



Overview of Deep Learning Innovations in Healthcare

Dr CK Lee, 2023

Leader of NVAITC-Taiwan; Senior Solution Architect

NVIDIA AI Technology Center (NVAITC) - Taiwan



My Mission today

Welcome to Join NVIDIA Developer

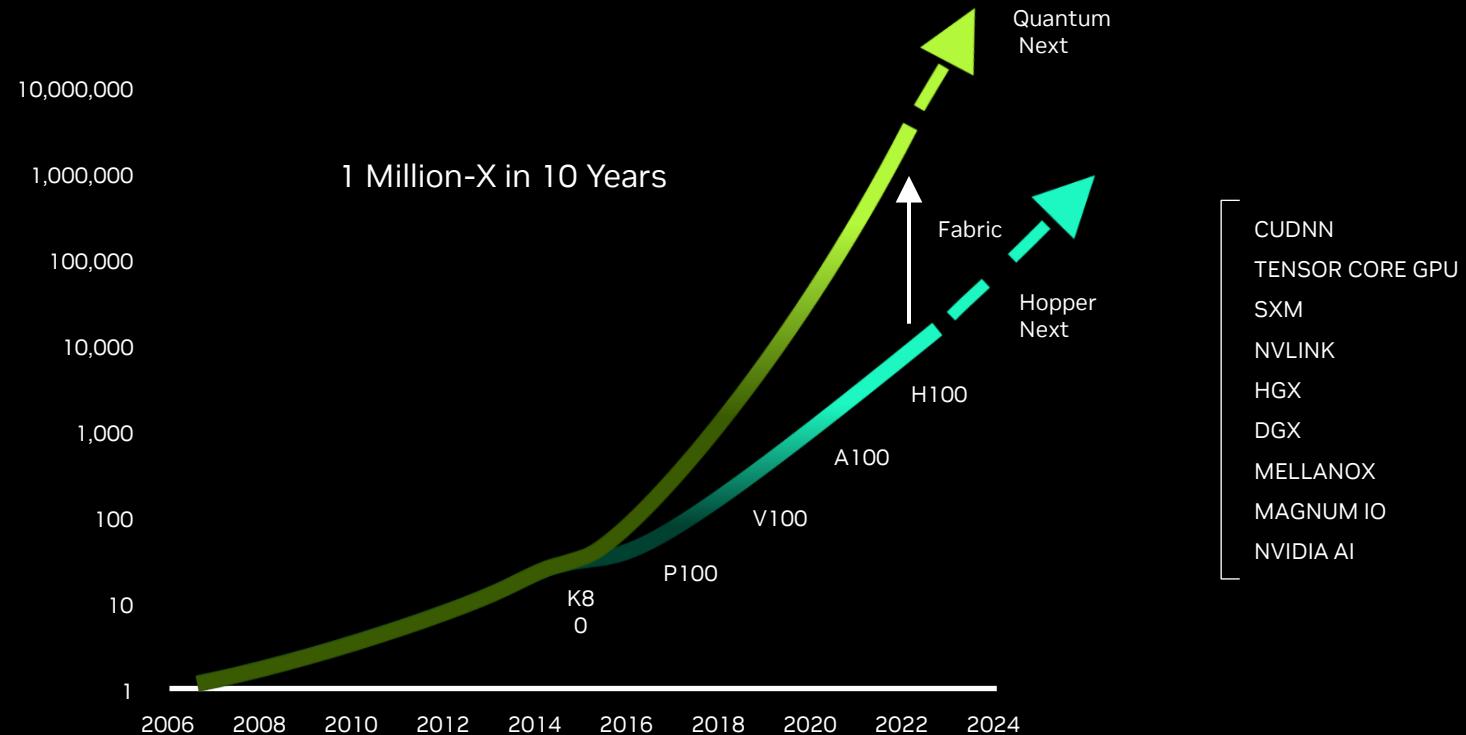
- Encourage you to join
<https://developer.nvidia.com/>
- What is the goal of NVIDIA?
- Overview of Deep Learning in Healthcare
 - Medical Imaging - MONAI
 - Medical Device - Holoscan
 - Drug discovery - BioNemo
 - Genomics - Parabricks





NVIDIA — A COMPUTING PLATFORM COMPANY

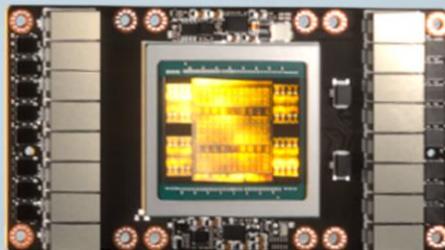
Reinventing Computing For Software 2.0



AT THE INTERSECTION OF GRAPHICS, HPC, AI

We simulate worlds, physics,
and intelligence in real time

We make computers for the da
Vincis and Einsteins of our age



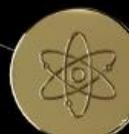


APPLICATION FRAMEWORKS

PLATFORM

SYSTEM SOFTWARE

HARDWARE



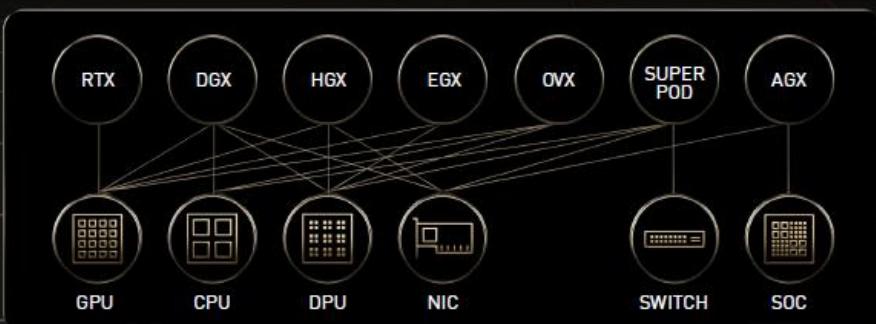
NVIDIA HPC



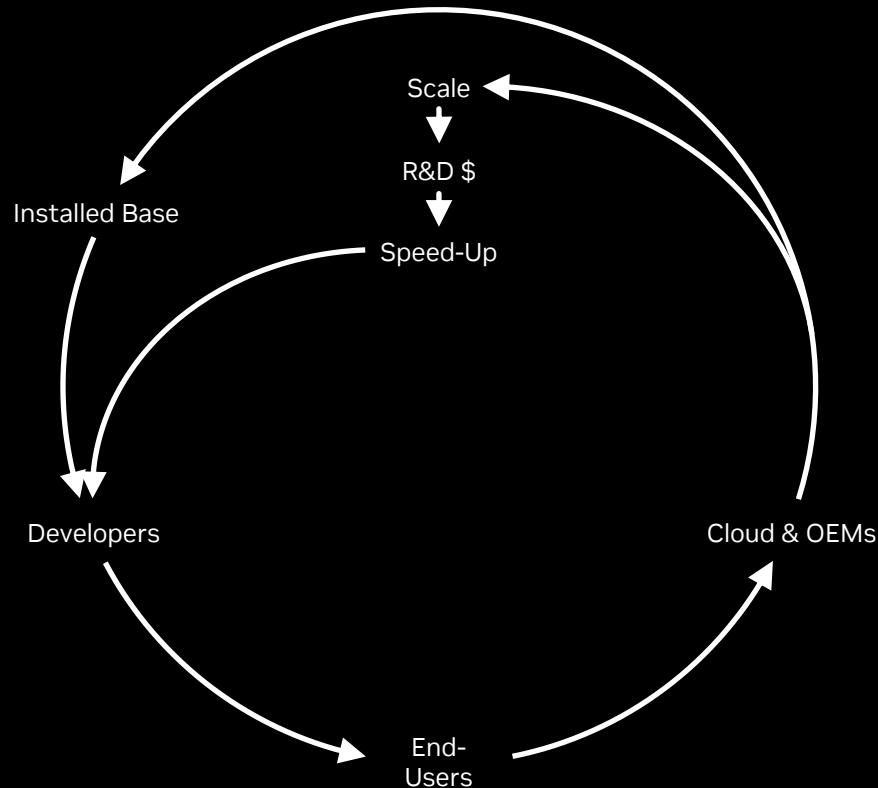
NVIDIA AI



NVIDIA OMNIVERSE



NVIDIA Accelerated Computing Virtuous Cycle

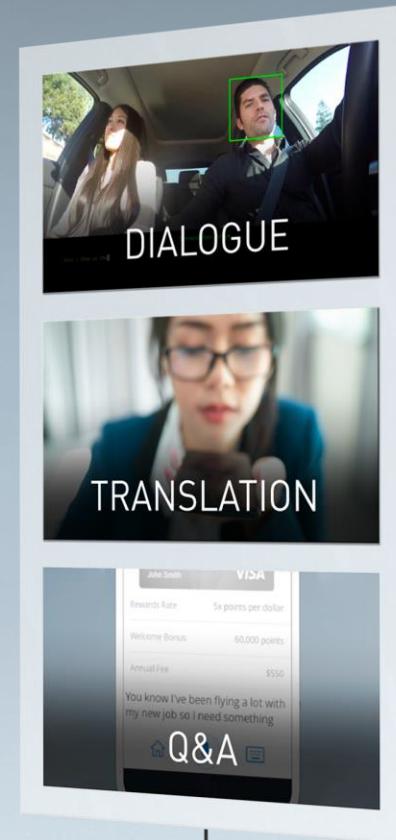
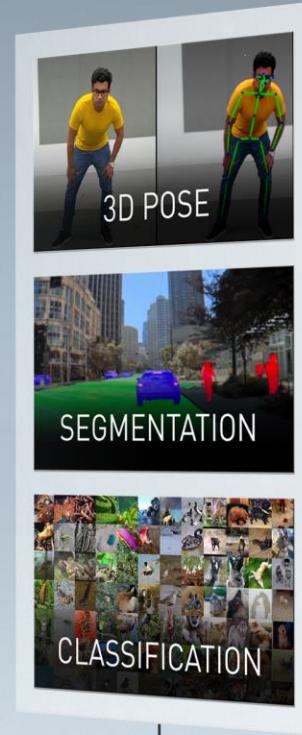


- 4 Million Developers
- 3000+ Accelerated Applications
- 40 Million+ CUDA Downloads – 25 Million in 2022
- 15,000+ Startups
- 40,000 Enterprises

TWO GIANT WAVES OF AI

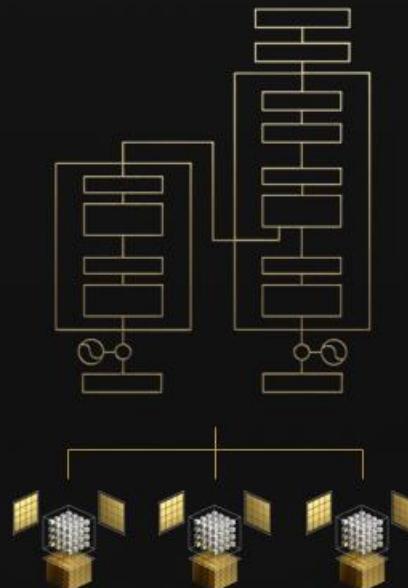
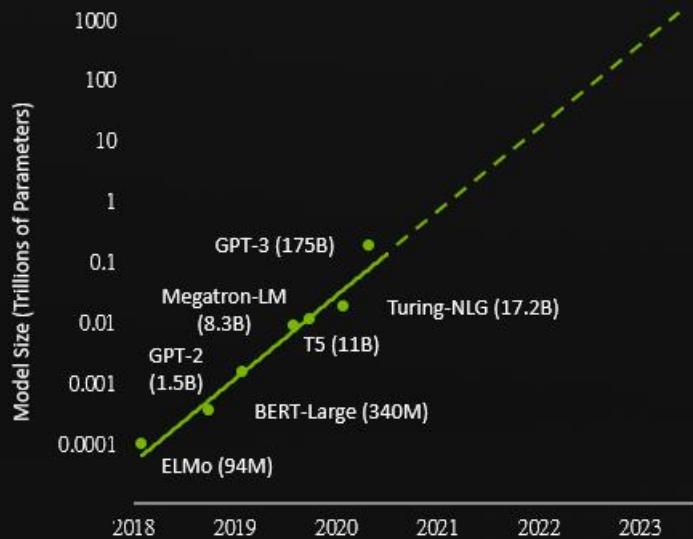
2012: GPU-accelerated AlexNet ushered in the era of superhuman image recognition

2018: BERT opens a new era of natural language understanding and conversational AI



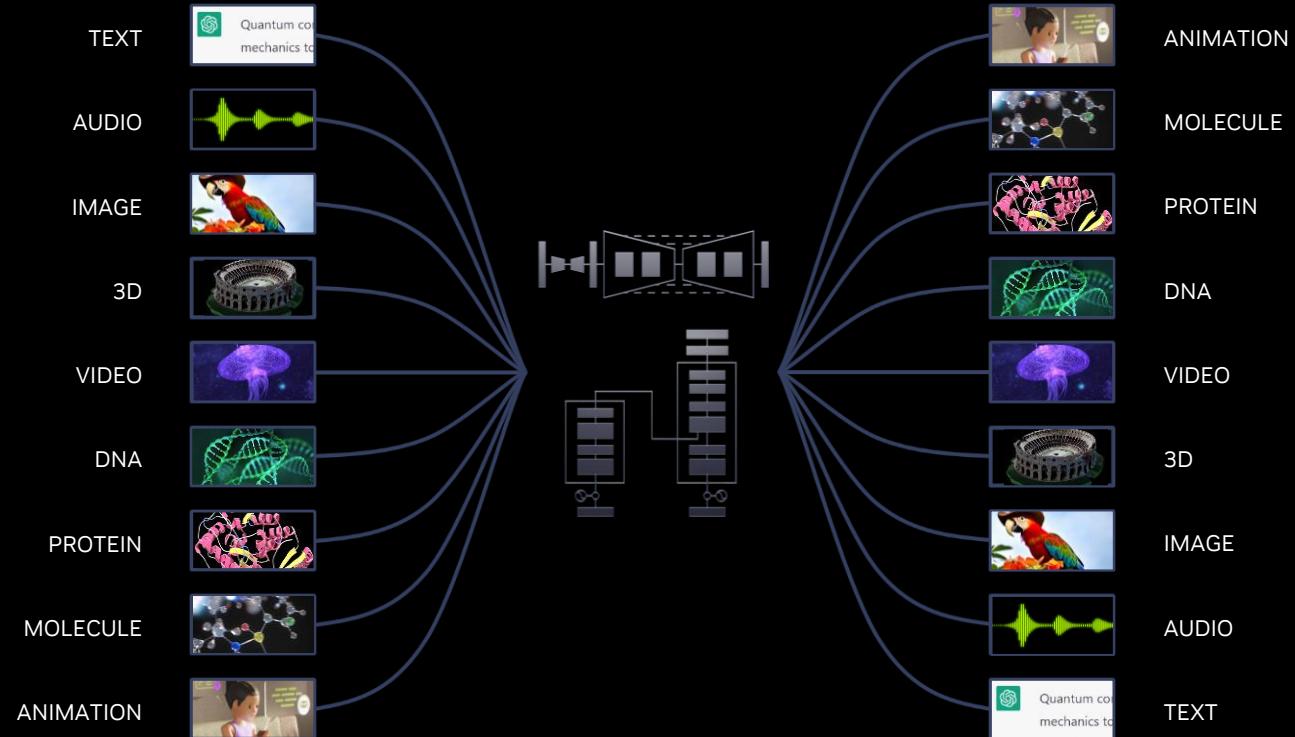
NVIDIA MEGATRON TRAINS TRANSFORMERS

100 TRILLION PARAMETERS MODELS BY 2023



Generative AI

The iPhone moment of AI



NVIDIA Clara for Healthcare and Life Sciences

World's Largest Data Industry | 36% CAGR by 2025



IMAGING



MEDICAL
DEVICES



DRUG
DISCOVERY



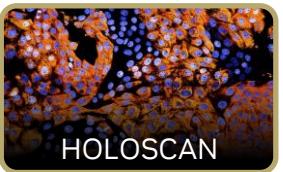
GENOMICS



NVIDIA CLARA
APPLICATION
FRAMEWORKS



FLARE



HOLOSCAN



BIONEMO



PARABRICKS

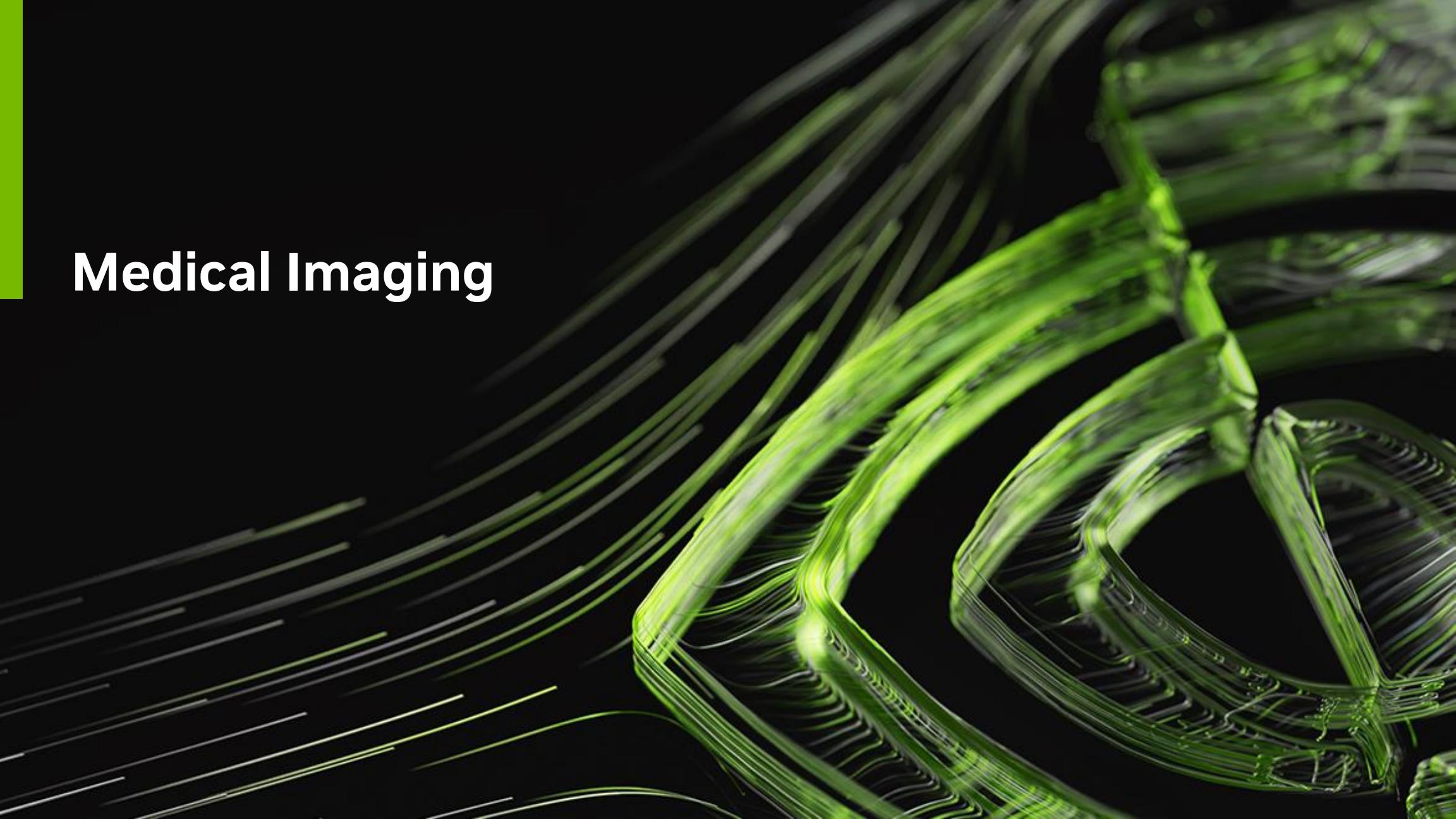


NVIDIA AI



NVIDIA Omniverse

Medical Imaging





What is MONAI?

Project MONAI is a collaborative open-source initiative built by academic and industry leaders for deep learning in healthcare imaging.

MONAI Advisory Board

Bringing Together Domain Experts & Leaders



Stephen Aylward

Chair of the Advisory Board

Sebastien Ourselin

Klaus Maier-Hein

Jayashree Kalpathy-Cramer

Jorge Cardoso

Daniel Rubin

Lena Maier-Hein



Kevin Zhou

Nassir Navab

Andrew Feng

Nasir Rajpoot

Justin Kirby

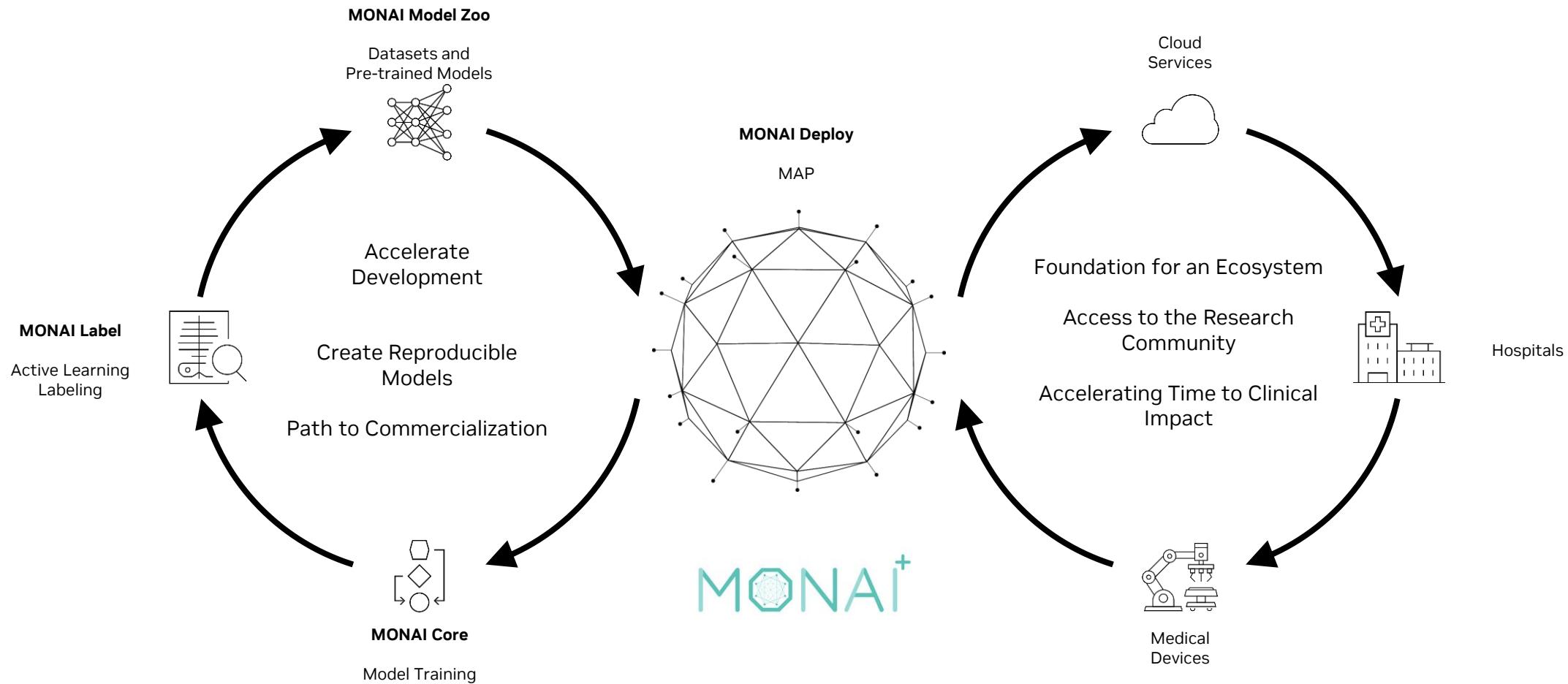
Keyvan Farahani

Selnur Erdal

MONAI stands for Medical Open Network for AI and is a collaborative open-source projects for building the best practices for deep learning in Medical Imaging

MONAI

An AI acceleration framework for medical imaging



NVIDIA Solutions for Building AI Imaging Solutions

Domain-Specific Models, Frameworks, SDKs, and Applications

Application
Frameworks

MONAI[†]
Label

MONAI[†]
Core

MONAI[†]
Deploy

Acceleration
Libraries



TensorFlow

Accelerated with
NVIDIA.

PyTorch

Accelerated with
NVIDIA.

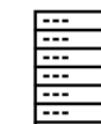


Systems

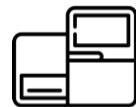
EGX | DGX
A100 | A40 | A10

DGX | HGX | EGX
A100 | A40 | RTX A6000

EGX | DGX
A100 | A40 | A10



Data Center



Edge



Workstation



Cloud



Embedded

MONAI Adoption Across the Healthcare Ecosystem

Academic Medical Centers | Enterprise | Cloud Service Providers | Clinical Platforms

AMCs and Enterprise



Frederick National Laboratory
for Cancer Research



NATIONAL
CANCER
INSTITUTE

Guy's and St Thomas'
NHS Foundation Trust



TUM
Technische Universität München

UNITED IMAGING



Hyperscaler Integration

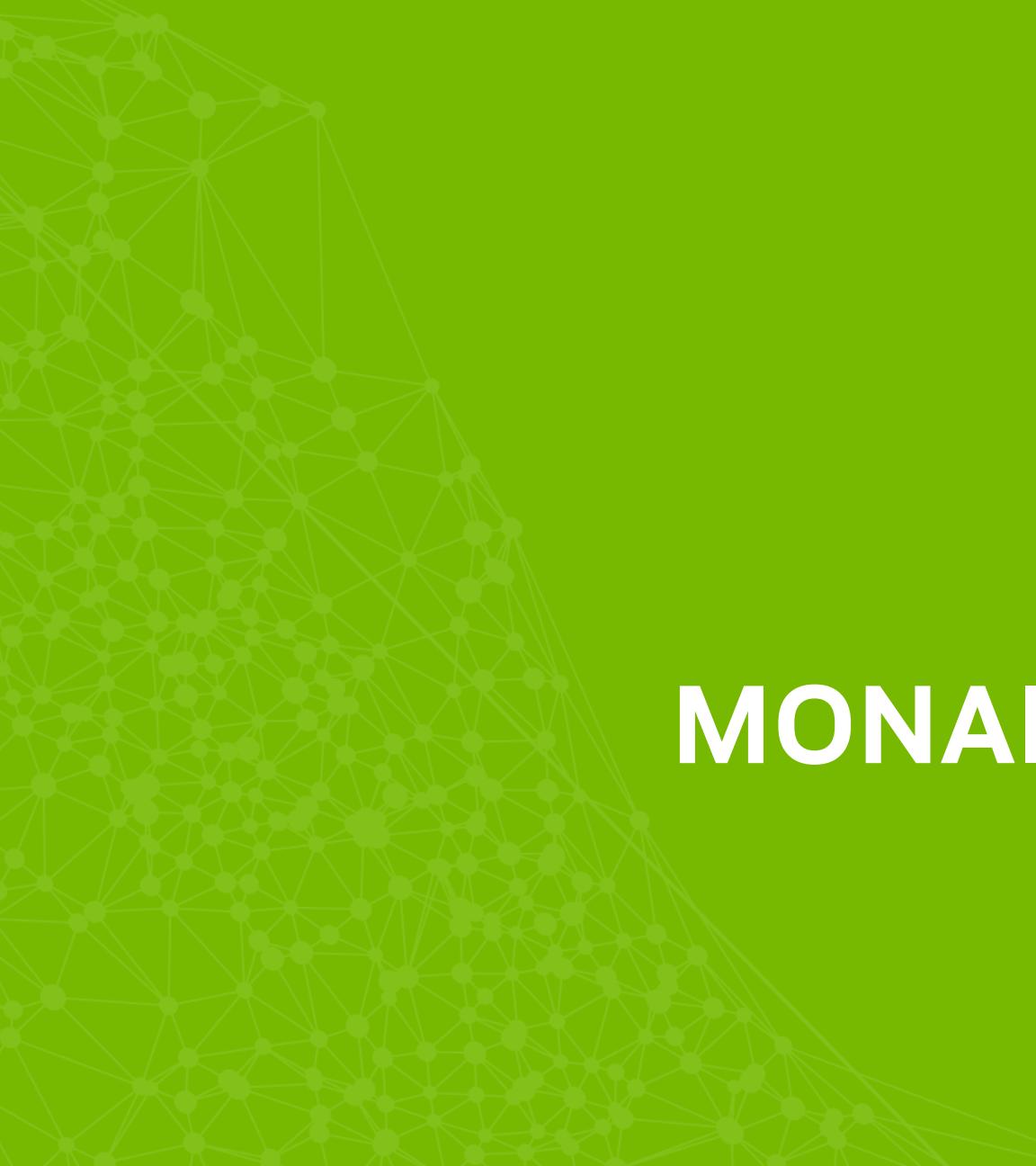


ORACLE CLOUD
Infrastructure

Clinical Deployment Platforms



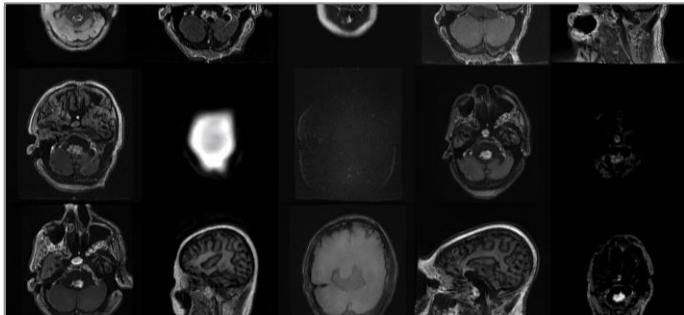
quare.ai



MONAI FEATURES AND FUNCTIONALITY

MONAI Solves 6 Key Workflows

Accelerating AI Solutions in Healthcare Imaging AI



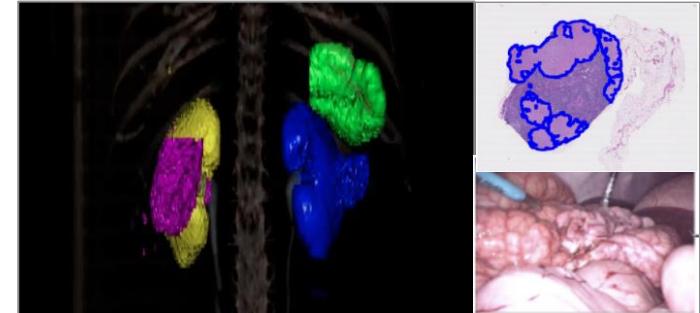
Generating Synthetic Data

Generative AI Models & Tooling



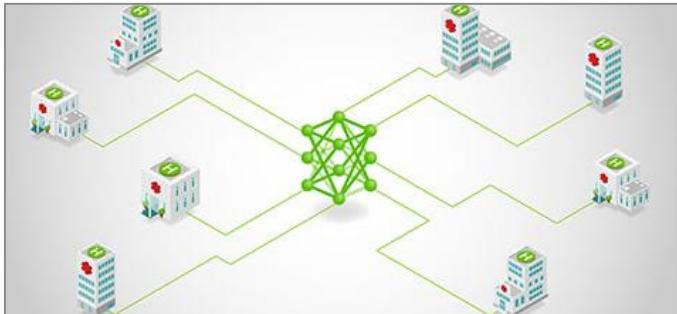
AI-Assisted Annotation

Intelligent and interactive AI labeling



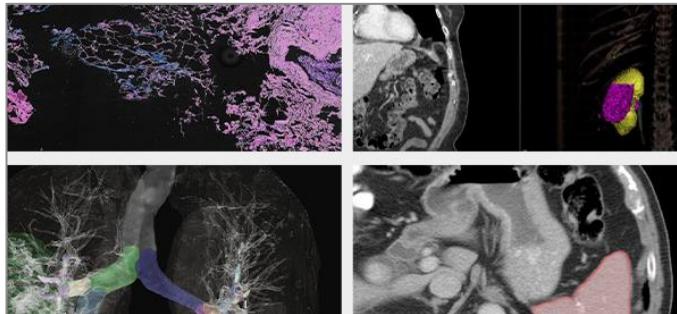
Training Imaging AI Models

Intelligent model selection and training



Federated Learning

Collaborative, privacy-preserving model development



Translating AI Models to Applications

Build and package models into deployable applications

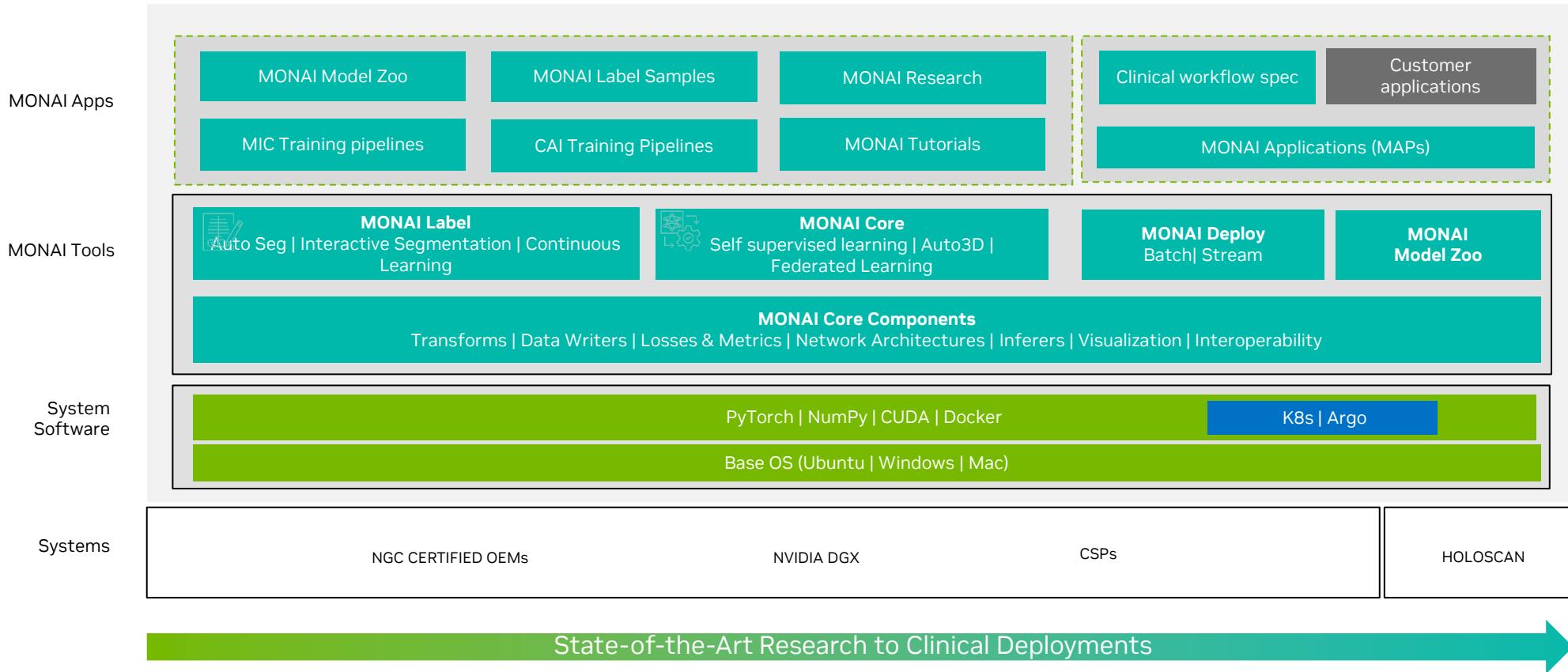


Deploying to Hospital Infrastructure

Standardize end-to-end clinical workflows

MONAI Stack

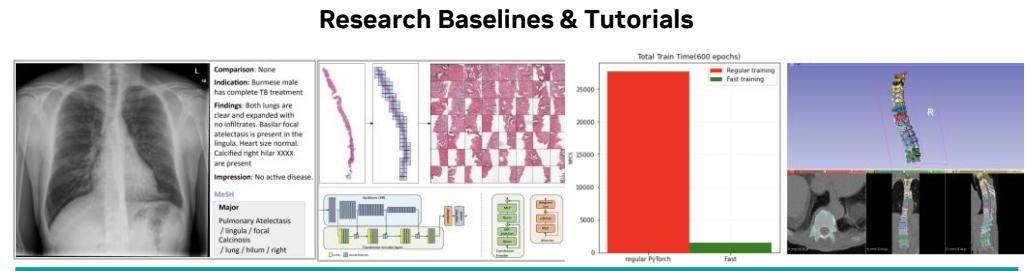
Model Zoo | Label | Core | Deploy



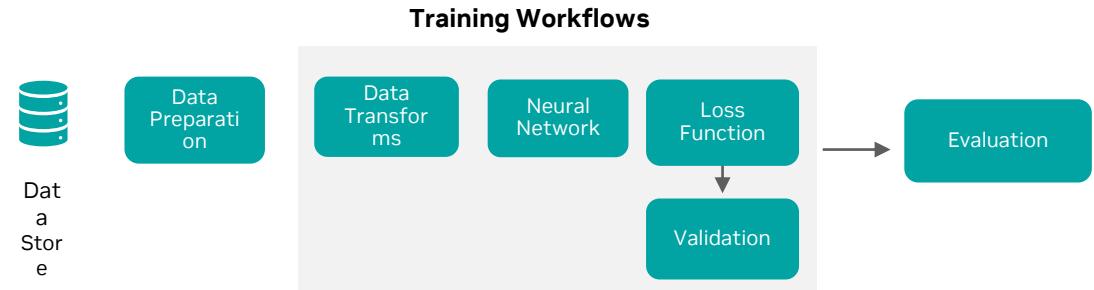
MONAI Core

Train with domain-optimized AI capabilities

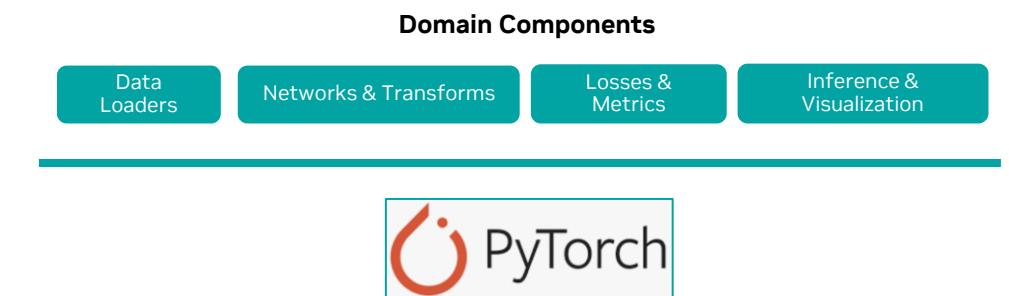
- Robust, backward compatible API design
- Supports new MetaTensor API to store Torch Tensors and meta information.
- Modular domain specialized components: network arch; transforms; losses; metrics; data handlers etc.
- Learning workflows: Supervised; Self-supervised; Federated Learning; Network architecture search
- Up to 15x faster performance as compared to vanilla implementations



Research Baselines & Tutorials



Training Workflows



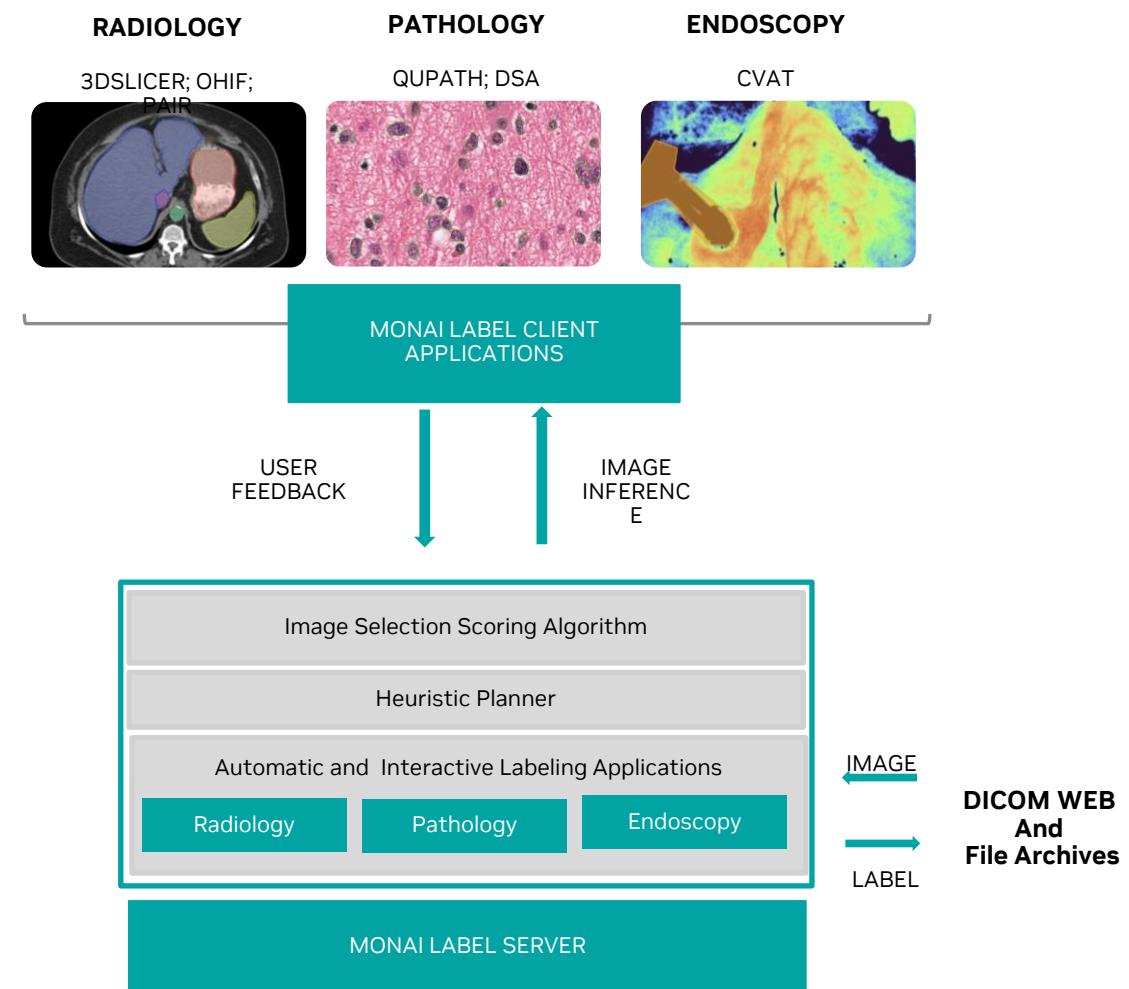
Domain Components



MONAI Label

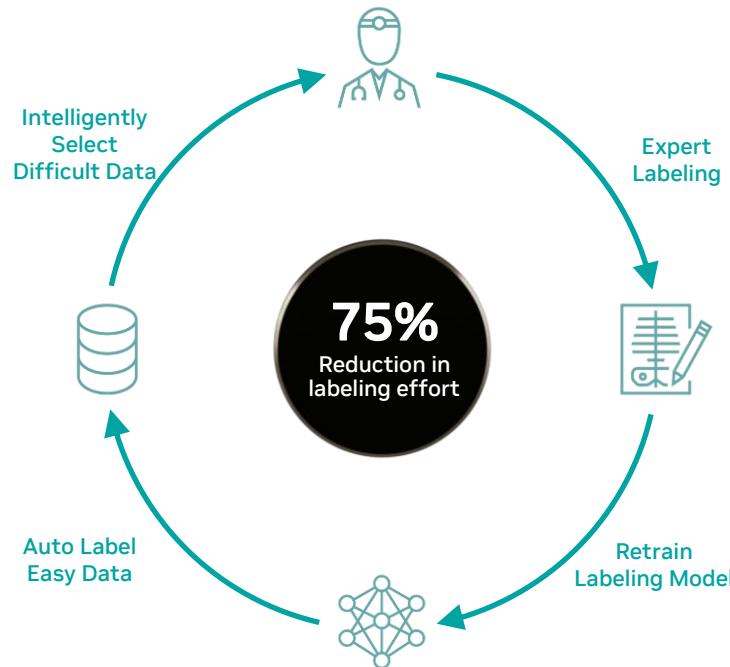
Save time with an intelligent data labeling

- Created annotated datasets and build AI annotation models for clinical evaluation
- Support for 3D segmentation (radiology), pathology (nuclei detection), and 2D segmentation (endoscopy video).
- Client Viewer Integrations: 3DSlicer, OHIF, PAIR, Digital Slide Archive, QuPath, CVAT (MONAI Label is a plug in into other viewers, not a viewer)
- MONAI Label Server makes applications ready to deploy and serve as a service



MONAI LABEL: Active Learning

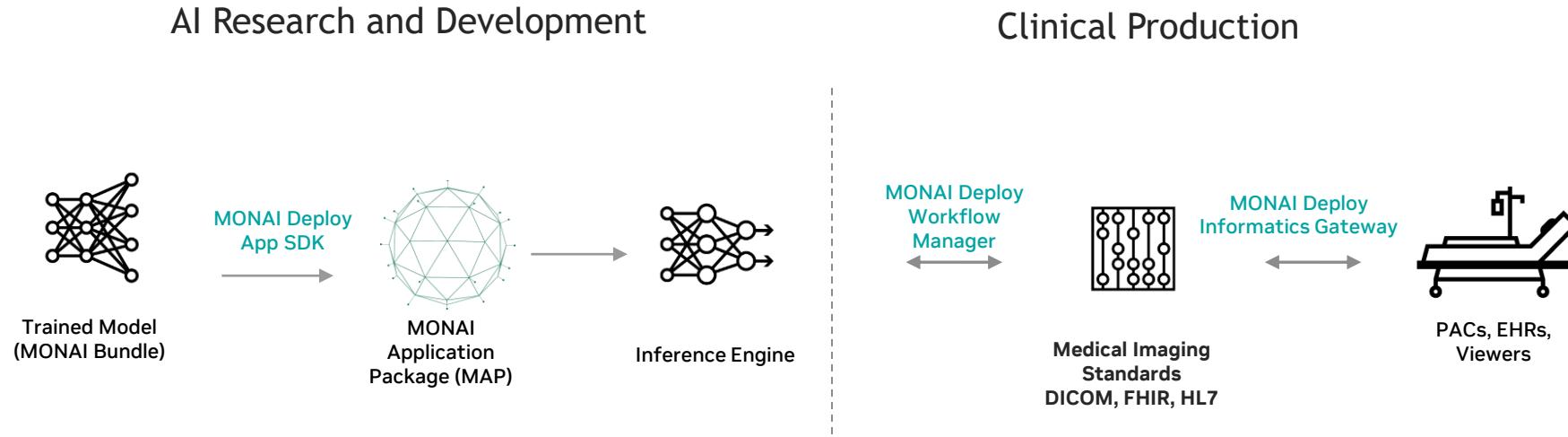
Reduce the cost of labeling imaging data



Features	Benefits
Support 3D Segmentation for Radiology & 2D Endoscopy <small>NEW</small>	2D Endoscopy benchmark showcase 75% reduction in need for labeled data Charite University reduced labeling cost by 50% using Epistemic algorithm
Image selection algorithms (Aleatoric, Epistemic and FAS) intelligently chooses which data to learn from	Train models that can learn more from less data.
Human-in-the loop to annotate the most difficult images for clinical inputs	Focus on annotations that provide the highest gain in model performance and address areas where the model has uncertainty
1-click Train from Viewer UI	Seamlessly initiate training with high value annotated data bringing clinicians into data science loop

MONAI Deploy

End-to-end workflow to deploy AI, from bench to bedside



Accelerate Application Development

Build AI applications with few lines of code and package into a MONAI Application Package (MAP) in < 20 min.

Run Anywhere

Integrate into health IT standards and deploy across data center, cloud and edge environments.

Streamline Hospital Operations

Maintain AI governance for Hospital IT

MONAI Bundle

Open standard for Model definition & Package

Execute training:

```
python -m monai.bundle run training \
--meta_file configs/metadata.json \
--config_file configs/train.json \
--logging_file configs/logging.conf
```



Override the `train` config to execute multi-GPU training:

```
torchrun --standalone --nnodes=1 --nproc_per_node=2 -m monai.bundle run training \
--meta_file configs/metadata.json \
--config_file "[configs/train.json,'configs/multi_gpu_train.json']" \
--logging_file configs/logging.conf
```

Override the `train` config to execute evaluation with the trained model:

```
python -m monai.bundle run evaluating \
--meta_file configs/metadata.json \
--config_file "[configs/train.json,'configs/evaluate.json']" \
--logging_file configs/logging.conf
```

Execute inference:

```
python -m monai.bundle run evaluating \
--meta_file configs/metadata.json \
--config_file configs/inference.json \
--logging_file configs/logging.conf
```

Verify the metadata format:

```
python -m monai.bundle verify_metadata --meta_file configs/metadata.json --filepath eval/schema.json
```

Verify the data shape of network:

```
python -m monai.bundle verify_net_in_out network_def --meta_file configs/metadata.json --config_file configs/inference.json
```

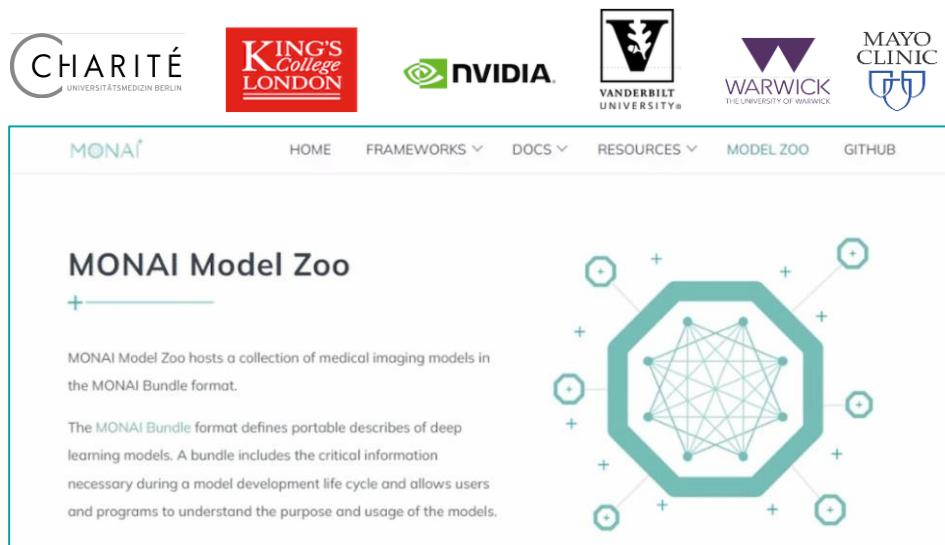
Export checkpoint to TorchScript file:

```
python -m monai.bundle ckpt_export network_def \
--filepath models/model.ts \
--ckpt_file models/model.pt \
--meta_file configs/metadata.json \
--config_file configs/inference.json
```

- MONAI Bundle is a self-contained model package with pre-trained weights and all meta data
- Build python workflows via structured configs
- Develop once and deploy anywhere
- Ease of use & flexibility to override & customize configs
- Hybrid programming with support for config to python conversion

MONAI Model Zoo

A Hub for Pre-Trained Imaging AI Models



The screenshot shows the MONAI Model Zoo homepage. At the top, there's a navigation bar with the MONAI logo, followed by links to HOME, FRAMEWORKS, DOCS, RESOURCES, MODEL ZOO, and GITHUB. Below the navigation, the title "MONAI Model Zoo" is displayed, along with a brief description of what it is. To the right of the text is a graphic of a neural network with a green octagonal border and green circular nodes with plus signs.

Features

15 and growing pre-trained models across CT, MR, Pathology, Endoscopy.

Benefits

Jumpstart training workflows

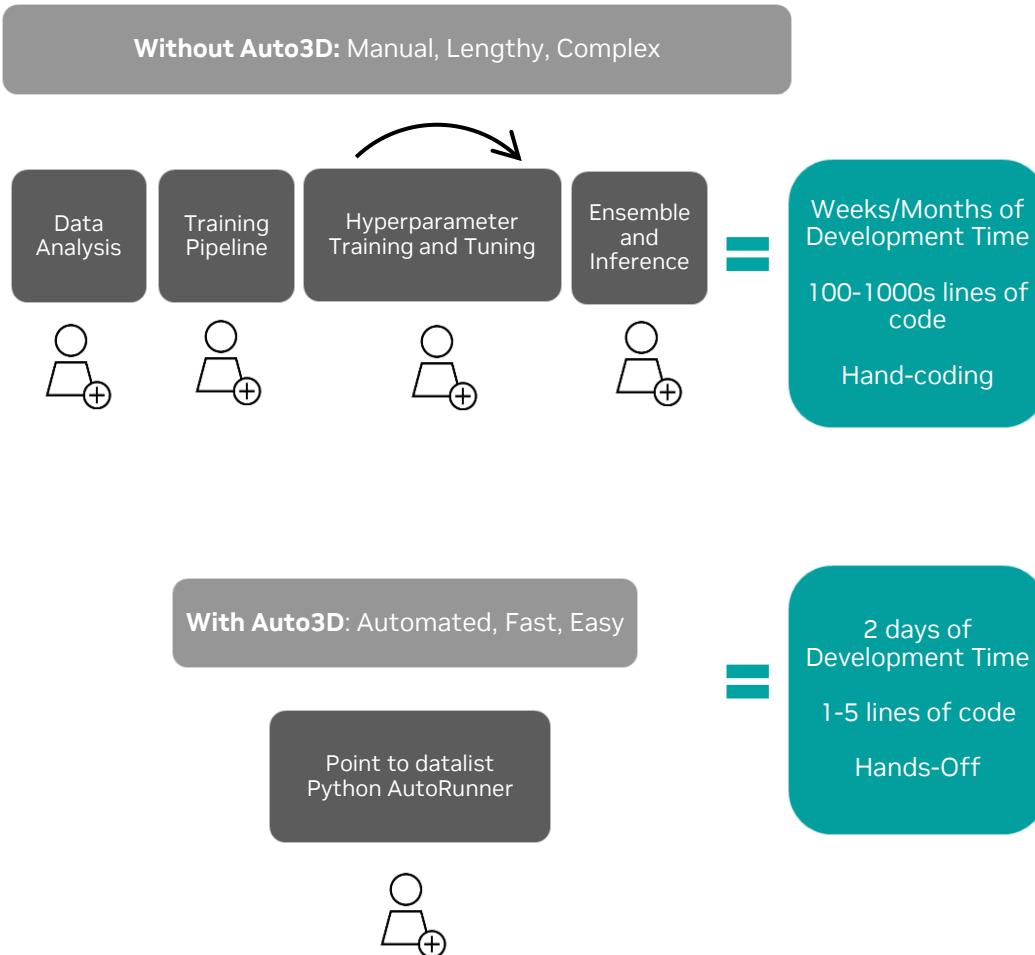
Establish common standard for reproducible research & collaboration

Broaden reach and impact of research

Leverage the SOTA pre-trained models for downstream clinical tasks

AUTO3D Segmentation

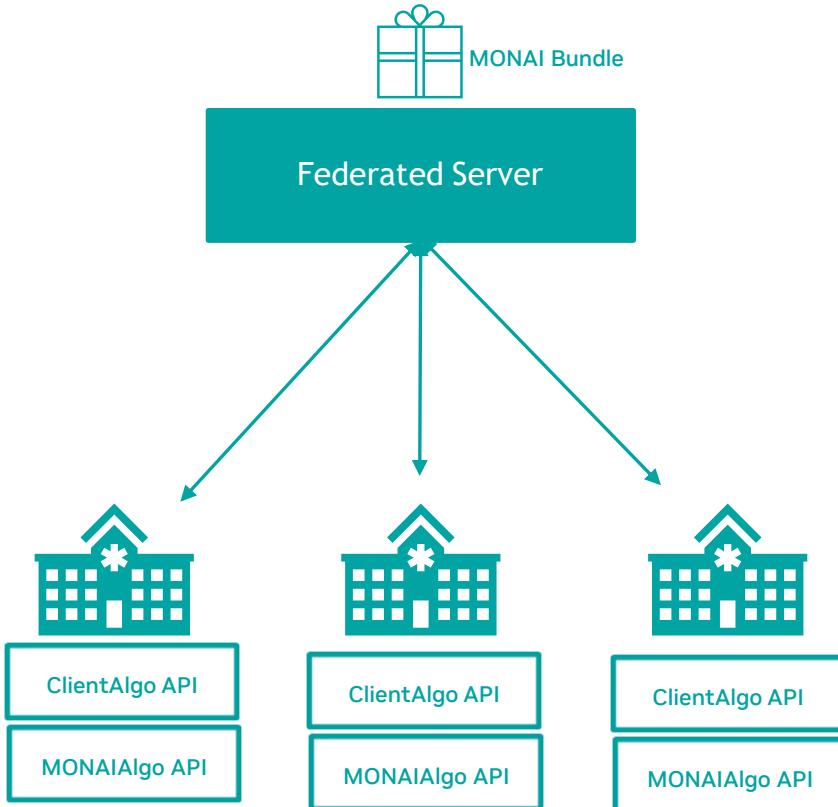
Low-code framework to train high-quality 3D Segmentation Models



- Auto3D is a self-contained solution to build 3D Segmentation models with minimal user inputs
- Supports state of the art algorithms: SwinUNETR; DiNTS; SegResNet out of the box and ability to customize
- Built for high Accuracy & efficiency Auto3D maximizes the productivity of data science teams.
- Auto3D based solution with no human interaction amongst MICCAI challenges winning solutions: HEKTOR22, Instance22, ISLES22

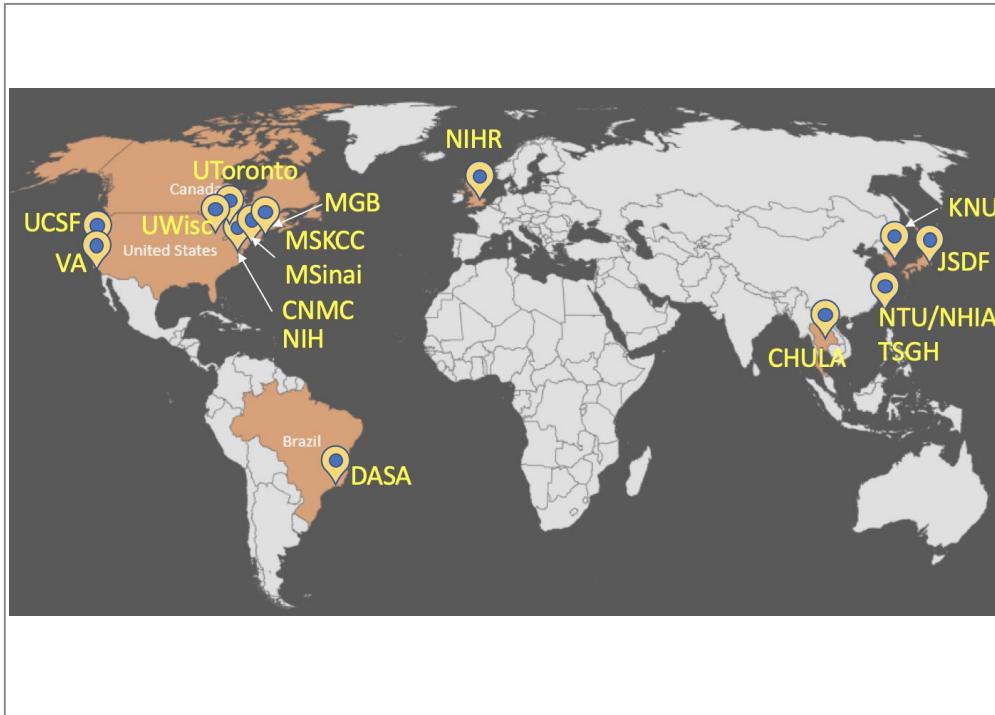
MONAI Federated Learning

Low Code Framework for high quality 3D Segmentation Models

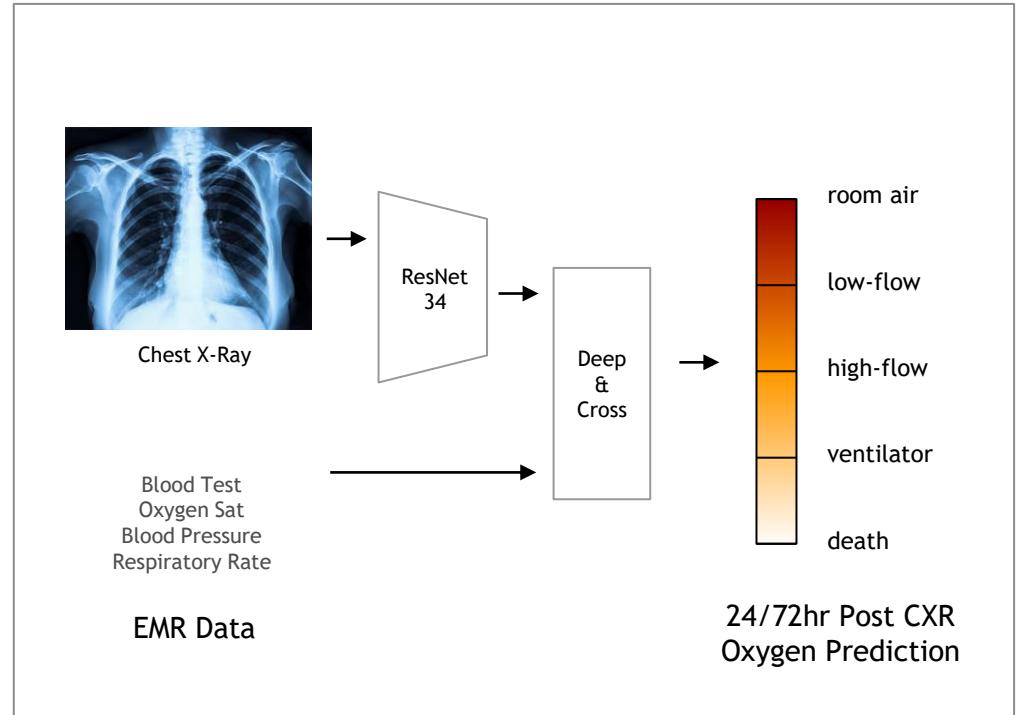


Features	Benefits
Federated Learning Client Algo APIs	Ability to defining a MONAI client app that can run on any FL platform. Enabler for FL Toolkits interoperability.
NVIDIA FLARE Integration	Nvidia FLARE, the federated platform developed by Nvidia already integrated with MONAI FL Client APIs
MONAI Bundle compatibility	Allow for seamlessly extending any bundle from MONAI zoo to a federated paradigm

CLARA FEDERATED LEARNING FOR COVID-19 PATIENT CARE “EXAM” AI MODEL



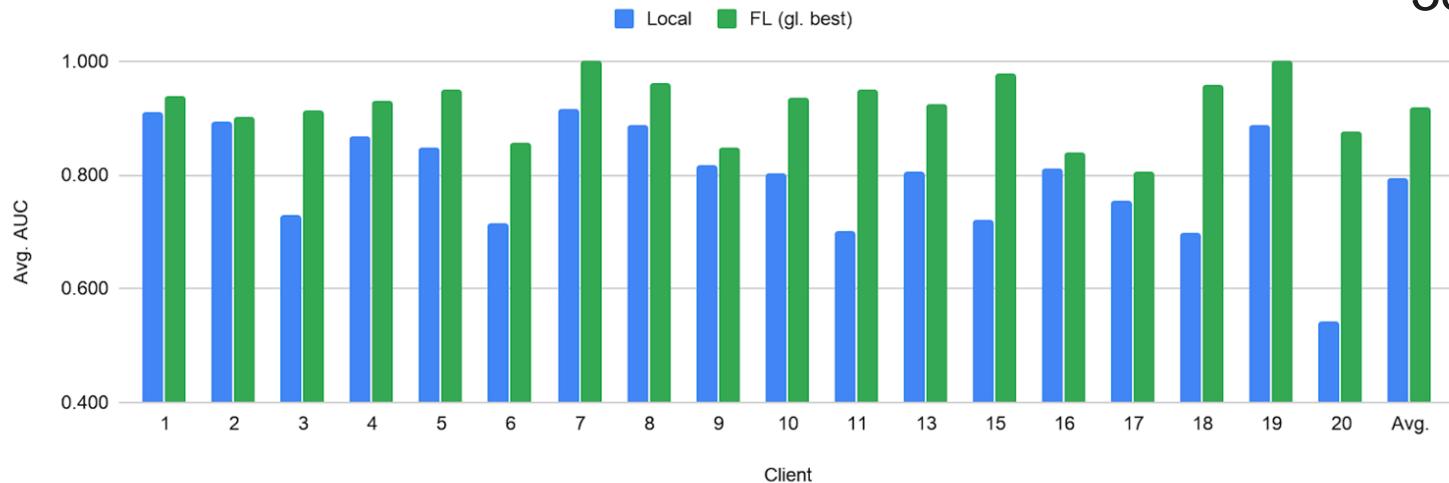
Clara Federated Learning
20 Sites | 8 Countries
COVID-19 Oxygen Prediction



Global Model Achieved .93AUC
>25% Relative Improvement
Every Site Benefited Regardless of Dataset Size

CLARA FL FOR COVID-19

size-based ordering



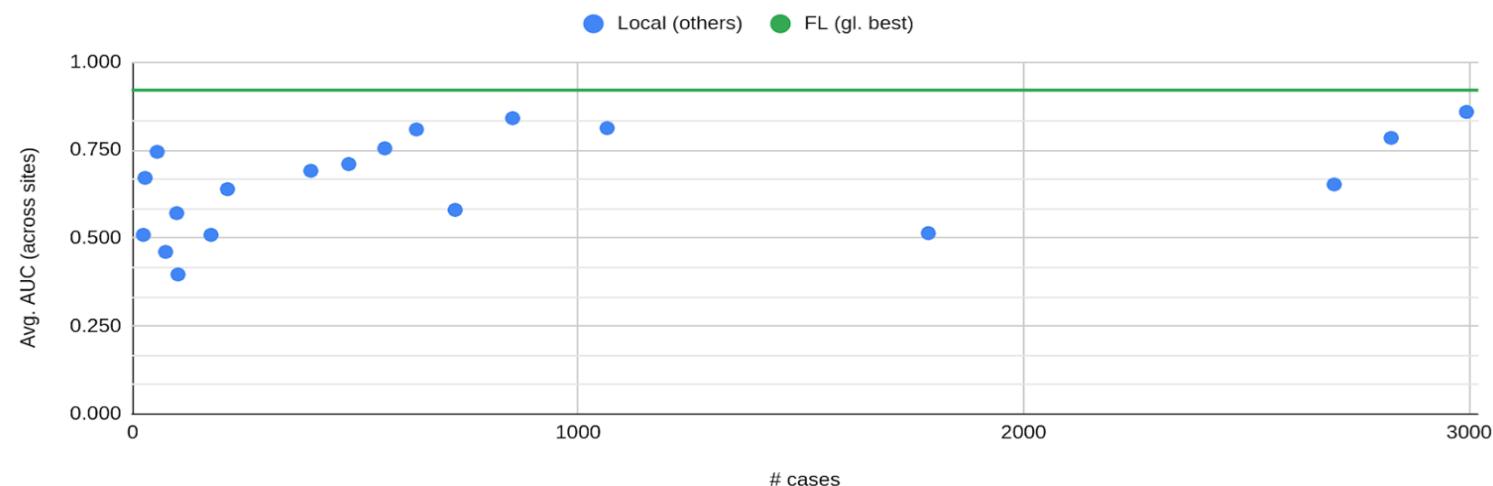
FL resulted on average in
16% performance improvement
38% generalizability improvement

EXAM Model:

24h avg. AUC: 0.94

72h avg. AUC: 0.91

Size vs. generalizability



NVIDIA Clara for Healthcare and Life Sciences

World's Largest Data Industry | 36% CAGR by 2025



IMAGING



MEDICAL
DEVICES



DRUG
DISCOVERY



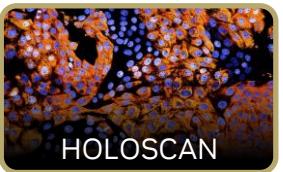
GENOMICS



NVIDIA CLARA
APPLICATION
FRAMEWORKS



FLARE



HOLOSCAN



BIONEMO



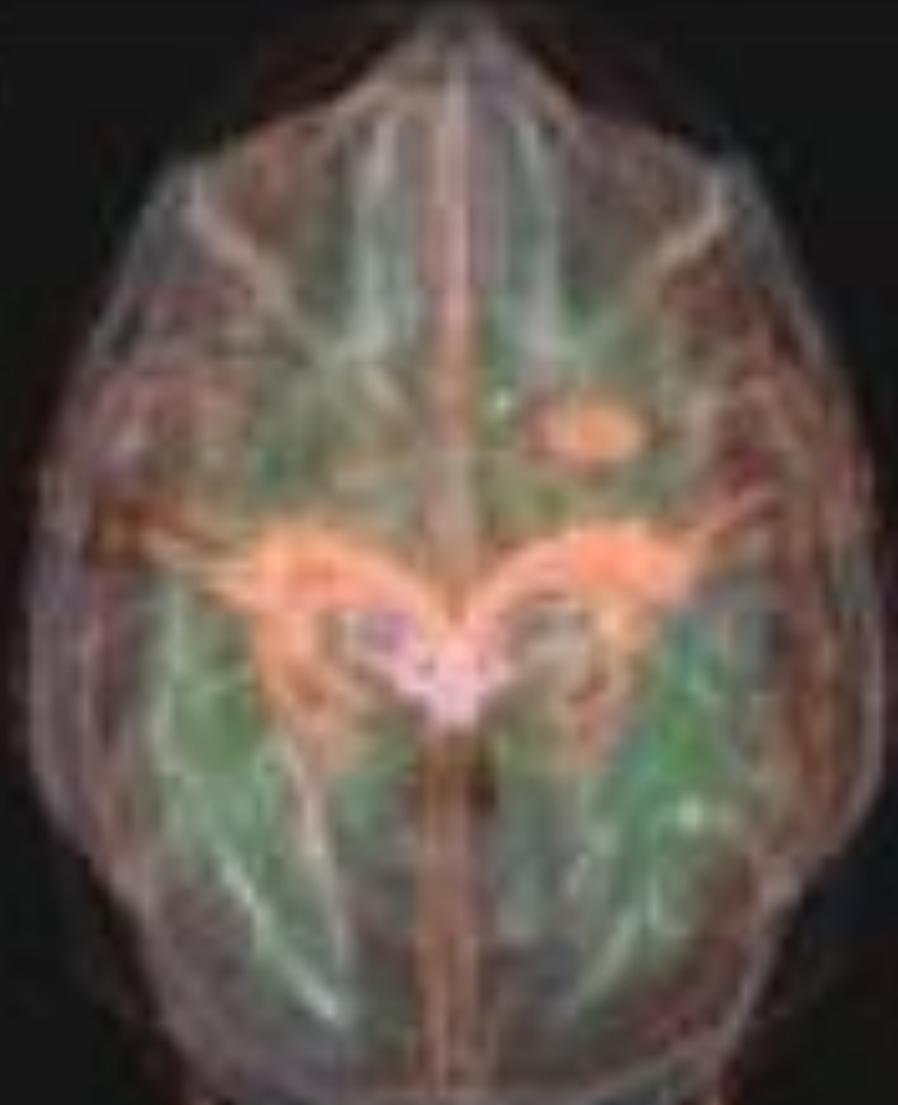
PARABRICKS



NVIDIA AI



NVIDIA Omniverse



<https://www.youtube.com/watch?v=VN5SOV7qOks>

Join MONAI's Growing Community

Collaborate with the top developers and industry experts in Medical AI

Install

<https://github.com/Project-MONAI>

MONAI Core

<https://github.com/Project-MONAI/MONAI>

MONAI Label

<https://github.com/Project-MONAI/MONAILabel>

MONAI Deploy

<https://github.com/Project-MONAI/monai-deploy>

Learn

MONAI Docs

<https://monai.io/docs.html>

MONAI YouTube

<https://www.youtube.com/c/Project-MONAI>

MONAI Medium

<https://monai.medium.com/>

Collaborate

MONAI Slack

<https://forms.gle/QTxJq3hFictp31UM9>

MONAI Twitter

<https://twitter.com/ProjectMONAI>

NVIDIA for Medical Devices



NVIDIA Clara for Healthcare and Life Sciences

World's Largest Data Industry | 36% CAGR by 2025



IMAGING



MEDICAL
DEVICES



DRUG
DISCOVERY



GENOMICS

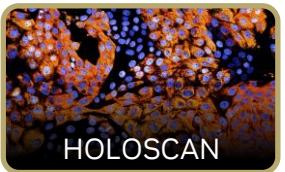


NVIDIA CLARA
APPLICATION
FRAMEWORKS



MONAI

FLARE



HOLOSCAN



BIONEMO



PARABRICKS



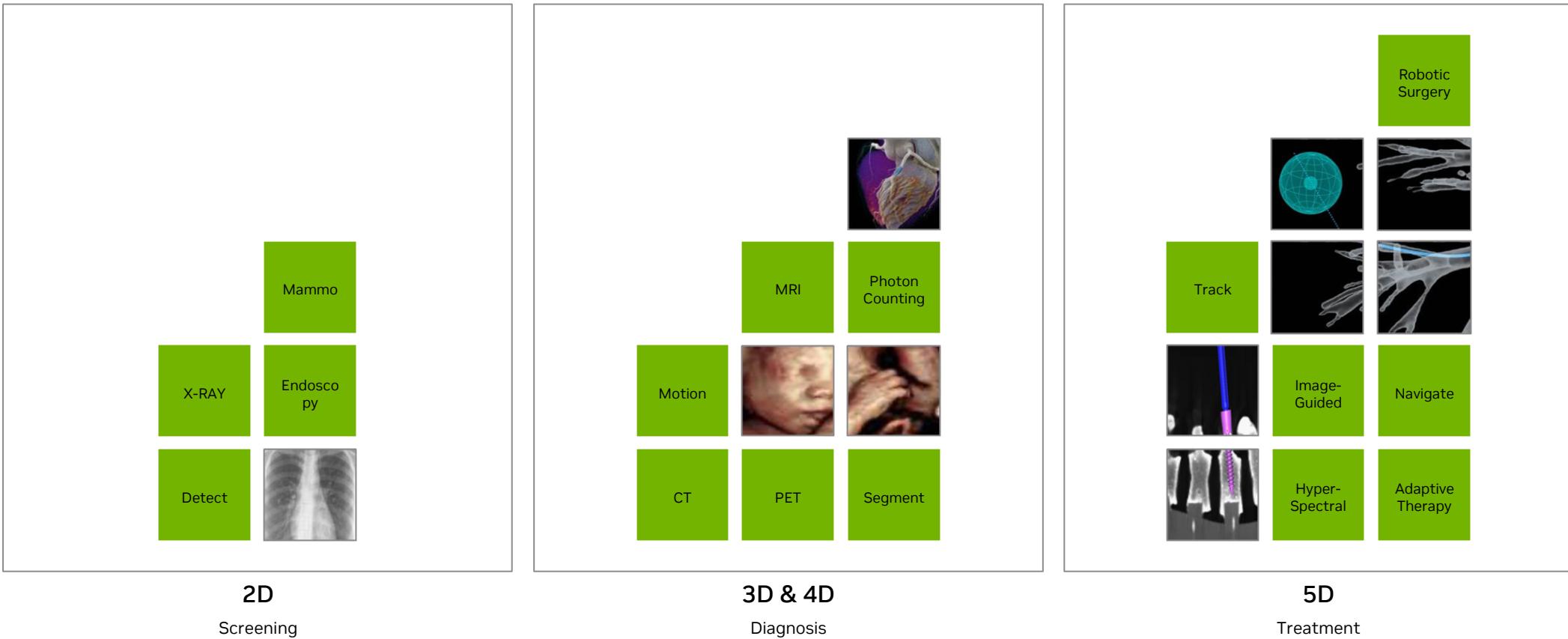
NVIDIA AI



NVIDIA Omniverse

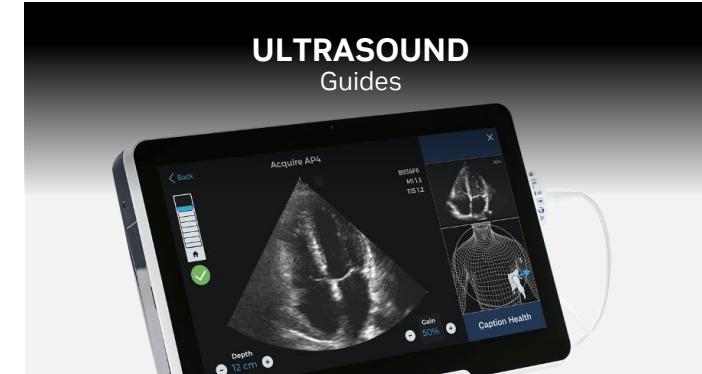
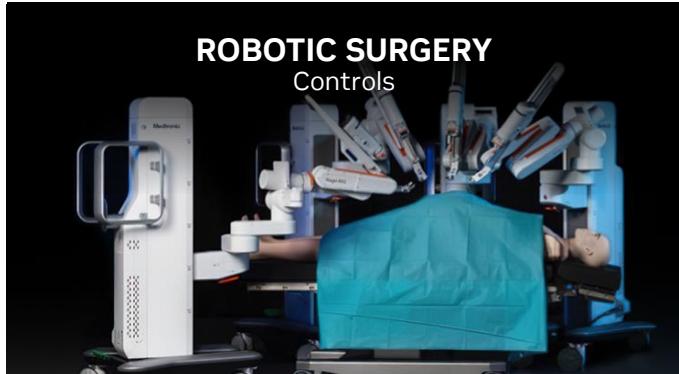
Medical Imaging is the Essential Instrument in Healthcare

2,000,000 Devices | 16,000 Companies



Every Medical Device will become Robotic

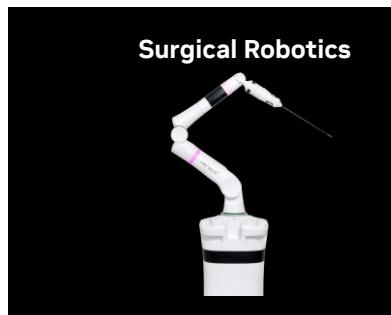
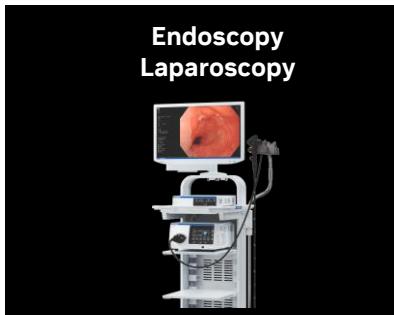
Performing AI Actions or Skills in Real-time



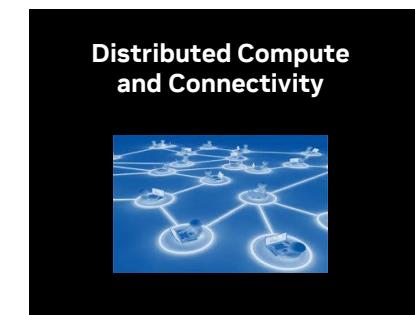
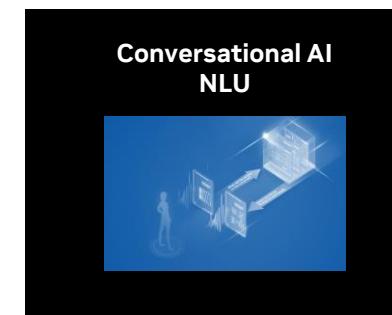
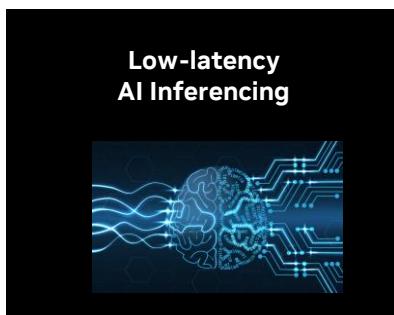
Target Medical Device Applications

Developing next gen medical devices that are edge-based and AI-enabled

Modalities



Applications



Medical Device Development Challenges and NVIDIA Solutions

Challenges	NVIDIA Solutions
Need to Deploy Low-latency Streaming AI Inference	NVIDIA's Accelerated HW and SW stack
Limited Long-life, Upgradable Hardware & Software Options	10 Year HW and SW Long Term Support (LTS)
Non-value Added Dev, Test & Commercialization for HW and SW	Pre-Certification Ready: 60601 & 624304
Difficult to Adopt New Technology	NVIDIA SDK's Algorithm development MONAI

NVIDIA - Accelerating Innovation in Medical Devices

NVIDIA is committed to accelerating AI-powered innovation for Medical Devices by enabling each of the operating businesses to easily and efficiently adopt artificial intelligence from device to edge to cloud by providing a *healthcare specific accelerated computing platform*.

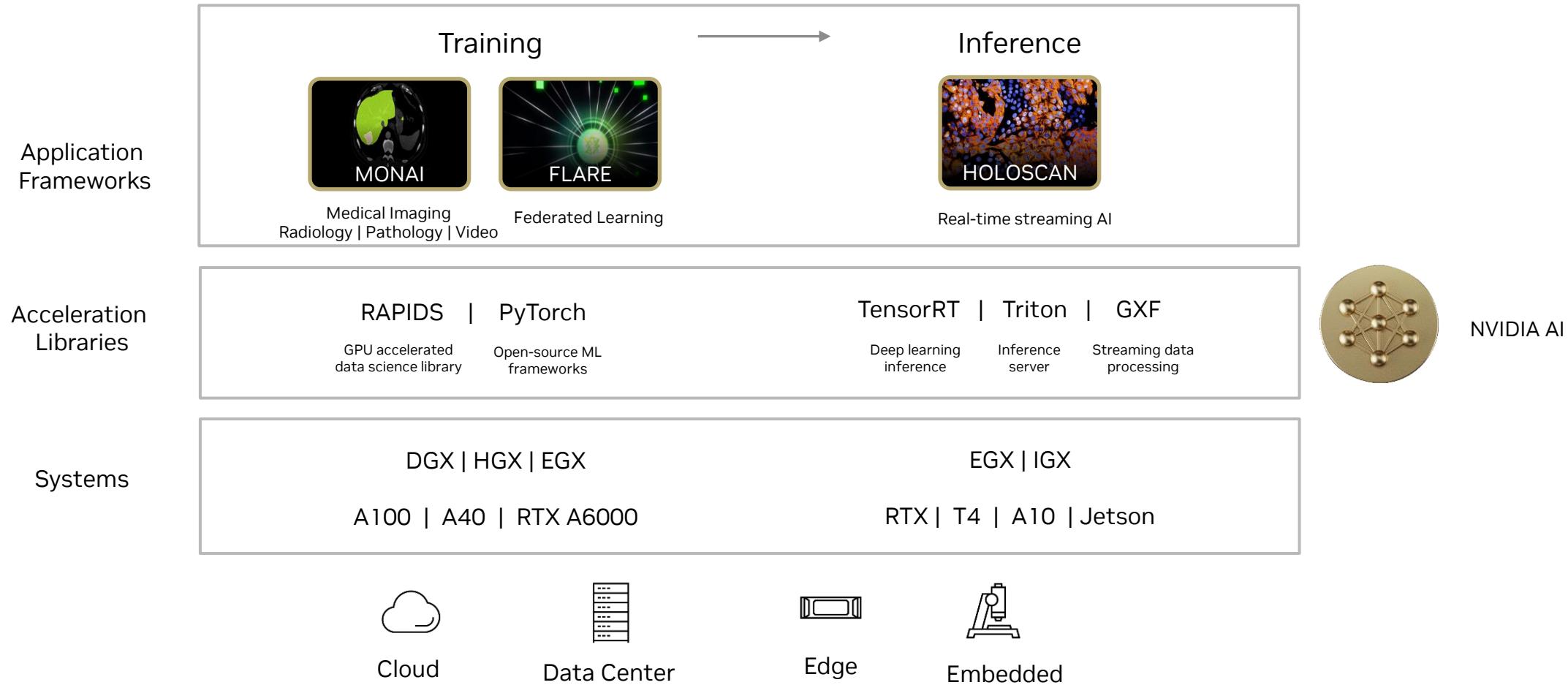
Exploring New
Product
Capabilities

Accelerate
Product to
Market

Optimize
Engineering
Resources

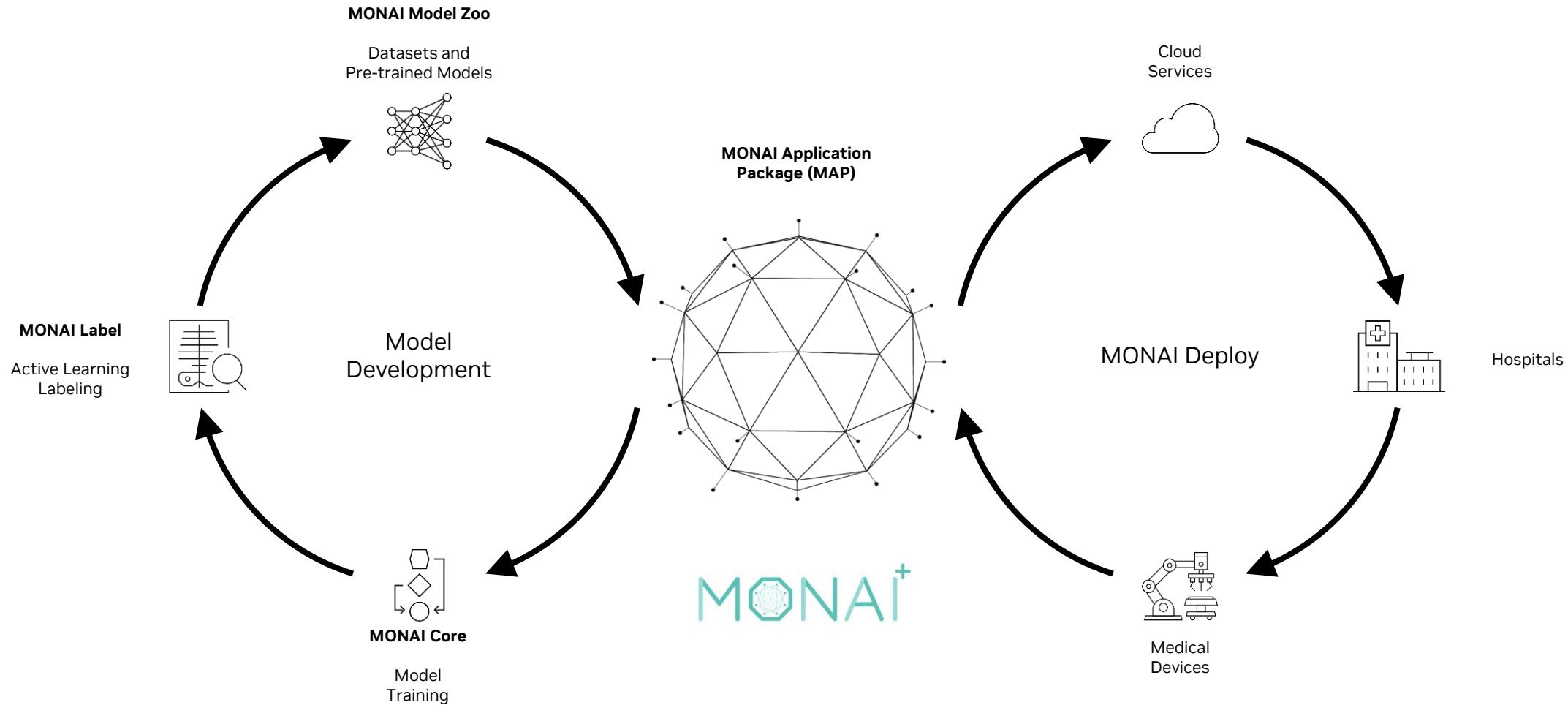
NVIDIA Solutions for Building Real-Time Streaming Medical Devices

Domain Specific and Horizontal Models, Frameworks, SDK's, and Applications



MONAI

A comprehensive Medical AI framework built by experts, accelerated by NVIDIA.

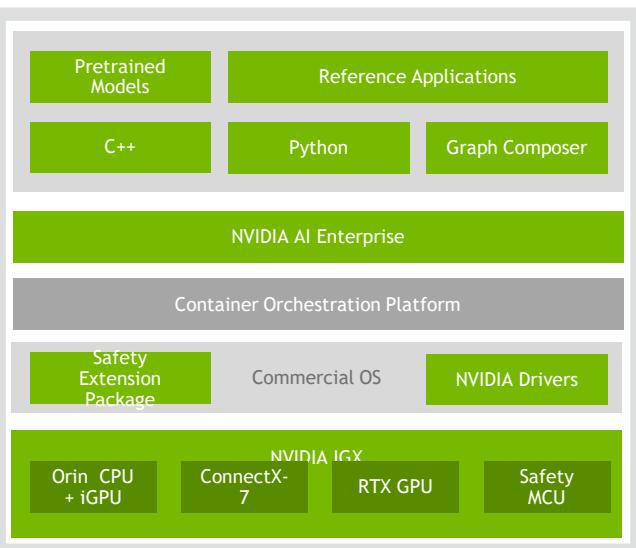


NVIDIA Clara Holoscan Platform

An AI Computing Platform for Medical Devices

Clara Holoscan SDK

Develop



Developer Kits

Validate



NVIDIA IGX DevKit (EA)
Orin, RTX A6000, ConnectX-7



Clara AGX DevKit (GA)
AGX Xavier, RTX 6000, ConnectX-6

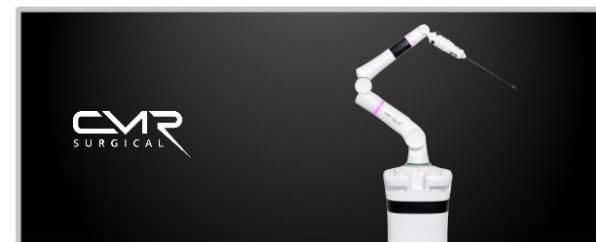
NVIDIA IGX

IEC 60601 Ready



Medical Devices Developing on Clara Holoscan

Over 70 leading medical device companies, startups, and medical centers



Digital Surgery Overview

Market trends and challenges

	Diagnostic Imaging	Endoscopy	OR / ICU Integration	Surgical Robotics	Interventional Radiology	Digital Microscopy
Application Areas	Visualization, Streaming, AI Pipeline					

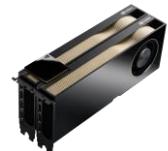
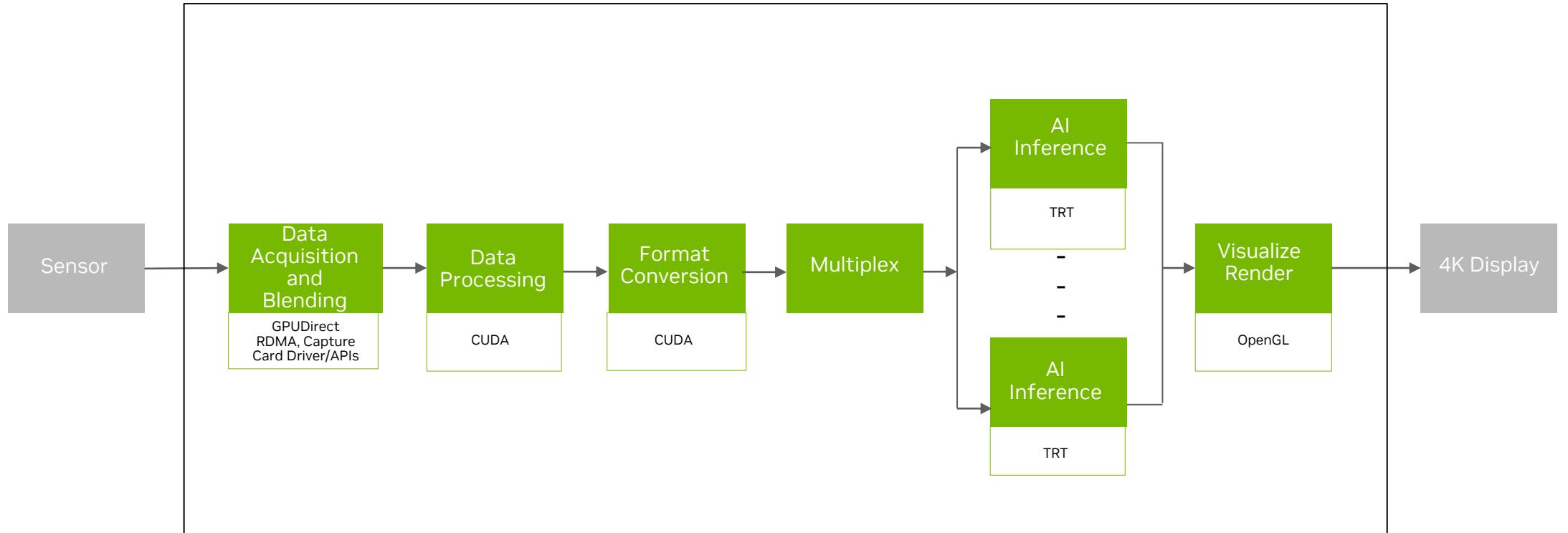
Trends

- 4K multiple video streams
- AI assisted surgery to provide real time insights during surgery
- Pre surgical work and surgeon training leveraging Digital Twins
- AI enabled surgical analytics

Challenges

- Running multiple AI enabled 4K video streams simultaneously in real time
- Designing a safe and secure system
- Meeting regulatory requirements (IEC 60601, 62304)

Digital Surgery Workflow & NVIDIA solutions

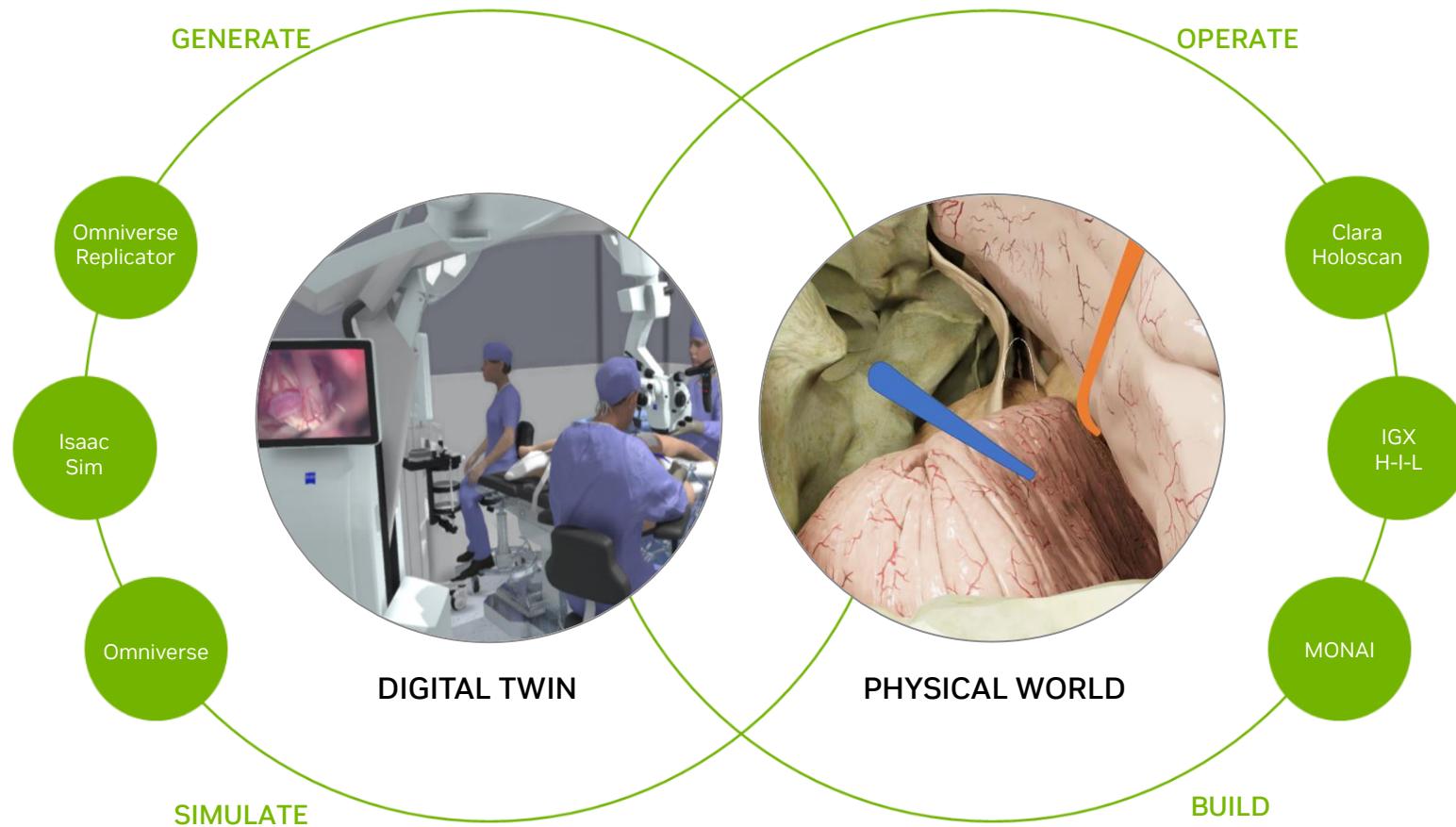


[RTX A6000/5000/4000](#)



[NVIDIA IGX Orin](#)

Digital Twins for Surgical Robotics





<https://www.youtube.com/watch?v=EV3UonTD8FE>

RESOURCES

-  **NVIDIA Clara Holoscan Platform**
nvidia.com/en-us/clara/medical-devices/
-  **NVIDIA Clara Holoscan SDK**
developer.nvidia.com/clara-holoscan-sdk
-  **NVIDIA Clara Holoscan on GitHub**
github.com/NVIDIA/clara-Holoscan
-  **NVIDIA Clara DevKit**
<https://developer.nvidia.com/clara-devkit-distributors>
-  **MONAI YouTube Channel**
[@ProjectMONAI](https://youtube.com/c/ProjectMONAI)
-  **Follow us on Twitter**
[@NVIDIAHealth](https://twitter.com/NVIDIAHealth)

NVIDIA for BioPharmaceuticals



NVIDIA Clara for Healthcare and Life Sciences

World's Largest Data Industry | 36% CAGR by 2025



IMAGING



MEDICAL
DEVICES



DRUG
DISCOVERY



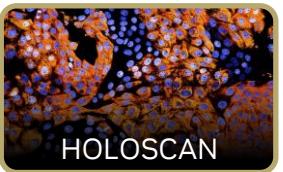
GENOMICS



NVIDIA CLARA
APPLICATION
FRAMEWORKS



FLARE



HOLOSCAN



BIONEMO



PARABRICKS



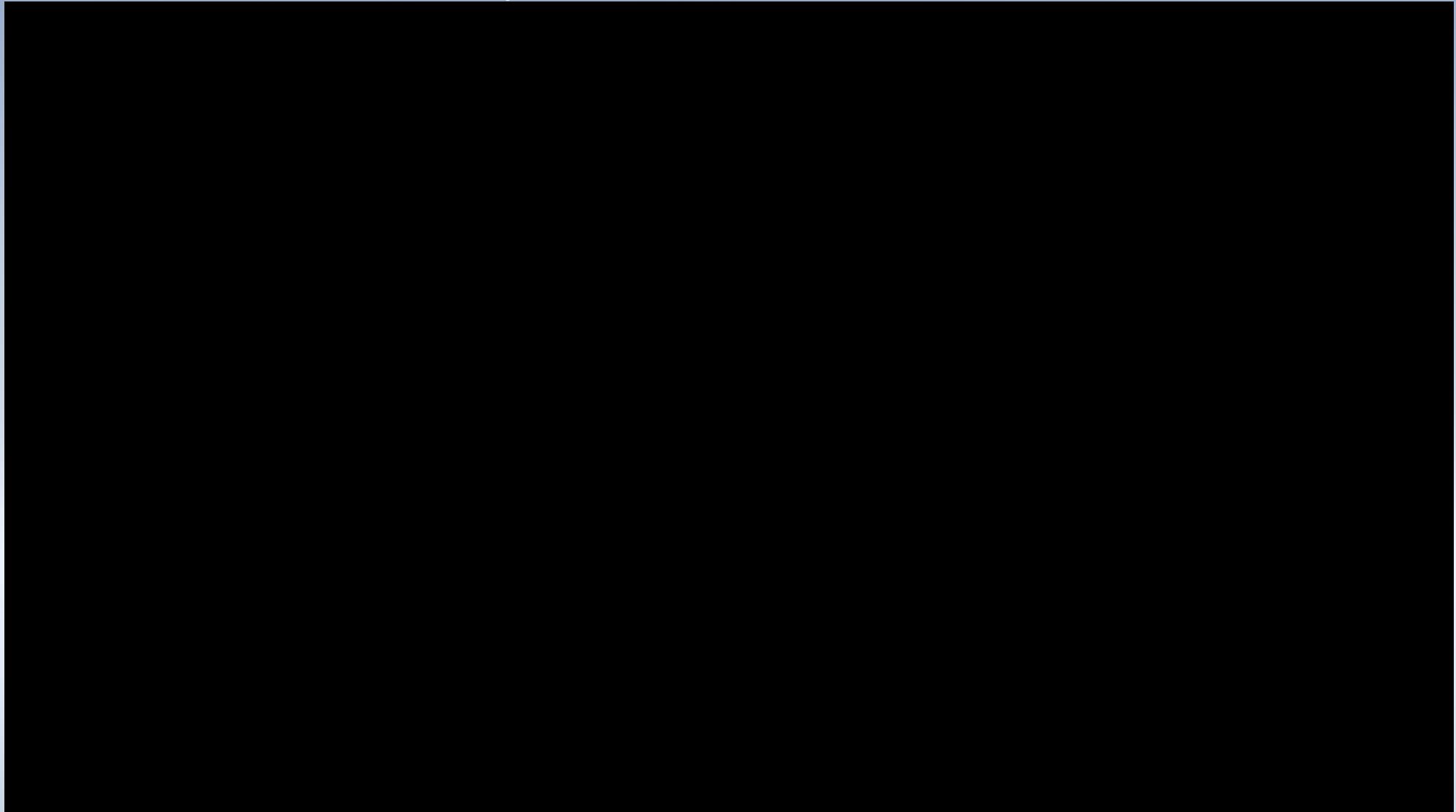
NVIDIA AI



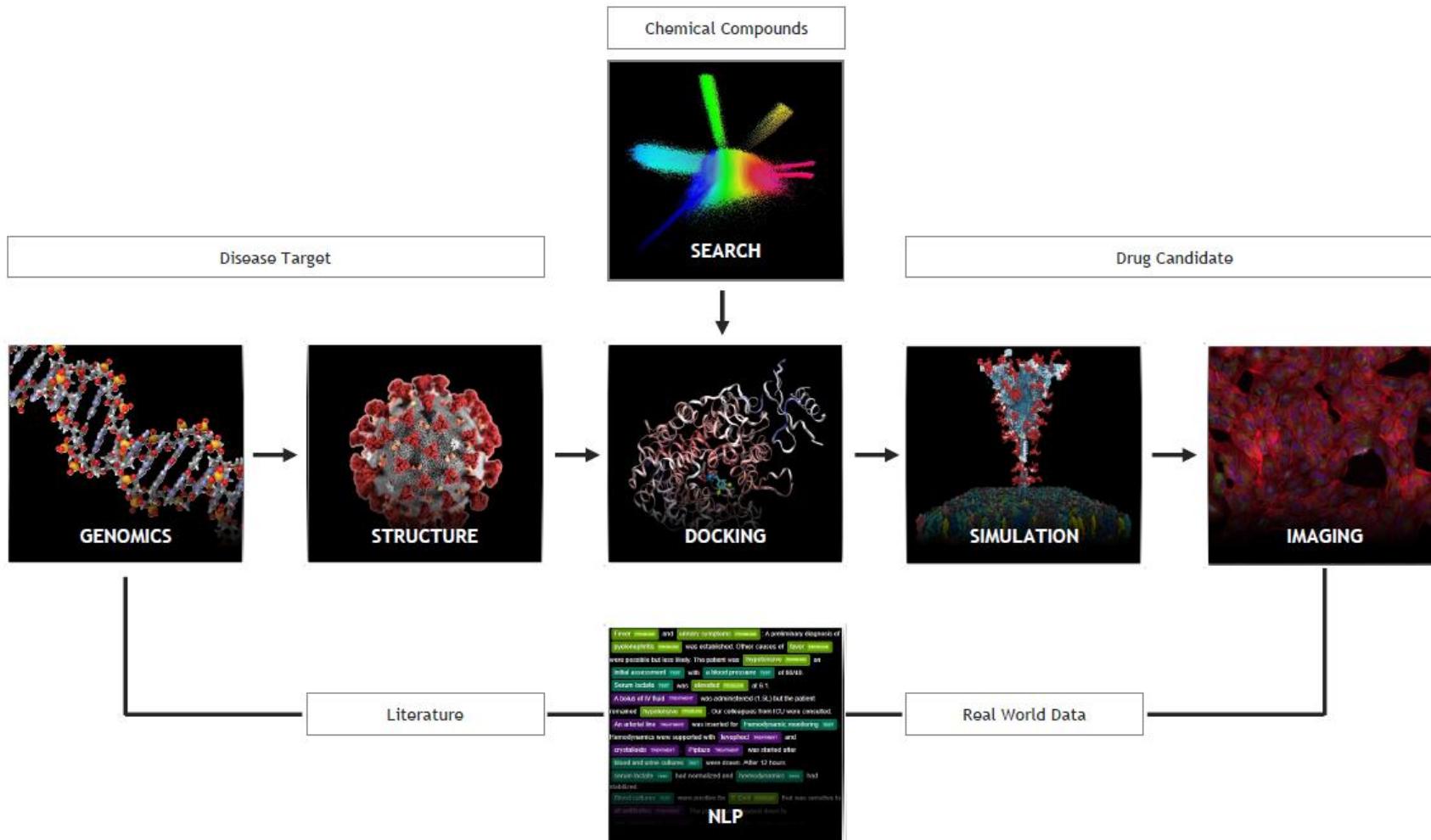
NVIDIA Omniverse

COVID-19 AUTODOCK DEMO

1B Compounds in 12 Hours on All of Summit



NVIDIA CLARA DISCOVERY



How AI is Transforming the Drug Discovery Process

Deep Learning is an Essential Tool for Modern R&D

EDITORIAL
machine intelligence

REVIEW ARTICLE
<https://doi.org/10.1038/s42256-022-00463-x>

The transformational role of GPU computing and deep learning in drug discovery

Mohit Pandey^{1,2*}, Michael Remondao^{3,4}, Francisco Gentile², Olegandr Isayev^{2,5}, Alexander Tropsha⁶, Abraham G. Stern^{1,2} and Arkin Cherkasov^{1,2}

Deep learning has transformed nearly every field of science, including those of direct importance to drug discovery, such as molecular chemistry and pharmacology. This revolution has largely been attributed to the unprecedented advances in highly parallelizable graph-based processing units (GPUs) and the development of efficient neural algorithms. In this review, we present an overview of recent developments in the application of deep learning to drug discovery, including its impact on the discovery of new drugs and drug targets. We also cover the state-of-the-art of deep learning architectures that have found practical applications in both early drug-discovery and cross-disciplinary applications, including the accelerations of molecular docking, the prediction of off-target effects and the prediction of pharmacological properties. We conclude by discussing the impact of AI on acceleration and deep learning models on the global drug discovery in the field of drug discovery that may lead to the next generation of chemical tools to accelerate the discovery of novel medicines.

Abstract Deep learning is a modern deep-dimensional graph on the basis of GPUs for parallel parallel computing very quickly gained by the scientific community. The first attempt to use GPU for scientific purpose employed the use of the C programming language and was made by NVidia, which Computer Unified Device Architecture (CUDA) as an extension of the C programming language, together with compilers and debuggers, is used by the developer for writing programs that can run on GPU. The main advantage of CUDA is that it is a standard API for parallel computation, which makes it easier to learn and use. In this paper, we will focus on the basic concepts of GPU computing and how it can be used for drug discovery. We will also discuss the challenges of GPU computing and how it can be used for drug discovery. Finally, we will conclude by discussing the future of GPU computing in drug discovery.

Keywords: GPU computing, Deep learning, Drug discovery, Machine learning, Deep learning, AI, Machine learning, Deep learning, GPU computing, AI for molecular simulation.

GPU computing and DL for molecular simulation

GPU acceleration comes from massive data parallelism, which is achieved by dividing the data into smaller pieces and then applying a similar operation on each piece. The most common support for training and testing processes enabled by GPU is related to the use of the Cuda API for parallel computation. In molecular simulation, the main challenge is the parallelization of the simulation of proteins and protein-ligand complexes. GPU acceleration of DL algorithms has already been demonstrated in cryo-electron microscopy (cryo-EM) and 3D structure prediction of proteins.

GPU computing and DL for molecular simulation

GPU acceleration comes from massive data parallelism, which is achieved by dividing the data into smaller pieces and then applying a similar operation on each piece. An example of a common data parallel operation is the use of a rotation matrix to rotate molecules, defining the position of objects as it is related. In molecular simulation, the main challenge is the parallelization of the calculation of atomic potential energies. Similarly, DL model training has been used and has been proven that it is currently superior to existing implementations that are usually parallelized in memory. While NVidia's CUDA is a more established GPU programming language, AMD's ROCm represents a similar model for GPU acceleration. ROCm provides a set of new frameworks to support common open-source machine learning libraries such as TensorFlow and PyTorch. It also provides libraries for parallel NVIDIA CUDA and AMD ROCm. In addition, there are many other GPU providers and software partners in the GPU computing ecosystem, but also recently introduced the new flagships GPU software are AMD's Instinct MI200 Series to compete with NVIDIA's Ampere A10 GPU and AMD's Instinct MI200 Series to compete with NVIDIA's Ampere A10 GPU. These changes promise a lot of opportunities in the future, including compute-aided drug discovery.

*Correspondence: Pandey, M., Department of Biological Sciences, University of Illinois Urbana-Champaign, Illinois, United States. E-mail: mohit.pandey@illinois.edu
†Department of Chemistry, Carnegie Mellon University, PA, USA.
‡School of Pharmacy, University of North Carolina at Chapel Hill, NC, USA.
§Eli Lilly and Company, Seattle, WA, USA.
¶These authors contributed equally to this work. M. Pandey, M. Remondao, A. Stern, and A. Cherkasov (✉). E-mail: arkin.cherkasov@illinois.edu

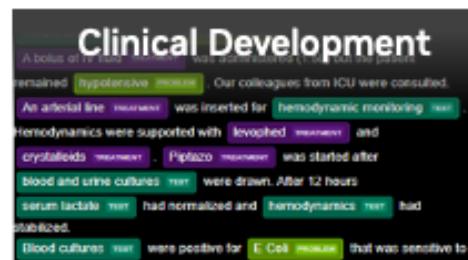
Nature Reviews Cancer (2023) 21:20–30. https://doi.org/10.1038/s42256-022-00463-x



Gene Expression Prediction, scRNA analysis
Accelerated cryo-em & protein structure prediction
Knowledge synthesis from scientific literature



Active-learning virtual screening
AI powered molecular property prediction & generation
Drug-target interaction prediction



NLP for Clinical trial matching
Adverse event monitoring
Histopathology/Radiology/OMICS biomarker ID

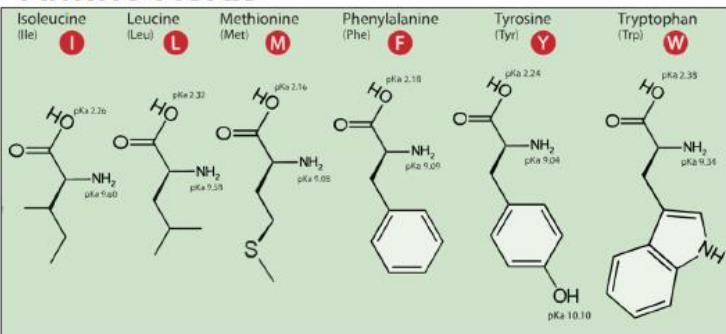
PROTEIN STRUCTURE PREDICTION

3D Structure of Protein Target Central to Structure-Based Drug Discovery

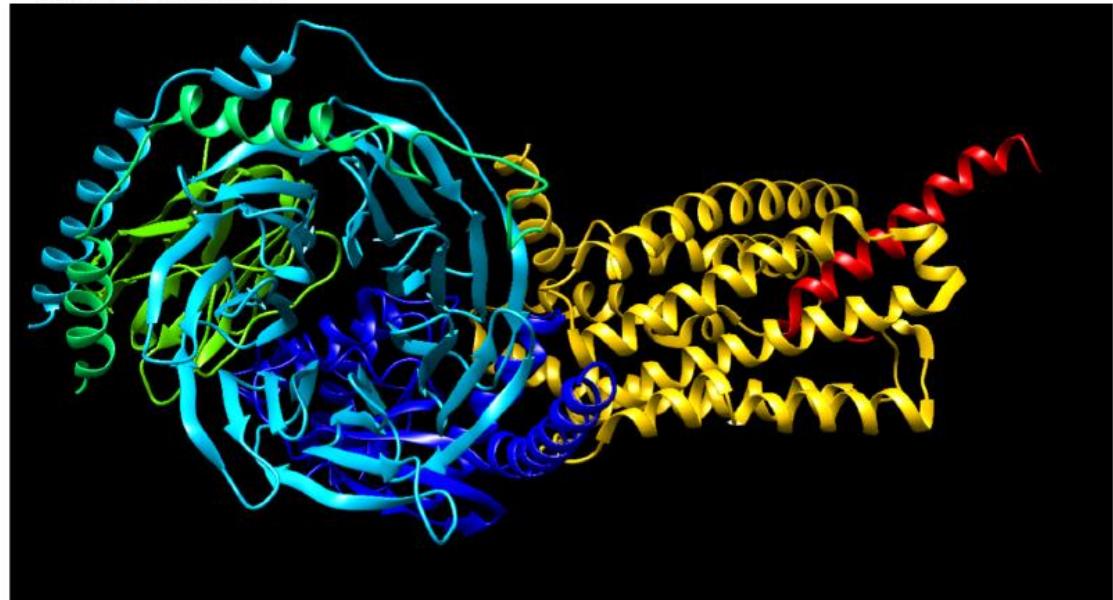
Sequence

MGCLGNSKTEDQRNEEKAQREANKKIEKQLQKDQ
VYRATHRLLLGGAGESGKSTIVKQMRLILHVNGFNG
EGGEEDPQAA...

Amino Acids



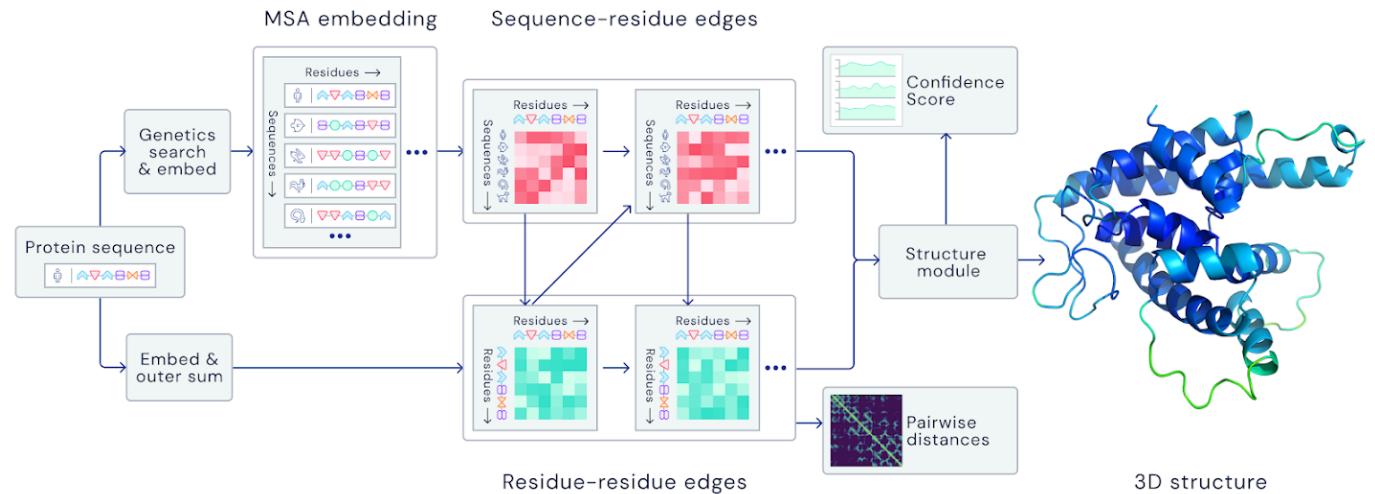
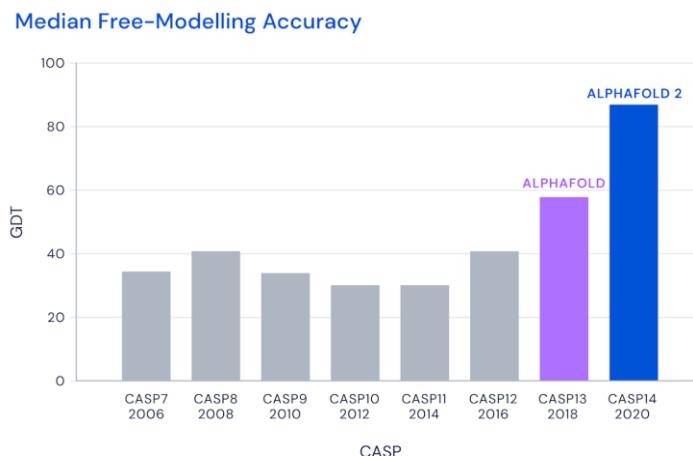
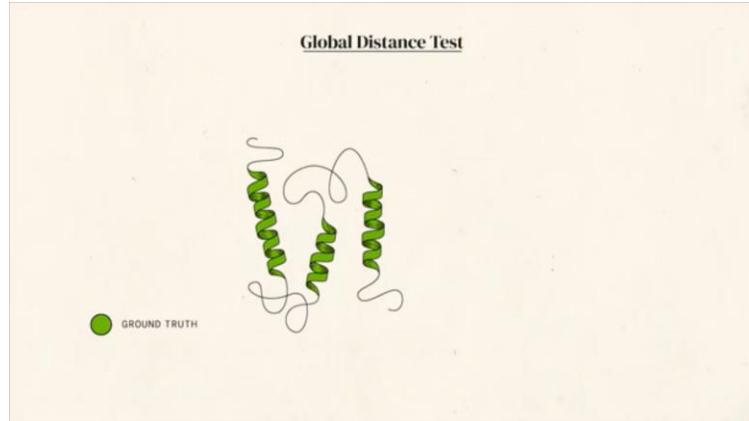
3D Structure



[6P9X] G-Protein Coupled Receptor

ALPHAFOLD 2 FROM DEEPMIND

CASP (Critical Assessment of protein Structure Prediction)



<https://github.com/lucidrains/alphafold2>

<https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

<https://www.youtube.com/watch?v=gg7WjuFs8F4>

NVIDIA BioNeMo Service

Cloud API for Biology and Chemistry | EA has OpenFold and ESM-1

BioNeMo Service > Playground

Playground

Report Bug Documentation

Baseline Models

OpenFold
 ESM-1b

View Code

Clear All Generate

Input

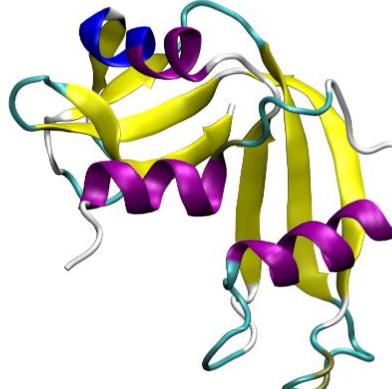
```
MNIFEMLRIDEGLRLKIYKDTEGYYTIGYLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAKLKPVYDSDLAVRRAALINMFQMGETGVAGFTNSLRMLQQKRWDEAVNLAKSRWYNQTPNRAKRVITTFRTGTWDAYKNL
```

Example Use Cases

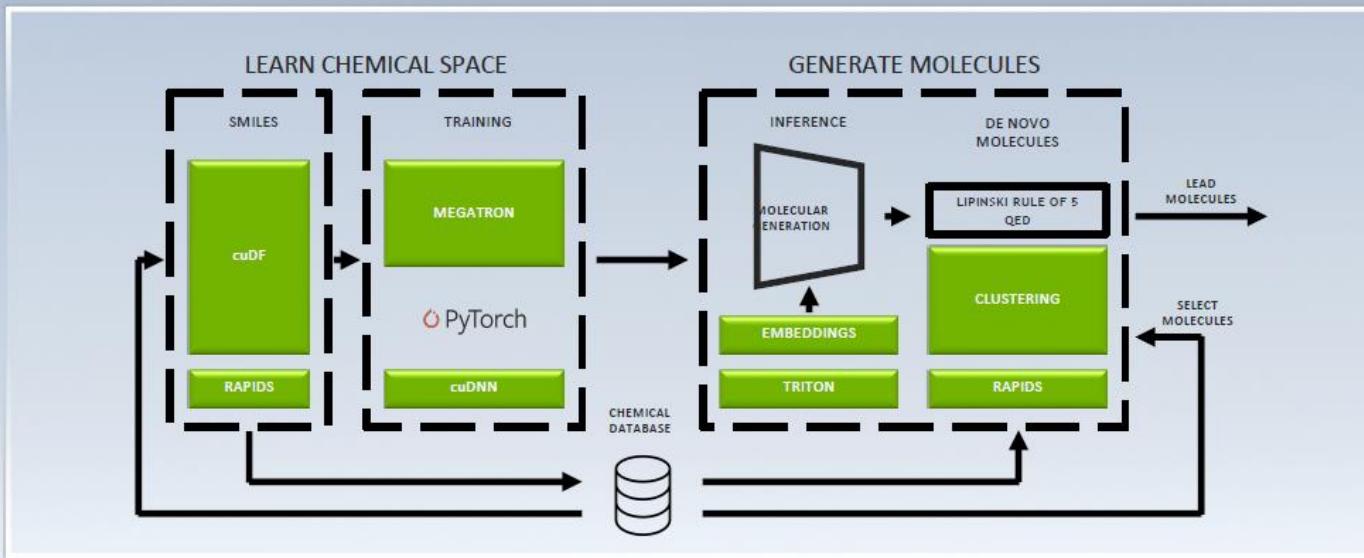
PDB ID: 102L
PDB ID: 91CJ
PDB ID: 6S2M
PDB ID: 2WFI
PDB ID: 3BCJ

Output (Preview)

Download PDB



ANNOUNCING MEGAMOLBART GENERATIVE MODEL FINDING MOLECULES BEYOND CHEMICAL DATABASES



Pre-Training with 1B Molecules in Zinc Database

Megatron Multi-GPU, Multi-Node Training Across 32 DGX A100

Triton Inference 250M Molecules/Day on Cambridge-1

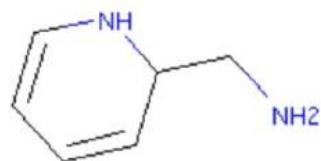
Property Prediction & Molecule Generation

Cambridge-1
DGX SuperPOD



MEGAMOLBART CHEMICAL TRANSFORMER

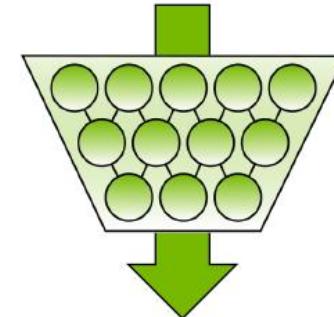
```
from rdkit import Chem  
from rdkit.Chem import Draw  
  
m = Chem.MolFromSmiles("C1=CC=CNC1CN")  
Draw.MolToImage(m)
```



C 1 = C C = C N C 1 C N
C [1 0 0 1 1 0 1 0 1 0 1 0
1 0 1 0 0 0 0 0 0 0 1 0 0
= 0 0 1 0 0 1 0 0 0 0 0 0 0
N 0 0 0 0 0 0 1 0 0 0 0 1

BERT Encoder

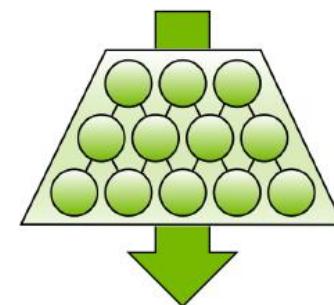
C1=CC=CNC1CN



Latent Representation

1 | 2 | 0 | 1 | 1 | 2 | 1 | 0 | 1 | 0 | 1 | 2

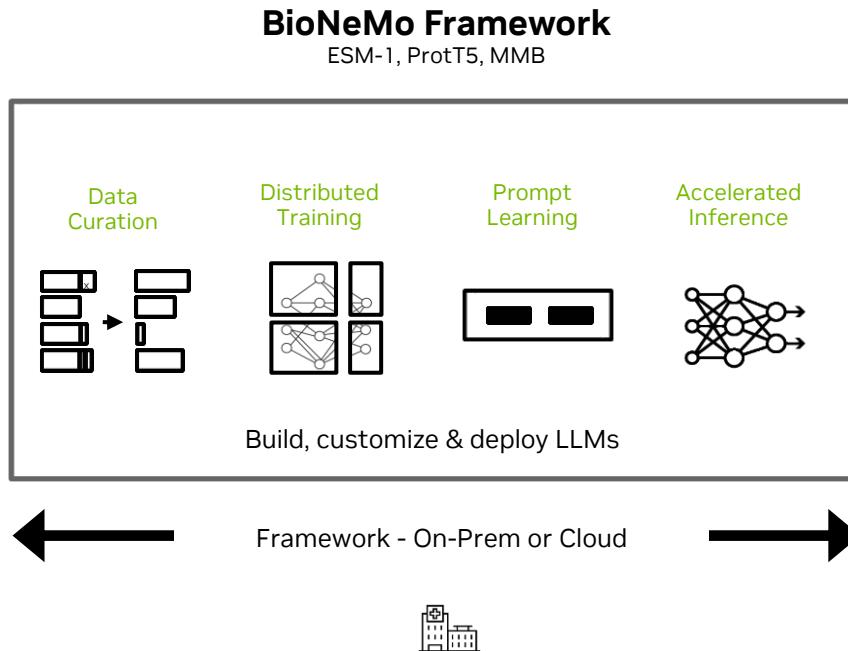
GPT Decoder



C1=CC=CNC1CN

NVIDIA BioNeMo Framework

AI Framework for BioMolecular Large Language Models



Domain Specific

Drug discovery researchers need an LLM framework that speaks the language of biology and chemistry.

Optimized For Scale

LLMs are massive requiring model parallelism and supercomputing resources to train.

Pre-Trained Models

Training LLMs takes enormous compute resources and is very time consuming. Pre-trained models are ready for production.

Cloud Native

Use pre-trained models, create powerful customized workflows and deploy in the cloud.

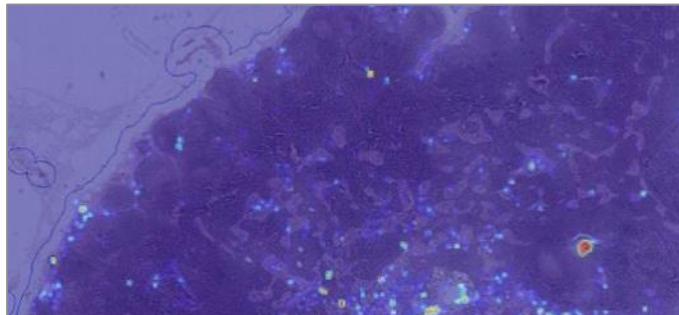
Leaders in Drug Discovery & Development use NVIDIA

DGX SuperPod is the Modern Instrument of Discovery & Innovation



Recursion Pharmaceuticals

AI powered phenotypic screening at scale
DGX SuperPod, BaseCommand



Paige.ai

Transforming Cancer Diagnostics & Biomarker Identification
DGX SuperPod, cuCIM



Janssen

Adverse Event Monitoring, Histopathology, CompChem, +
DGX SuperPod, BaseCommand



G42

Population Scale short and long read sequencing
DGX SuperPod, Parabricks



World's Leading Cancer Centers

Medical Imaging, NLP, and Genomics Research
DGX SuperPod, MONAI, Parabricks

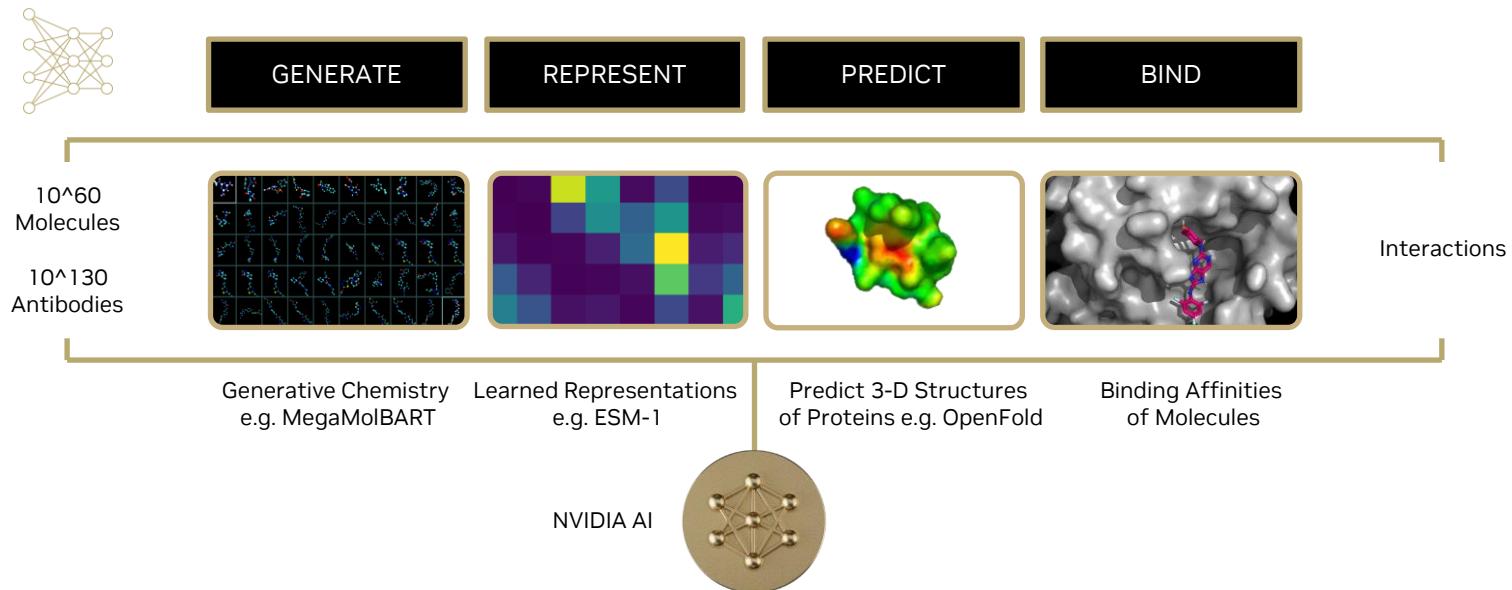


Univ. Florida Health System

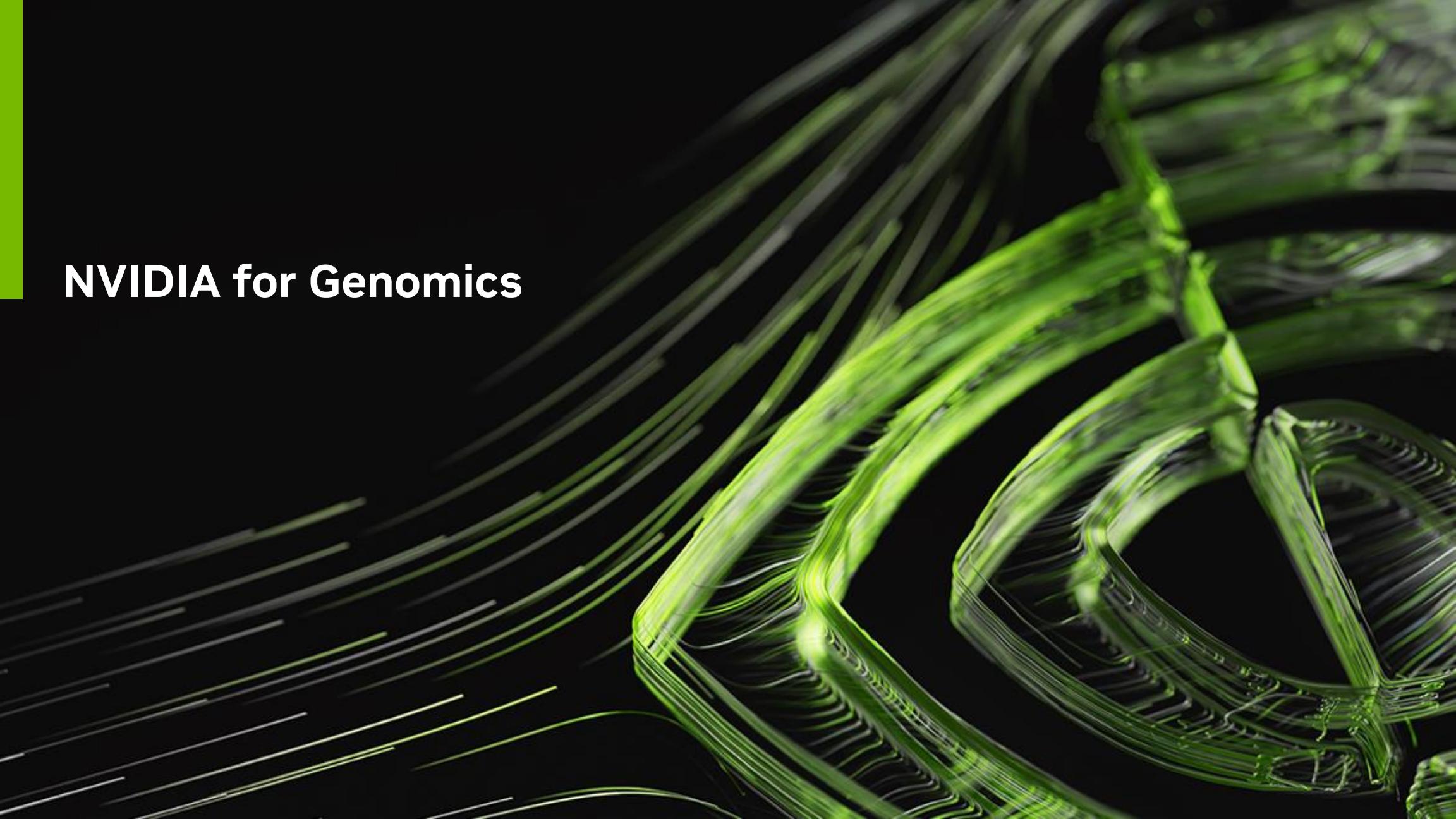
Clinical Language Models, Synthetic Clinical Data Generation
DGX SuperPod

The Future of Drug Discovery is *In silico*

Turning Drug Discovery into an Information Science



NVIDIA for Genomics



NVIDIA Clara for Healthcare and Life Sciences

World's Largest Data Industry | 36% CAGR by 2025



IMAGING



MEDICAL
DEVICES



DRUG
DISCOVERY



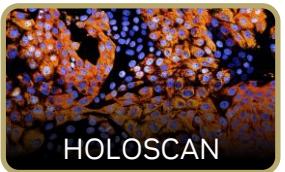
GENOMICS



NVIDIA CLARA
APPLICATION
FRAMEWORKS



FLARE



HOLOSCAN



BIONEMO



PARABRICKS



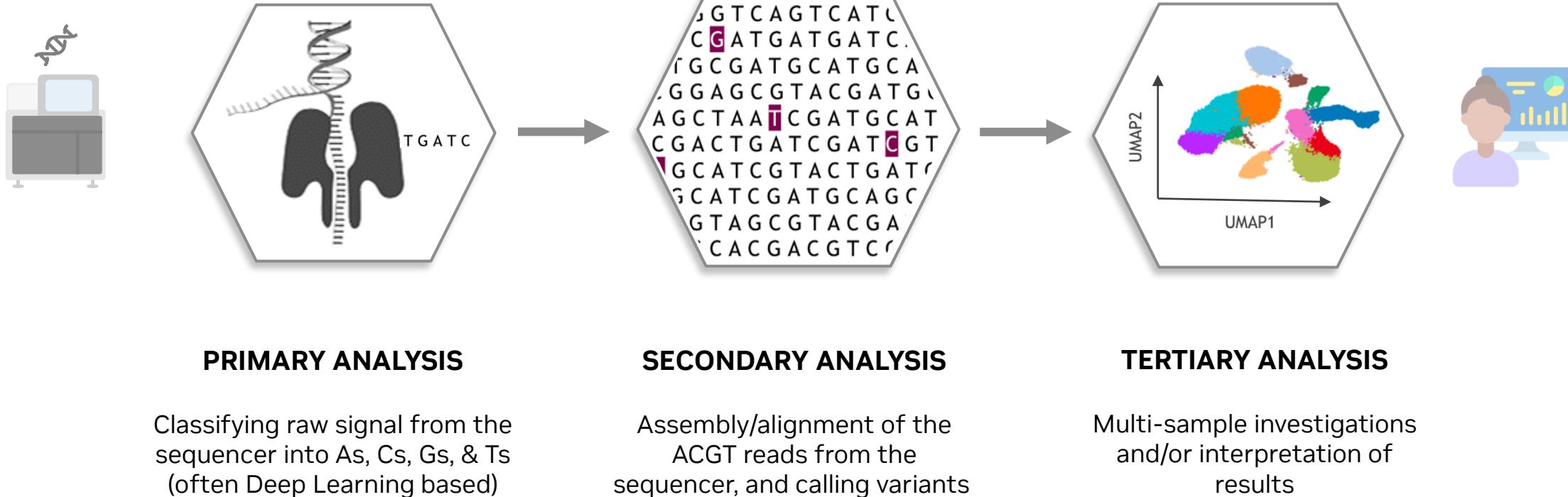
NVIDIA AI



NVIDIA Omniverse

Computation in the Genomics Workflow

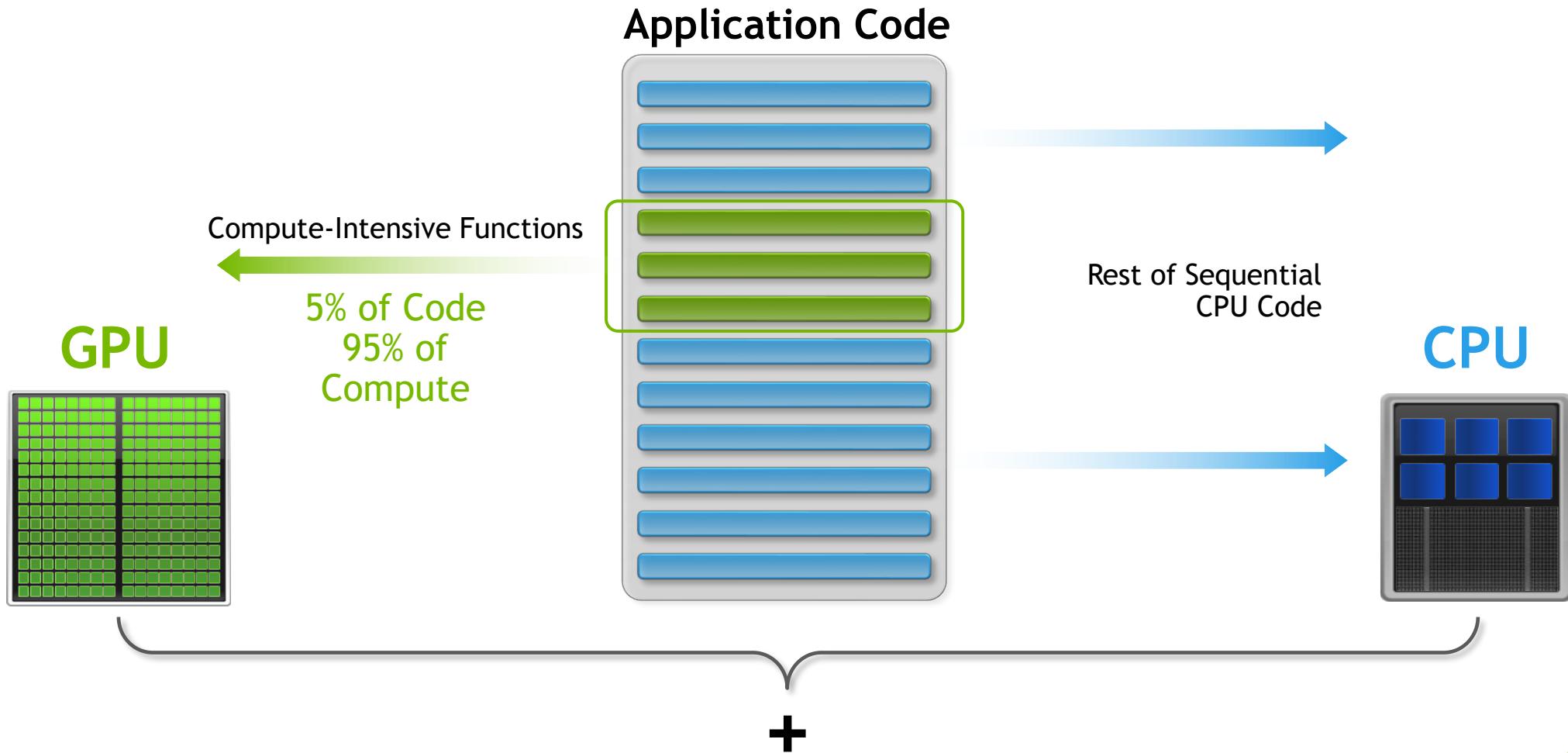
Background



NVIDIA ADDRESSES CHALLENGES ACROSS THE FULL GENOMICS ANALYSIS WORKFLOW

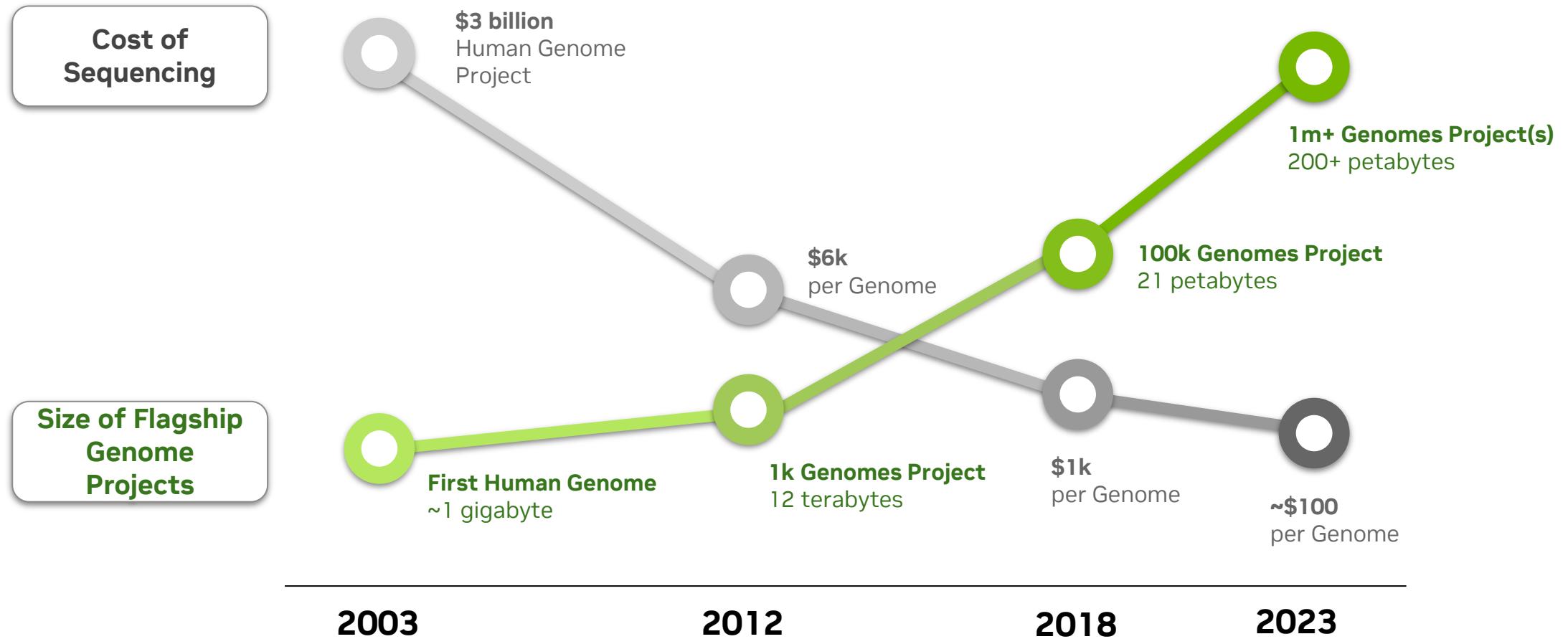
Small Changes, Big Speed-up

GPUs are Extremely Well Suited for Highly Parallel, Compute Intensive Tasks



The Big Data Problem: Volume & Velocity

As sequencing gets cheaper, the data deluge grows

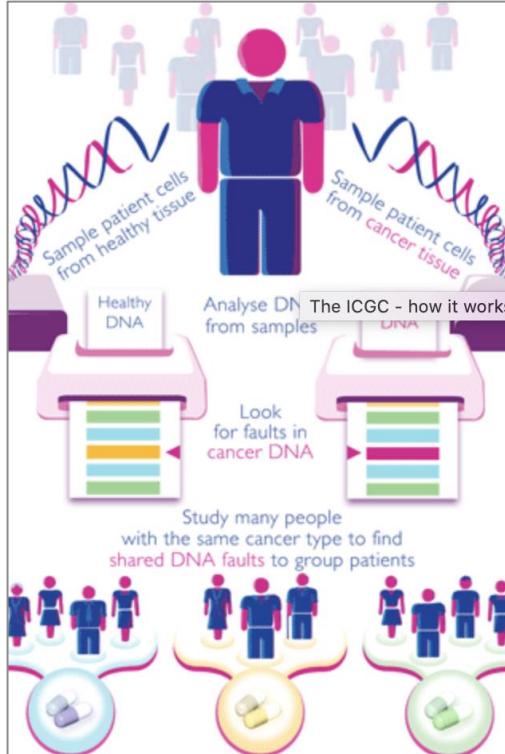


Why Do We Need Accelerated Genomic Analysis?

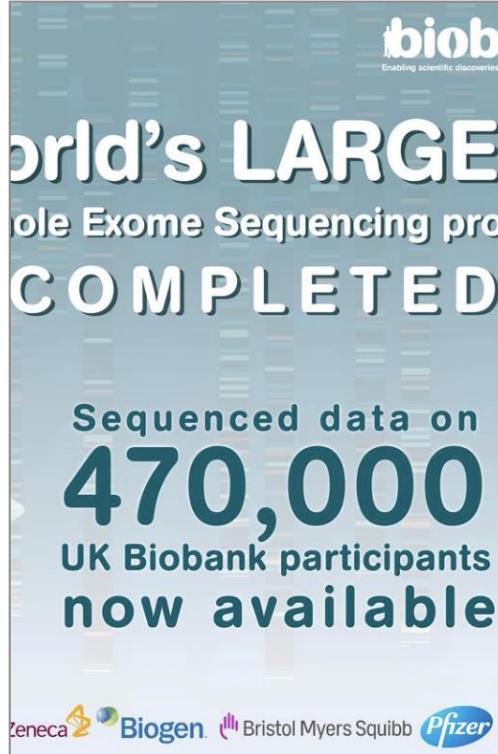
Precision Medicine Evolution; What is important about gleaning insights quickly?



Newborns, Critical Care



Cancer



Large Population Studies

A screenshot of a preprint article from bioRxiv. The title is "Novo variant calling identifies cancer mutation profiles in the 1000 Genomes Project" by K. Ng, Pankaj Vats, Elyn Fritz-Waters, Stephanie Sarkar, Eleanor I. Sams, Evin M. Padhi, Zachary L. Leonard, Marc A. West, Chandler Prince, Lee Tran, Marshall Jansen, George Vacek, Mehrzad Samadi, T. Harkins, Craig Pohl, and Tychele N. Turner. The abstract discusses the development of a GPU-based workflow for calling de novo variants (DNVs) in whole genome sequenced parent-child trios. It highlights the identification of 78 ± 15 DNVs per individual, 18% ± 5% at CpG sites, 75% ± 9% phased to the paternal chromosome of origin, and an average allele balance of 0.49. The article also notes the presence of excess DNVs in the 1000G dataset compared to a control dataset, which were hypothesized to be cell line artifacts. The study found that 40% of individuals in 1000G did not have random DNV profiles, with some having profiles matching B-cell lymphoma. The article concludes with significant excess of protein-coding DNVs in the gene *IGLL5* that has already been implicated in cancer.

Research

POWERING RAPID SEQUENCING

NVIDIA GPUs are already onboard many sequencing instruments, accelerating primary and secondary analysis at source

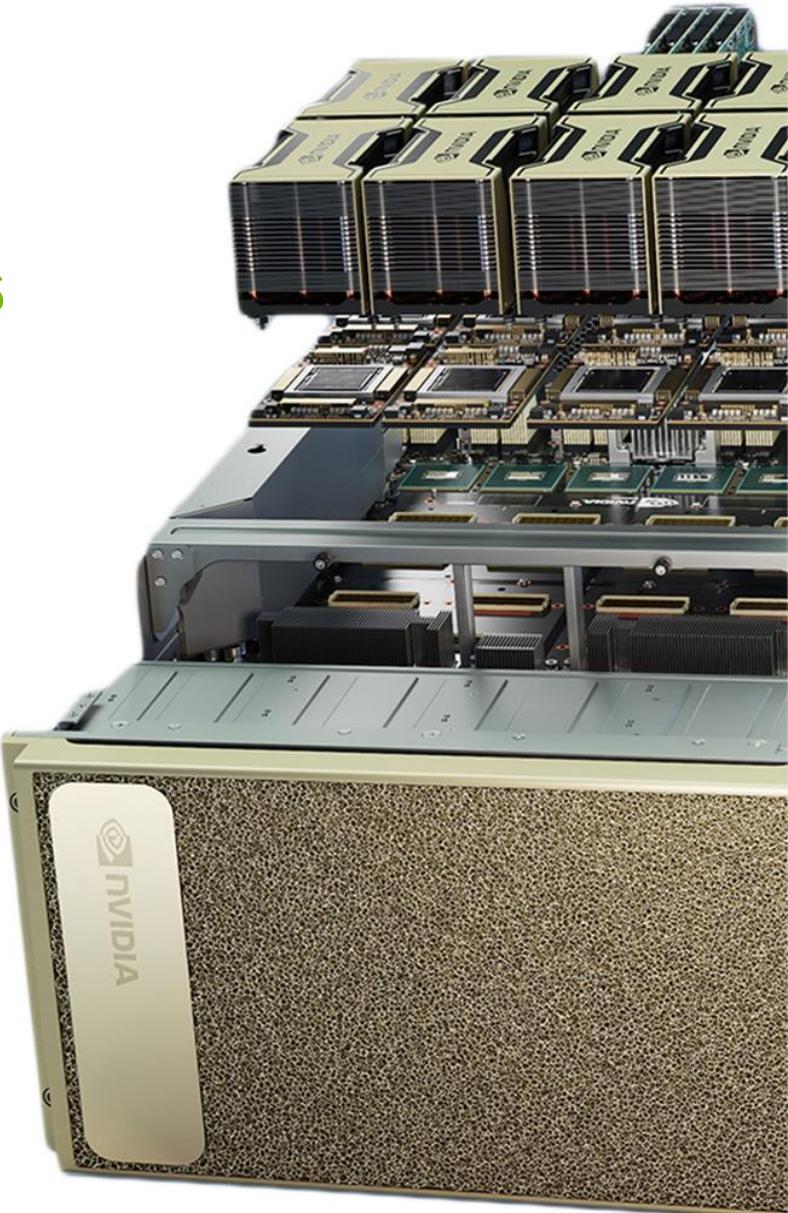
- Including Oxford Nanopore, MGI, Singular, Ultima, and others
- Our DevTechs and SAs work with these providers to accelerate primary analysis, often involving deep learning models and rapid inference
- NVIDIA Clara Parabricks can also bring accelerated secondary analysis directly onto the instrument



ENABLING HIGH-THROUGHPUT ANALYSIS

NVIDIA Clara Parabricks is designed to deliver high speed secondary analysis at data center scale

- Perfect for national genome projects, BioBanks, and large institutes, that need to operate at scale
- Clara Parabricks can perform secondary analysis for up to 60 whole genomes per day (under 25 mins each) on a single DGX server
- Fully compatible with common workflow languages and managers, for flexible and scalable deployment



ACCELERATING DATA SCIENCE

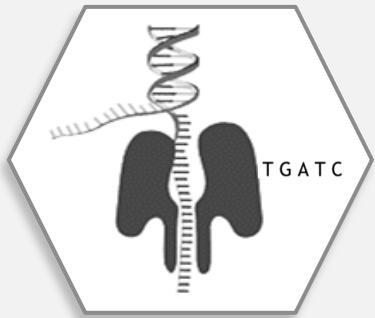
NVIDIA data science libraries already provide the ideal solution for faster tertiary analysis

- Accelerated Python data science with RAPIDS and CuPy means minimal code changes and no new tools to learn
- Array manipulation and computation, regression, clustering, machine learning, visualization, +more
- Open source and deployable from laptops to clusters



NVIDIA Across the Full Genomics Workflow

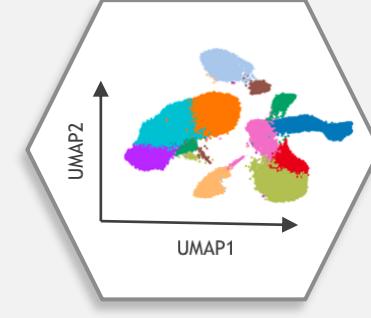
All the tools required for full stack analysis



Primary



Secondary



Tertiary

E2E
Genomic
Analysis
Workflow



TensorFlow

PyTorch



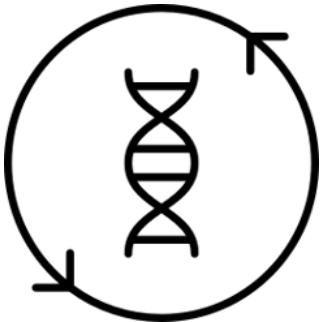
RAPIDS



NVIDIA AI
Enterprise
Supported

Benefits of Clara Parabricks v4.0

Speed, Accuracy, Flexibility



Key Applications

Accelerating tooling for gold-standard germline, somatic and RNA analysis, at speed.

Up to 80x Acceleration

Up to 80x faster for WGS than CPU-only solutions, reducing computing costs by up to 50 percent.

Better Accuracy

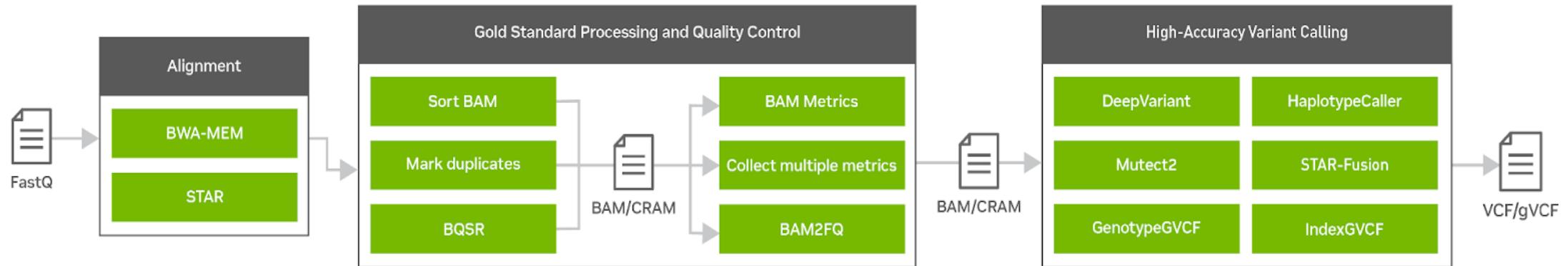
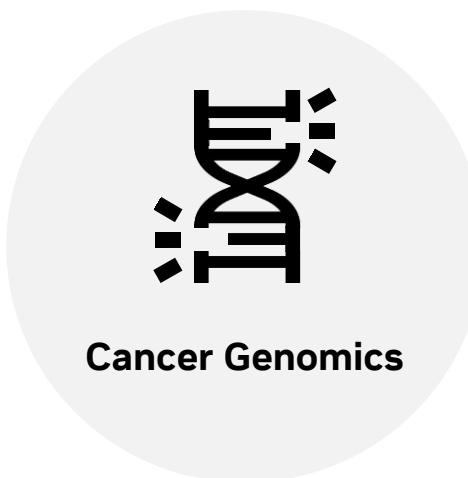
Bring the power of deep learning to your genomic analysis with Clara Parabricks and GPUs.

Flexible Workflows

Create powerful customized workflows by configuring tools in WDL and NextFlow.

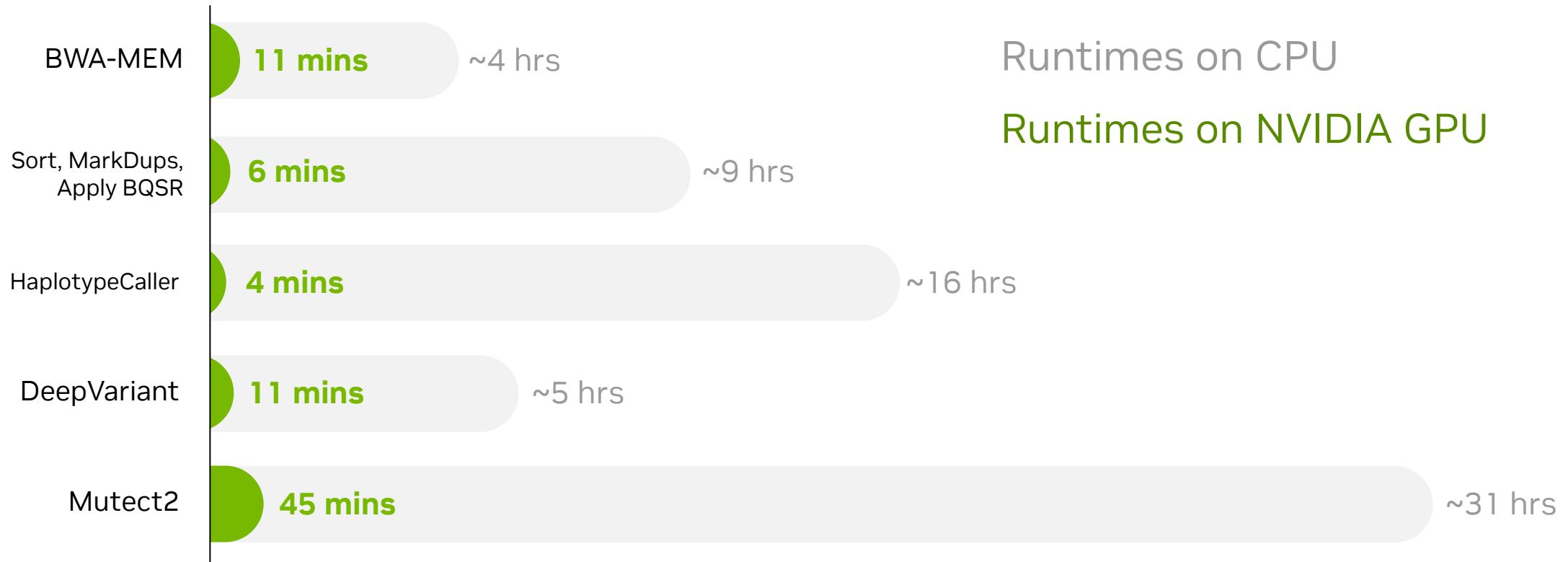
Key Applications of Clara Parabricks

Accelerated and Deep Learning Genomic Analysis



Up to 80x Acceleration

Gold-standard results, faster



v3.8 Benchmarks

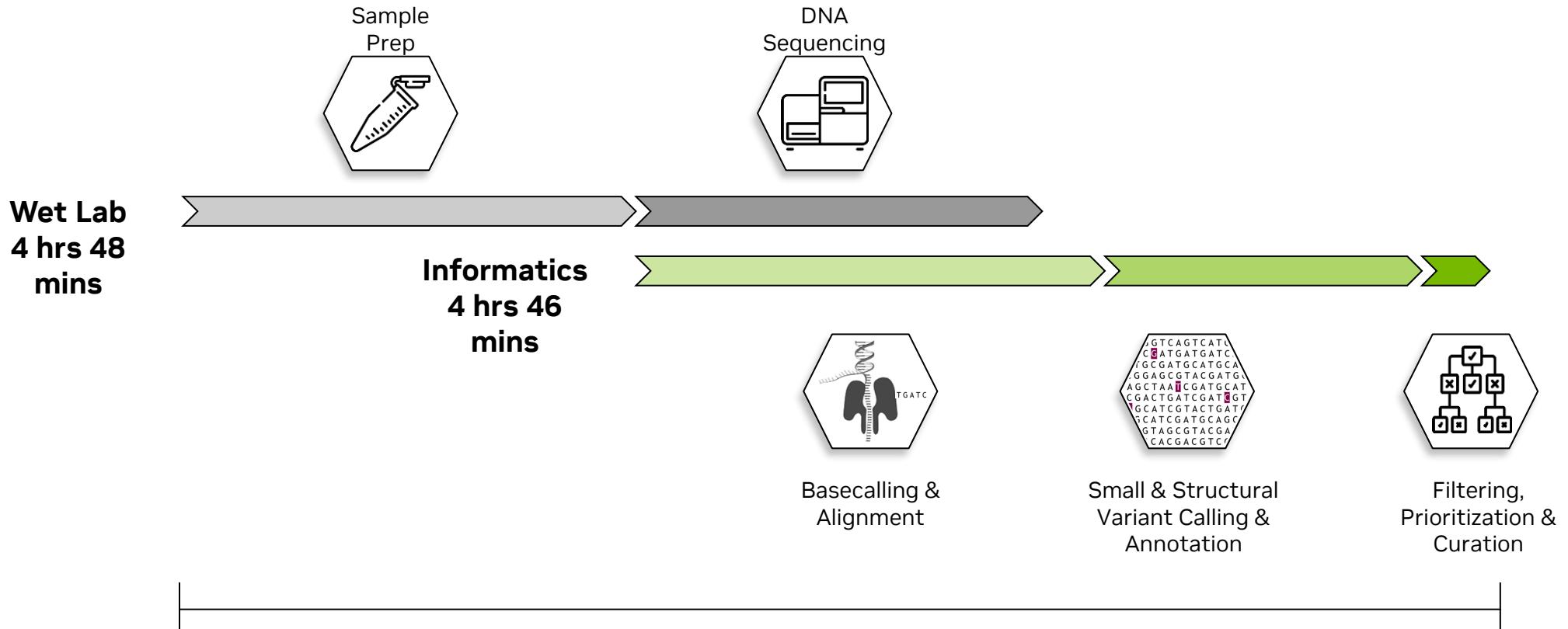
Dataset: HG002 30x WGS, except Mutect2 on SEQC2 50x WGS

CPU: m5.24xlarge; GPU: 8xA100, except DeepVariant & Mutect2 on 8xV100



The Big Data Problem: Volume & Velocity

As sequencing gets faster, informatics becomes the bottleneck

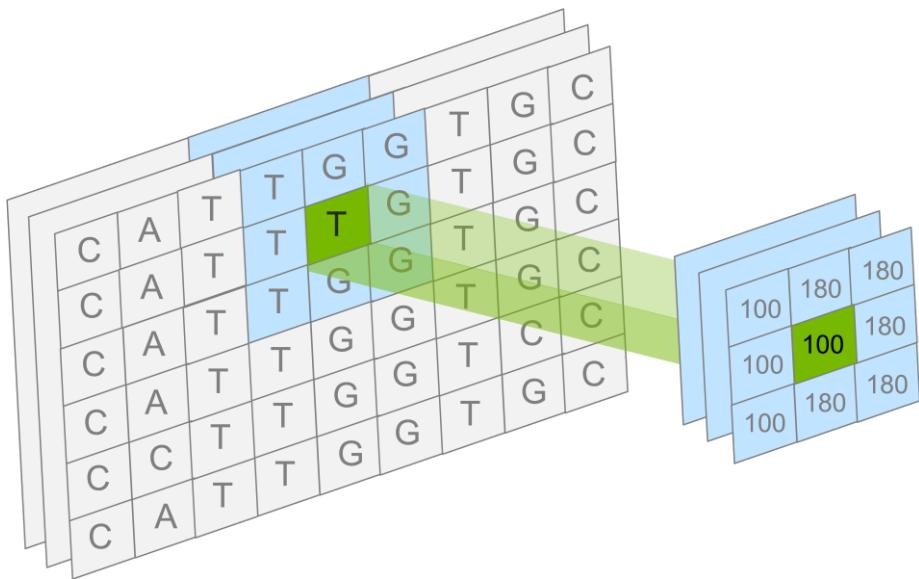


World Record for Fastest DNA Sequencing Technique (2022)
7 hrs 18 mins

Stanford University, Oxford Nanopore, Google & NVIDIA

Improve Accuracy with DeepVariant

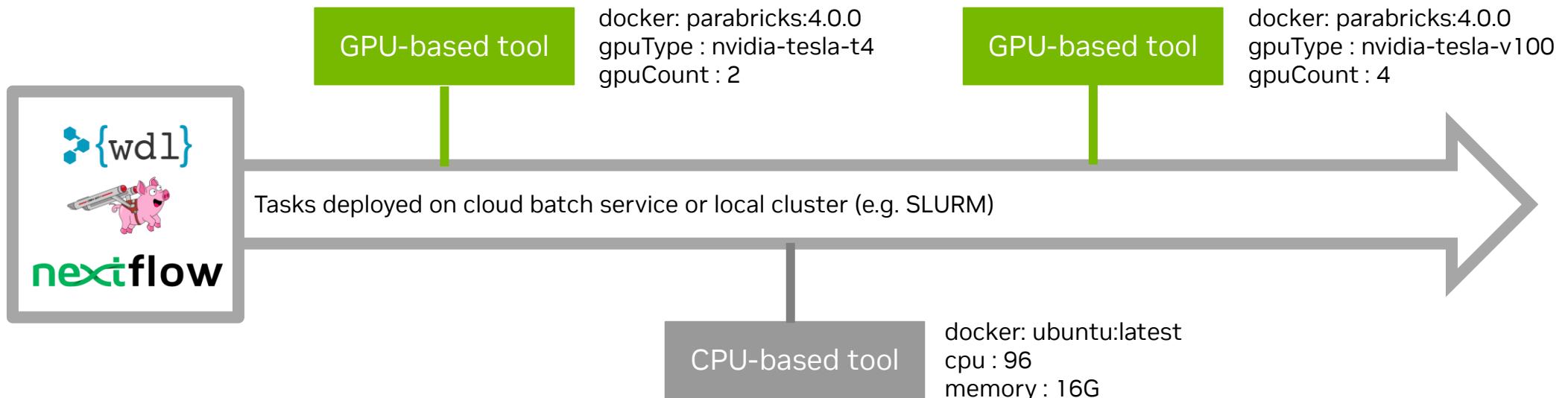
Deep learning for genomic variant calling



- DeepVariant is a deep learning based Variant Caller from Google
- Utilizes CNN's for Image Classification to identify variants
- Greater Accuracy and Precision compared to other industry-standard tools
- Models can be re-trained for improved accuracy

Workflow Manager Compatible

Customize and deploy Parabricks at scale

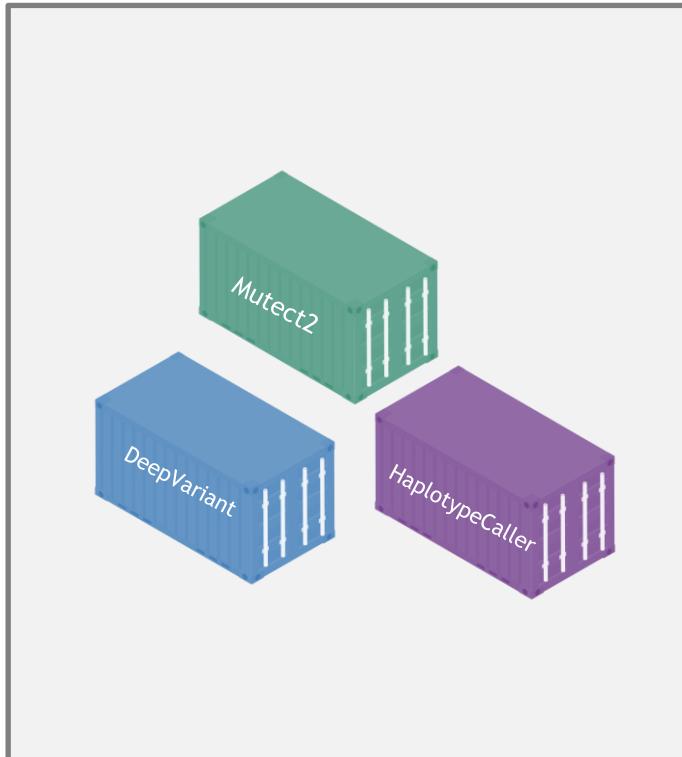


Parabricks is fully compatible with common workflow managers WDL and NextFlow for deploying at scale

- Intertwine GPU and CPU powered tasks with different compute requirements
- Reference workflows and recommended compute configs at: github.com/clara-parabricks-workflows

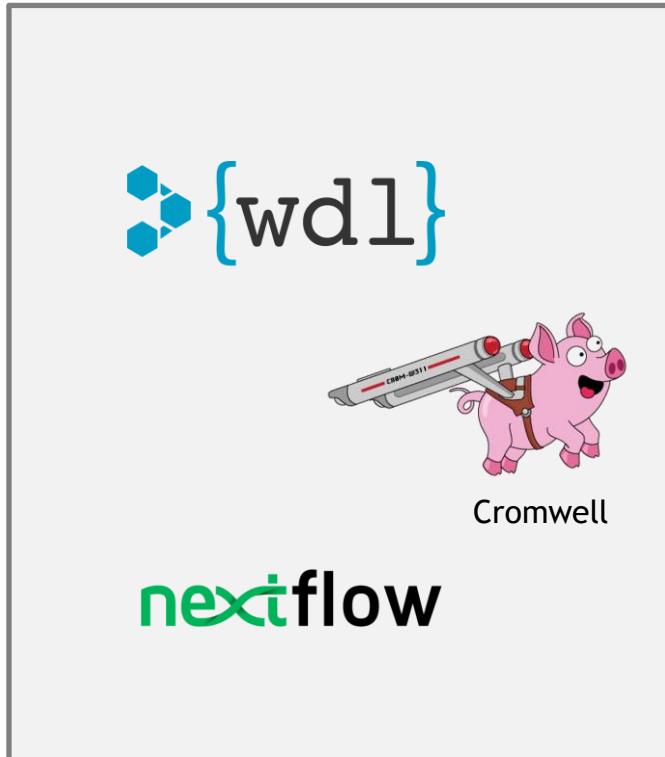
Top 3 Features of Clara Parabricks v4.0

Modular Tools, Workflow Support, DeepVariant v1.4



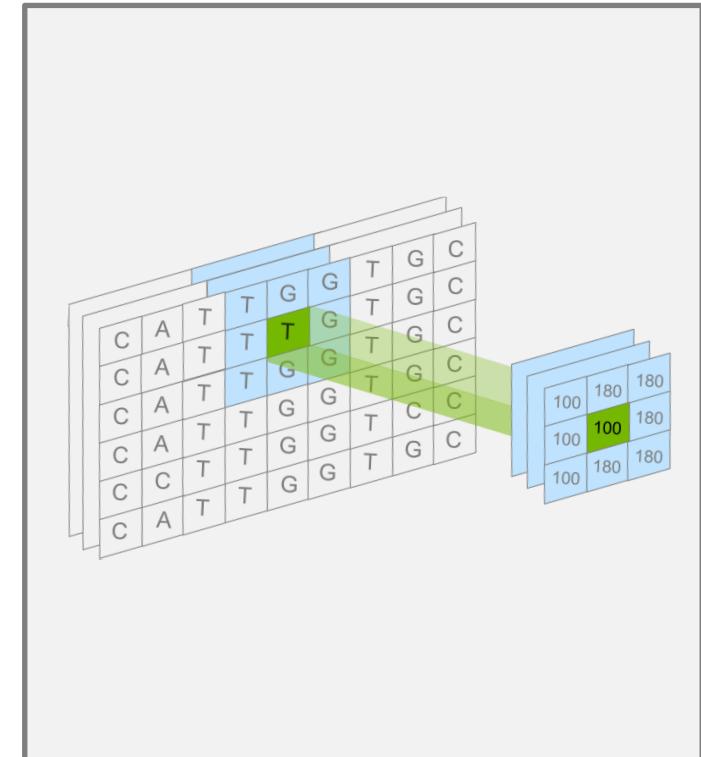
Modular Tools for Flexible Deployment

Accelerated Tools are available for Free on NGC, with option to purchase NVIDIA AI Enterprise to obtain Enterprise Support



Support for Leading Workflow Managers

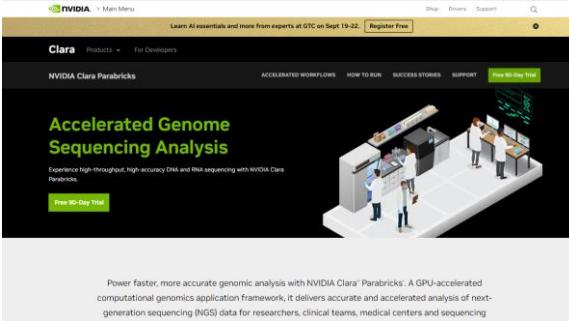
WDL (Cromwell) and NextFlow workflow languages make it easy to scale and customize



DeepVariant 1.4 Germline Caller

DL-Based Variant Caller for long and short read data; with improved accuracy

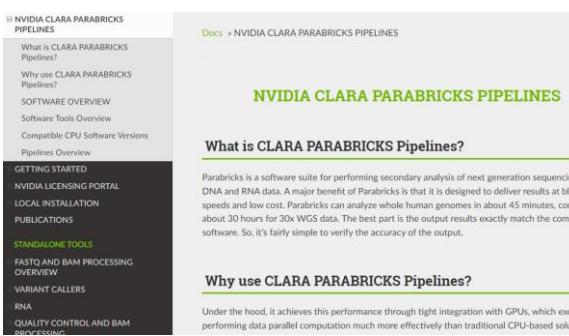
Help and Resources



NVIDIA Clara Parabricks Landing Page



NVIDIA Clara Parabricks Product Sheet & Solution Briefs



NVIDIA Clara Parabricks Pipelines Documentation

RAPIDS

GPU DATA SCIENCE

ⓘ ACCELERATED DATA SCIENCE

The RAPIDS suite of open source software libraries gives you the freedom to execute end-to-end data science and analytics pipelines entirely on GPUs.

[Learn about RAPIDS »](#)

ⓘ TOP MODEL ACCURACY

Increase machine learning model accuracy by iterating on models faster and deploying them more frequently.

[Learn about RAPIDS for model optimization »](#)

⚙ SCALE OUT ONGPUS

Seamlessly scale from GPU workstations to multi-GPU servers and multi-node clusters with Dask.

[Learn about Dask »](#)

⚡ PYTHON INTEGRATION

Accelerate your Python data science toolchain with minimal code changes and no new tools to learn.

[Learn about our libraries »](#)

ⓘ REDUCED TRAINING TIME

Drastically improve your productivity with more interactive data science tools like XGBoost.

[Learn about XGBoost »](#)

[Learn about accelerated ML with cuML »](#)

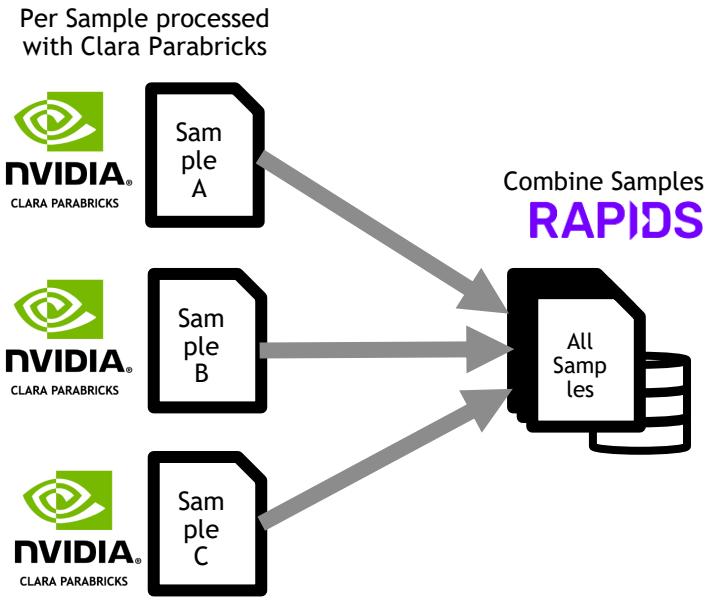
⚡ OPEN SOURCE

RAPIDS is an open source project. Supported by NVIDIA, it also relies on Numba, Apache Arrow, and many more open source projects.

[Learn about our projects »](#)

RAPIDS for Tertiary Analysis in Genomics

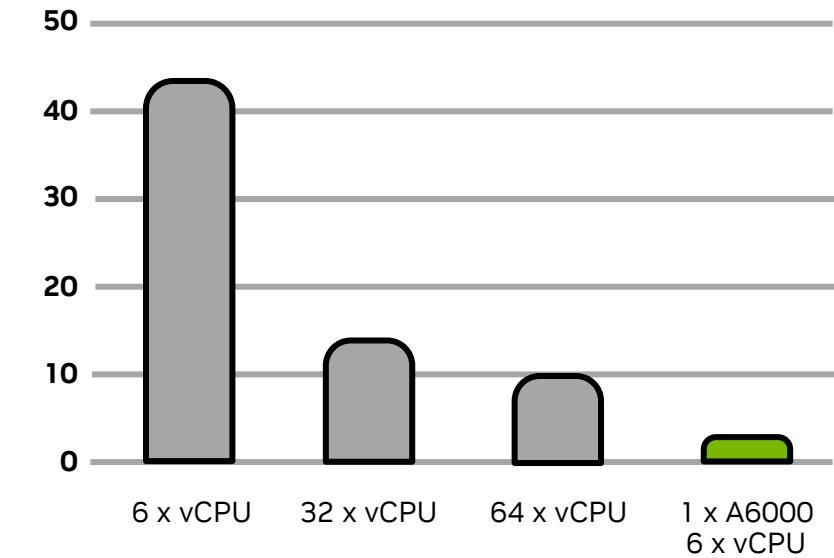
Accelerated Data Frame Manipulation



RAPIDS is built for end-to-end data science on GPU, including dataframe manipulation, math functions, ML, and visualization.

All are important to bioinformaticians. We can start with merging the processed samples ready for tertiary analysis

Merging samples with gINexus methodology
Runtime per 100k sites on 10 samples

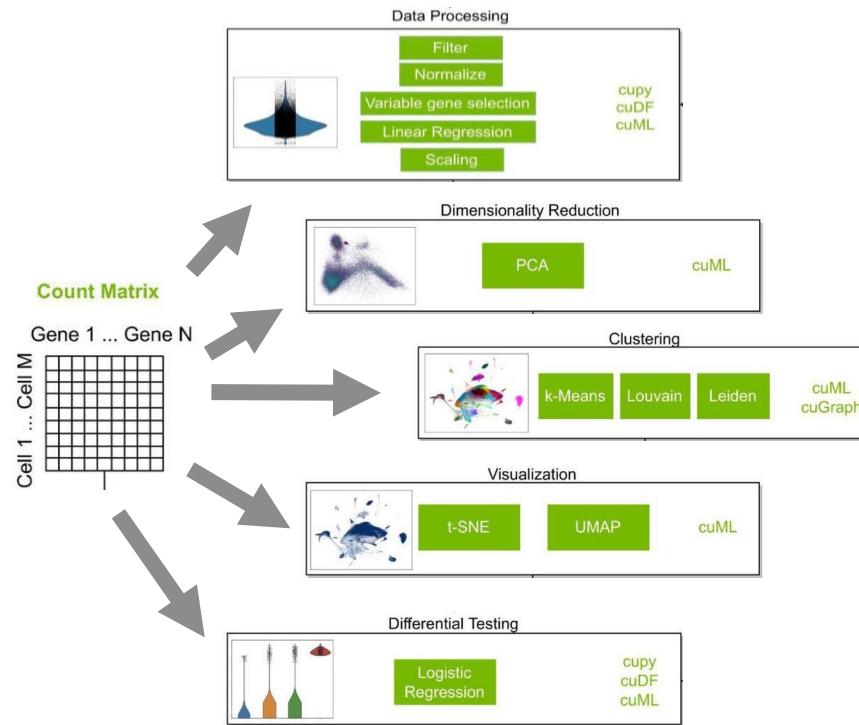


Loading the gVCF and merging the dataframe by allele can be done with cuDNN

This runs in just 3 seconds per 100k sites on a single A6000

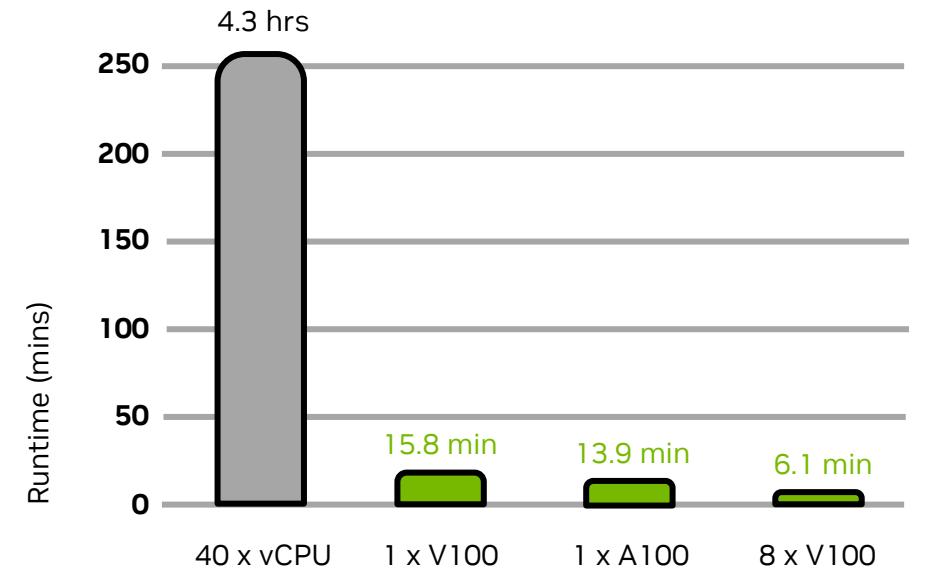
RAPIDS for Tertiary Analysis in Genomics

Accelerated processing, clustering and visualization



Single cell tertiary analysis relies heavily on regression, clustering and visualization methods

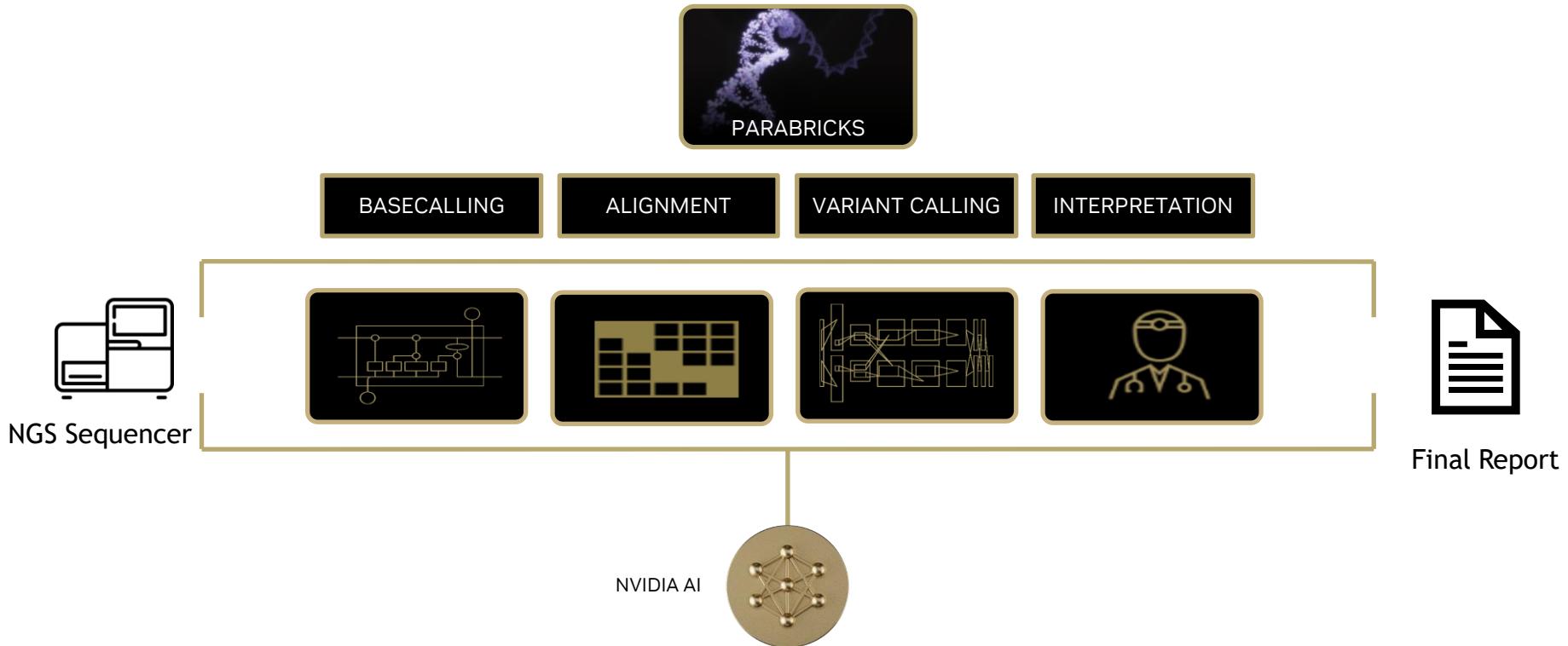
scRNA-Seq processing, clustering, regression & visualization
End-to-end runtime for 1.3 million cells



End-to-end runtime of all the processes on the left can be accelerated from 4.3 hours to 6.1 mins (42x acceleration)

The Future of Genomics is on GPU's

With Deep Learning and Acceleration



SUMMARY

NVIDIA Clara for Healthcare and Life Sciences

World's Largest Data Industry | 36% CAGR by 2025



IMAGING



MEDICAL
DEVICES



DRUG
DISCOVERY



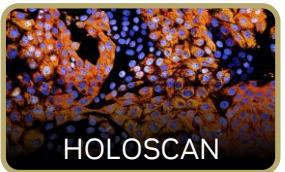
GENOMICS



NVIDIA CLARA
APPLICATION
FRAMEWORKS



FLARE



HOLOSCAN



BIONEMO



PARABRICKS



NVIDIA AI



NVIDIA Omniverse

My Mission today

Welcome to Join NVIDIA Developer

- Encourage you to join
<https://developer.nvidia.com/>
- What is the goal of NVIDIA?
- Overview of Deep Learning in Healthcare
 - Medical Imaging - MONAI
 - Medical Device - Holoscan
 - Drug discovery - BioNemo
 - Genomics - Parabricks



