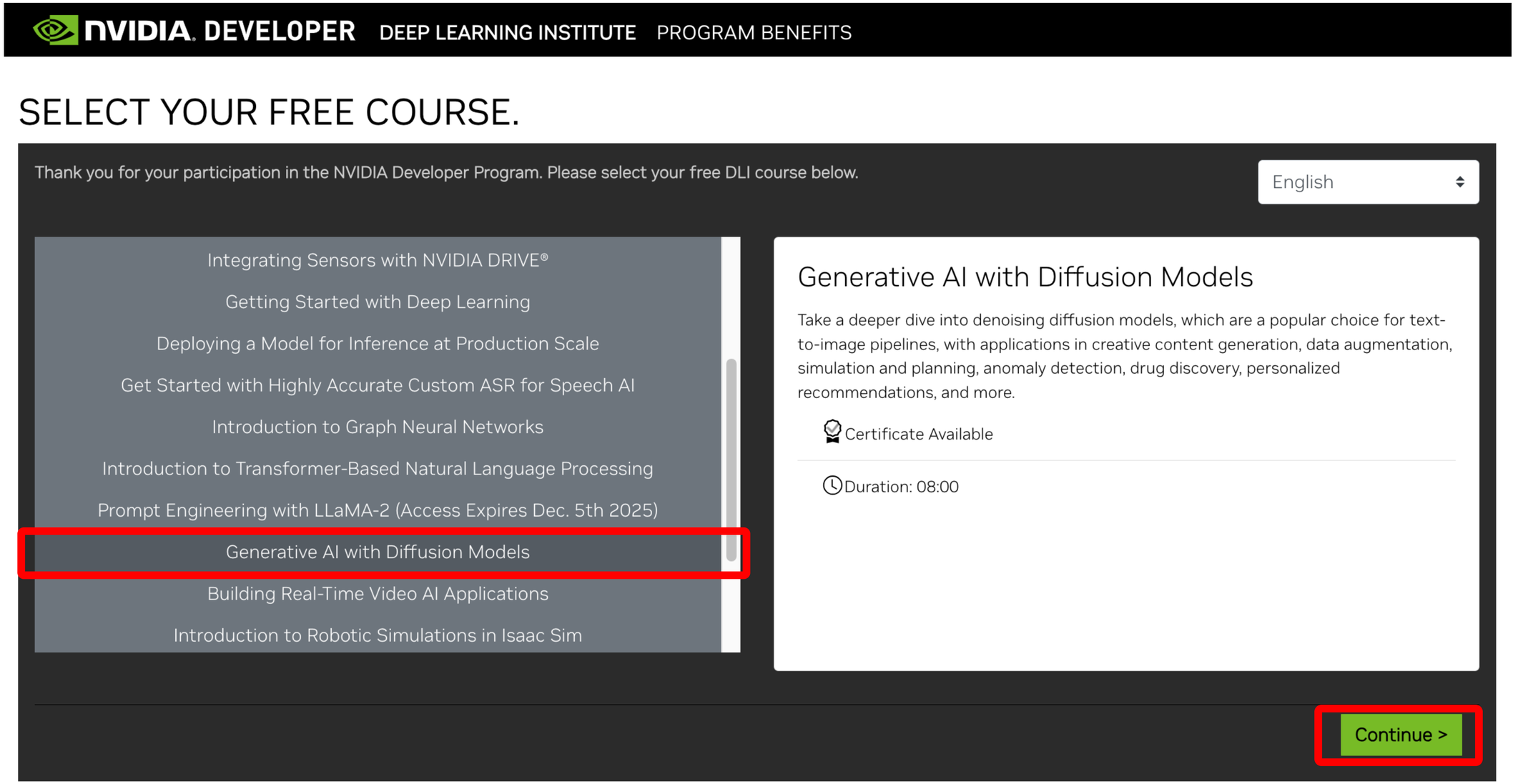
take one of the complimentary technical self-paced courses (worth up to $90)



https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

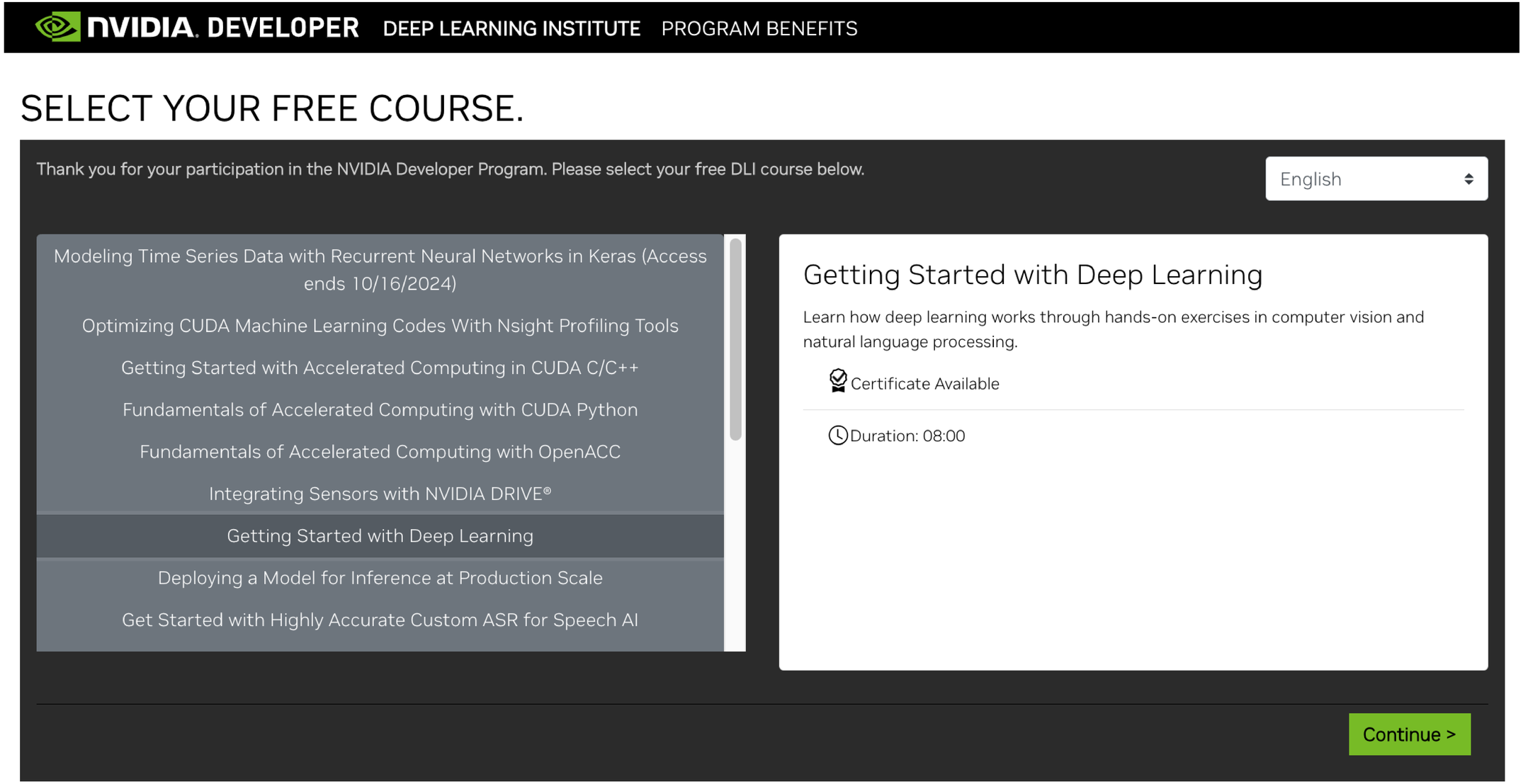
take one of the complimentary technical self-paced courses (worth up to $90)



https://developer.nvidia.com/join-nvidia-developer-program

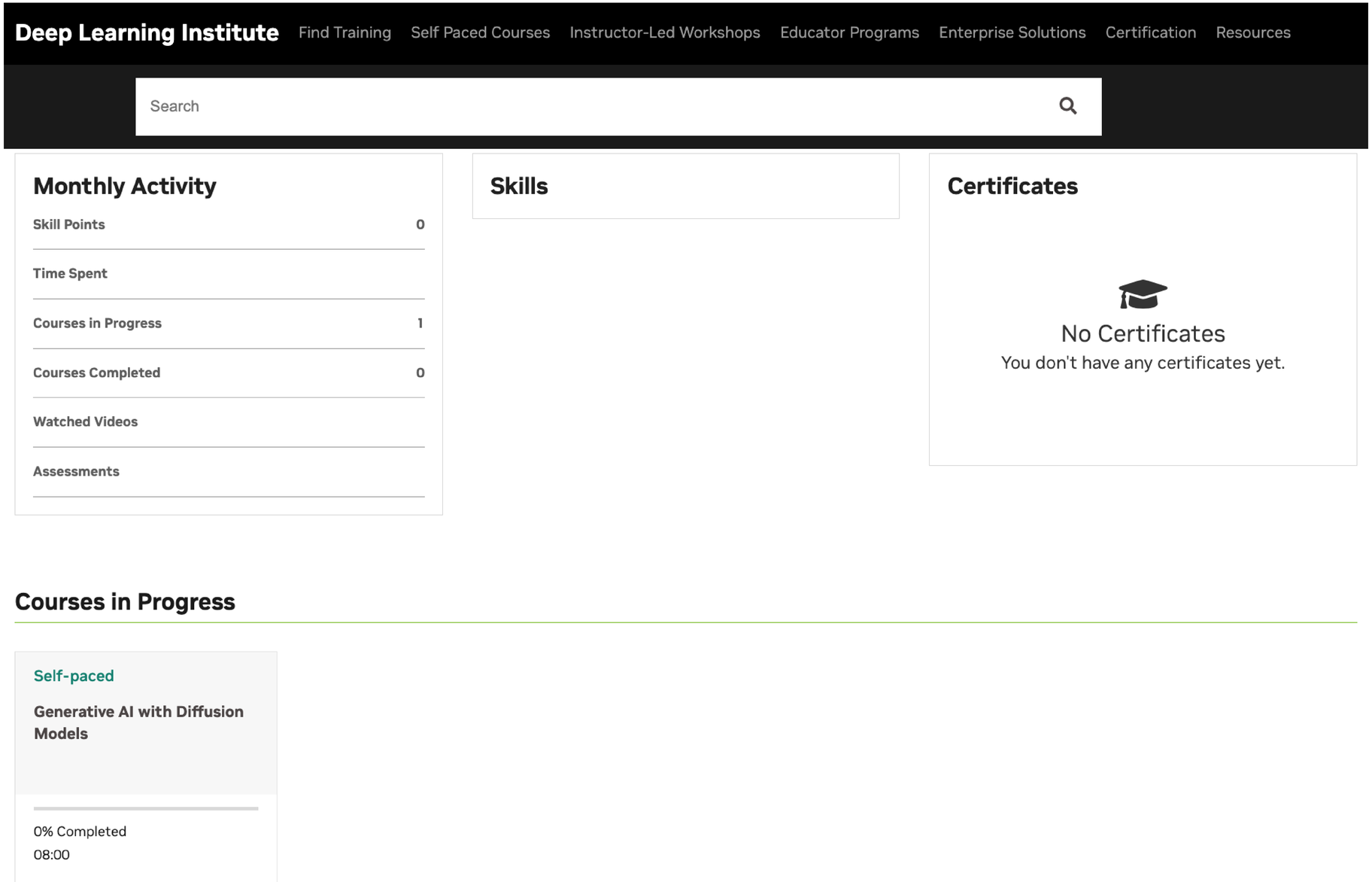
Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)



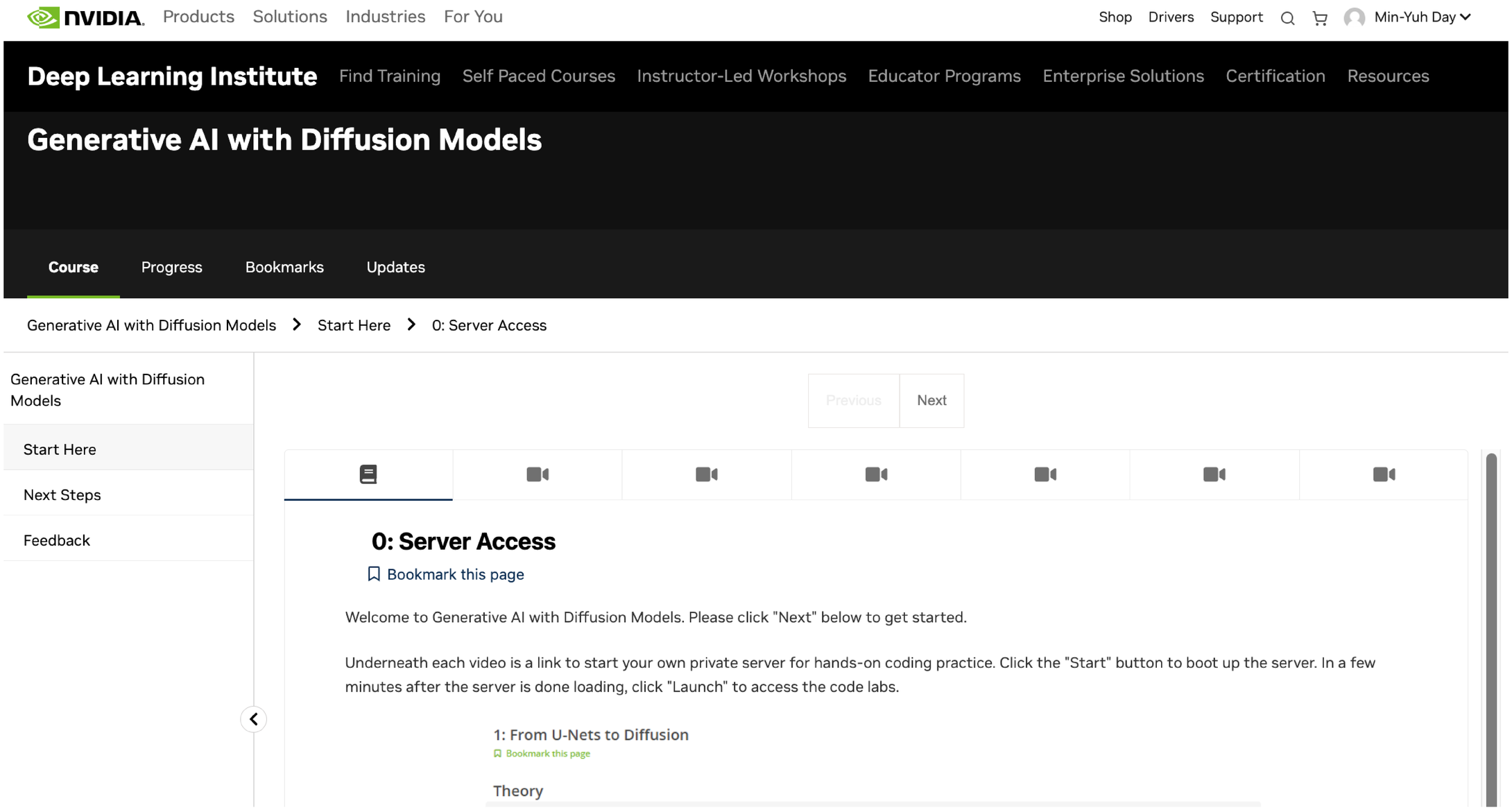
https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA Deep Learning Institute (DLI)



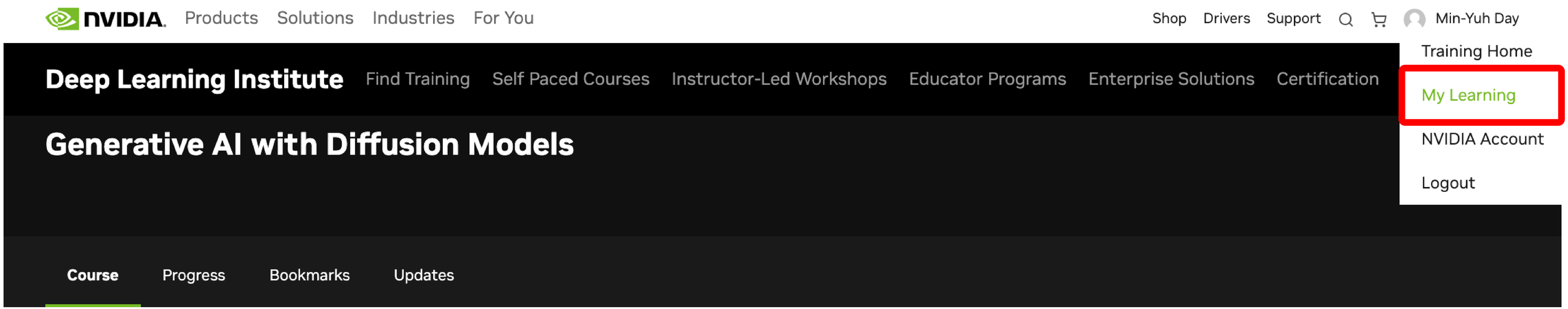
https://learn.nvidia.com/my-learning

NVIDIA Deep Learning Institute (DLI)



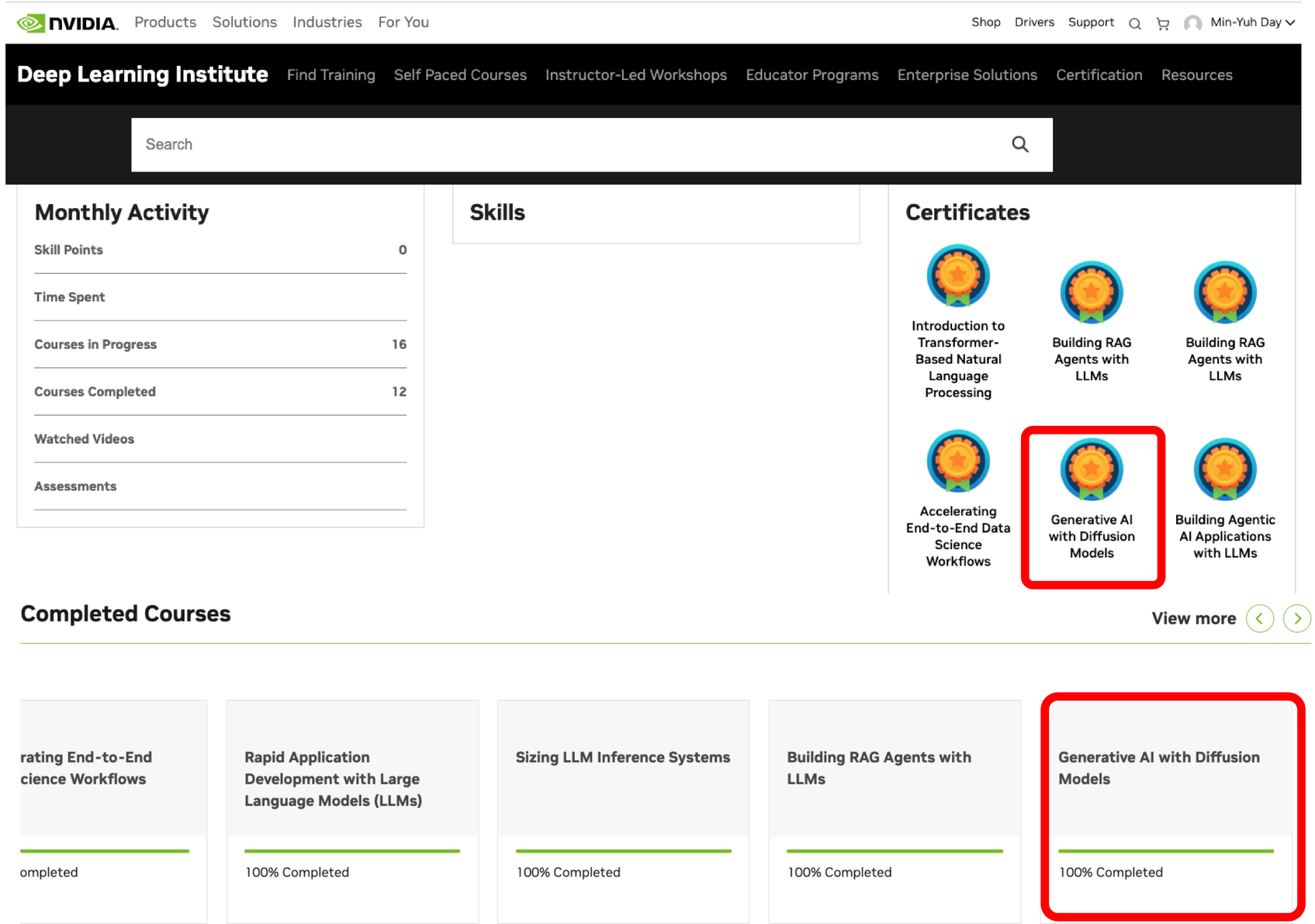
https://learn.nvidia.com/my-learning

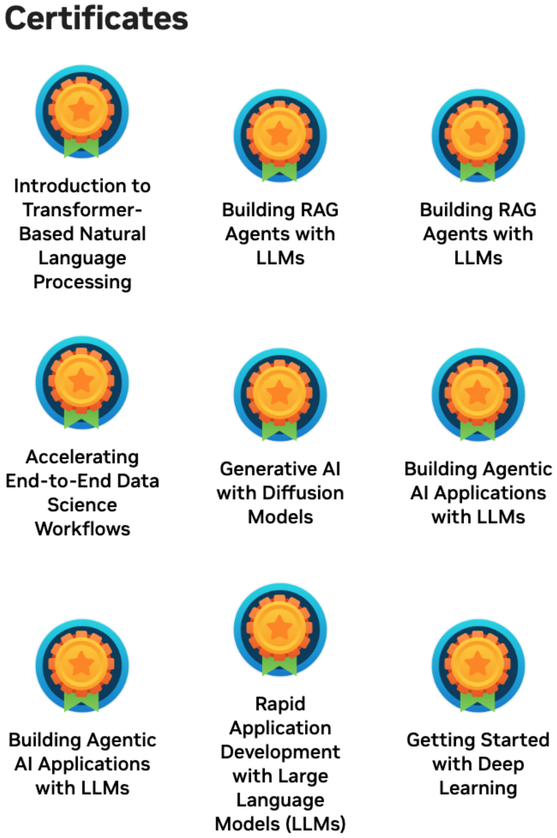
NVIDIA Deep Learning Institute (DLI)



https://learn.nvidia.com/my-learning

Deep Learning Institute (DLI)





https://learn.nvidia.com/my-learning

NVIDIA

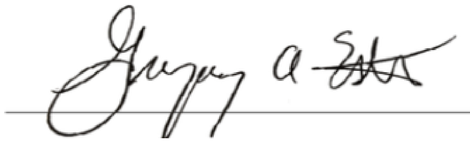
Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Generative AI with Diffusion Models



Greg Estes

Vice President, NVIDIA

Issue Date: : February 28, 2025

Certification ID: q3oo-oBhTQKtyCCote2E-Q

https://learn.nvidia.com/certificates?id=q3oo-oBhTQKtyCCote2E-Q

NVIDIA

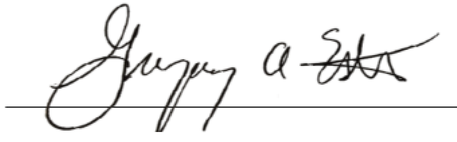
Certificate of Completion

This certificate is awarded to

Min-Yuh Day

for successfully completing

Building RAG Agents with LLMs



Greg Estes

Vice President, NVIDIA

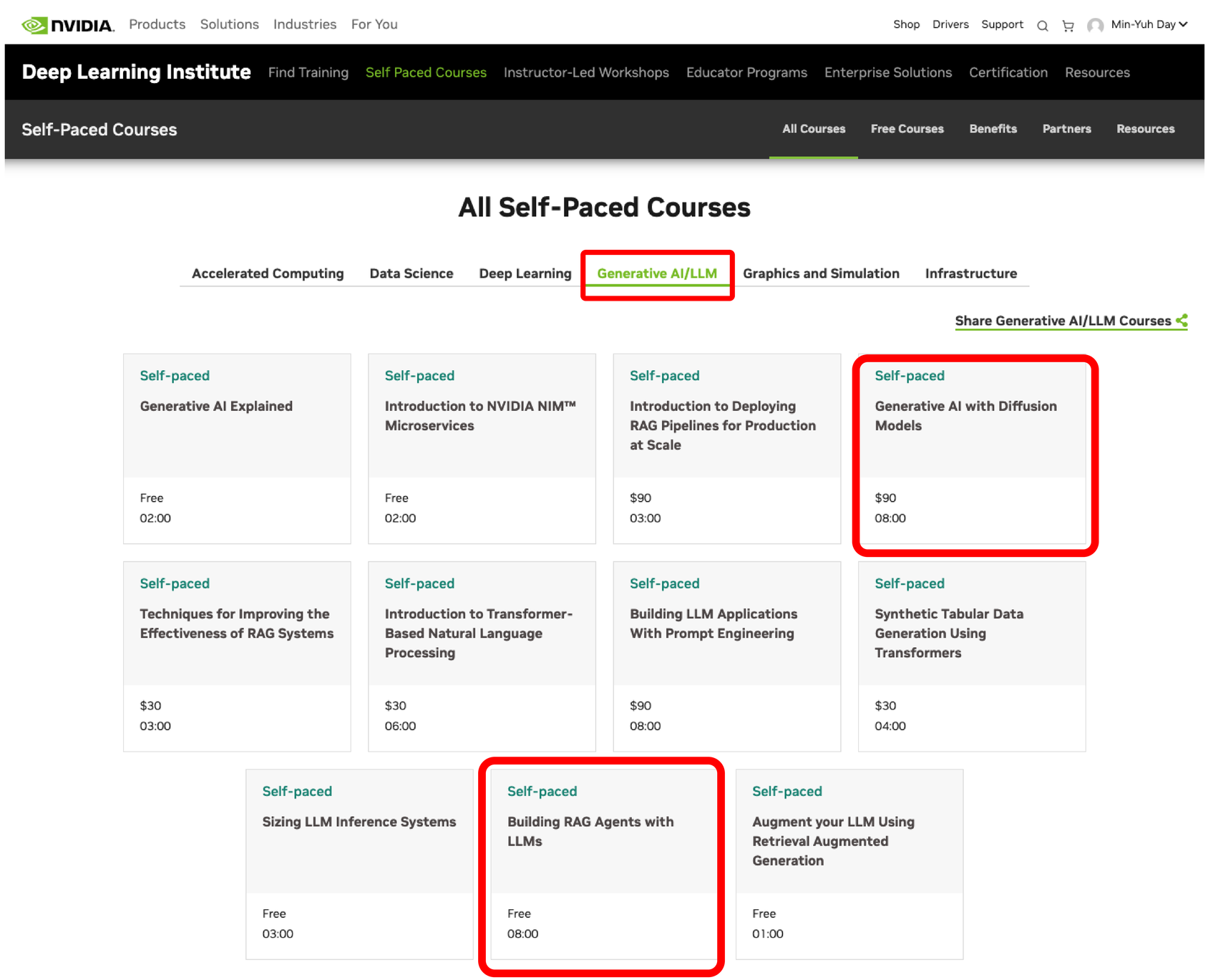
Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQaatzU8SNUNxgw |

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw/courses/course?course_id=course-v1:DLI+S-FX-15+V1

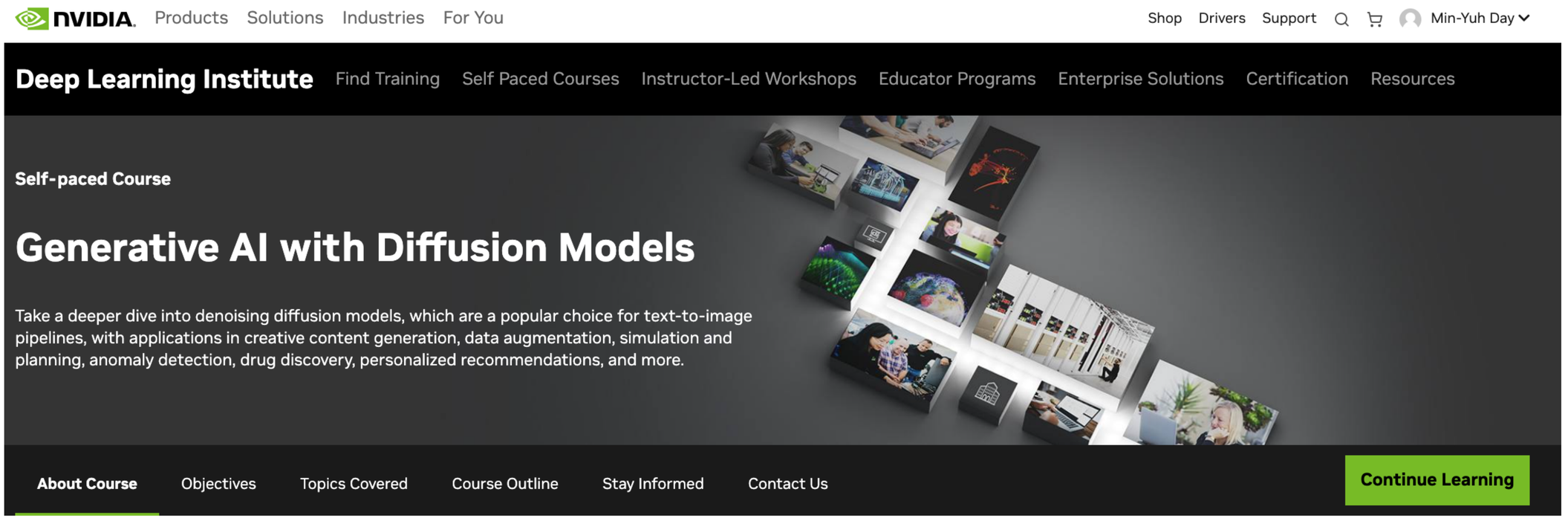
https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

NVIDIA Deep Learning Institute (DLI)



https://learn.nvidia.com/en-us/training/self-paced-courses

Generative AI with Diffusion Models



About this Course

Thanks to improvements in computing power and scientific theory, generative AI is more accessible than ever before. Generative AI plays a significant role across industries due to its numerous applications, such as creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more. In this course, learners will take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines.

Course Details

Duration: 08:00

Price: $90

Subject: Generative AI/LLM

Language: English

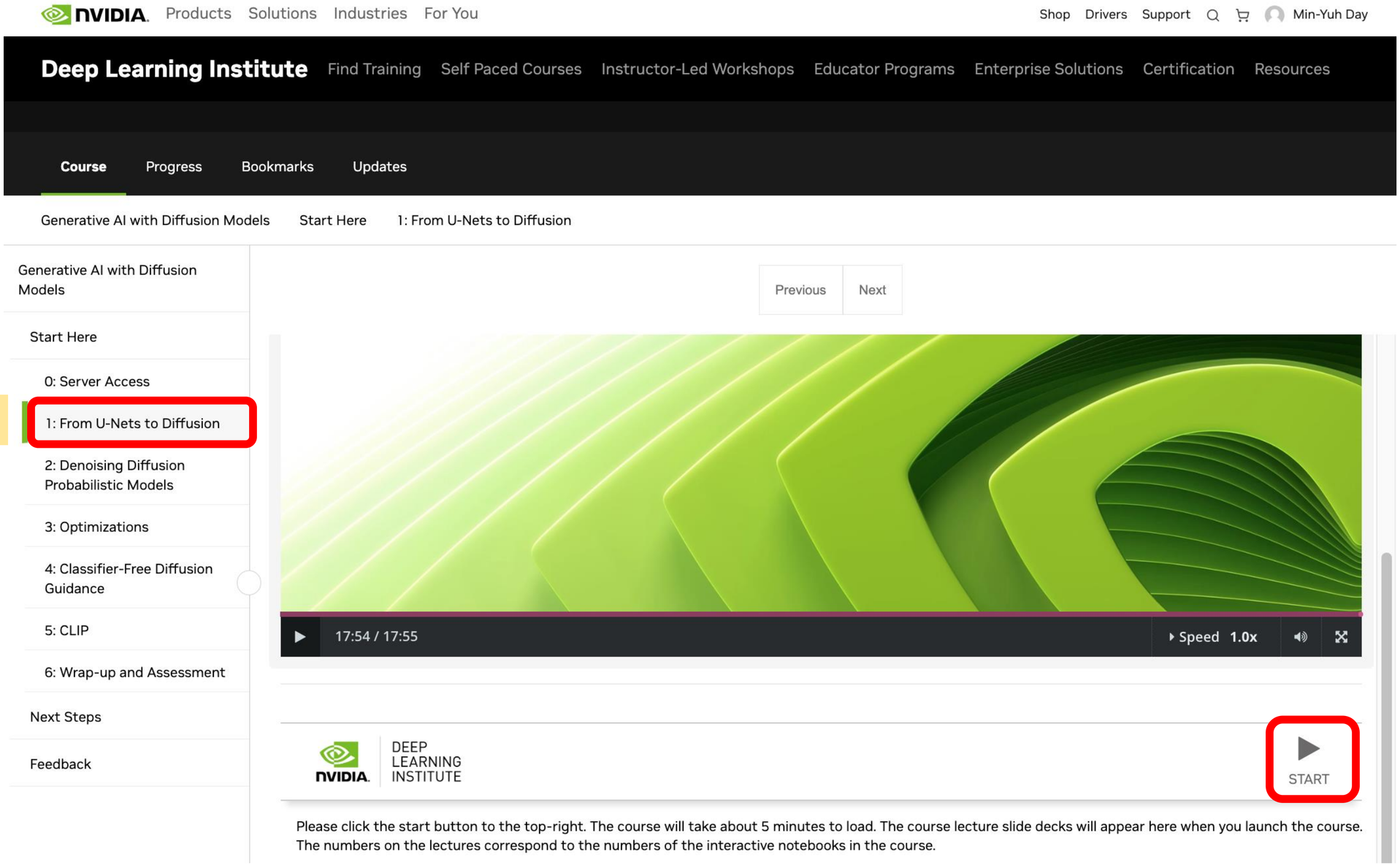
Course Prerequisites:
A basic understanding of Deep Learning Concepts.

Learning Objectives

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

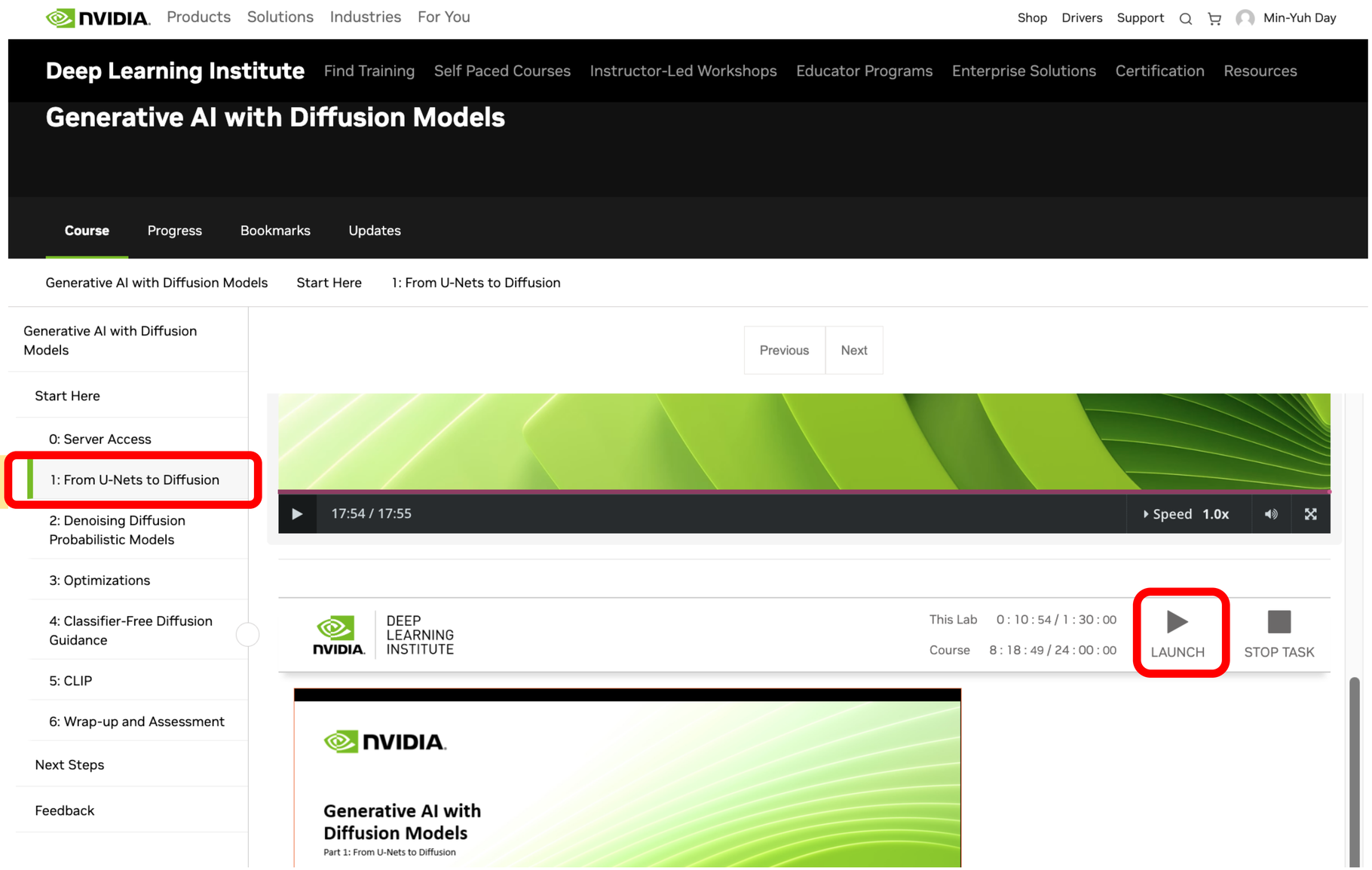
1: From U-Nets to Diffusion



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

1: From U-Nets to Diffusion

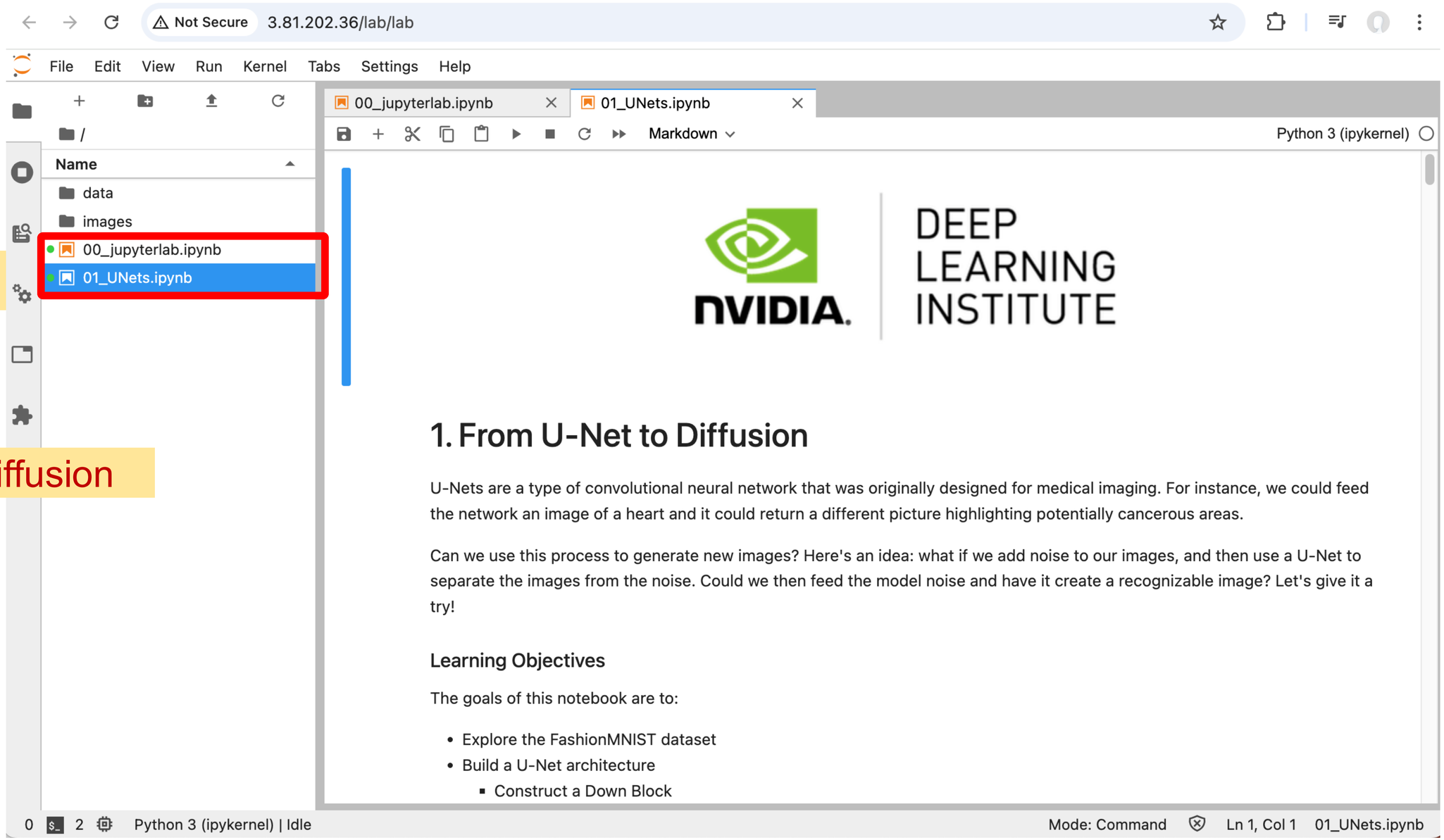


https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

01_UNets.ipynb

1: From U-Nets to Diffusion



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

Part 1: From U-Nets to Diffusion

Agenda

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier Free Diffusion
- Part 5: CLIP
- Part 6: Wrap-up & Assessment

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Prerequisites

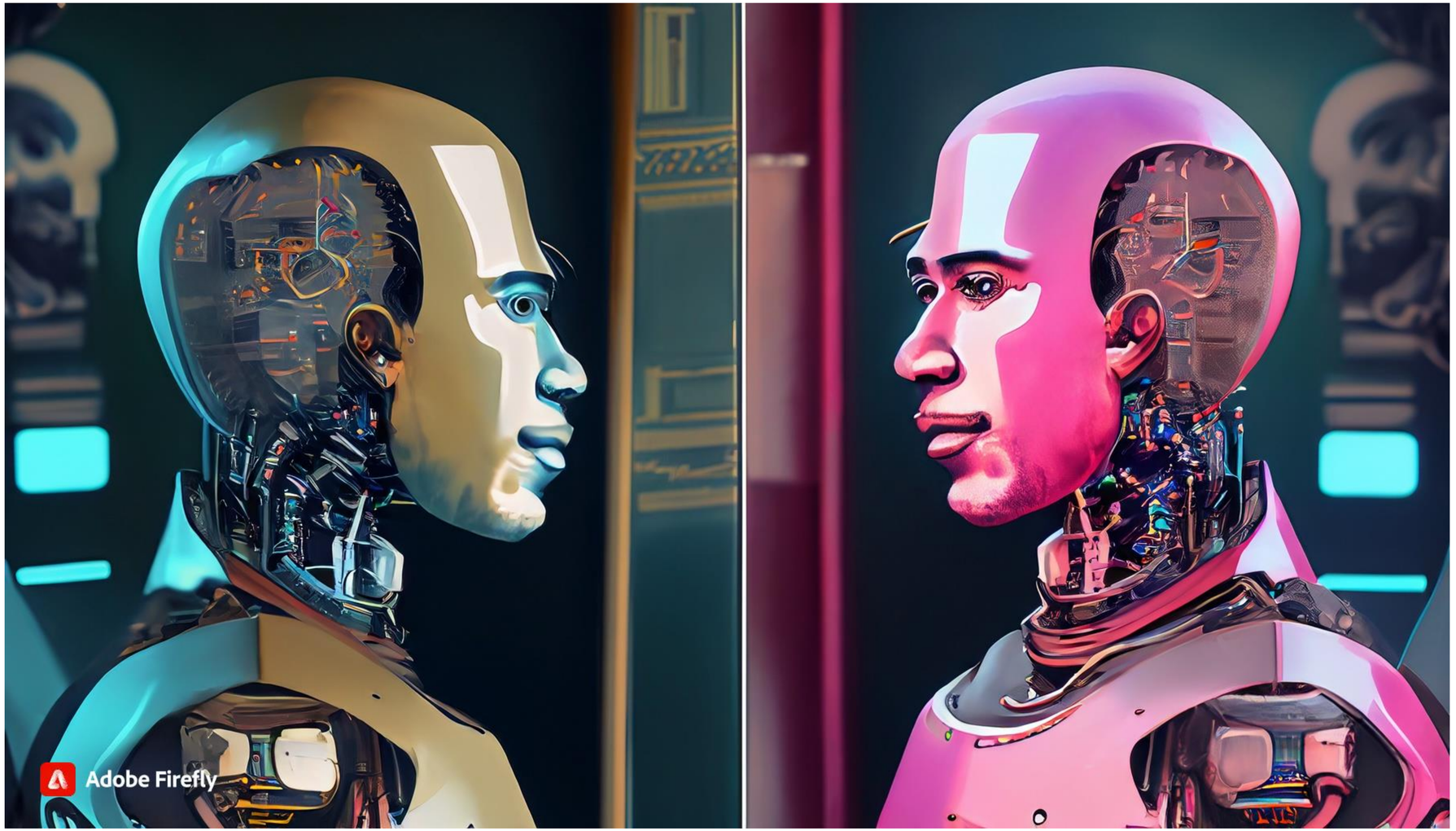
- Basic familiarity with convolutional neural networks (CNNs)
- Basic familiarity with a deep learning framework such as:
  - PyTorch
  - TensorFlow

A Brief History of Generative AI

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

The Imitation Game

A.K.A The Turing Test

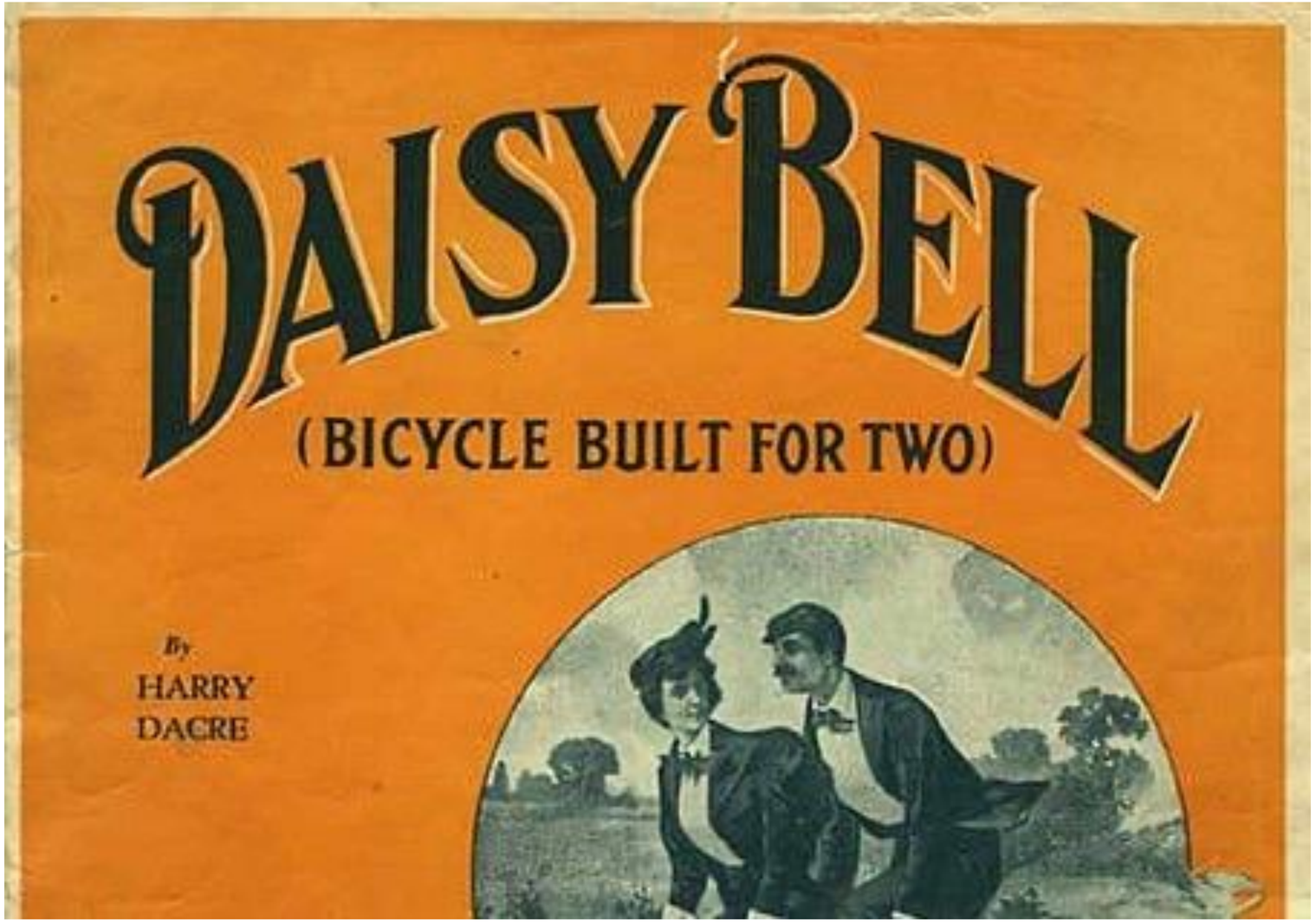


A robot looks in a mirror and the reflection is human, cyberpunk

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

IBM 704

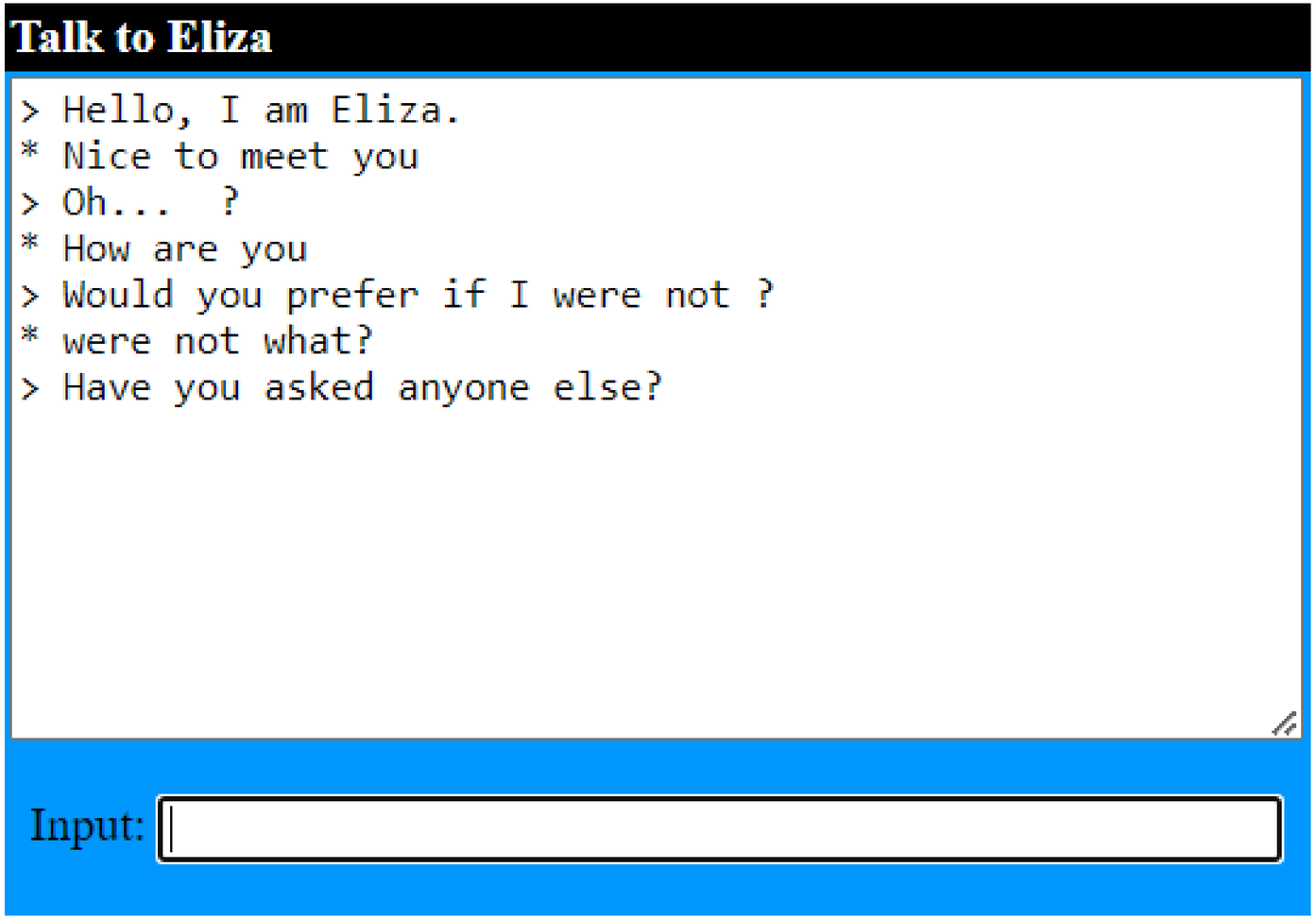
The First Singing Computer



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Eliza

The First Gen AI Chatbot?



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI of the 70’s, 80’s and 90s?

- Electronic music
- Video games graphics
- Video game AI
- Computer animation
- Instant messaging chatbots

An 80’s arcade with lots of machines

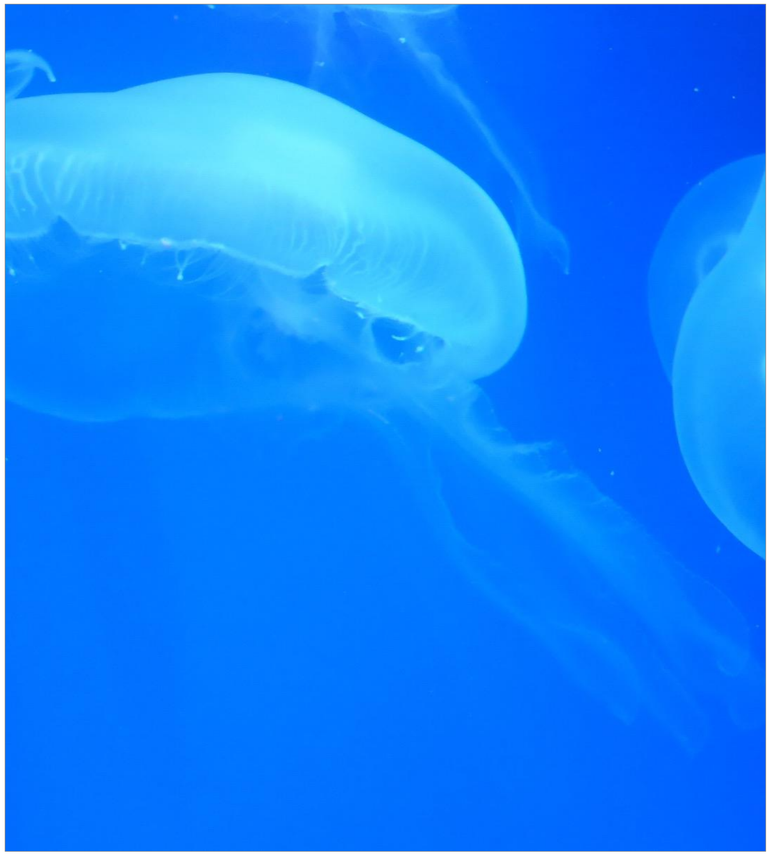
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

The Rise of Neural Networks

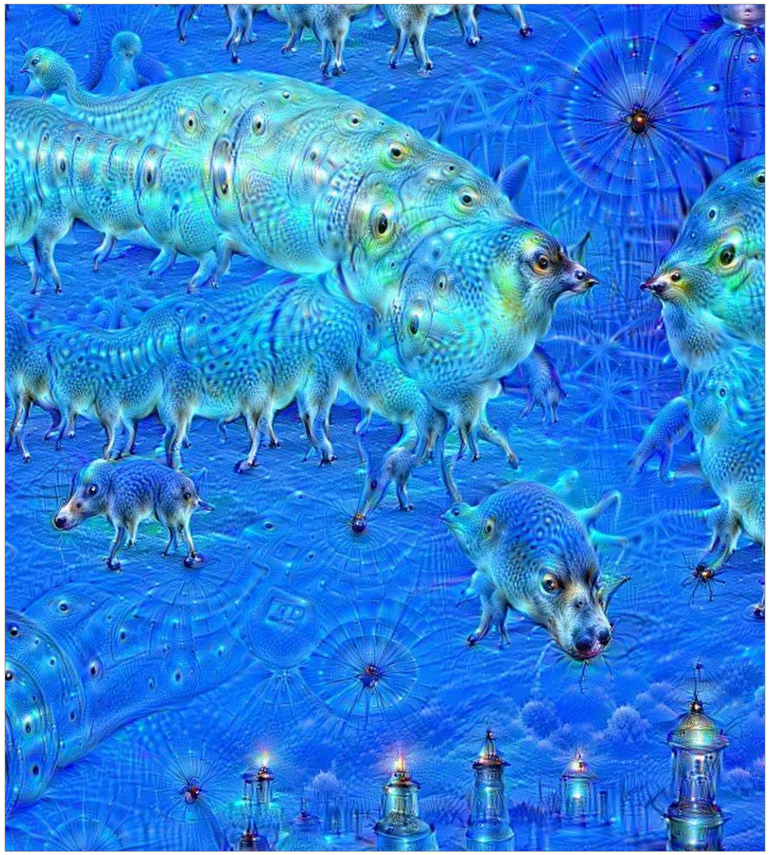
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Deep Dreaming

Images by Martin Thoma



Original



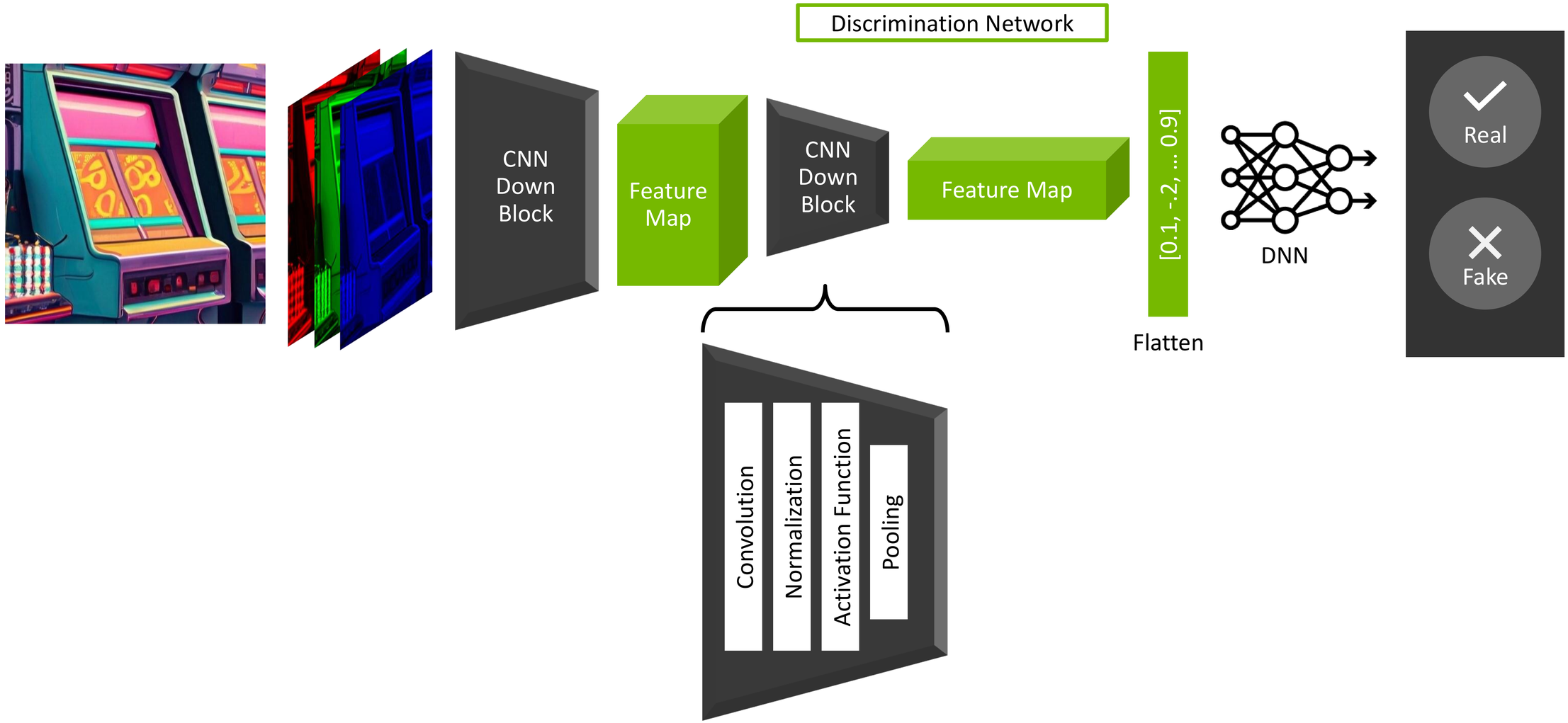
10 Iterations

50 Iterations

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

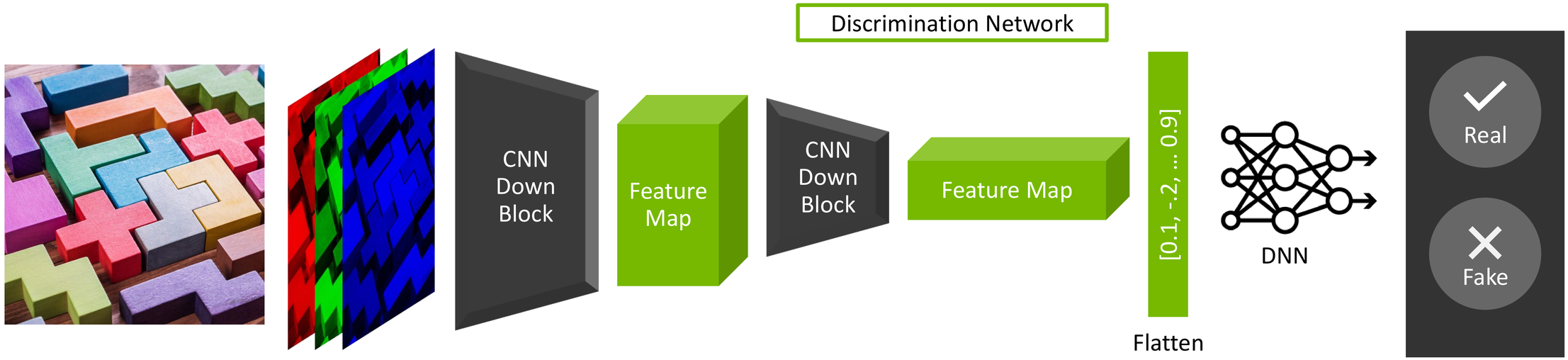
GANs

Generative Adversarial Networks



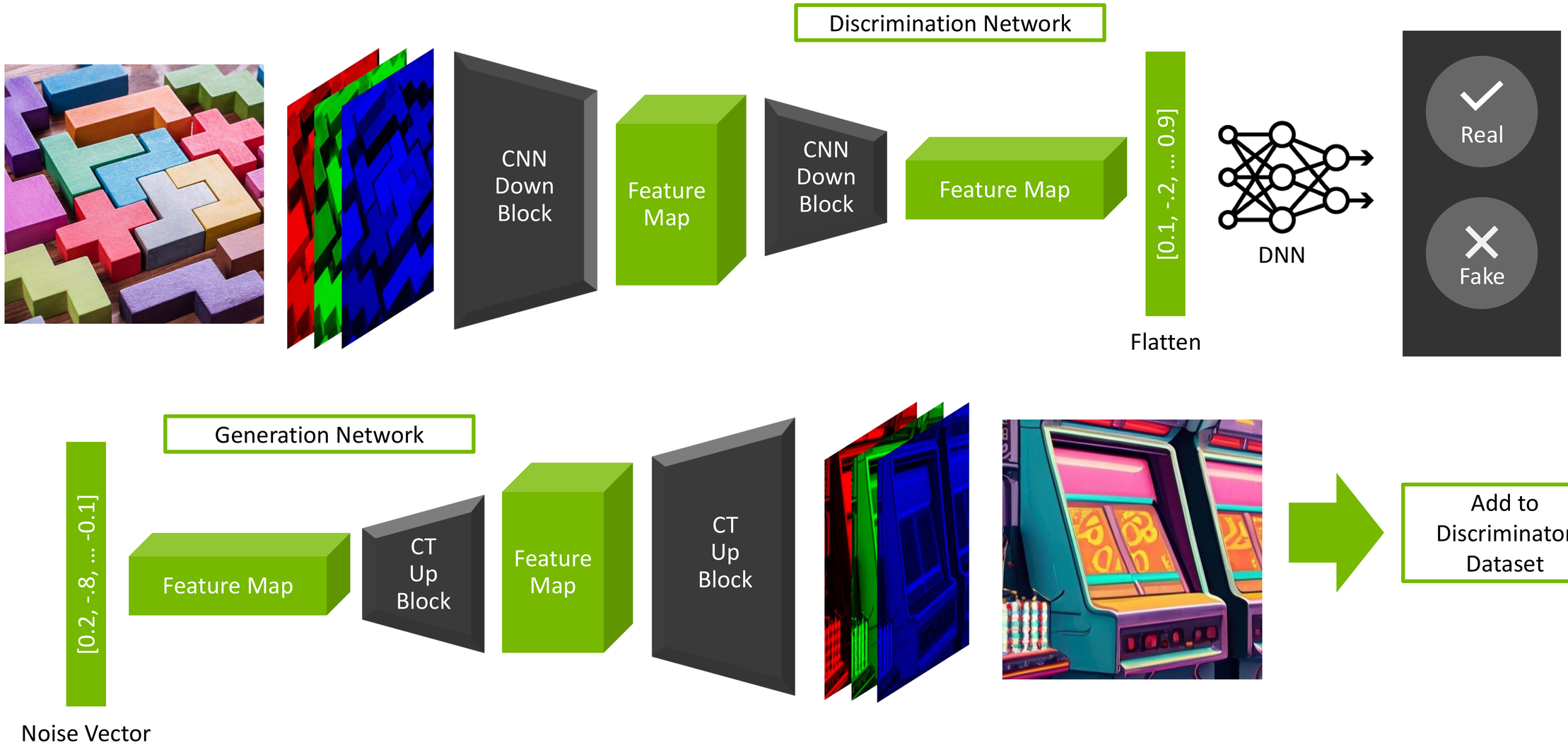
GANs

Generative Adversarial Networks



GANs

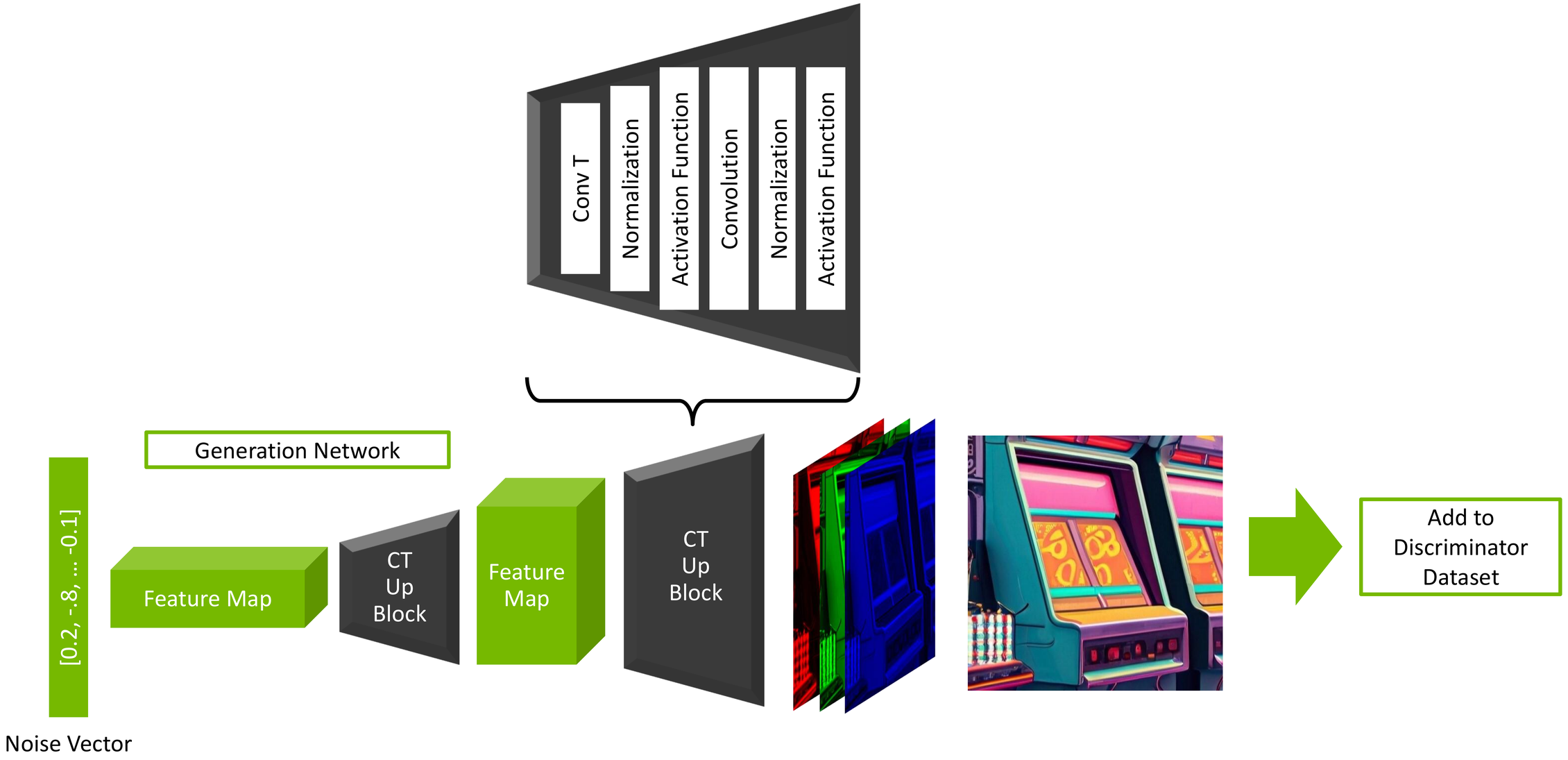
Generative Adversarial Networks



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

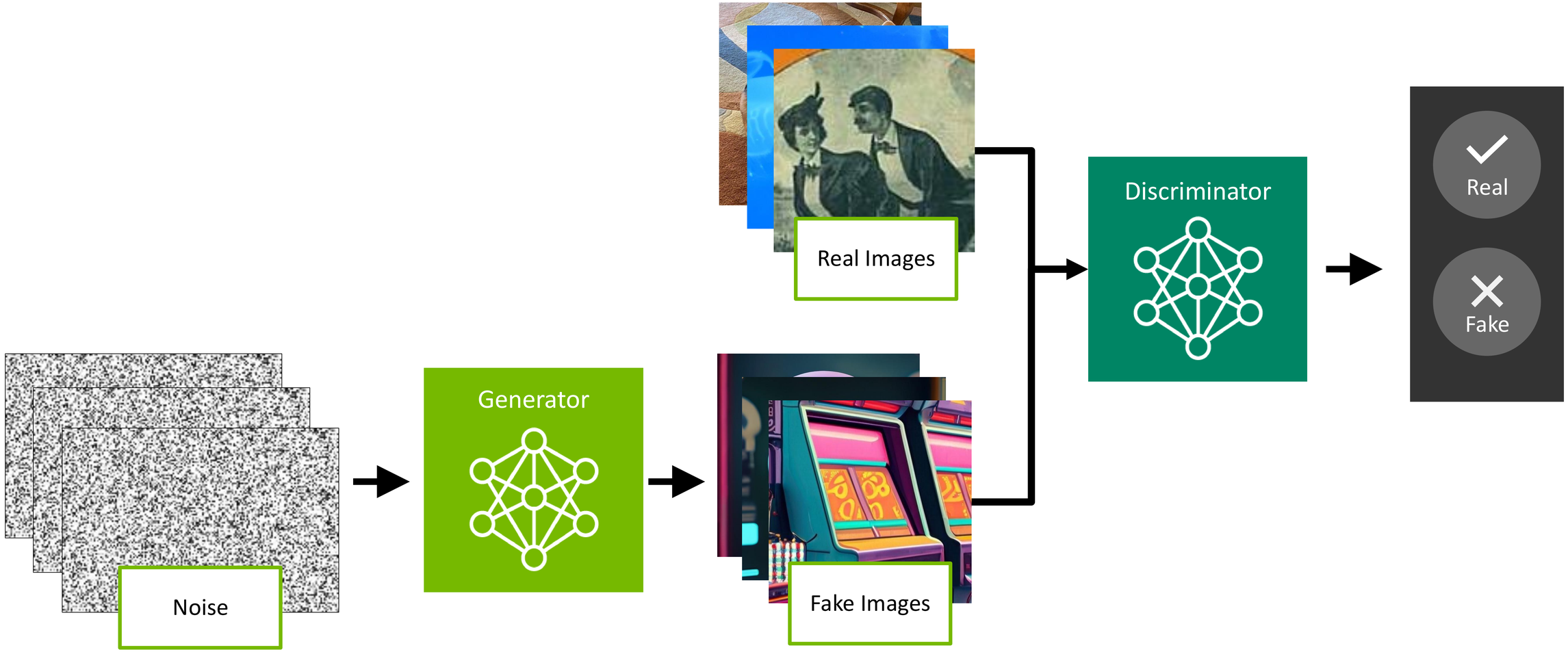
GANs

Generative Adversarial Networks



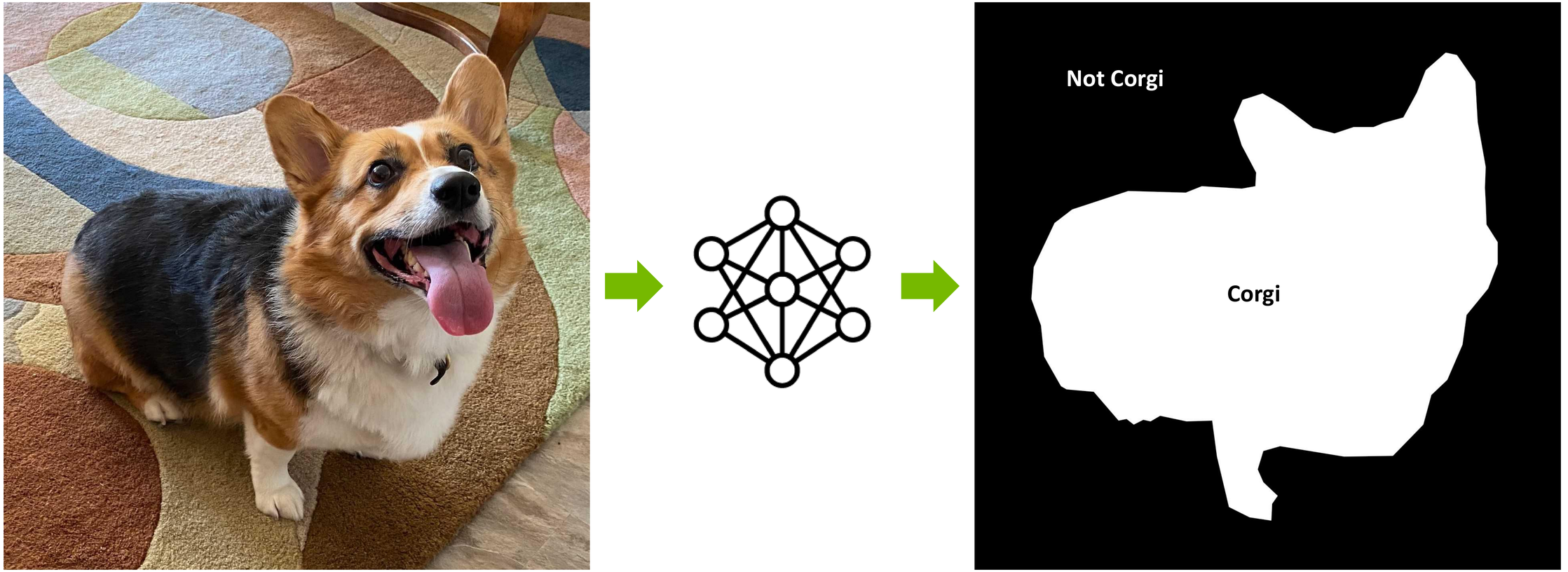
GANs

Generative Adversarial Networks



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

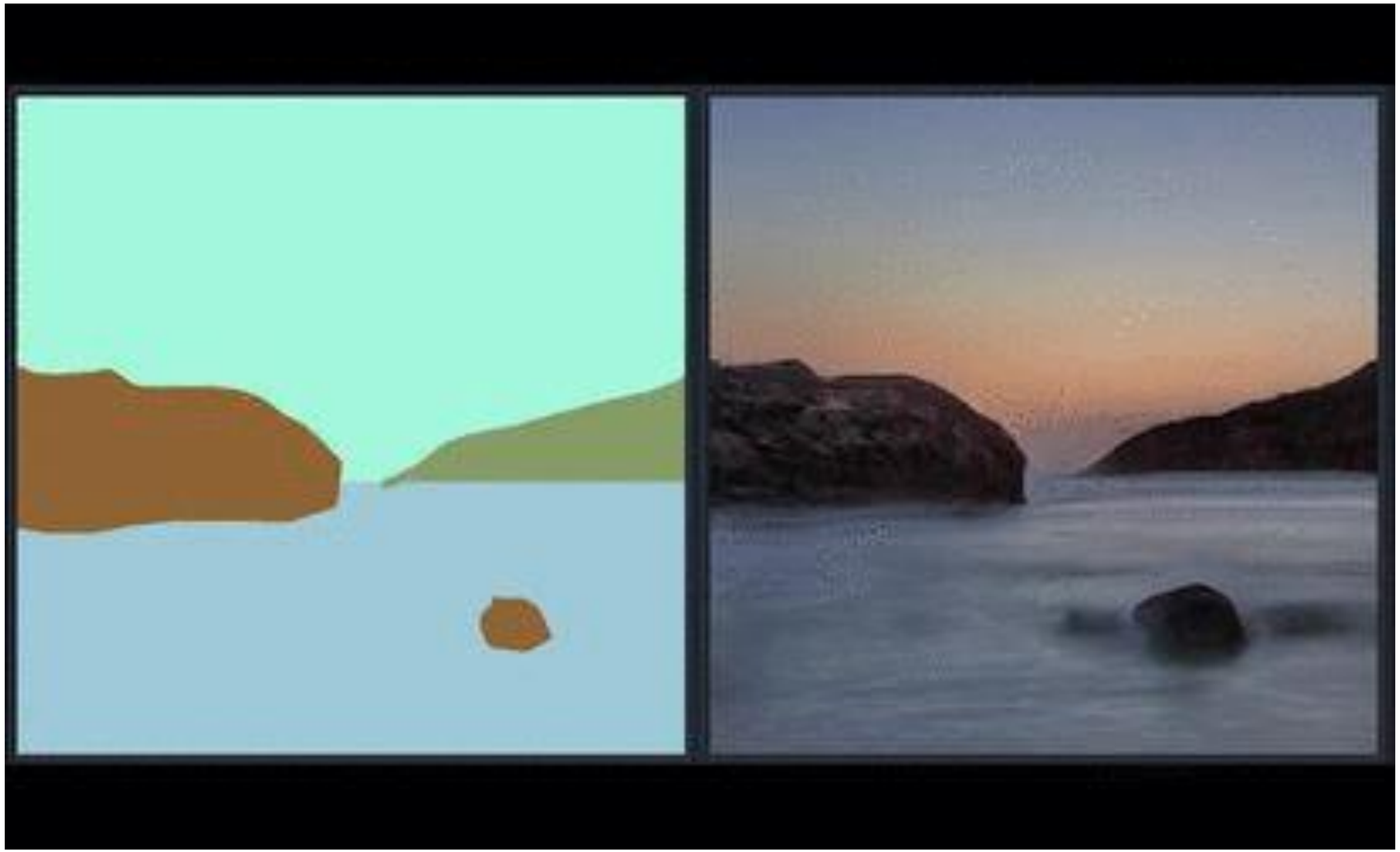
Image Segmentation



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Image Segmentation + GANs

NVIDIA Spade

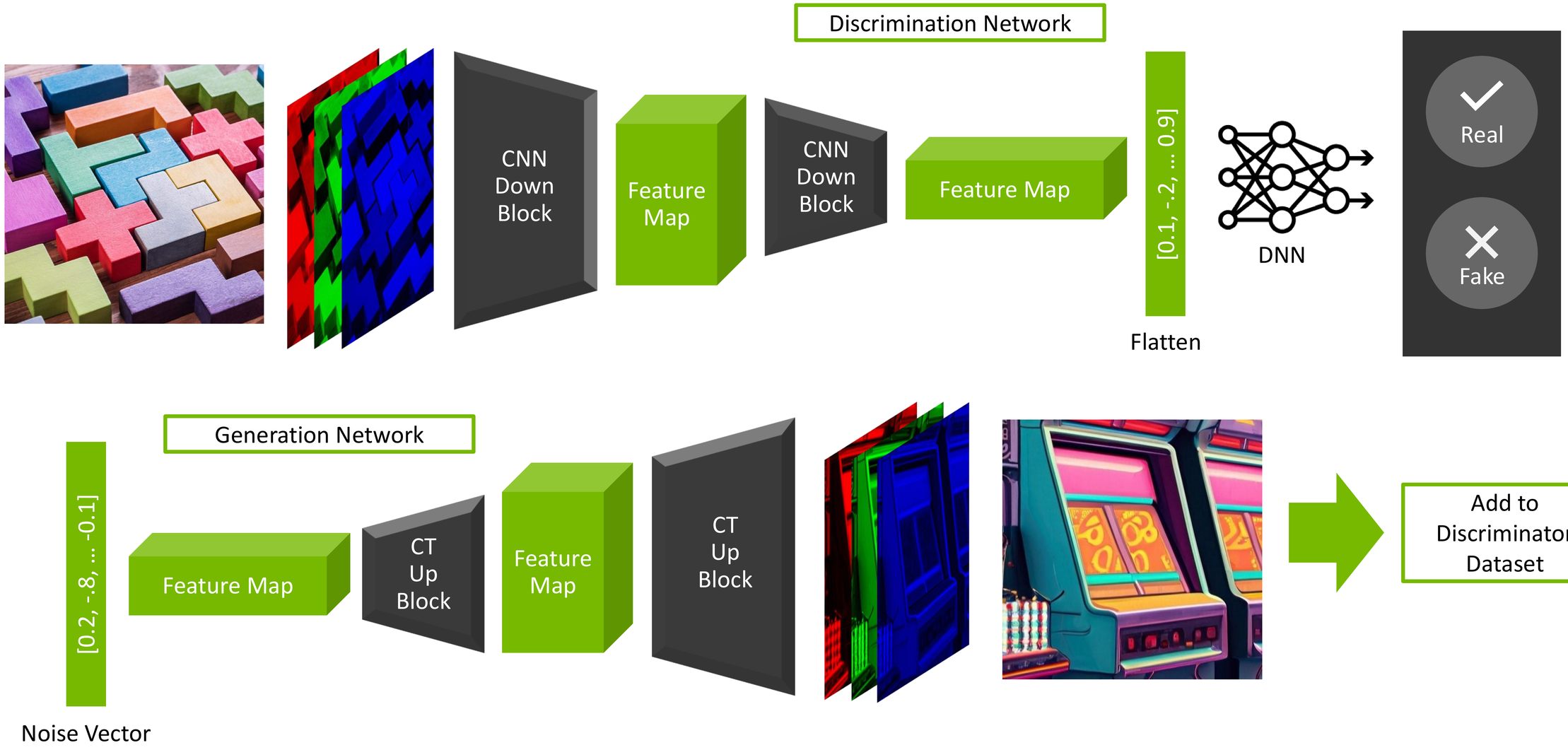


Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

U-Nets

GANs

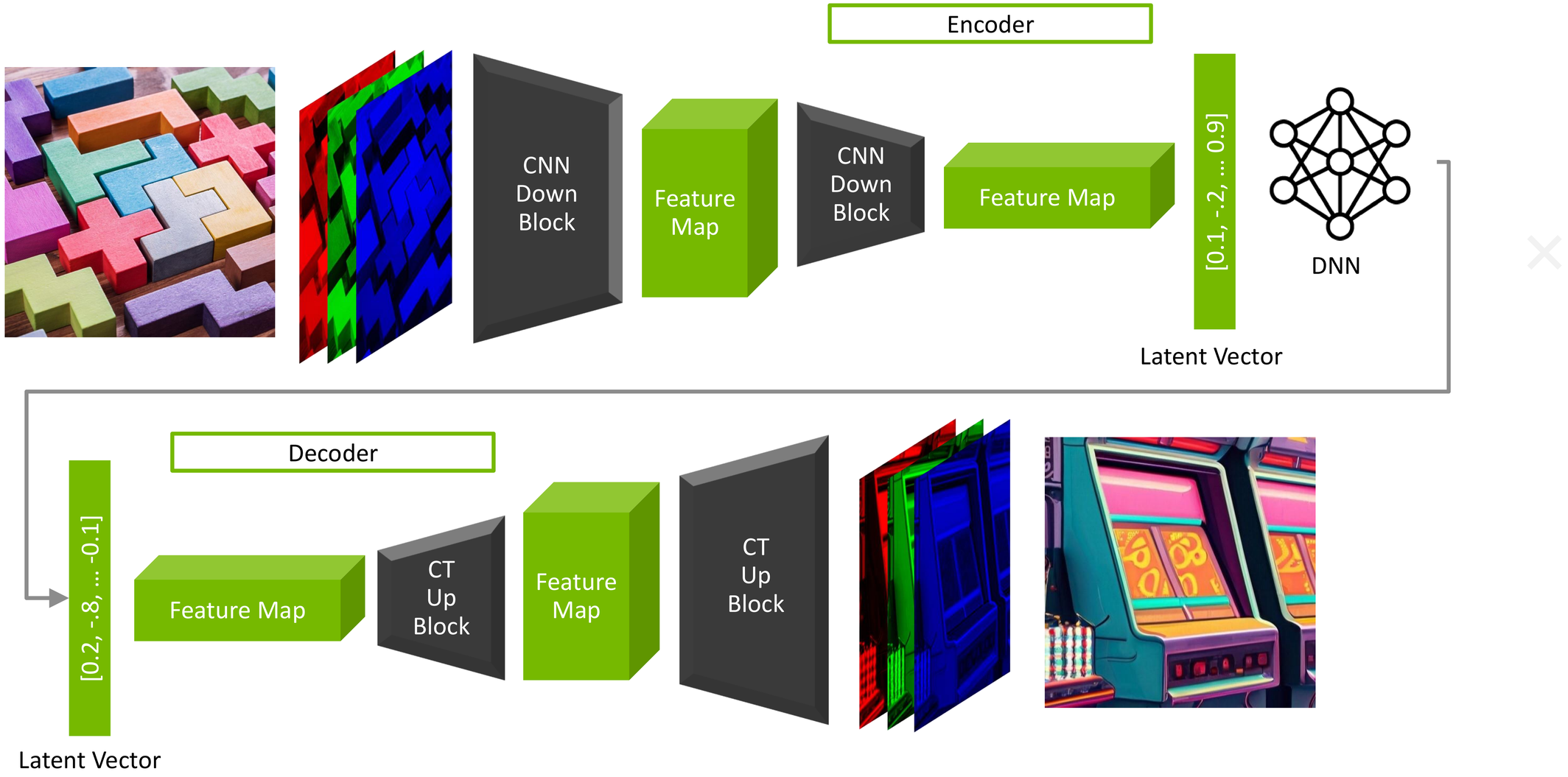
Generative Adversarial Networks



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

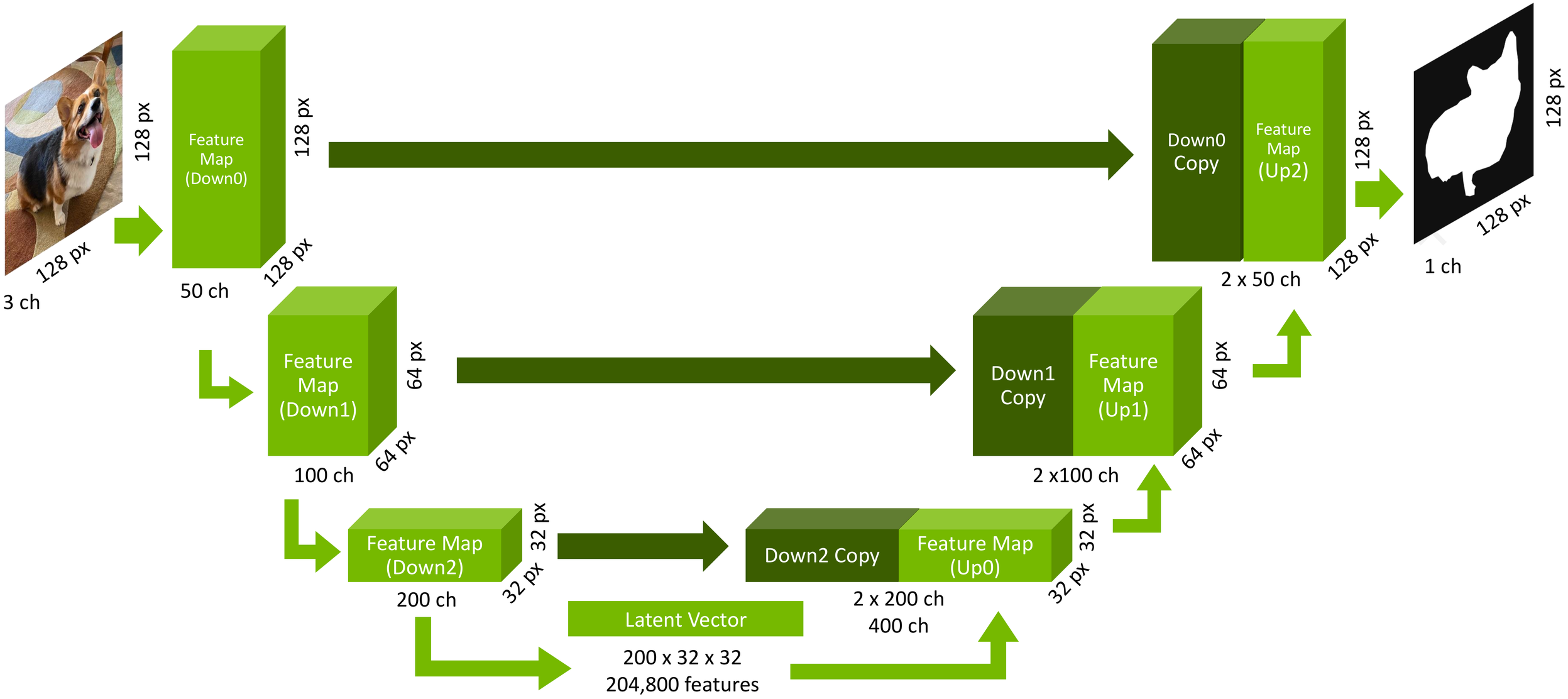
~~GANs~~ U-Nets

The U shaped Autoencoder



U-Nets

The U shaped Autoencoder



Transposed Convolution

Transposed Convolution

Convolution Review

Kernel

.25	.25
.25	.25

Image

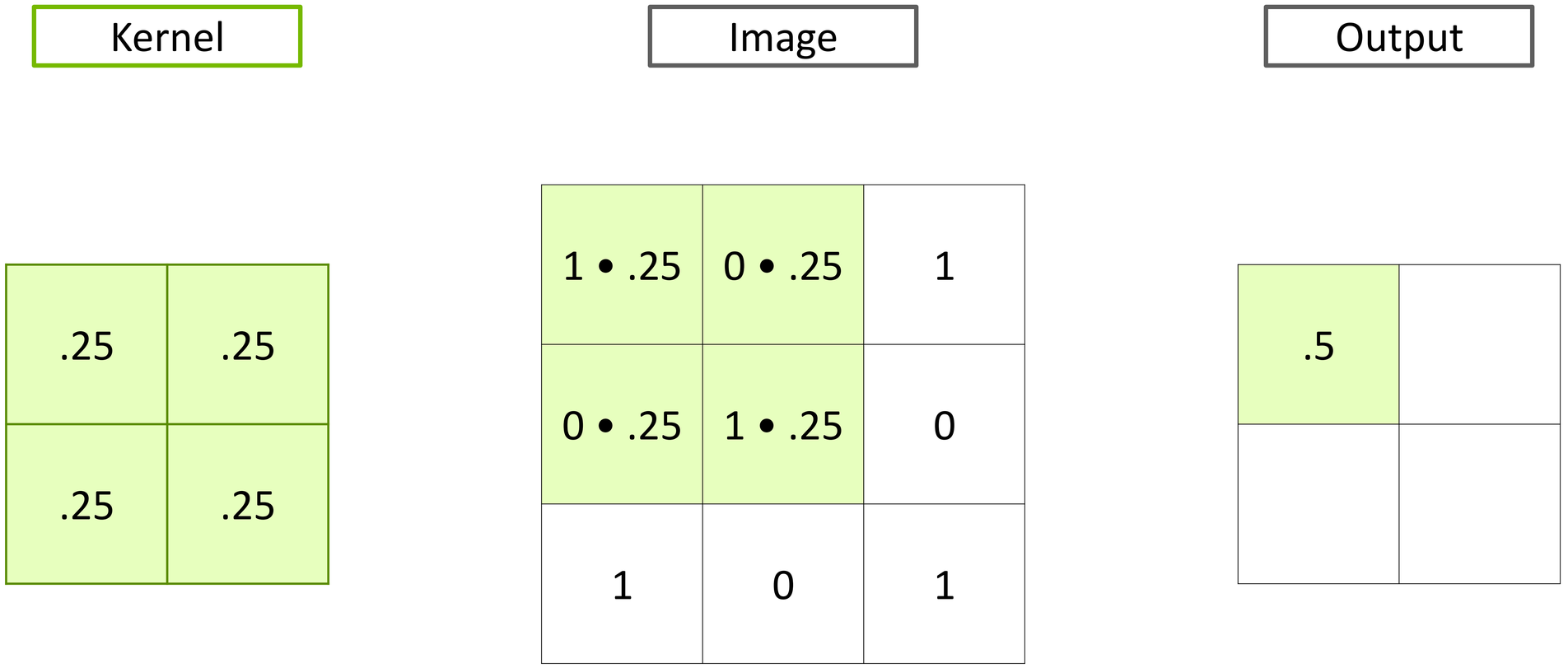
1	0	1
0	1	0
1	0	1

Output

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Convolution Review



Transposed Convolution

Convolution Review

Kernel

.25	.25
.25	.25

Image

1	0 • .25	1 • .25
0	1 • .25	0 • .25
1	0	1

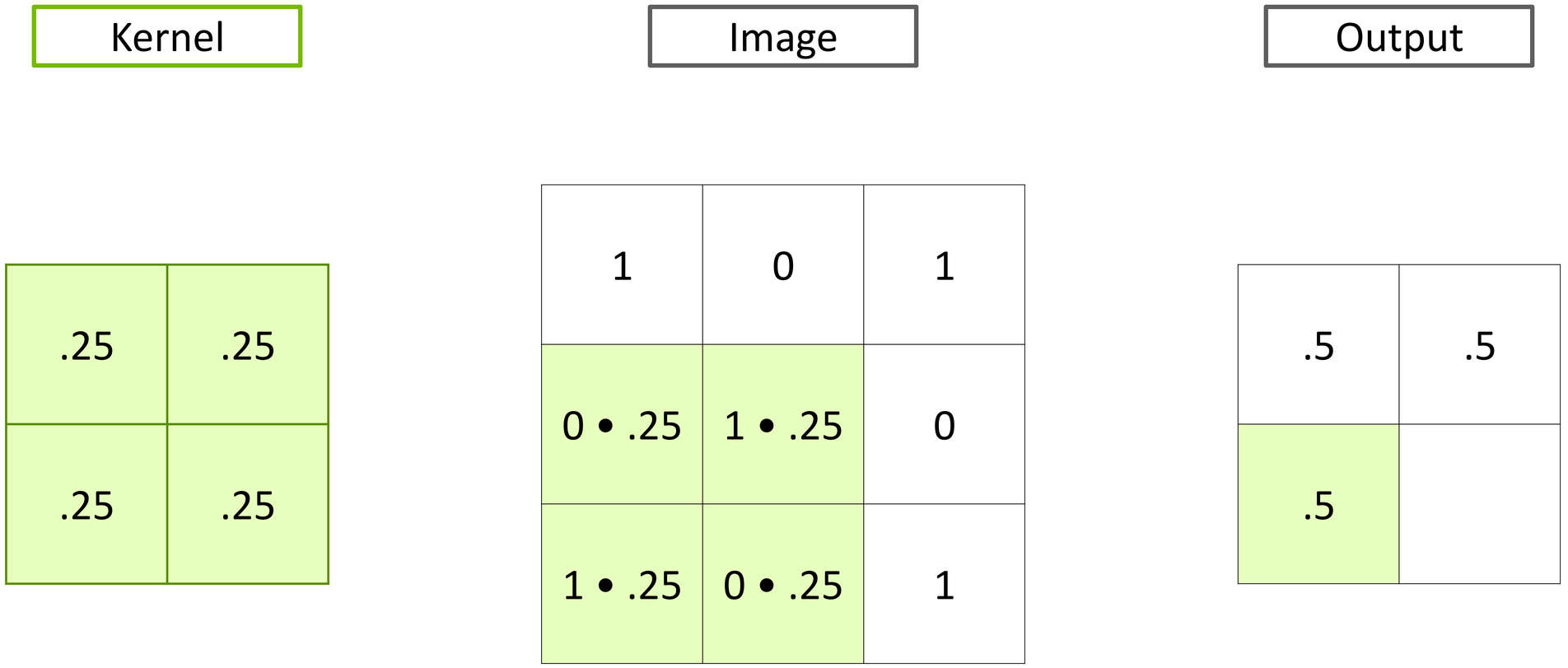
Output

.5	.5

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

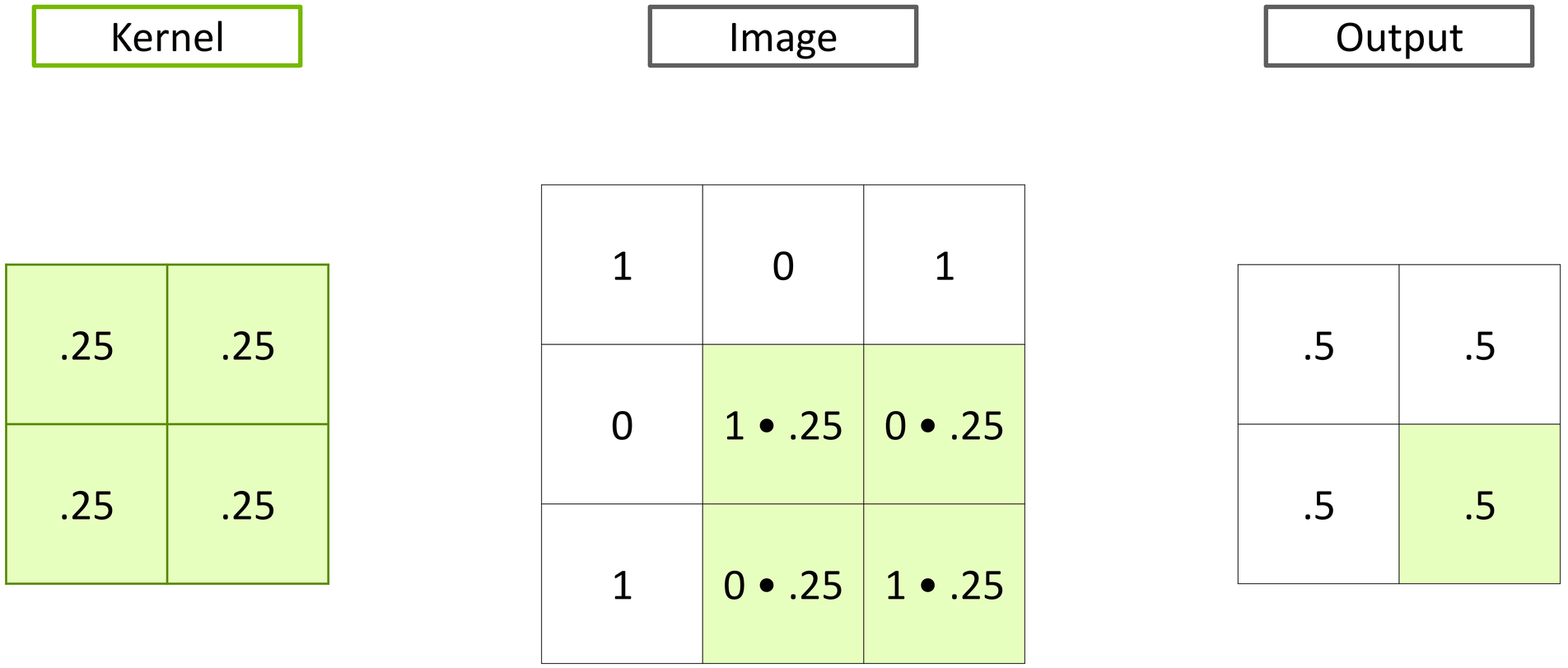
Transposed Convolution

Convolution Review



Transposed Convolution

Convolution Review



Transposed Convolution

Image Upscaling

Kernel

.25	.25
.25	.25

Image

1	0	1
0	1	0
1	0	1

Output

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Image Upscaling

Kernel

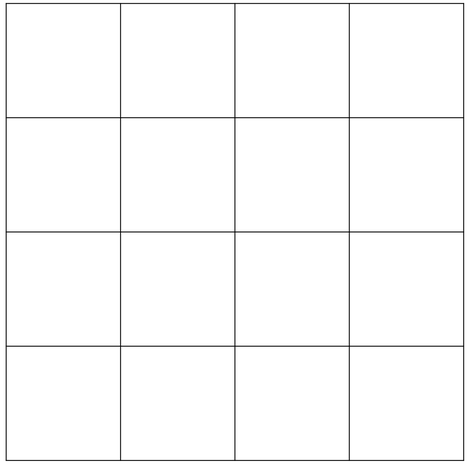
.25	.25
.25	.25

Image

Stride = 2

1	0	0	0	1
0	0	0	0	0
0	0	1	0	0
0	0	0	0	0
1	0	0	0	1

Output



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Image Upscaling

Kernel

.25	.25
.25	.25

Image

Stride = 2

1 • .25	0 • .25	0	0	1
0 • .25	0 • .25	0	0	0
0	0	1	0	0
0	0	0	0	0
1	0	0	0	1

Output

.25			

Transposed Convolution

Image Upscaling

Kernel

.25	.25
.25	.25

Image

Stride = 2

1	0 • .25	0 • .25	0	1
0	0 • .25	0 • .25	0	0
0	0	1	0	0
0	0	0	0	0
1	0	0	0	1

Output

.25	0		

Transposed Convolution

Image Upscaling

Kernel

.25	.25
.25	.25

Image

Stride = 2

1	0	0	0	1
0	0	0	0	0
0	0	1	0	0
0	0	0	0	0
1	0	0	0	1

Output

.25	0	0	.25
0	.25	.25	0
0	.25	.25	0
.25	0	0	.25

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Stride

Image

Stride = 2

1	0	0	0	1
0	0	0	0	0
0	0	1	0	0
0	0	0	0	0
1	0	0	0	1

Image

Stride = 3

1	0	0	0	0	0	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
1	0	0	0	0	0	1

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Padding

Image

Stride = 3

Padding = 0

1	0	0	0	0	0	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
1	0	0	0	0	0	1

Image

Stride = 3

Padding = 2

		0	0	0		
		0	1	0		
		0	0	0		

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Padding

Image

Stride = 3

Padding = 0

1	0	0	0	0	0	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
1	0	0	0	0	0	1

Image

Stride = 3

Padding = 2

0	0	0
0	1	0
0	0	0

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Transposed Convolution

Out Padding

Image

Out Padding = 0

1	0	1
0	1	0
1	0	1

Image

Out Padding = 1

1	0	1	0
0	1	0	0
1	0	1	0
0	0	0	0

Image

Out Padding = 2

1	0	1	0	0
0	1	0	0	0
1	0	1	0	0
0	0	0	0	0
0	0	0	0	0

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Image Resizing

Upsampling

64 px

64 px

128 px

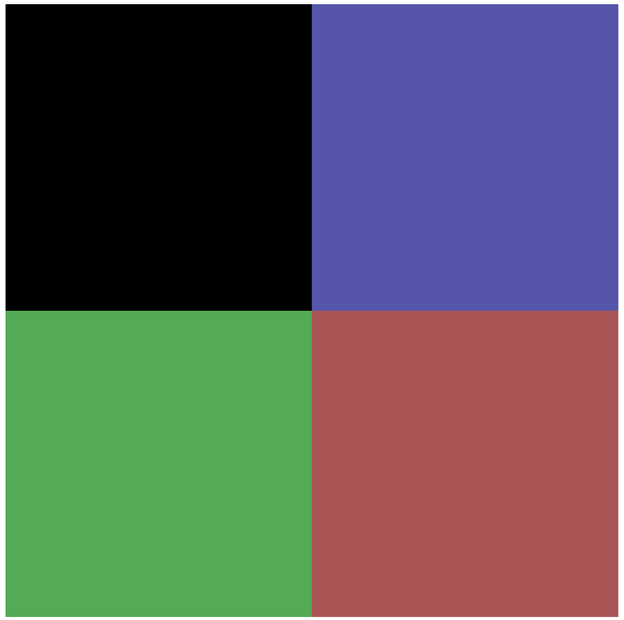
128 px

192 px

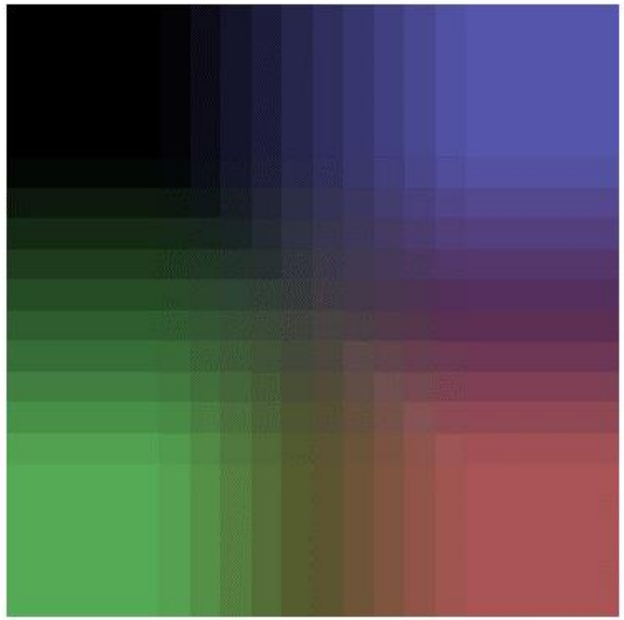
192 px

Image Resizing

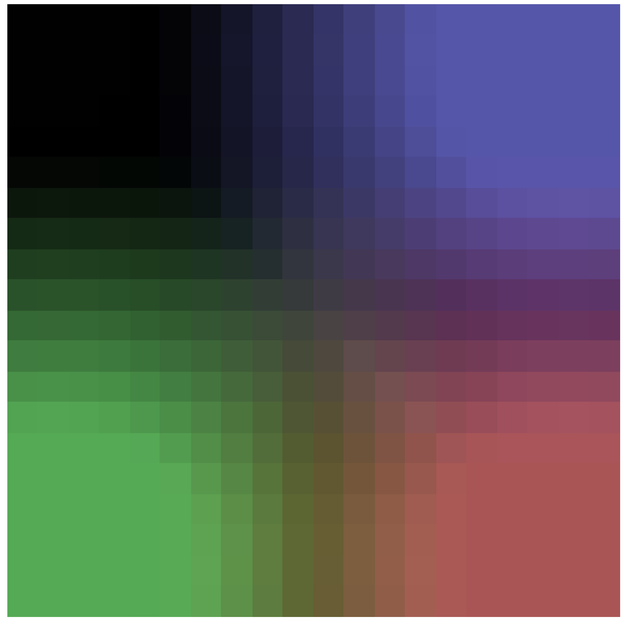
Upsampling



Nearest



Bilinear

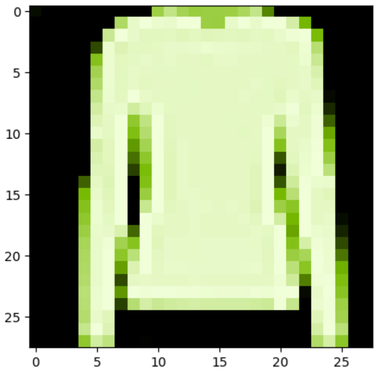


Bicubic

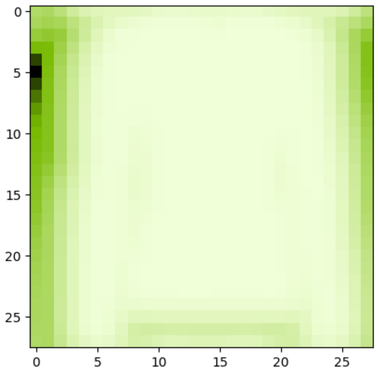
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Deconvolution?

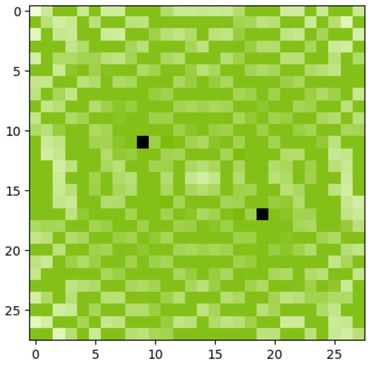
Same as Transposed Convolution?



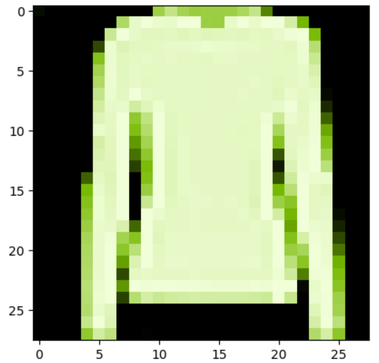
Image



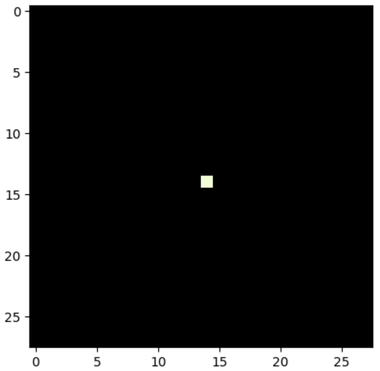
Convolved Image



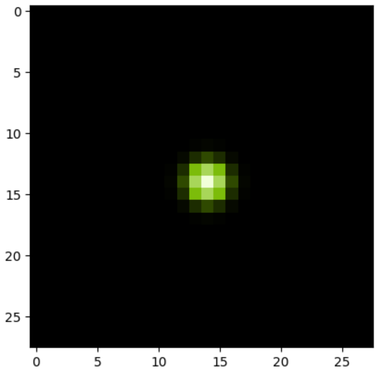
FFT Convolved



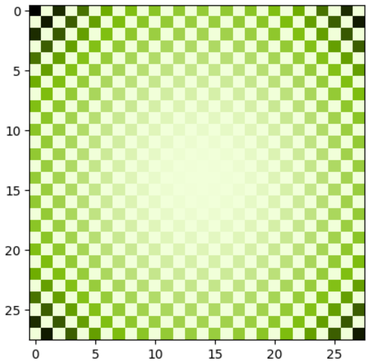
Deconvolved Convolved Image



Point



Point Spread Function



FFT Point Spread Function

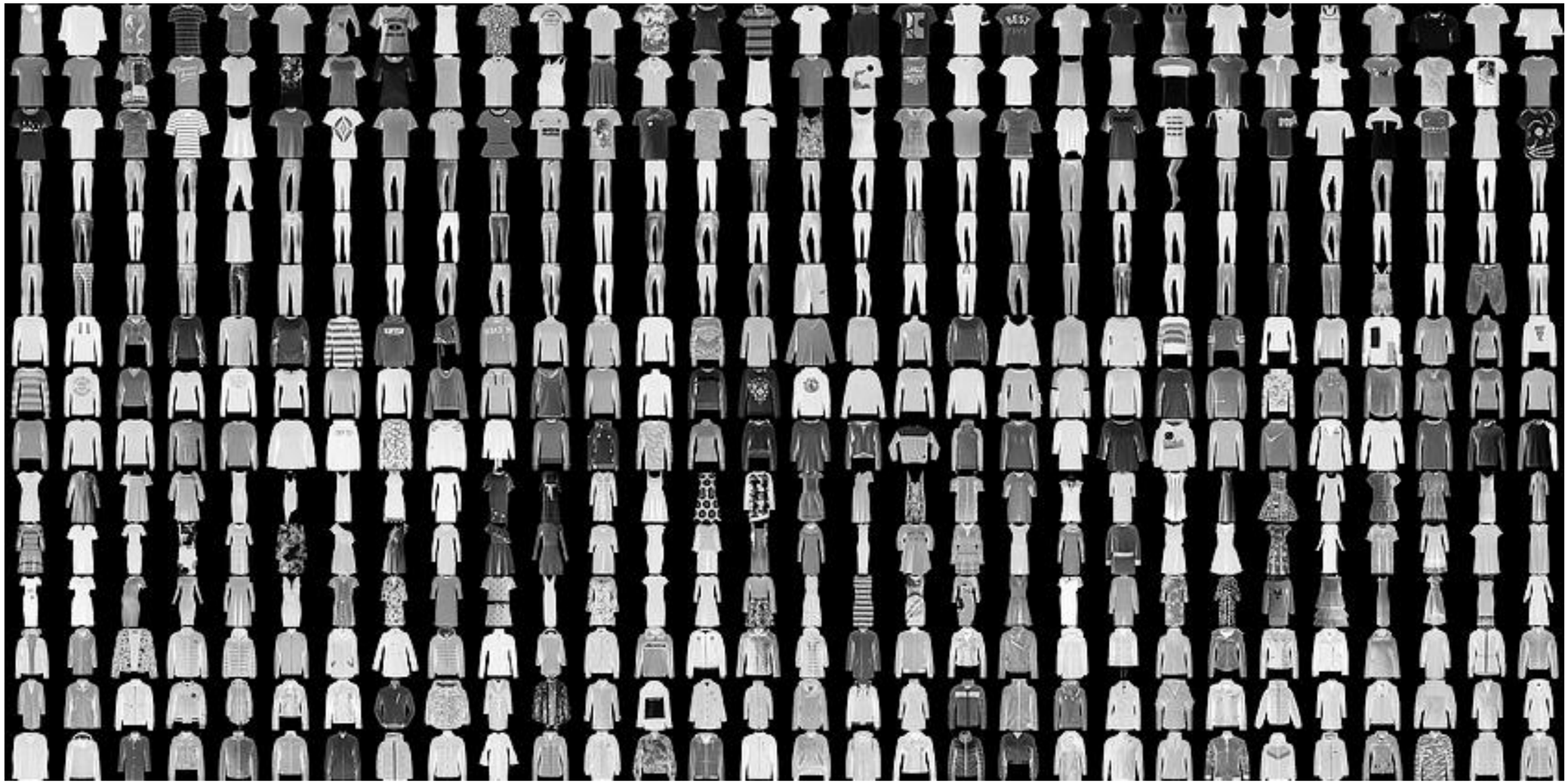
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Lab

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

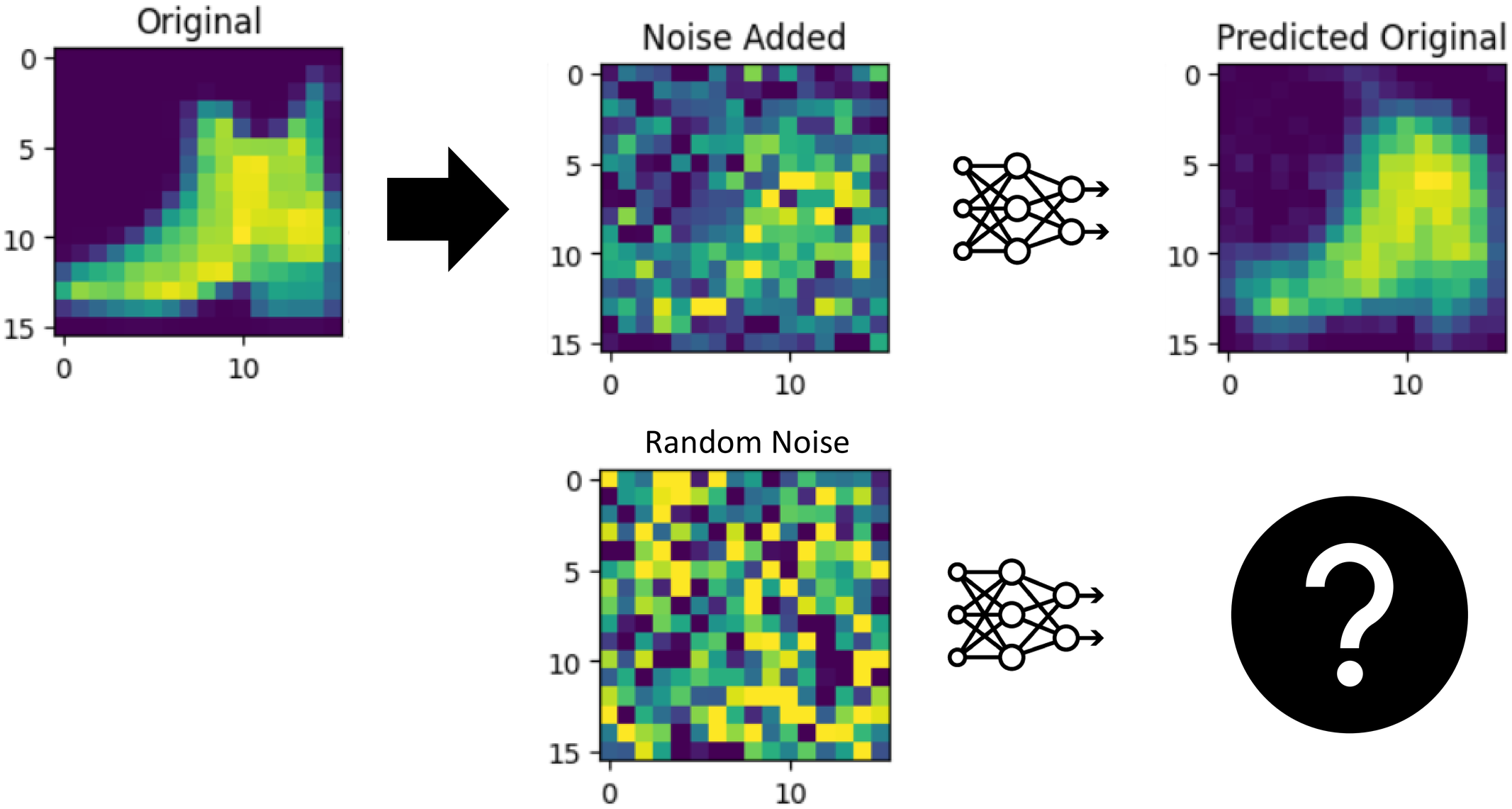
FashionMNIST

Convolutional Neural Network “Hello World”

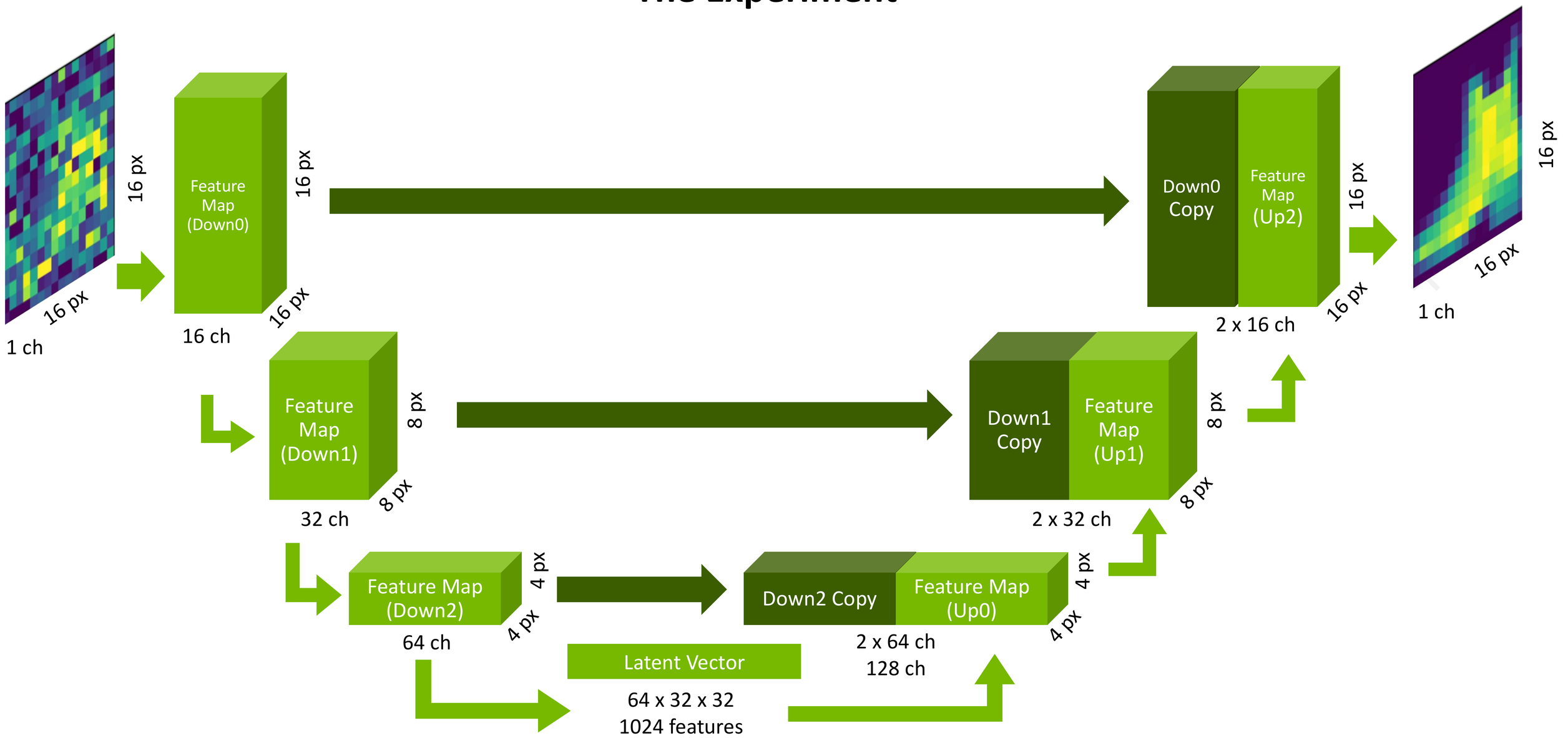


Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Hypothesis: Generate an image from Noise



The Experiment



Let’s get started!

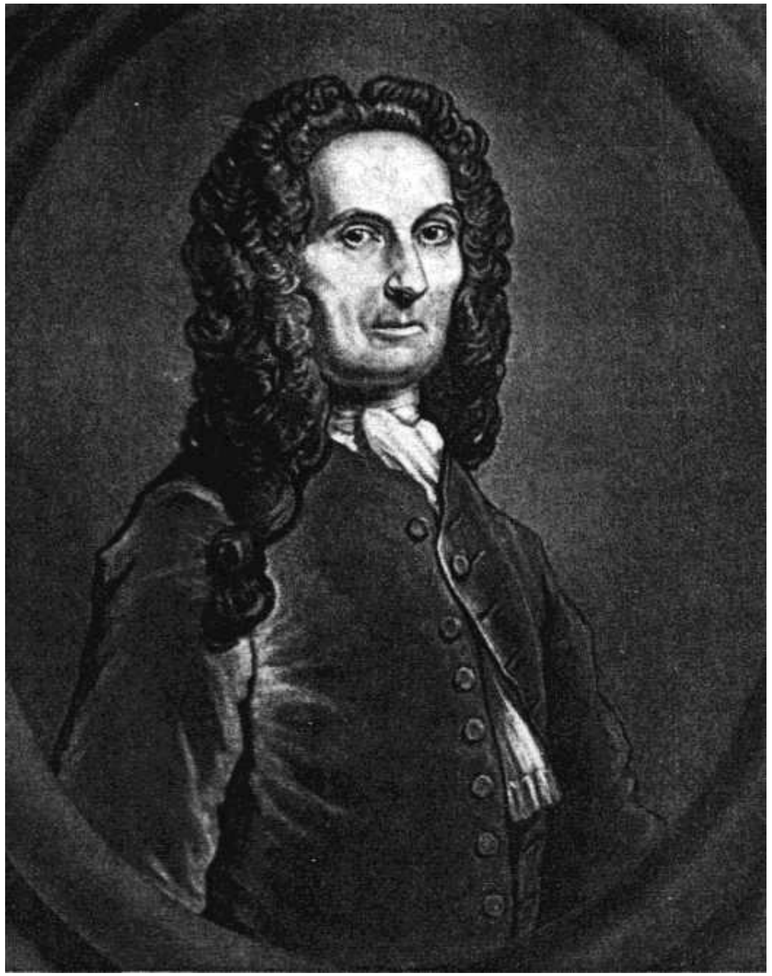
1: From U-Nets to Diffusion

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Appendix:
The Normal Distribution

De Moivre

From Coin Flips to Bells



$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\Pr(X = k) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$\Pr(X = 2) = \frac{4!}{2!\,(4-2)!} \left(\frac{1}{2}\right)^2 (1-\frac{1}{2})^{4-2}$$

$$\Pr(X = 2) = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 2 \cdot 1} \left(\frac{1}{4}\right)\left(\frac{1}{4}\right)$$

$$\Pr(X = 2) = \frac{6}{16}$$

$$p = \frac{1}{2}$$

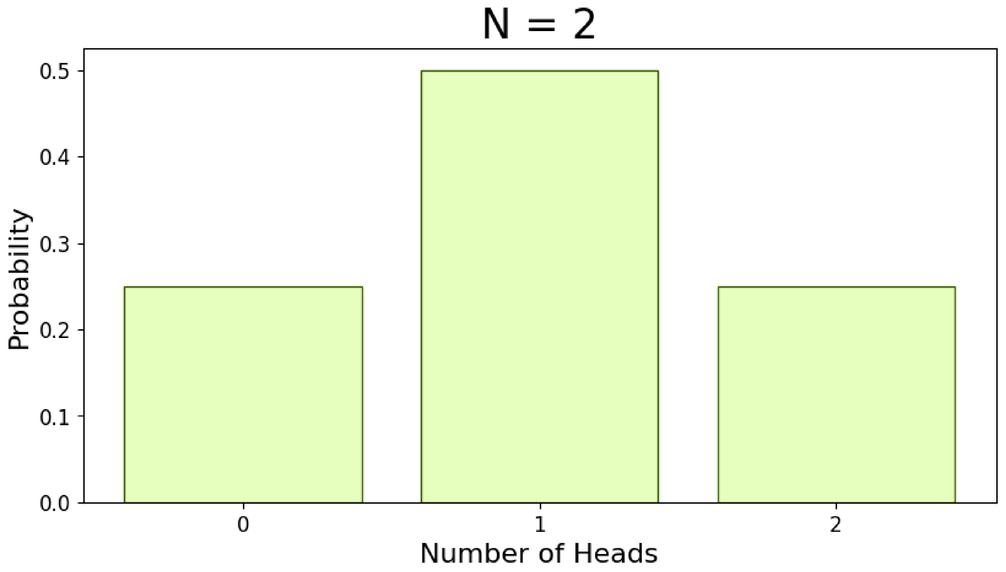
$$n = 4$$

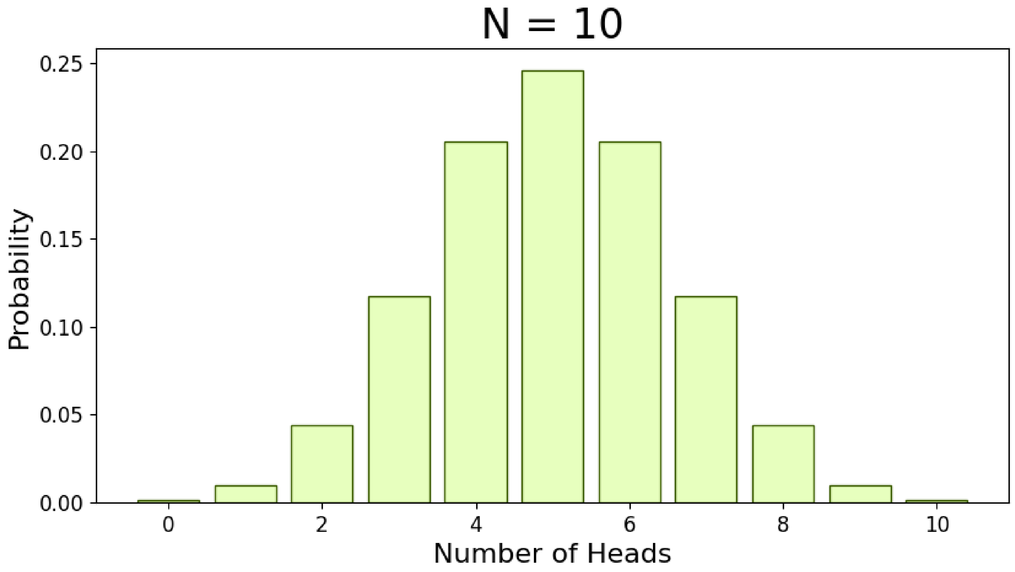
$$k = 2$$

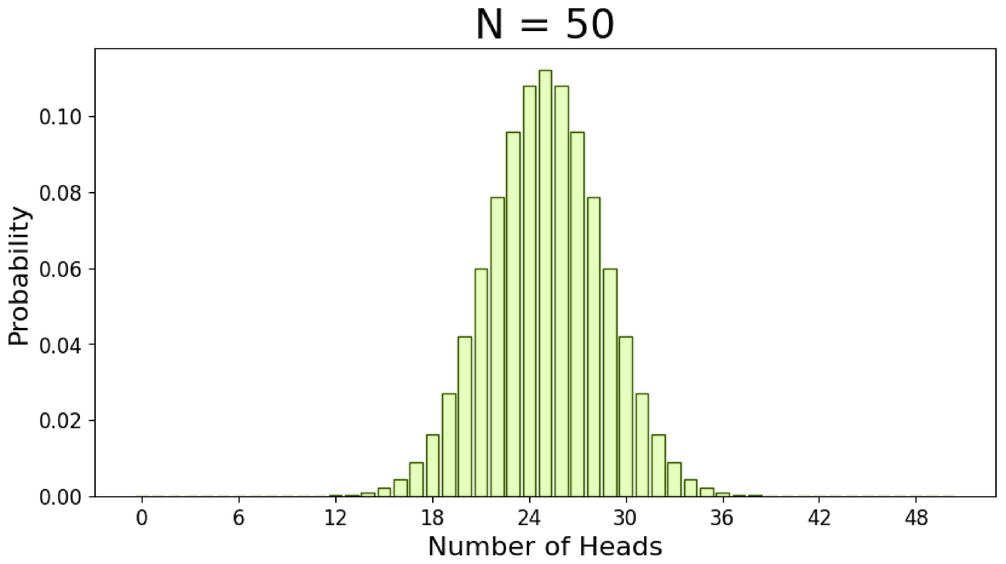
A weighted coin flipping through the air like a cartoon

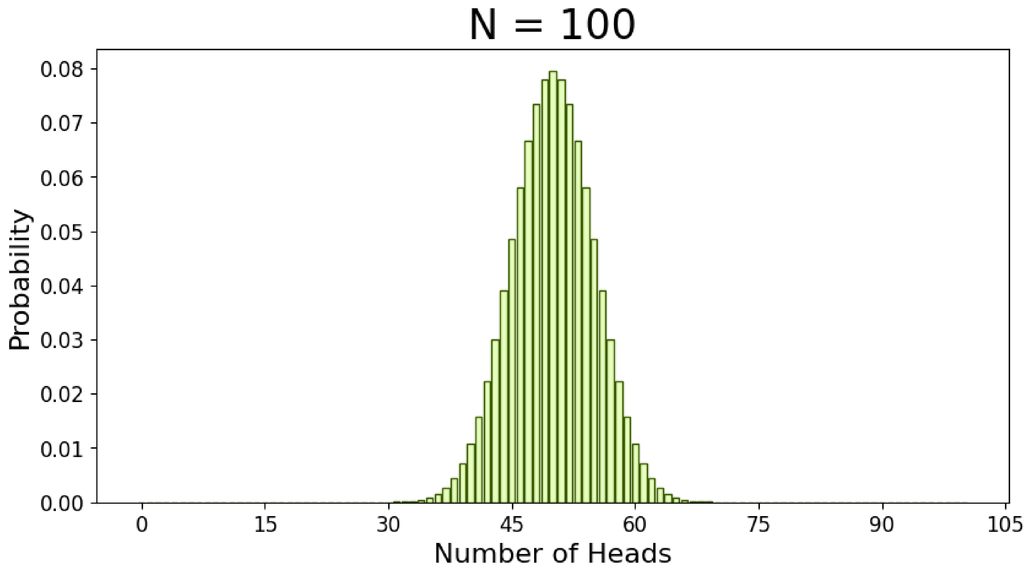
De Moivre

From Coin Flips to Bells





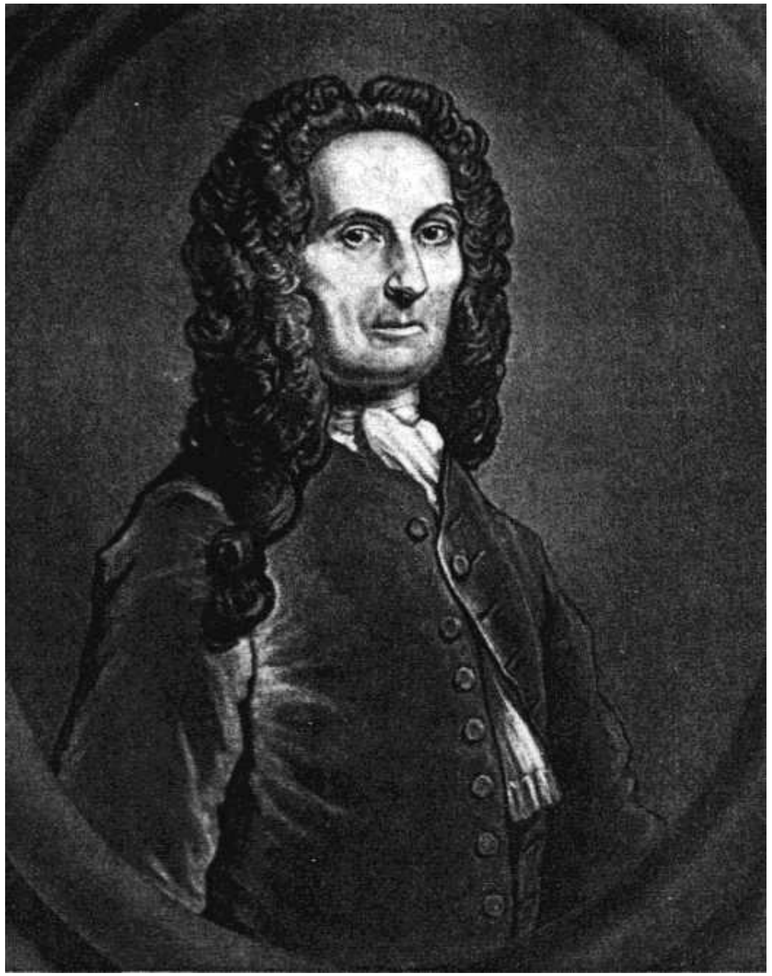




Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

De Moivre

From Coin Flips to Bells



$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\Pr(X = k) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$n! \approx \sqrt{2\pi}\left(\frac{n}{e}\right)^n$$

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

$$N(\mathrm{x};\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

A weighted coin flipping through the air like a cartoon

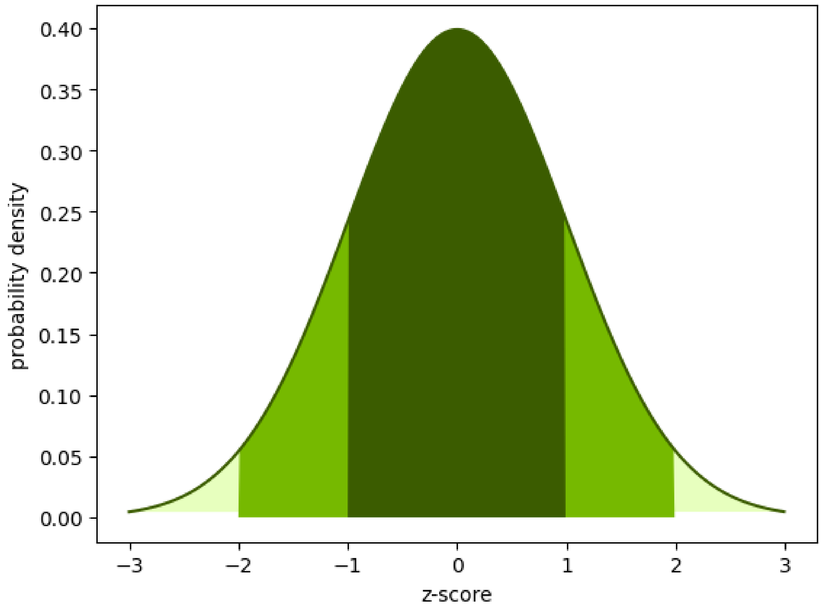
De Moivre

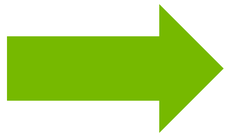
From Coin Flips to Bells

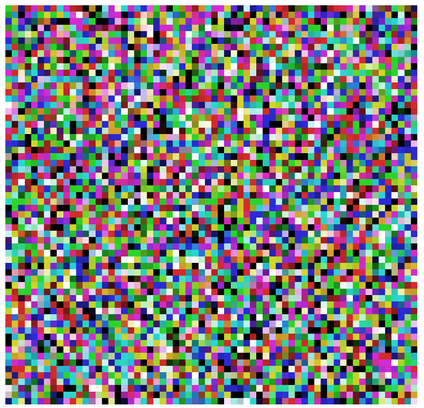
$$N(\mathrm{x};\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\mu = mean, a.k.a. average$ $\sigma = standard\ devaition, a.k.a. spread$

$$z = \frac{x\ -\mu}{\sigma}$$

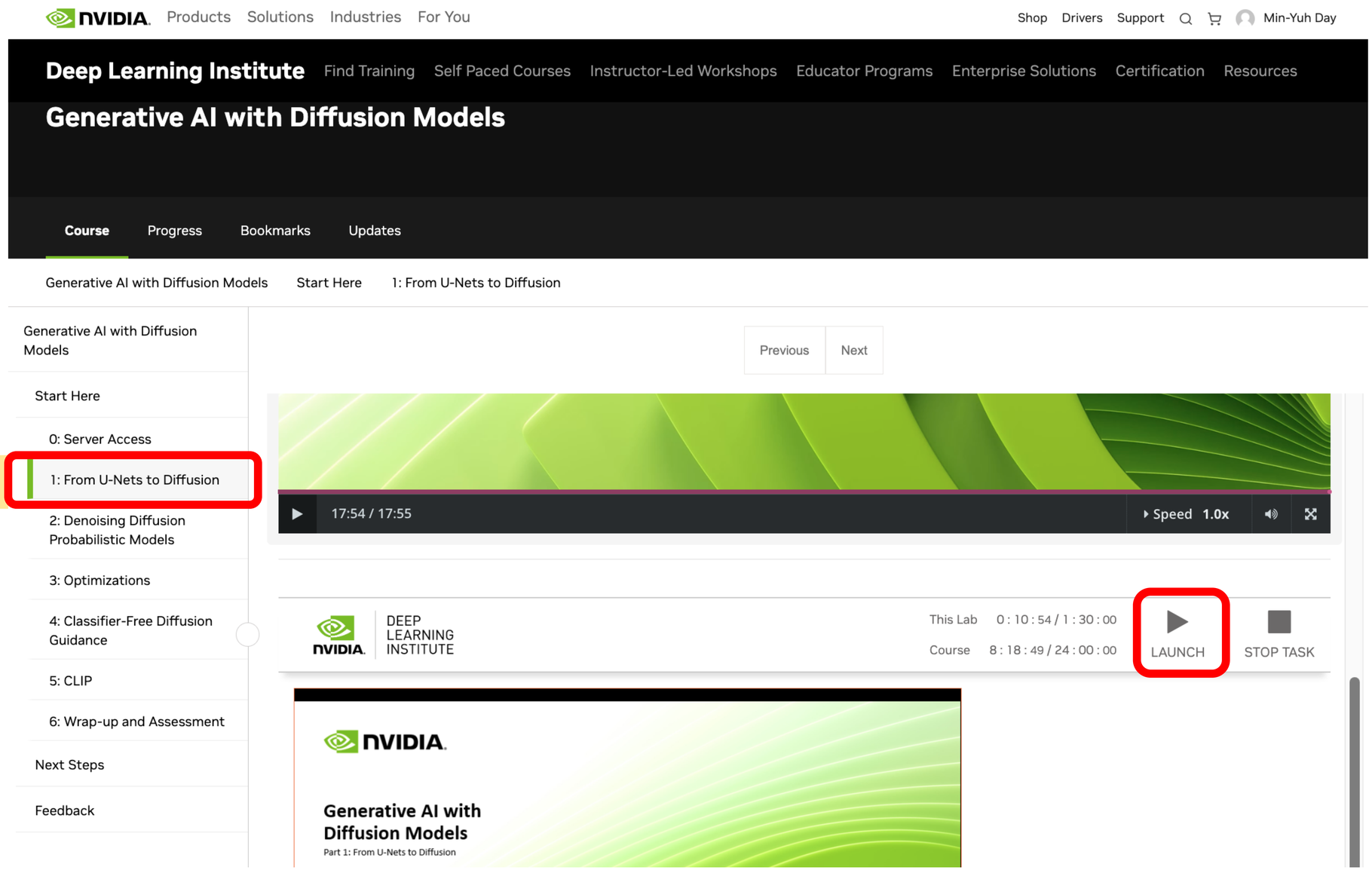






Generative AI with Diffusion Models

1: From U-Nets to Diffusion

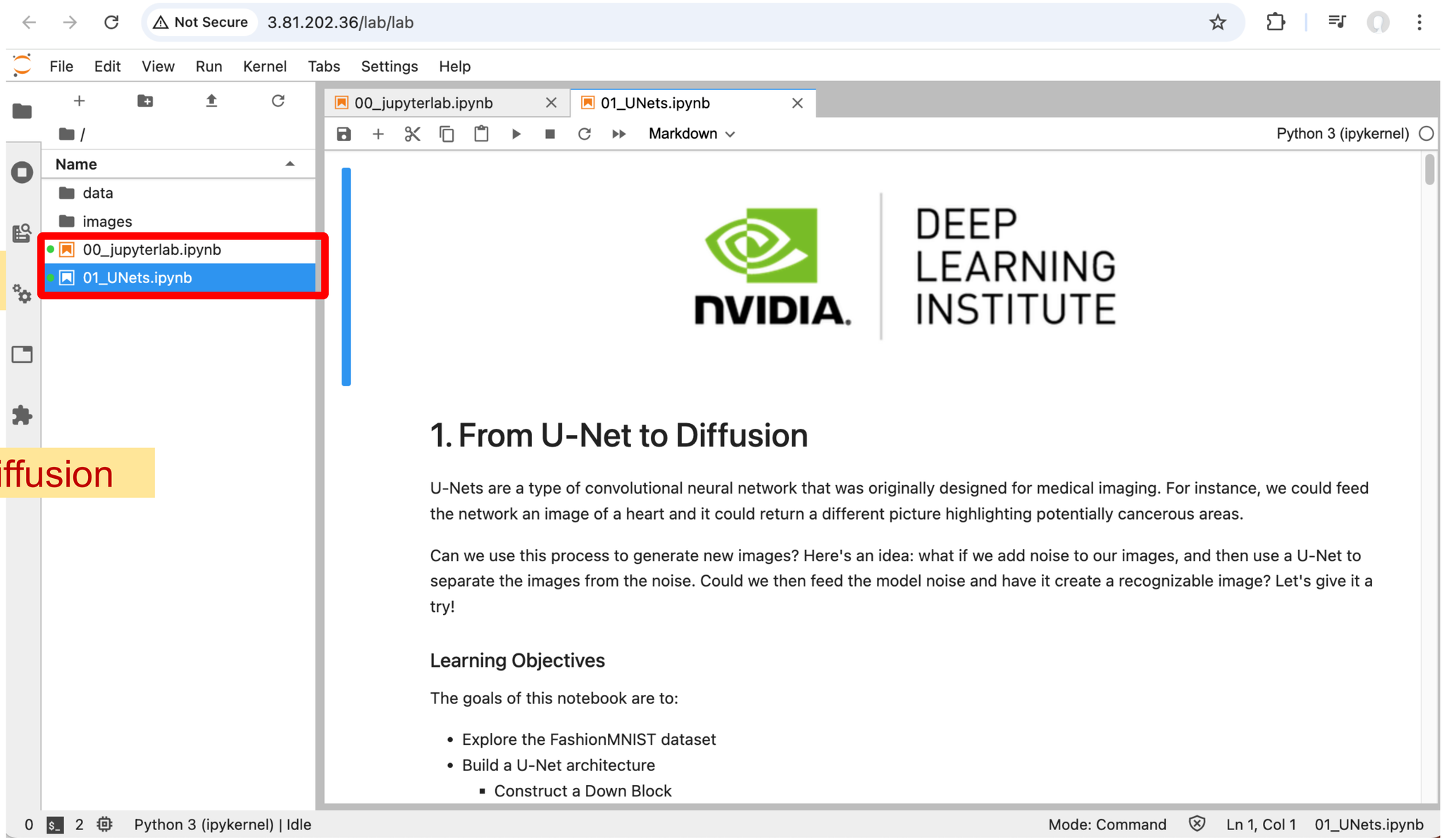


https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

01_UNets.ipynb

1: From U-Nets to Diffusion



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

Part 2: Denoising Diffusion Probabilistic Models

Agenda

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance
- Part 5: CLIP
- Part 6: Wrap-up & Assessment

Diffusion Inspiration

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Thermodynamic Origins

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein JASCHA@STANFORD.EDU
Stanford University

Eric A. Weiss EAWEISS@BERKELEY.EDU
University of California, Berkeley

Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University

Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University

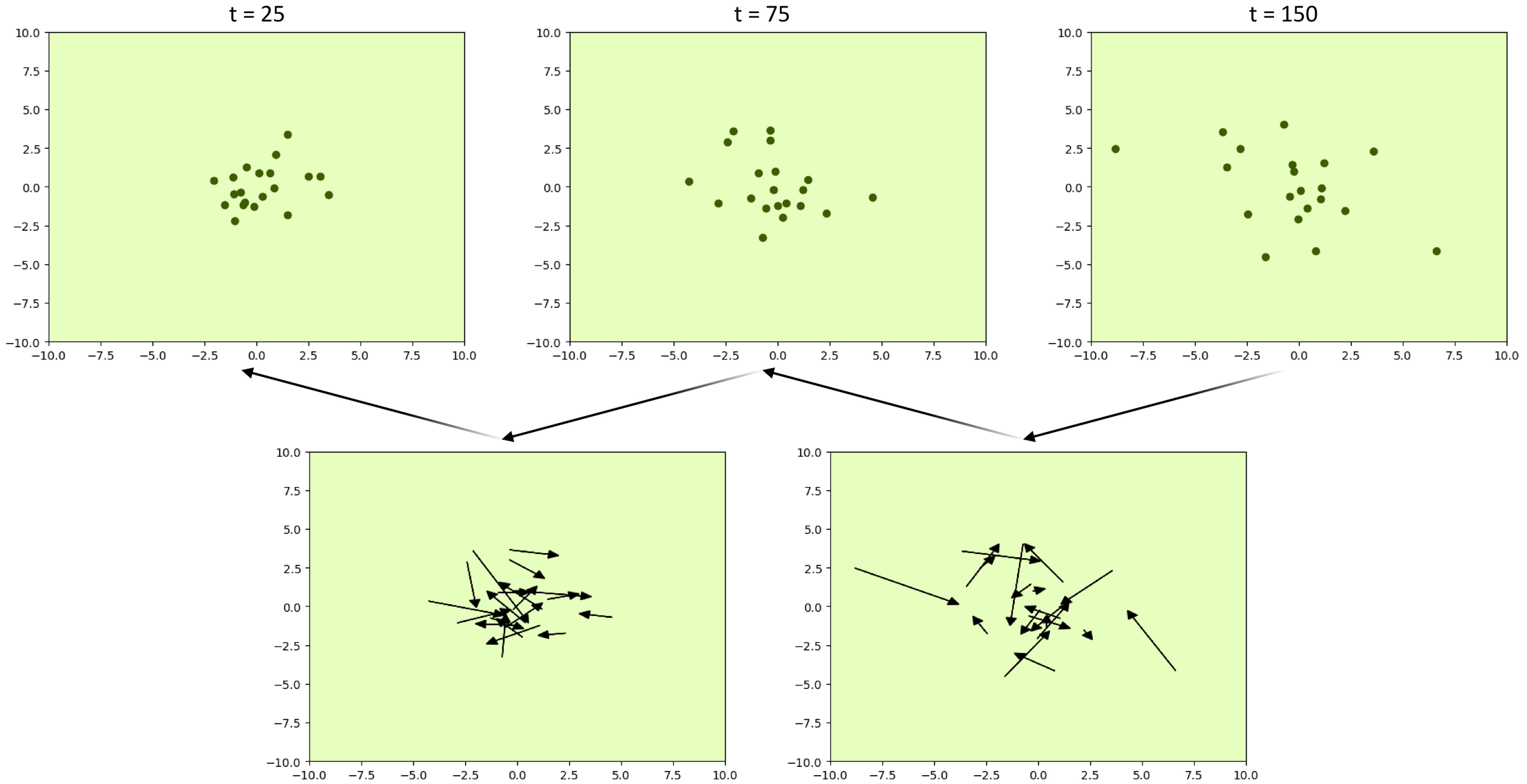
Abstract

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation

these models are unable to aptly describe structure in rich datasets. On the other hand, models that are *flexible* can be molded to fit structure in arbitrary data. For example, we can define models in terms of any (non-negative) function $\phi(\mathbf{x})$ yielding the flexible distribution $p(\mathbf{x}) = \frac{\phi(\mathbf{x})}{Z}$, where Z is a normalization constant. However, computing this

Green food coloring dissolving in a glass of water

Thermodynamic Origins



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Denoising Diffusion Probabilistic Models

Getting Lost in the Noise

Denoising Diffusion Probabilistic Models

Jonathan Ho
UC Berkeley
jonathanho@berkeley.edu

Ajay Jain
UC Berkeley
ajayj@berkeley.edu

Pieter Abbeel
UC Berkeley
pabbeel@cs.berkeley.edu

Abstract

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic models and denoising score matching with Langevin dynamics, and our models naturally admit a progressive lossy decompression scheme that can be interpreted as a

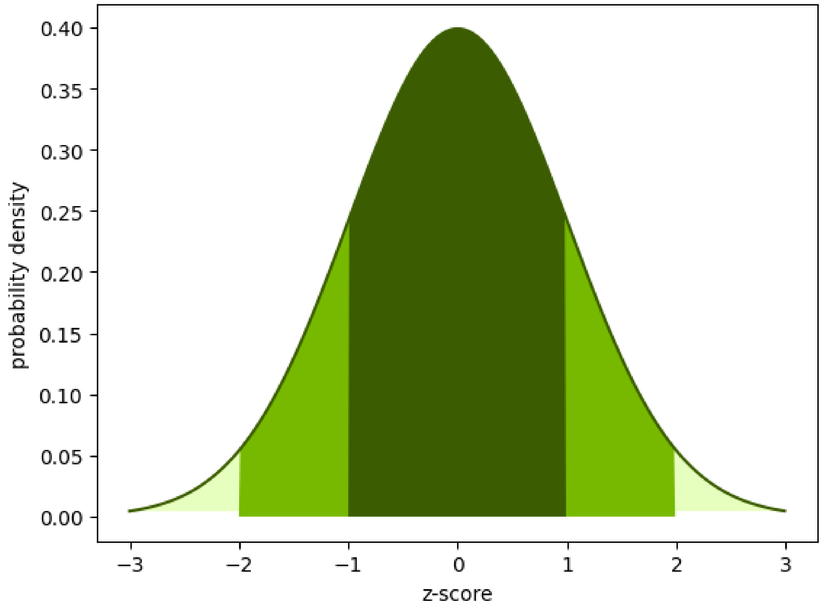
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

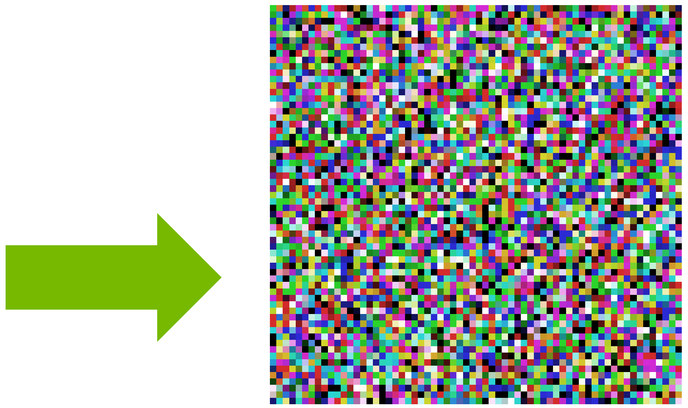
Forward Diffusion

Getting Lost in the Noise

Forward Diffusion

t = 0 t = 1 t = 2 t = 3 t = 4 t = 5 t = 6 t = 7 t = 8 t = 9 t = 10



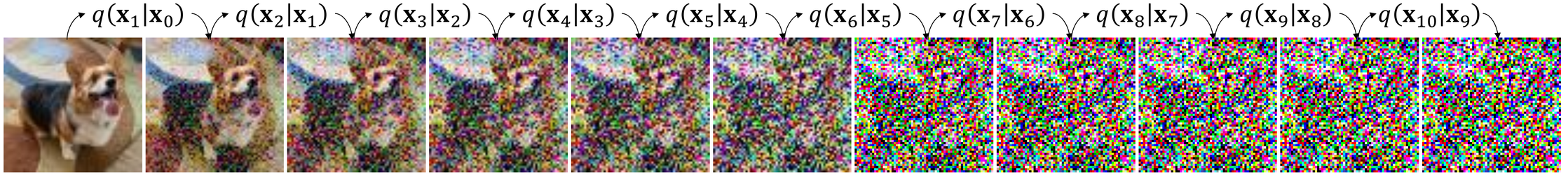


$$N(\mathrm{x};\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Getting Lost in the Noise

Forward Diffusion



$$\beta = [0.0001, 0.0023, 0.0045, 0.0067, \ldots, 0.0200]$$

t = 0 t = 1 t = 3 t = 4 t = 10

$$T = 10$$

Total number of timesteps

$$\alpha = 1 - \beta$$

Convenient shorthand

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Getting Lost in the Noise

Forward Diffusion

t = 0 t = 1 t = 2 t = 3 t = 4 t = 5 t = 6 t = 7 t = 8 t = 9 t = 10

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = N(\mathbf{x}_t; \sqrt{1-\beta_t} \cdot \mathbf{x}_{t-1}, \beta_t \cdot \mathbf{I})$$

Code:

```
noise = torch.randn_like(x_t)
x_t = torch.sqrt(1 - B[t]) * x_t
  + torch.sqrt(B[t]) * noise
```

Step 1:

- Generate noise from a standard normal distribution
- Multiply result by $\sqrt{\beta_t}$

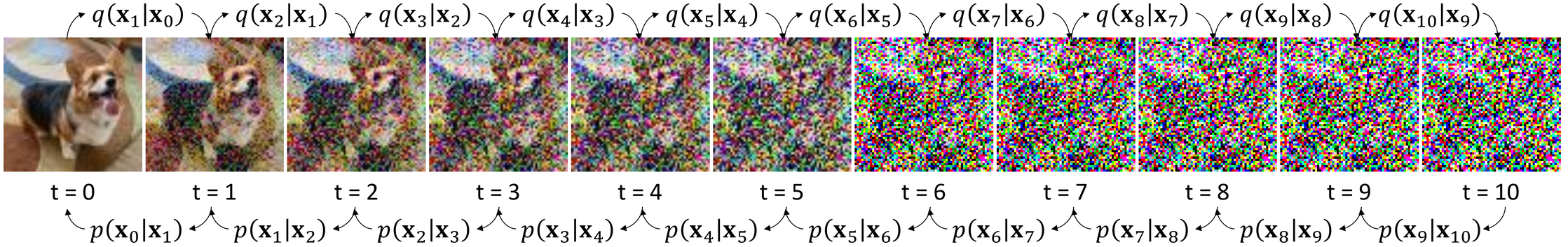
Step 2:

- Multiply image at previous step by $\sqrt{1-\beta_t}$
- Add it to the result from step 1

Reverse Diffusion

Getting Lost in the Noise

Reverse Diffusion



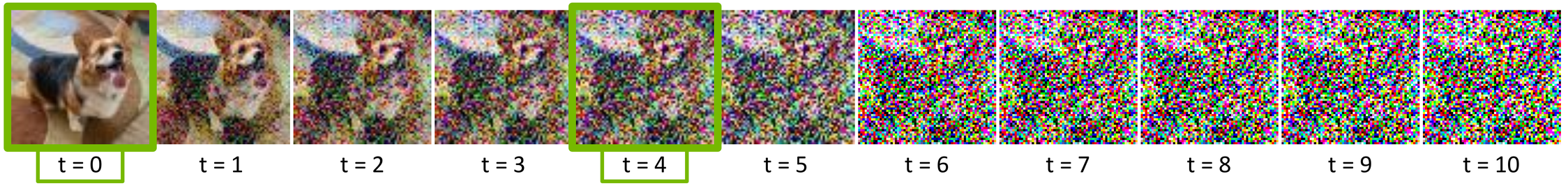
$$p(\mathbf{x}_{t-1}|\mathbf{x}_t) = N(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$$

Difficult to calculate

Approximate average used to create $\mathbf{x}_t$

Getting Lost in the Noise

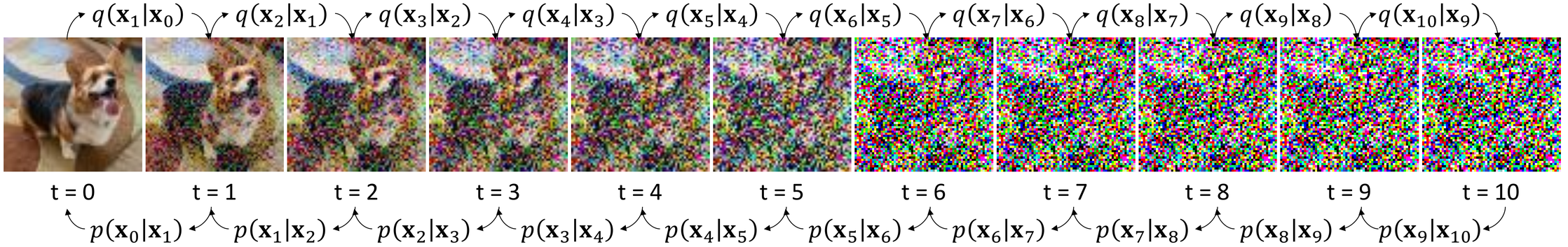
Reverse Diffusion



Can generate “evidence” of p using q

Getting Lost in the Noise

Reverse Diffusion



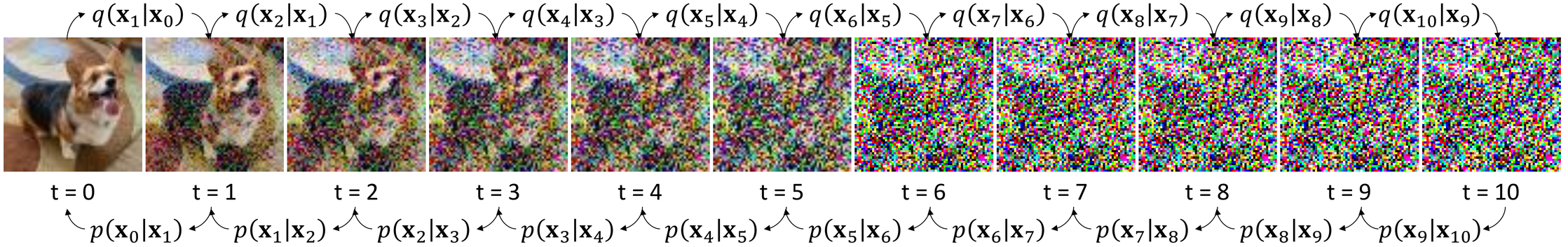
$$p(\mathbf{x}_{t-1}|\mathbf{x}_t) = N(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = N(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \cdot \mathbf{I}) \qquad \tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t$$

Use Bayes' Rule and …

Getting Lost in the Noise

Reverse Diffusion



$$p(\mathbf{x}_{t-1}|\mathbf{x}_t) = N(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = N(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \cdot \mathbf{I})$$

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}}\epsilon_t\right)$$

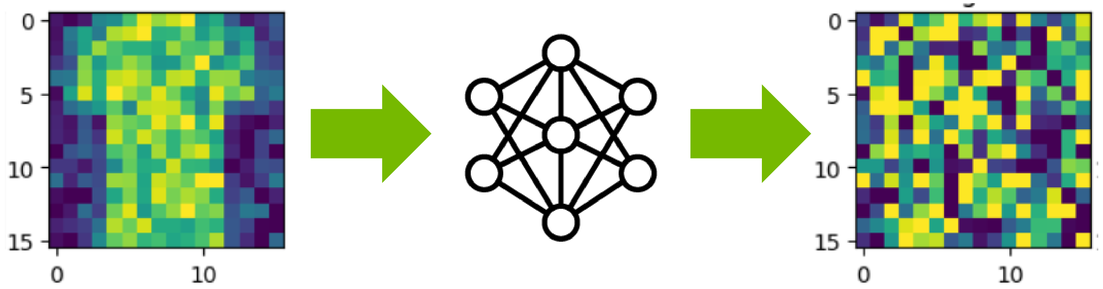
The noise added at time t
This is what the neural network will estimate

Model Training

Model Training

Training Data

Noisy image at time t

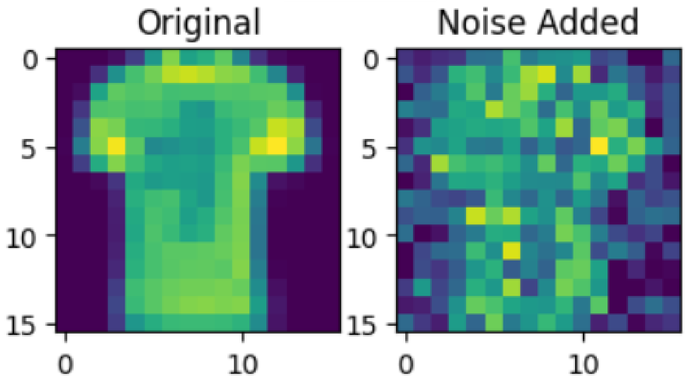


Noise added at time t

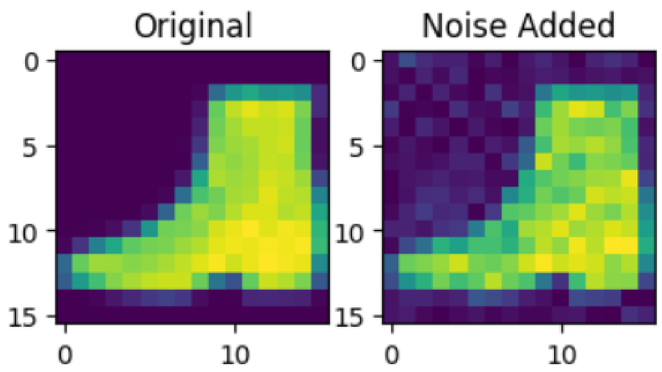
$$q(\mathbf{x}_t|\mathbf{x}_0) = N(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \cdot \mathbf{x}_0, (1 - \bar{\alpha}_t) \cdot \mathbf{I})$$

"Skip ahead" function

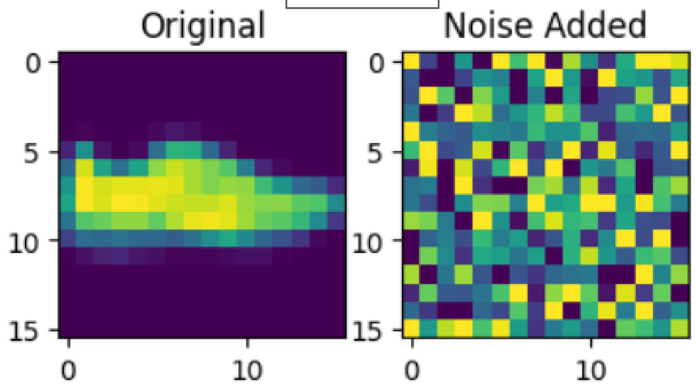
$t = 5$



$t = 25$



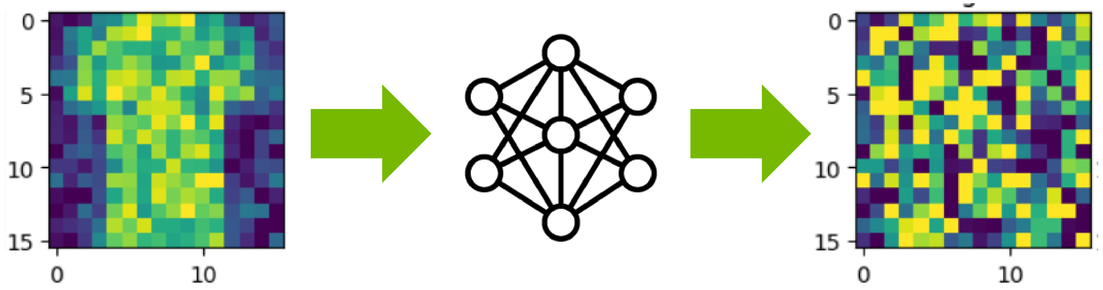
$t = 100$



Model Training

ELBO Loss

Noisy image at time t



Noise added at time t

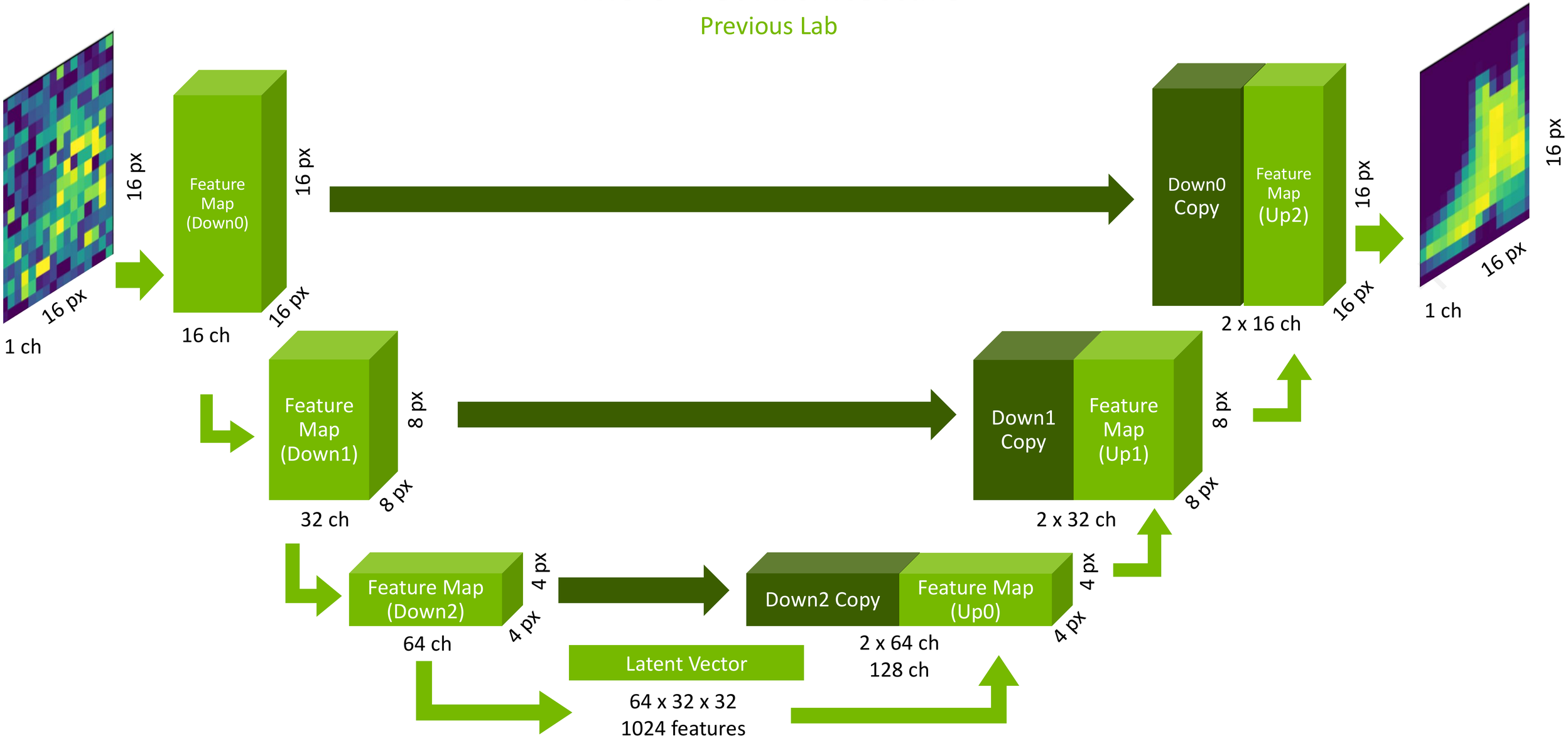
$$ELBO \dots$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Model Architecture

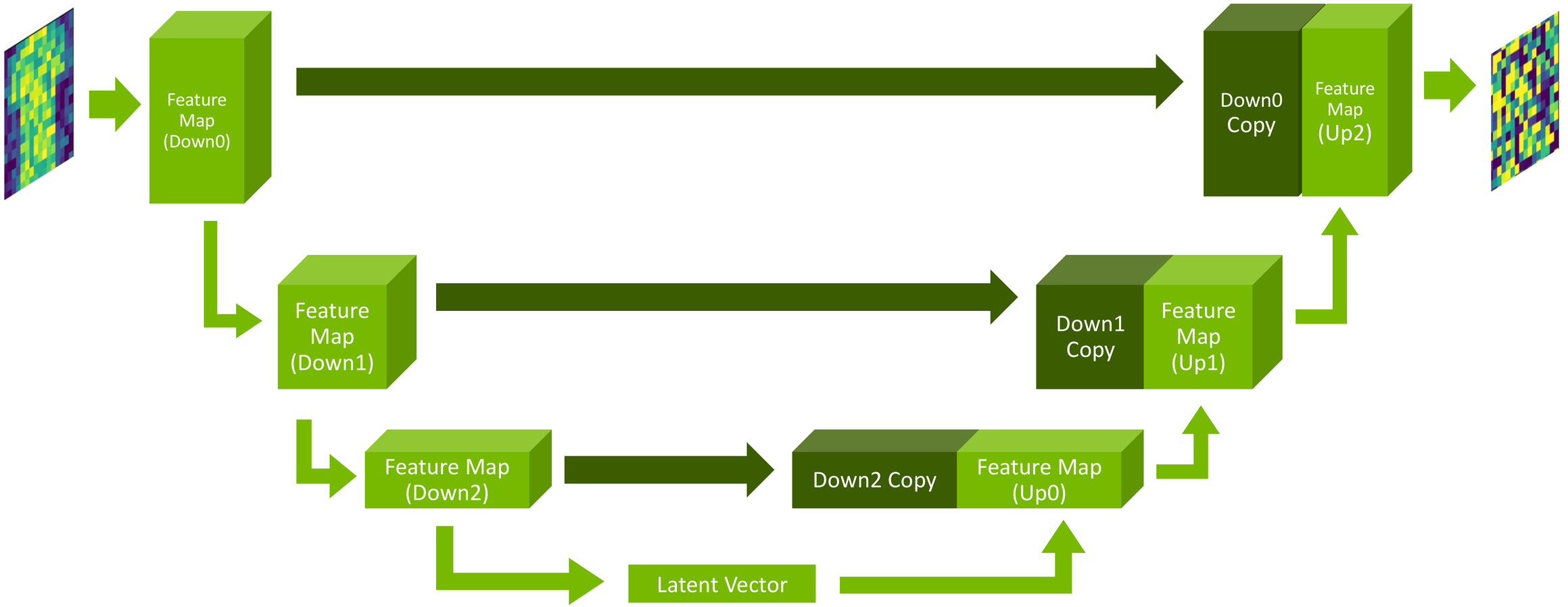
The U-Net Architecture

Previous Lab



The U-Net Architecture

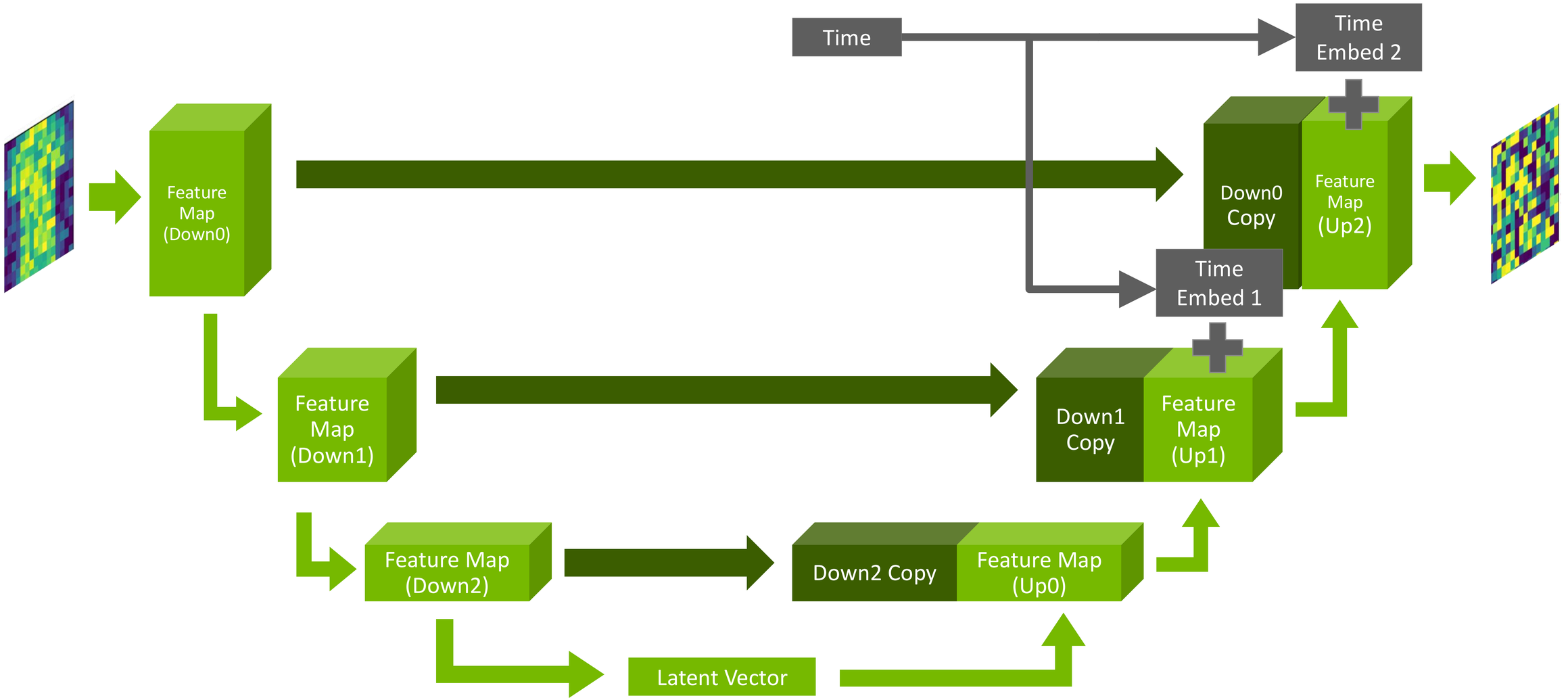
Next Lab



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

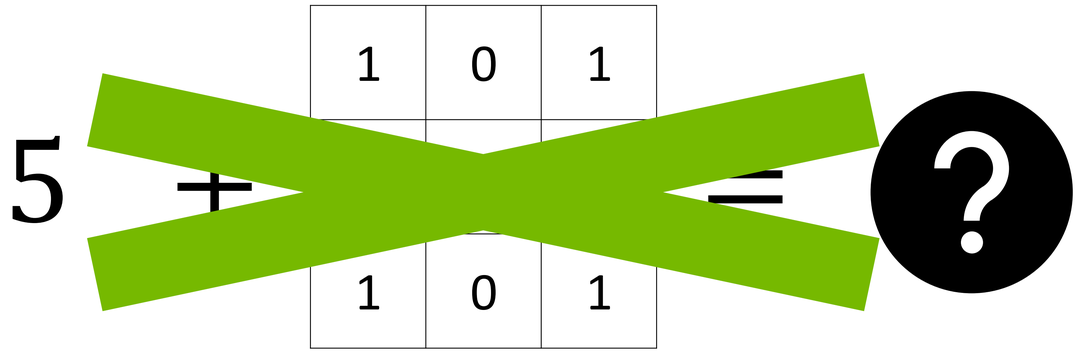
It's About Time

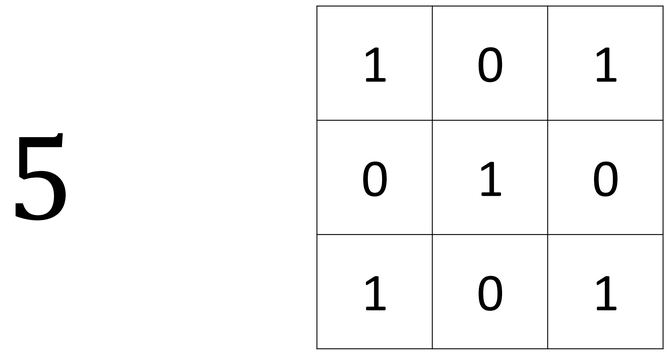
Adding Time



It’s About Time

Broadcasting





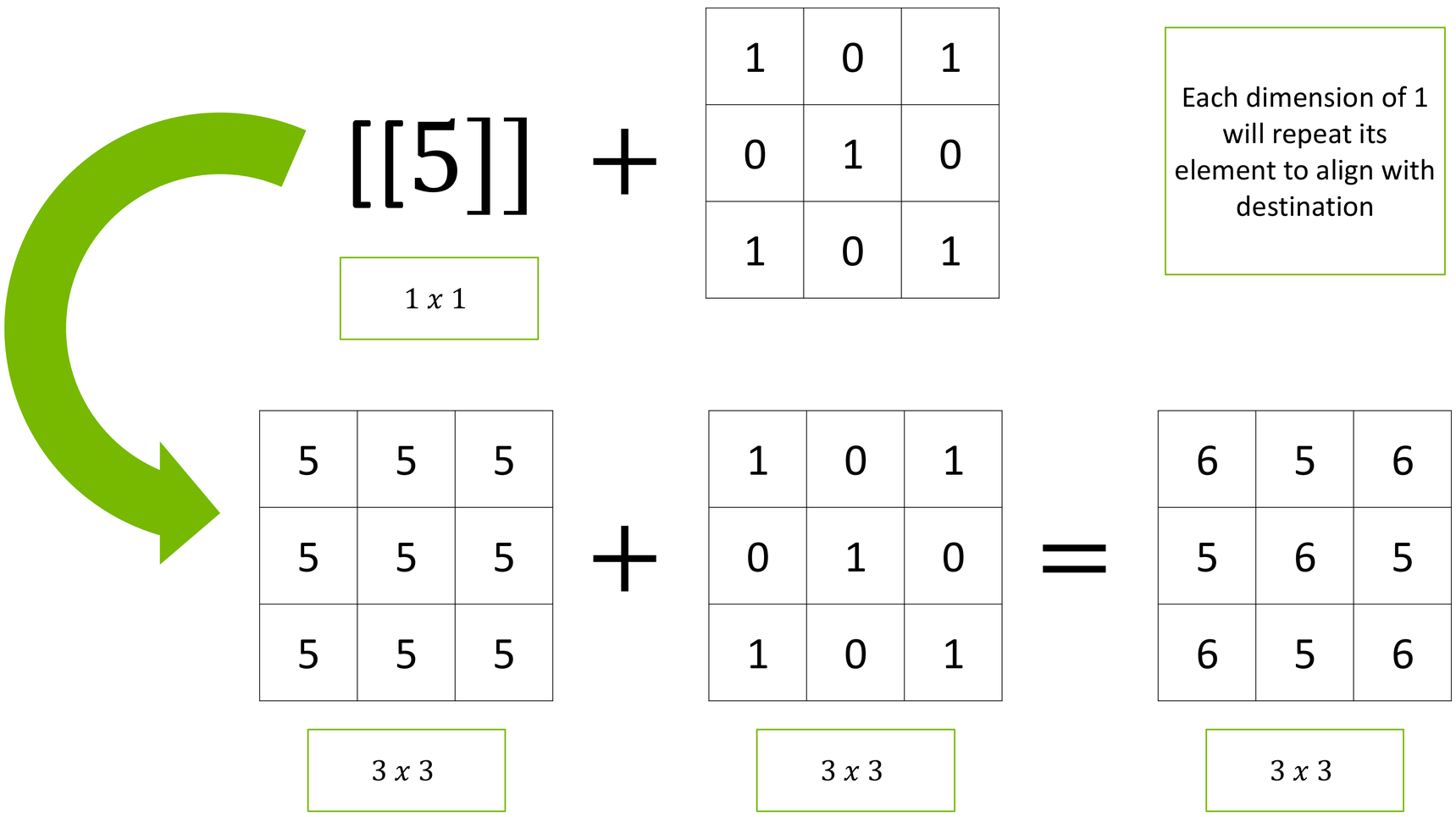
0 or 1

$3 \; x \; 3$

Dimensions

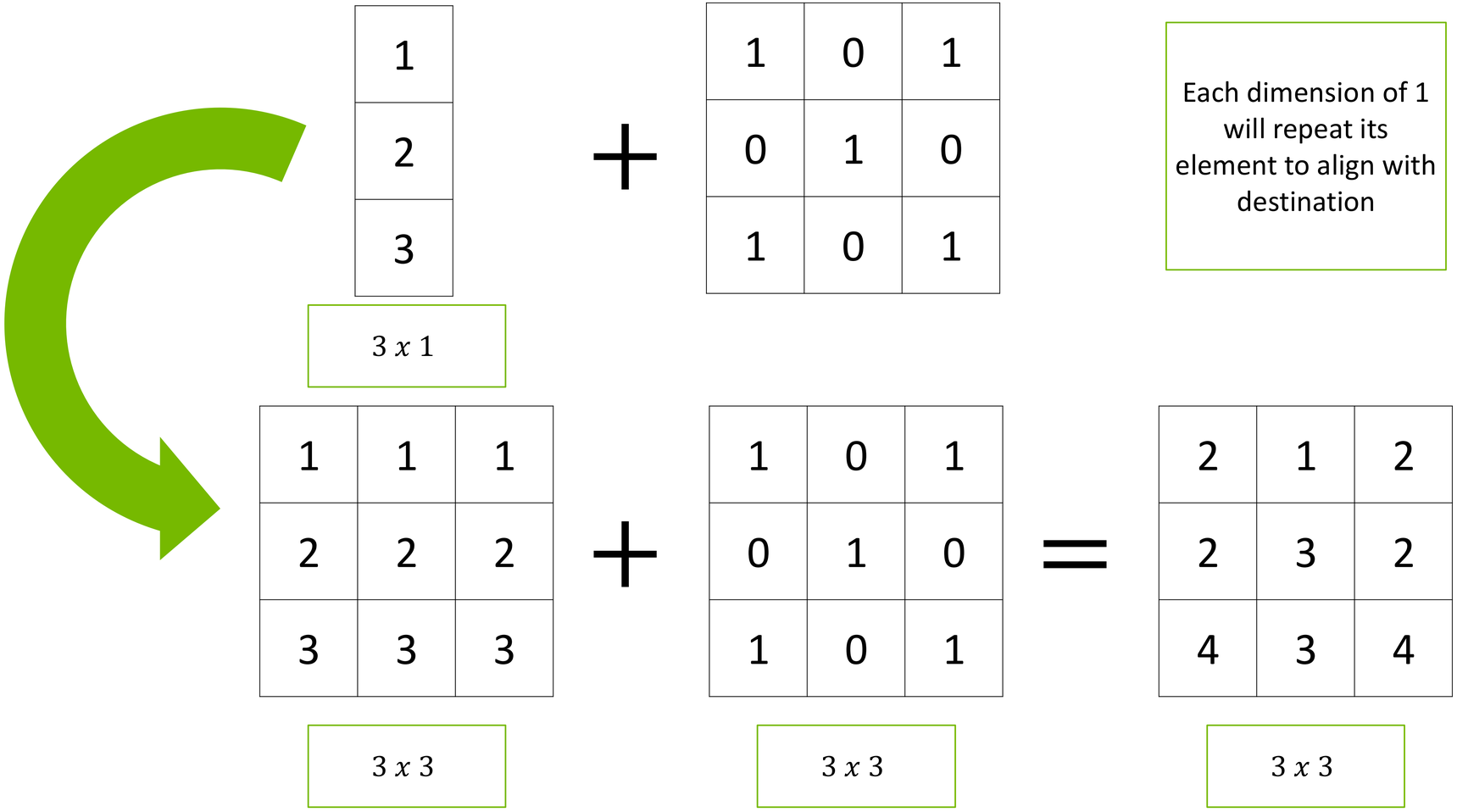
It’s About Time

Broadcasting



It’s About Time

Broadcasting



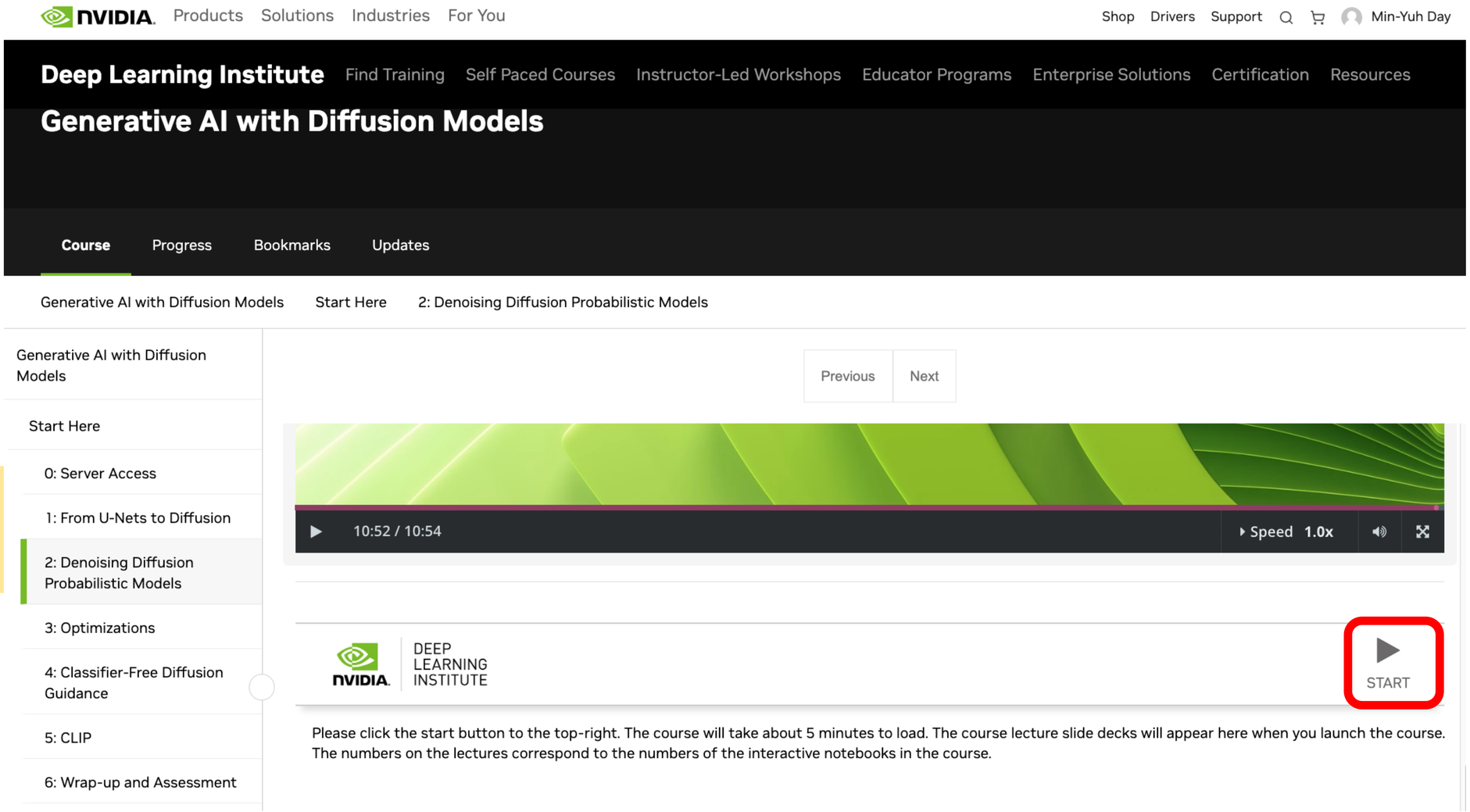
Let’s get started!

2: Denoising Diffusion Probabilistic Models

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

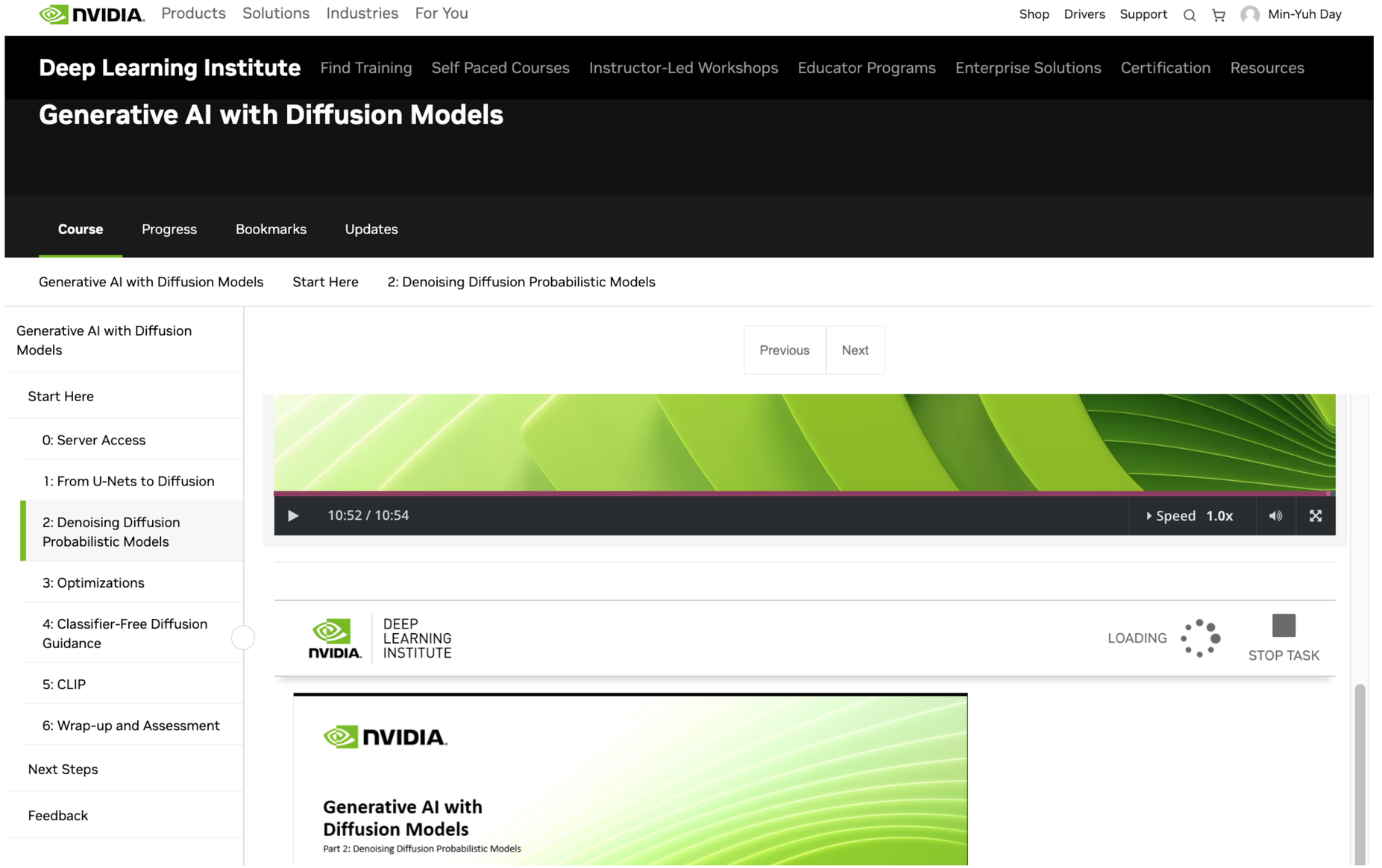
2: Denoising Diffusion Probabilistic Models



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

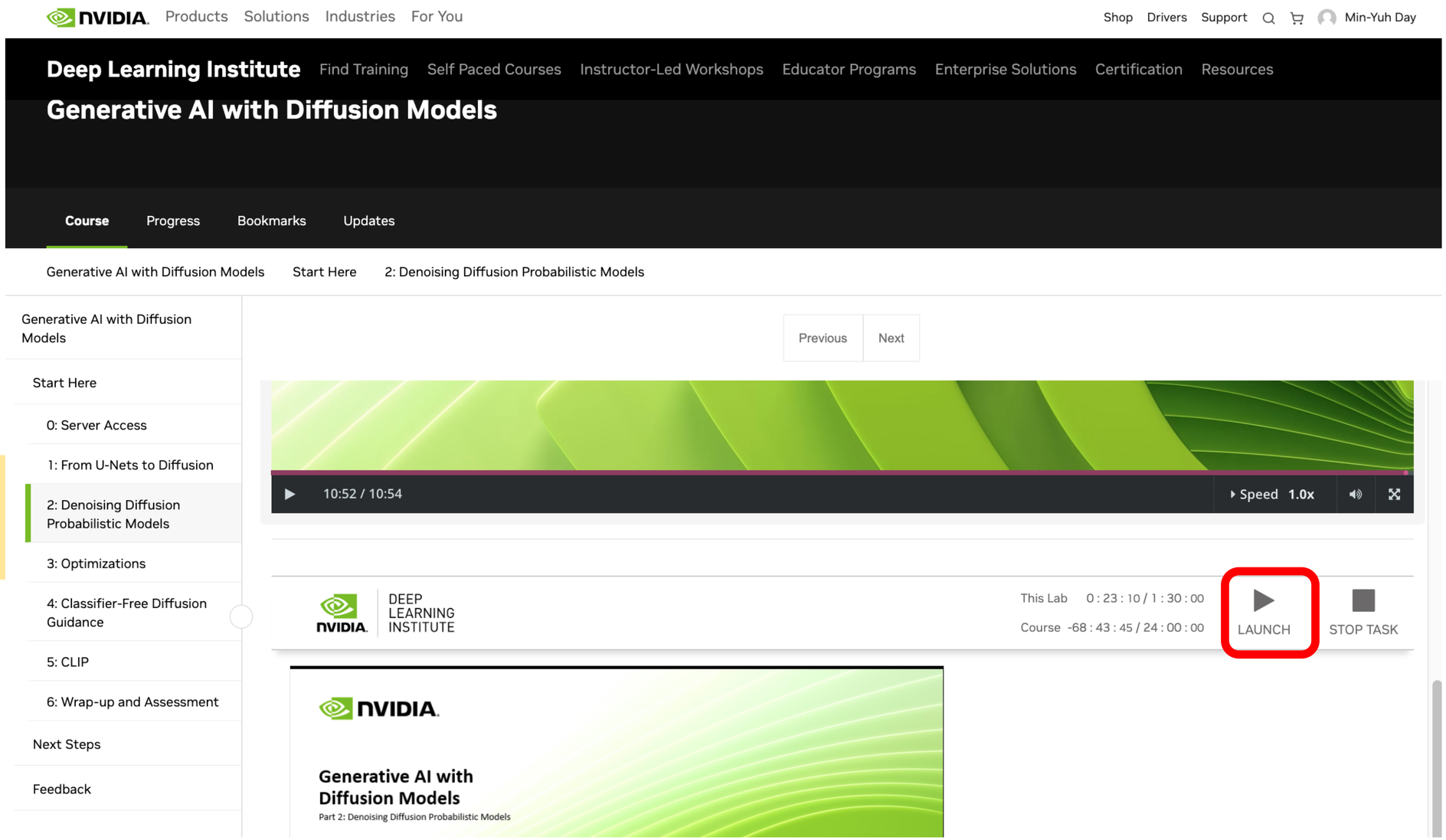
2: Denoising Diffusion Probabilistic Models



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

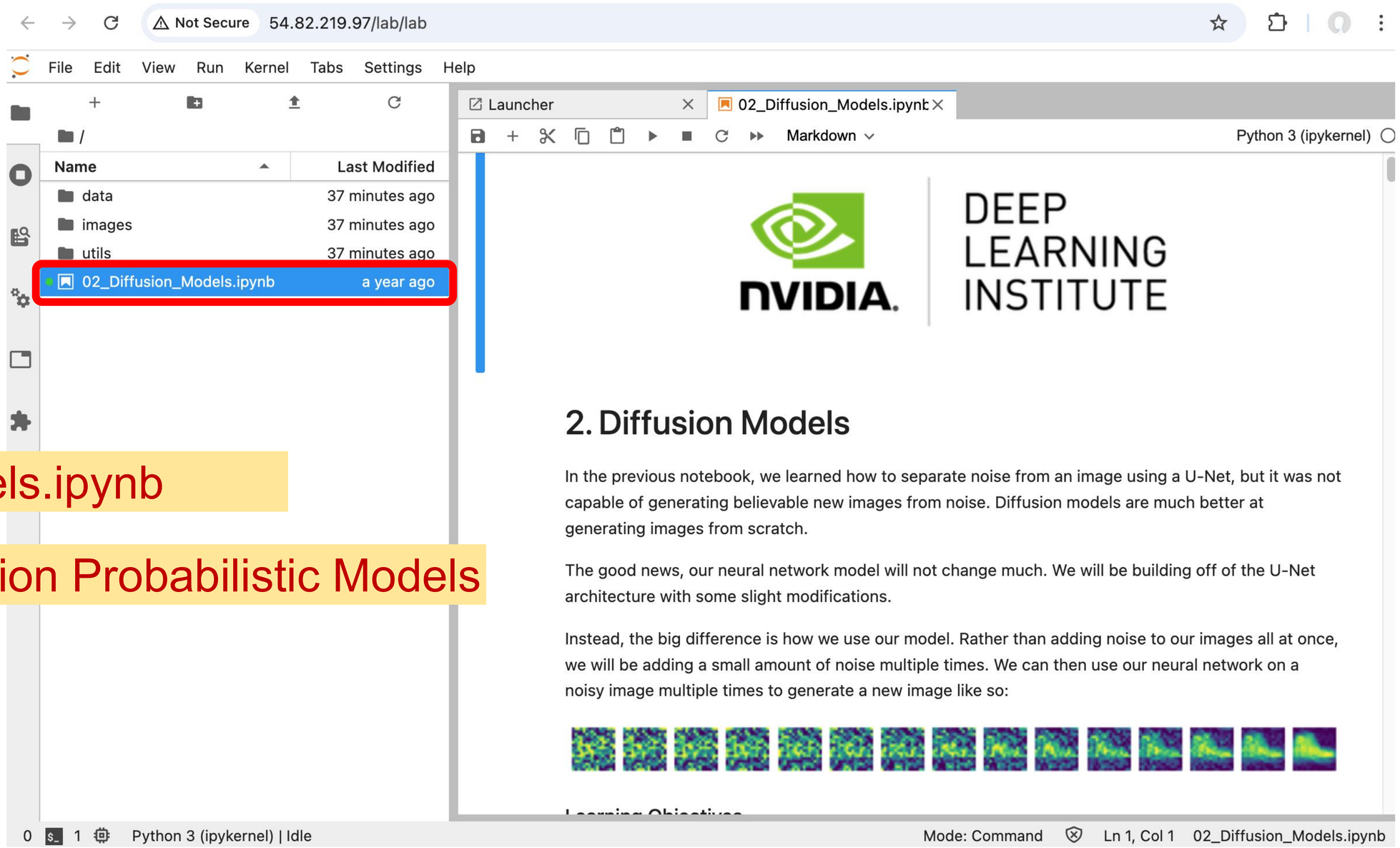
Generative AI with Diffusion Models

2: Denoising Diffusion Probabilistic Models



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models



02_Diffusion_Models.ipynb

2: Denoising Diffusion Probabilistic Models

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA
Generative AI with Diffusion Models
Part II

[0,1,2] [3, 4] [5, 6]

Generative AI with Diffusion Models

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance (CFG)
- Part 5: CLIP (Contrastive Language-Image Pre-Training)
- Part 6: Wrap-up & Assessment

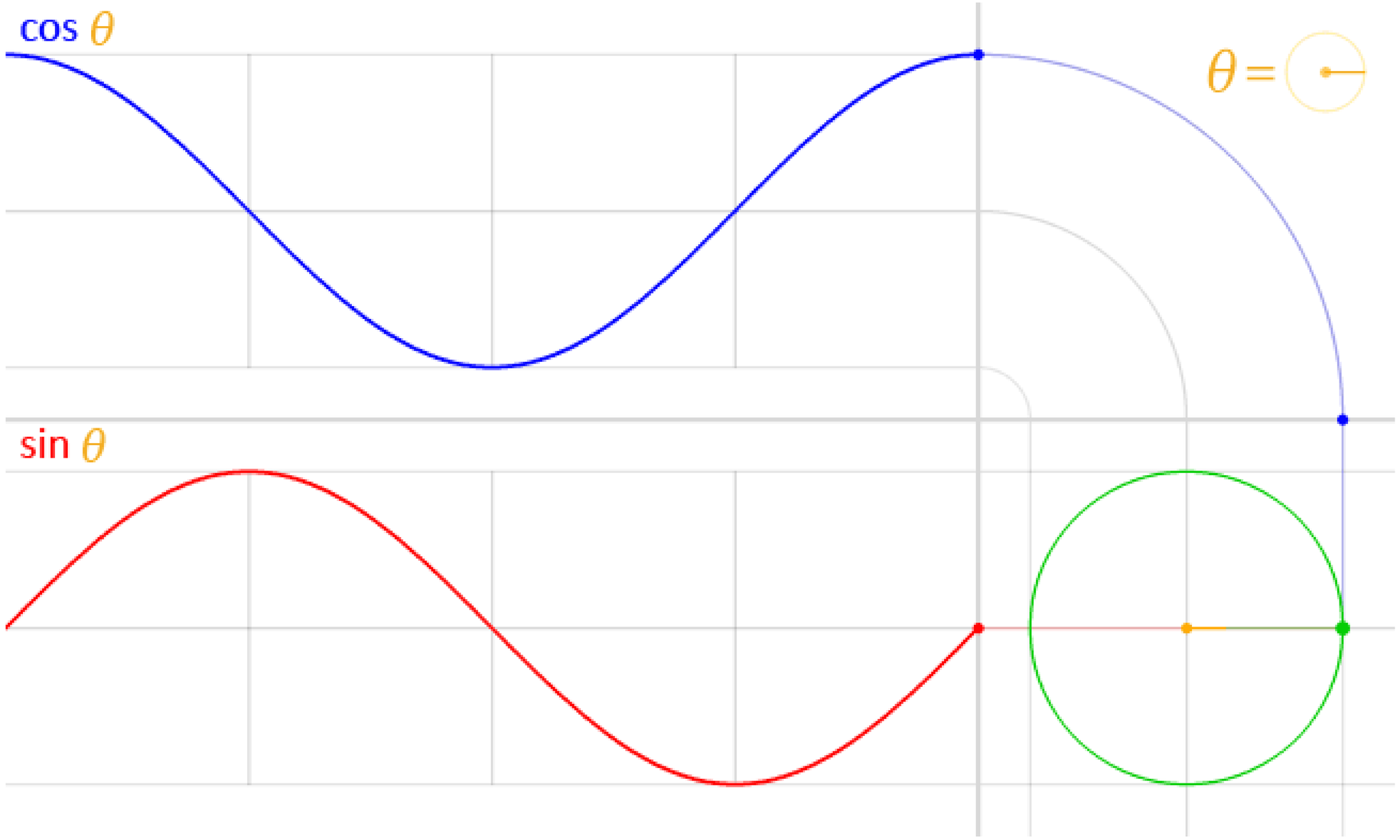
https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models
Part 3: Optimizations

Technique	Key Benefit
Group Normalization	Stable training
Gaussian Error Linear Unit (GELU) Activation	Smoother learning
Rearrange Pooling	Smooth upsampling
Sinusoidal Position Embeddings (Sine Wave)	Better timestep understanding
Deeper Networks	Finer image generation

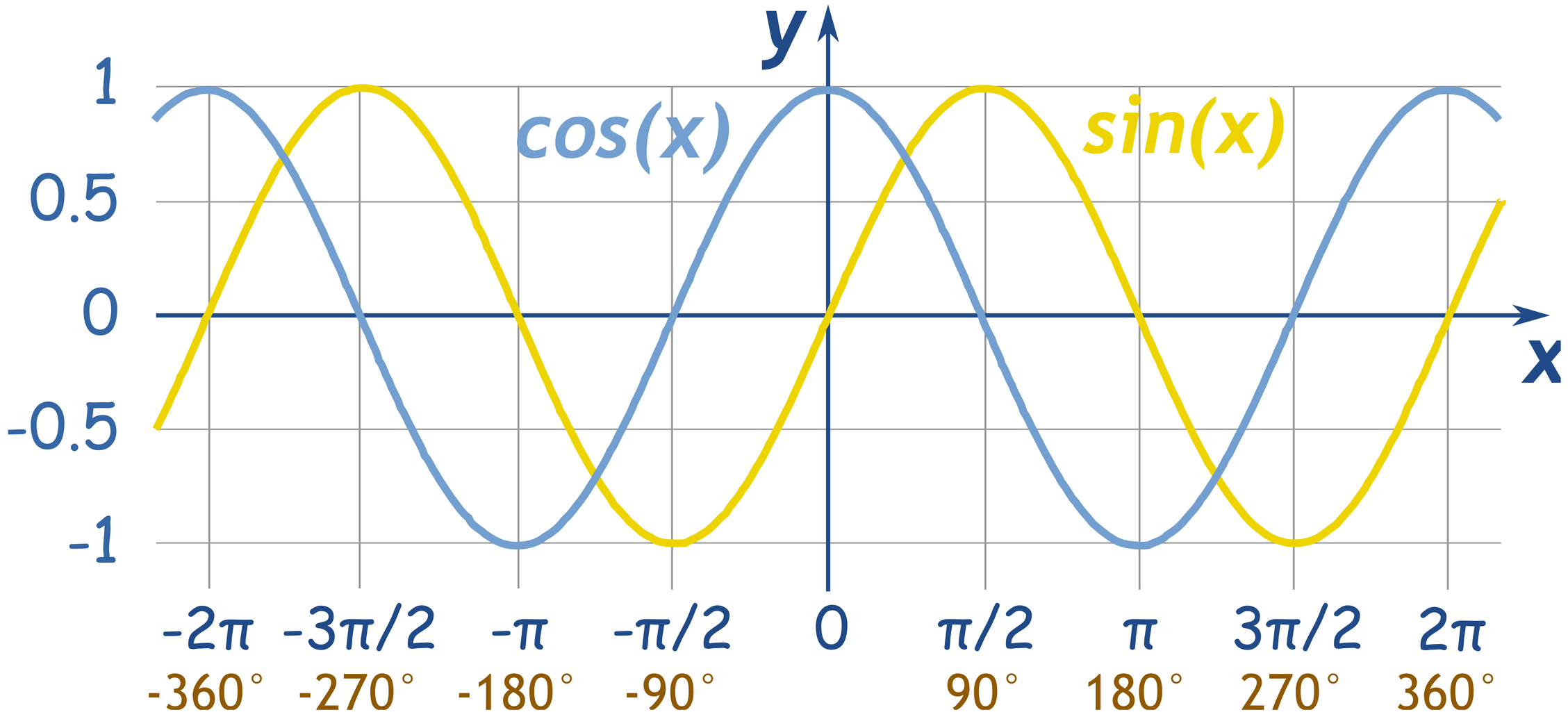
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Sinusoidal Position Embeddings for Diffusion Model
Sine and Cosine Waves to Represent Time as a Sequence



Source: https://en.m.wikipedia.org/wiki/File:Circle_cos_sin.gif

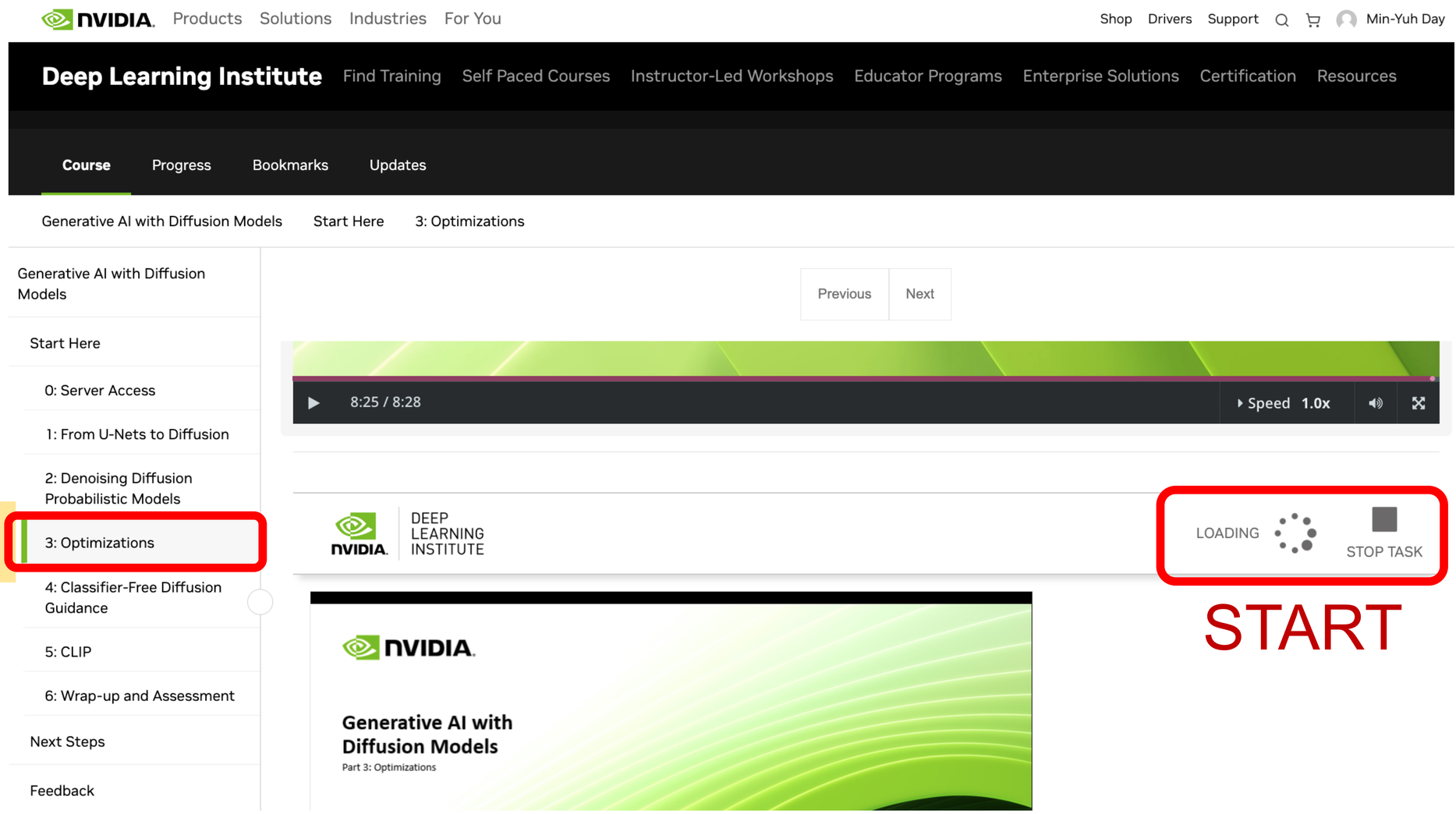
Sinusoidal Position Embeddings for Diffusion Model
Sine and Cosine Waves to Represent Time as a Sequence



Source: https://www.mathsisfun.com/algebra/trig-sin-cos-tan-graphs.html

Generative AI with Diffusion Models

3: Optimizations



START

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

Part 3: Optimizations

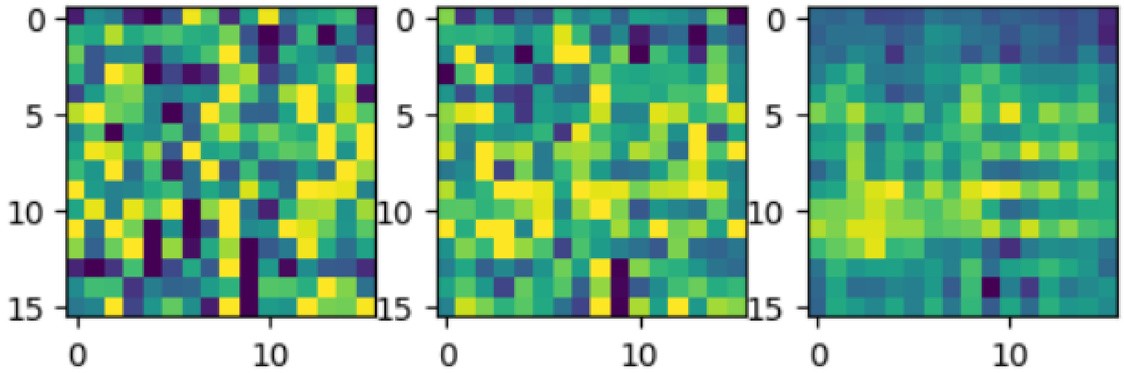
Agenda

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance
- Part 5: CLIP
- Part 6: Wrap-up & Assessment

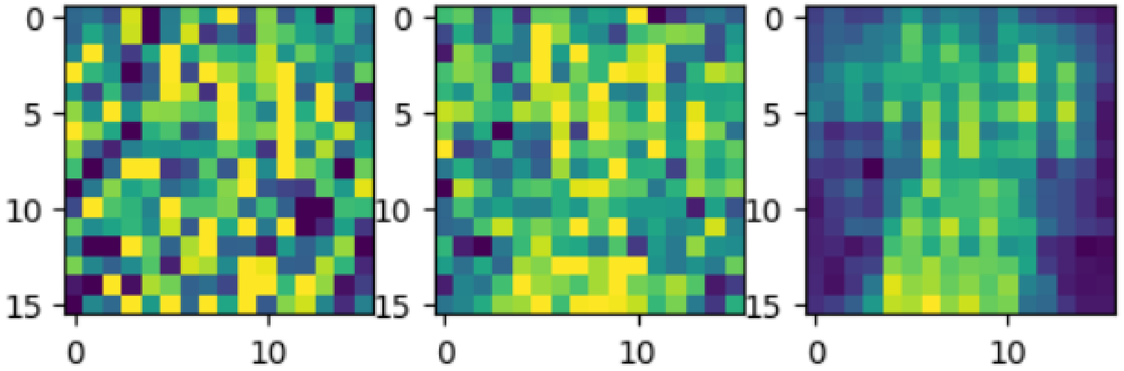
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

The Checkerboard Problem

The Checkerboard Problem



Shoe?

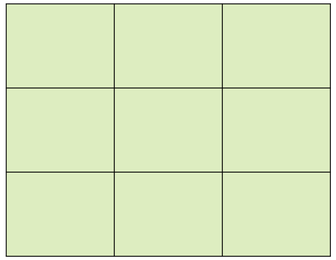


Shirt?

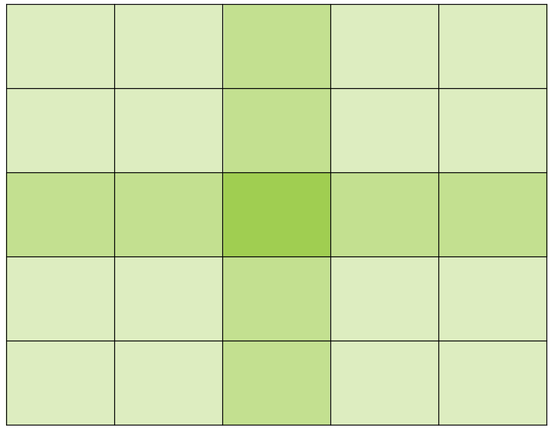
The Checkerboard Problem

The Impact of Stride

Kernel Size = 3



Stride = 2

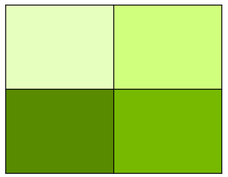


A cozy green plaid blanket, fairy tale drawing

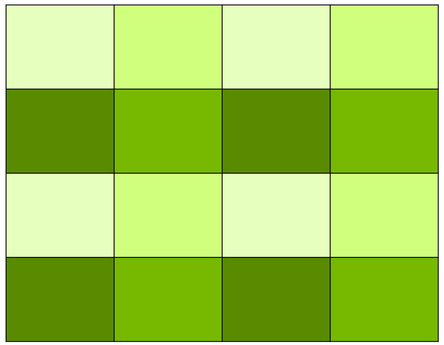
The Checkerboard Problem

The Impact of Stride

Kernel Size = 2



Stride = 2

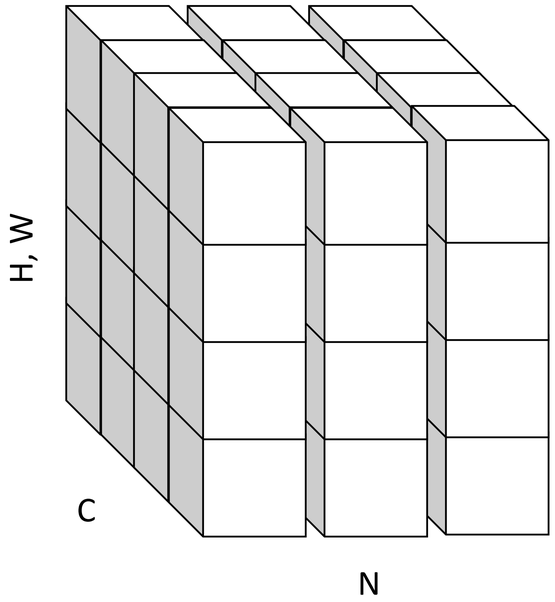


Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Group Normalization

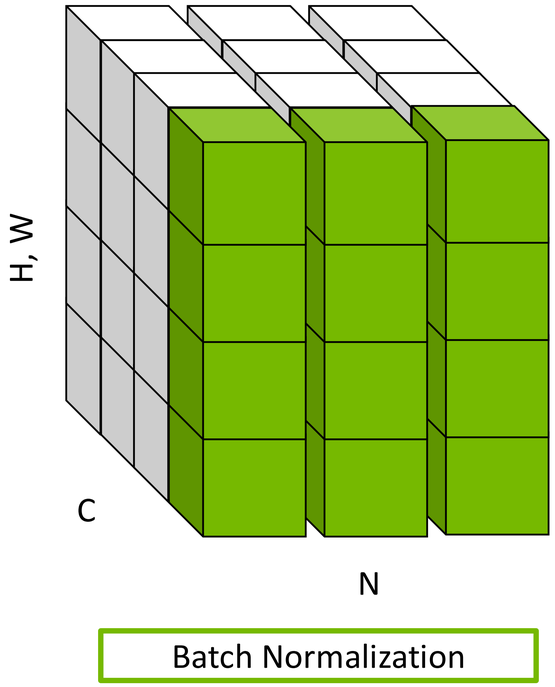
Group Normalization

Batch Normalization Review



Group Normalization

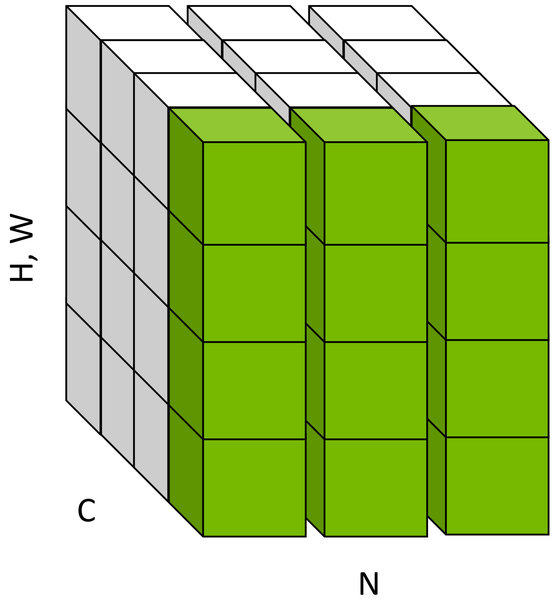
Batch Normalization Review



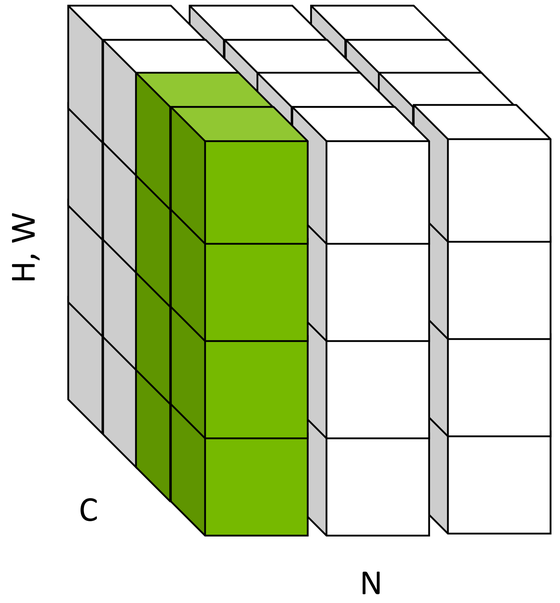
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Group Normalization

Types of Group Normalization



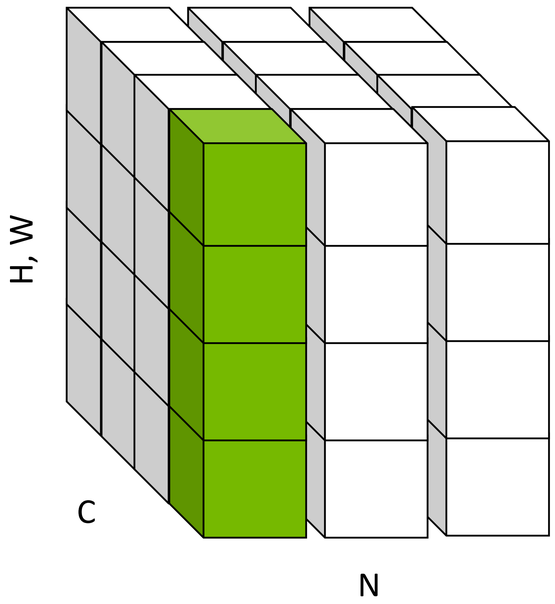
Batch Normalization



Group Normalization

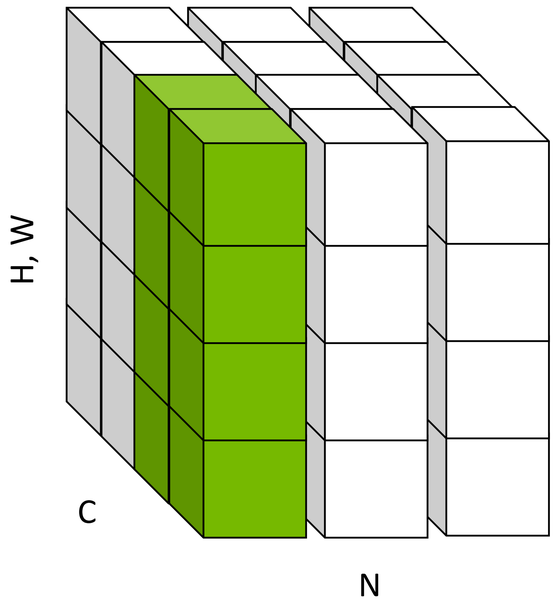
Group Normalization

Types of Group Normalization



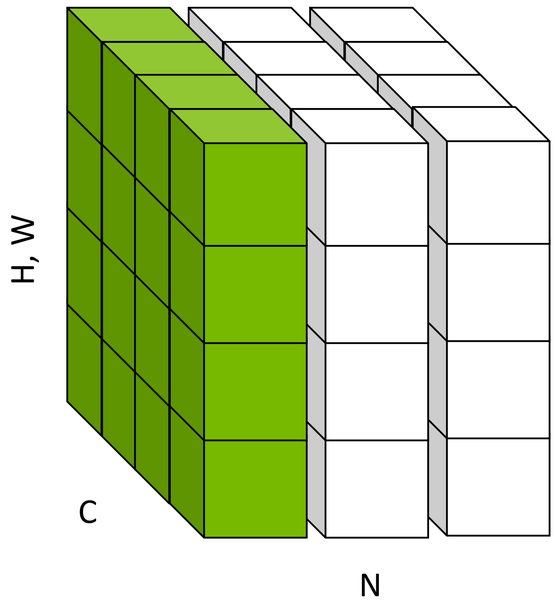
Instance Normalization

Group Size = 1



Group Normalization

Group Size = 2



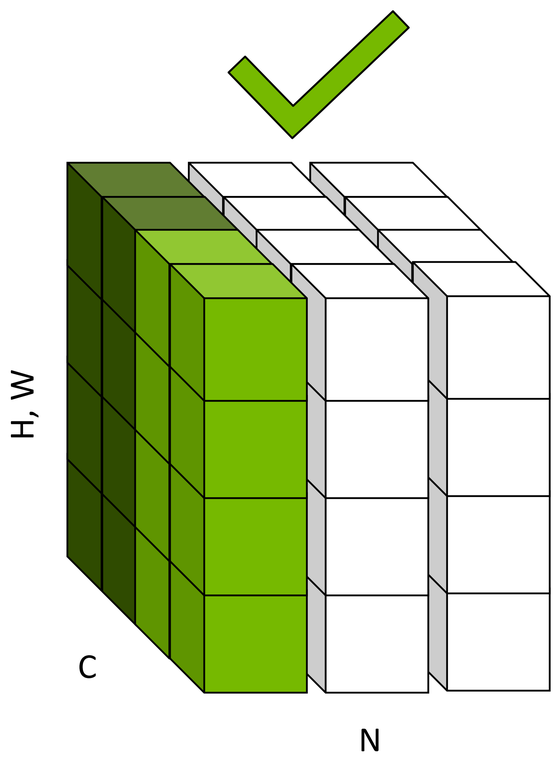
Layer Normalization

Group Size = All

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

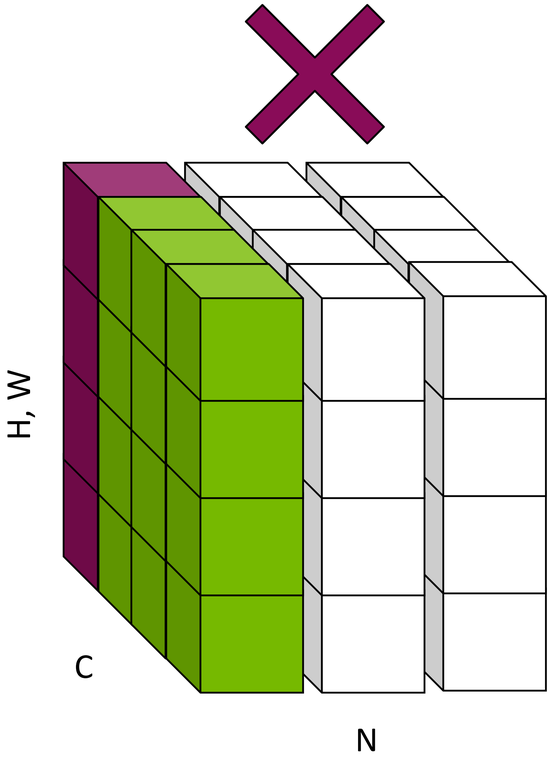
Group Normalization

Types of Group Normalization



Group Normalization

Group Size = 2



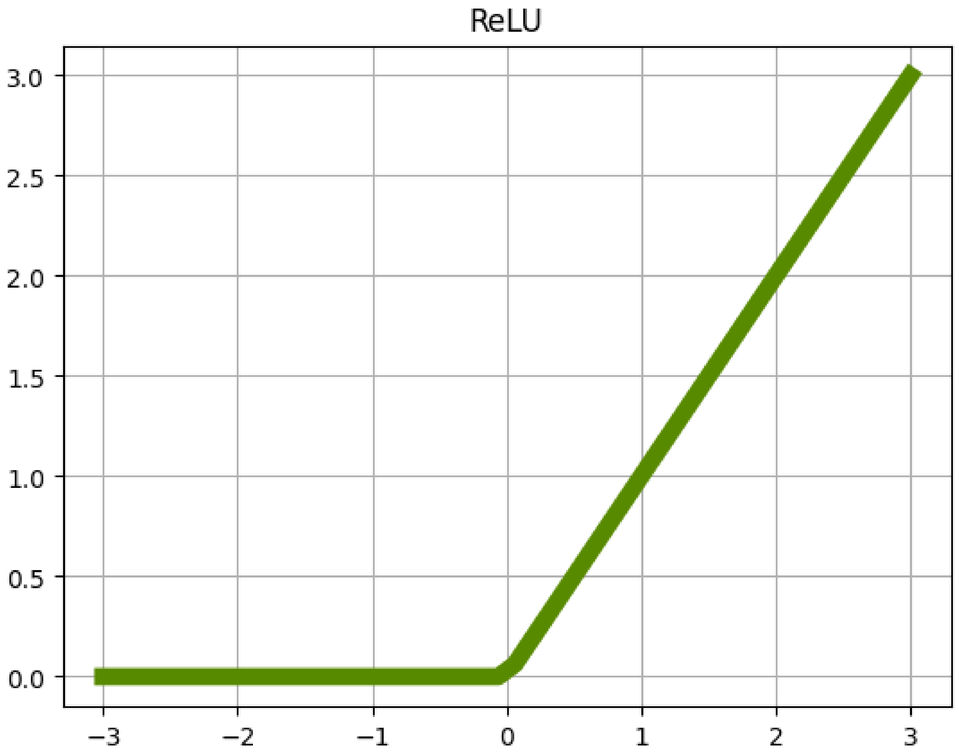
Group Normalization

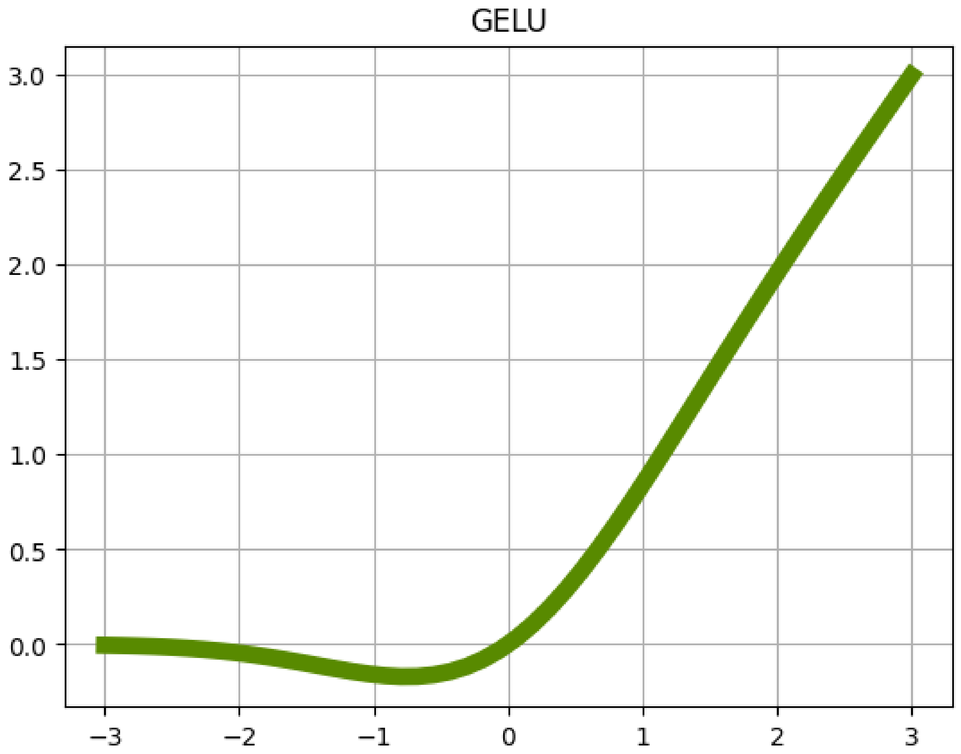
Group Size = 3

Gaussian Error Linear Unit (GELU)

GELU

Gaussian Error Linear Unit



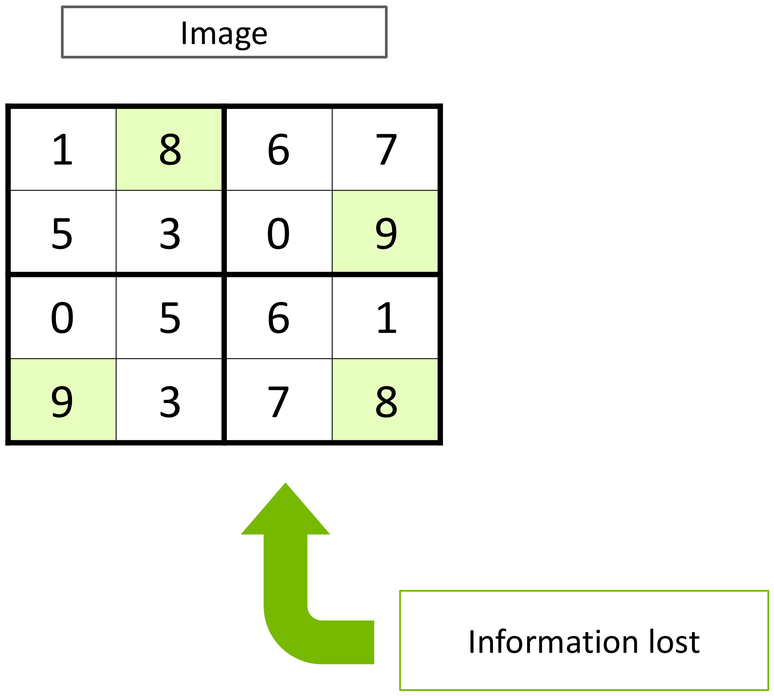


Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Rearrange Pooling

Pooling

Max Pooling Review



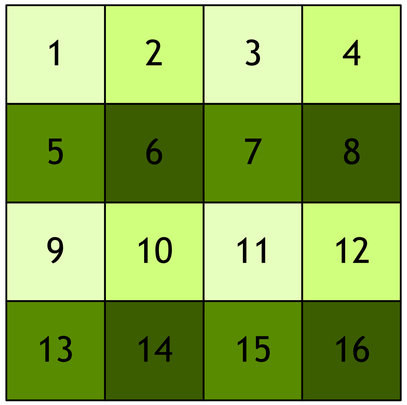
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Einops

Better Dimension Manipulation

```
Rearrange(
  "c (h p1) (w p2) -> (c p1 p2) h w",
  p1=2, p2=2
)
```

Cut image into strips and stack



Einops: Einstein Operations

Flexible and powerful tensor operations for readable and reliable code.

Supports numpy, pytorch, tensorflow, jax, and others.

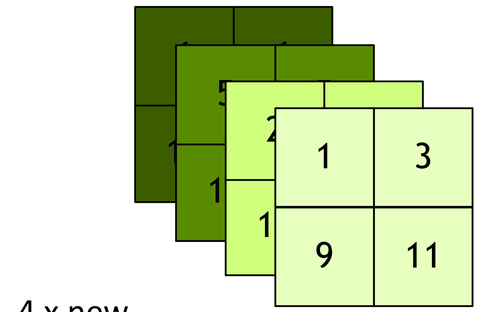
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Einops

Better Dimension Manipulation

```
Rearrange(
  "c (h p1) (w p2) -> (c p1 p2) h w",
  p1=2, p2=2
)
```

Cut image into strips and stack



$^1/_2$ Height

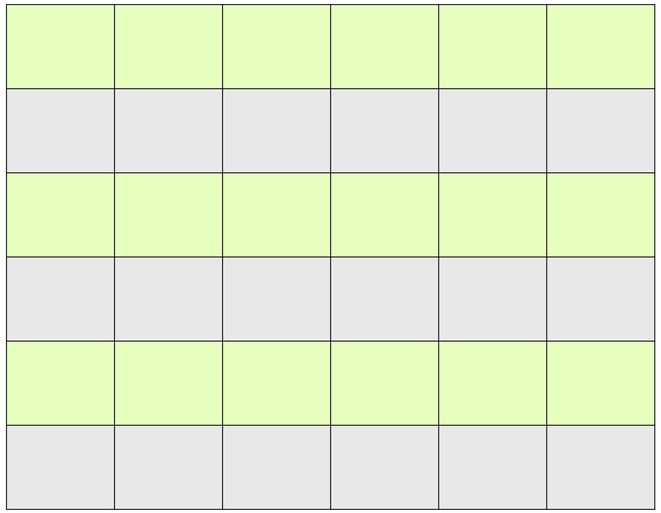
4 x new Channels

$^1/_2$ Width

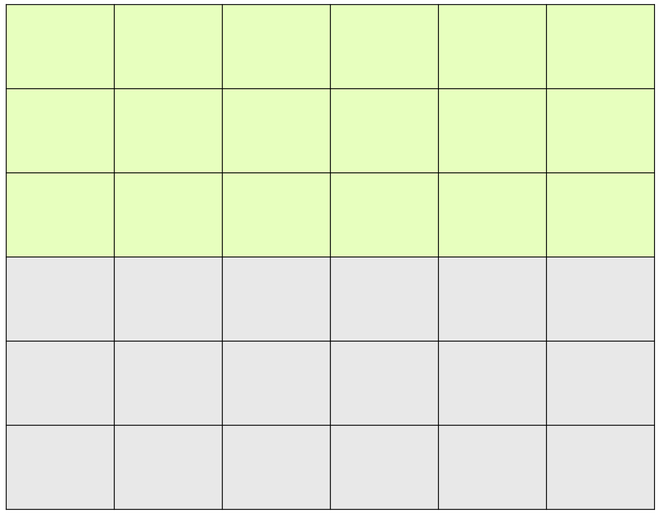
Einops

Order Matters

```
Rearrange(
  "c (h p1) w -> (c p1) h w", p1=2
)
```



```
Rearrange(
  "c (p1 h) w -> (c p1) h w", p1=2
)
```



Sinusoidal Position Embeddings

Time as a Sequence

How to Represent Time as Discrete Steps?

t

0	1	2	3	4	5	6	7	8	9	…

As a one-hot encoding?

t = 7

0	0	0	0	0	0	0	1	0	0	…

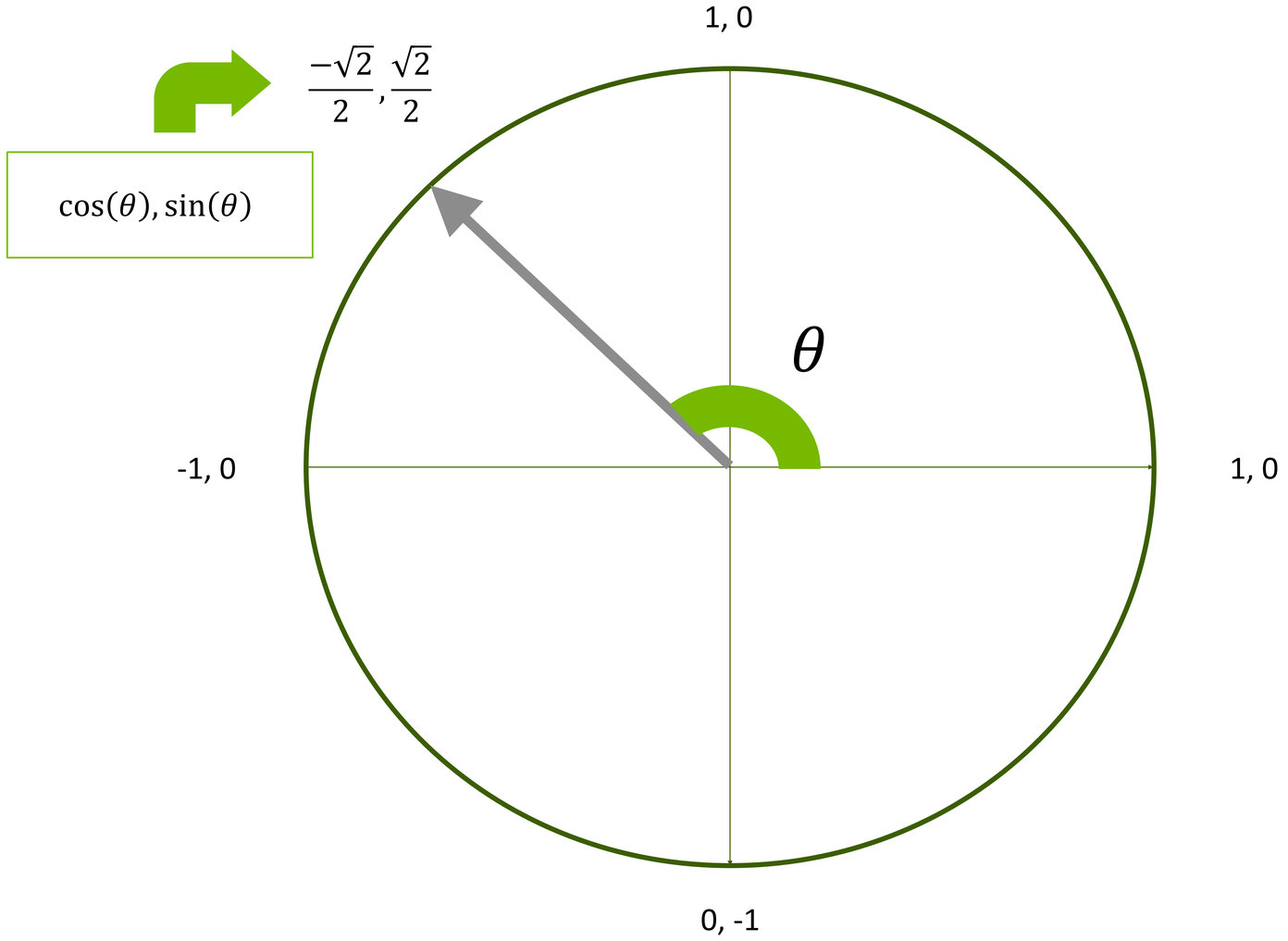
As binary?

0	0	0	0	0	0	0	0	1	1	…
0	0	0	0	1	1	1	1	0	0	…
0	0	1	1	0	0	1	1	0	0	…
0	1	0	1	0	1	0	1	0	1	…

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

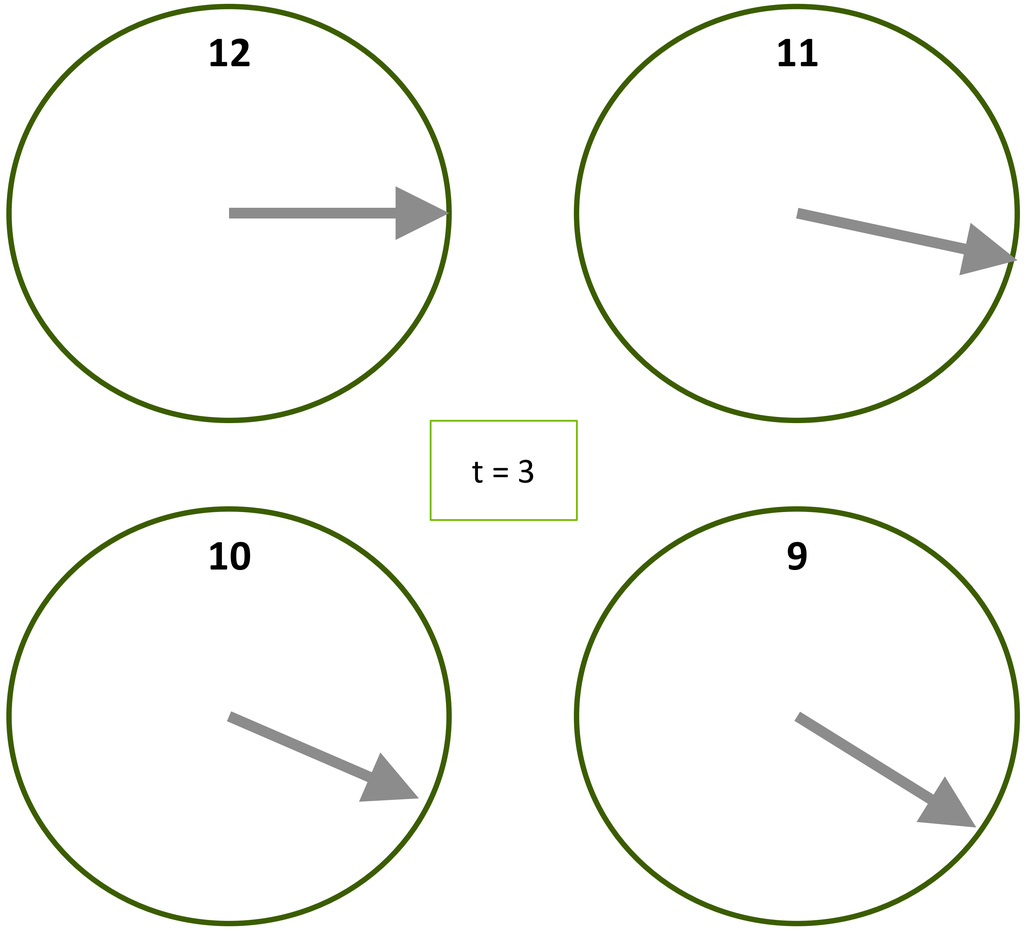
Time as a Sequence

As a Unit Circle?



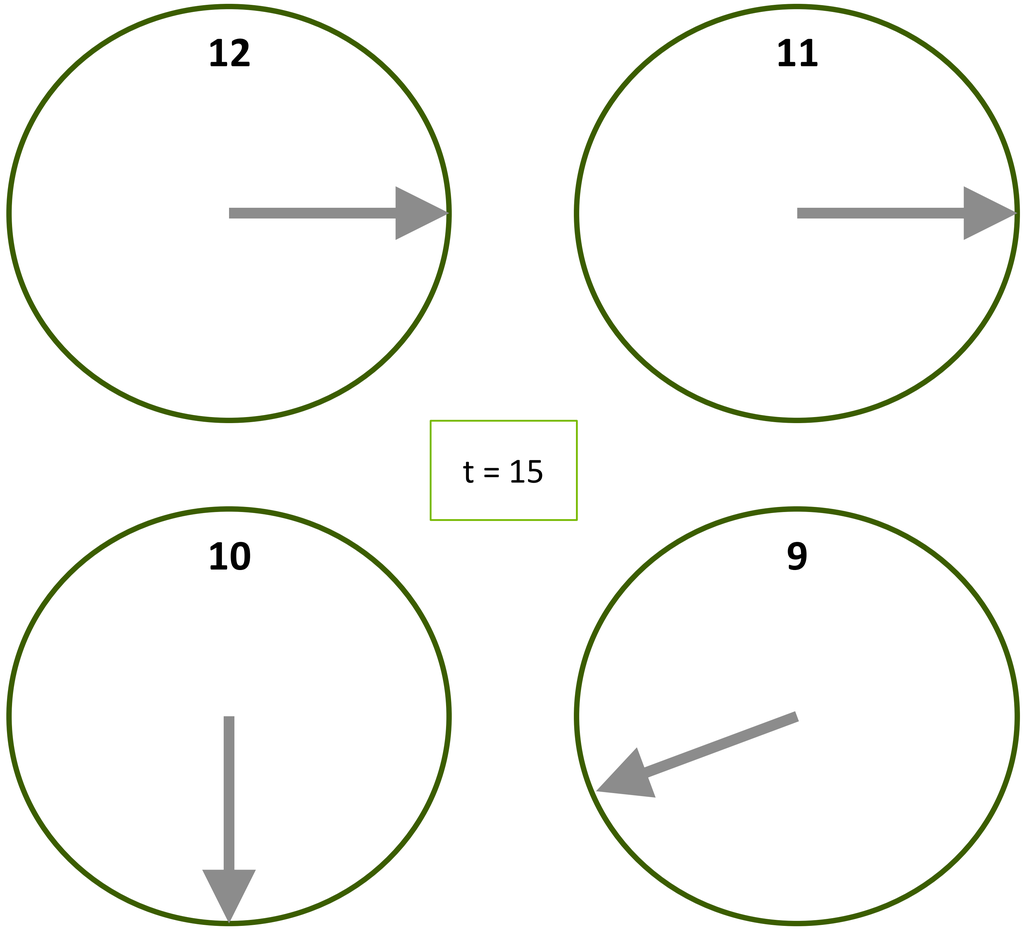
Time as a Sequence

As Unit Circles?



Time as a Sequence

As Unit Circles?

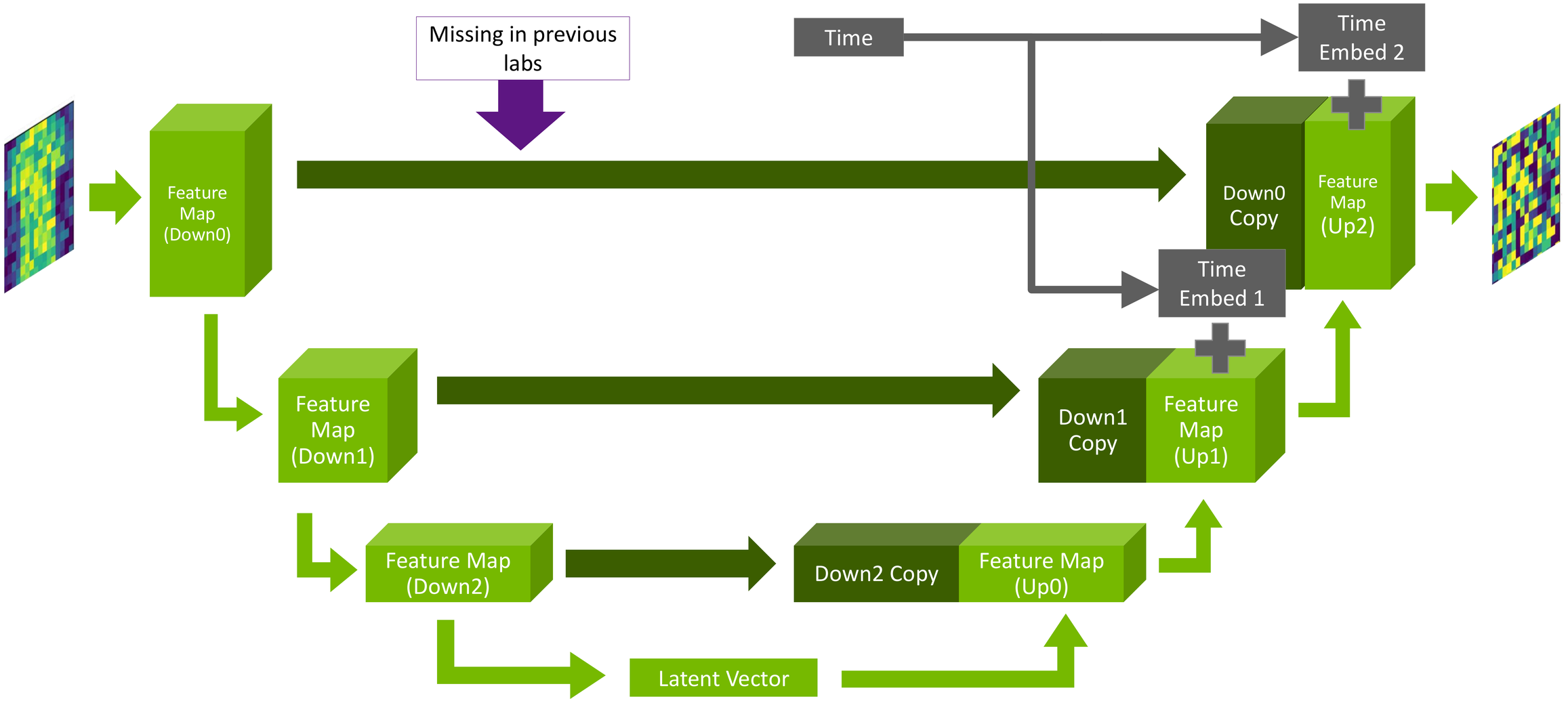


A bunch of abstract clocks with different numbers on them

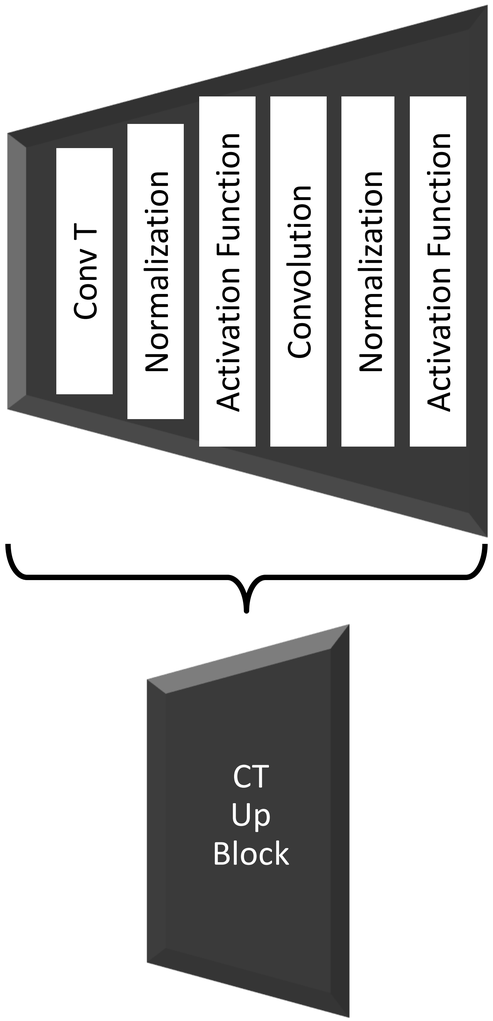
Deeper Networks

The U-Net Architecture

Adding Time

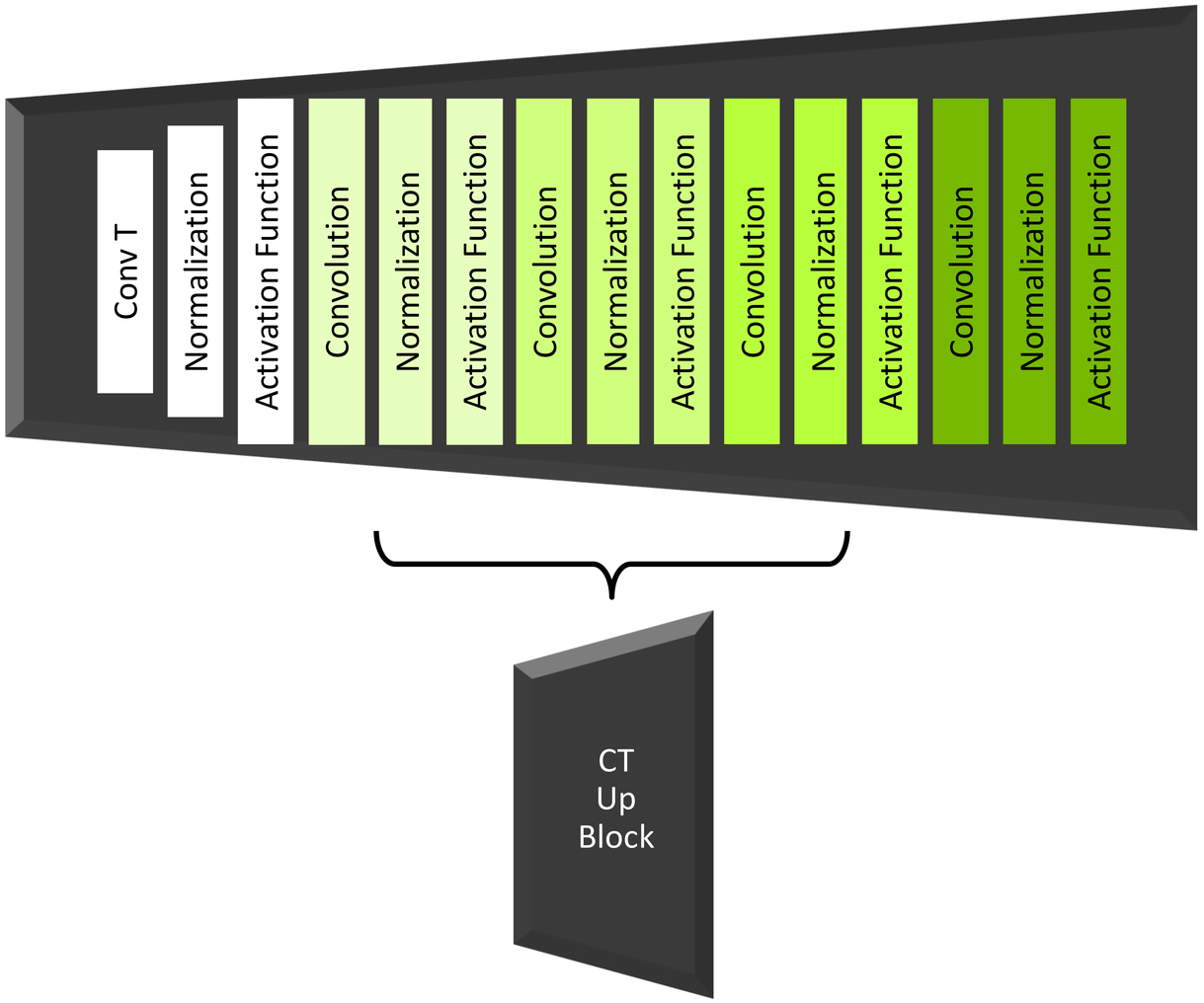


Adding Depth



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Adding Depth



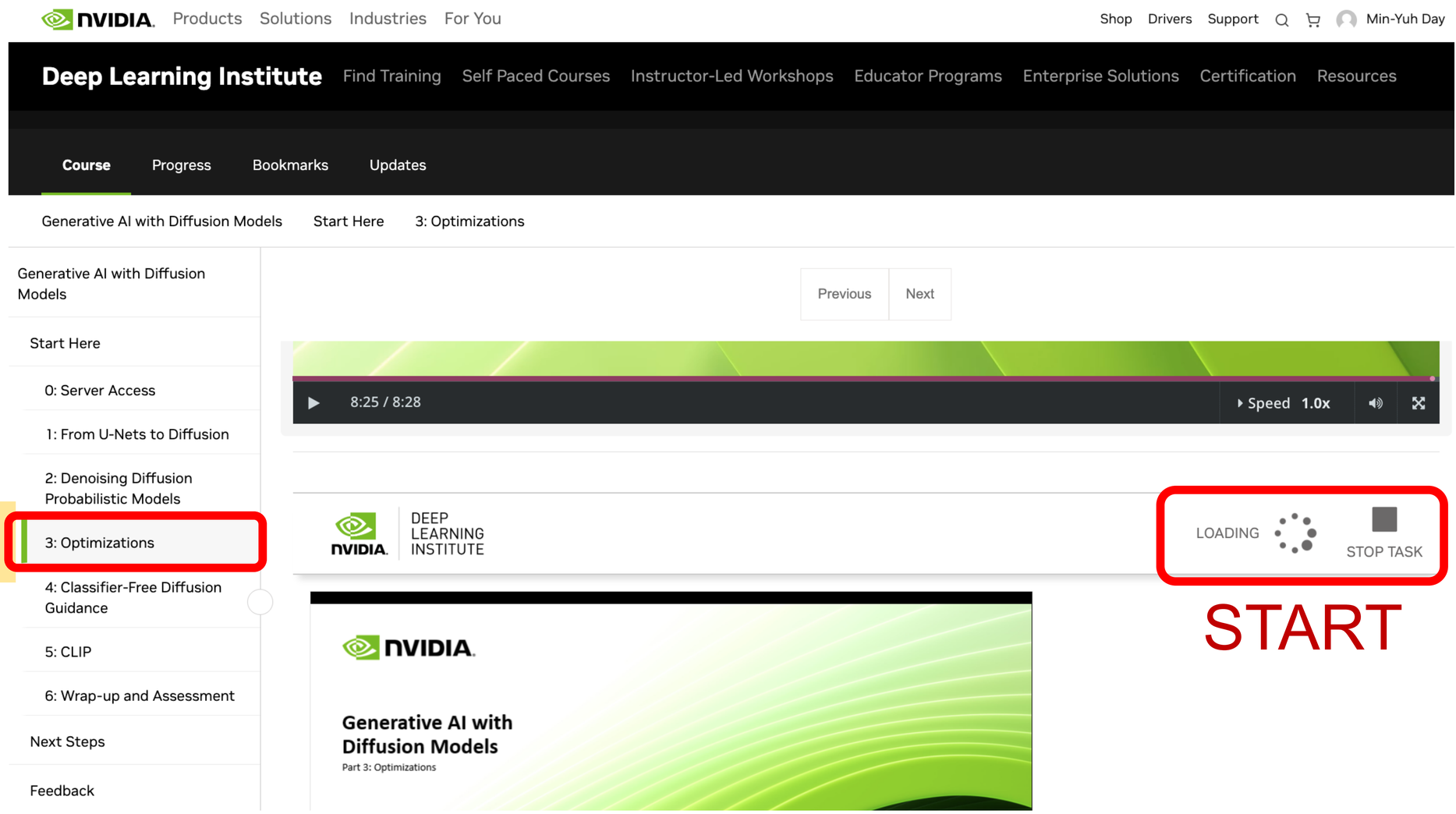
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Let’s get started!

3: Optimizations

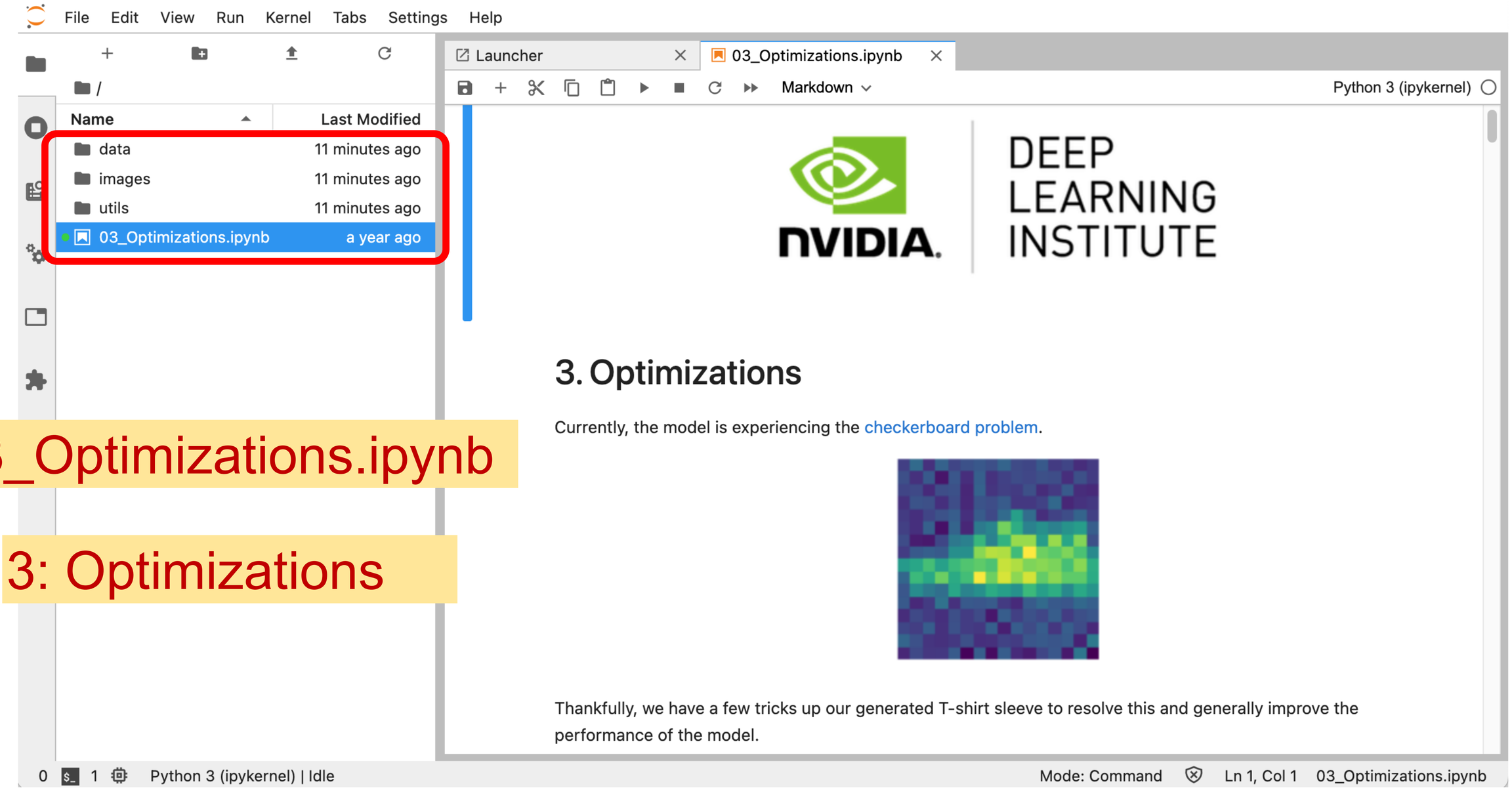
Generative AI with Diffusion Models

3: Optimizations



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models



03_Optimizations.ipynb

3: Optimizations

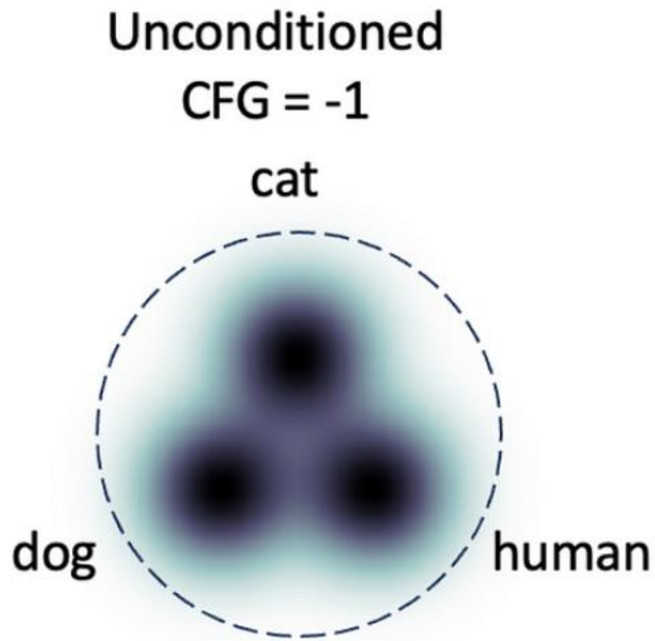
https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

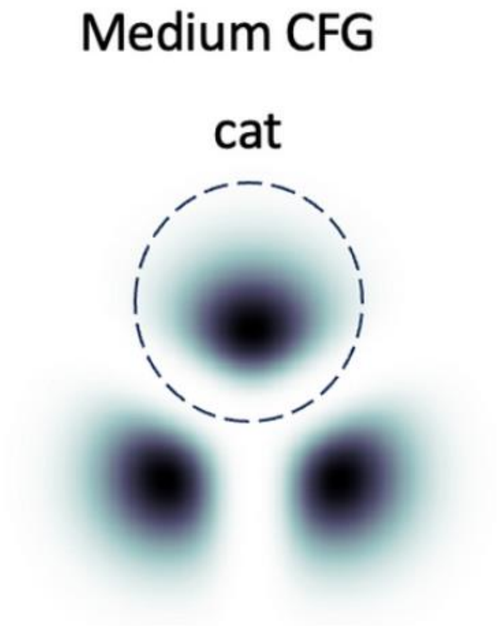
Generative AI with Diffusion Models

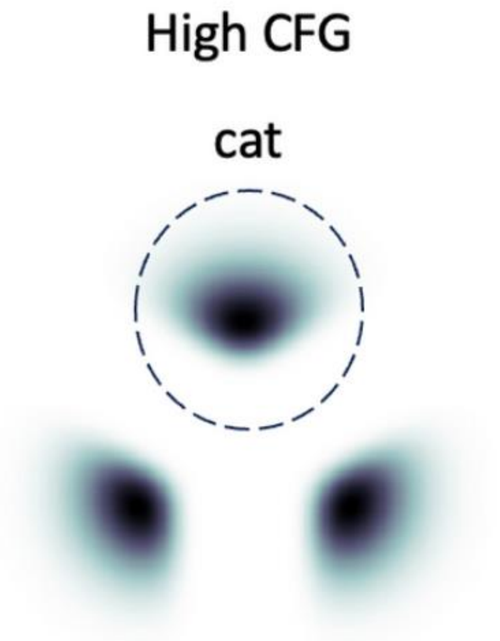
- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance (CFG)
- Part 5: CLIP (Contrastive Language-Image Pre-Training)
- Part 6: Wrap-up & Assessment

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Classifier-Free Diffusion Guidance (CFG)
CFG Scale (Condition)







Low CFG

More Freedom

Better Quality

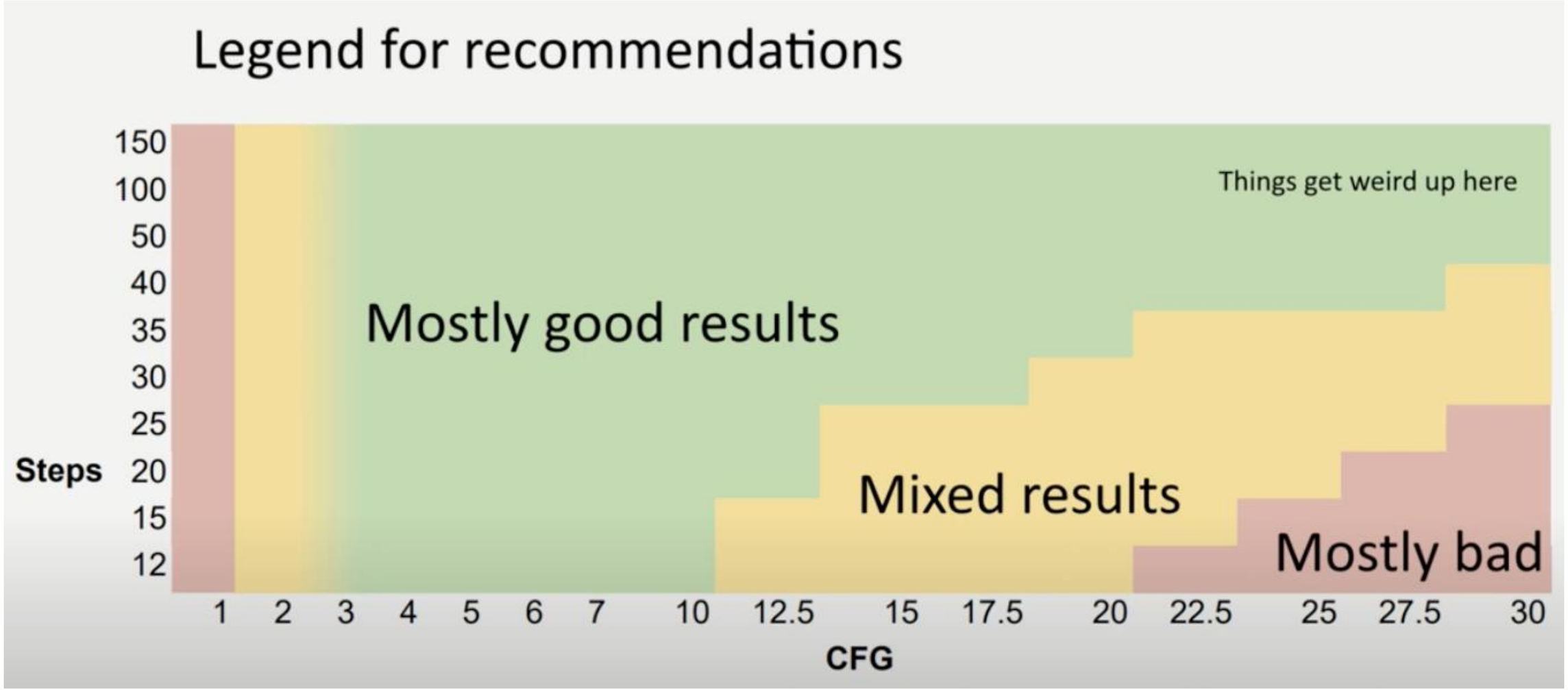
High CFG

Less Freedom

Math Prompt

Source: https://blogs.novita.ai/understanding-cfg-scale-in-stable-diffusion/

Classifier-Free Diffusion Guidance (CFG)
CFG and Steps



Source: SiliconThaumaturgy (2023), Stable Diffusion Deep Dive - CFG - Don't Accidentally Fry Your Images, https://www.youtube.com/watch?v=kuhO9zAzetk

Generative AI with Diffusion Models

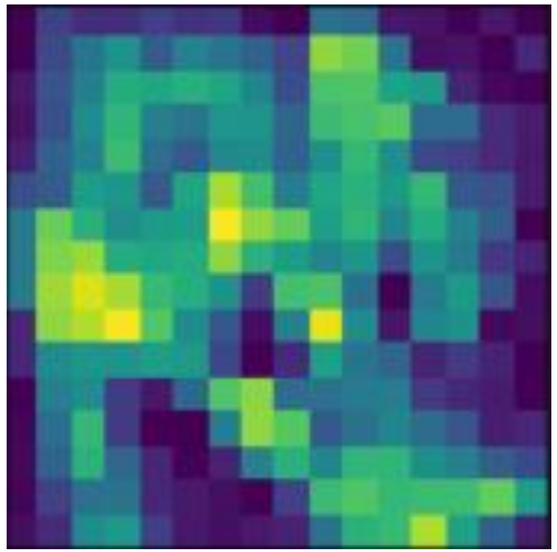
Part 4: Classifier-Free Diffusion Guidance

Agenda

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance
- Part 5: CLIP
- Part 6: Wrap-up & Assessment

Adding Context

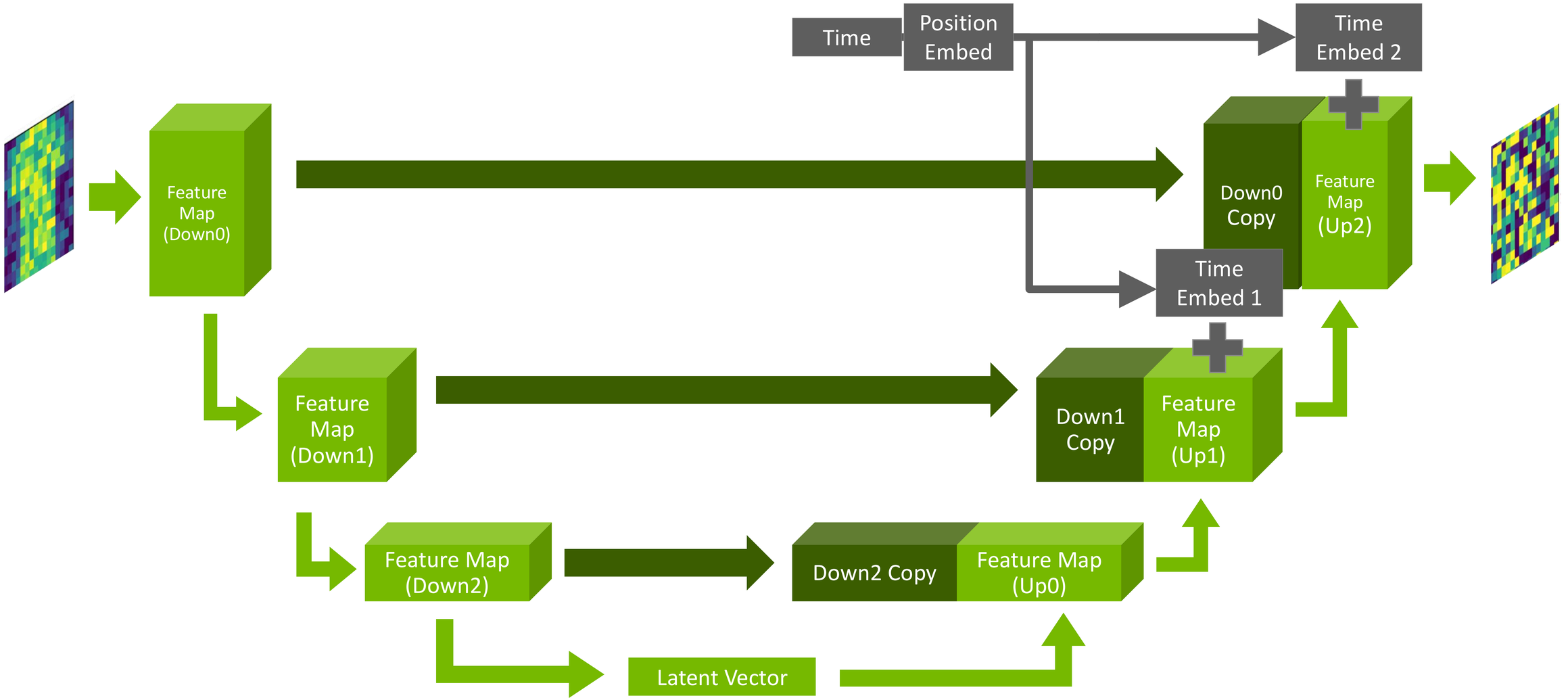
Adding Context



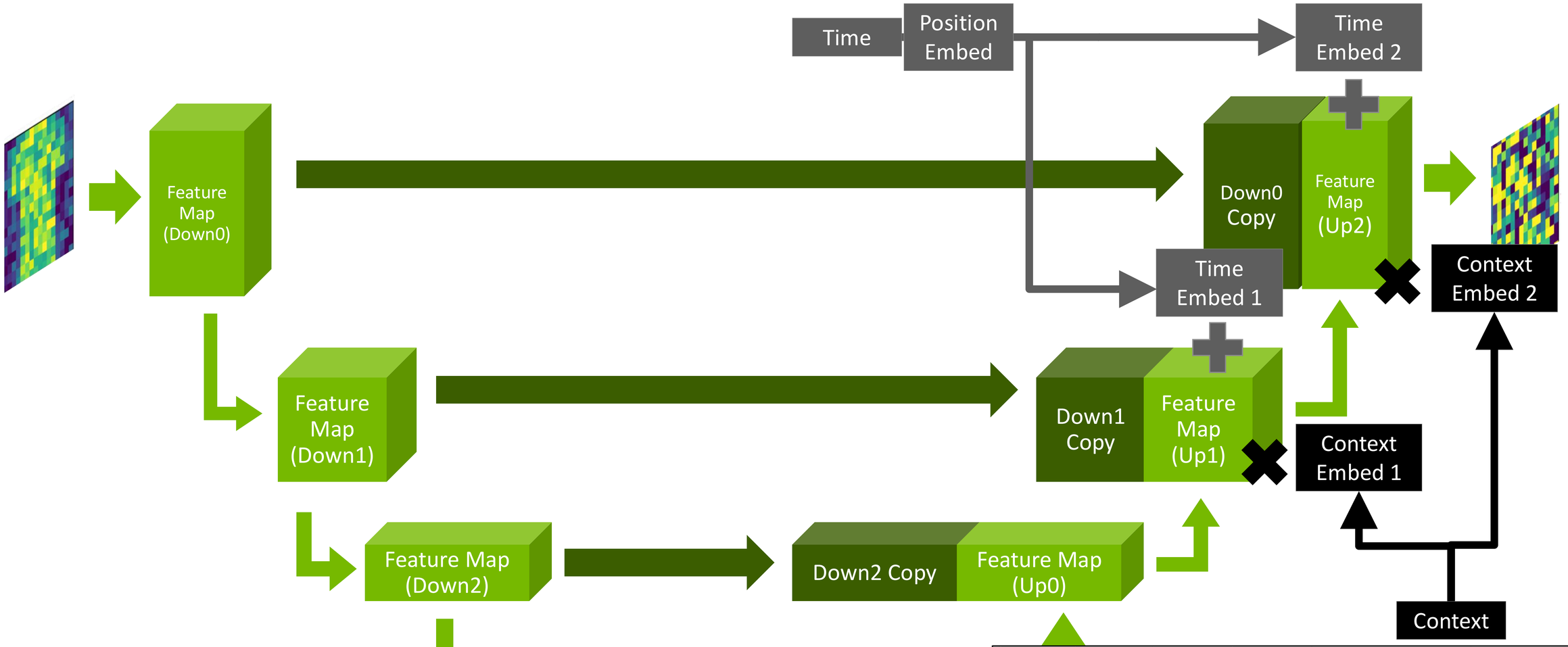
Shirt sleeve?

Shoe heel?

Adding Context

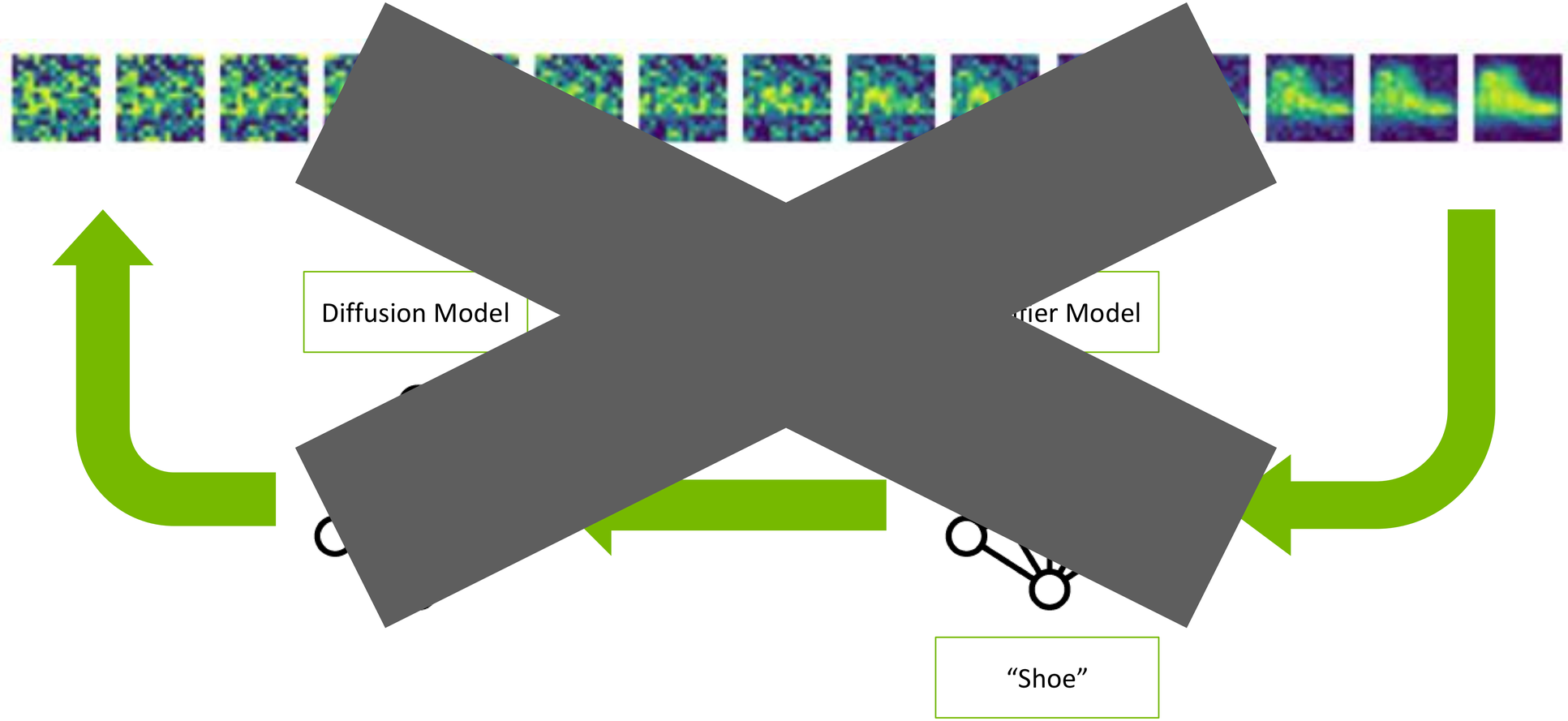


Adding Context



Classifier-Free Diffusion Guidance

Make a Classifier Model?



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Classifier-Free Diffusion Guidance

CLASSIFIER-FREE DIFFUSION GUIDANCE

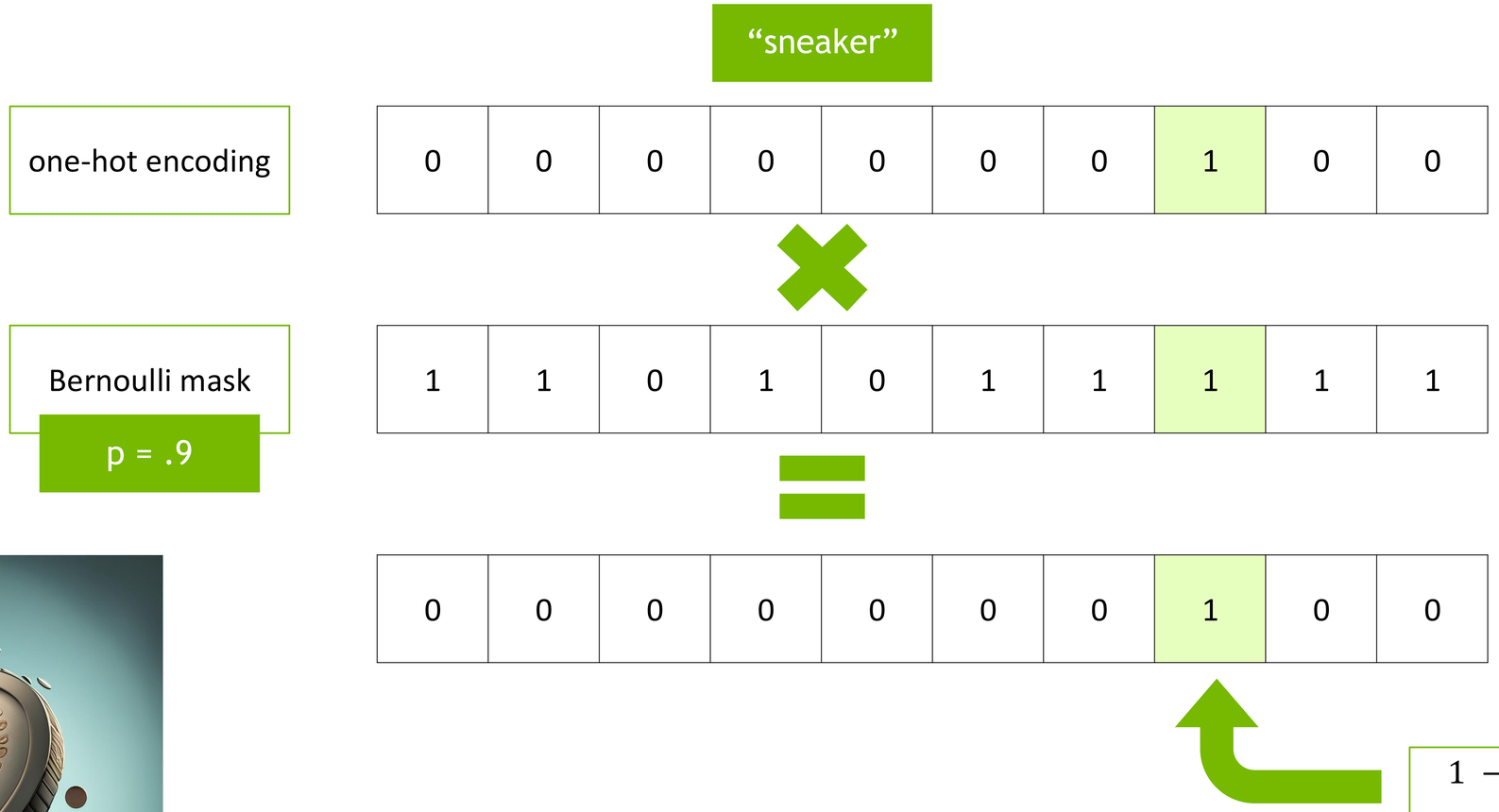
Jonathan Ho & Tim Salimans
Google Research, Brain team
{jonathanho,salimans}@google.com

ABSTRACT

Classifier guidance is a recently introduced method to trade off mode coverage and sample fidelity in conditional diffusion models post training, in the same spirit as low temperature sampling or truncation in other types of generative models. Classifier guidance combines the score estimate of a diffusion model with the gradient of an image classifier and thereby requires training an image classifier separate from the diffusion model. It also raises the question of whether guidance

Classifier Free Diffusion Guidance

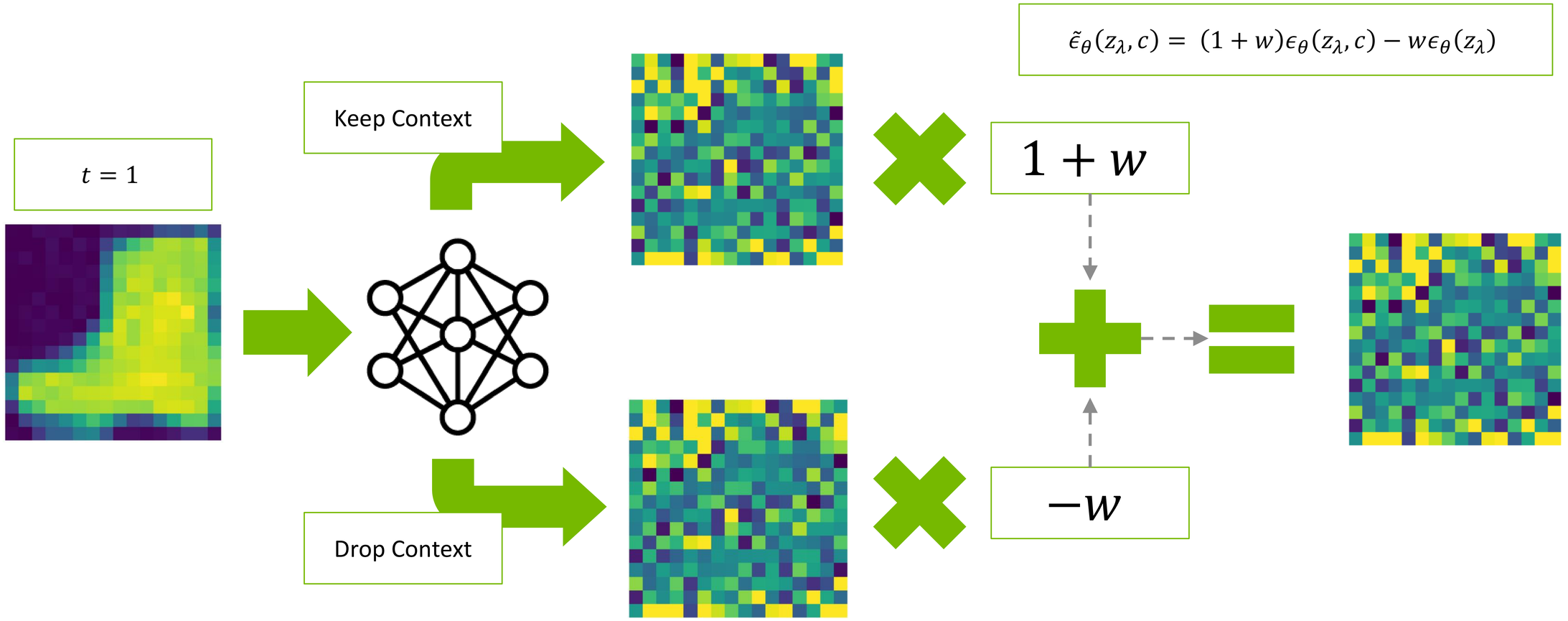
Bernoulli Masks



A weighted coin flipping through the air like a cartoon

Classifier Free Diffusion Guidance

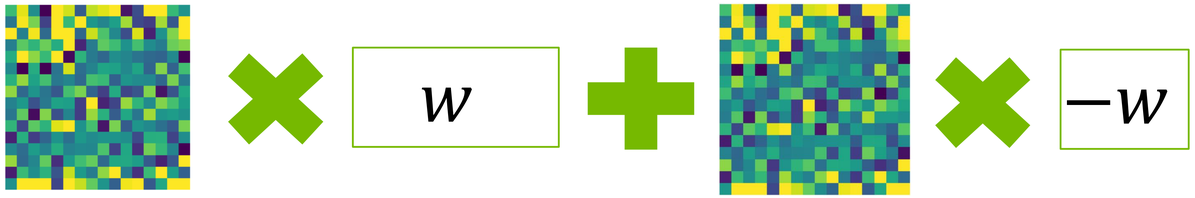
Weighted Reverse Diffusion



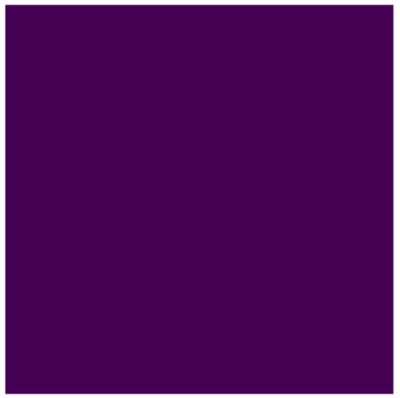
Classifier Free Diffusion Guidance

Weighted Reverse Diffusion

$$(w)\epsilon_\theta(z_\lambda, c) - w\epsilon_\theta(z_\lambda)$$

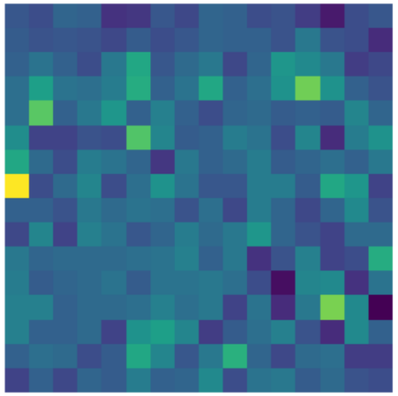


w = 0



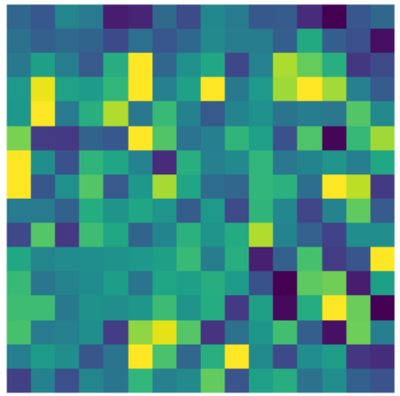
No context feature exaggeration

w = 1



Some context feature exaggeration

w = 10



Huge context feature exaggeration

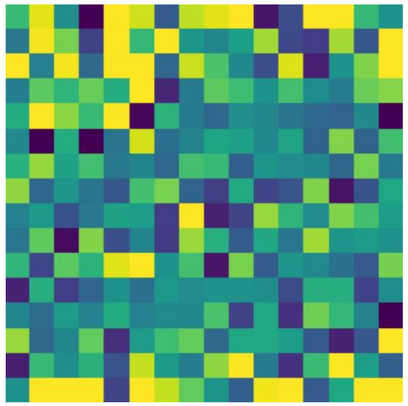
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Classifier Free Diffusion Guidance

Weighted Reverse Diffusion

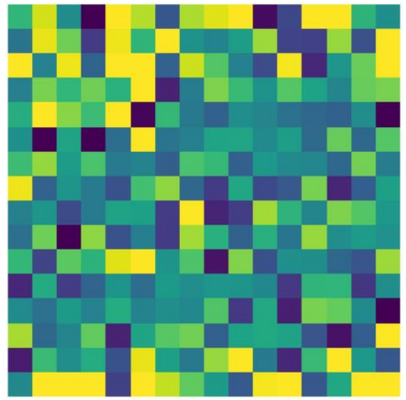
$$(1 + w)\epsilon_\theta(z_\lambda, c) - w\epsilon_\theta(z_\lambda)$$

w = 0



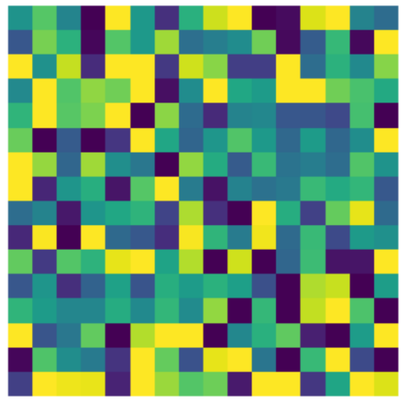
No context feature exaggeration

w = 1



Some context feature exaggeration

w = 10



Huge context feature exaggeration

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Classifier Free Diffusion Guidance

Weighted Reverse Diffusion

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Modified TF Flowers

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Modified TF Flowers

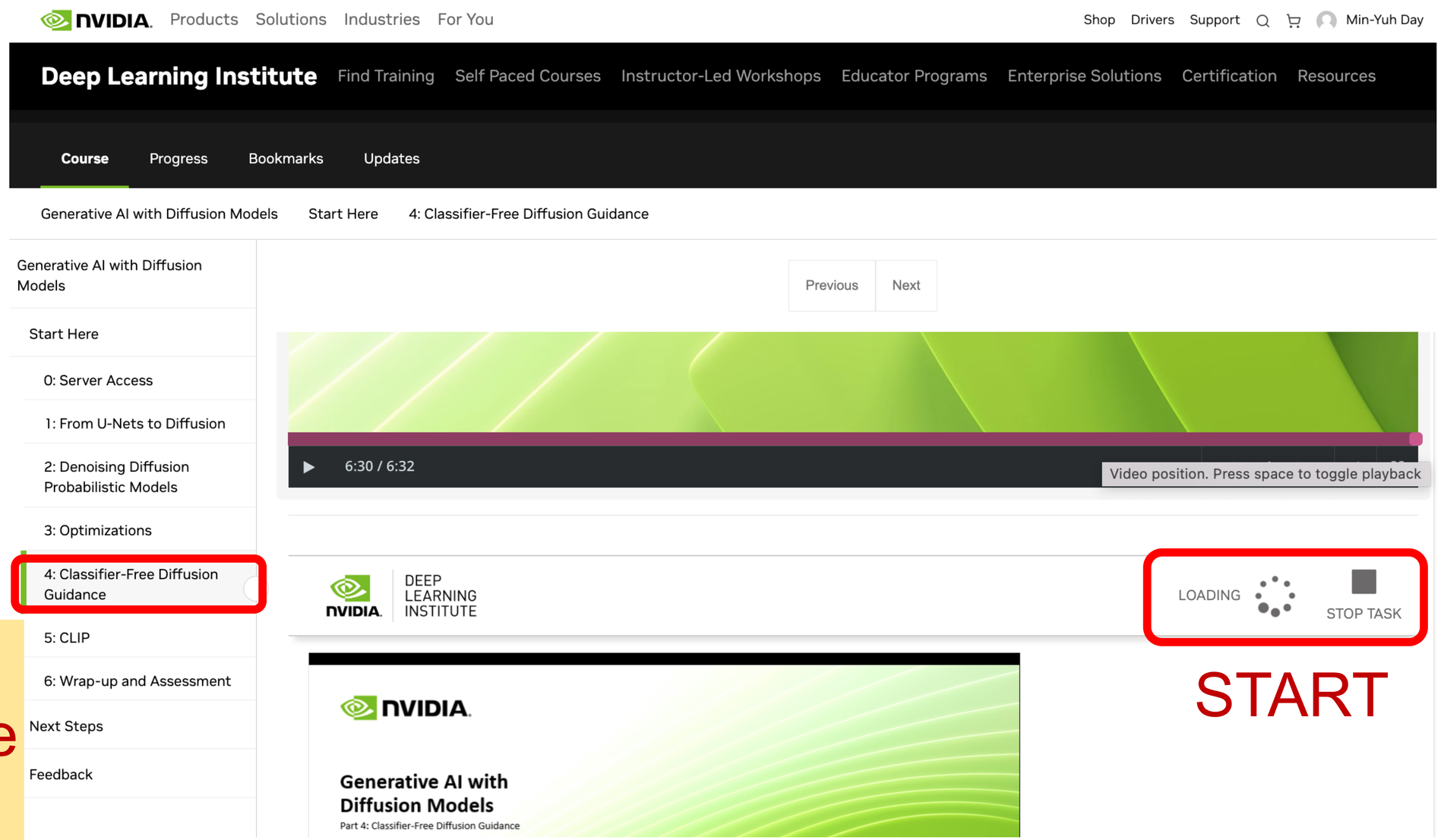
Modified for Image Generation

Photo by _e.t

Let’s get started!

4: Classifier-Free Diffusion Guidance (CFG)

Generative AI with Diffusion Models

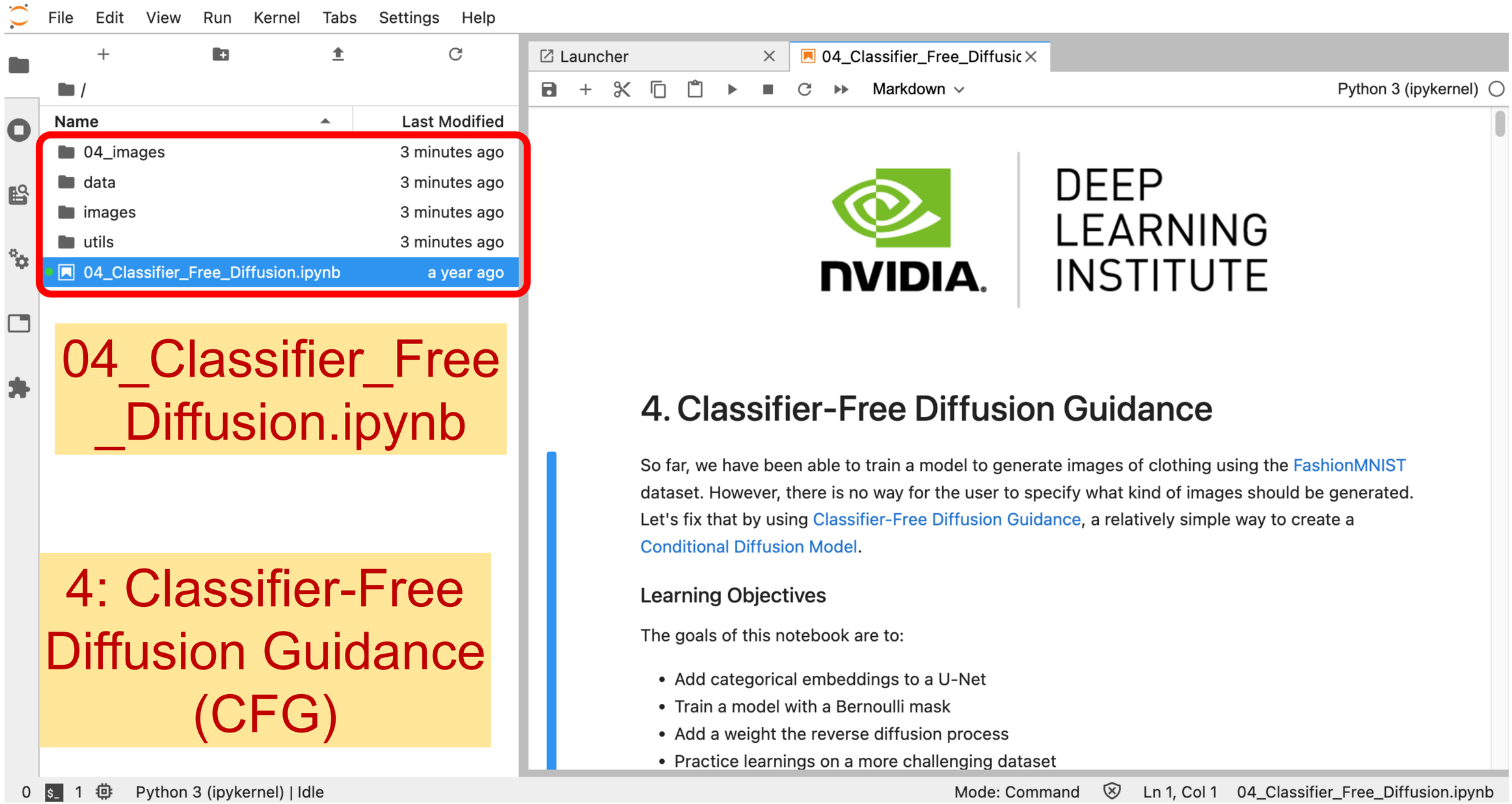


START

4: Classifier-Free Diffusion Guidance (CFG)

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models



04_Classifier_Free_Diffusion.ipynb

4: Classifier-Free Diffusion Guidance (CFG)

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA
Generative AI with Diffusion Models
Part III

[0,1,2] [3, 4] [5, 6]

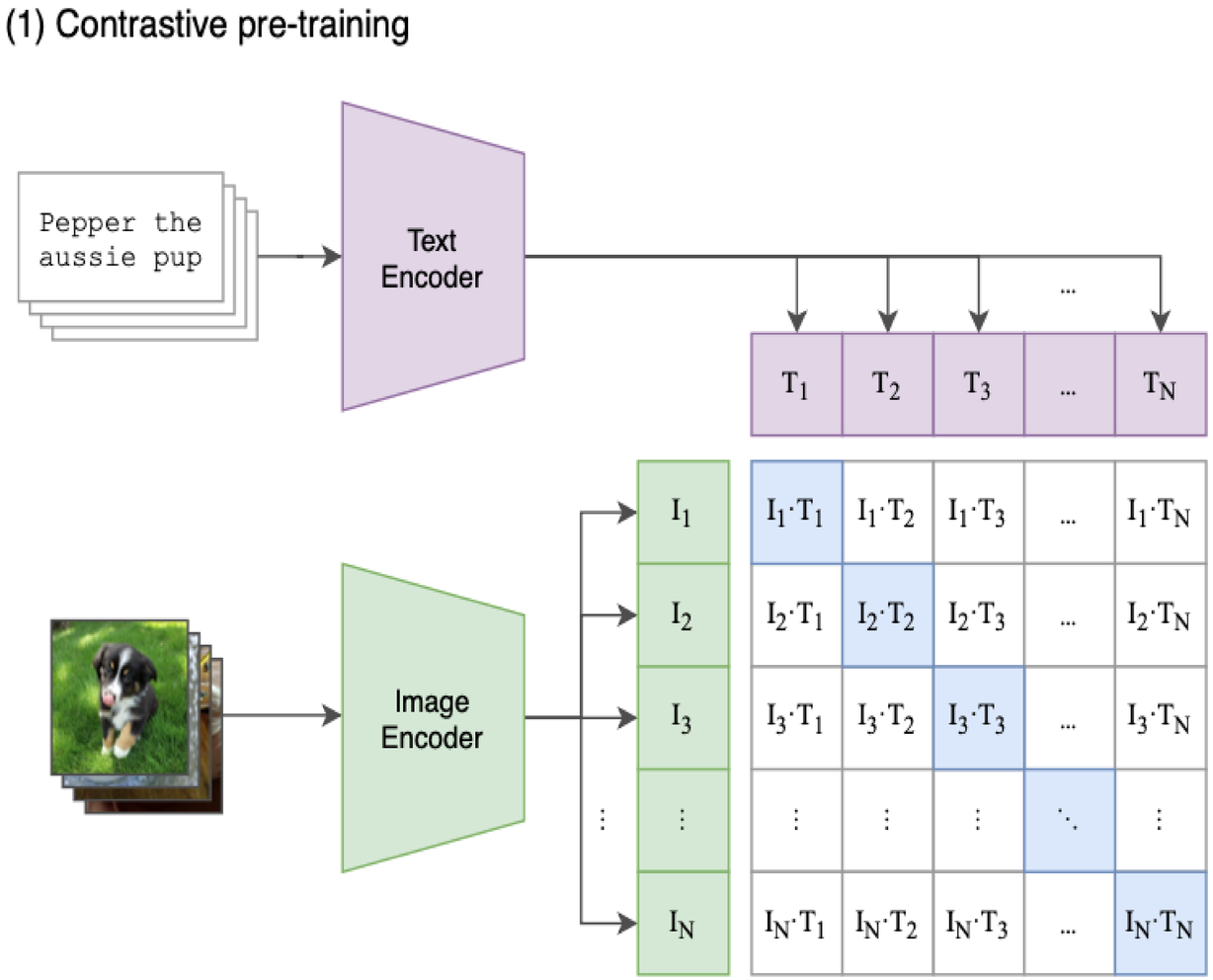
Generative AI with Diffusion Models

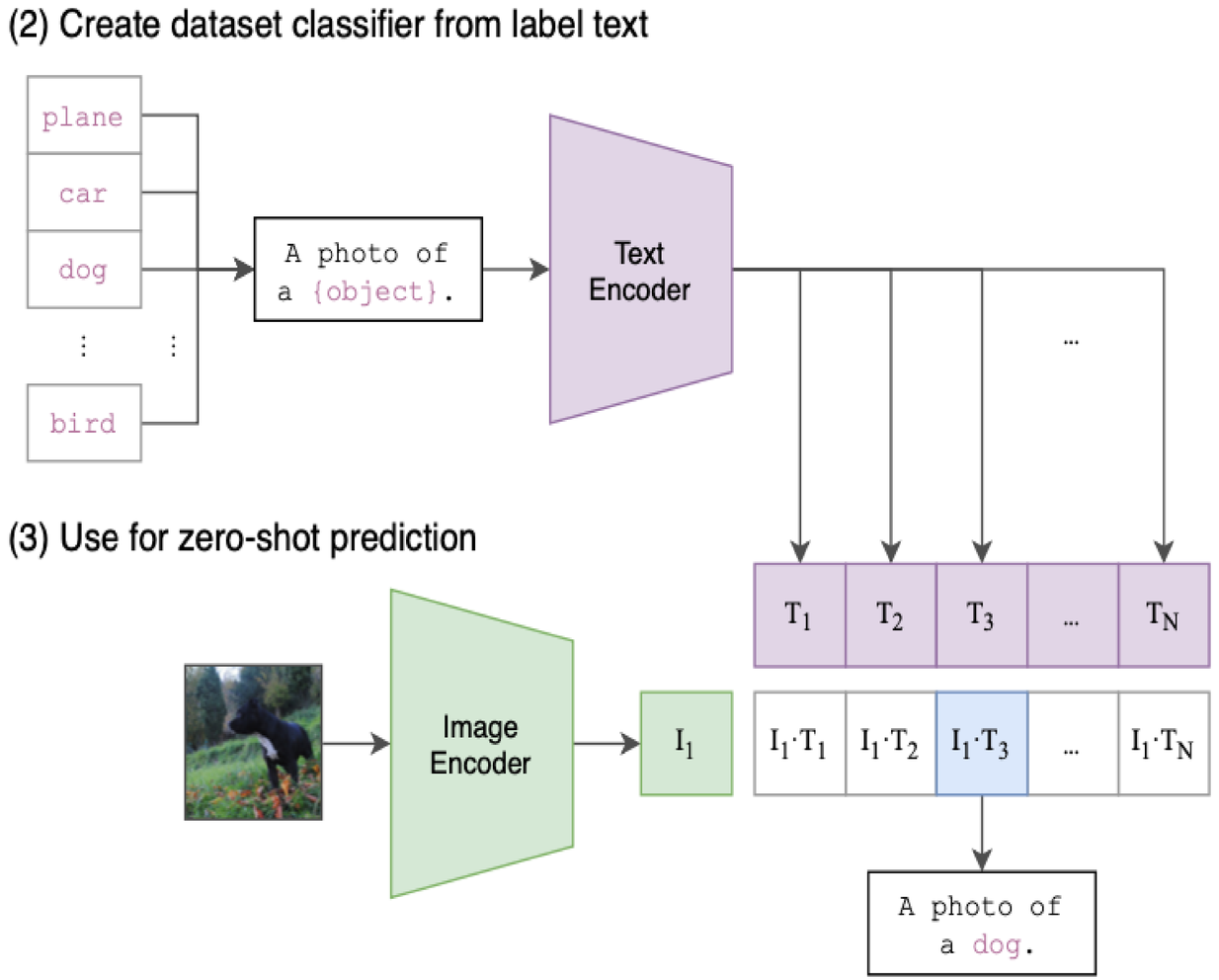
- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance (CFG)
- Part 5: CLIP (Contrastive Language-Image Pre-Training)
- Part 6: Wrap-up & Assessment

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

CLIP (Contrastive Language-Image Pre-Training)

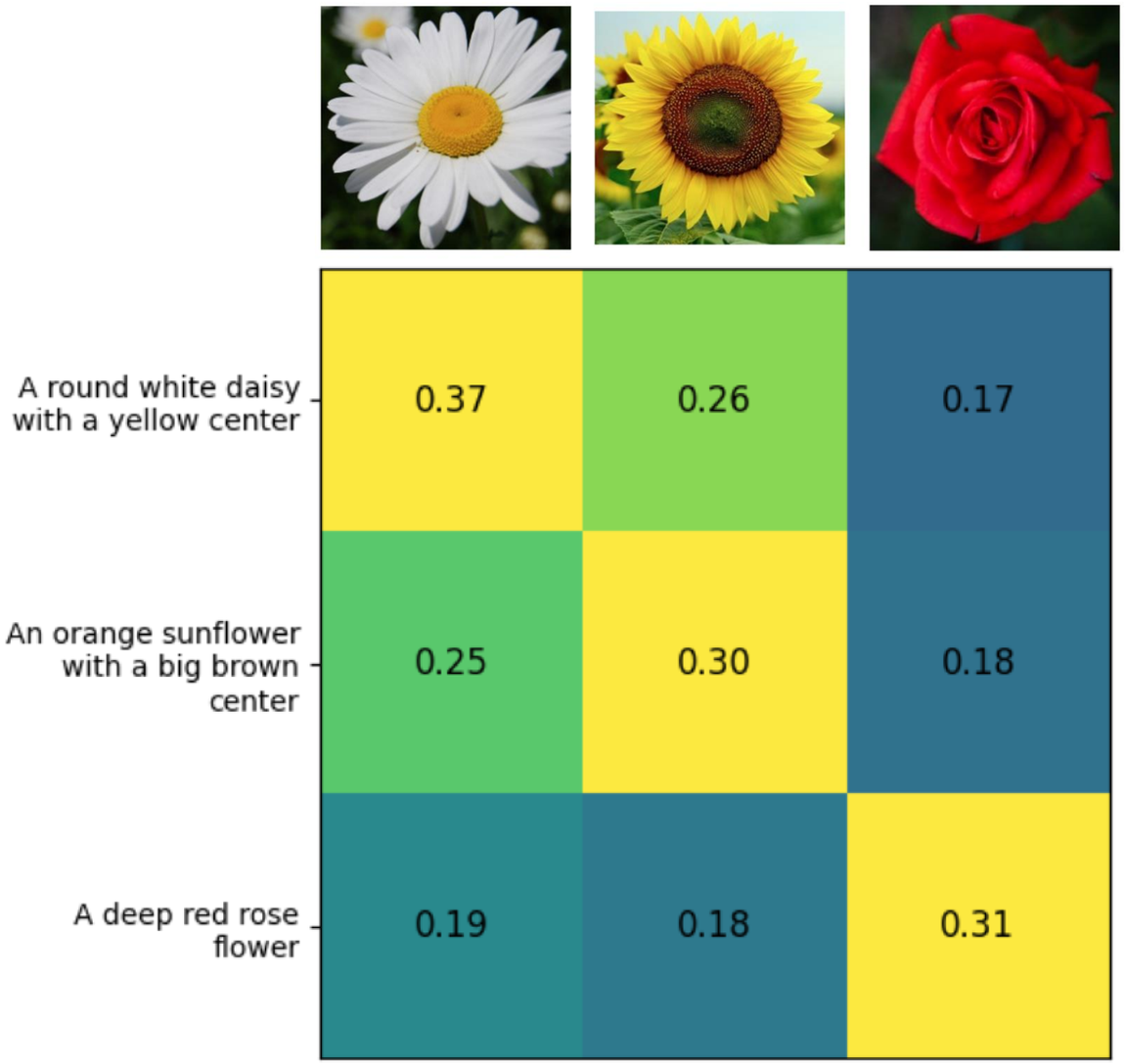
Learning Transferable Visual Models From Natural Language Supervision





Source: Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry et al. (2021) "Learning transferable visual models from natural language supervision." In International Conference on Machine Learning, pp. 8748-8763. PMLR.

CLIP
(Contrastive Language-Image Pre-Training)



https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Agenda

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance
- Part 5: CLIP
- Part 6: Wrap-up & Assessment

NVIDIA.

Contrastive Language-Image Pre-Training (CLIP)

Matching Text to Image

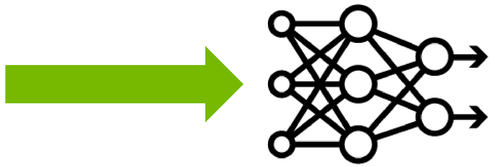
Is it Possible?

[0.8, -0.6, 0.7]

“A bunch of different marbles”

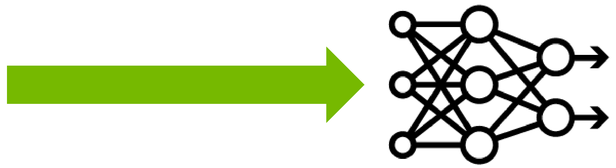
NVIDIA.

Cosine Similarity



[0.0, 1.0]

“A bunch of different marbles”



[1.0, 1.0]

NVIDIA.

Cosine Similarity

[0.0, 1.0]

[1.0, 1.0]

"A bunch of different marbles"

45°

$cos(45^{\circ}) = \frac{\sqrt{2}}{2}$

$cos(90^{\circ}) = 0$

$cos(270^{\circ}) = 0$

$cos(0^{\circ}) = 1$

$cos(180^{\circ}) = -1$

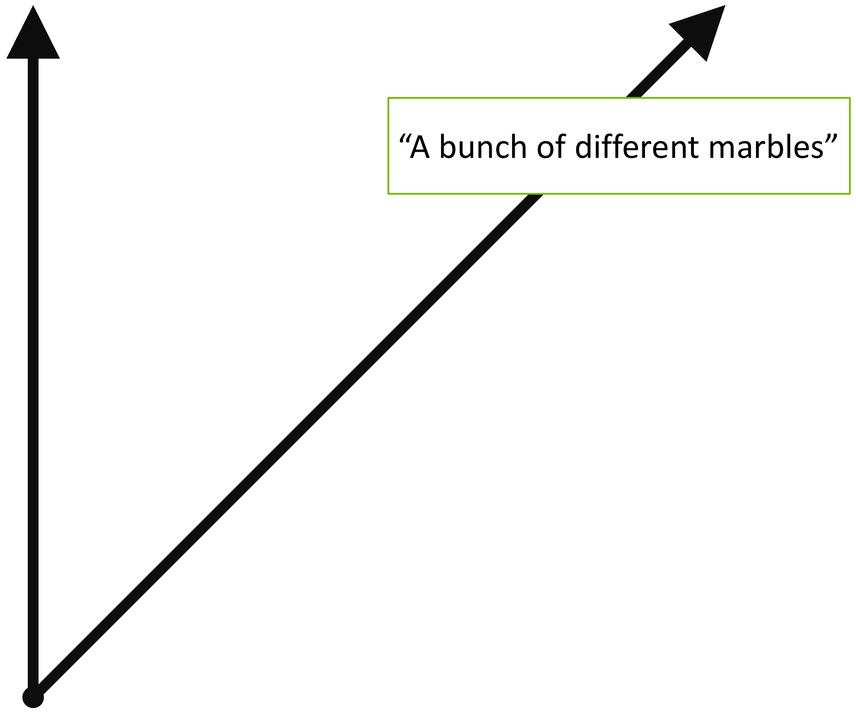
NVIDIA.

Dot Product

[0.0, 1.0]

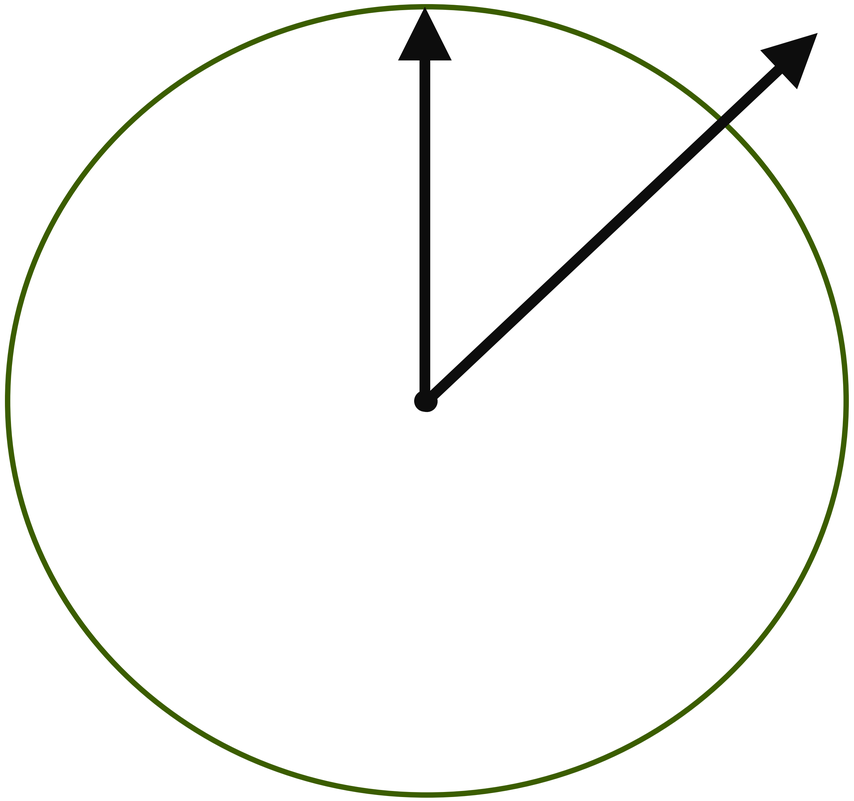
[1.0, 1.0]

“A bunch of different marbles”



[0.0, 1.0]

[1.0, 1.0]



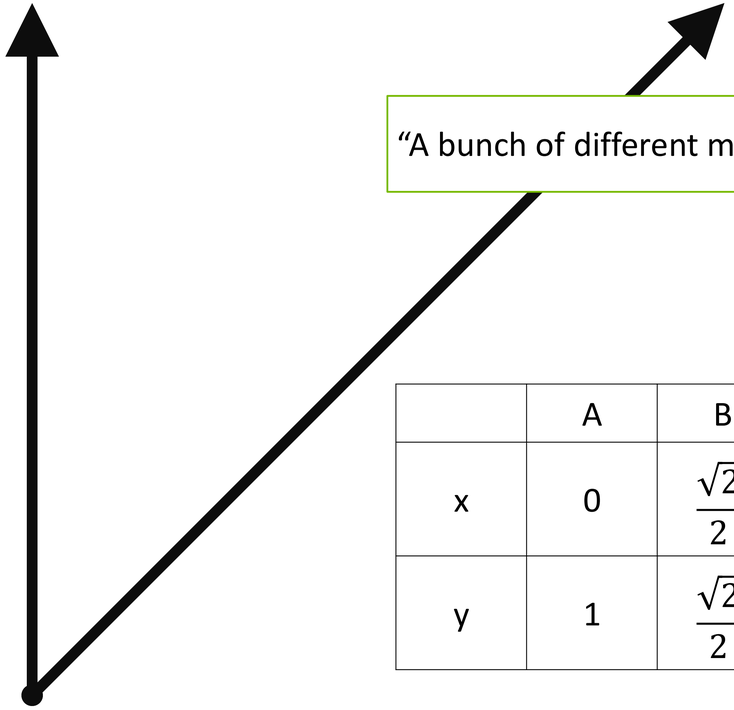
NVIDIA.

Dot Product

[0.0, 1.0]

[1.0, 1.0]

“A bunch of different marbles”



	A	B	A x B
x	0	$\frac{\sqrt{2}}{2}$	0
y	1	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$

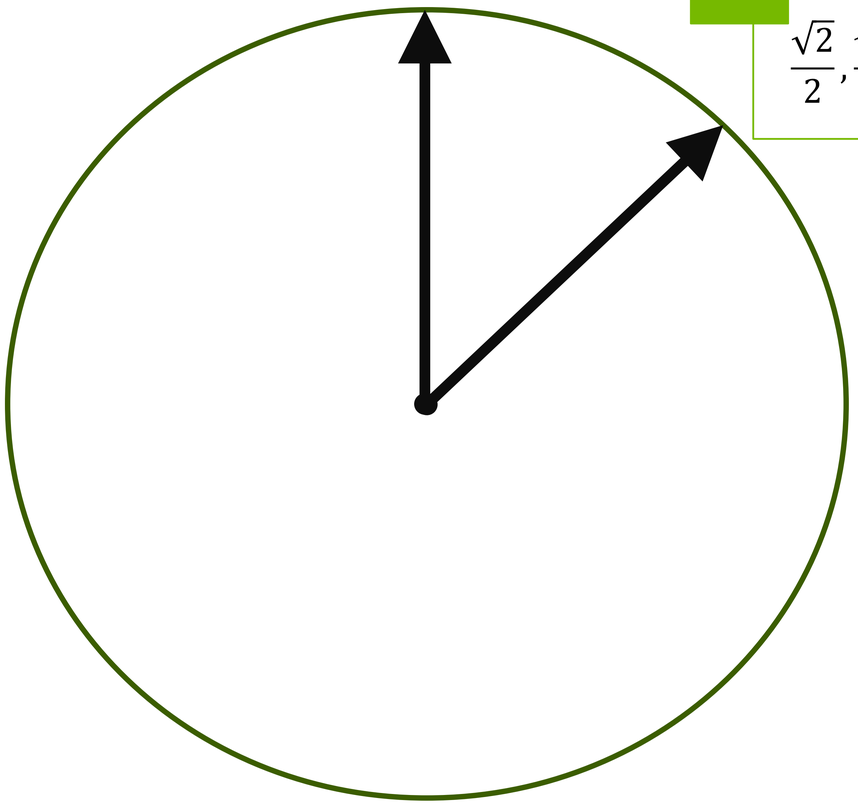
$$cos(45^{\circ}) = \frac{\sqrt{2}}{2}$$

A

[0.0, 1.0]

B

$\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}$



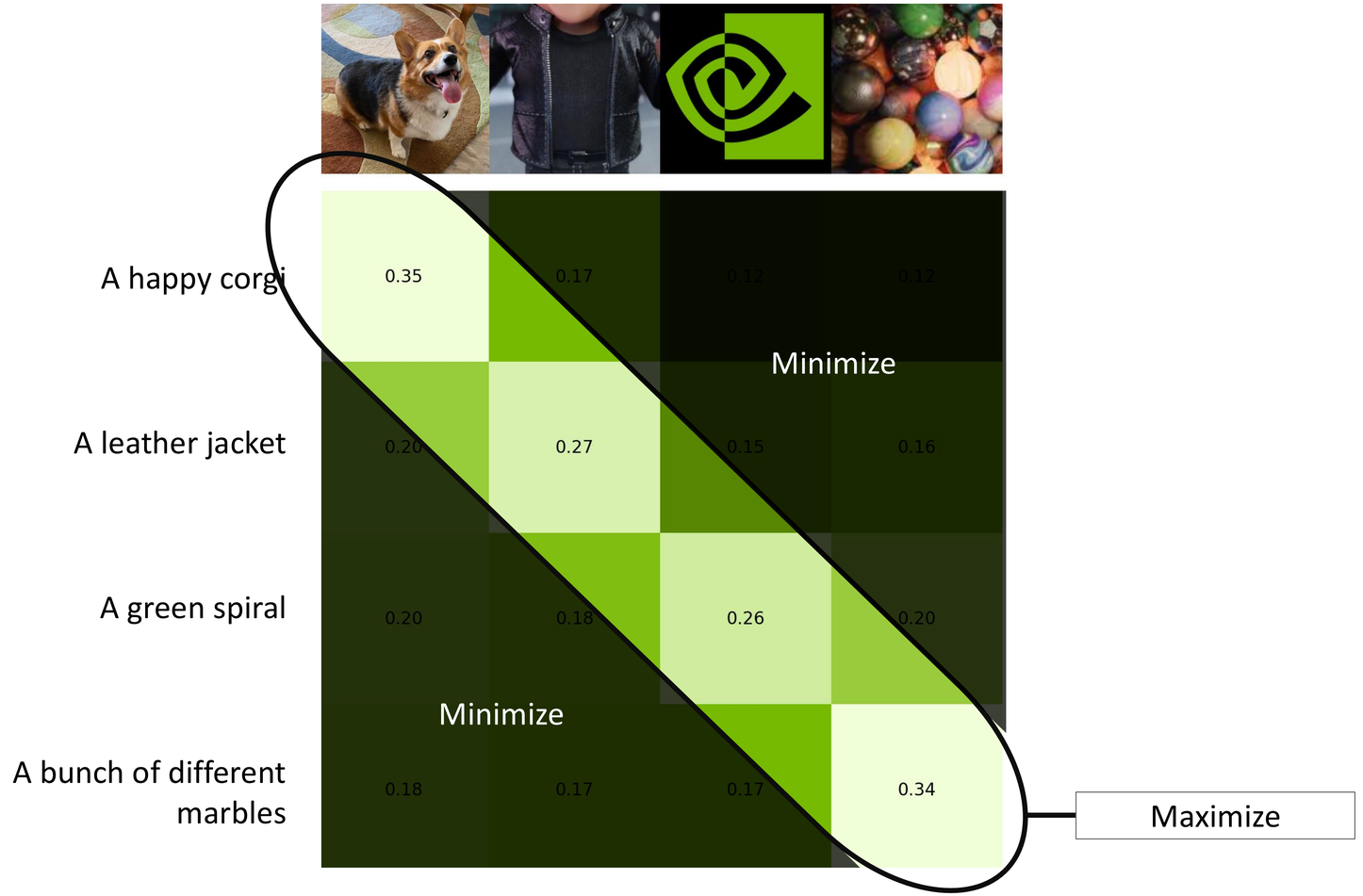
CLIP Training

A happy corgi	0.35	0.17	0.12	0.12
A leather jacket	0.20	0.27	0.15	0.16
A green spiral	0.20	0.18	0.26	0.20
A bunch of different marbles	0.18	0.17	0.17	0.34

Cosine similarity between encoding for “A happy corgi” and encoding for each image

Cosine similarity between encoding for the NVIDIA logo and encoding for each text

CLIP Training



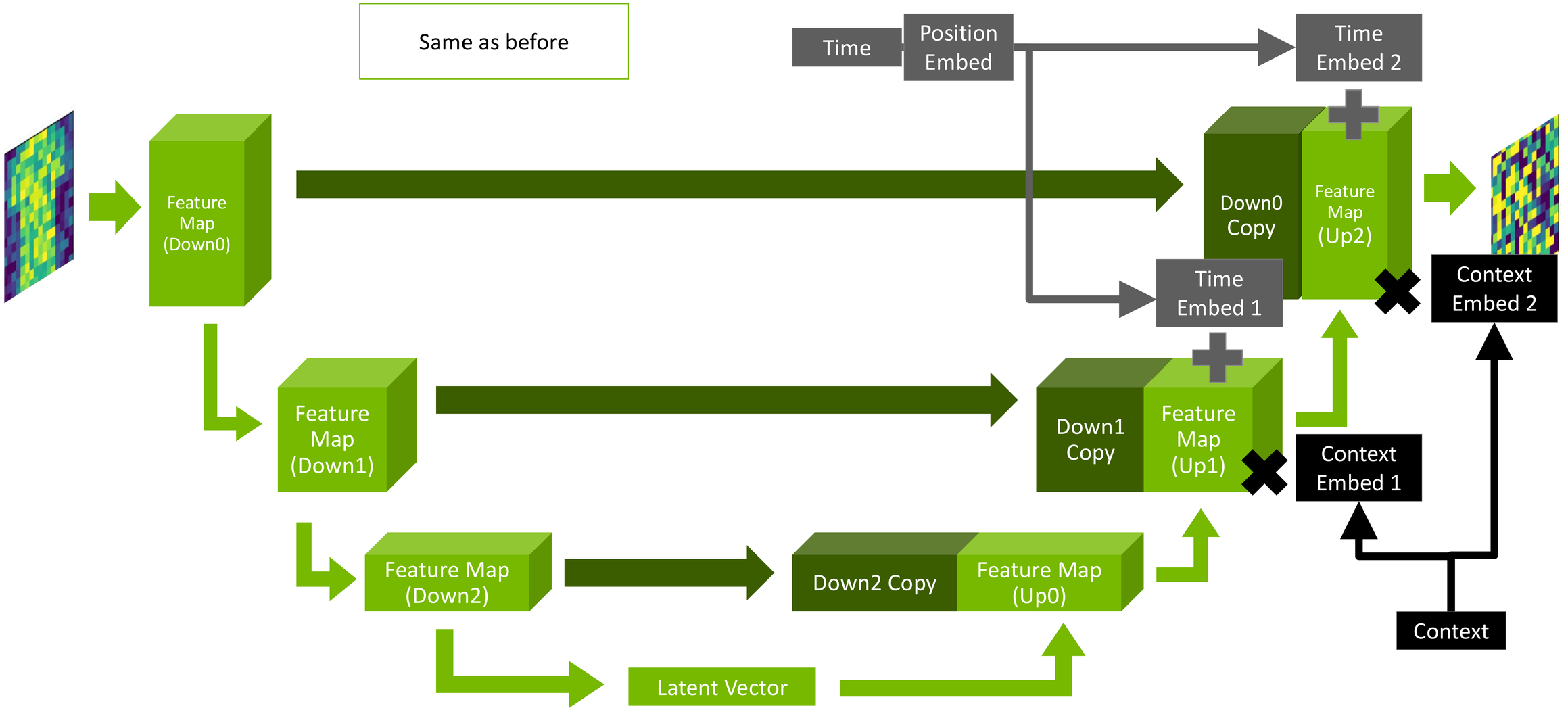
An Experiment

“

If CLIP is a pretrained model, do we need text labels to make a text-to-image model?

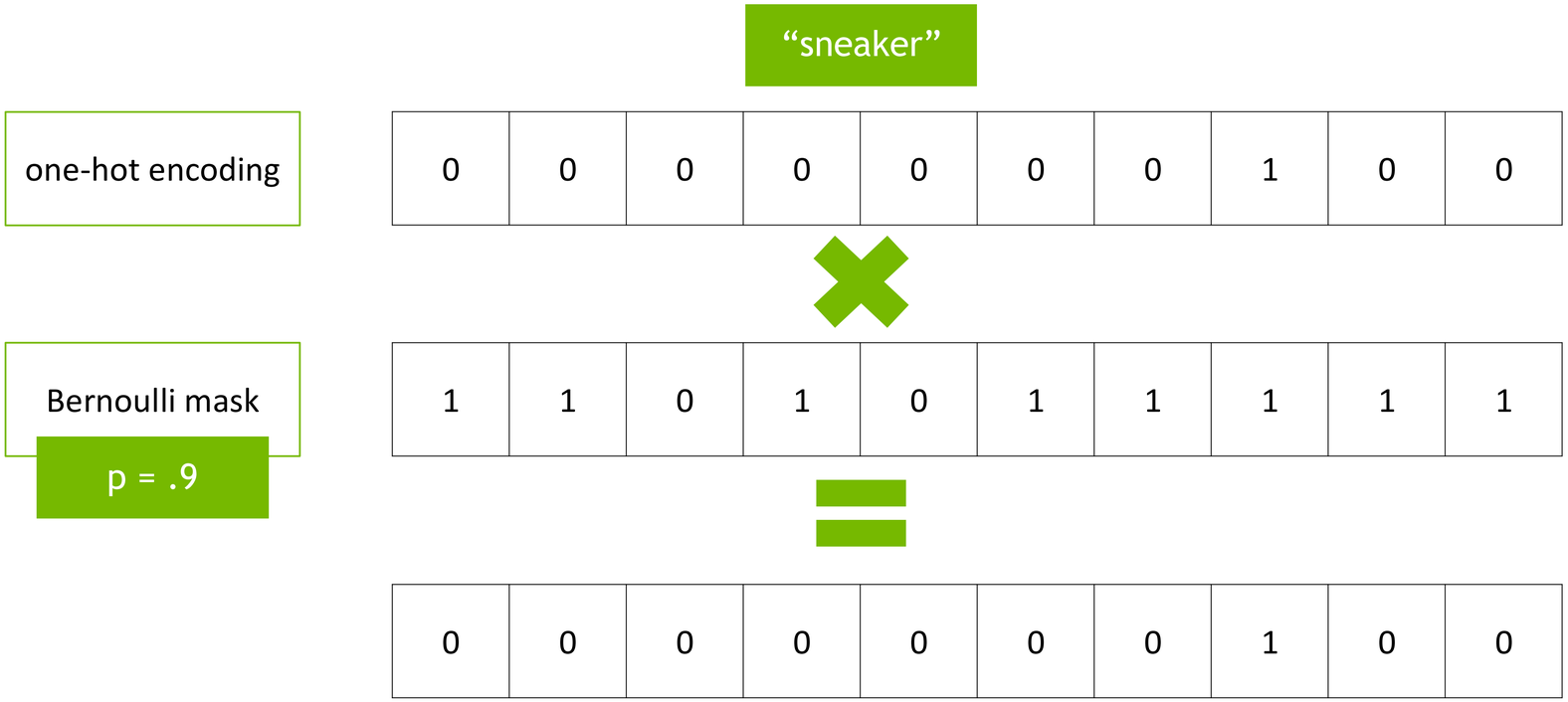
”

The Final Model

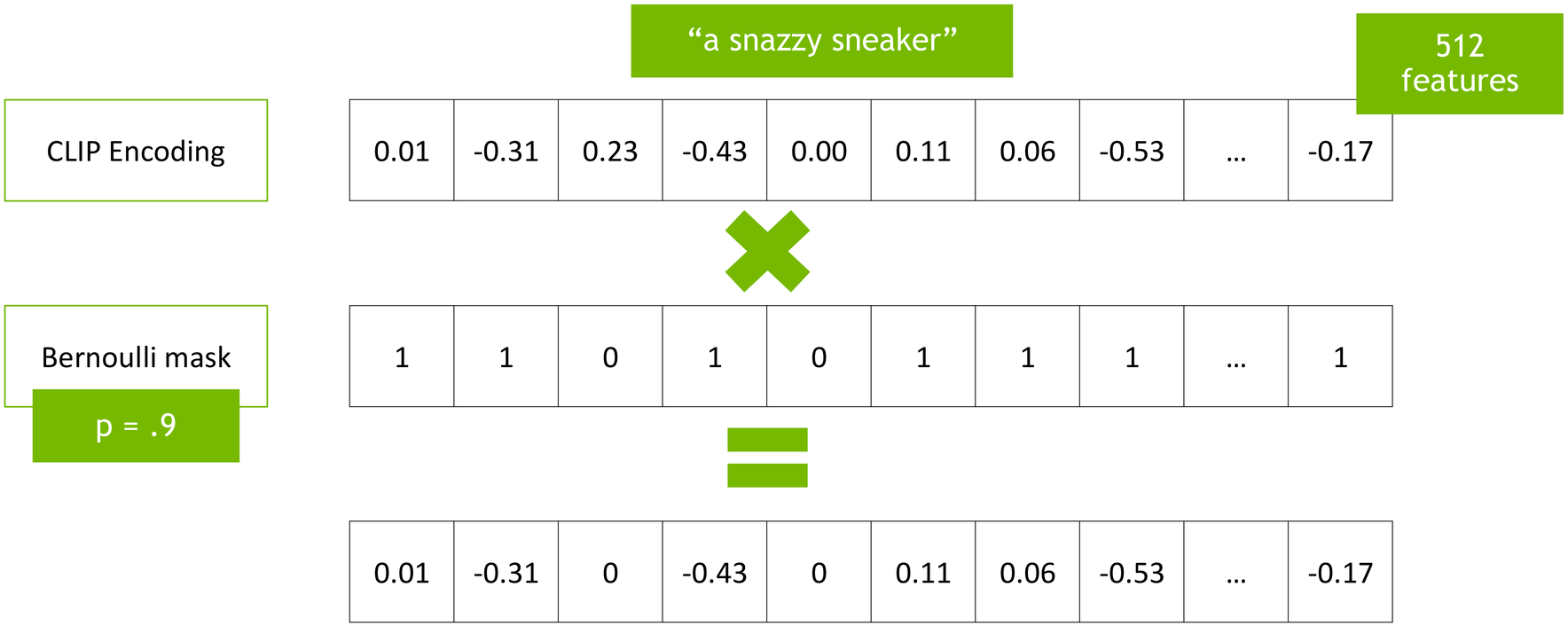


NVIDIA.

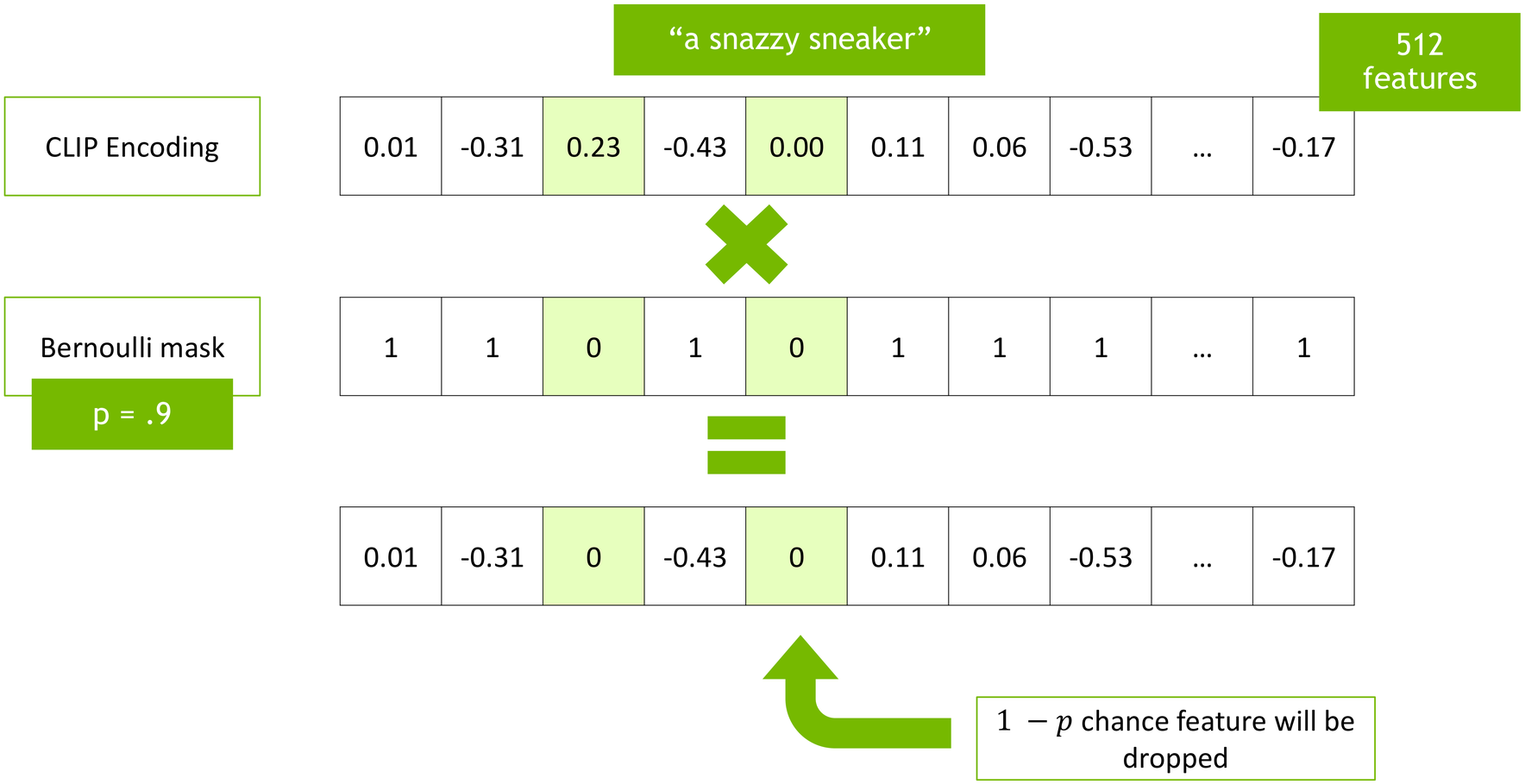
From Class to Context



From Class to Context

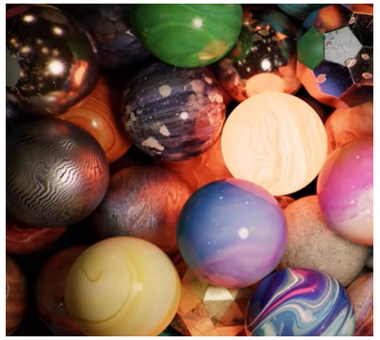


From Class to Context



Experimenting with CLIP

Training



CLIP

[-0.13, …, 0.00]

Prediction

“A bunch of different marbles”

CLIP

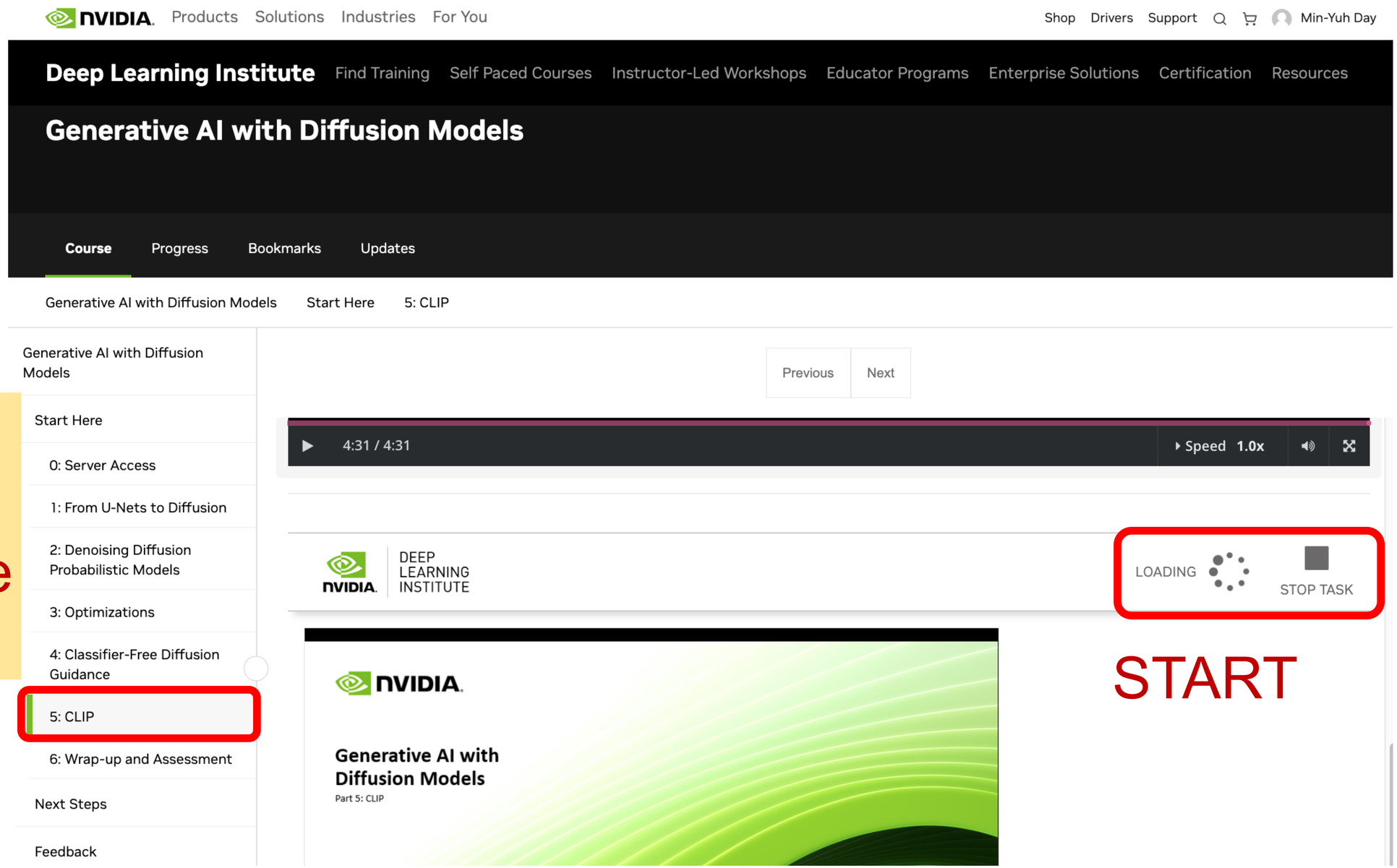
[-0.17, …, -0.11]

NVIDIA.

Let’s get started!

5: CLIP (Contrastive Language-Image Pre-Training)

Generative AI with Diffusion Models

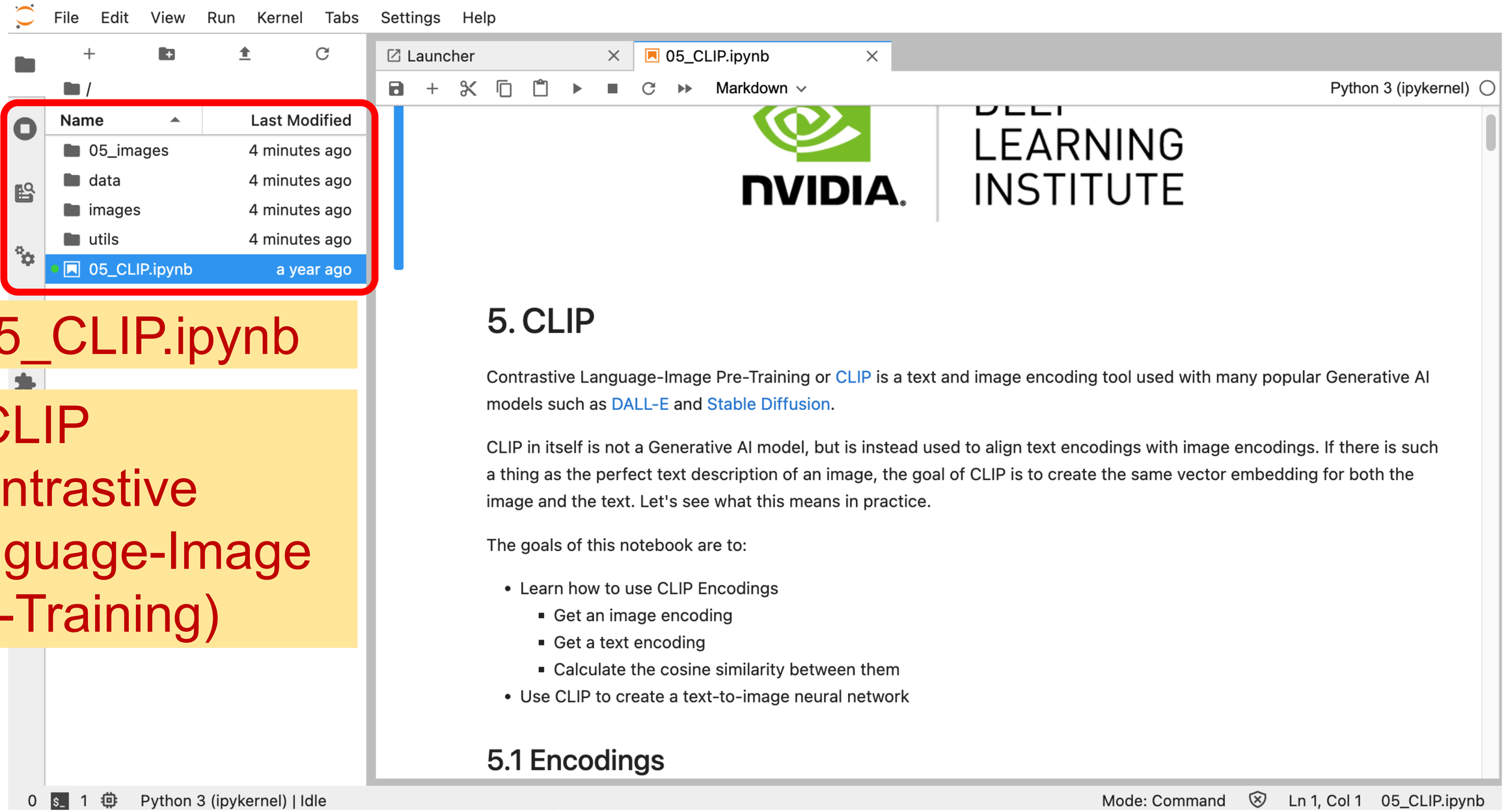


5: CLIP (Contrastive Language-Image Pre-Training)

START

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models



05_CLIP.ipynb

5: CLIP (Contrastive Language-Image Pre-Training)

Generative AI with Diffusion Models

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance (CFG)
- Part 5: CLIP (Contrastive Language-Image Pre-Training)
- Part 6: Wrap-up & Assessment

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Generative AI with Diffusion Models

Part 6: Wrap-up & Assessment

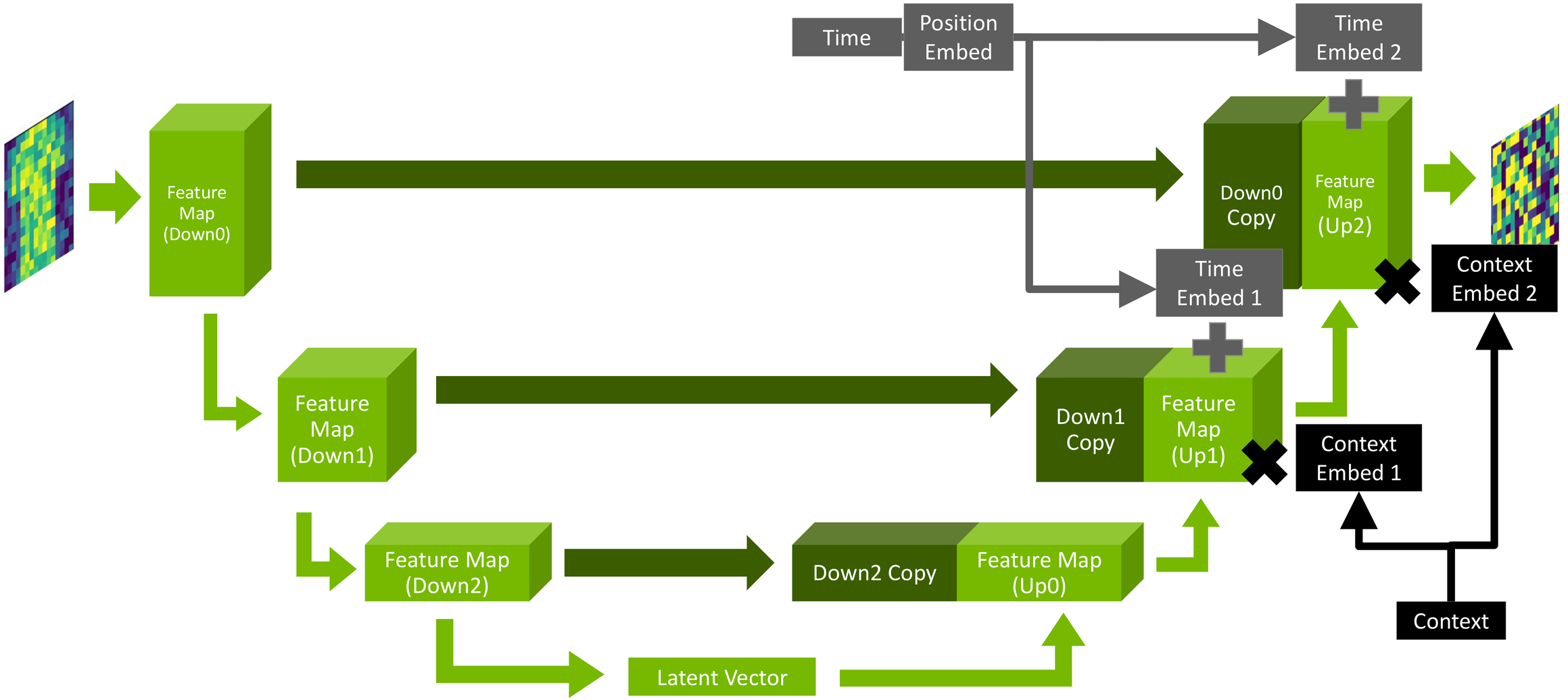
Agenda

- Part 1: From U-Nets to Diffusion
- Part 2: Denoising Diffusion Probabilistic Models
- Part 3: Optimizations
- Part 4: Classifier-Free Diffusion Guidance
- Part 5: CLIP
- Part 6: Wrap-up & Assessment

NVIDIA.

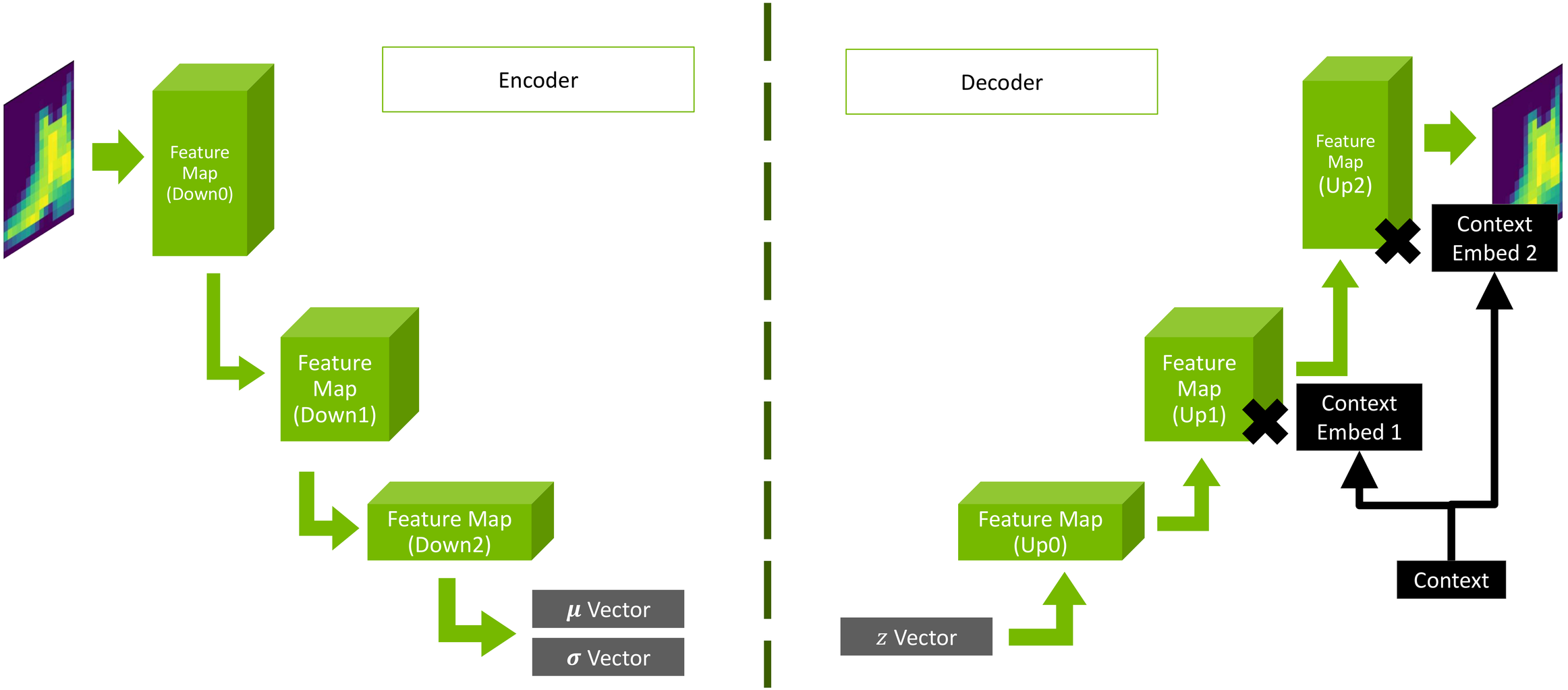
Variational Autoencoders (VAE)

U-Net


Time
Position Embed
Time Embed 2
Feature Map (Down0)
Down0 Copy
Feature Map (Up2)
Time Embed 1
Context Embed 2
Feature Map (Down1)
Down1 Copy
Feature Map (Up1)
Context Embed 1
Feature Map (Down2)
Down2 Copy
Feature Map (Up0)
Context
Latent Vector

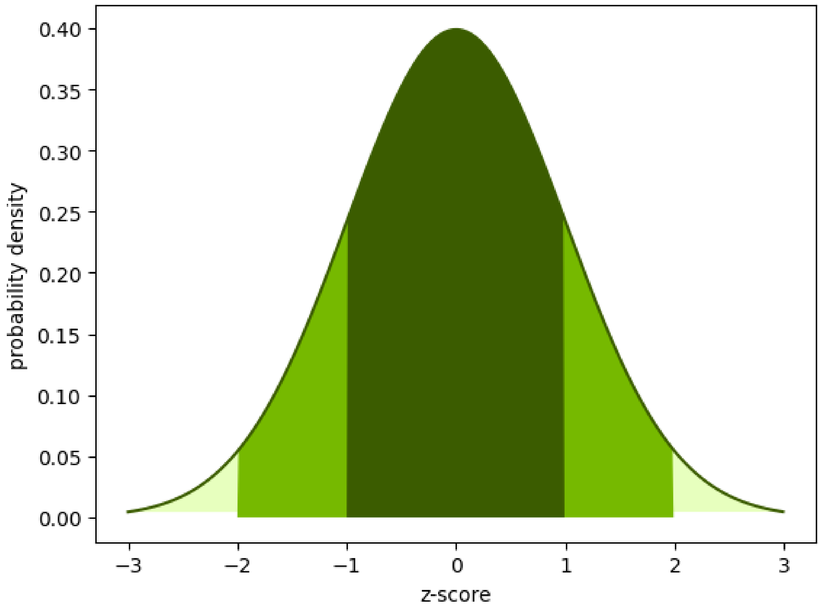
NVIDIA.

Variational Autoencoder

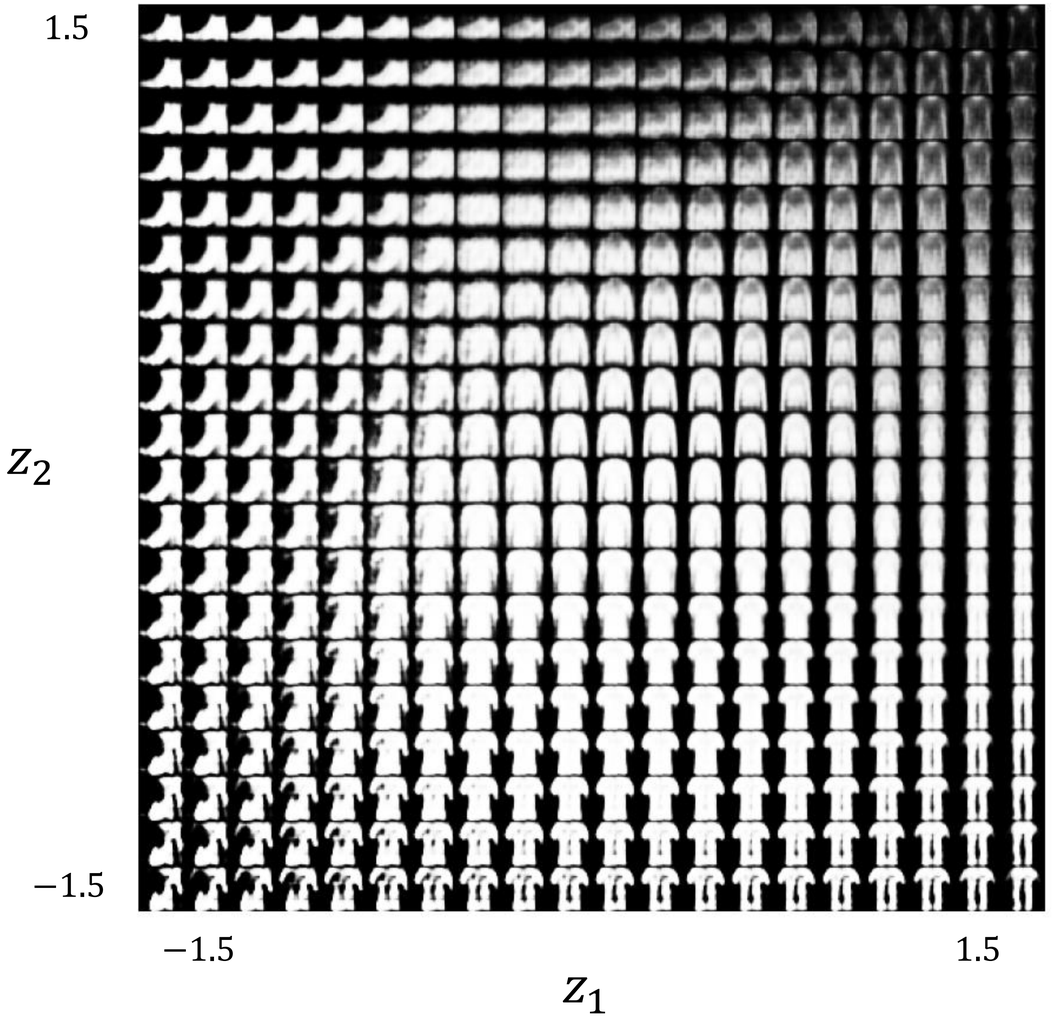


NVIDIA.

Variational Autoencoder

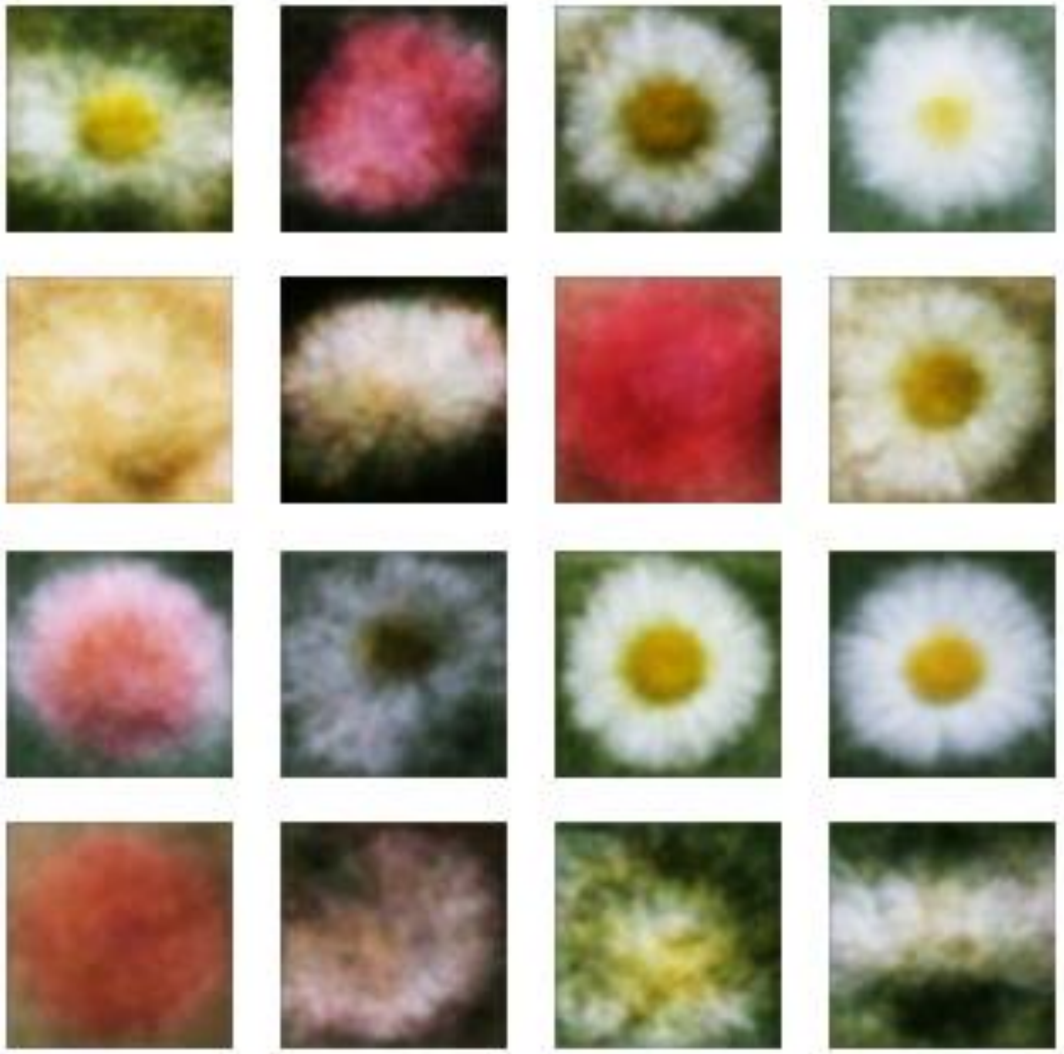


FashionMNIST
2 Latent Variables



NVIDIA.

Variational Autoencoder



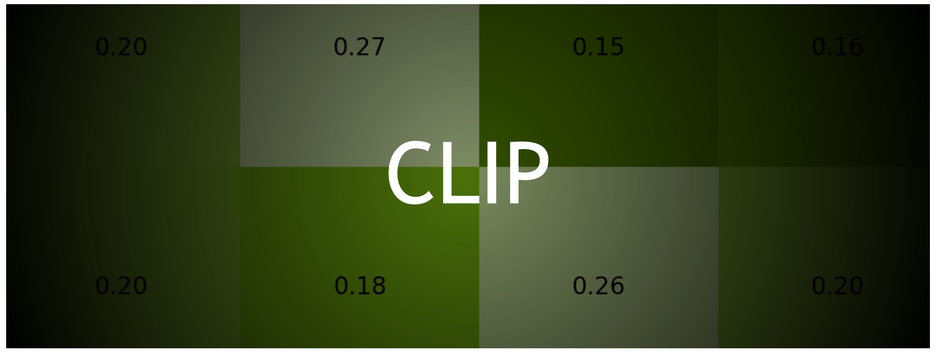
Modified TF Flowers
20 Latent Variables
16 Random Images

NVIDIA.

State of the Art Models

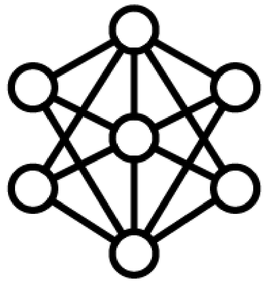
State of the Art Models

All of them are used



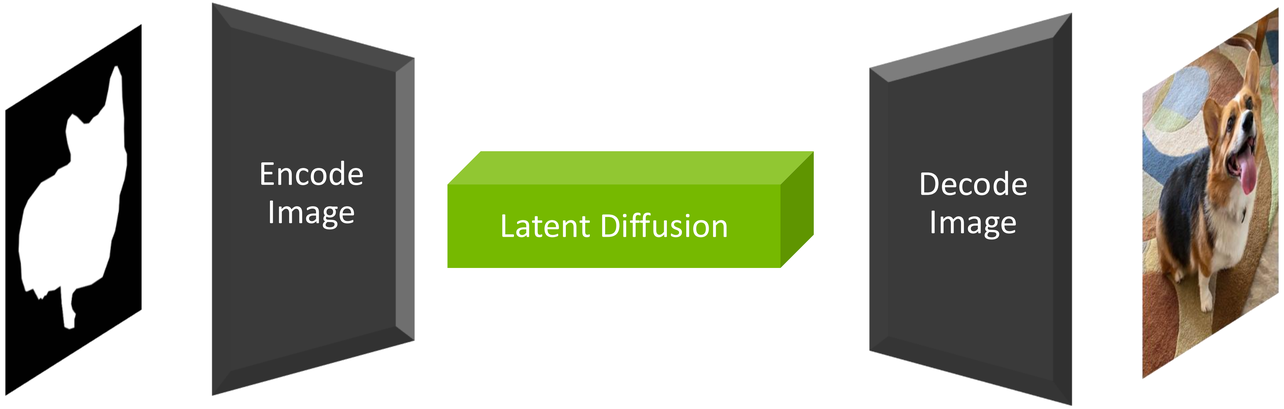
NVIDIA.

Latent Diffusion



VAE

Diffusion



Tiled Upsampling

Transposed Convolution Review

Kernel

.25	.25
.25	.25

Image

1	0	1
0	1	0
1	0	1

Output

.25	0	0	.25
0	.25	.25	0
0	.25	.25	0
.25	0	0	.25

NVIDIA.

Tiled Upsampling

Convolving with a Neural Network

Kernel?

Image

1	0	1
0	1	0
1	0	1

Output

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

NVIDIA.

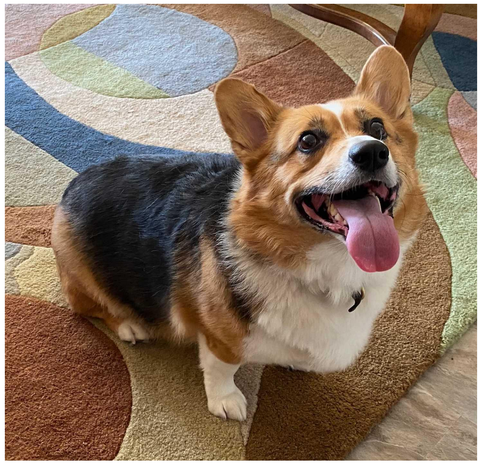
Tiled Upsampling

Convolving with a Neural Network

Kernel?

Image

Output



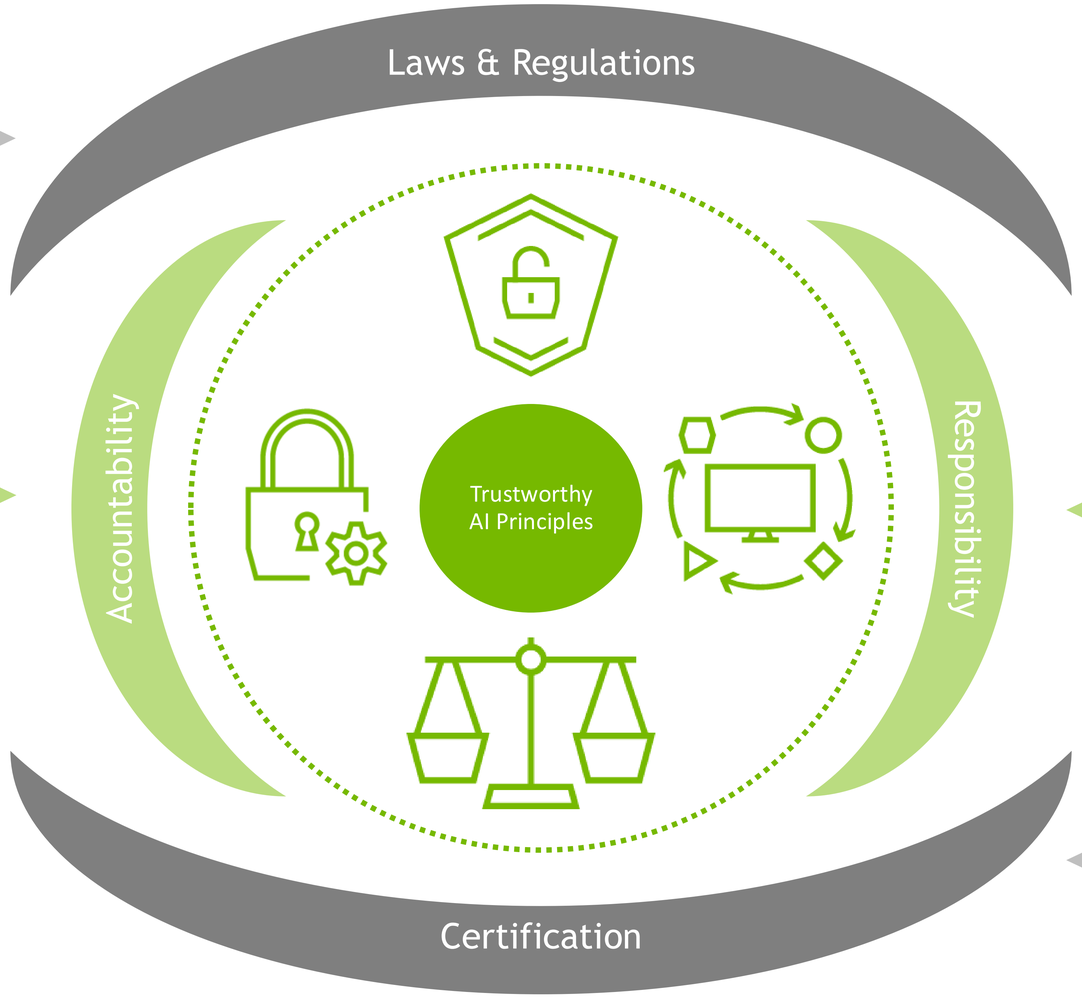
NVIDIA.

Trustworthy AI

Foundations of Trustworthiness

An AI model complies with relevant laws, regulations, and standards for the region and industry where it is deployed

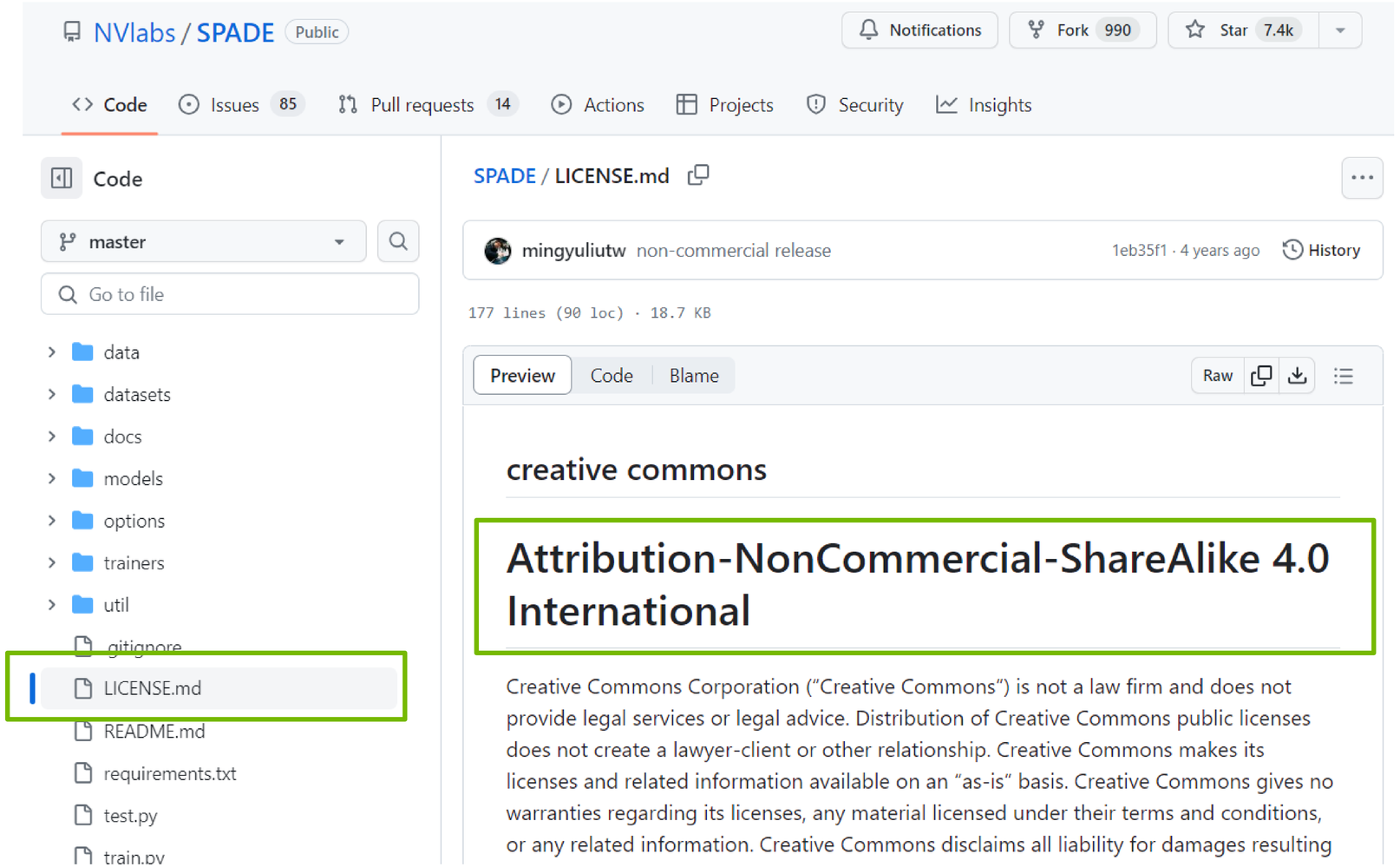
Accountability means being answerable for the impact of an AI on the societal system that uses it and being thoughtful about how it might be misused



Responsibility can be legal or moral. Legal responsibility requires a violation of a law or regulation. Moral responsibility occurs from an action or an intentional failure to act.

An AI model has been certified as fit for purpose as required by the region or industry where it is deployed

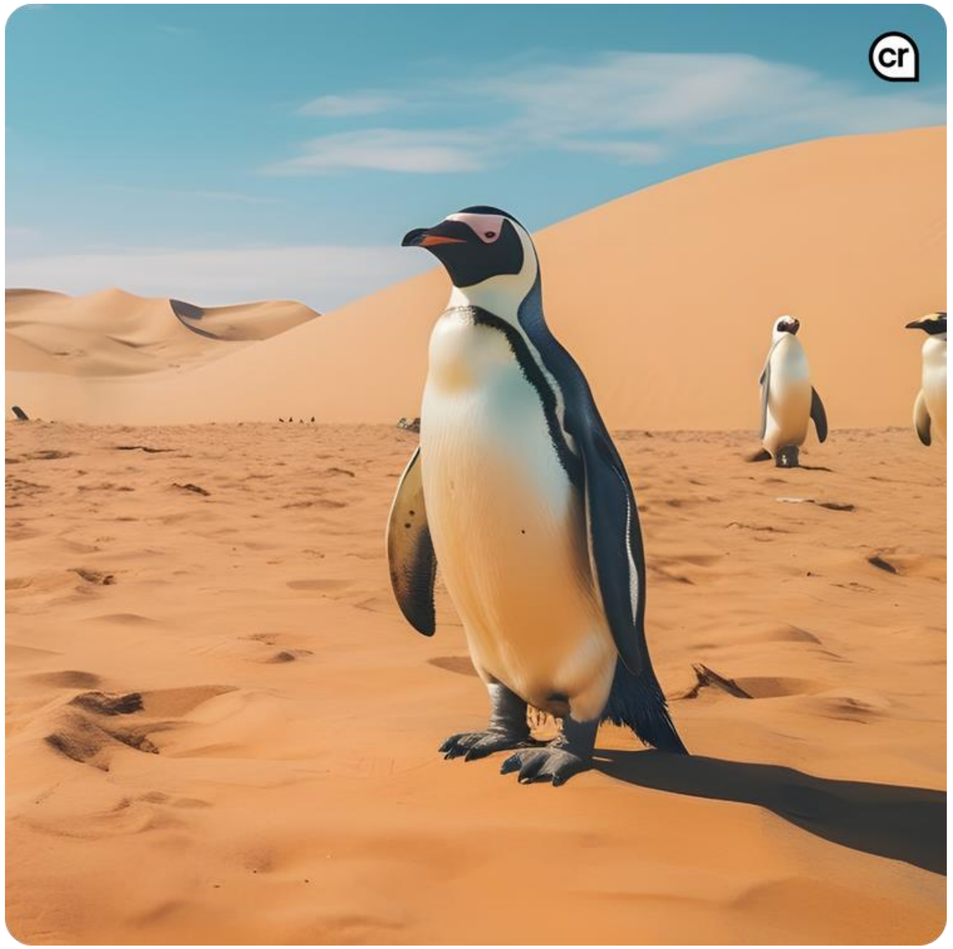
Check the License



Many code repositories and datasets have a license file which describe the use limitations such as:

- Whether the asset can be used for commercial purposes
- How the authors would like to be recognized or attributed

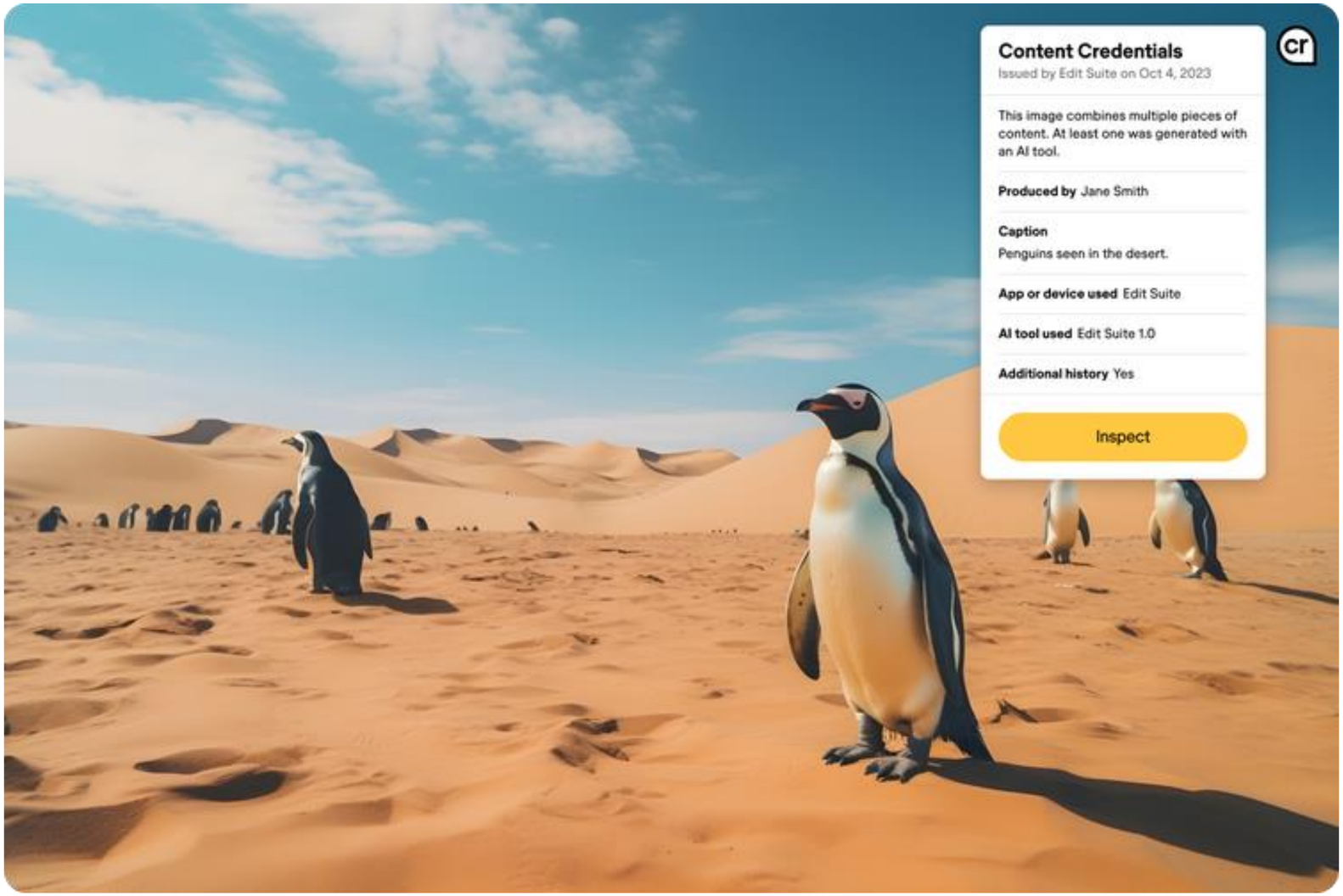
Trustworthy AI Tools



Content Authenticity Initiative

content credentials

Trustworthy AI Tools



All content generated with Adobe Firefly is clearly labeled and can be inspected

content credentials

Select another file from your device or drag and drop anywhere

Generated image
Sep 27, 2023

Search for possible matches

Generated image
Sep 27, 2023

Compare

Change language

Generated image

Sep 27, 2023

Content summary

This image was generated with an AI tool.

Credit and usage

The producer chose to share the following info:

AI tool used

Adobe Firefly Pipeline

Process

The app or device used to produce this content recorded the following info:

App or device used

Adobe Firefly 1.0

Actions

Created

content credentials

Inspect content to dig deeper

Drag content into Verify to inspect its Content Credentials in detail and see how it has changed over time.

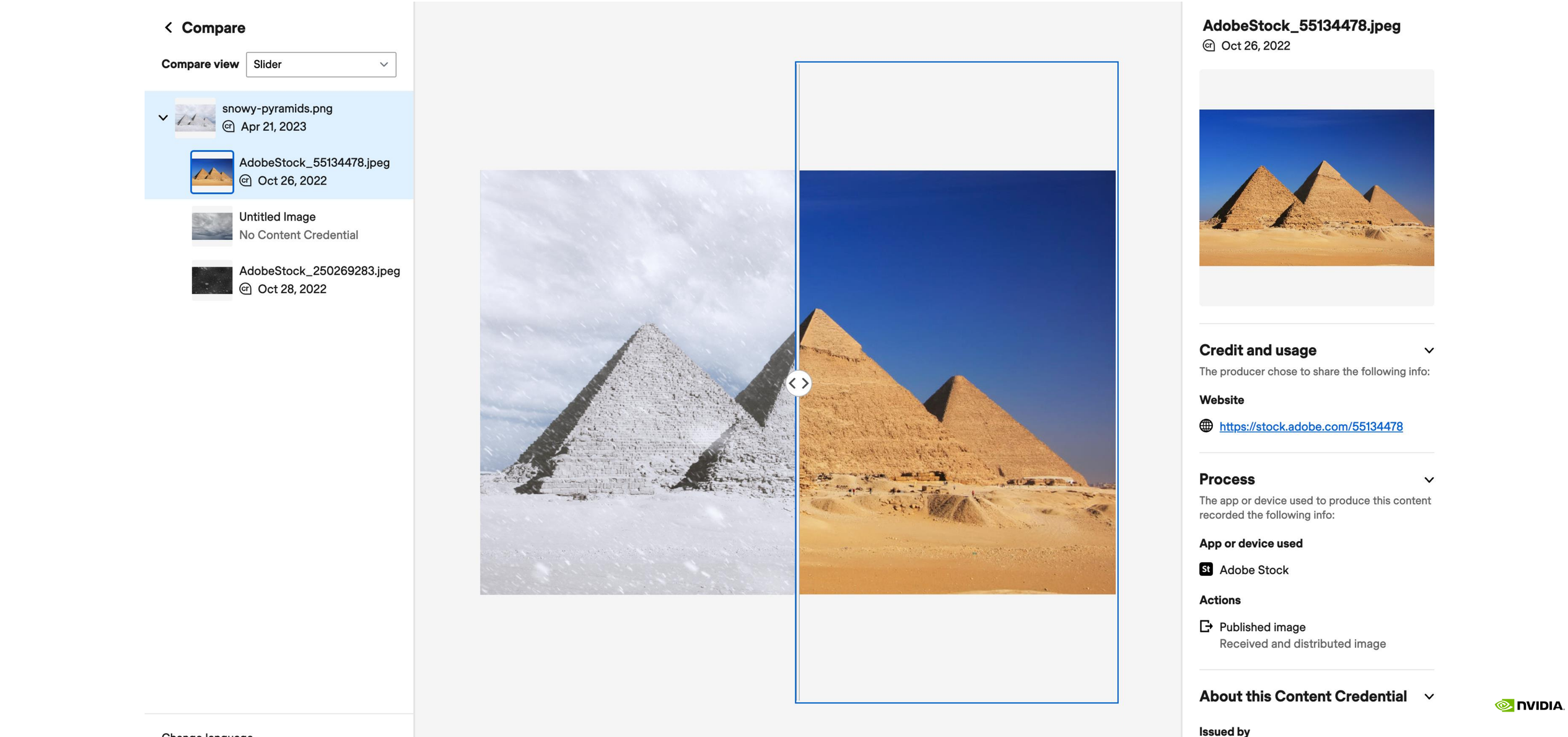
Content Credentials are still rolling out, so the content you choose to inspect may not have information to view.

Select a file from your device
or drag and drop anywhere

Supported formats: AVIF, DNG, HEIC, JPEG, M4A, MP4, PNG, SVG, TIFF, WAV, WebP

Change language

The Verify inspection tool allows us see the provenance of a digital file

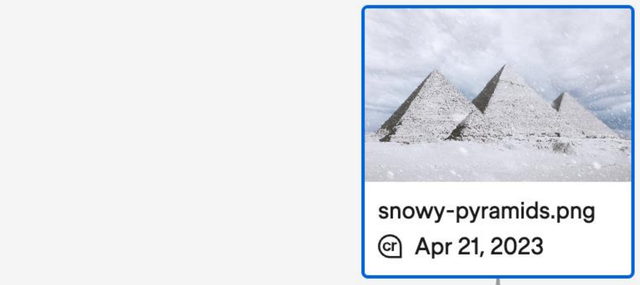


content credentials

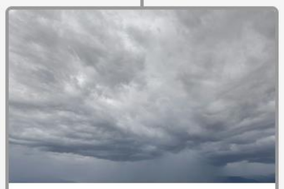
Select another file from your device or drag and drop anywhere

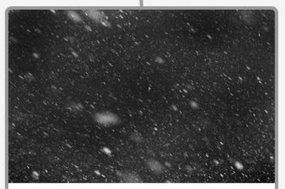
snowy-pyramids.png
Apr 21, 2023

Search for possible matches


snowy-pyramids.png
Apr 21, 2023

AdobeStock_5513447...
Oct 26, 2022


Untitled Image
No Content Credential


AdobeStock_2502692..
Oct 28, 2022

Compare

Change language

snowy-pyramids.png
Apr 21, 2023

Content summary

This image was generated with an AI tool.

Credit and usage

The producer chose to share the following info:

Produced by

Pia Blumenthal

Social media accounts

Pia Blumenthal
Behance

pia___b
Twitter

AI tool used

undefined

The Content Credentials pin will appear next to content online, indicating that provenance information is present.

Content
Authenticity
Initiative

Adobe Firefly

Adobe Firefly

https://firefly.adobe.com/

Text to Image

Generate images from a detailed text description.

Create with Firefly

Generative Fill

Use a brush to remove objects or paint in new ones.

Create with Firefly

Text Effects

Apply styles or textures to words and phrases.

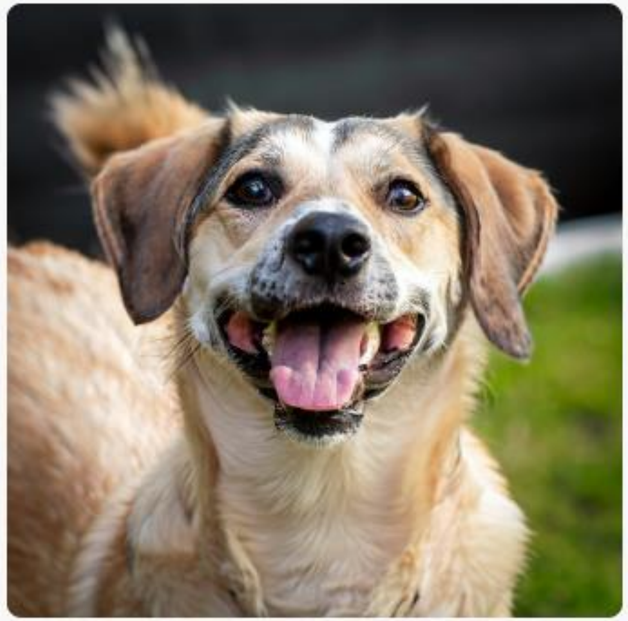
Create with Firefly

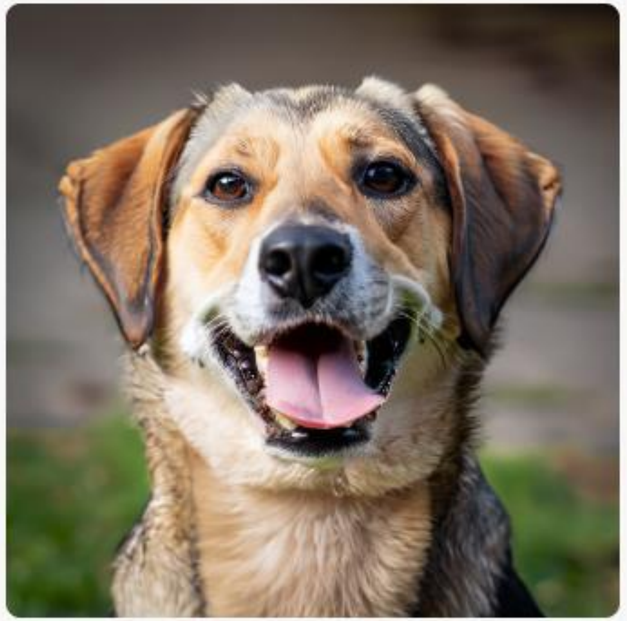
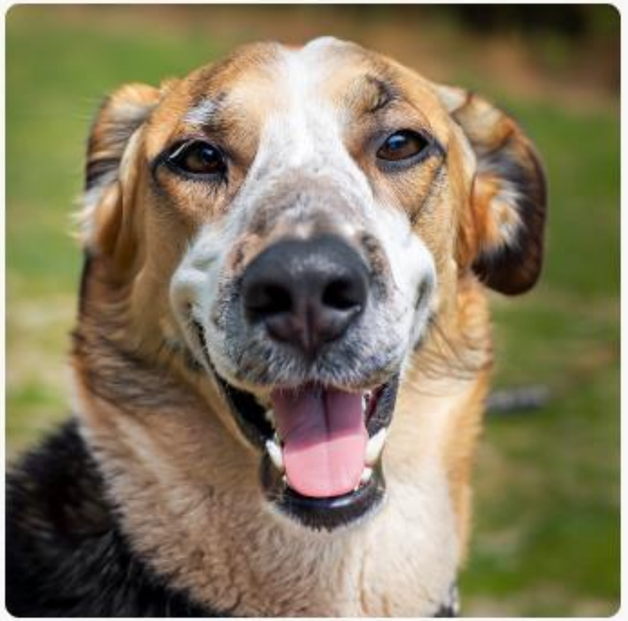
NVIDIA.

Prompt Engineering

State-of-the-art Models

Better understanding the vocabulary the model was trained will make it easier to influence the results.



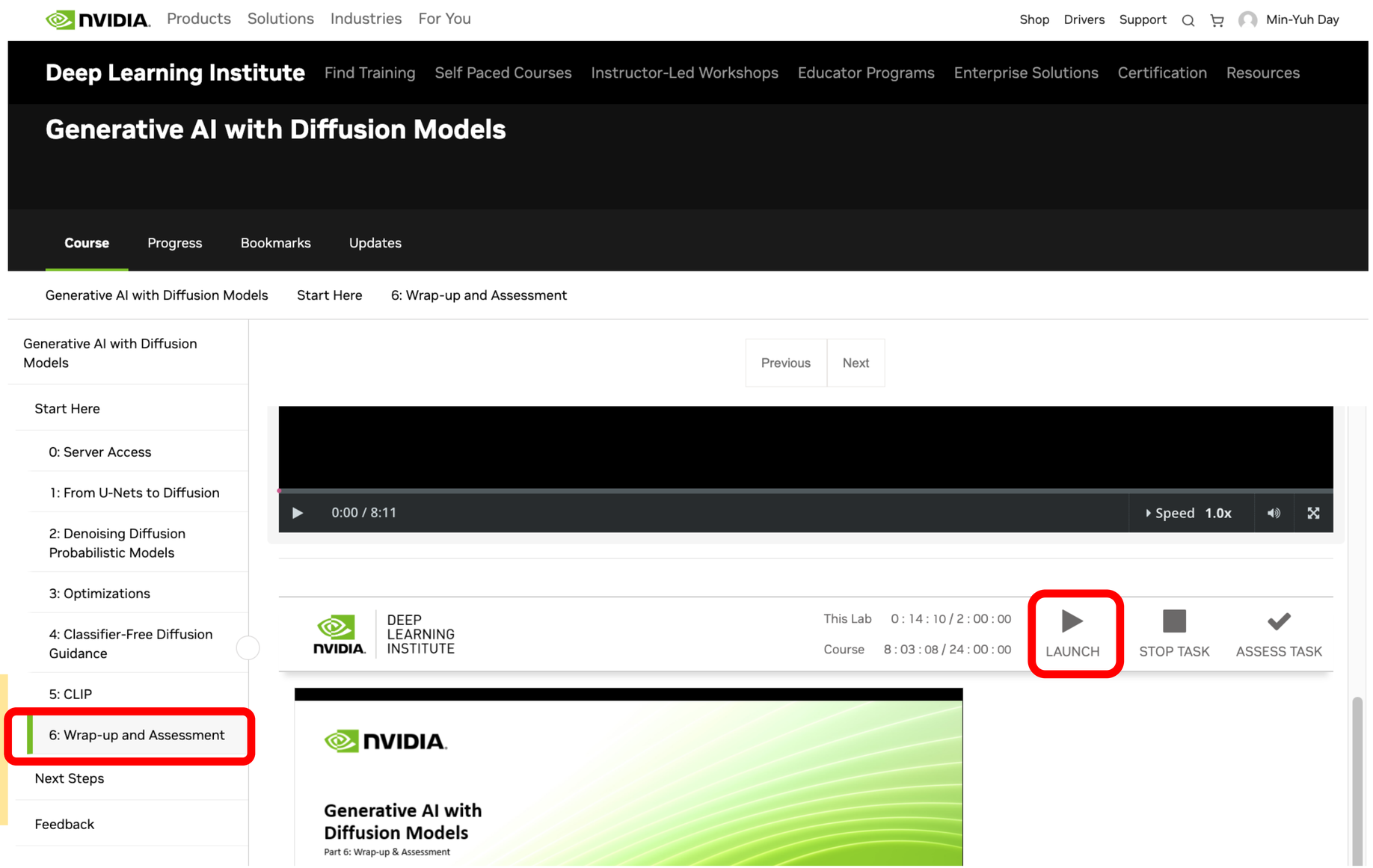
“Portrait of a happy canine with a 17mm lens”

NVIDIA.

Good Luck!

6: Wrap-up and Assessment

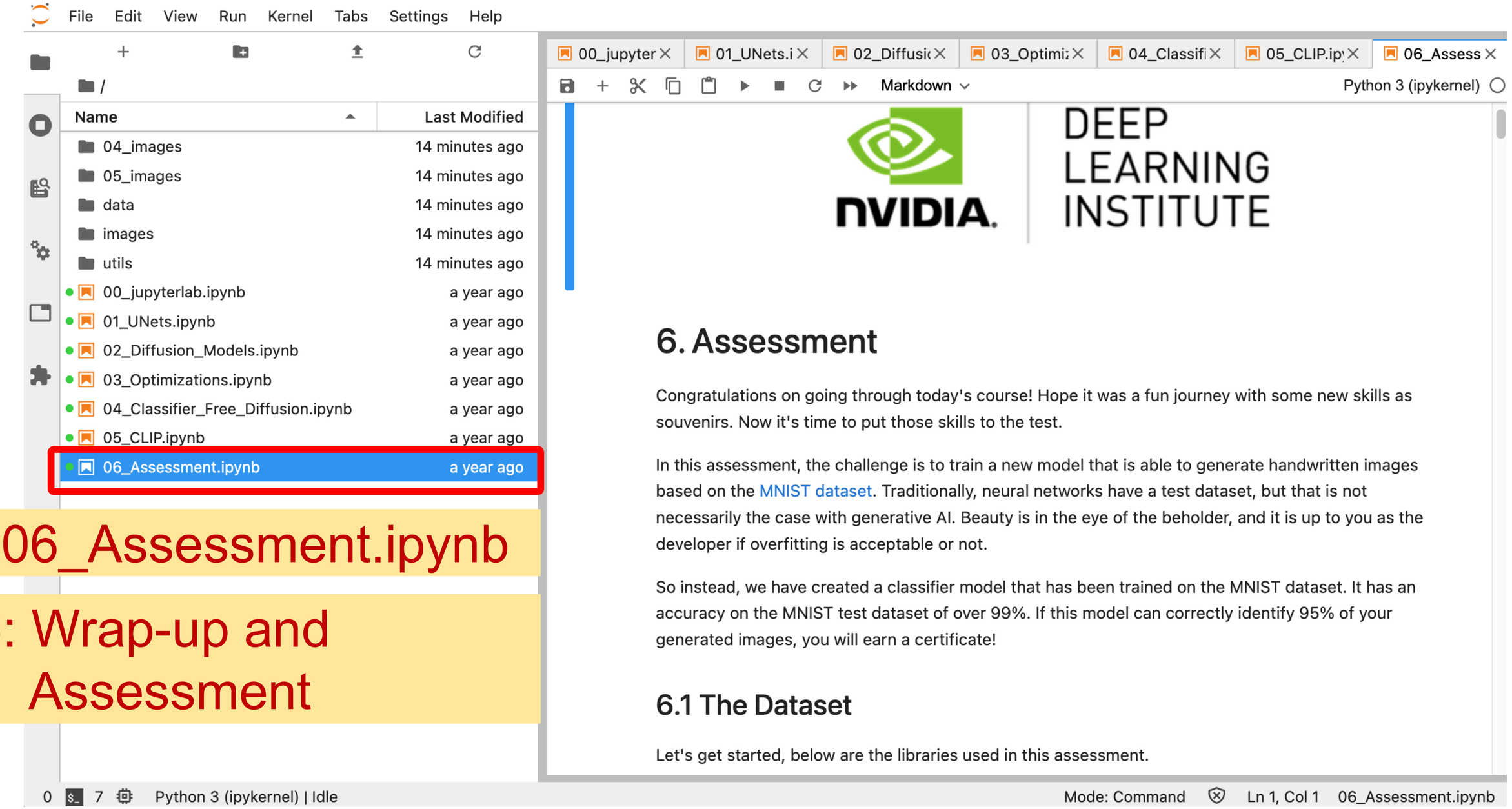
Generative AI with Diffusion Models



6: Wrap-up and Assessment

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



06_Assessment.ipynb

6: Wrap-up and Assessment

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models

00_jupyterlab.ipynb | 01_UNets.ipynb | 02_Diffusion_Models. | 03_Optimizations.ipyn | 04_Classifier_Free_Di | 05_CLIP.ipynb | 06_Assessment.ipynb

Markdown | Python 3 (ipykernel)

06_Assessment.ipynb

```
[31]: model.eval()
      w = 2.0  #0.0  # Change me
      c = torch.arange(N_CLASSES).to(device)
      c_drop_prob = 0
      c_hot, c_mask = get_context_mask(c, c_drop_prob)

      x_0 = sample_w(model, c_hot, 1.0)
      other_utils.to_image(make_grid(x_0.cpu(), nrow=N_CLASSES))
```

[31]:

This is important for the automated grader. Is the output shape `[10, 1, 28, 28]` ? If yes, you're ready to test the model!

```
[32]: x_0.shape
```

```
[32]: torch.Size([10, 1, 28, 28])
```

https://tinyurl.com/nvdligaidms

6.6 Run the Assessment

To assess your model run the following two cells.

NOTE: `run_assessment` assumes your model is named `model` and your diffusion weight is called `w`. If for any reason you have modified these variable names, please update the names of the arguments passed to `run_assessment`. If your results are close, but not quite there, try changing `w` above.

```
[33]: from run_assessment import run_assessment
```

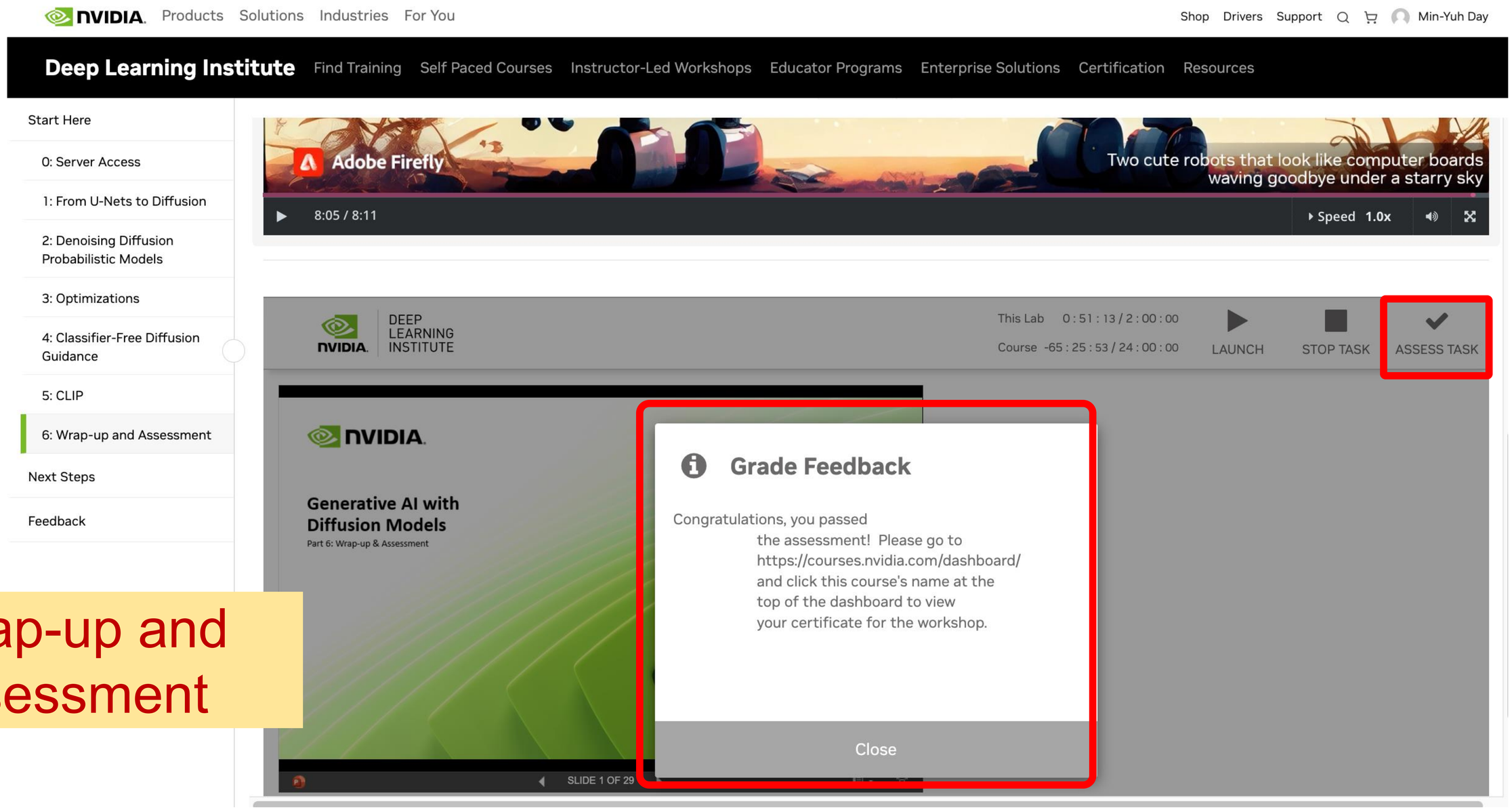
```
[34]: run_assessment(model, sample_w, w)
```

```
Evaluating model...

Classifier prediction for generated `0`: [0, 0, 0, 0, 0] - 100.0% Accuracy
Classifier prediction for generated `1`: [1, 1, 1, 1, 1] - 100.0% Accuracy
Classifier prediction for generated `2`: [2, 2, 2, 2, 2] - 100.0% Accuracy
Classifier prediction for generated `3`: [3, 3, 3, 3, 3] - 100.0% Accuracy
Classifier prediction for generated `4`: [4, 4, 4, 4, 4] - 100.0% Accuracy
Classifier prediction for generated `5`: [5, 5, 5, 5, 5] - 100.0% Accuracy
Classifier prediction for generated `6`: [6, 6, 6, 6, 6] - 100.0% Accuracy
Classifier prediction for generated `7`: [7, 7, 7, 7, 7] - 100.0% Accuracy
Classifier prediction for generated `8`: [8, 8, 8, 8, 8] - 100.0% Accuracy
Classifier prediction for generated `9`: [9, 9, 9, 9, 9] - 100.0% Accuracy
Final Accuracy - 100.0%

Accuracy required to pass the assessment is 95% or greater.
Congratulations! You passed the assessment!
See instructions below to generate a certificate.
```

Saving completed | Mode: Command | Ln 1, Col 1 | 06_Assessment.ipynb

6: Wrap-up and Assessment

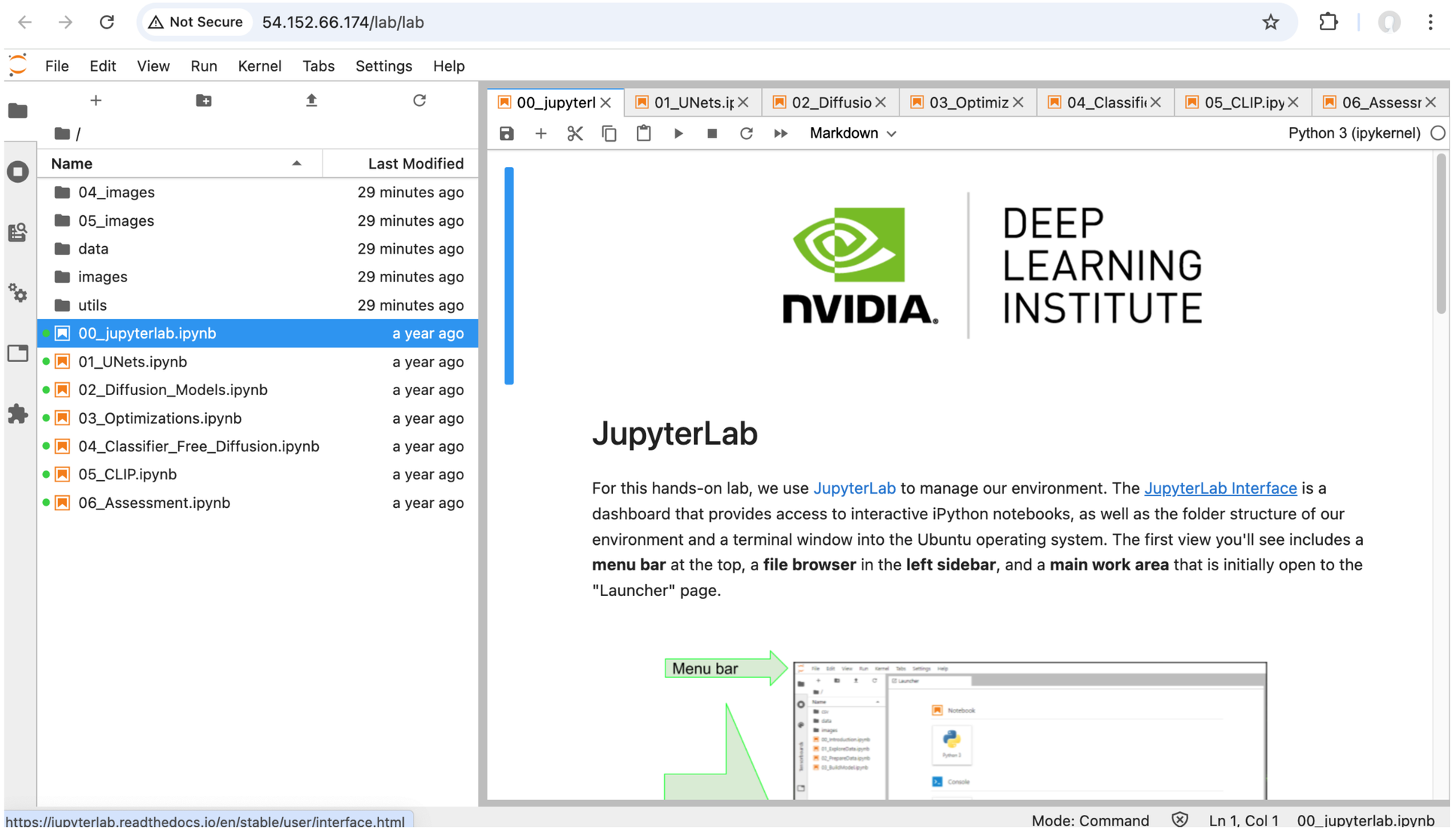
NVIDIA Generative AI with Diffusion Models



6: Wrap-up and Assessment

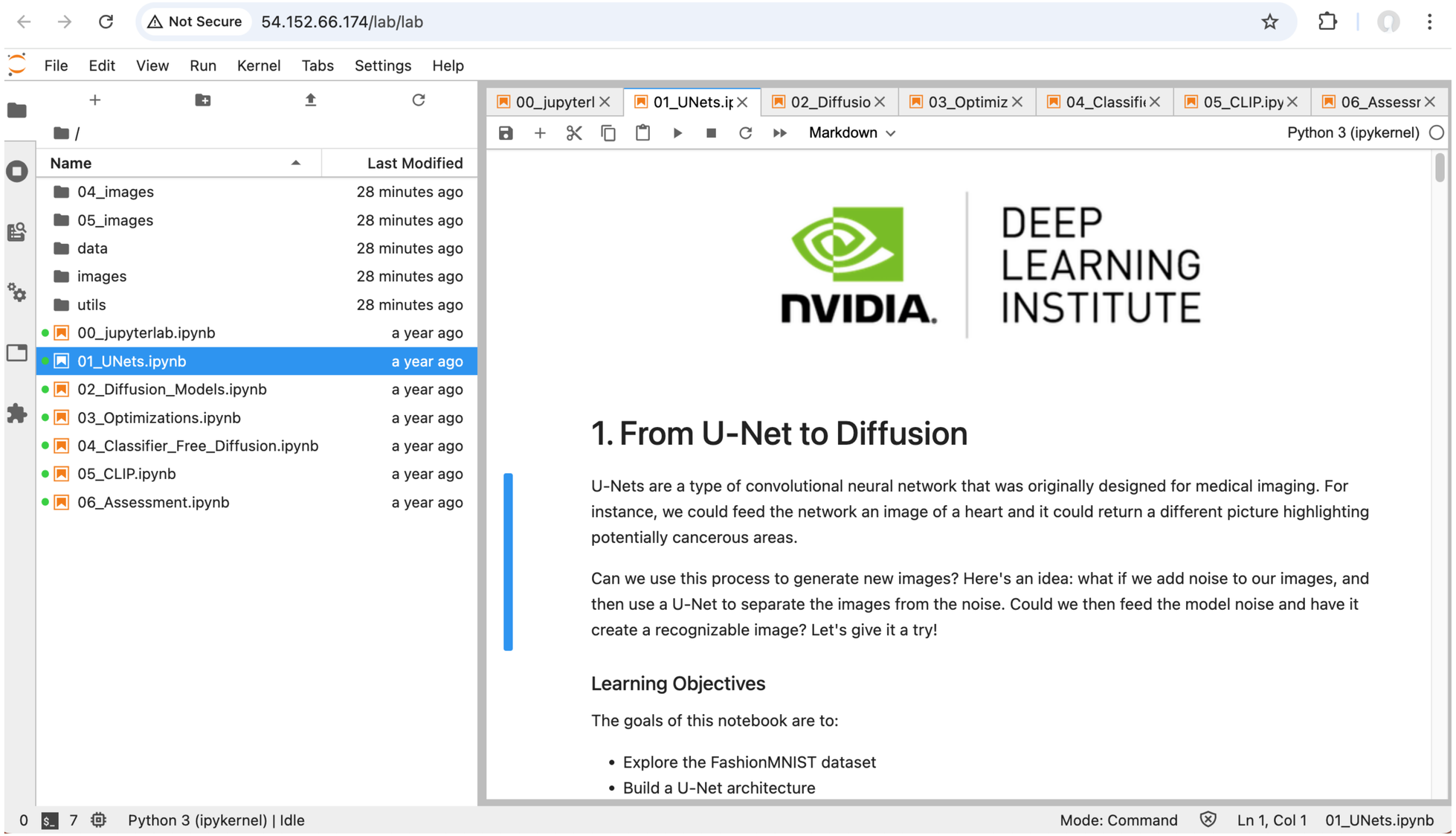
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



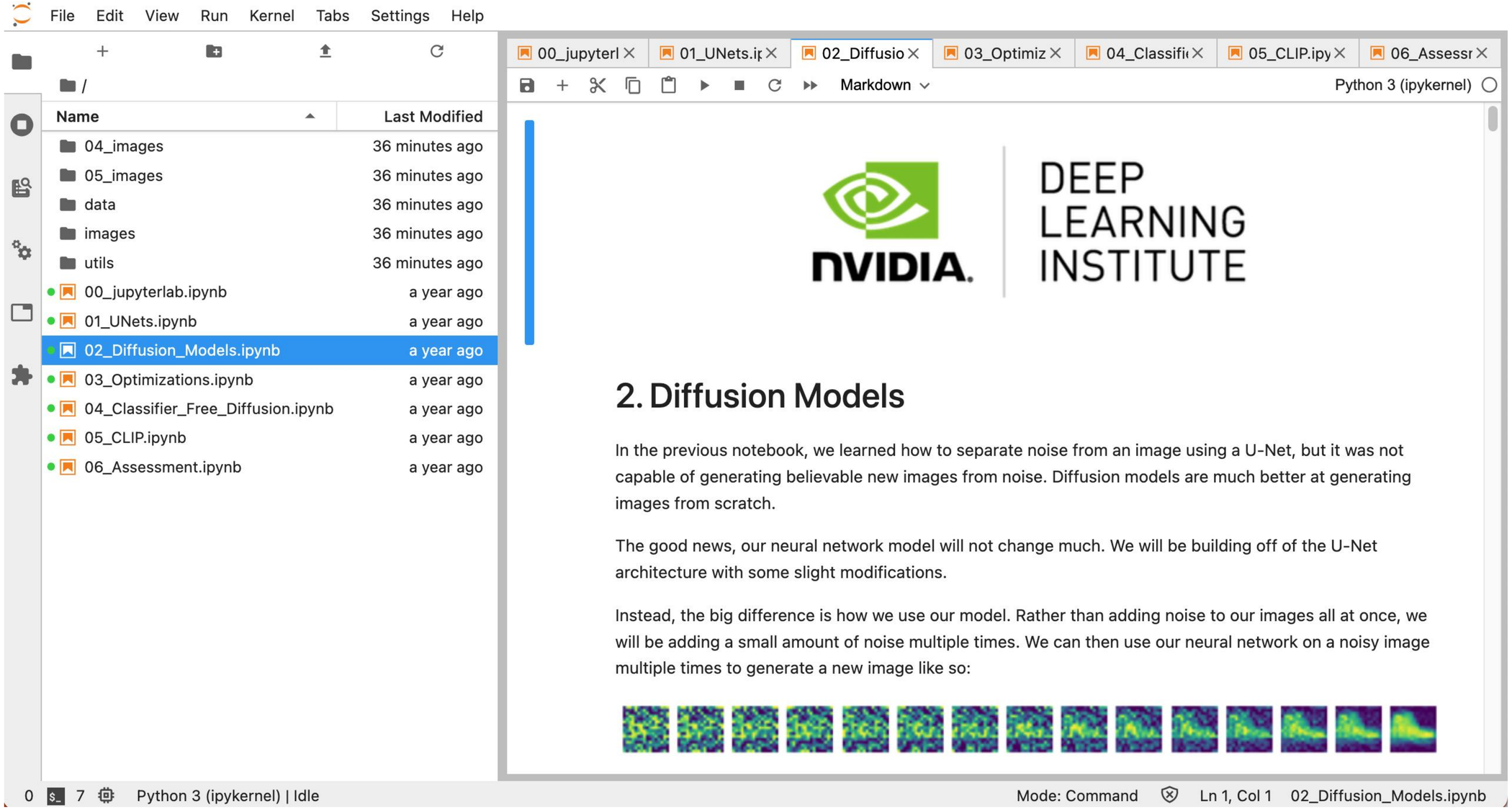
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



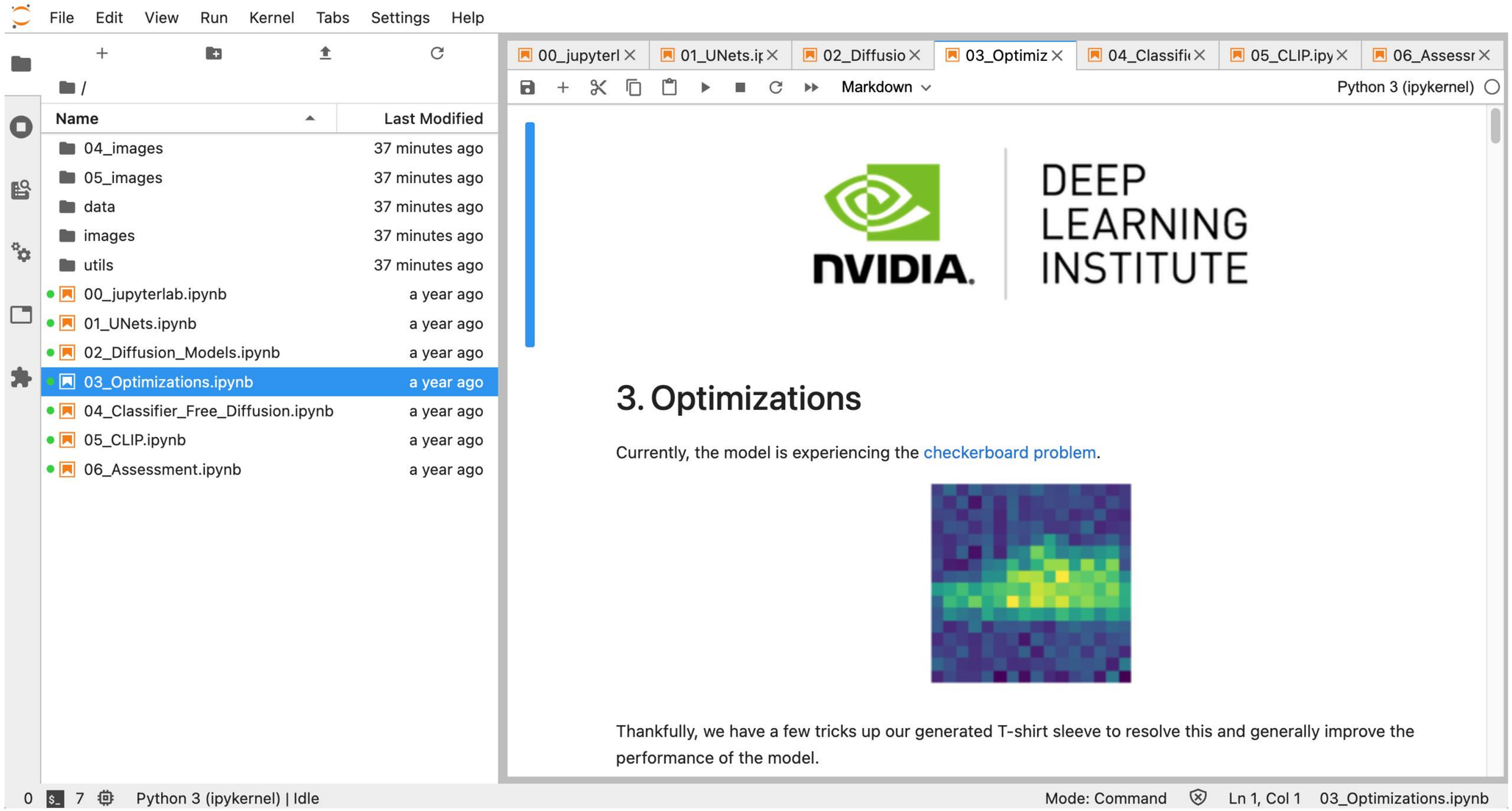
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



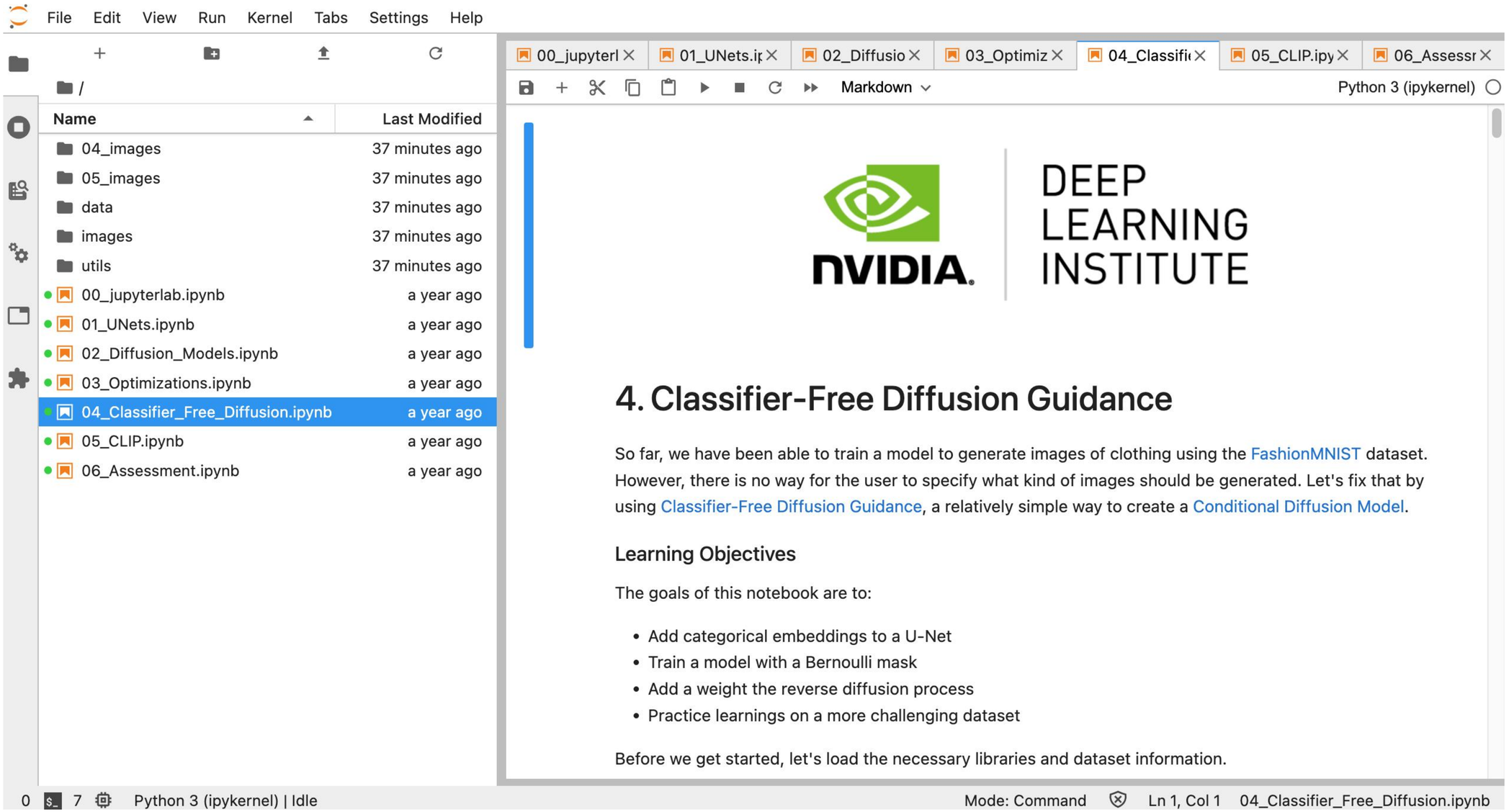
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



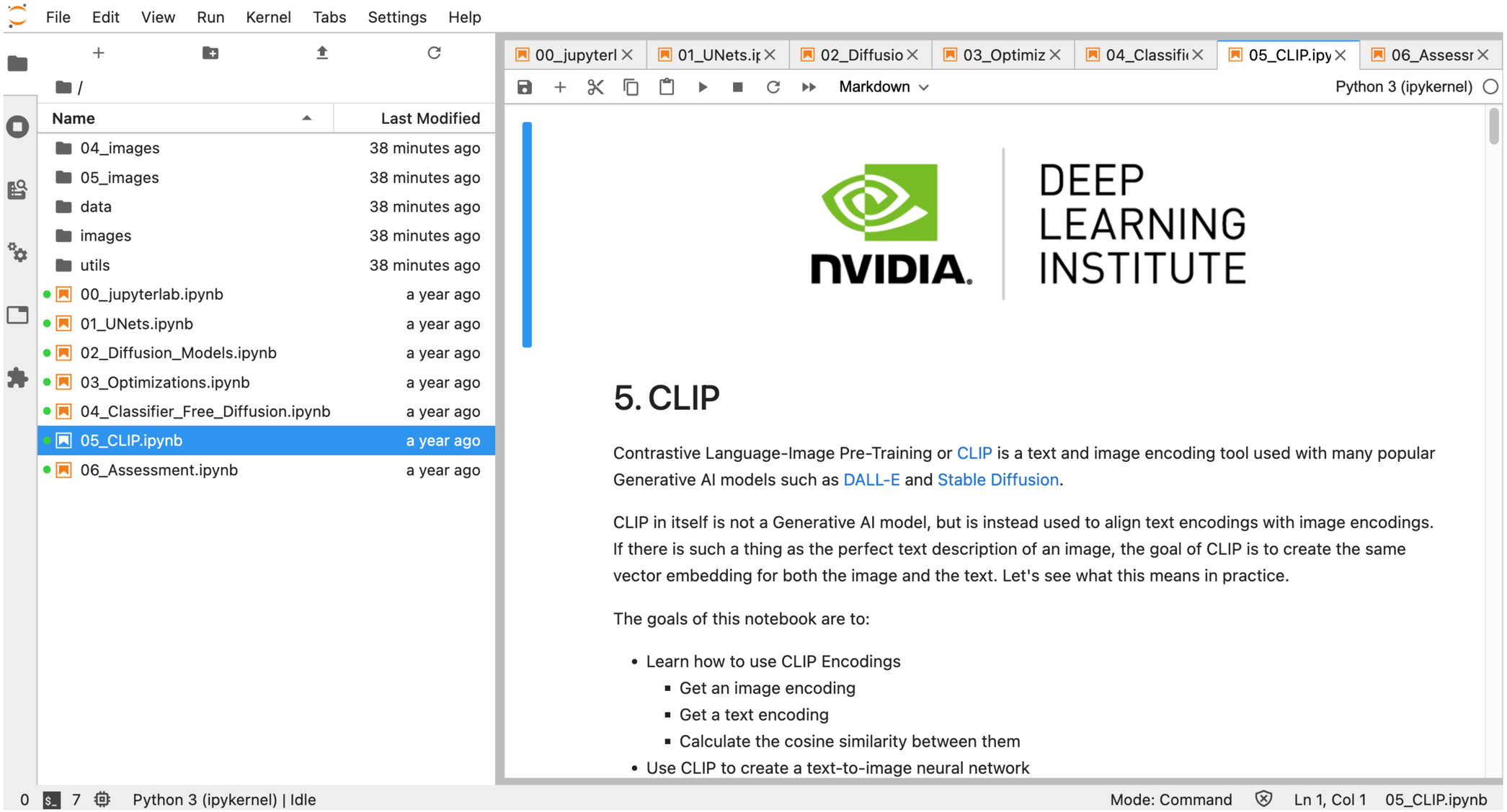
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



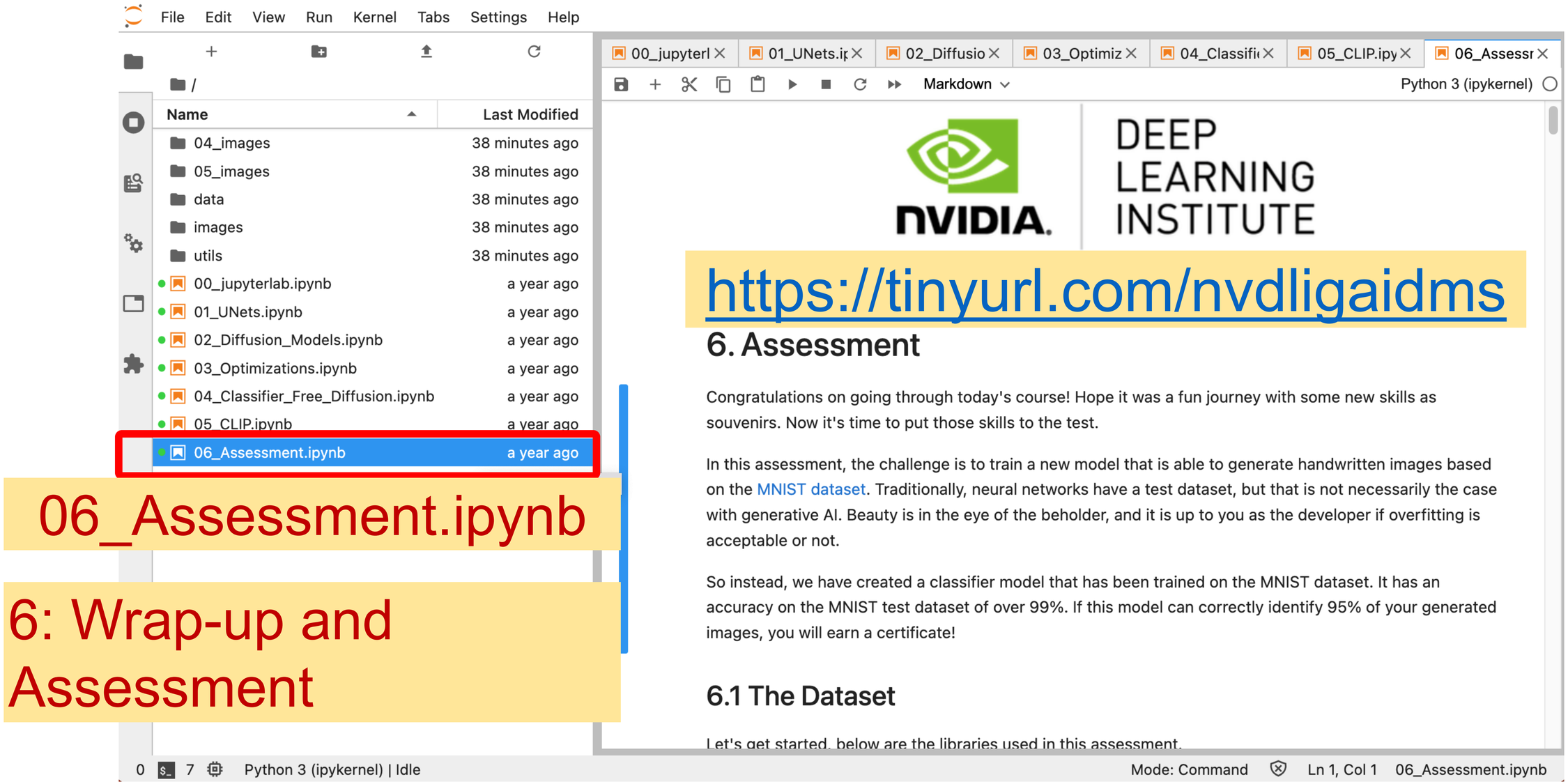
Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA Generative AI with Diffusion Models



06_Assessment.ipynb

6: Wrap-up and Assessment

Source: NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

NVIDIA

Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Generative AI with Diffusion Models

Greg Estes
Vice President, NVIDIA

Issue Date: : February 28, 2025
Certification ID: q3oo-oBhTQKtyCCote2E-Q

https://learn.nvidia.com/certificates?id=q3oo-oBhTQKtyCCote2E-Q

AI Agents and Large Multimodal Agents (LMAs)

Generative AI, Agentic AI, Physical AI

Physical AI
Self-driving cars
General robotics

Agentic AI
Coding assistants
Customer service
Patient care

Generative AI
Digital marketing
Content creation

Perception AI
Speech recognition
Deep recommender systems
Medical imaging

2012 AlexNet
Deep learning breakthrough

Source: NVIDIA (2025), GTC March 2025 Keynote with NVIDIA CEO Jensen Huang, https://www.youtube.com/watch?v=_waPvOwL9Z8

Generative AI LLMs (2017-2025)

Key Organizations

- Google
- OpenAI
- Meta
- Mistral AI
- Alibaba
- xAI
- Anthropic
- NVIDIA
- BigScience

Key Milestones

2017: Transformer architecture

2018: First-gen GPT, BERT

2020: GPT-3 (175B parameters)

2022: Emergent abilities, instruction tuning

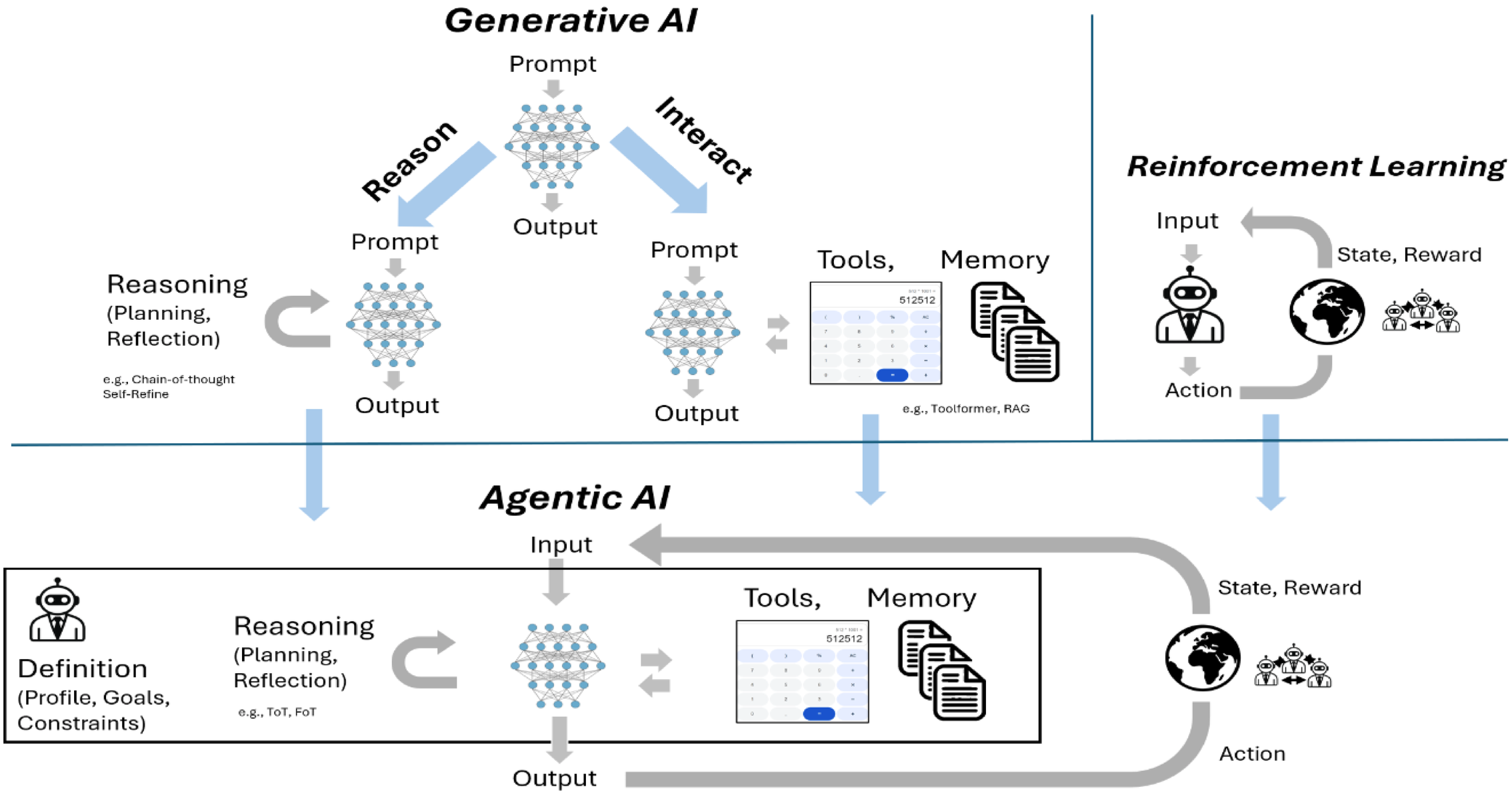
2023: GPT-4, multimodal models

2024: Open-weights race, Mamba2

2025: DeepSeek-R1, Grok 3
Claude 3.7, Gemma 3

Source: https://github.com/Hannibal046/Awesome-LLM

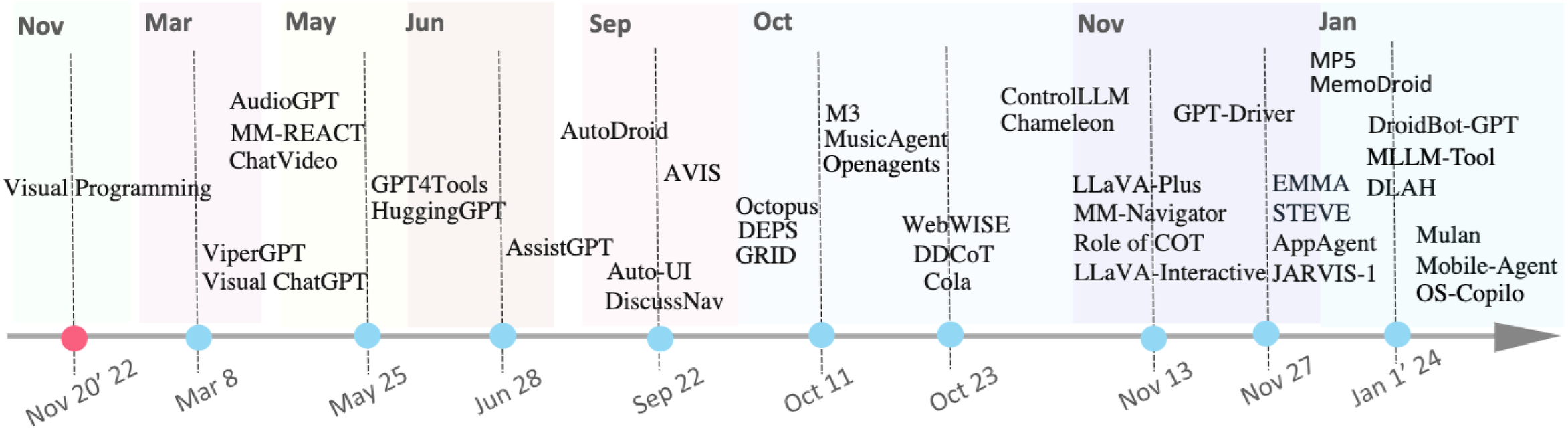
From Generative AI to Agentic AI



Source: Schneider, Johannes. "Generative to Agentic AI: Survey, Conceptualization, and Challenges." arXiv preprint arXiv:2504.18875 (2025).

LLM-powered Multimodal Agents

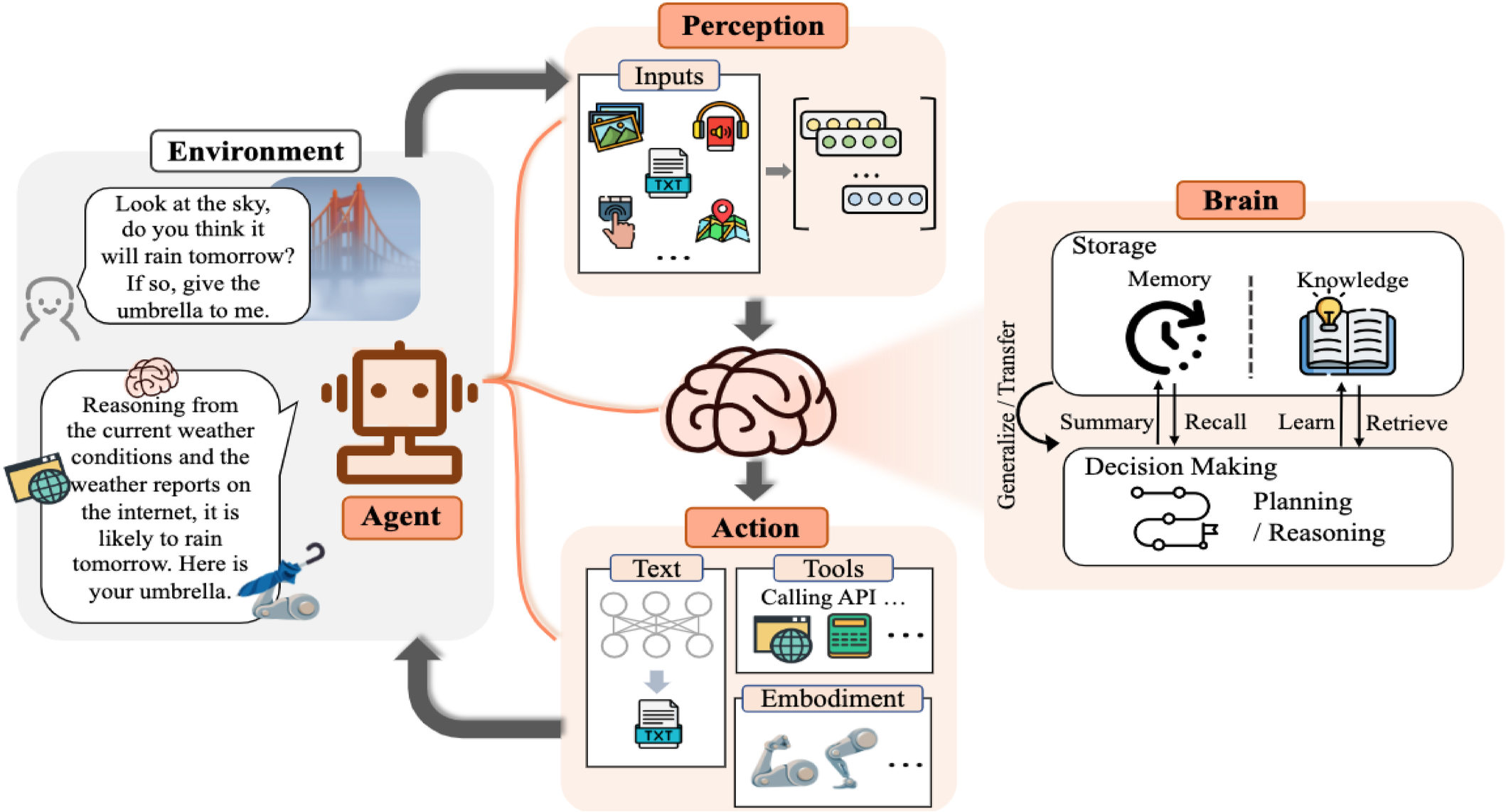
Large Multimodal Agents (LMAs)



2022 2023 2024

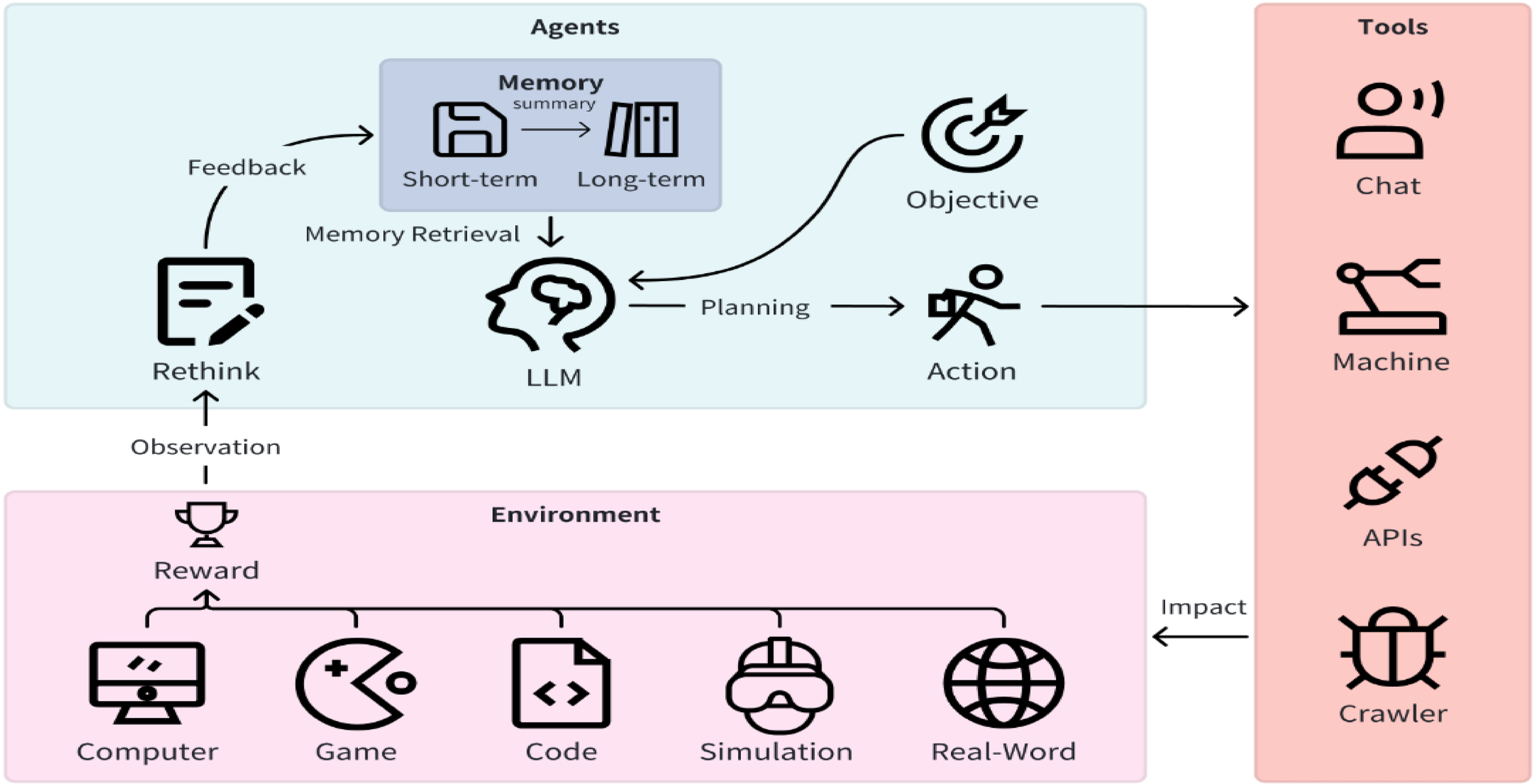
Source: Xie, Junlin, Zhihong Chen, Ruifei Zhang, Xiang Wan, and Guanbin Li. "Large Multimodal Agents: A Survey." arXiv preprint arXiv:2402.15116 (2024).

Large Language Model (LLM) based Agents



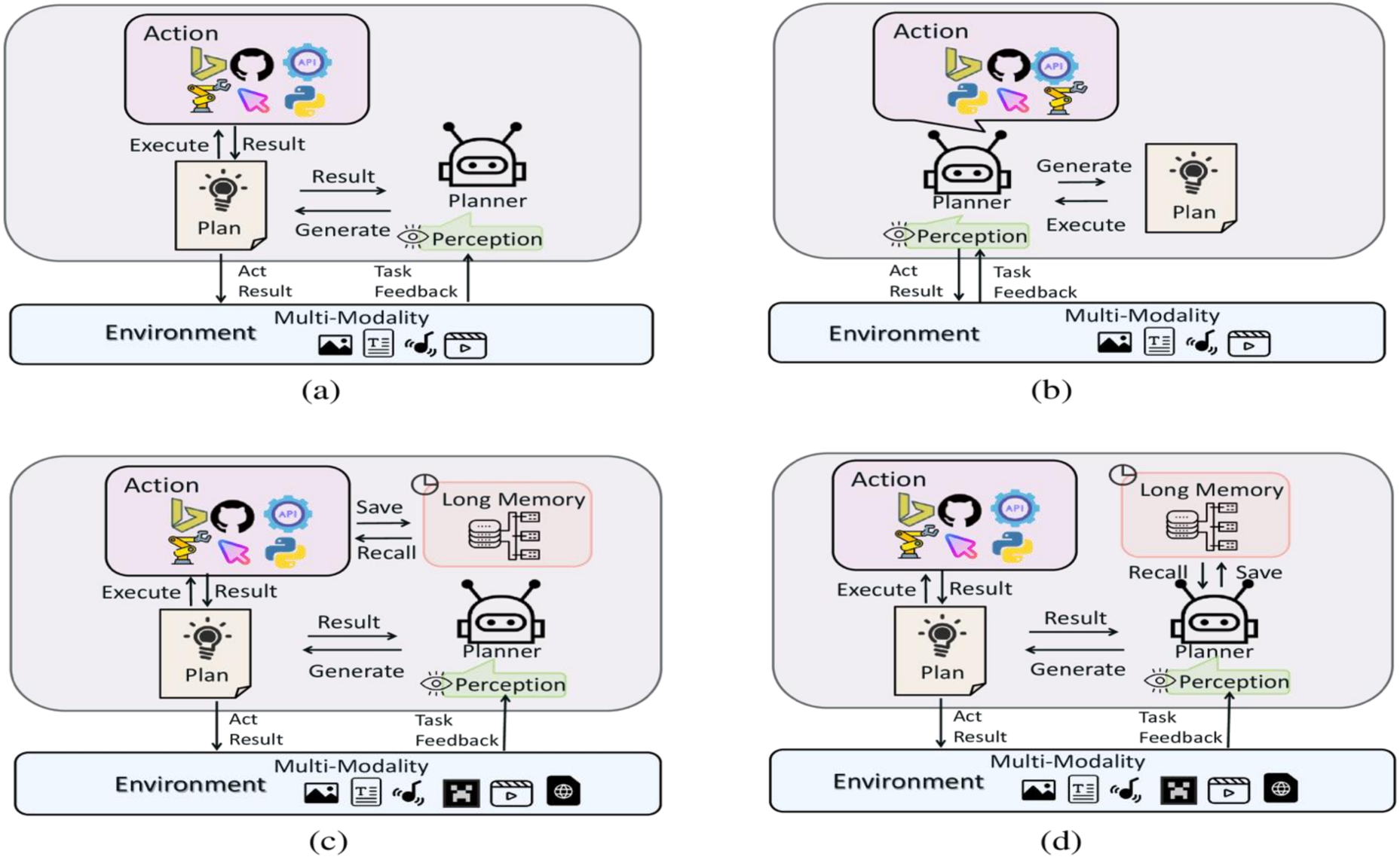
Source: Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2023). The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864.

LLM-based Agents



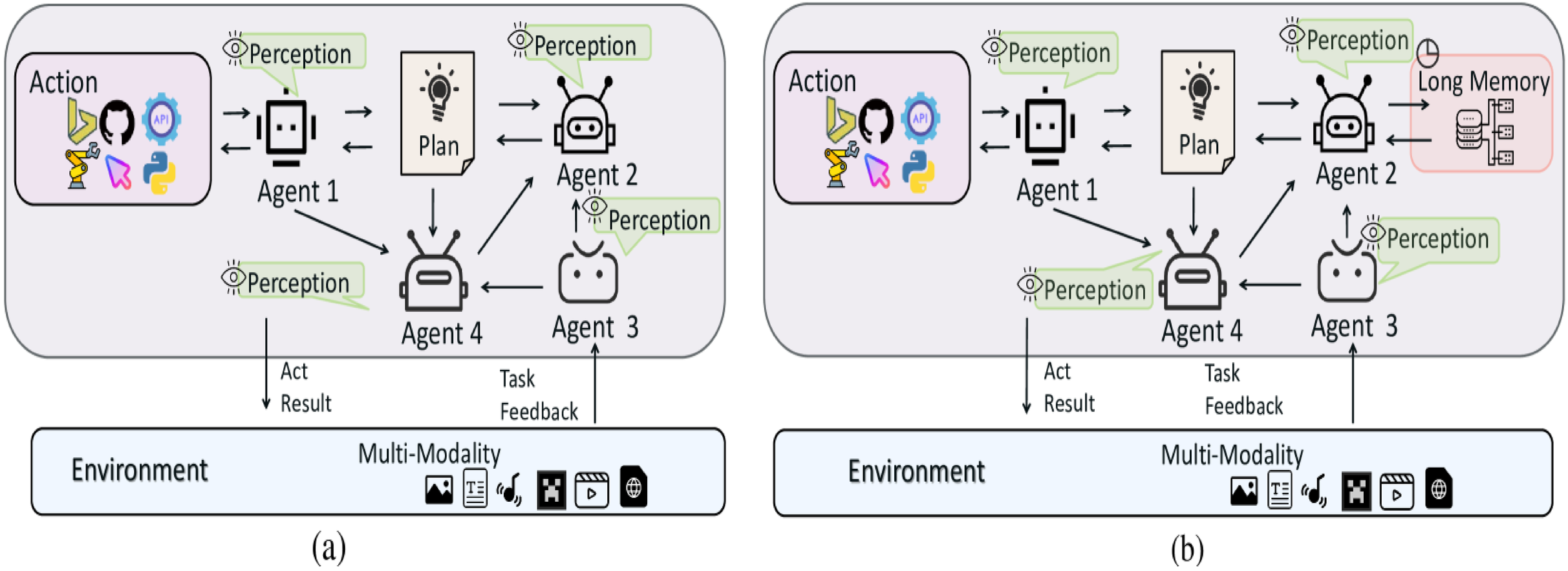
Source: Cheng, Yuheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang et al. "Exploring large language model based intelligent agents: Definitions, methods, and prospects." arXiv preprint arXiv:2401.03428 (2024).

Large Multimodal Agents (LMA)



Source: Xie, J., Chen, Z., Zhang, R., Wan, X., & Li, G. (2024). Large Multimodal Agents: A Survey. ArXiv, abs/2402.15116.

Large Multimodal Agents (LMA)



Source: Xie, J., Chen, Z., Zhang, R., Wan, X., & Li, G. (2024). Large Multimodal Agents: A Survey. ArXiv, abs/2402.15116.

Agentic AI Cloud Architecture

Microservices and Serverless Architecture

Containers (Docker, Kubernetes)

Serverless platforms (AWS Lambda, Google Cloud Functions)

APIs and Tooling Integration via MCP

Agents access tools (e.g., databases, APIs, CRMs, payment gateways)
using Model Context Protocol (MCP)

Enhances tool-using behavior of LLM agents

Tools and Frameworks

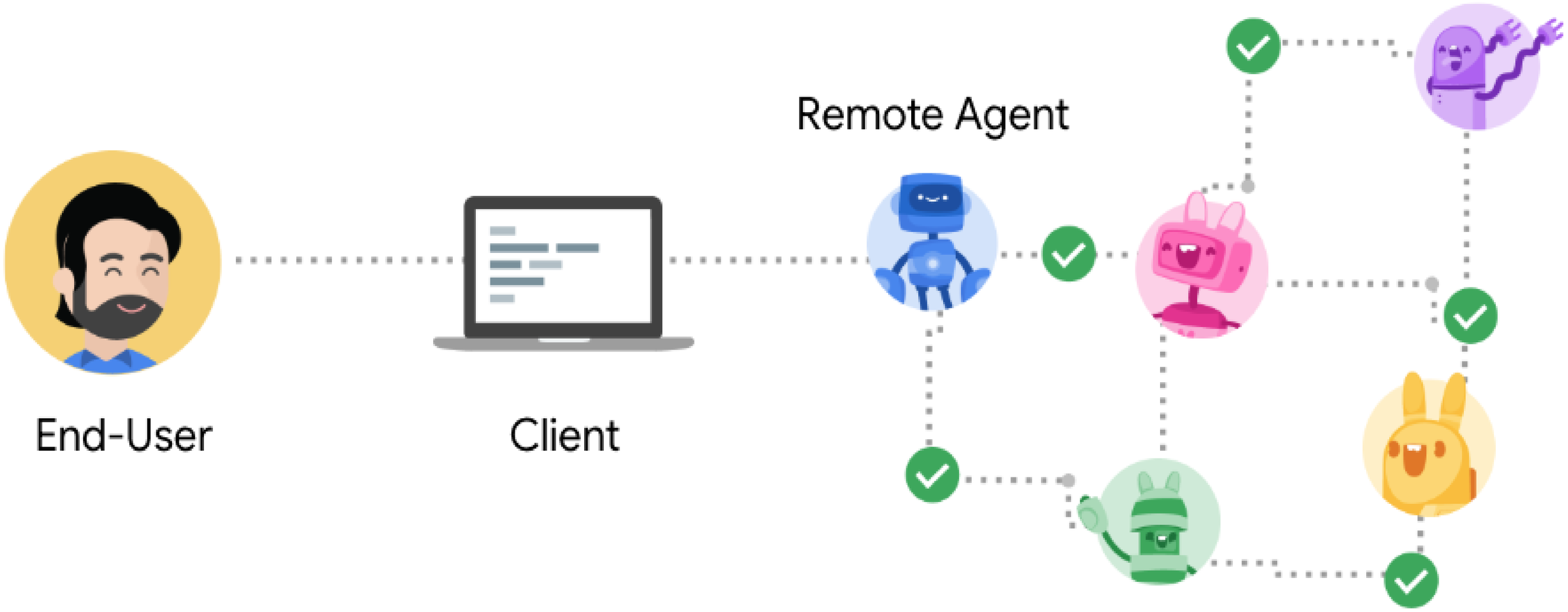
LangChain, AutoGen, CrewAI: for orchestrating LLM agents

Anthropic’s MCP, Google’s A2A: communication protocols

Vector DBs (Pinecone, Weaviate): for agent memory

Agent2Agent Protocol (A2A)

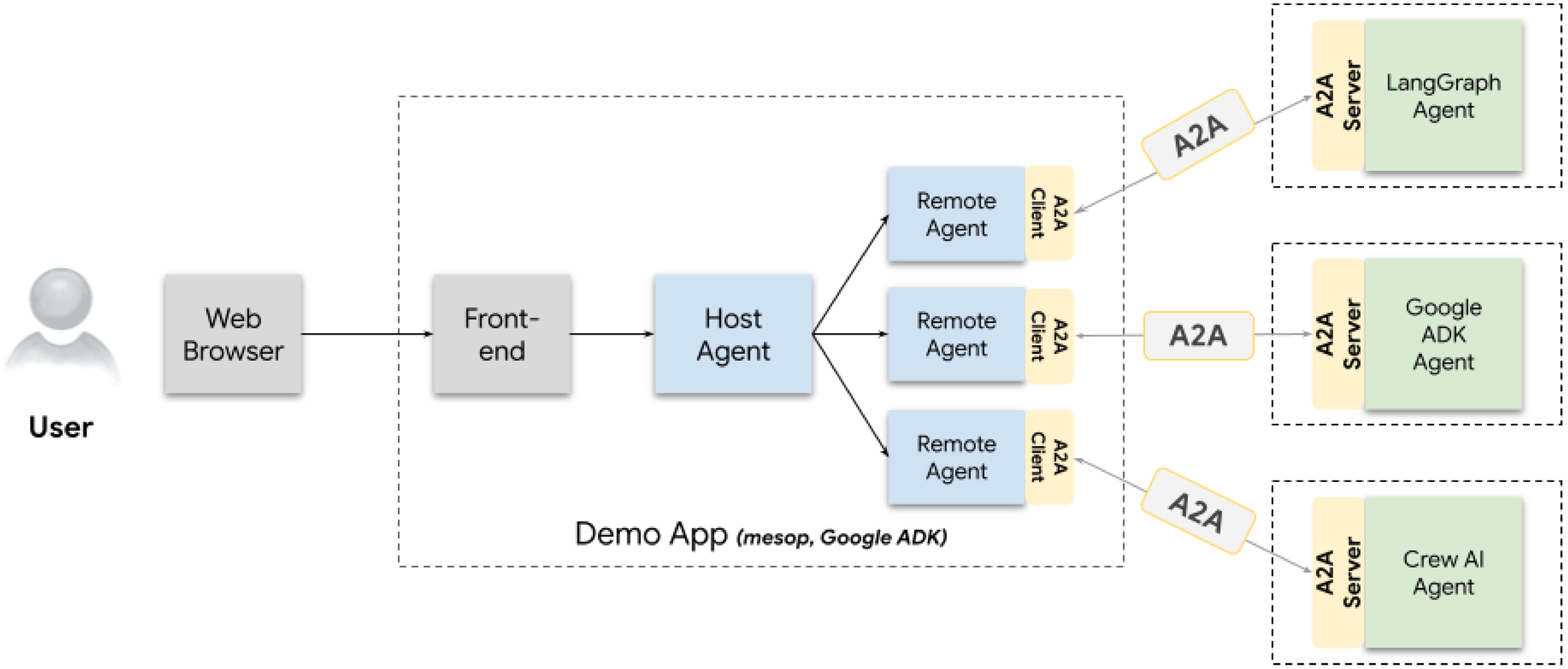
An open protocol enabling Agent-to-Agent interoperability, bridging the gap between opaque agentic systems



Source: https://google.github.io/A2A/

A2A Demo Web App

Agents talking to other agents over A2A



Source: https://github.com/google/A2A/blob/main/demo/README.md

A2A

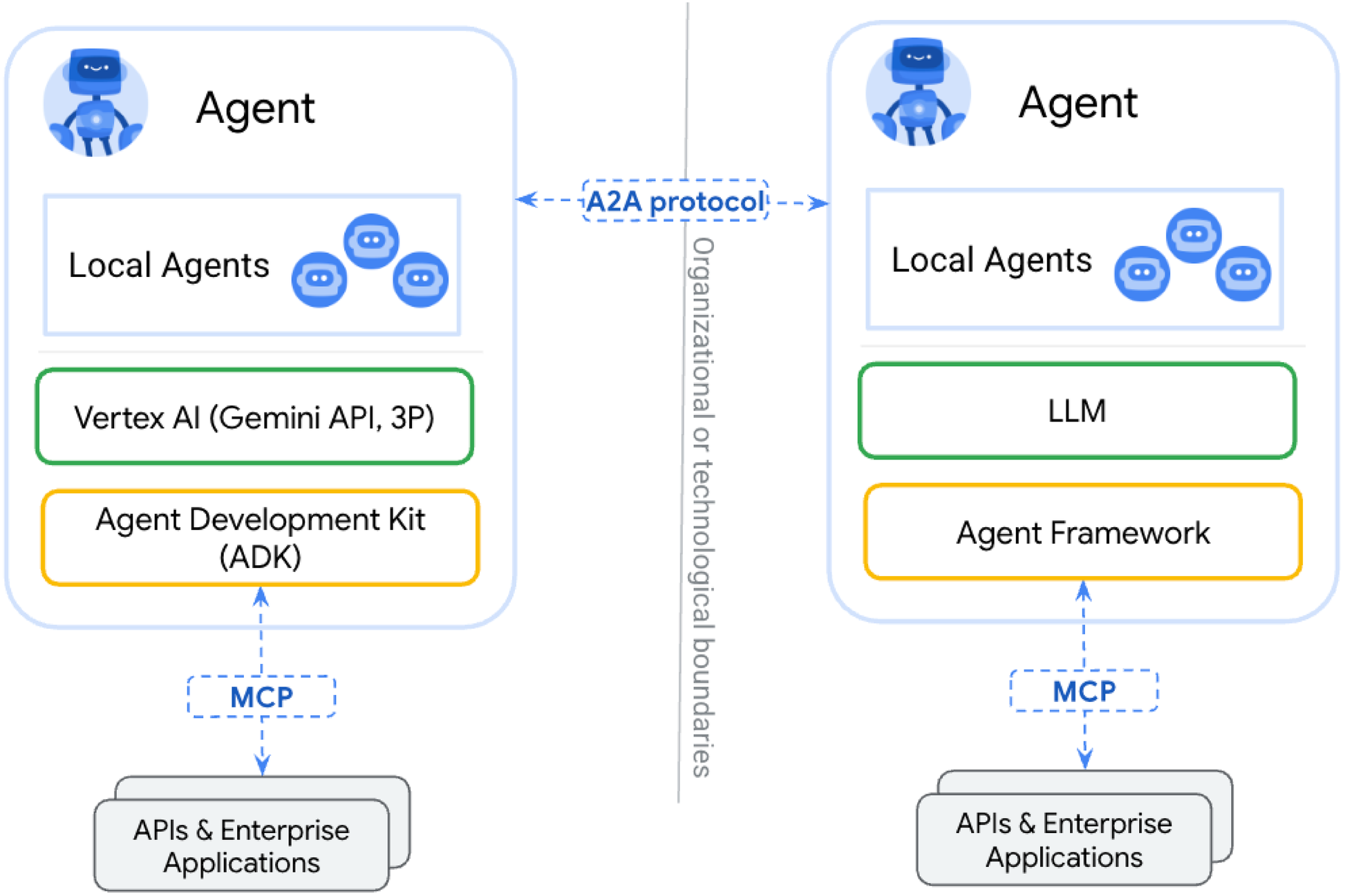
(Agent2Agent Protocol)

for agent-agent collaboration

MCP

(Model Context Protocol)

for tools and resources



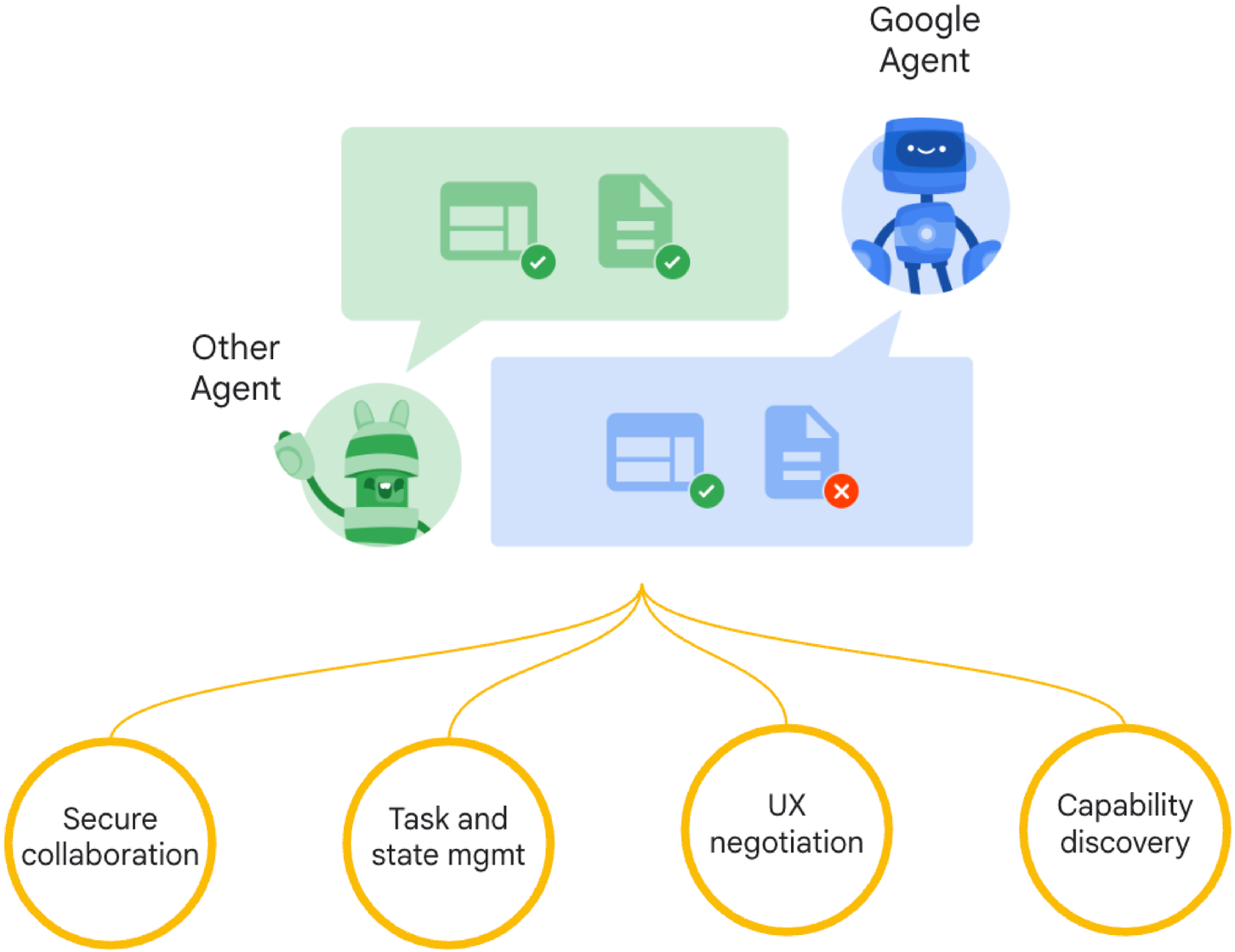
Source: https://google.github.io/A2A/

Google A2A (Agent2Agent Protocol)

Seamless Agent Collaboration

Simplifies Enterprise Agent Integration

Supports Key Enterprise Requirements

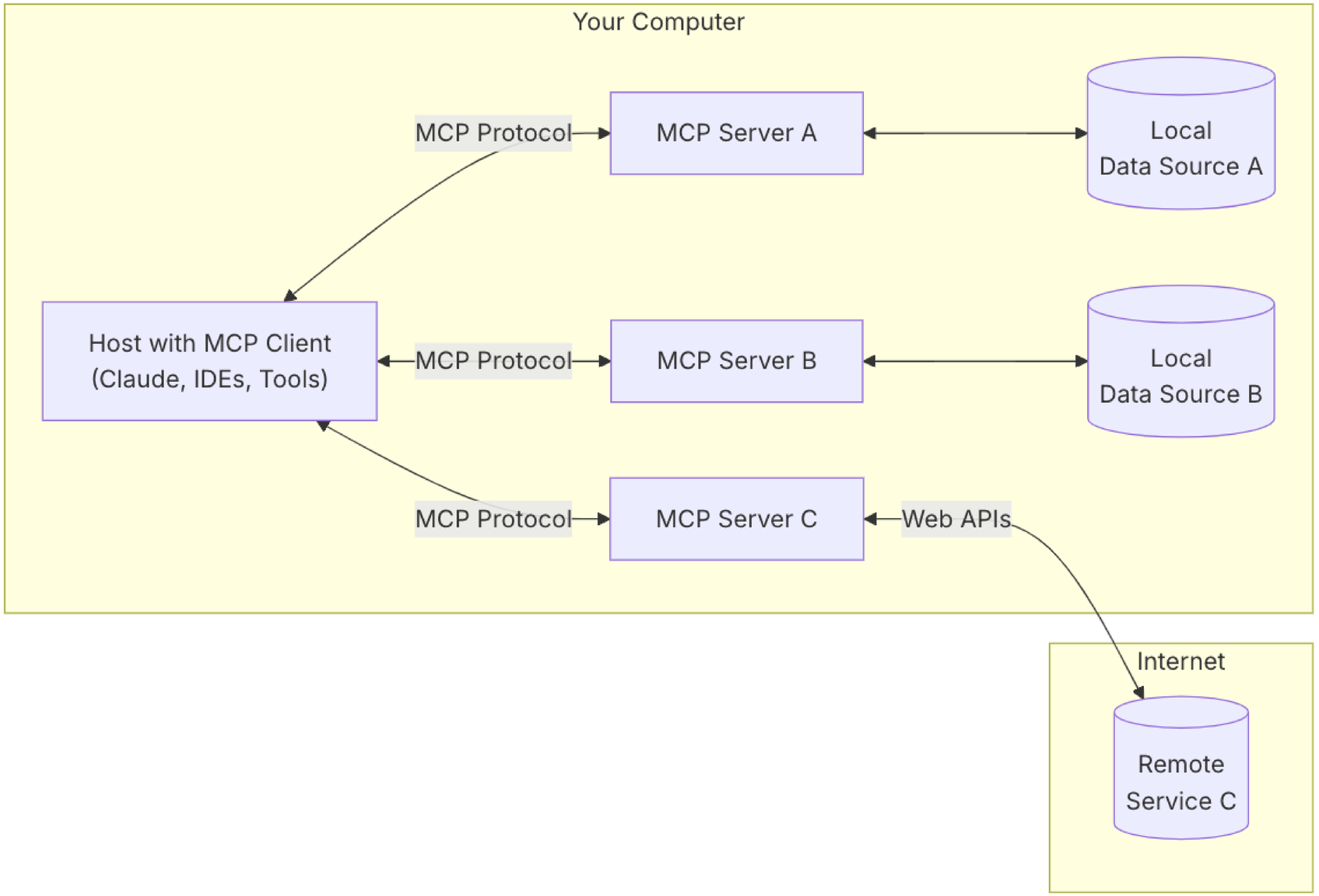


Source: https://google.github.io/A2A/

MCP (Model Context Protocol)

MCP is a open protocol that standardizes how applications provide context to LLMs.

MCP: USB-C port for AI applications.



Source: https://modelcontextprotocol.io/introduction

MCP and A2A

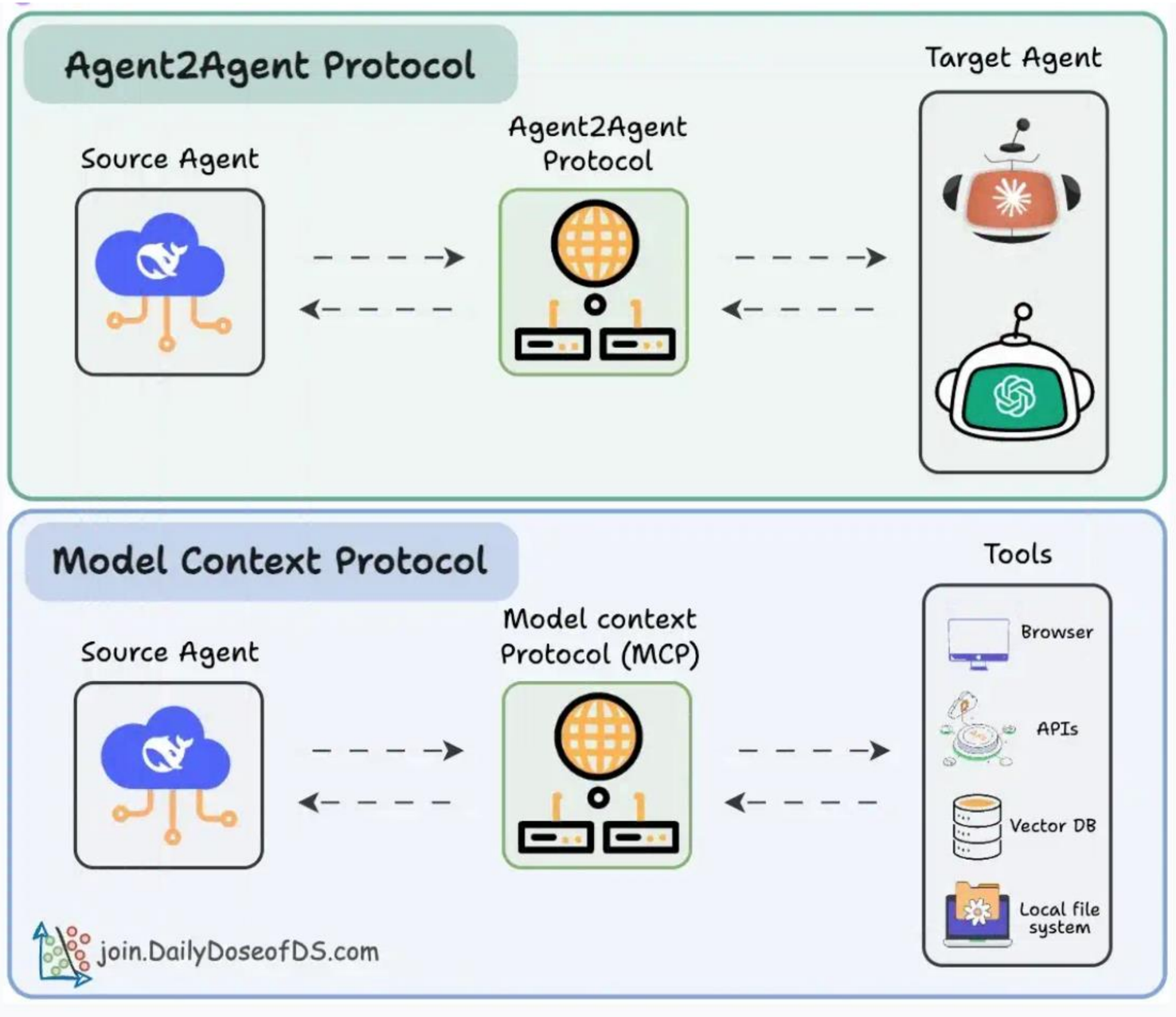
- MCP (Model Context Protocol) for tools and resources
  - Connect agents to tools, APIs, and resources with structured inputs/outputs.
  - Google ADK supports MCP tools. Enabling wide range of MCP servers to be used with agents.
- A2A (Agent2Agent Protocol) for agent-agent collaboration
  - Dynamic, multimodal communication between different agents without sharing memory, resources, and tools
  - Open standard driven by community.
  - Samples available using Google ADK, LangGraph, Crew.AI

Source: https://google.github.io/A2A/

Agentic applications require both A2A and MCP

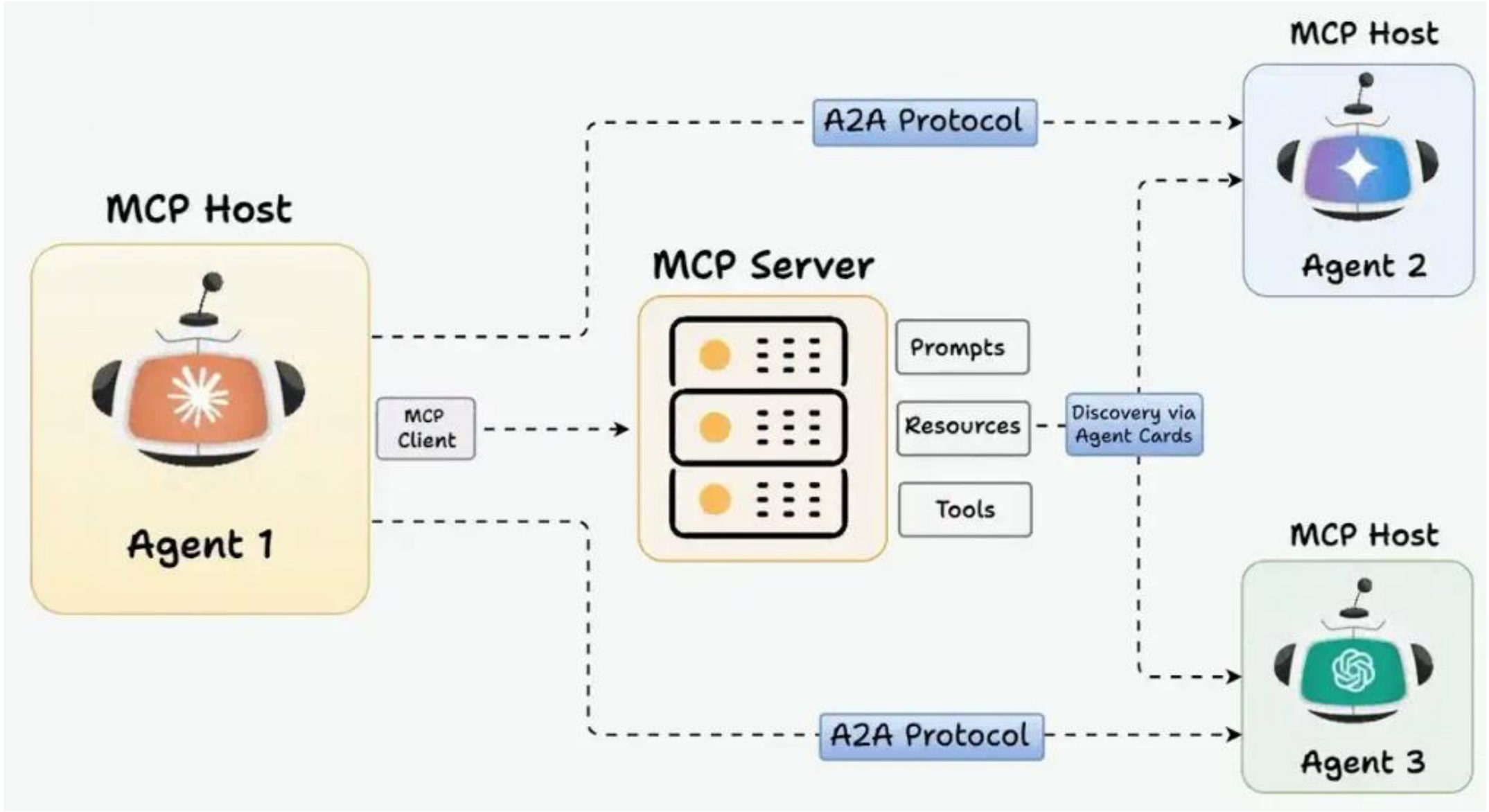
A2A allows agents to connect with other agents and collaborate in teams.

MCP provides agents with access to tools



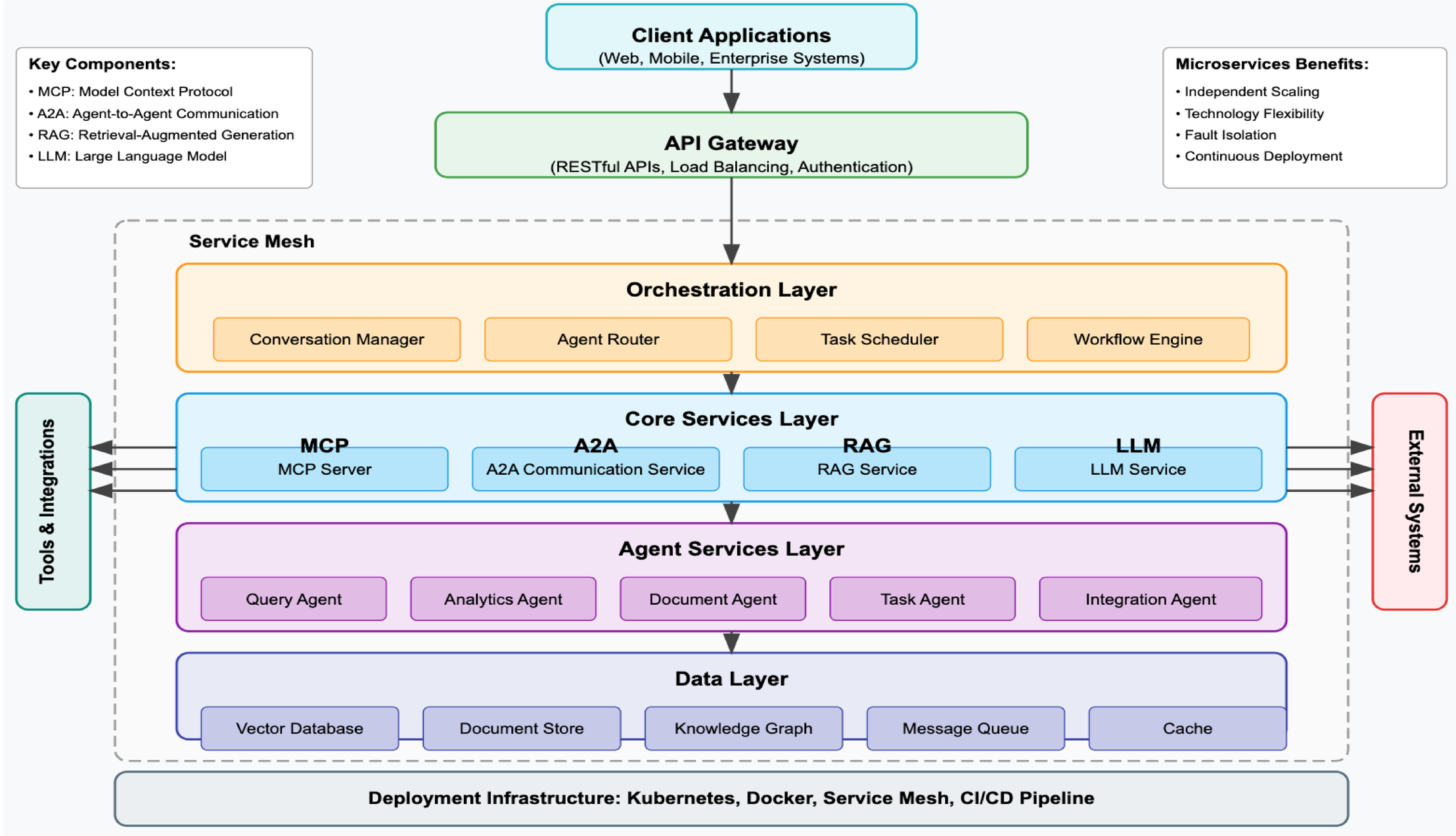
Source: https://blog.dailydoseofds.com/p/a-visual-guide-to-agent2agent-a2a

MCP and A2A Protocol for AI Agents

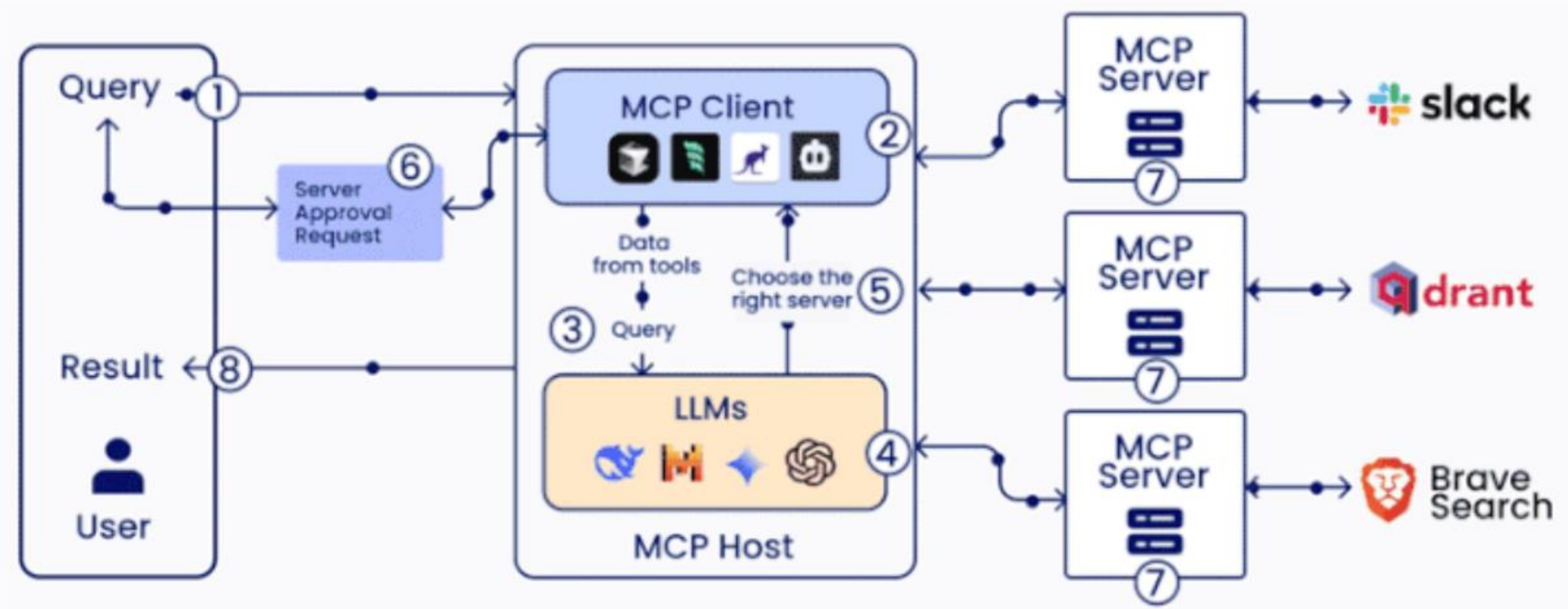


Source: https://blog.dailydoseofds.com/p/a-visual-guide-to-agent2agent-a2a

Agentic AI System with Microservices Architecture

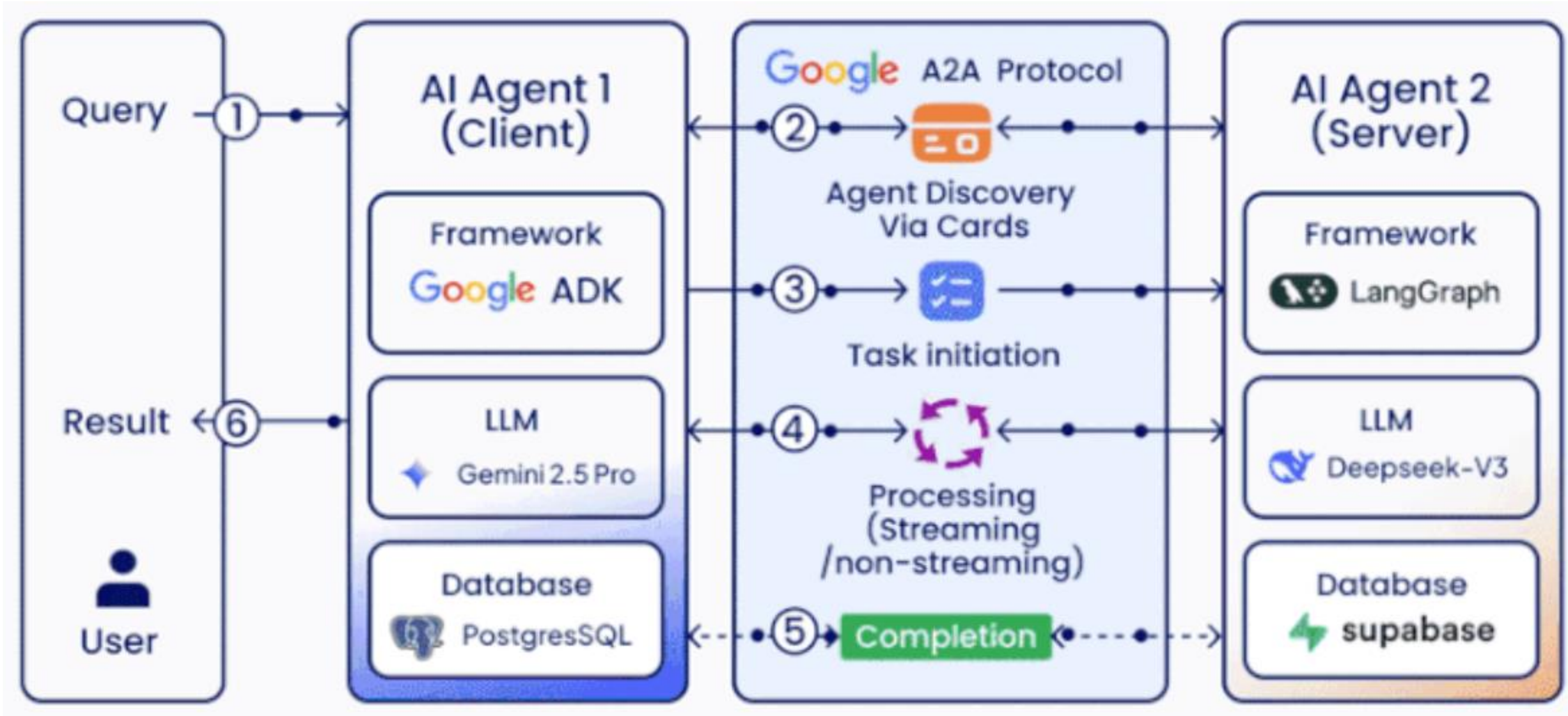


MCP (Model Context Protocol)



Source: Houssem Eddine Lassoued (2025), How MCP and A2A Protocols Are Revolutionizing AI Agent Development, https://medium.com/@houssemeddinelassoued/

A2A (Agent2Agent Protocol)



Source: Houssem Eddine Lassoued (2025), How MCP and A2A Protocols Are Revolutionizing AI Agent Development, https://medium.com/@houssemeddinelassoued/

References

- NVIDIA DLI (2024), Building RAG Agents with LLMs, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1
- NVIDIA DLI (2024), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1