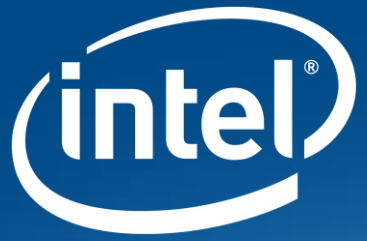


ACCELERATE DEEP LEARNING INFERENCE USING INTEL® TECHNOLOGIES

INTRODUCTION: SMART VIDEO

INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT 2020.R4 VERSION

July 2020

NOTICES AND DISCLAIMER

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

OPTIMIZATION NOTICE

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer to learn more.

The benchmark results reported in this deck may need to be revised as additional testing is conducted. The results depend on the specific platform configurations and workloads utilized in the testing, and may not be applicable to any particular user's components, computer system or workloads. The results are not necessarily representative of other benchmarks and other benchmark results may show greater or lesser impact from mitigations.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Other names and brands may be claimed as the property of others. Any third-party information referenced on this document is provided for information only. Intel does not endorse any specific third-party product or entity mentioned on this document. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries. Copyright Intel Corporation.

Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

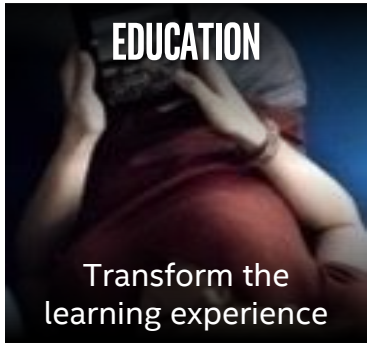
Notice revision #20110804

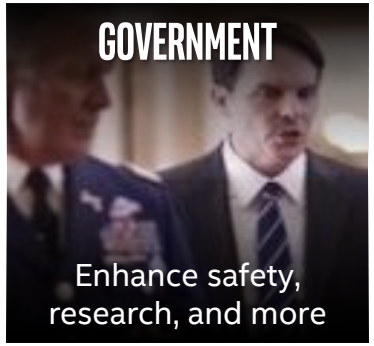
AI CHANGING AND ENABLING EVERY INDUSTRY

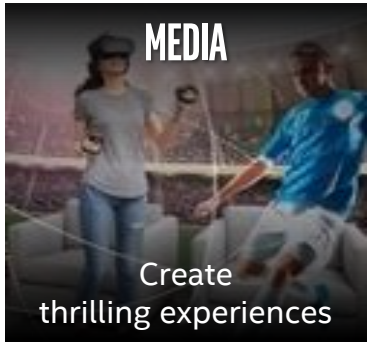
AI software market is projected to reach USD 126.0 billion in annual worldwide revenue by 2025[3]

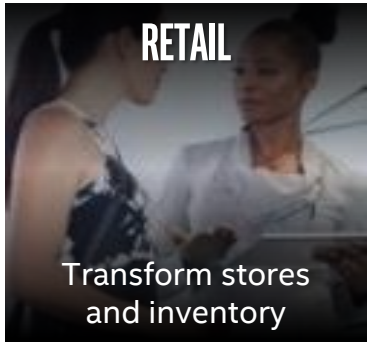
Deep learning software revenue is estimated to grow to USD 67.2 billion by 2025[4]

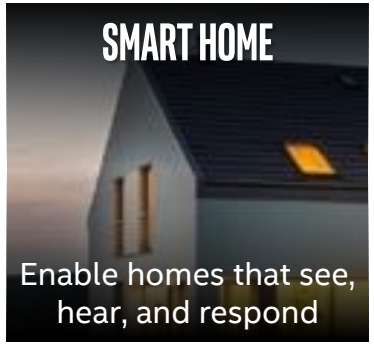
Global deep learning chip market is expected to reach USD 29.4 billion by 2025[5]

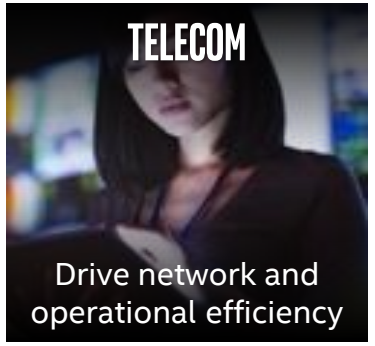


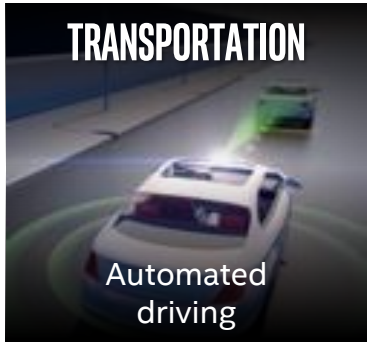












3. Tractica, Artificial Intelligence Software Market, 2020

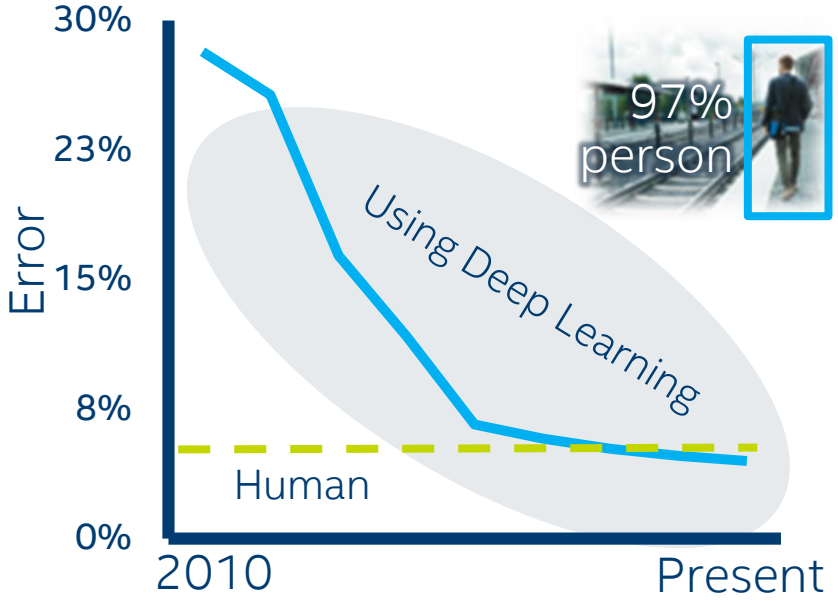
4. Tractica, deep learning research, 2018

5. AlliedMarketResearch, Deep Learning Chip Market, 2018

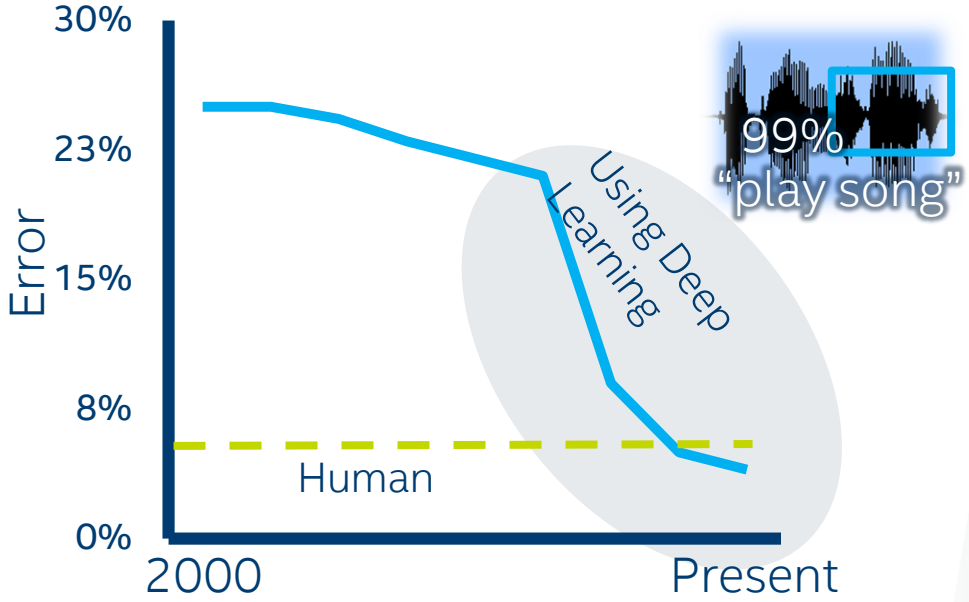
DEEP LEARNING BREAKTHROUGHS AND OPPORTUNITIES

Machines able to meet or exceed human image and speech recognition

Image Recognition



Speech Recognition



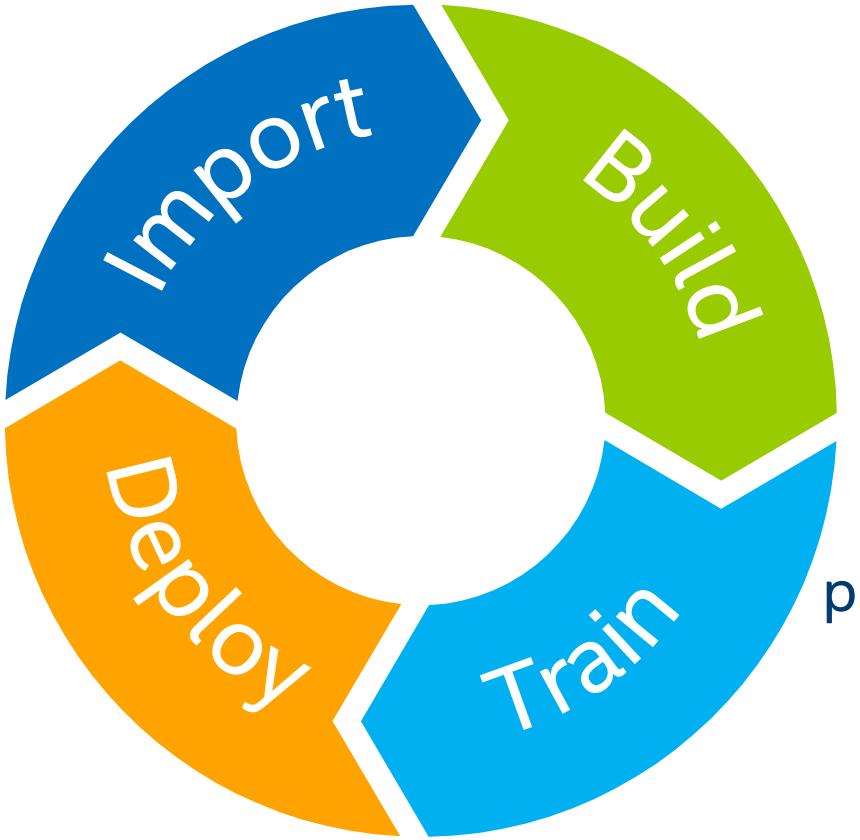
ADDITIONAL ECONOMIC IMPACT DRIVEN BY AI
$13 TRILLION IN 2030

Source: ILSVRC ImageNet winning entry classification error rate each year 2010-2016 (Left), https://www.microsoft.com/en-us/research/blog/microsoft-researchers-achieve-new-conversational-speech-recognition-milestone/ (Right)
Source: https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning

DEEP LEARNING DEVELOPMENT CYCLE

Data acquisition and organization

Create models



Integrate trained models with application code

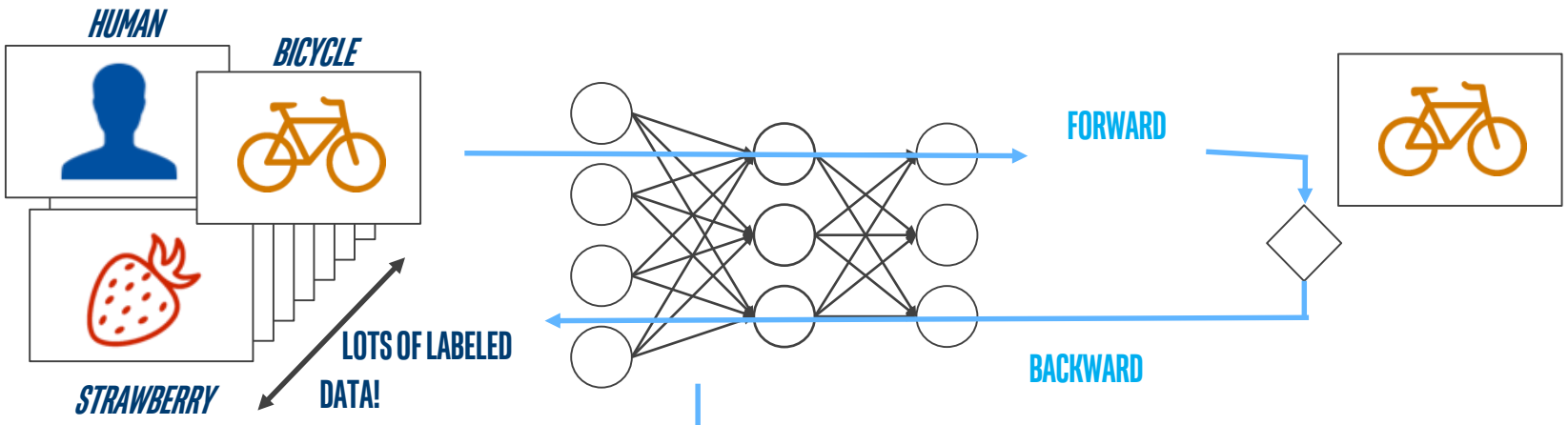
Adjust models to meet performance and accuracy objectives

Intel® Distribution OpenVINO™ Toolkit Provides Deployment from Intel® Edge to Cloud

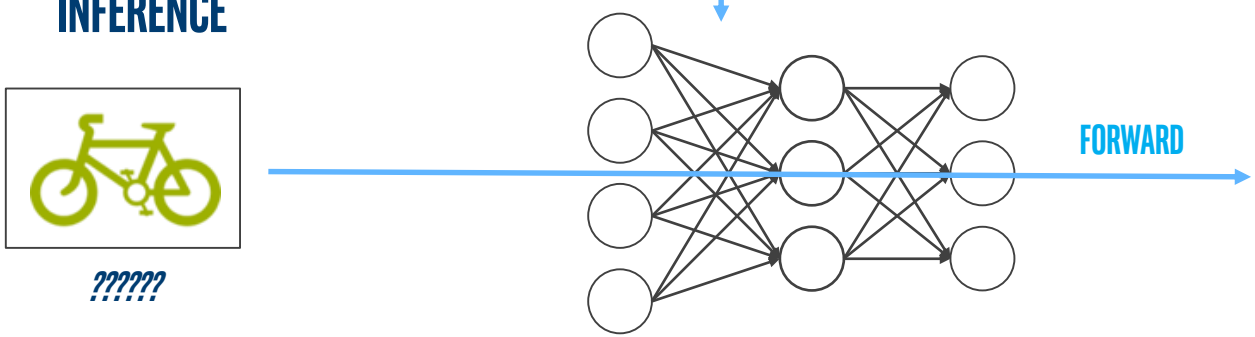
intel

DEEP LEARNING: TRAINING VS. INFERENCE

TRAINING

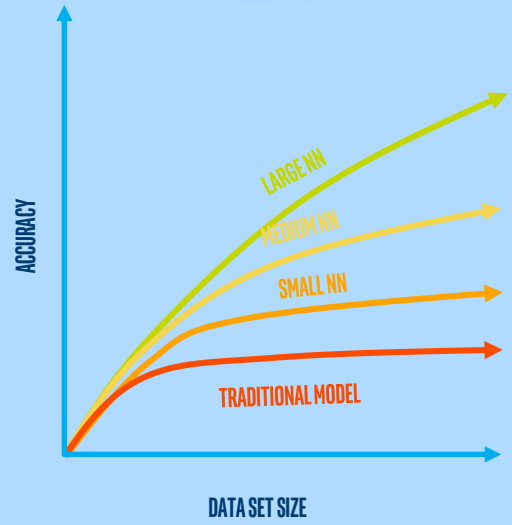


INFERENCE



DID YOU KNOW?

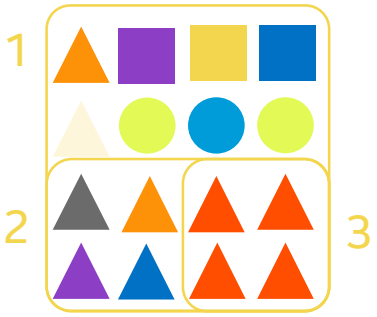
Training requires a very large data set and deep neural network (many layers) to achieve the highest accuracy in most cases



AI COMPUTE CONSIDERATIONS

How do you determine the right computing for your AI needs?

WORKLOADS

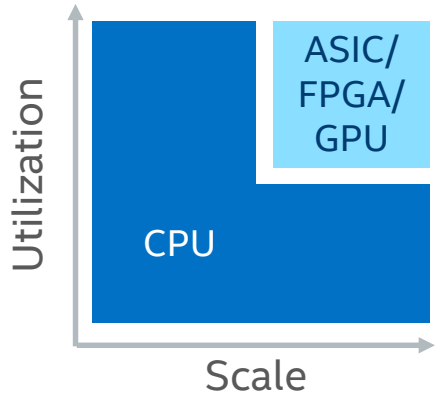


What is my workload profile?

REQUIREMENTS

What are my use case requirements?

DEMAND

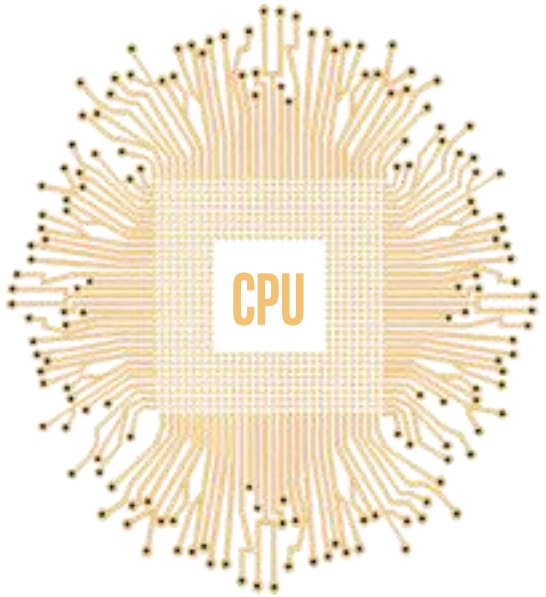


How prevalent is AI in my environment?

OpenVINO

WHY INTEL AI COMPUTE?

MAXIMIZE



Get the most out of the foundation for AI from the CPU leader

OPTIMIZE

Choose the right compute for you from the one with all the options

SIMPLIFY

OPTIMIZED SW
DATA PIPELINE
ANALYTICS & AI
SUPPORT
MOVE/STORE

Reduce “moving parts” by building on an optimized AI platform

LEAD

Lead your industry by aligning with the builder of next-gen AI solutions

INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT

Tool Suite for High-Performance, Deep Learning Inference

Fast, accurate real-world results using high-performance, AI and computer vision inference deployed into production across Intel® architecture from edge to cloud

High-Performance,
Deep Learning Inference

Streamlined Development,
Ease of Use

Write Once,
Deploy Anywhere

USING THE INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT

Advanced capabilities to streamline deep learning deployments

1. BUILD

Trained Model

TensorFlow Caffe
KALDI mxnet
ONNX

Open Model Zoo
100+ open sourced and optimized pre-trained models; 80+ supported public models

2. OPTIMIZE

Model Optimizer
Converts and optimizes trained model using a supported framework

Read, Load, Infer

IR Data

Intermediate Representation (.xml, .bin)

Post-Training Optimization Tool

Deep Learning Workbench

Deep Learning Streamer

OpenCV **OpenCL™**

Code Samples & Demos
(e.g. Benchmark app, Accuracy Checker, Model Downloader)

3. DEPLOY

Inference Engine
Common API that abstracts low-level programming for each hardware

Deployment Manager

CPU Plugin
GPU Plugin
GNA Plugin
Myriad Plugin
For Intel® NCS2 & NCS
HDDL Plugin
FGPA Plugin