SNIA NSF | NETWORKING STORAGE

Storage for AI Applications

Live Webcast

October 5, 2021

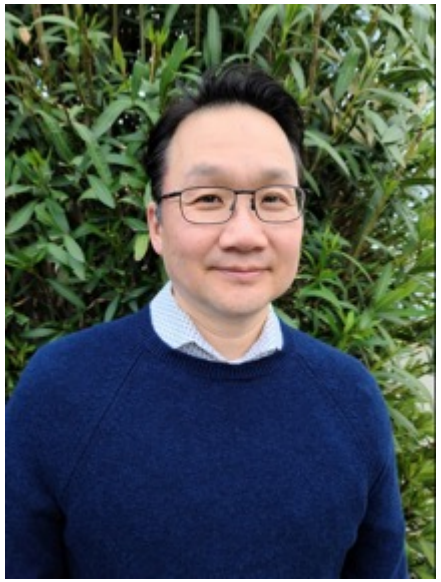
10:00 am PT / 1:00 pm ET

Today's Presenters

Tom Friend
Moderator
Chief Engineer
Illuminosi

Craig Tierney
Principal Product Architect
NVIDIA

Brien Porter
Solutions Lead
Intel



Young Paik
Senior Director
Product Planning
Samsung

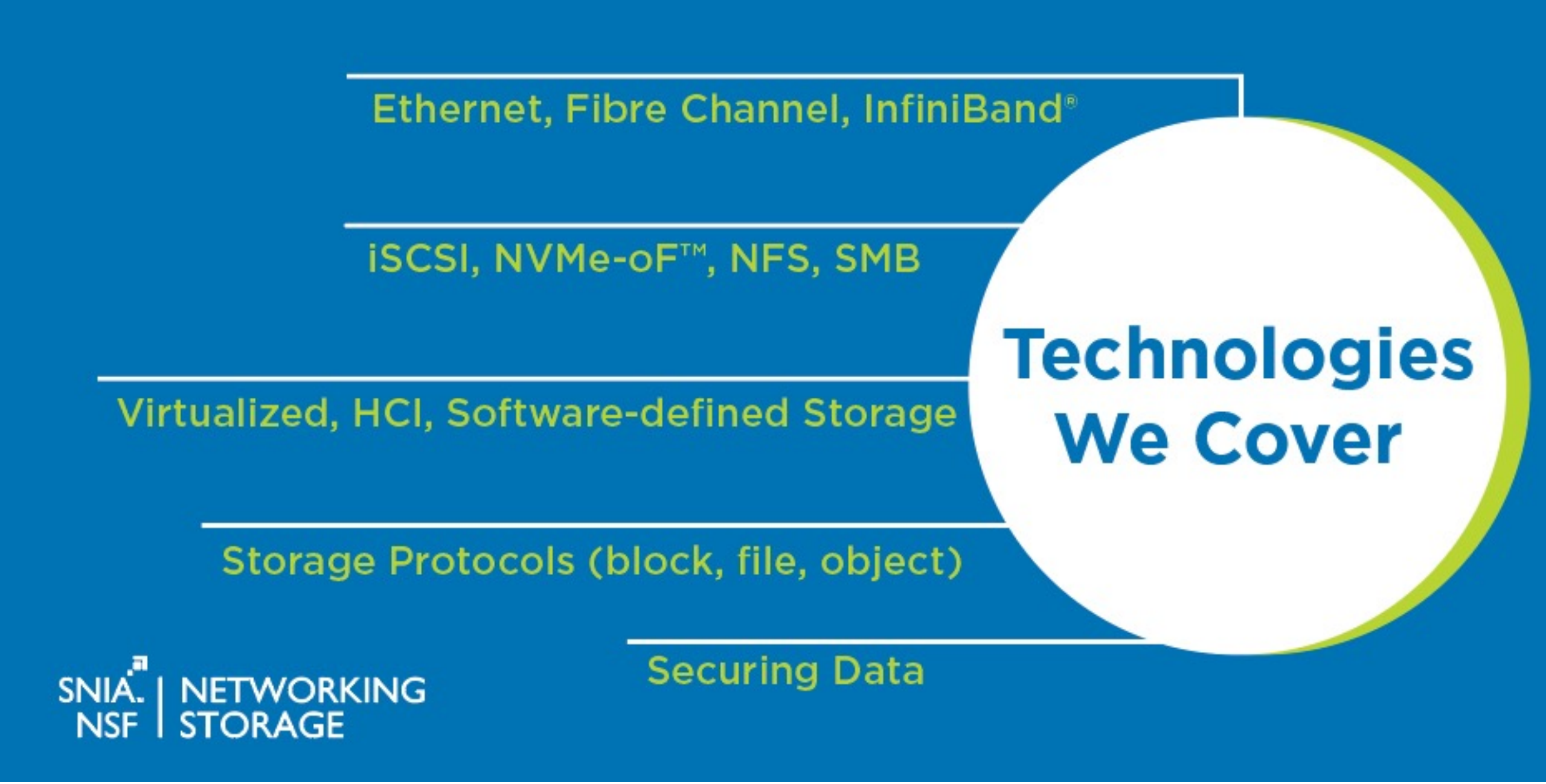
SNIA-at-a-Glance

180
industry leading organizations

2,500
active contributing members

50,000
IT end users & storage pros worldwide

Learn more: **snia.org/technical** **@SNIA**



SNIA Legal Notice

- The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

- Introduction to Data Science
- The Data and AI Lifecycle
- Mapping Business Questions to Algorithms
- Model training and data usage

- Roundtable Q&A

Accelerated Data Science

DATA ANALYTICS

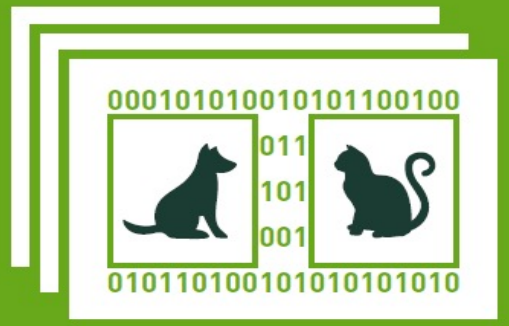
Extracting insights from big data

MACHINE LEARNING

Learning from examples in the data

DEEP LEARNING

Automating feature engineering



2.2 exabytes (2.2B GB) of data created daily – McKinsey

$260B annual revenue by 2022 for big data and business analytics – IDC

Design an Intelligent AI Data Services Platform

Modernize → **Transform**

Storage Intelligence

Drive efficiencies for traditional architecture

Thin-provisioning

De-duplication

Compression

Encryption

Storage Optimization

Maximize performance of new data center technologies

Data Inventory

Data Tiering

Optimize Performance

All Flash/NVM
Storage Memory Technology
Hybrid
Legacy SAS

Lower TCO

Automation

Architecture & workload unifying cloud & enterprise

Scale Out Architecture

Hyperconverged Infrastructure

Software Defined Storage

Orchestration

Seamless integration across architecture, media, vendors

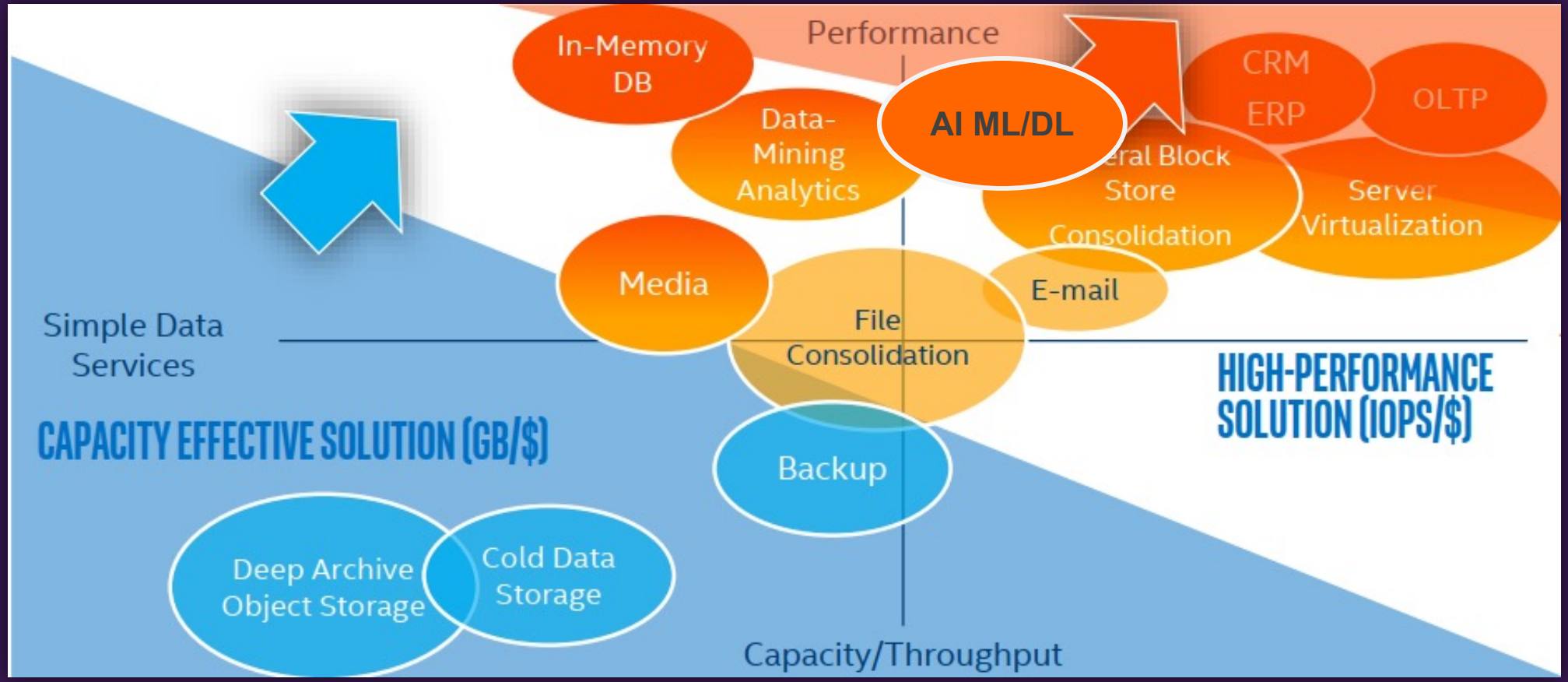
Network Function Virtualization

Software Defined Infrastructure

Shift of the IO Bottleneck

Business Alignment

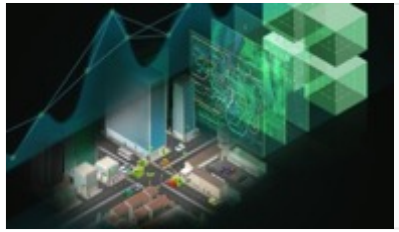
Different Workloads Drive Different Solutions



Cost-effective solutions are best suited for cold storage

High-performance solutions are best suited for warm and some hot tiers

The NEED FOR Data Science



CONSUMER INTERNET

Ad Personalization
Click Through Rate Optimization
Churn Reduction

FINANCIAL SERVICES

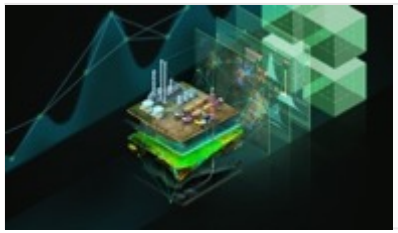
Claim Fraud
Customer Service Chatbots/Routing
Risk Evaluation

HEALTHCARE

Improve Clinical Care
Drive Operational Efficiency
Speed Up Drug Discovery

RETAIL

Supply Chain & Inventory Management
Price Management / Markdown Optimization
Promotion Prioritization And Ad Targeting



OIL & GAS

Sensor Data Tag Mapping
Anomaly Detection
Robust Fault Prediction

MANUFACTURING

Remaining Useful Life Estimation
Failure Prediction
Demand Forecasting

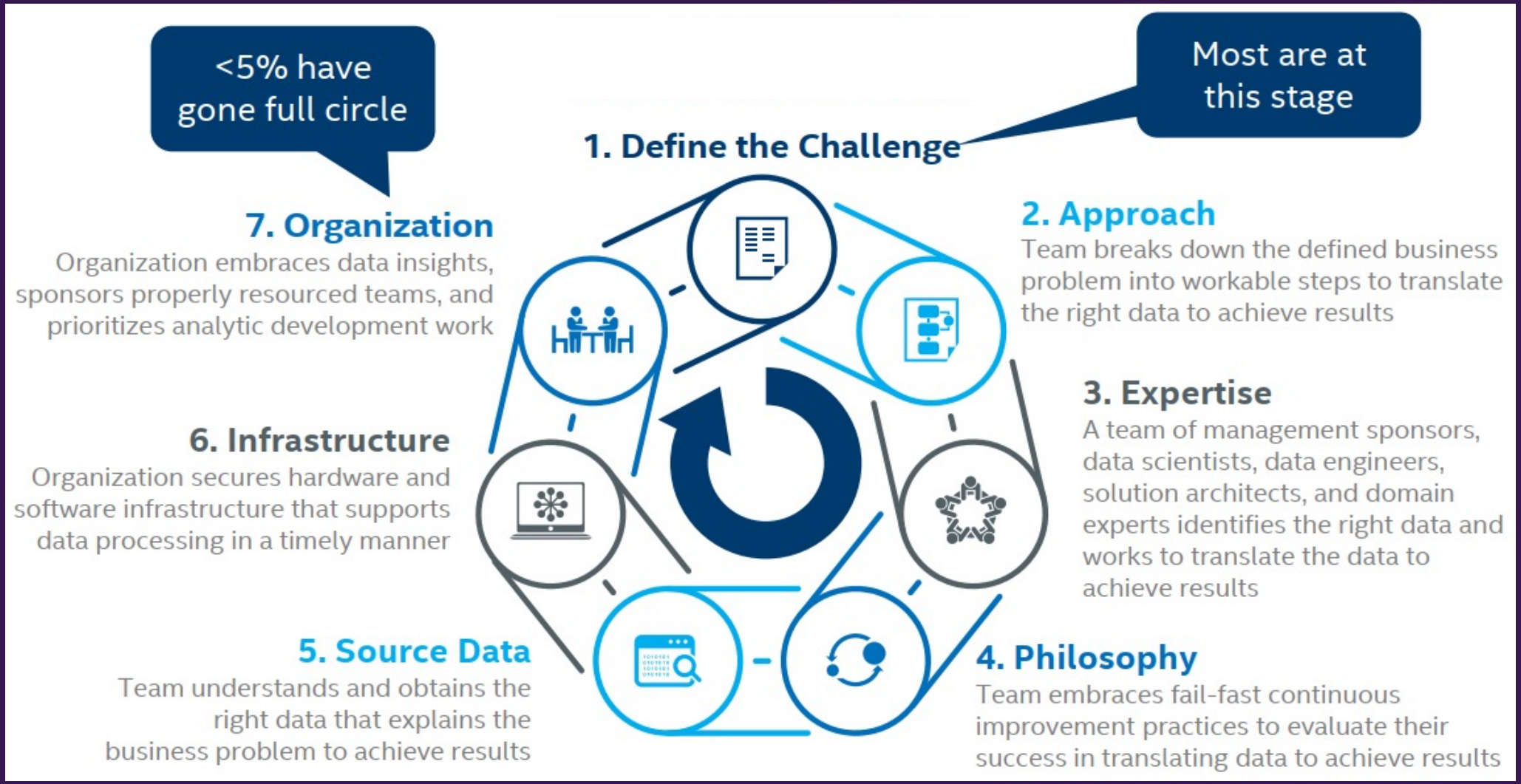
TELECOM

Detect Network/Security Anomalies
Forecasting Network Performance
Network Resource Optimization (SON)

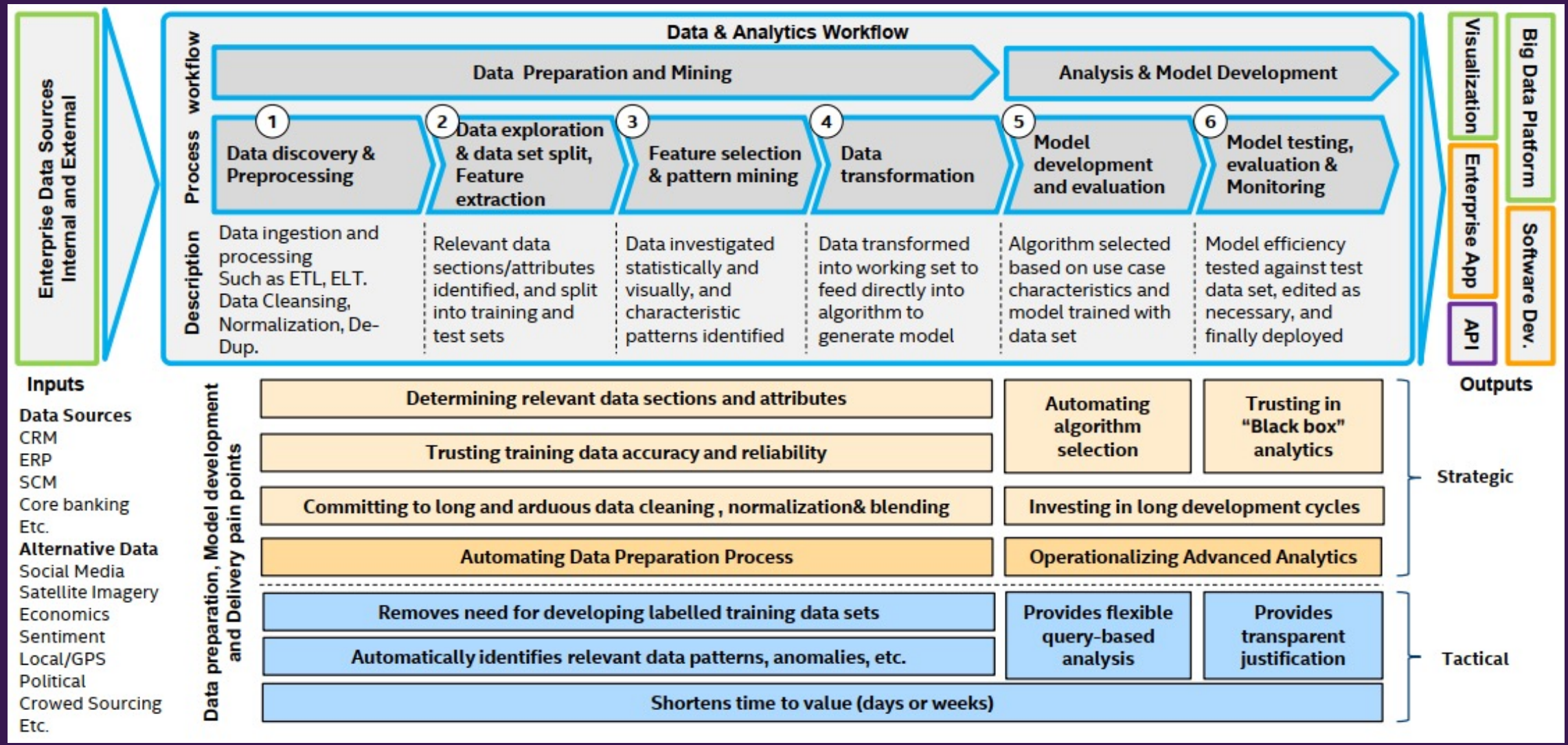
AUTOMOTIVE

Personalization & Intelligent Customer Interactions
Connected Vehicle Predictive Maintenance
Forecasting, Demand, & Capacity Planning

The AI Lifecycle



Understanding Enterprise Data & Analytics Workflow



WHAT PROBLEM ARE YOU SOLVING?

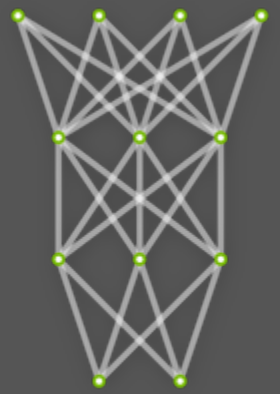
Defining the AI/DL task

INPUTS	BUSINESS QUESTIONS	AI / DL TASK	EXAMPLE OUTPUTS		
			HEALTHCARE	RETAIL	FINANCE
 Text Data  Images  Video  Audio	Is “it” **present** or not?	Detection	Cancer Detection	Targeted Ads	Cybersecurity
	What **type** of thing is “it”?	Classification	Image Classification	Basket Analysis	Credit Scoring
	To what **extent** is “it” present?	Segmentation	Tumor Size / Shape Analysis	Build 360° Customer View	Credit Risk Analysis
	What is the likely **outcome**?	Prediction	Survivability Prediction	Sentiment & Behavior Recognition	Fraud Detection
	What will likely satisfy the objective?	Recommendations	Therapy Recommendation	Recommendation Engine	Algorithmic Trading

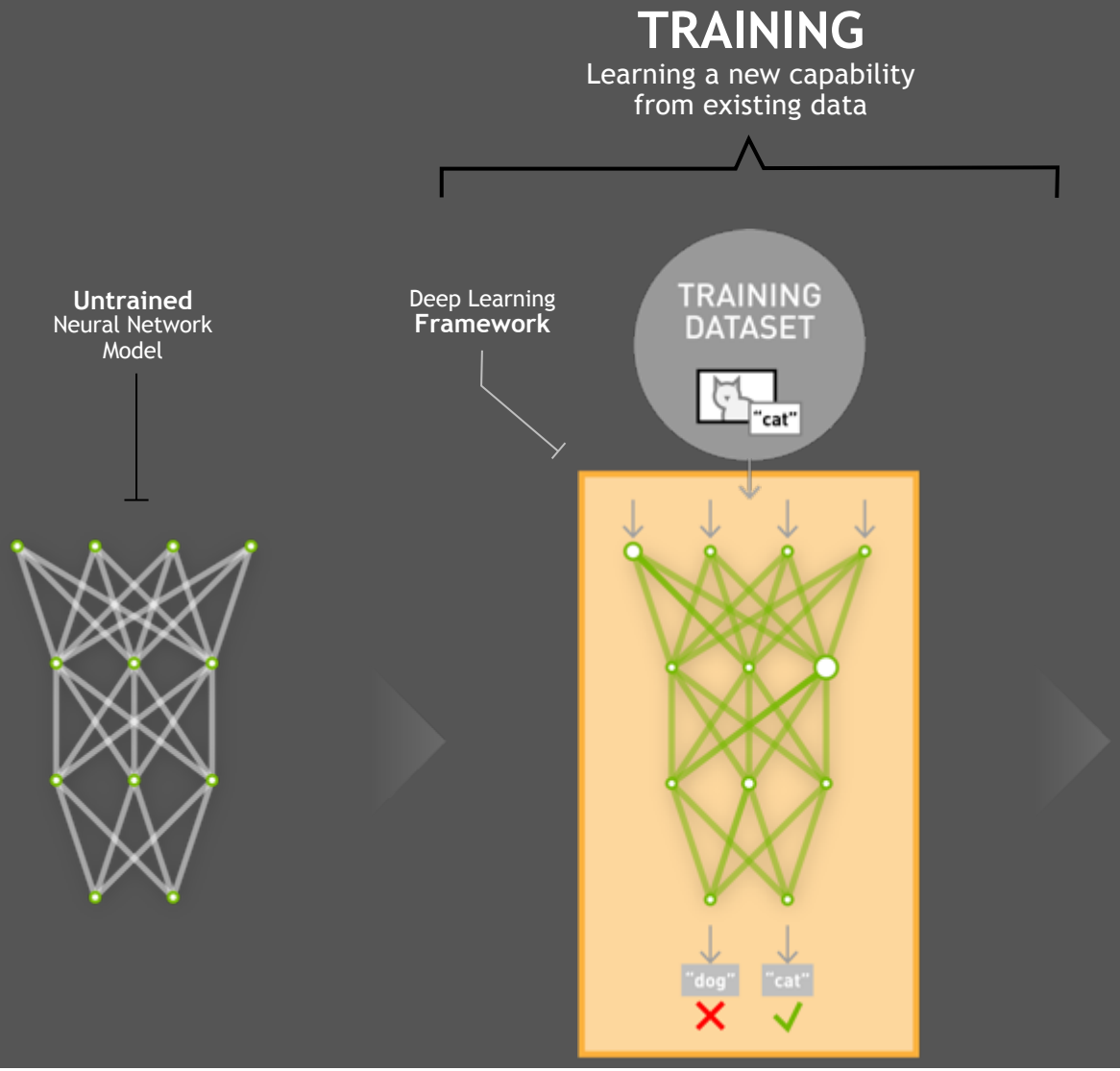
AI/ML/DL Application Development

AI/ML/DL Application Development

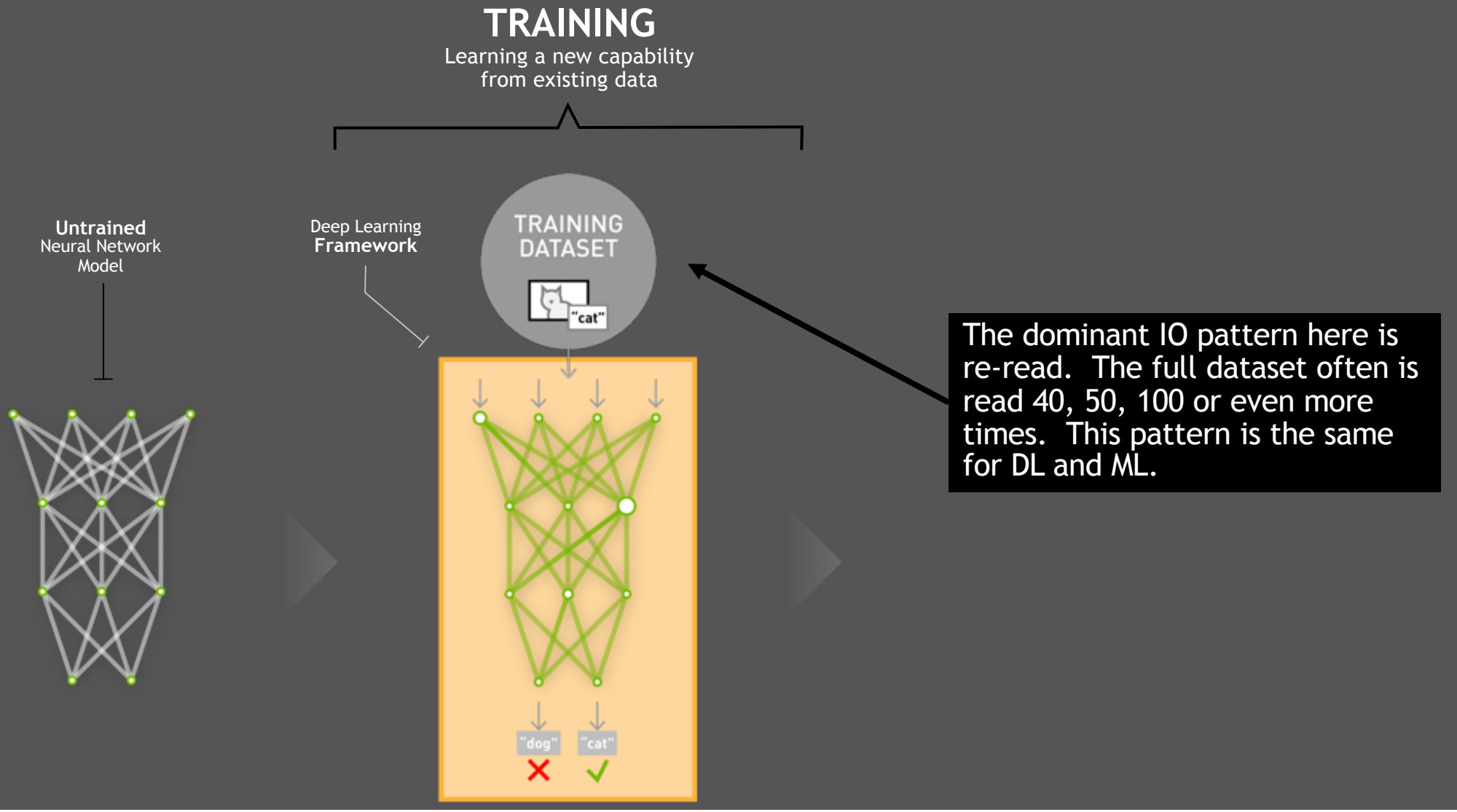
Untrained
Neural Network
Model



AI/ML/DL Application Development

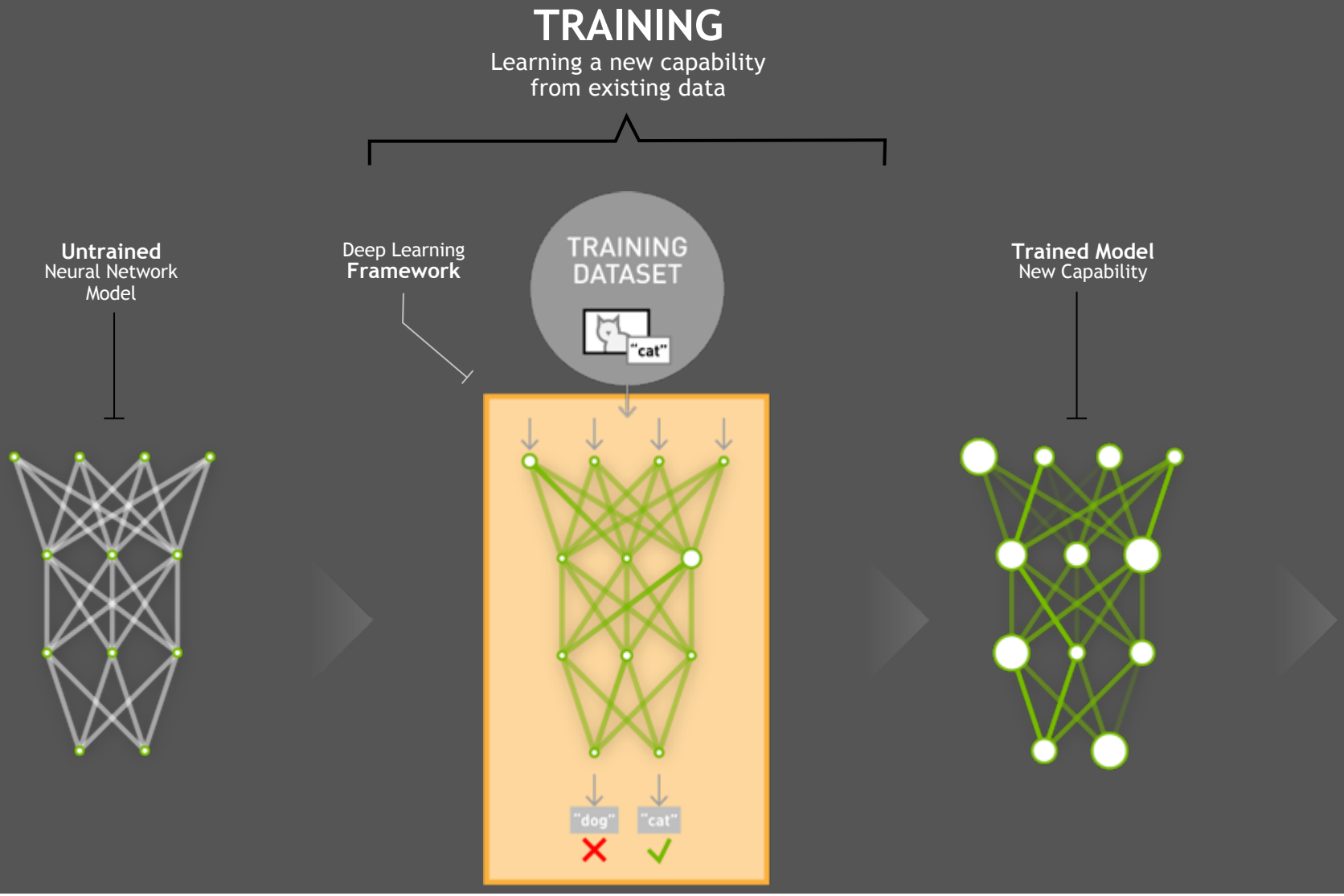


AI/ML/DL Application Development

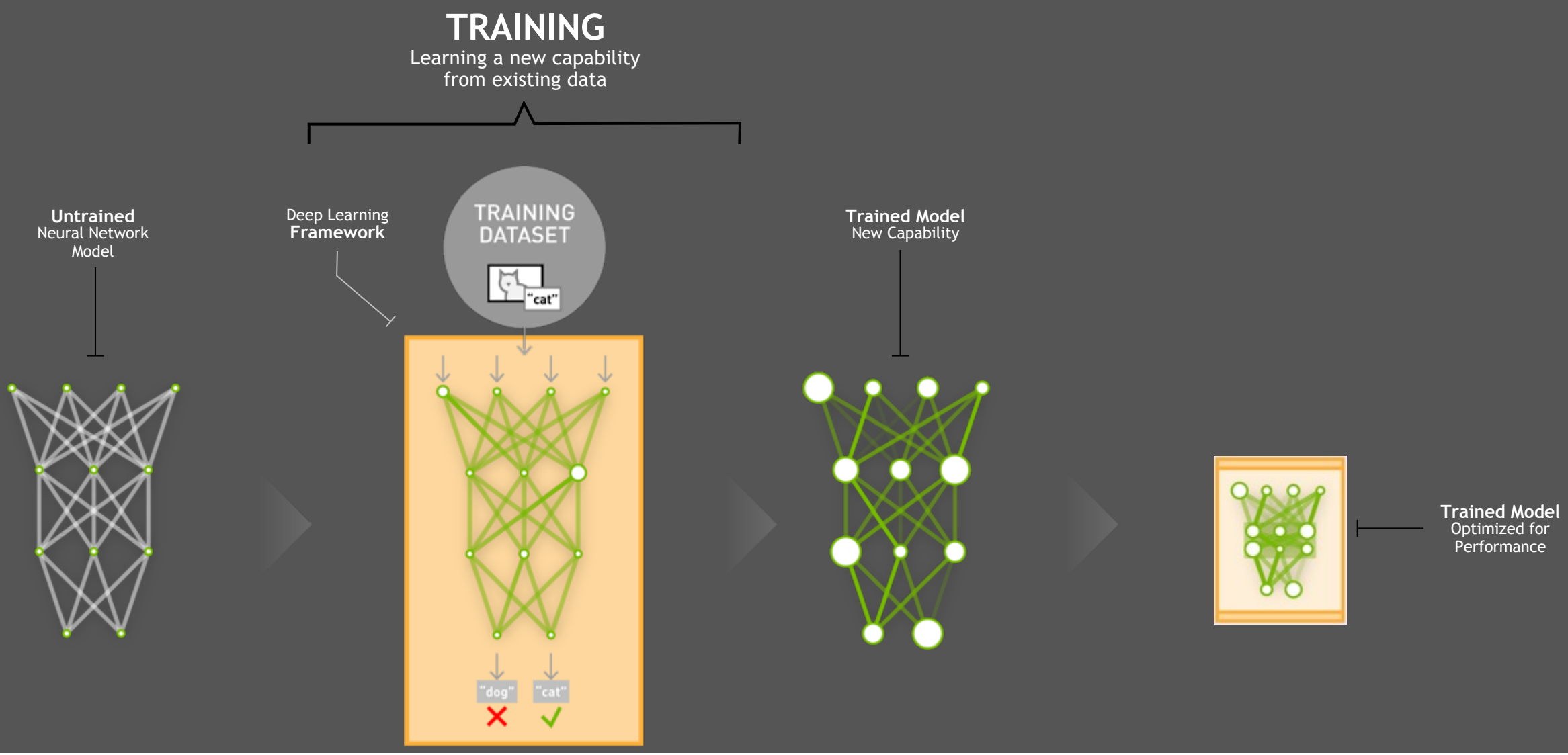


The dominant IO pattern here is re-read. The full dataset often is read 40, 50, 100 or even more times. This pattern is the same for DL and ML.

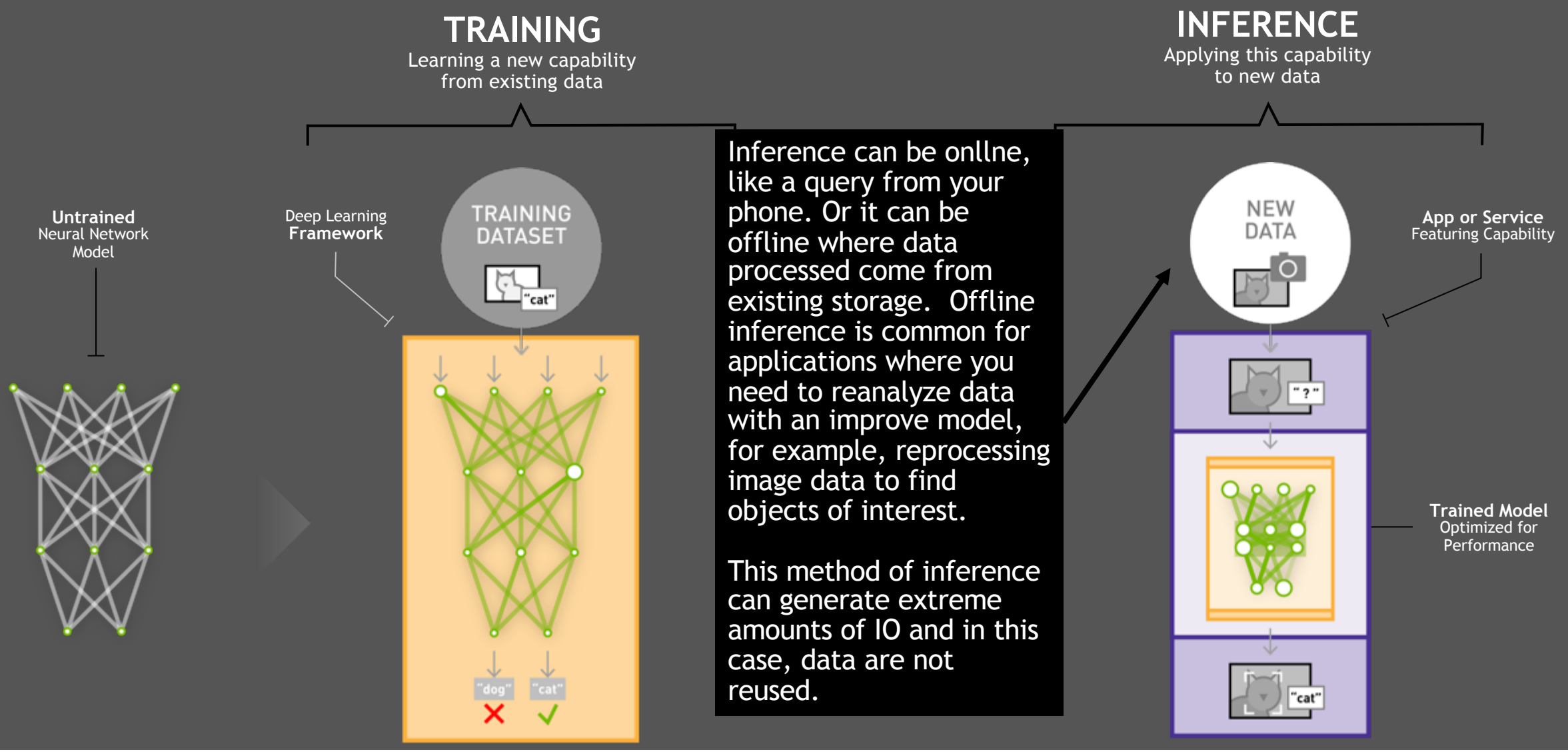
DEEP LEARNING Application Development



AI/ML/DL Application Development



AI/ML/DL Application Development



Roundtable Q&A

After this Webcast

- Please rate this webcast and provide us with your feedback
- This webcast and a copy of the slides will be available at the SNIA Educational Library https://www.snia.org/educational-library
- A Q&A from this webcast, including answers to questions we couldn't get to today, will be posted on our blog at https://sniansfblog.org/
- Follow us on Twitter @SNIANSF

Thank You

