Accelerating the

Open AI SW Ecosystem

for AI Everywhere

Wei Li – VP/GM AI & Analytics

Imagine a world with AI everywhere…

AI Applications Today

Faster Nerve Detection

Through Ultrasound

Healthcare

Improving Customer Experience

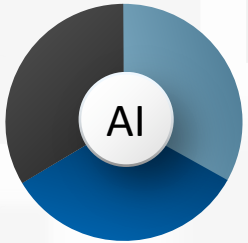
Through Recommender Systems

Finance

Identifying Clean Water Sources

With Cameras & AI Inferencing

Environment

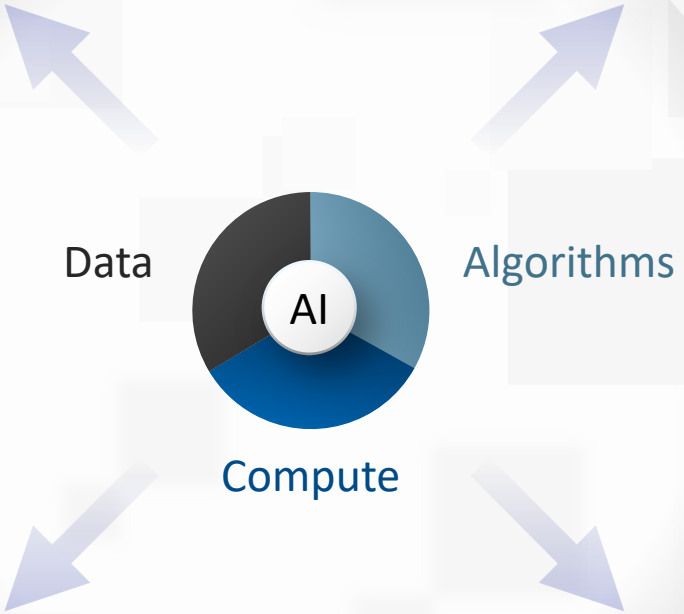


Find more of our many AI success stories at intel.com/SoftwareFirst

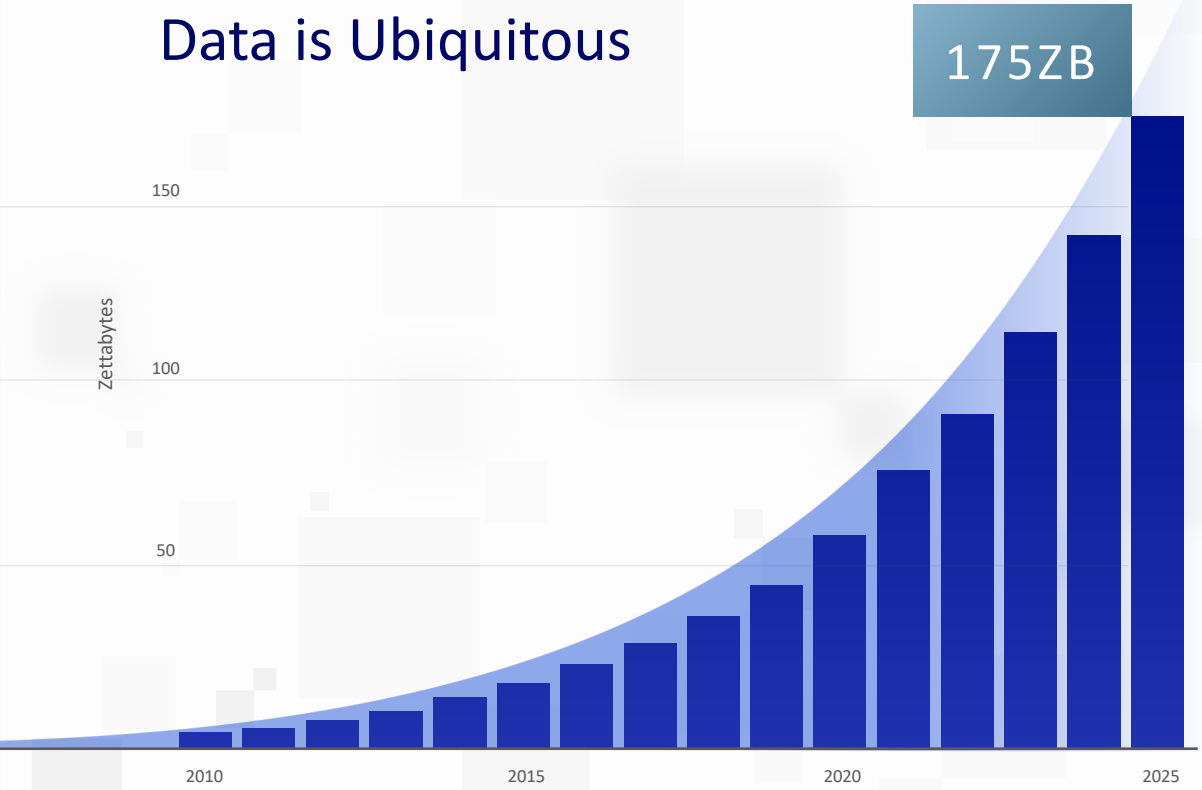
… And Several Others

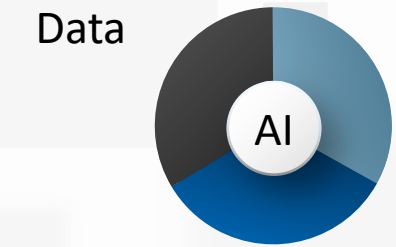
Bringing

AI Everywhere



Data is Ubiquitous

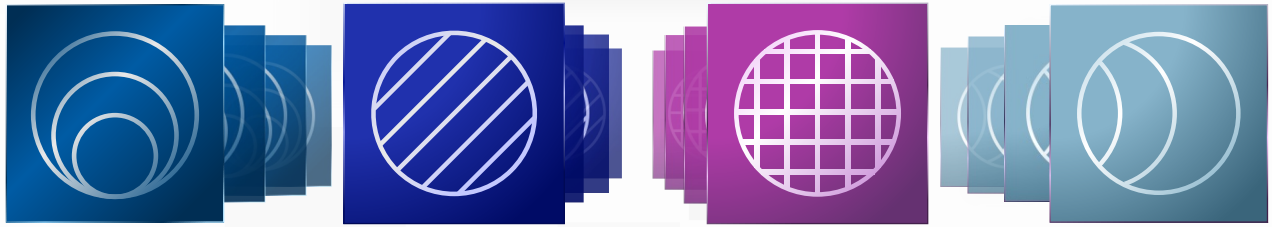




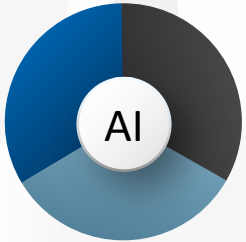
Source: IDC (2021)

… we are building the right Compute

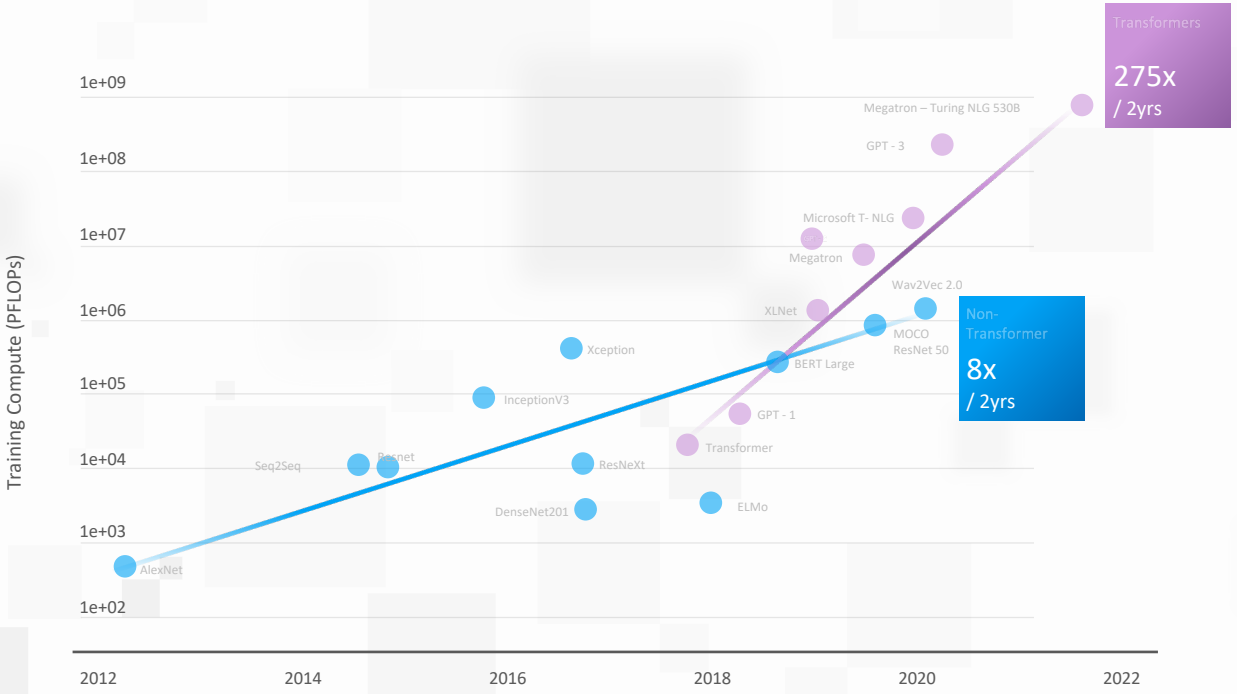
Scalar Vector Matrix Spatial



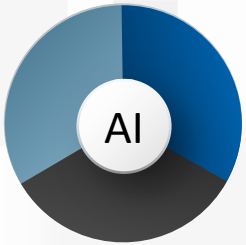
Compute



…and models are growing in sizes & variety



Algorithms



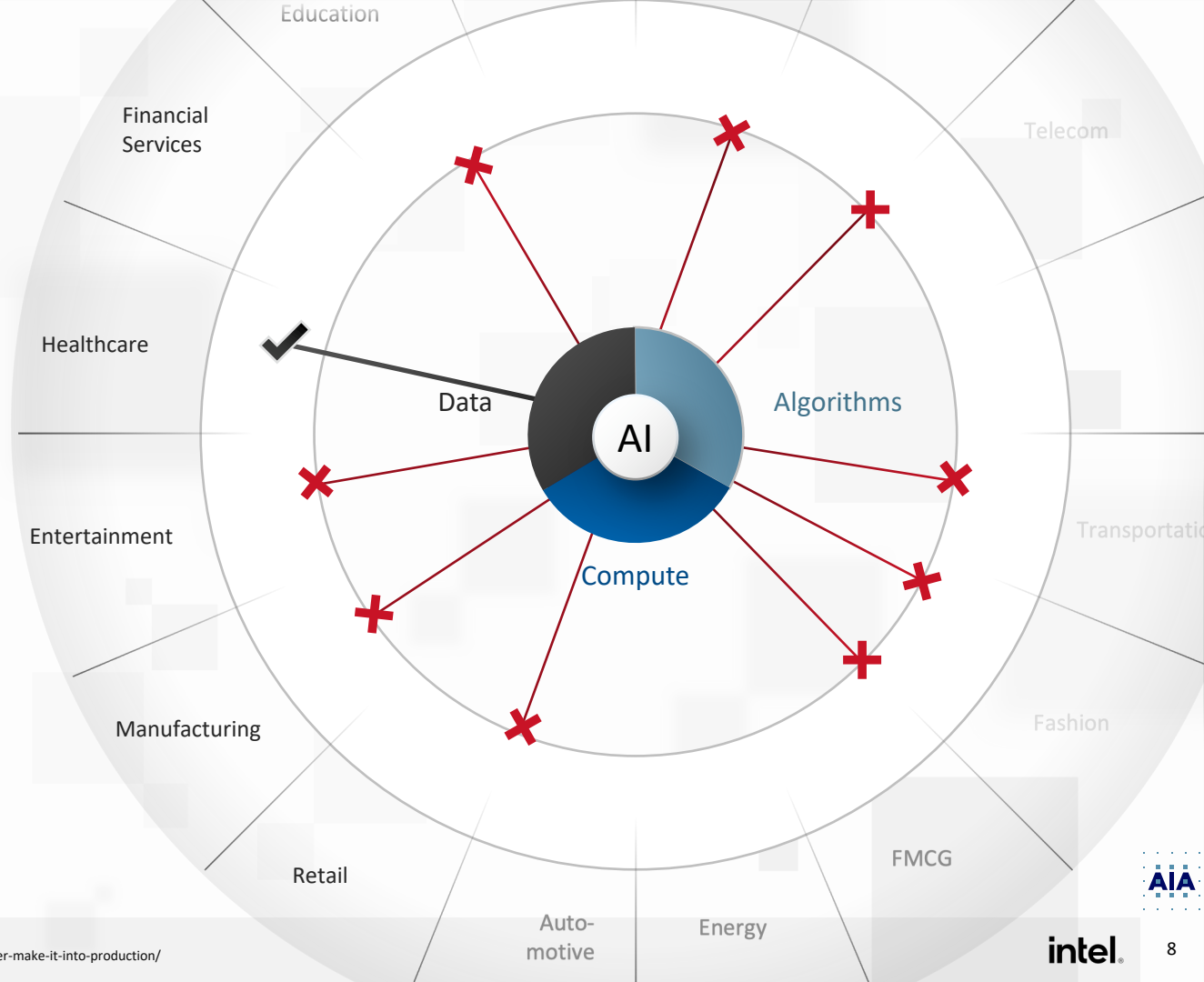
Source: Moore, S. (2022), "Nvidia's Next GPU Shows That Transformers Are Transforming AI", IEEE Spectrum.

But…

87%

of AI concepts
do not make it

- Lack of Performance
- Infrastructure Maturity
- SW Complexity
- Lack of Tools

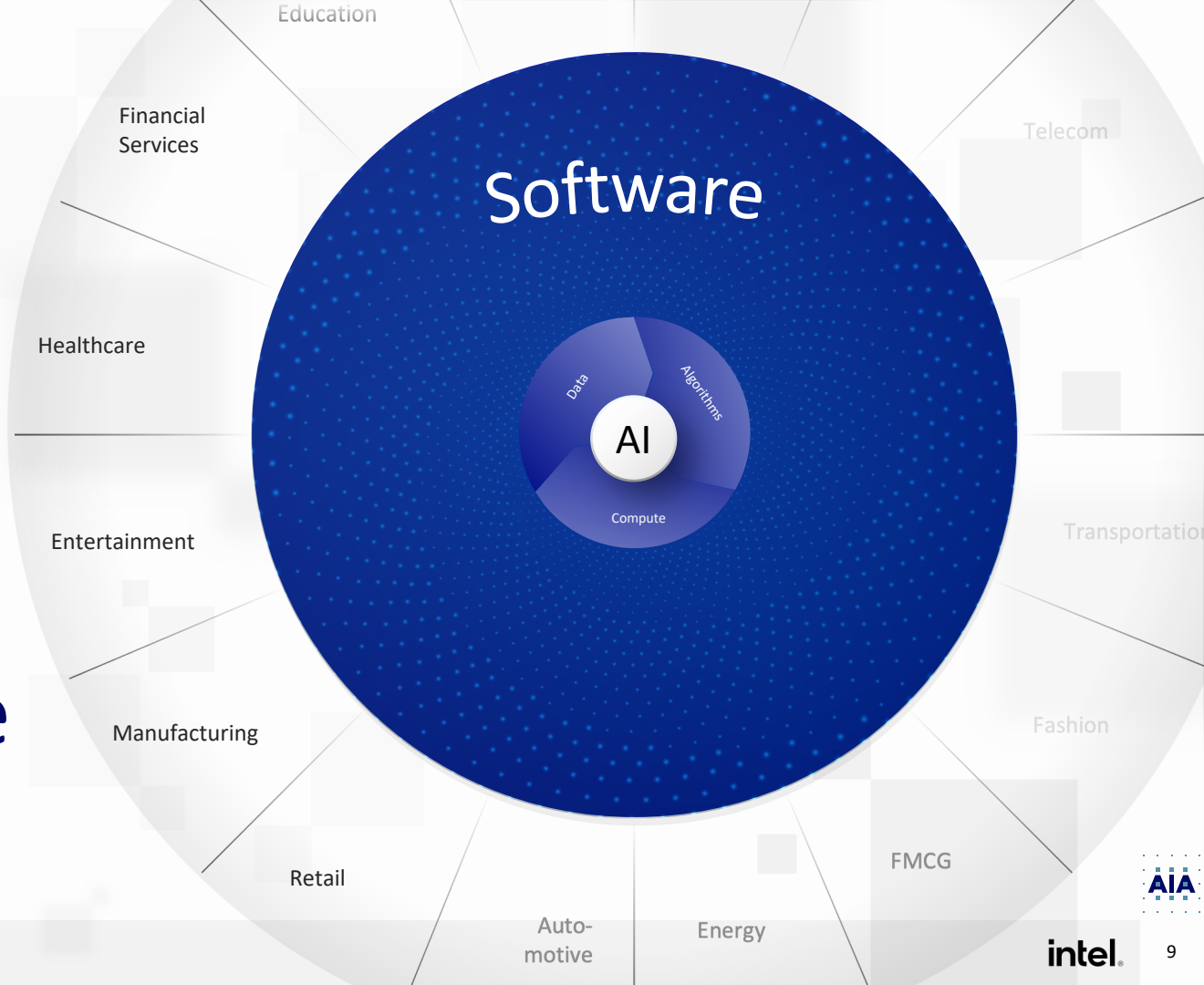


Source: VentureBeat, 2019
https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/

Software is the bridge

to bring

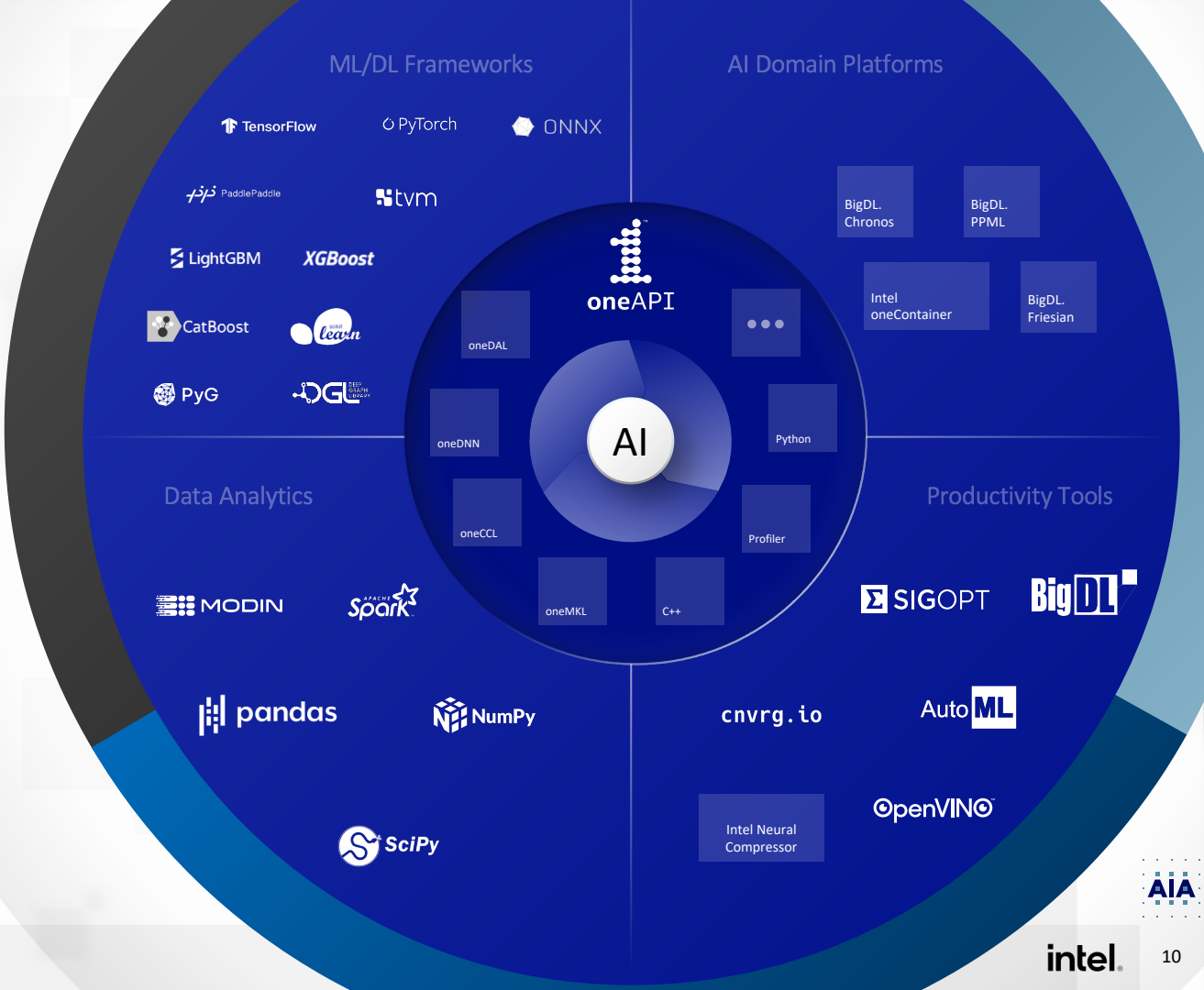
AI Everywhere



Open

AI Software Ecosystem

Key Enabler



* Other names and brands may be claimed as the property of others

oneAPI

An Open Project & Intel's Product

Open Specification for Accelerated Computing

Standards-Based Data Parallel Language

Standard Interfaces for Common Accelerator Libraries

Open-source implementations on diverse non-Intel CPU, GPU, FPGA, and AI solutions

Intel's Implementation of the oneAPI Specification

First Customer Shipment – Dec 2020

Productive, Performant, Cross-Platform

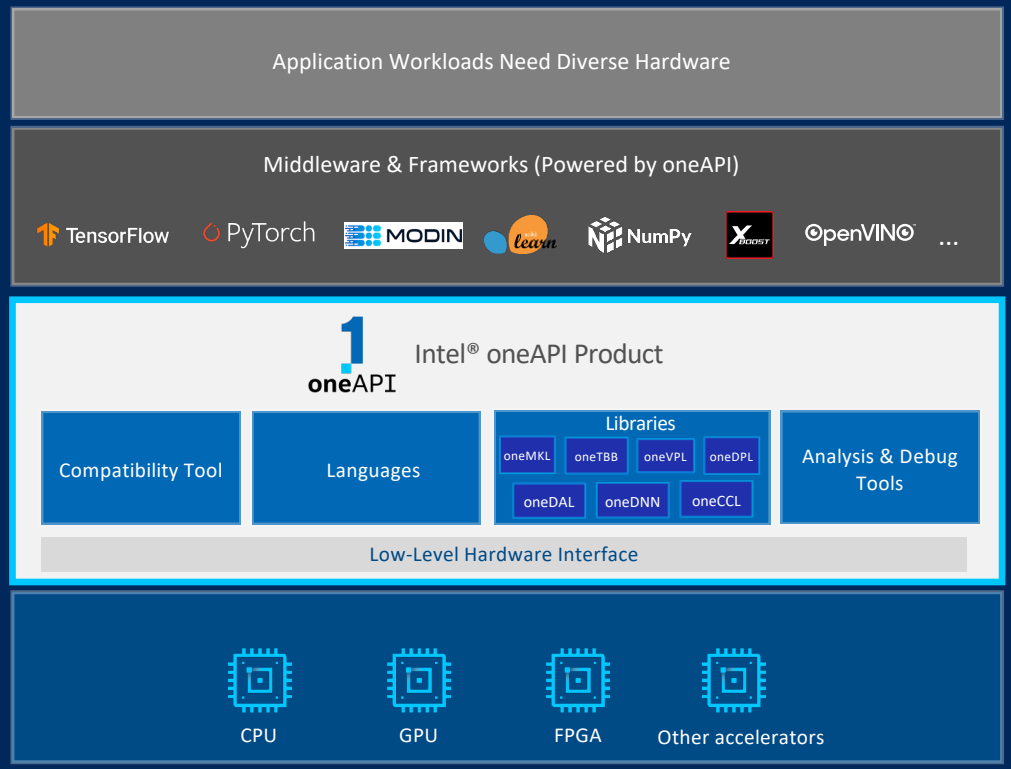
Supports Intel CPU, GPU (integrated & discrete), and FPGA today

Realizing the vision of productive programming for accelerators, free from proprietary lock-in

Intel® oneAPI Software Tools for AI & Analytics

Popular AI frameworks and middleware are extended and optimized using one or more of the oneAPI industry specification elements

Can target CPUs, GPUs, and other accelerators

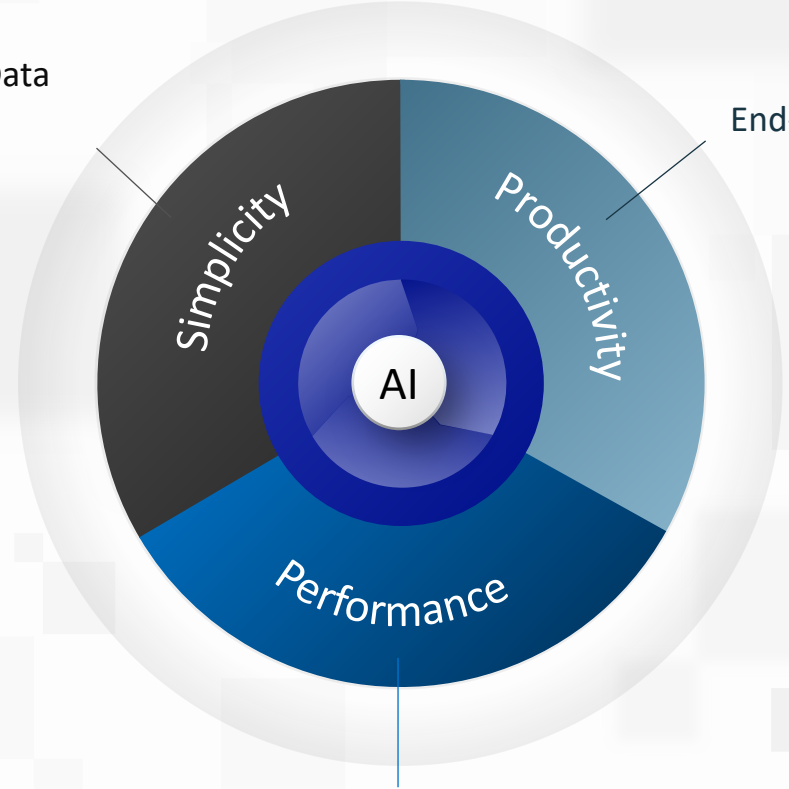


Visit software.intel.com/oneapi for more details
Some capabilities may differ per architecture and custom-tuning will still be required. Other accelerators to be supported in the future.

Intel AI Software Strategy

to deliver

Simplicity, Productivity, and Performance



Performance

for Every AI workload

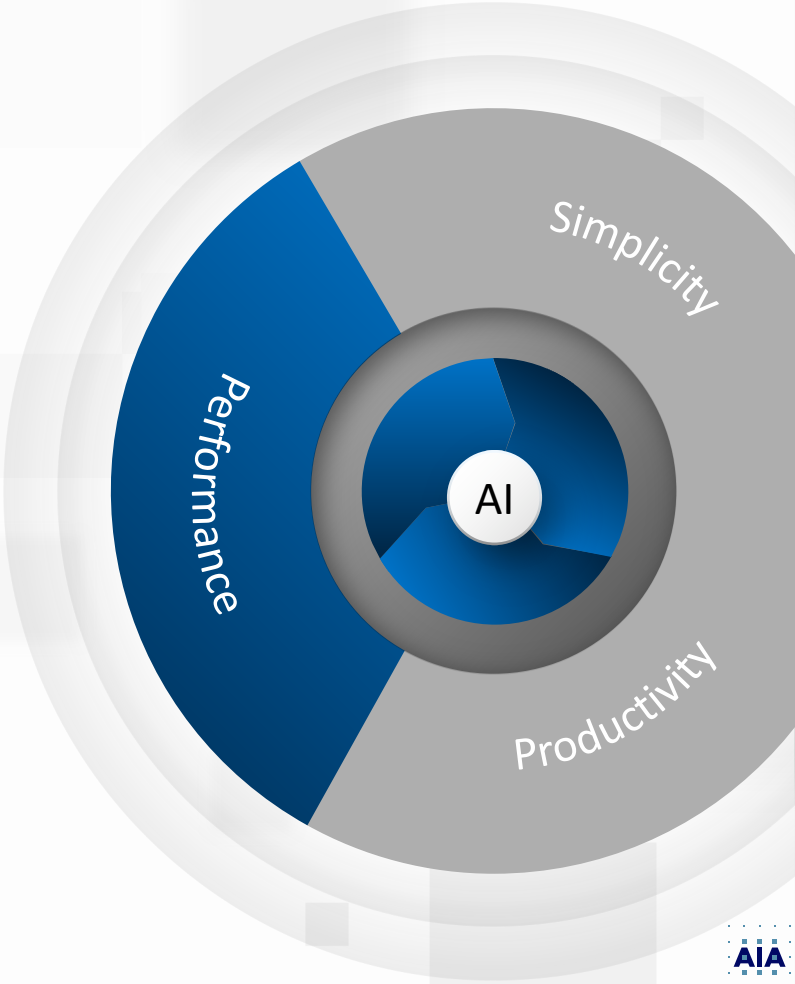
PyTorch TensorFlow dmlc XGBoost scikit-learn MODIN NumPy •••

oneAPI

General Purpose — Purpose Built

intel CORE | intel XEON | Xe | intel MOVIDIUS | habana An Intel Company

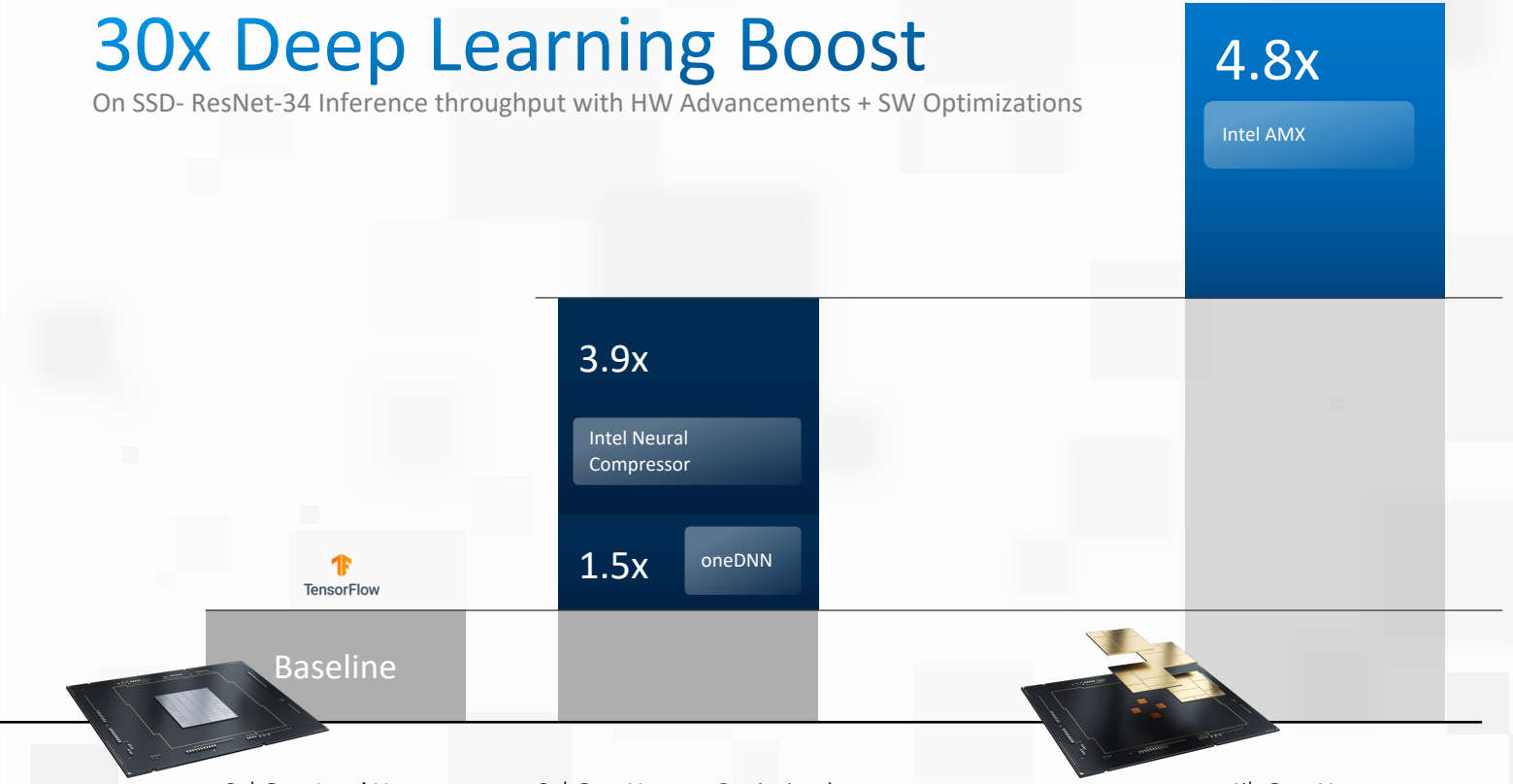
CPU GPU Accelerators

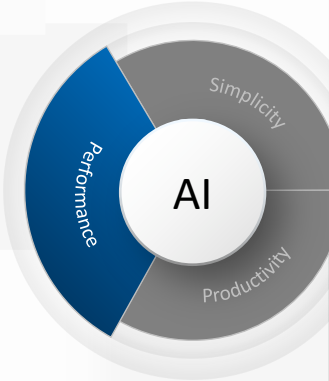


* Other names and brands may be claimed as the property of others

30x Deep Learning Boost

On SSD- ResNet-34 Inference throughput with HW Advancements + SW Optimizations

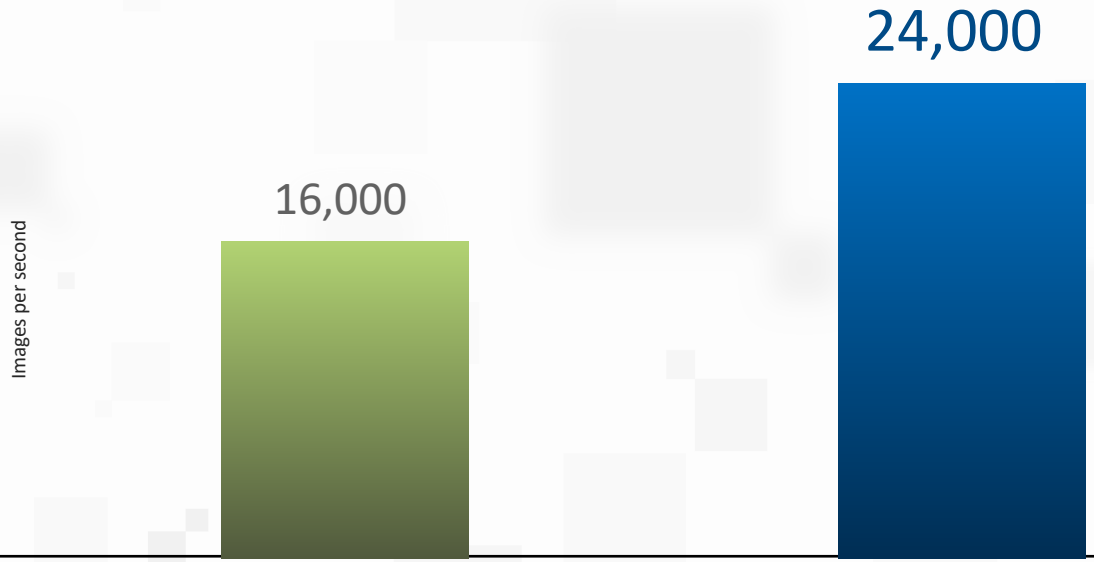


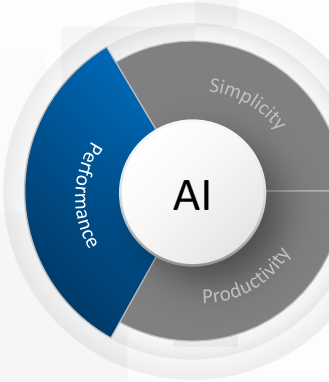


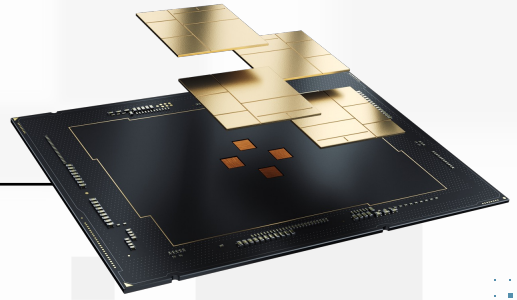
SSD-ResNet-34 Inference Throughput (Batch Size =1) For workloads and configurations visit www.intel.com/InnovationEventClaims. Results may vary.

Xeon with AMX 1.5x Faster

vs. Nvidia A30 on Resnet50



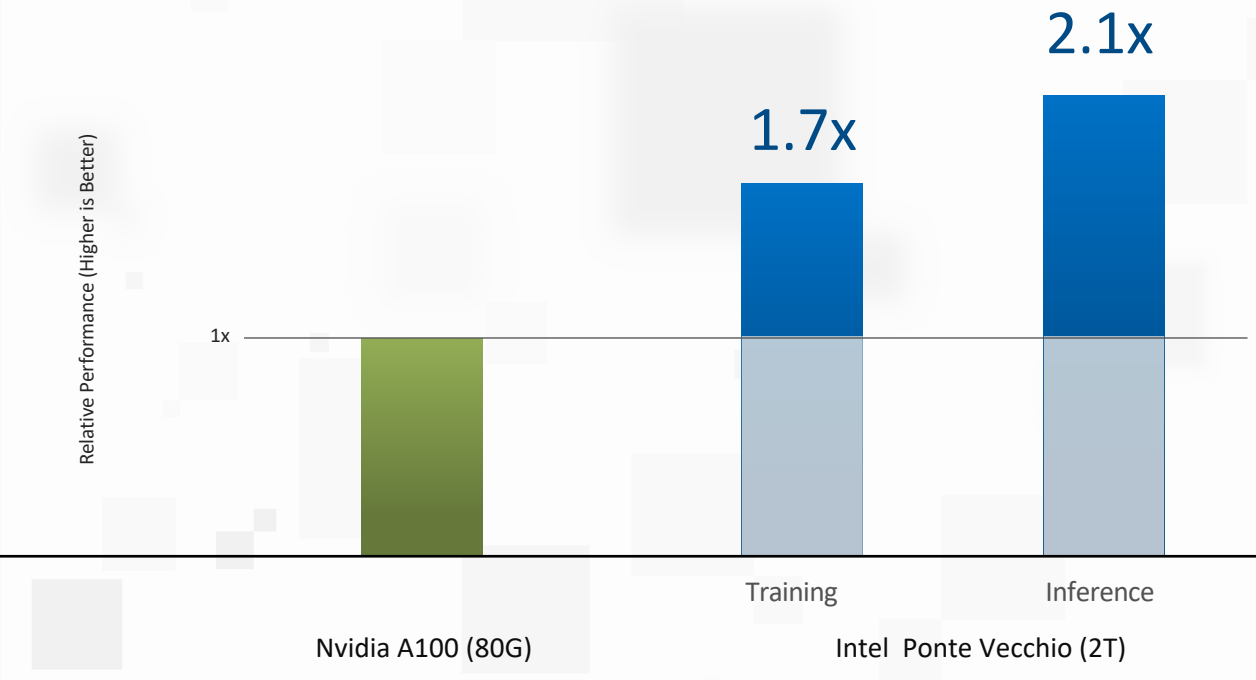


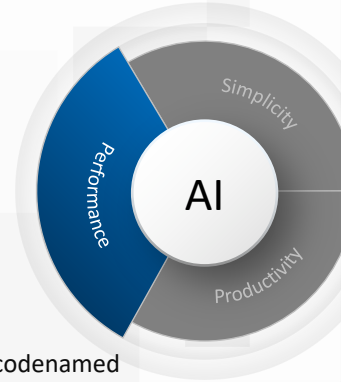


Based on pre-production measurements. See backup for workloads and configurations or visit www.intel.com/innovationeventclaims. Results may vary.
Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Leadership AI Performance

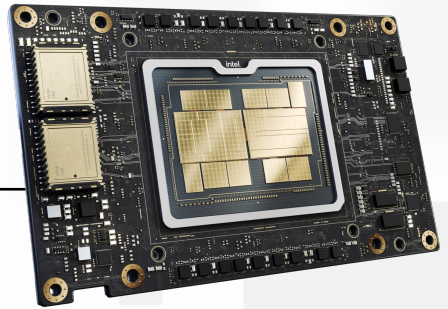
ResNet50 - Ponte Vecchio B step @1.4GHz vs A100(80G)





Intel data center GPU codenamed

Ponte Vecchio



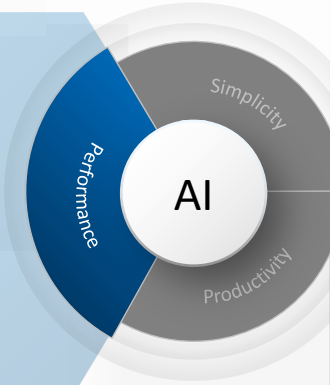
Based on pre-production measurements. See backup for workloads and configurations. Results may vary.
Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

new

TensorFlow + intel ai

TensorFlow 2.9 now accelerated with Intel oneDNN by Default

- Increased Productivity
- Faster Inference & Training
- Efficient Compute Utilization



Accelerating workflows for

LAIKA IBM Watson Google Health LILT ...

AIA

Simplicity

End-to-End Solutions For Every Industry

Apply Analytics & Machine Learning

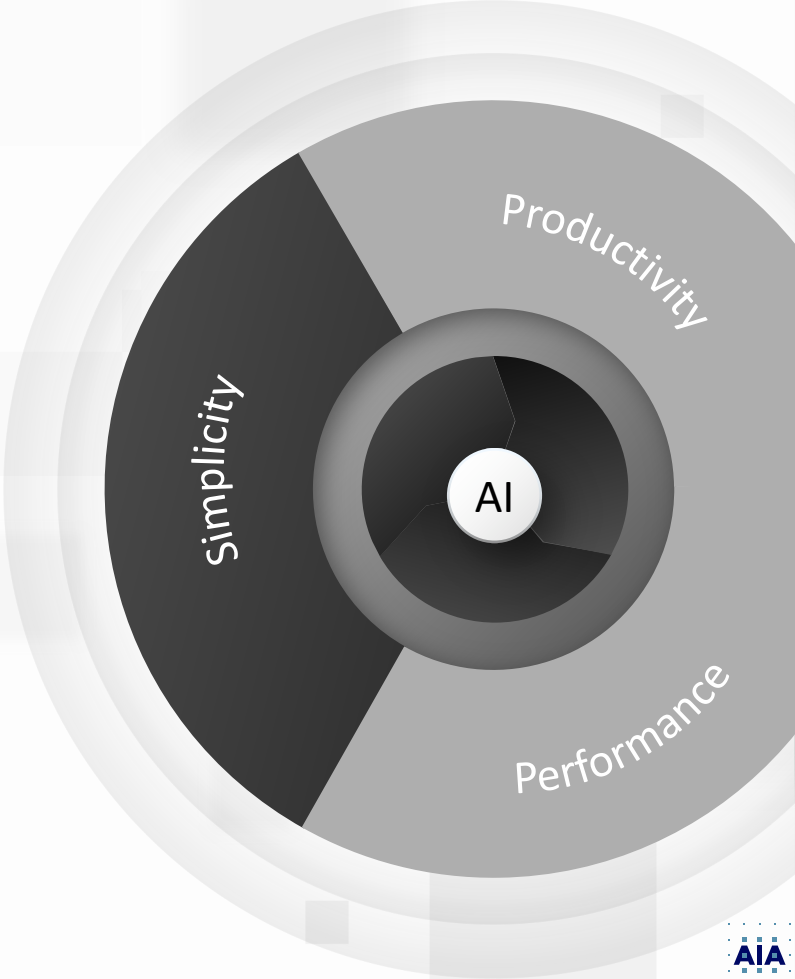
on existing Intel environment

Build & Scale Quickly

With Optimized, Ready-to deploy solutions

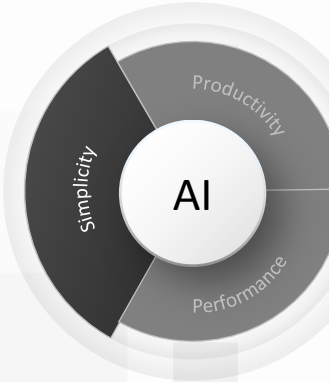
Tangible Results

Without unnecessary complexity and specialized hardware



new

intel + accenture



Industry Use Case AI Reference Kits

Open-Source and Prebuilt AI with Meaningful Enterprise Contexts

30+

AI Reference Kits

- Predictive Asset Analytics
- Intelligent Document Indexing
- Customer Care Chatbot
- Quality Visual Inspection
- Purchase Prediction
- Fraud Detection
- Hyper-Personal Targeting
- Customer Churn Prediction
- Customer Lifetime Valuation
- Product Recomm Systems

For Companies that want to…

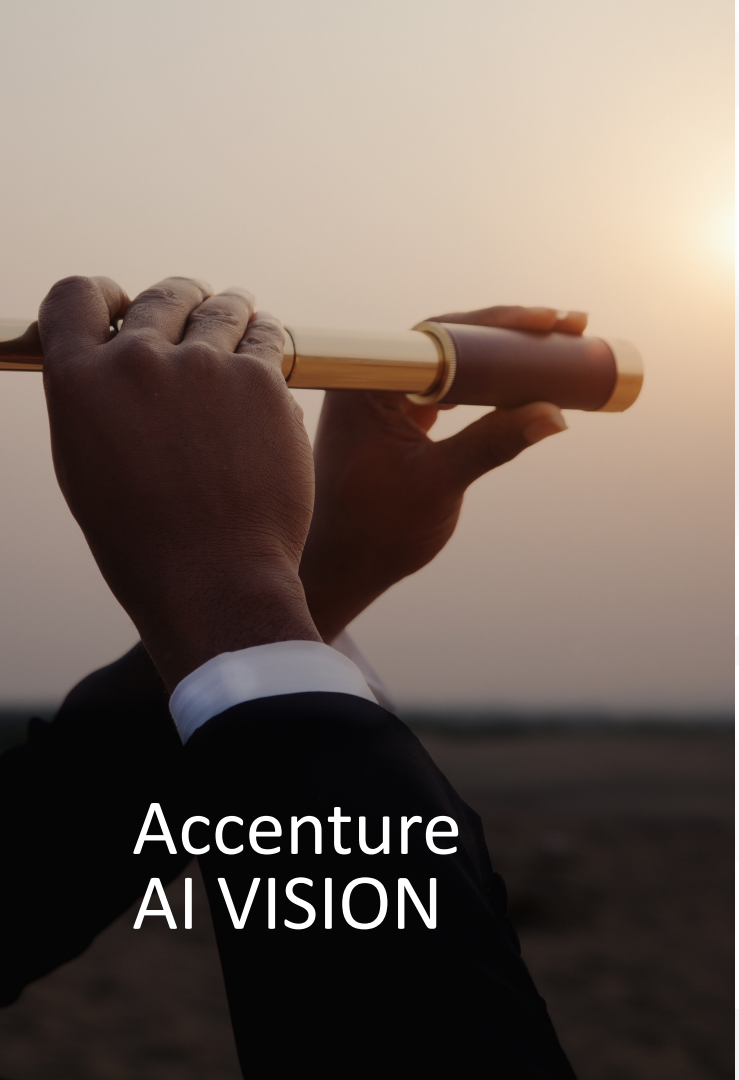
Boost Existing AI/ML Solution | Introduce Greenfield AI | Change AI Solution Strategy

AIA

intel

Jean-Luc Chatelain, CTO, Accenture Applied Intelligence

At Accenture, Jean-Luc Chatelain leads innovation and technology strategy focused on artificial intelligence, advanced analytics and cognitive automation. His team leads the development and roadmap of their flagship system of intelligence solution, the Applied Intelligence (AIP+).



Accenture AI VISION

All enterprises today must become digital

AI is the core engine that support this data driven transformation

Our AI vision is to enable our clients to deploy pragmatic AI to securely, responsibly augment the human business decision process to achieve best outcomes

oneAPI & AI

Speed is the #1 KPI for business

Enables you to deliver that speed at scale

Helps you

- Hide plumbing complexity
- Simplify access speed
- Deploy everywhere

Intel + Accenture deliver acceleration with Industry Reference Kits

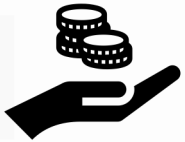
First Set of 4 Toolkits Releasing Today!

https://github.com/oneapi-src

Visual Quality Inspection	Predictive Asset Maintenance
Intelligent Document Indexing	Customer Care Chatbot

Reduced Time to Deployment

Improved Training and Inference Performance



Lower TCO

Inside Our AI Reference Kits

intel + accenture

Millions of Developers

General Developers | Data Scientists | Edge & IoT Developers | AI Developers | DL & Inference Developers | ...

AI Reference Kits

Predictive Asset Analytics | Intelligent Document Indexing | Customer Care Chatbot | Quality Visual Inspection | ...

Solution Brief | **Developer Guide** | **Code & Data Repository** | **Compute Architecture** | **Optimization results**

Leading Industry Frameworks

python | pandas | dmlc XGBoost | scikit-learn | PyTorch | daal4py | ...

\+

Intel Optimizations & Tools

oneAPI (oneMKL, oneDAL, oneDNN, oneCCL) | ... | OpenVINO | SIGOPT | Intel Neural Compressor | AI Analytics Toolkit | ...

Download the reference kits from GitHub today!

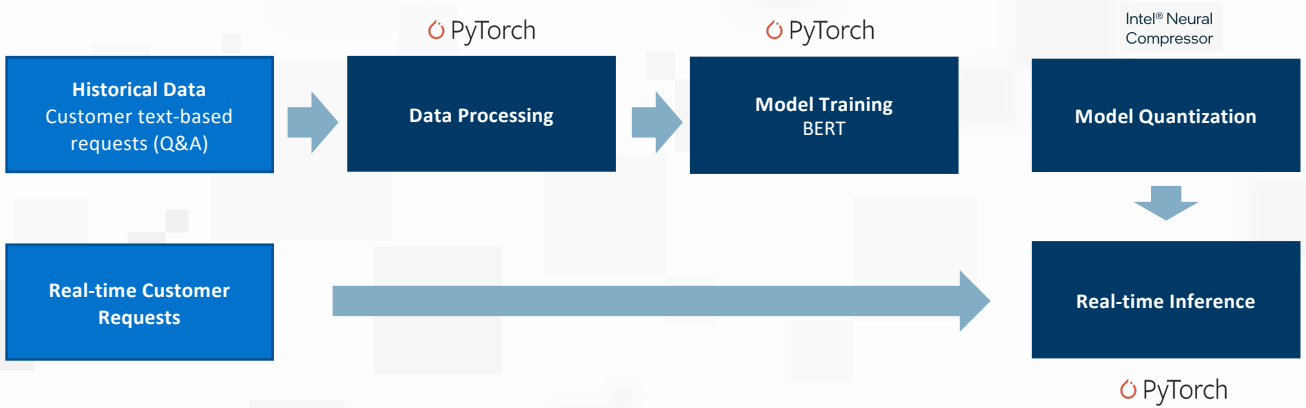
https://github.com/oneapi-src

* Other names and brands may be claimed as the property of others

Customer Care Agent Intent Enablement

To enable virtual agents to understand user intents in automated conversations using Natural Language Understanding (NLU). Customer care organizations need to reduce operational costs and yet offer a more natural and engaging conversational experience.

Experiment: build the NLU ML pipeline using airline travel dataset and PyTorch/BERT and use Intel Extension for PyTorch and Intel Neural Compressor boost inference times and reduce training cycles.



Results

- Up to **61%** reduction in customer interaction time
- Up to **16%** faster training cycles

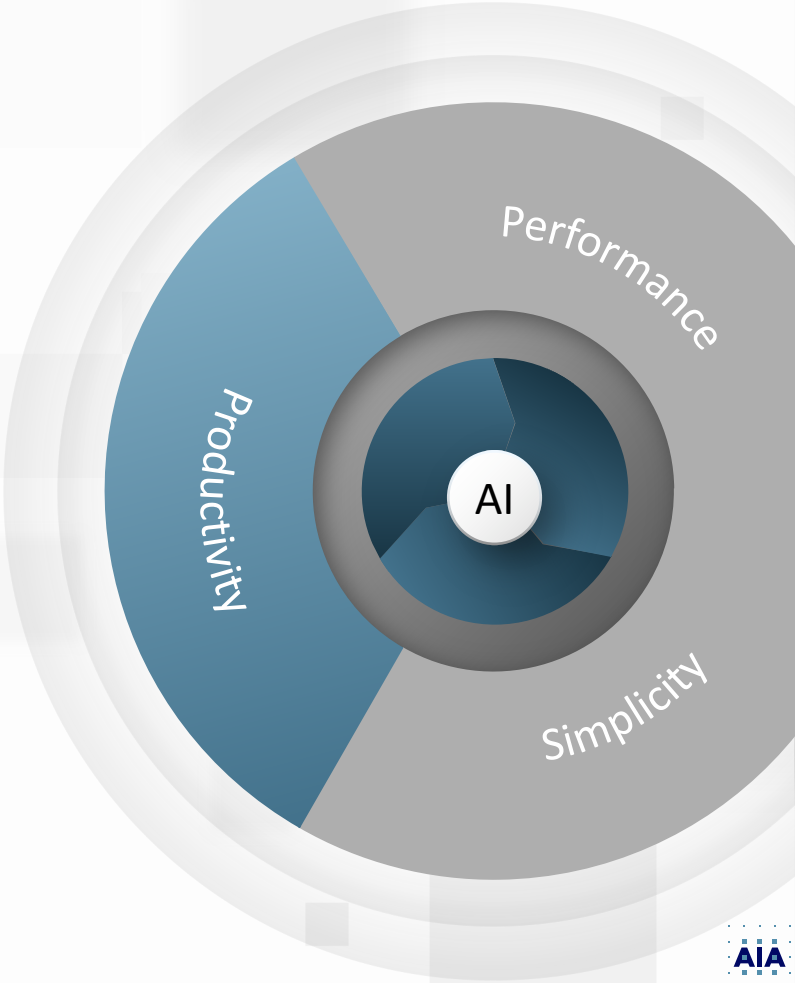
Productivity

Build Faster and Smarter

End-to-End Software Accelerators

Familiar Data Science Tools

Optimized libraries and frameworks



One line of code

Unlocks End-to-End Performance gains

```
import modin.pandas as pd
```

pandas MODIN

Engineer Data

90x

performance

```
from sklearnex import patch_sklearn
patch_sklearn()
```

Intel Extension for scikit-Learn

Create ML and DL Models

38x

performance

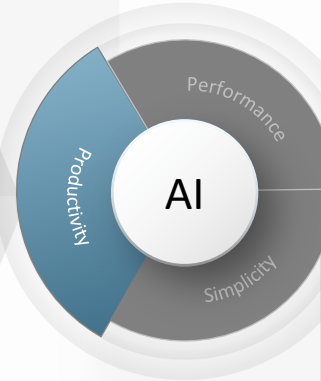
No Change Needed

TensorFlow

Deploy

3x

throughput



Source: https://techdecoded.intel.io/resources/one-line-code-changes-to-boost-pandas-scikit-learn-and-tensorflow-performance/#gs.bzkn2n

new

Hugging Face + intel

Democratizing Accelerated Transformers

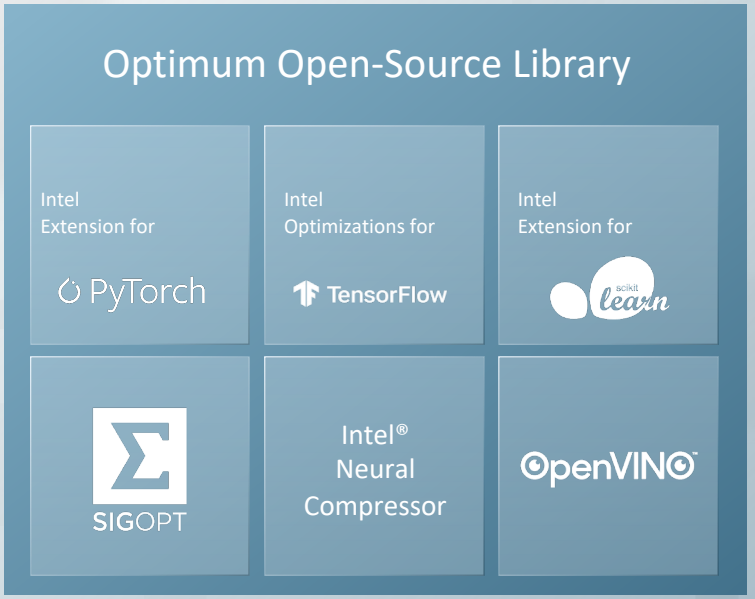
on Intel Platforms

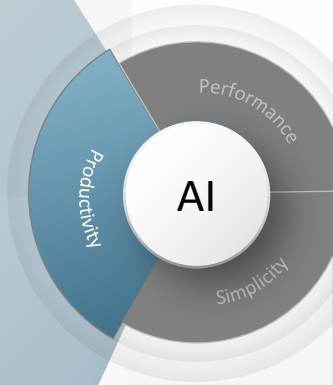
Inference Optimization Process from Days to Hours with Up to 4x Performance Speedup

Distributed Training through efficient compute scaling

intel XEON

habana An Intel Company





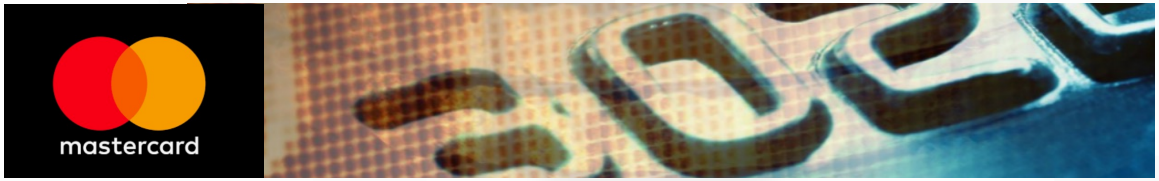
AIA

* Other names and brands may be claimed as the property of others

intel

AI & Big Data: Mastercard Recommender AI Service

with Intel's BigDL for Accelerated End-to-End Development



+

BigDL

Intel's Open-Source Big Data AI Project

Domain Specific Solutions

- PPML Privacy Preserving Machine Learning
- Chronos Time Series
- Friesian Recommendation System

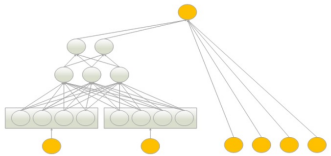
End-to-End Distributed Pipelines

- Orca: Distributed TensorFlow/PyTorch/OpenVINO/Ray on Apache Spark
- DLlib: Distributed Deep Learning Library for Apache Spark

Laptop | K8s | Apache Hadoop/Spark | Cloud

Data → Model → Deployment

Large History of Transaction Records *(>100 Billion)*



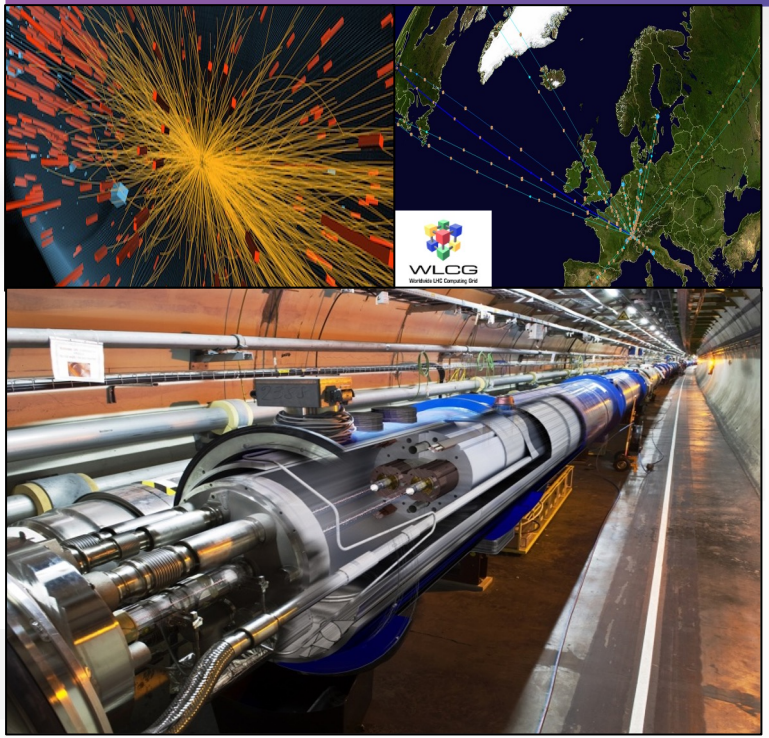
Personalized Recommendations

Improving the customer experience, new product campaign performance, and better personalized recommendation accuracy by using the deep learning-based neural recommendation models

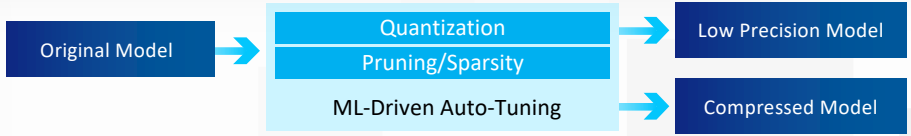
Source: https://software.intel.com/content/www/us/en/develop/articles/deep-learning-with-analytic-zoo-optimizes-mastercard-recommender-ai-service.html

AI & High Performance Computing: CERN Large Hadron Collider (LHC)
with Intel Neural Compressor for 10X Productivity

High Performance AI Inferencing made Easy



Intel Neural Compressor



- Simulations are essential to all high energy physics experiments
- Complex physics and geometry modeling requires >50% power of worldwide LHC Computing Grid (WLCG)

- Deep Generative Adversarial Networks (GAN) models can replace Monte Carlo simulation to significantly save computation needs and ensure computing requirements remain manageable

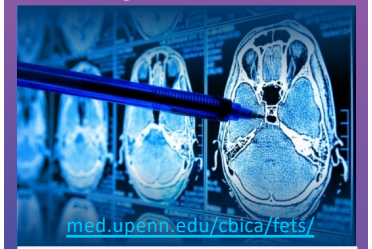
- Faster inference via Intel Neural Compressor allows GAN models to generate data on the fly delivering more timely simulations

Sources: Conference paper: 10th International Conference on Pattern Recognition Applications and Methods
Blog: https://www.nextplatform.com/2021/02/01/cern-uses-dlboost-oneapi-to-juice-inference-without-accuracy-loss

AI & Security: UPenn Federated Tumor Segmentation Initiative

featuring Intel Federated Learning (OpenFL)

Finding tumors from MRIs



nature research SCIENTIFIC REPORTS

OPEN **Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data**

Micah J. Sheller[1], Brandon Edwards[1], G. Anthony Reina[1], Jason Martin[1], Sarthak Pati[2,3], Aikaterini Kotrotsou[2,4], Mikhail Milchenko[5], Weilin Xu[6], Daniel Marcus[2,4], Rivka R. Colen[2,4,7] & Spyridon Bakas[2,3,4,7]

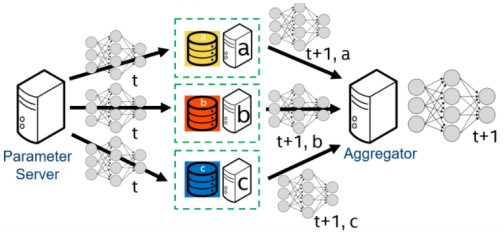
Perelman School of Medicine
University of Pennsylvania

How much better does each institution do when training on the full data vs. just their own data?

17% BETTER on the hold-out BraTS data

2.6% BETTER on their own validation data

Intel Federated Learning (OpenFL)
and Intel SGX



Secure Distributed Machine Learning

- Machine Learning collaboration without sharing sensitive data
  (such as patient records, financial data, IP)
- Intel Software Guard Extensions protect data in use with HW-hardened enclaves

TensorFlow PyTorch Keras

https://pypi.org/project/openfl/

Source: med.upenn.edu/cbica/fets/. Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex

intel AI Builders

AI solutions accelerated by Intel

Vertical Partners

Retail

Healthcare

BFSI

Transportation

Media, Travel & Entertainment

Sec & Govt

SW Tools & Services

Prof. Services

Agri

Telecom

Horizontal Partners

BI & Analytics

Vision

Conv. Bots

NLU/NLP

AI Tools & Consulting

AI PaaS

Big Data

Intel® Disruptor Initiative

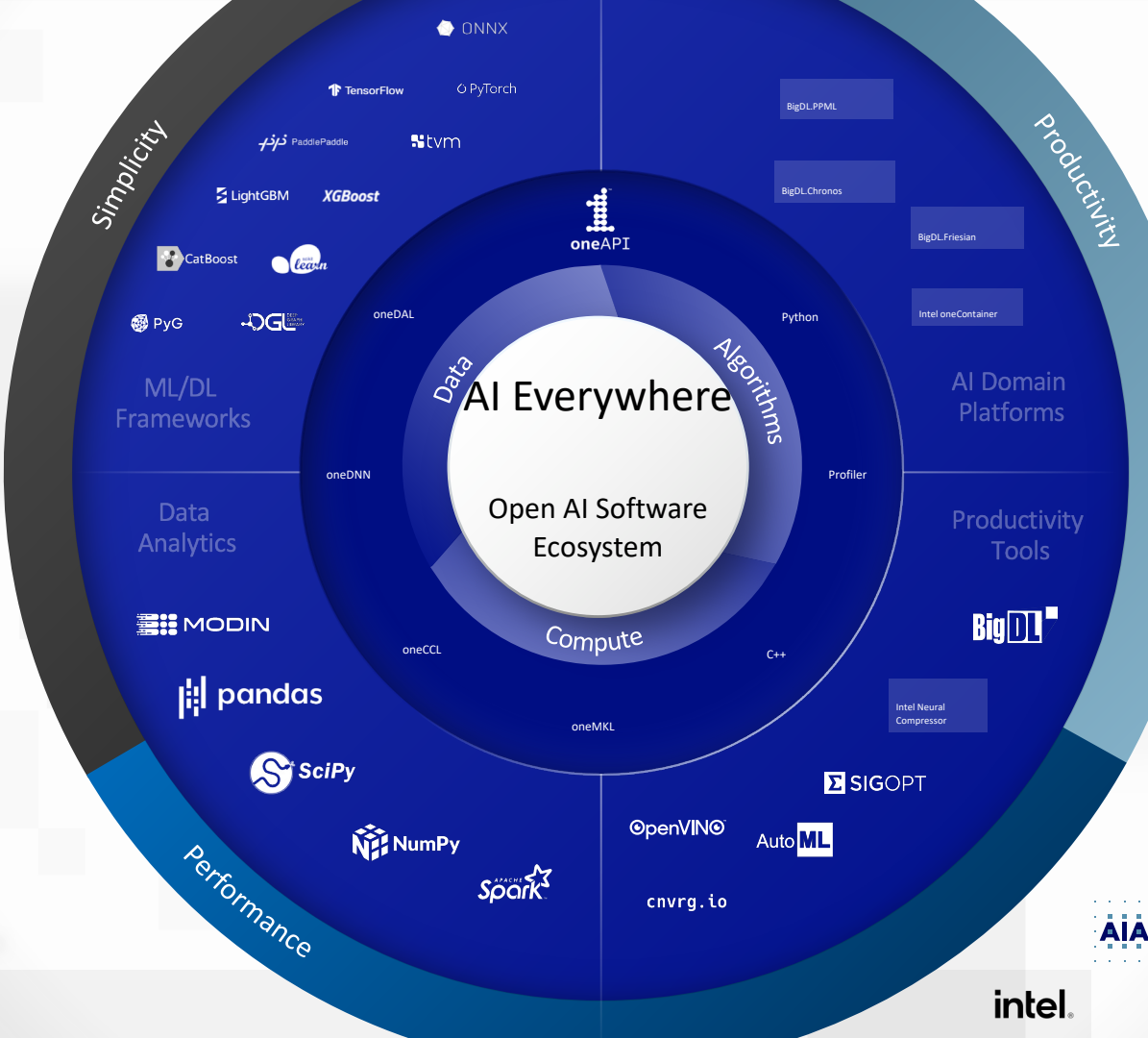
Strategic Partnerships with Innovative AI and Analytics Companies to Provide Best-In-Class Joint Solutions

+ Many Additional Participating Partners

www.intel.com/disruptor


Let’s work together to bring AI Everywhere

Visit **developer.intel.com/ai** for more info



* Other names and brands may be claimed as the property of others

Thank You