NVIDIA AI TECHNOLOGY CENTER (NVAITC)

Yang-Hsien Lin (yanghsienl@nvidia.com)

NVIDIA AI TECHNOLOGY CENTER (NVAITC)

Outcome-Focused Collaborations for Research & Talent Development

SELECTED NVAITC COLLABORATORS

Singapore

Nanyang Technological University Singapore
Singapore University of Technology and Design
SiT
IMDA Infocomm Media Development Authority
National Research Foundation Prime Minister's Office
National Supercomputing Centre
Temasek Polytechnic
NUS National University of Singapore
EDB Singapore
Republic Polytechnic
SkillsFuture SG
Newcastle University UK | Malaysia | Singapore

Europe

uni.lu Université du Luxembourg
LIST
CSC
cini consorzio interuniversitario nazionale per l'informatica
CINECA
FCAI Finnish Center for Artificial Intelligence

Korea

GIST Gwangju Institute of Science and Technology
KiSTi www.kisti.re.kr

Taiwan

National Taiwan University
National Tsing Hua University
National Cheng Kung University
NCHC

Hong Kong

香港科技大學 The Hong Kong University of Science and Technology
HKSTP 香港科技園
The Hong Kong Polytechnic University 香港理工大學
香港大學 The University of Hong Kong
香港浸會大學 Hong Kong Baptist University

Thailand

จุฬาลงกรณ์มหาวิทยาลัย

Brazil

SENAI CIMATEC
FIEB Federação das Indústrias do Estado da Bahia
UFRGS Universidade Federal do Rio Grande do Sul
Petrobras

Indonesia

epsindo PT. Epsindo Prima Sinergi

AI Acceleration Advisory Board

Monash University

India

IIT Hyderabad
Indian Institute of Technology Madras
CDAC
Indian Institute of Technology Kanpur
Mahindra University Global Thinkers. Engaged Leaders.
NCRA • TIFR

Japan

Kyoto University Founded 1897 Kyoto Japan
Waseda University
Kanazawa University 1946
University of Tsukuba Center for Computational Sciences 筑波大学 計算科学研究センター

NALA

UF University of Florida

China

Chinese Academy of Sciences
Nanjing University
Peking University
Fudan University
Shanghai Jiao Tong University
The University of Nottingham United Kingdom · China · Malaysia

NVIDIA

NV AI & NV DOMAIN TECH CENTER ENGAGEMENT DRIVERS

Address critical needs across the whole national AI journey

Similar focus across countries

- Science, Research and Development
- Technology Adoption
- Developing an AI workforce

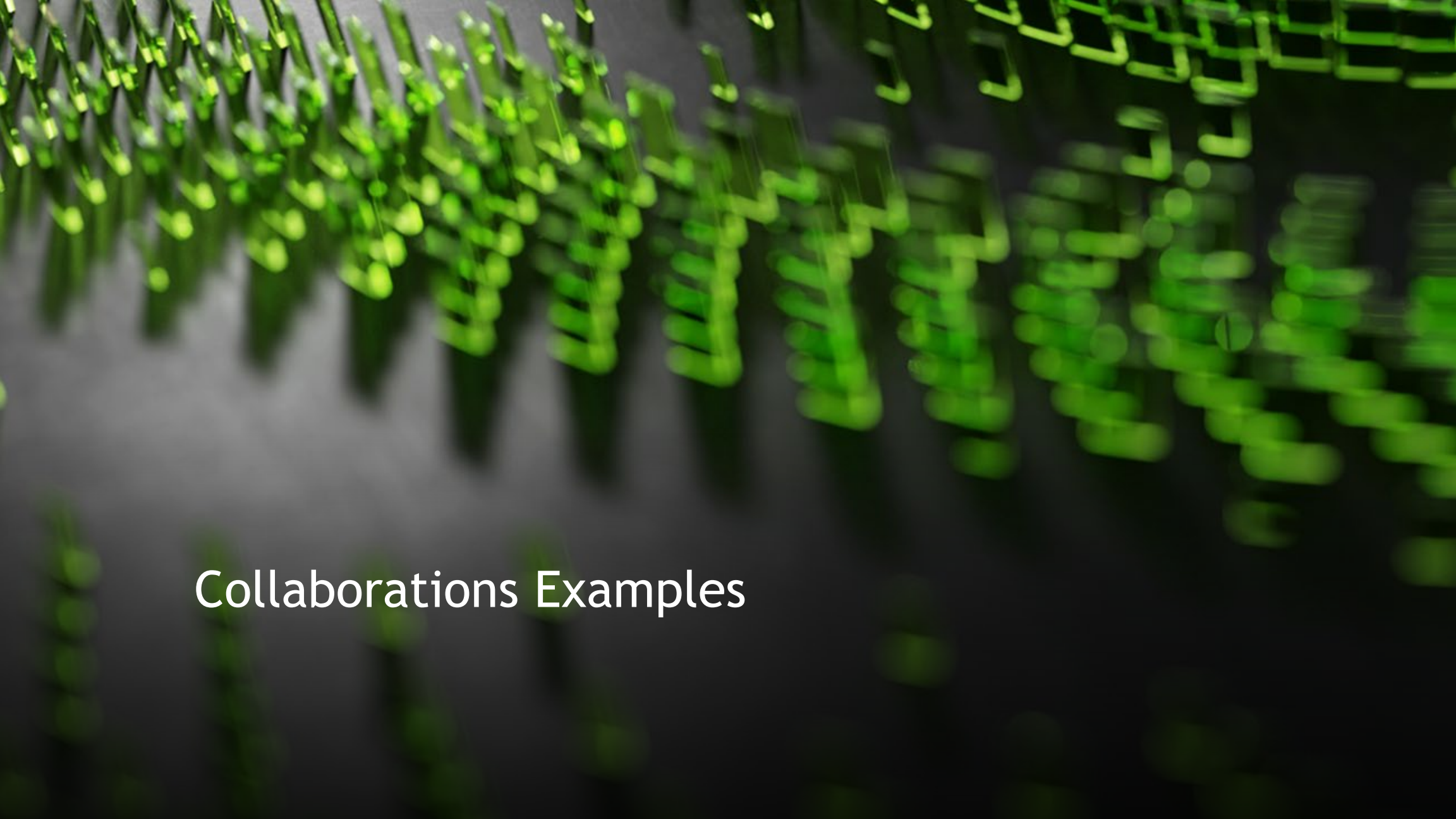
Unique Local Implementations

- Staffing Model
- Government Agencies
- Talent Development
- Funding Sources
- Supercomputing Centers
- Level of Research Focus
- Interns & Graduate Students
- Universities
- Collaborate vs Support Balance


Collaboration Model

NVAITC GENERAL RESEARCH GUIDELINES

Project Selection	• Jointly agreed by Host and NVAITC • NV criteria include technology stack, computing scale, novelty • an agreed-upon “statement of work”
PI Contribution	• Access to existing compute platform resources • Funding • Researchers
NV Contribution	• GPU/ML/DL technology / adoption enablement • Staff effort is bounded in both total time and duration • No joint IP • Beyond NVAITC, other NVResearch or teams may or may not collaborate as per any other academic collaboration
Expected Outcome	• scientific publication, with NVAITC acknowledgement or co-authorship • open-source code release



Collaborations Examples

NCHC X NVIDIA WORKSHOPS/BOOTCAMPS/HACKATHONS

Since 2020

Events

35+

Developers

1600+

RAG-BASED LLM COURSE

Start from Q1 2024

RAG-based LLM Course

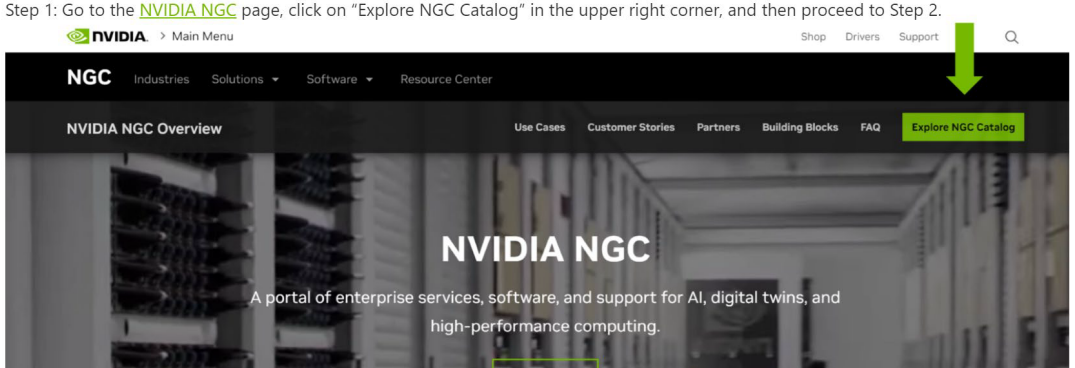
Yang-Hsien Lin

We are thrilled to share our upcoming plans to partner with various academic institutions in delivering instruction on RAG-based LLM, while simultaneously offering students the chance to gain hands-on experience with RAG.
We've harnessed NVIDIA's resources to create instructional slides, incorporating Generative AI Examples provided by NVIDIA as our case studies (kindly note that these examples are subject to ongoing updates).

Recognizing that many academic institutions lack the necessary GPU capabilities to run LLM, we've opted for a Cloud-based LLM approach to fulfill our educational and experiential objectives. We warmly invite any innovative solutions or creative course content suggestions you may wish to contribute.

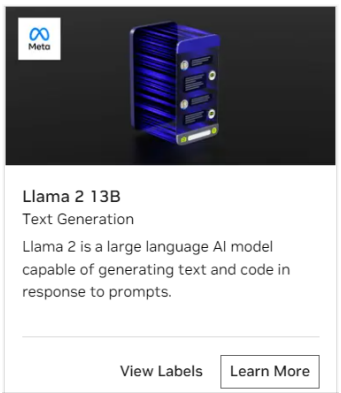
Outlined below are the tasks that students are expected to complete in preparation for the course.
This test can be roughly divided into four steps (students must achieve the first three points before the class):

1. **Apply for an NVIDIA NGC account**: The purpose of this is to obtain the NGC API Key and get the required docker image from NGC.
    - Step 1: Go to the NVIDIA NGC page, click on "Explore NGC Catalog" in the upper right corner, and then proceed to Step 2.



Document

- **Increase by 1000+ students**
    - NGC accounts
    - NVIDIA AI Endpoint users
- **Promote RAG to campus**
    - National Taiwan University
    - Chung Yuan Christian University
    - National Tsing Hua University
    - China Medical University
- **Experience RAG**
    - Based on our public Generative AI examples
    - NVIDIA Cloud based LLM



NVIDIA

NVAITC

Outcome Highlighted

Journals & Magazines > IEEE Access > Volume: 12

Solar Irradiance Forecasting Using a Hybrid Quantum Neural Network: A Comparison on GPU-Based Workflow Development Platforms

Publisher: IEEE | Cite This | PDF

IEEE Access

Ying-Yi Hong ; Dylan Josh Domingo Lopez ; Yun-Yuan Wang All Authors

---

Maximum Entropy Reinforcement Learning via Energy-Based Normalizing Flow

NeurIPS

Chen-Hao Chao*[1,2] Chien Feng*[1] Wei-Fang Sun[2]

Cheng-Kuang Lee[2] Simon See[2] Chun-Yi Lee†[1]

[1] Elsa Lab, National Tsing Hua University, Hsinchu City, Taiwan
[2] NVIDIA AI Technology Center, NVIDIA Corporation, Santa Clara, CA, USA

Conferences > ICASSP 2025 - 2025 IEEE Inter...

3D Gaussian Splatting with Grouped Uncertainty for Unconstrained Images

Publisher: IEEE | Cite This | PDF

ICASSP

Hao-Yu Hou ; Chia-Chi Hsu ; Yu-Chen Huang ; Mu-Yi Shen ; Wei-Fang Sun ; Cheng Sun All Authors

---

PAPER • OPEN ACCESS

Validating large-scale quantum machine learning: efficient simulation of quantum support vector machines using tensor networks

IOP Science

Kuan-Cheng Chen*, Tai-Yue Li, Yun-Yuan Wang, Simon See, Chun-Chieh Wang, Robert Wille, Nan-Yow Chen, An-Cheng Yang and Chun-Yu Lin

Published 20 February 2025 • © 2025 The Author(s). Published by IOP Publishing Ltd

Machine Learning: Science and Technology, Volume 6, Number 1

Conferences > 2024 International Conference...

Quantum-Classical-Quantum Workflow in Quantum-HPC Middleware with GPU Acceleration

Publisher: IEEE | Cite This | PDF

IEEE QCNC

Kuan-Cheng Chen ; Xiaoren Li ; Xiaotian Xu ; Yun-Yuan Wang ; Chen-Yu Liu All Authors

---

PCIe Bandwidth-Aware Scheduling for Multi-Instance GPUs

HPC Asia

Yan-Mei Tang
National Tsing Hua University
Hsinchu, Taiwan
ymtang@lsalab.cs.nthu.edu.tw

Wei-Fang Sun
NVIDIA AI Technology Center
Santa Clara, USA
johnsons@nvidia.com

hsutzu.ting@lsalab.cs.nthu.edu.tw

Ming-Hung Chen
IBM Research
Yorktown Heights, NY, USA
minghungchen@ibm.com

I-Hsin Chung
IBM Research
Yorktown Heights, NY, USA
ihchung@us.ibm.com

Jerry Chou
National Tsing Hua University
Hsinchu, Taiwan

Home > International Conference on Biomedical and Health Informatics 2024 > Conference paper

Pilot Study of Retrieval-Augmented Generation Model in Recommending Traditional Chinese Medicine Formulations

IFMBE

Conference paper | First Online: 28 March 2025

pp 331–341 | Cite this conference paper

Ya-Chuan Chan, Po-Yu Huang, Zhi-Liang Chen, Chih-Nung Wang, Wen-Chen Lin, Jung-Peng Chiu, Yi-Chun Chiu, Yang-Hsien Lin, Eddie T. C. Huang, Simon See & Kang-Ping Lin

---

Harnessing Stable Diffusion: A Leap Forward in Rib Suppression for Chest X-Rays (C-10746)

ECR 2025

was presented within the framework of the Educational and Scientific Programme at

ECR 2025
FEBRUARY 26 – MARCH 2, 2025

GTC 2025 Posters

Driving Smarter Financial Intelligence: The Power of Deeply Trained Domain-Specific Llama Models [P71458]

Pathogenic or Benign? AI-Powered Predictions in Genomic Variant Classification [P73461]

RTeachBot: Transforming Health Education in Radiation Oncology [P73437]

Adaptive Distribution Generator: Powering Split-Step Quantum Walks with CUDA-Q [P72789]

CARDIA3D: Comprehensive Anatomical Recognition and Deep-Learning Imaging Algorithm for 3D Cardiac CT [P72960]

Join us as we explore new solutions for multiclass semantic segmentation in 3D cardiac CT images to address the critical lack of models capable of annotating both large structures — like atria and ventricles — and fine details such as coronary arteries, fat, and plaques. By enhancing recognition of multiple classes in high-resolution CT data, we aim to improve feature extraction for heart-focused AI applications and pave the way for future developments, such as training a cardiac-specific language model to interpret 3D CT images. This work is timely and important, filling a significant gap in cardiac imaging tools essential for advancing precision medicine.

Topic: Computer Vision / Video Analytics - Medical Imaging

Industry Segment: Healthcare & Life Sciences

Po Hsuan Tseng, Professor, National Taipei University of Technology

Shu-Yu Hsu, Graduate Researcher, National Taipei University of Technology


NVIDIA Tech

APPLICATION FRAMEWORKS

PLATFORM

NVIDIA AI

NVIDIA OMNIVERSE

ACCELERATION LIBRARIES

RTX

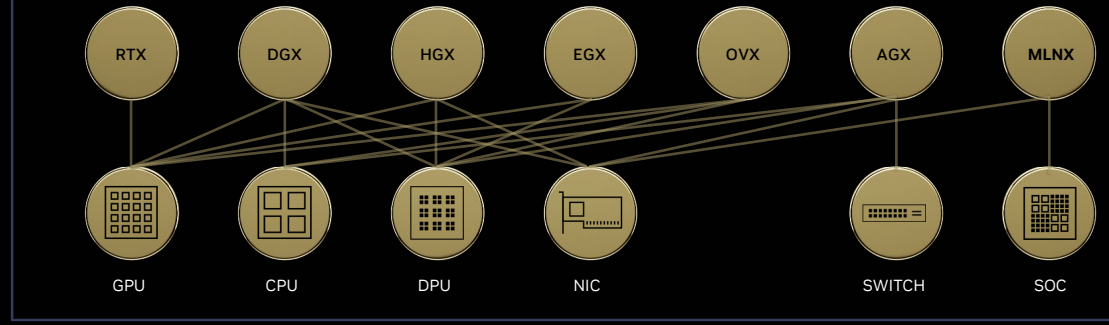
CUDA-X

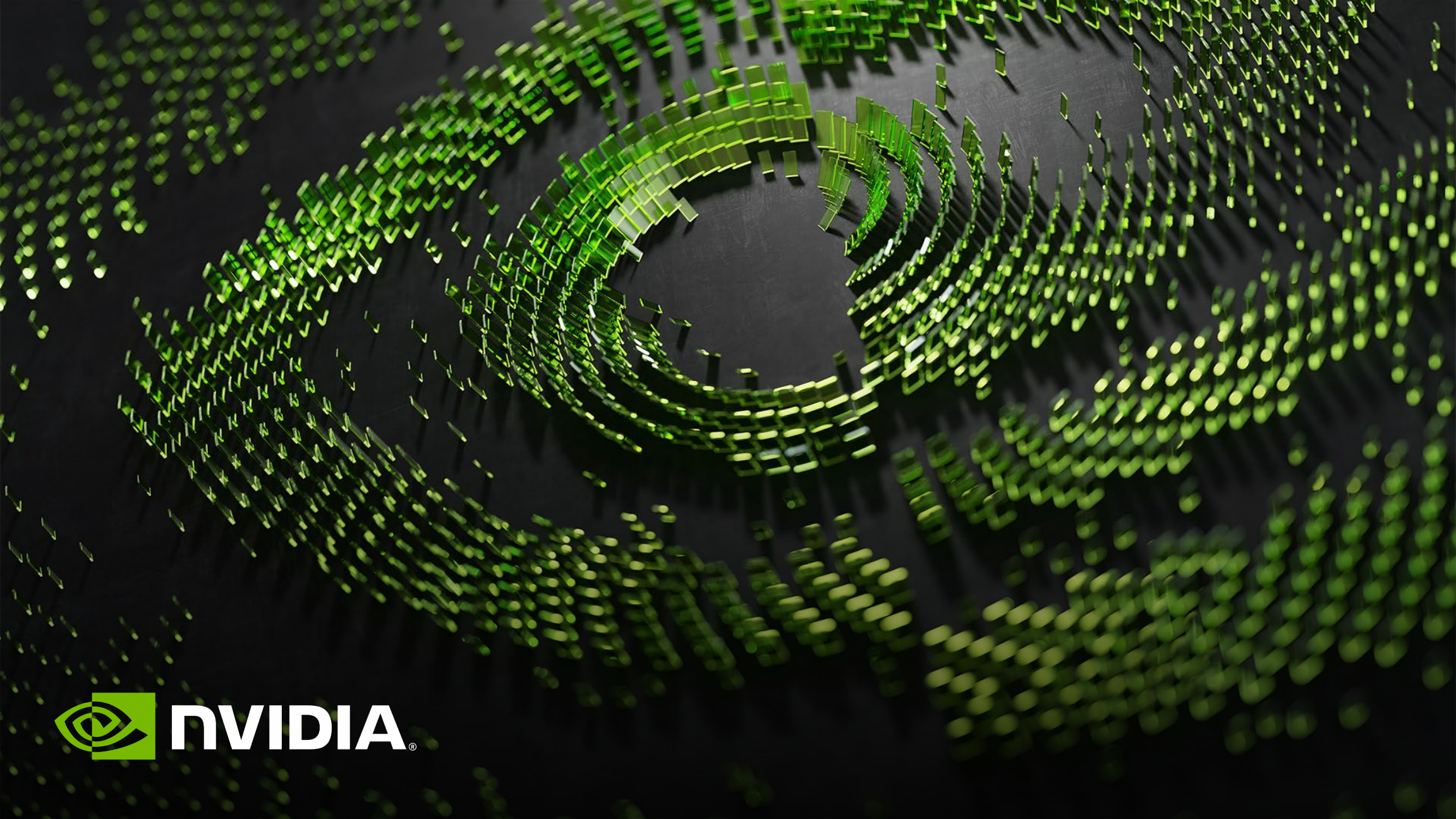
CUDA

SYSTEM SOFTWARE

Magnum IO | DOCA | Base Command | Forge

HARDWARE




NVIDIA