國家實驗研究院
國家高速網路與計算中心
National Center for High-performance Computing

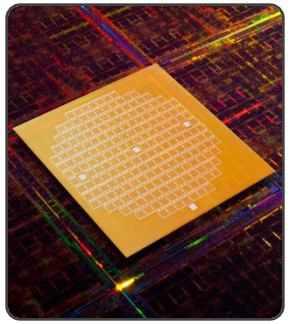
NCHC CUDA-Q Bootcamp

July 23, 2025

NCHC X OpenACC X NVIDIA

CUDA-X Accelerates Every Industry

6M+ Developers & 900+ SDKs/Models

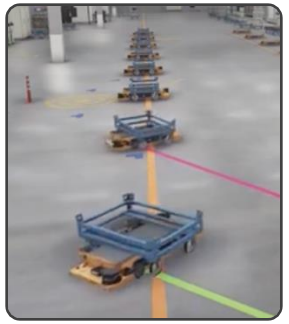


cuLitho

Computational Lithography

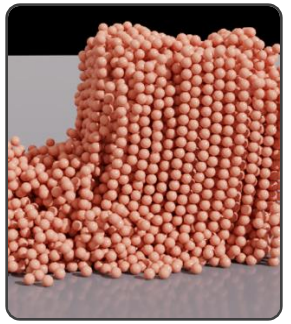
cuDSS

CAE



cuOpt

Decision Optimization



Warp

Physical Simulation

cuDF

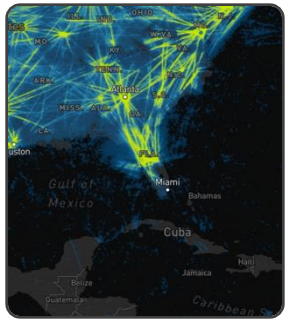
Data Processing

PhysicsNeMo

AI Physics

CUDA-Q

Quantum Computing

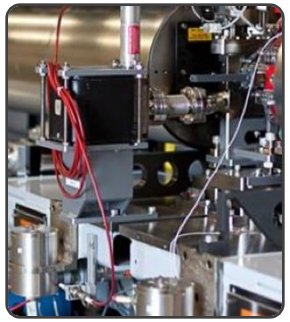


cuEquivariance

Drug & Materials Discovery

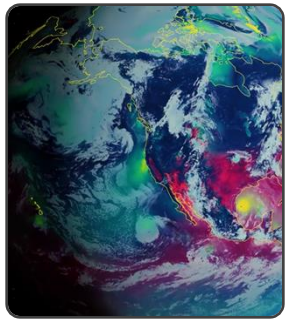
ALCHEMI

AI Materials Science



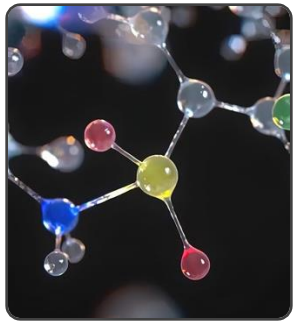
Holoscan

Edge HPC



Earth-2

Weather Analytics



Parabricks

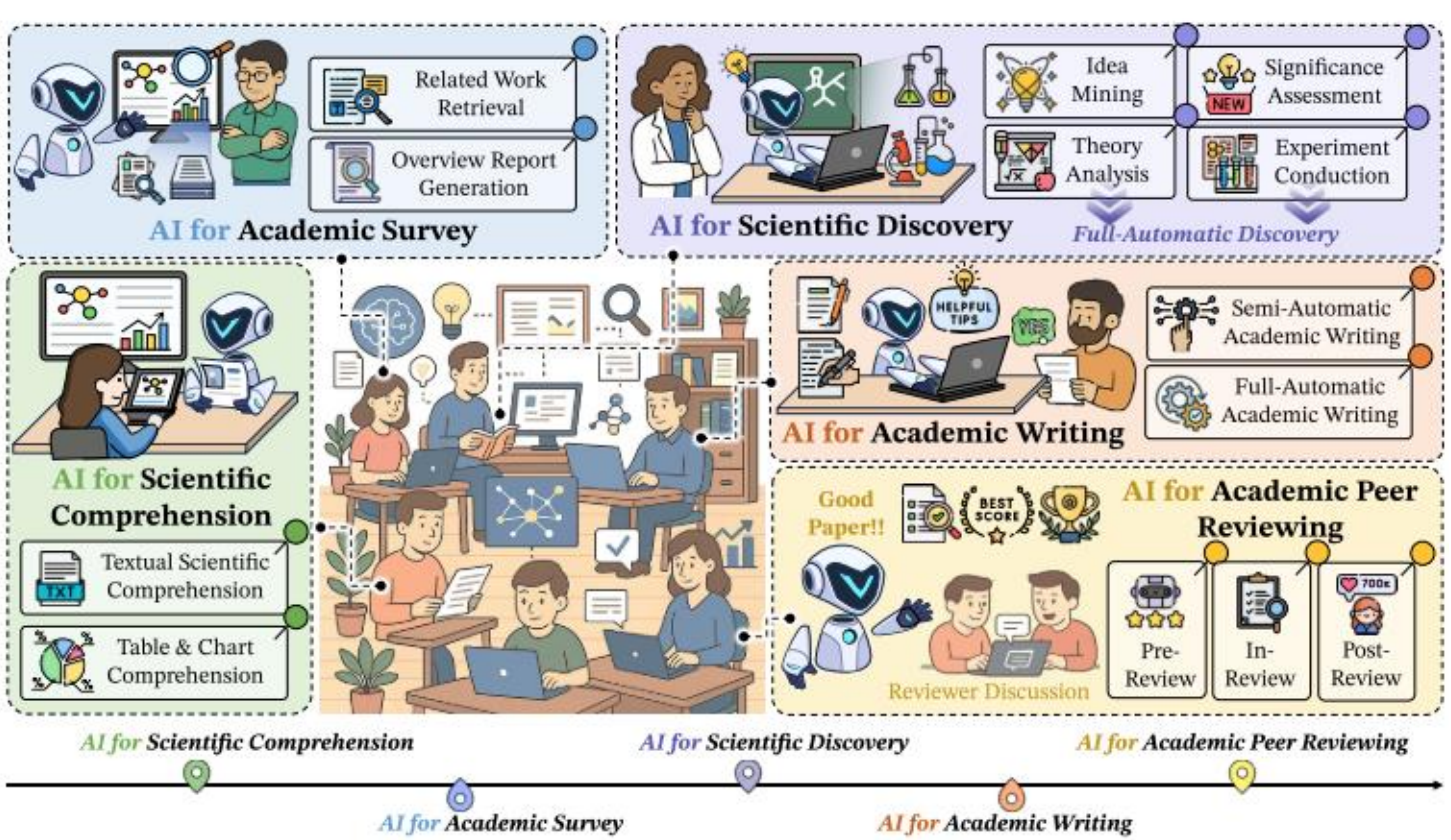
Gene Sequencing

cuPyNumeric

Numerical Computing

AI4Research: A Survey of Artificial Intelligence for Scientific Research

https://arxiv.org/abs/2507.01903
by Q. Chen, M. Yang et al.



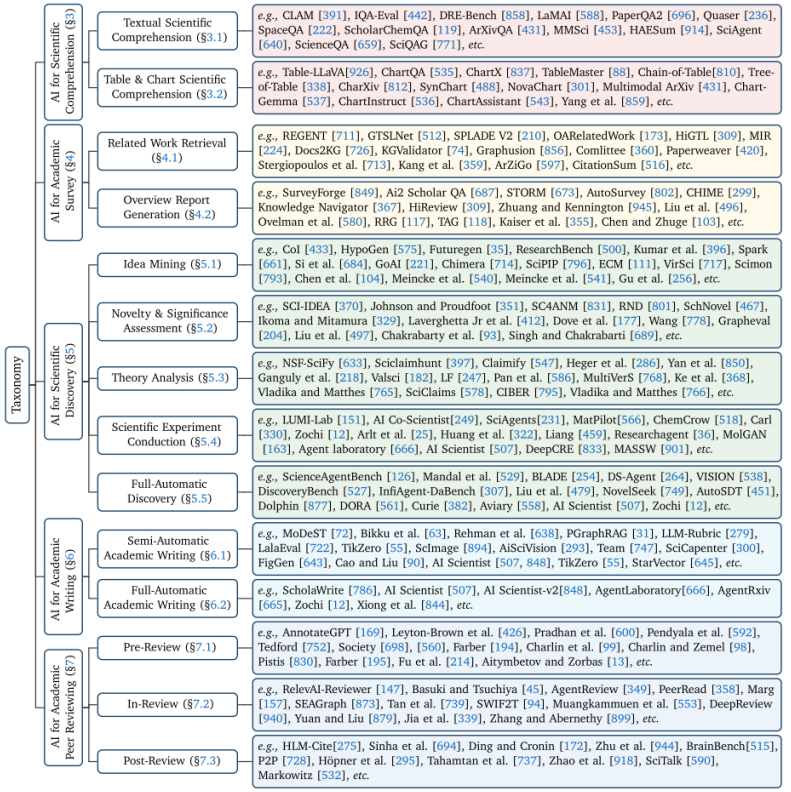


Figure 2: The taxonomy of AI in research (AI4Research) is categorized into five key areas. Each area is subdivided into specific tasks, underscoring the varied roles of AI in the entire research process.

NCHC-NVIDIA Joint Lab

https://github.com/nqobu/nvidia/

Repository for NCHC-NVIDIA Joint Lab

- tutorials - NVIDIA online courses/tutorials in AI/HPC
- 20210412 - NVIDIA Techniques Sharing 2021
- 20210706 - NCHC Techniques Sharing 2021
- 20211202 - NVIDIA Techniques Update 2021
- 20211221 - AI+HPC: 利用 NVIDIA Modulus 實踐 PINN 於物理模擬
- 20220415 - NVIDIA Techniques Update 2022
- 20220530 - NCHC-NVIDIA Techniques Sharing 2022
- 20220629 - PINN 與 NVIDIA Modulus 實作訓練營
- 20221111 - Quantum Computing Workshop / 量子計算模擬實作
- 20230413 - NVIDIA Techniques Sharing 2023
- 20230517 - NVIDIA Techniques Briefing: NVIDIA Federated Learning
- 20230525 - AI for Science: NVIDIA Modulus 及 NVIDIA Omniverse 實作
- 20230727 - N-Way to GPU Programming Bootcamp / 多 GPU 程式設計訓練課程
- 20230821 - NVIDIA Techniques Salon 2023: Programming the NVIDIA Superchip
- 20231207 - NCHC Open Hackathon 2023
- 20240410 - NCHC Quantum Computing Bootcamp 2024 - NVIDIA CUDA-Q and cuQuantum
- 20240506 - AI for Science: NVIDIA Modulus, NVIDIA Omniverse, and NVIDIA Earth-2
- 20240508 - NCHC Techniques Sharing 2024
- 20240626 - NCHC AI for Science Bootcamp 2024 - NVIDIA Modulus 物理模擬計算
- 20240806 - NCHC End-to-end LLM Bootcamp 2024 - NVIDIA NeMo 大型語言模型框架
- 20240924 - NCHC N-Way Bootcamp 2024 – NVIDIA GPU 加速運算
- 20241114 - Earth-2 Overview
- 20241129 - NCHC × NTU - NVIDIA BioNeMo Protein Design Workshop 2024
- 20241204 - NCHC Open Hackathons 2024
- 20250218 - NCHC Grace Workshop 2025
- 20250415 - NCHC N-Way Bootcamp 2025 – NVIDIA GPU 加速運算

Co-innovate with Developers

3 Ways

CUDA-X Bootcamp	Open Hackathon	NVAITC Projects
Training	Acceleration	Collaboration

Co-innovate with Developers

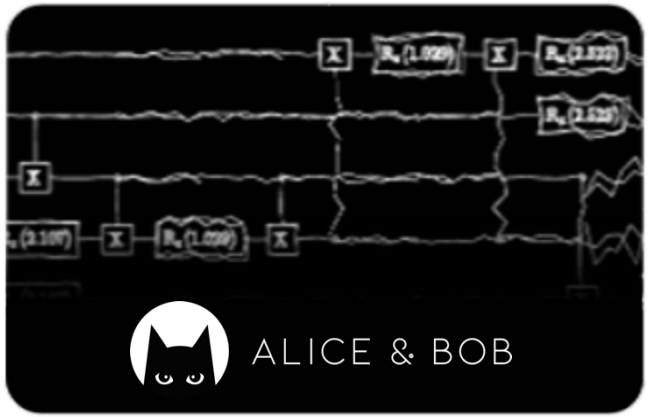
3 Ways

CUDA-X Bootcamp	Open Hackathon	NVAITC Projects
Training	Acceleration	Collaboration

Accelerating Quantum Workloads

Quantum Algorithm Development
40 qubits CFD research

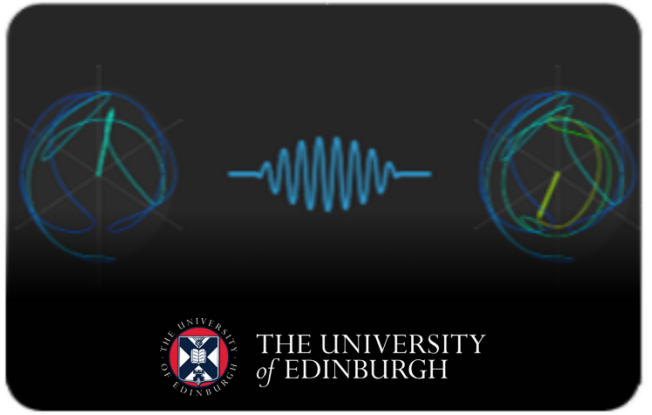
Qubit Design EDA
Dynamiqs software with CUDA-Q



Quantum Data Generation

Hybrid Applications

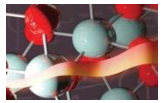
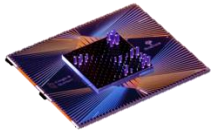
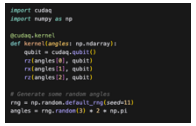
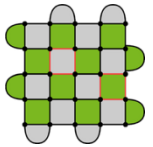
Quantum Error Correction



NVIDIA GB200 NVL72 Powers Quantum

State of the art accelerated computing provides speedups across quantum workloads

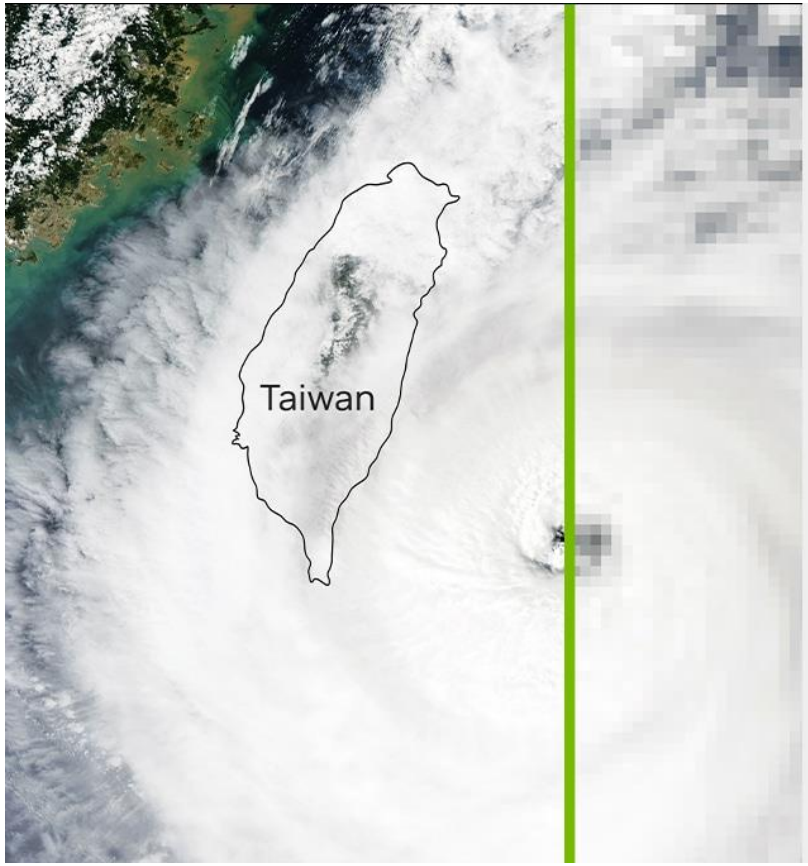
Speedup vs. CPU

800X	1,200X	4,000X	1,300X	500X
				
cuQuantum	cuQuantum	cuQuantum	CUDA-Q	CUDA-Q
Quantum Algorithm Development	Qubit Design EDA	Quantum Data Generation	Hybrid Applications	Quantum Error Correction

NVIDIA 驅動的超級電腦為台灣研究帶來大躍進

1700+ H200 + 2 GB200-NVL72

https://blogs.nvidia.com.tw/blog/taiwan-research-supercomputer/



NVIDIA GB200 NVL72

Delivers New Unit of Compute

GB200 NVL72

- 36 GRACE CPUs
- 72 BLACKWELL GPUs
- Fully Connected NVLink Switch Rack

HPC FP64	2.88 PFLOPs
Training FP8	720 PFLOPs
Inference FP4	1,440 PFLOPs
NVL Model Size	27T params
Multi-Node All-to-All	130 TB/s
Multi-Node All-Reduce	260 TB/s

Co-innovate with Developers

3 Ways

CUDA-X Bootcamp	Open Hackathon	NVAITC Projects
Training	Acceleration	Collaboration

OpenACC Open Hackathon

Celebrating 12 years and continue building the communities.





Ecosystem Development

Training/Education

OpenACC Specification

AI 人工智慧

在 3 週內實現高達萬倍的運算效能提升！NCHC、NVIDIA、OpenACC「NCHC Open Hackathon」黑客松，提供開發者實現 AI 創新最佳平台

NVIDIA 2025-03-13

https://buzzorange.com/techorange/2025/03/13/nchc-open-hackathon-nvidia/

OpenACC More Science, Less Programming
OPEN HACKATHONS
NCHC 國家實驗研究院 國家高速網路與計算中心 National Center for High-performance Computing
NVIDIA

2024/11/13-12/04: Open Hackathon (12 teams)

Team	Mentor	Core Area of Focus	Domain	Languages/ Libs	How much Speedup?	Why acceleration matters?
1-Dream Chaser	Anthony Chang (Engrg-Hardware 1) TW Ying-Ja Chen TW	Protein-small molecule docking	Healthcare Bioinformatics	CUDA	7.8X	Compared to the original AutoDocker-GPU, OmegaDocker supports larger molecular docking simulations. In addition to its high computational intensity, it also demands greater bandwidth.
2-NYCU HPC team2	Shijie Wang CN	Accelerate NVLM 1.0 inference	LLM Multimodal	python/pytorch	41.7X	Multimodal large language models face numerous challenges in inference acceleration, including high computational resource consumption and slow response times. By leveraging the latest inference acceleration technologies, GPU computing power can be fully utilized, effectively reducing inference latency, improving interactive experiences, and expanding application scenarios.
3-氣象署-興大應數聯隊	Leo Chen (Engrg-Hardware 1) TW	Accelerate physics parameterization in weather forcasting model	Weather	Fortran	70.6X	The global weather forecasting model TCo is divided into the dynamical core (GPU) and physical parameterization (CPU). Computationally intensive and data-independent subroutines are ported to the GPU to effectively utilize its computing power.
4-NTUT_BirdSong	Virginia Chen TW Iven Fu TW	Accelerate audio foundation model pretraining	Audio	python/transformer_engine	3.6X	To create a foundation model for bird songs, pre-trainning speed really matters.
5-Parallel Minds	Reese Wang TW	Accelerate firefly algorithm	HPC	CUDA	9X	The Firefly Algorithm has a wide range of applications. It can be applied not only to path prediction (navigation) but also to optimizing renewable energy systems (maximizing solar cell efficiency), gene regulatory network modeling, drug design, image processing, and more.
6-NTHU_LSALAB	Kevin Chen TW Sungta Tsai TW	Acclerate inference process performance	DPU	C/DOCA	1.23X	In current inference and model computations, the CPU is responsible for controlling and transmitting the data to be processed, which limits the GPU's computational power due to data handling speeds. By utilizing DPU I/O to directly access GPU memory, the processing capabilities of inference and models are enhanced.
7-NoLab	Pika Wang TW Ikko Hamamura JP Tian Zheng (Engrg-Hardware 1) CN	Variational quantum eigensolver (VQE)	Quantum Chemistry	CUDA-Q	8282X	The Variational Quantum Eigensolver (VQE) is a promising quantum algorithm for determining the optimal ground-state energy of molecules, a fundamental of chemical reactions and drug discovery. Efficient VQE simulation can help design quantum algorithms, accelerate scientific research, and reducing development cycles.
8-Elsa Robotics	Johnson Sun TW Frank A. Lin (Engrg-Hardware 1) TW	Accelerate robotics navigation pipeline	Robotics	Python/CuPy/TensorRT	11X	Low latency and high frame rates are critical for a robot's real-time responsiveness and operational precision in dynamic environments. Relying solely on CPU computation makes it difficult to meet real-time processing demands; therefore, GPU assistance is required in specific computational stages to enhance frame rates.
9-GBA-VVM	Min Yu CN Leo Chen (Engrg-Hardware 1) TW	Accelerate advection subroutine	Weather	Fortran/OpenACC	18X	The current atmospheric forecasting accuracy is typically at the kilometer scale, while VVM has improved it to the meter scale. This requires extensive computational resources.
10-smile lab	Ken Liao Yang-Hsien Lin TW	Federated Learning for Pathology	Healthcare Histopathology	Python/cuCIM and Pytorch Lightning	5.5X	Whole Slide Images (WSIs) are massive, making patch extraction computationally intensive. Efficient extraction is key for preprocessing in federated learning, where distributed nodes handle large datasets. Accelerating this process with optimized pipelines (e.g., caching, GPU acceleration) reduces preparation time and avoids I/O-related performance losses.
11-Plantmen	Cliff Chiu TW	Acclerate RAG inference pipeline performance	LLM Multimodal RAG	Python, TensorRT	10X	The extisting pipeline for RAG is too slow for the user to get the response on the LINE chatbot.
12-CYCU_Quantum	Pika Wang TW Ikko Hamamura JP Anderson Meng TW	Quantum PageRank	Quantum Machine Learning	CuPy	10000X	Quantum PageRank leverages quantum interference to reveal complex relationships between nodes, enhancing rankings in applications for websites and social networks. However, simulating noisy interference is much more complex than typical statevector simulation (2^2n vs. 2^n). Further speedup is required to investigate Quantum PageRank in real-world scenarios, which has not yet been achieved.

Other Publications

- 工商: https://www.ctee.com.tw/news/20250408700939-431204?utm=LINE_share_btn
- 經濟: https://money.udn.com/money/story/5635/8659767
- 引新聞: https://innews.com.tw/223620/
- Line: https://today.line.me/tw/v2/article/yzvjnGz

NVIDIA Developer Program

NVIDIA 開發者計畫

Program Benefits:

Tools

- 900+ exclusive SDKs and models
- GPU-optimized software, model scripts, and containerized apps
- Early access programs

Training

- Research papers, technical documentation, webinars, blogs, and news
- Technical training and certification opportunities
- 1,000s of technical sessions from industry events On-Demand

Community

- NVIDIA developer forums
- Exclusive meetups, hackathons, and events

Special Program (Present to Jan 2024)

- Join NVIDIA Developer program now, you will get one NVIDIA Training

Join the Community

Feedback Survey

https://forms.office.com/r/y2gbvErzcd


NVIDIA
Thank You