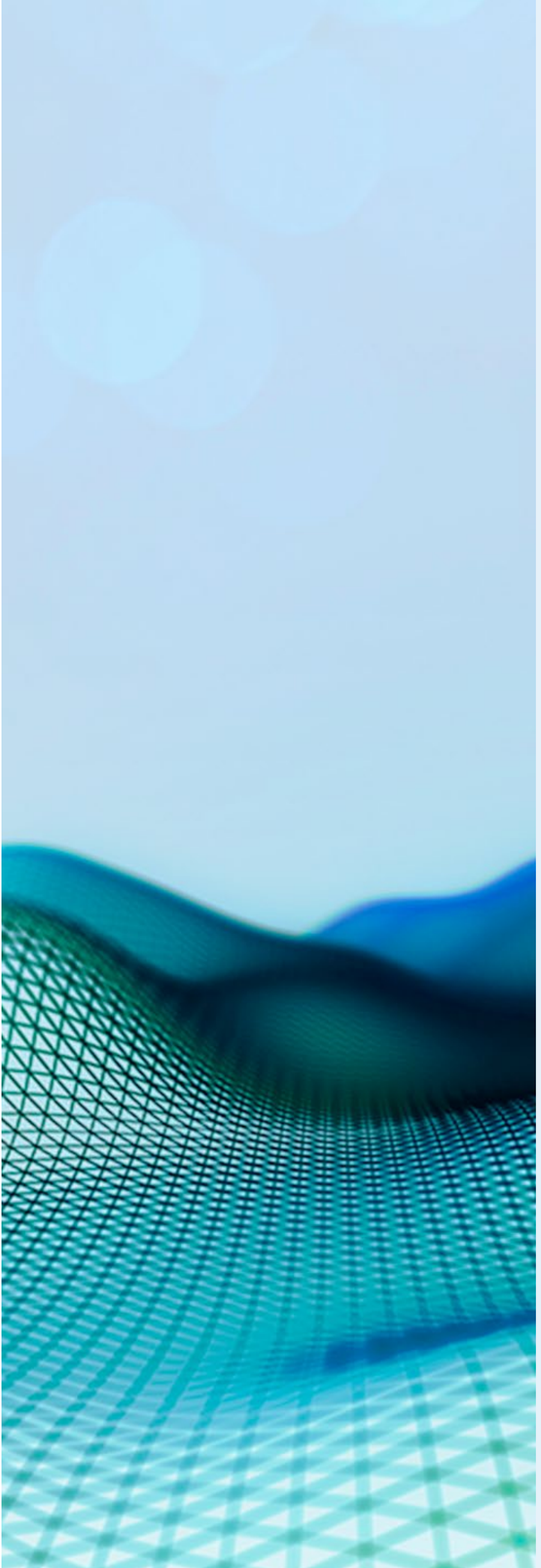
EXPLORING THE PATH FROM AI THEORY TO REAL BUSINESS VALUE

Infrastructure considerations for IT leaders



CONTENTS

Introduction: Deep learning is coming of age 2

Determining AI readiness 4

Developing and deploying data governance and security policies 5

Infrastructure strategies for the shift to deep learning inference at scale 6

The magnifying impact of optimized software 9

Next steps: Breaking barriers between model and reality 11

Further reading 12

Skip to discover the benefits of the latest generation Intel® Xeon® Scalable processors and Intel® Deep Learning Boost

1. DEEP LEARNING IS COMING OF AGE... AND IT'S HAPPENING FAST

By 2020, deep learning will have reached a fundamentally different stage of maturity. Deployment and adoption will no longer be confined to experimentation, becoming a core part of day-to-day business operations across most fields of research and industries.

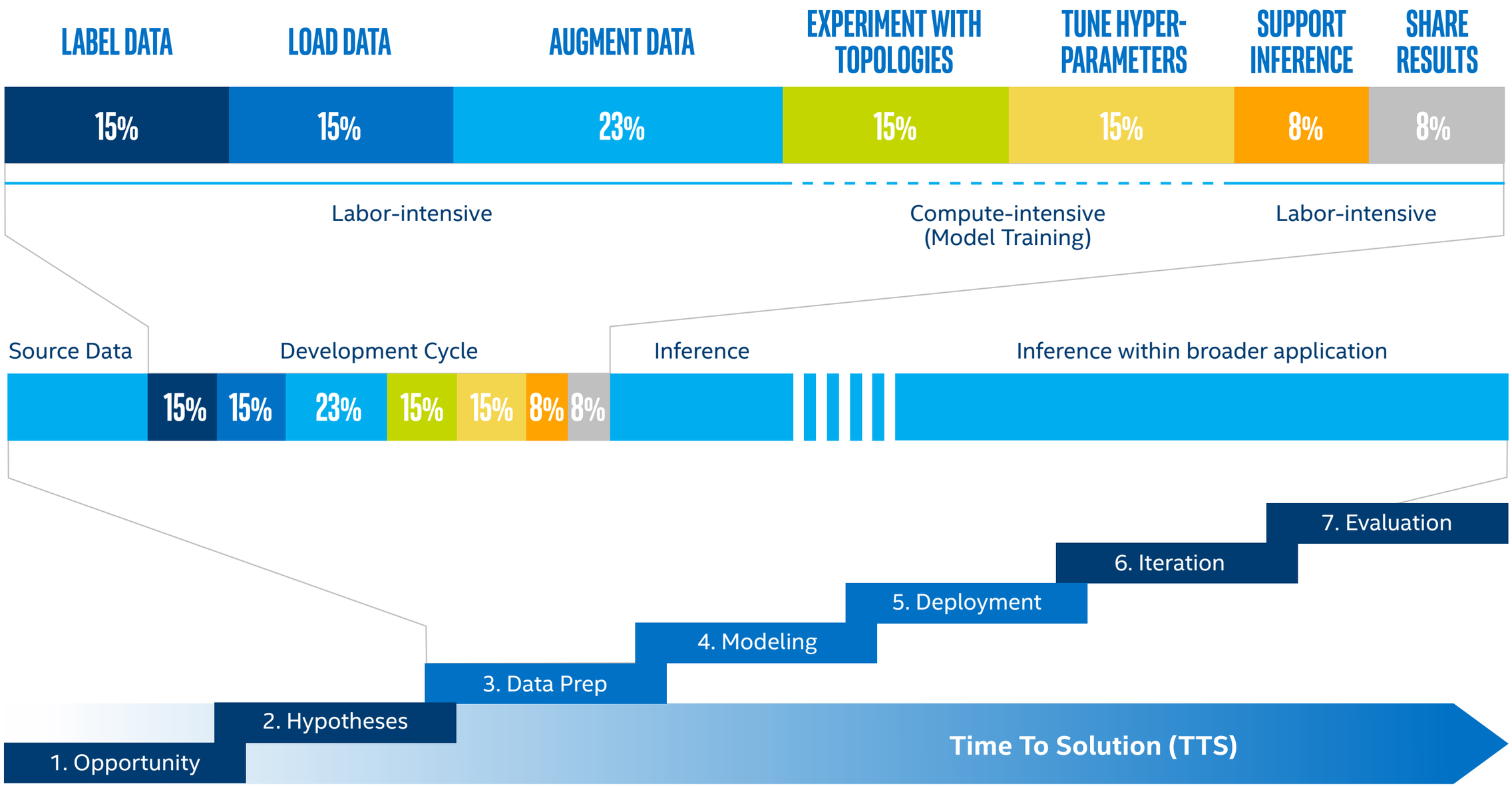
Why? Because advancements in the speed and accuracy of the hardware and software that underpin deep learning workloads have made it both viable and cost-effective. Much of this added value will be generated by deep learning inference – that is, using a model to infer something about data it has never seen before. Models can be deployed in the cloud or data center, but more and more we will see them on end devices like cameras and phones.

Intel predicts that there will be a shift in the ratio between cycles of inference and training from 1:1 in the early days of deep learning, to well over 5:1 by 2020[1]. Intel calls this the shift to 'inference at scale' and, with inference also taking up almost 80 percent of artificial intelligence (AI) workflows (Figure 1, Page 3), it follows that the path to true AI readiness starts with selecting hardware architectures that are well-suited to this task.

However, as the AI space is becoming increasingly complex, a one-size-fits-all solution cannot address the unique constraints of each environment across the AI spectrum. In this context, critical hardware considerations include availability, ease of use, and operational expense. What type of infrastructure do you use for your edge devices, workstations or servers today? Do you want to deal with the complexities of multiple architectures?

Exploring these challenges is the subject of this paper.

FIGURE 1: A TYPICAL AI WORKFLOW



2. DETERMINING AI READINESS

Understanding where your organization sits in terms of AI readiness is critical to prioritizing actions and smoothing the path from experimenting with AI to real-world deployments. Organizations can be grouped into three categories of AI readiness: foundational, operational or transformational, and progressing to the next stage or achieving ongoing success depends on having the right elements in place across skills and resources, infrastructure and technology, processes and models.

Among the defining characteristics of operational and transformational AI-ready enterprises is – to varying degrees – their ability to support better decision making or automate business processes/responses with AI through inference at scale.

At the foundational stage, enterprises should be prioritizing developing and deploying proof of concepts (PoCs) to establish and build the infrastructure, skills and executive buy-in required to scale with AI.

These topics are dealt with in detail in our white papers Five Steps to an AI Proof of Concept and The AI Readiness Model.

.

Stage of readiness	Top IT priority	Readiness indicators
Foundational	Once data strategy and target use cases have been identified, ensure appropriate infrastructure and interfaces for an AI proof of concept (PoC).	• AI starts with data. Ensure critical data sources are both available and accessible. • Capacity and capabilities of data center facilities to support a PoC with the massively scalable processing required for AI. • Might be exploring cloud-based services for testing use cases, due to low-entry point and pay-per-use services. • Considering how open source and commercial AI software packages will integrate with other tools for data management and visualization.
Operational	Establish suitable management and governance mechanisms to develop AI solutions sustainably.	• Best practice models such as DevOps are in place to help respond to quickly changing requirements. • If necessary, use-case specific skills are being built in-house to expand beyond an initial PoC. • The security of data, infrastructure and algorithms is being prioritized to reduce risks from corrupted data inputs, model tampering and unauthorized access to the resulting insights.
Transformational	Prioritize the ability of the enterprise to maximize the value it gets from AI – does it support better decision making at a senior level, or automate business processes/drive automated responses?	• Internal stakeholders are fully engaged on work to build an organizational structure that identifies AI-driven business opportunities that could improve processes or engage with customers. • A clear, costed business case is in place for what success looks like with AI. • Focus on achieving acceptance for AI to ensure the solution is adapted to business needs right through to the daily activities of front-line staff, and the people impacted.

3. DEVELOPING AND DEPLOYING DATA GOVERNANCE AND SECURITY POLICIES

Security in the context of AI can be viewed from two different perspectives. First, it's critical to secure the AI itself in the form of algorithms, parameters, and data. Second, AI has great potential to be used for the detection of advanced exploits.

The relationship between AI, security and governance is complex and many-faceted. In earlier stages of an organization's journey, governance questions will not be any different to other data-centric IT projects – can the project deliver, is customer privacy protected and so on. As its use is extended, AI brings additional ramifications: for example, in predictive planning and maintenance, how much human involvement is required in purchasing decisions (if any)?

Furthermore, where AI-driven decisions impact people's lives, the reputational risks associated with inaccurate or biased AI outputs will easily dwarf those of consumer data breaches seen today.

Threats here might take the form of 'model poisoning', when a model is biased via outlier inputs or the insertion of back doors in unused parameters. Hardware-based trusted execution environments (TEEs) should be considered to enable trusted models to be built between the end point and the aggregator where updates are protected, minimizing the risk of model poisoning.

4. INFRASTRUCTURE STRATEGIES FOR THE SHIFT TO DEEP LEARNING INFERENCE AT SCALE

AI is a complicated mix of getting raw data ready to use, creating, securing and then fine-tuning models, and deploying solutions at scale in the real world, where they must continually be refined and sometimes operate outside the relative safety of an on-premise data center, often with severe power and space constraints.

This means that architecting an on-premise and/or hybrid cloud solution for AI requires a brand-new approach, including creating flexible data centers capable of pooling huge resources of on-demand compute. It also requires storage, and connectivity, and – perhaps – networks that are able to move data at high speeds with minimum latency.

Further difficulties arise from the fact that AI is not one size fits all. However, for IT leaders considering which infrastructure strategy is right for their organization, their options broadly fall into four categories (these are explored in more detail in our paper Select the Best Infrastructure Strategy to Support Your AI Solution).

Repurpose existing hardware

- **What:** Enterprises nearer the start of their AI journeys often look to use 'spare cycles' in their data centers to run AI workloads, or they develop solutions based on a single, 'spare' server or workstation node, or a small cluster.
- **The benefits:** Makes use of existing hardware resources, enables research to focus on a tightly bound environment and reinforces the benefit of AI as 'augmenting' existing capabilities.
- **The disadvantages:** Won't always integrate easily with broader solutions, and unless the hardware is aligned to the need, overheads can be incurred converting or redirecting less appropriate resources.

Build a broader platform

- **What:** Organizations with more AI experience may want to adopt a broader infrastructure solution that supports more general AI workloads.
- **The benefits:** A platform-based approach offers a single point of configuration and unique deployment target. It is also easier to manage organizationally and provides a focus for building expertise.
- **The disadvantages:** The initial build could be seen as more complex, require hard-to-come-by in-house skills and create greater risk if the architecture proves too small or too big for the actual need.

Buy a one-off solution

- **What:** A custom solution procured to meet a well-defined use case.
- **The benefits:** Potentially ensures faster deployment, efficiency and performance because the solution has been designed with a predetermined use case in mind.
- **The disadvantages:** A one-off approach can result in multiple AI 'silos' which may need to be managed in parallel. This can prove more expensive overall.

Outsource solution delivery

- **What:** Organizations at various stages of their AI journeys may look to use third-party resources (including cloud-based options) to deliver a full-stack solution or work with existing resources.
- **The benefits:** Capabilities may be available 'off the shelf', minimizing deployment and configuration issues. External services can be used to augment and test new solutions before they are brought in-house.
- **The disadvantages:** Additional cost and inefficiency with managing an outsourcer. The resulting infrastructure architectures may also create data bottlenecks, depending on where data is sourced – for example, the organization may need to upload data from its internal systems into the cloud.

BEYOND THE DATA CENTER: TAKING AI TO THE NETWORK EDGE

Most AI today happens in data centers, or the cloud. As billions of devices get connected to the internet and our need for real-time intelligence grows, more AI inference will move to the edge of the network to avoid the need for data transfer to the cloud.

One of the most important ways we'll be able to safely move AI to the edge is with federated learning. This process enables edge devices to collaboratively learn a shared prediction model while keeping all training data on the device, decoupling the ability to enhance models from the need to store data in the cloud. This also enables devices to be used for model training. The device downloads the latest model, improves it by learning from data on the device, and then summarizes the changes as a small, focused update. Only this update is sent to the cloud, using encrypted communication, where it is immediately averaged with other user updates to improve the shared model. All training data remains on the edge device, and no individual updates are stored in the cloud.

Intel offers hardware and software tools to help enterprises deploying AI on edge devices, including:

- The Intel® Distribution of OpenVINO™ toolkit is software designed for deploying neural networks for video across multiple types of Intel® hardware, from data centers to devices at the edge.

FIND OUT MORE >

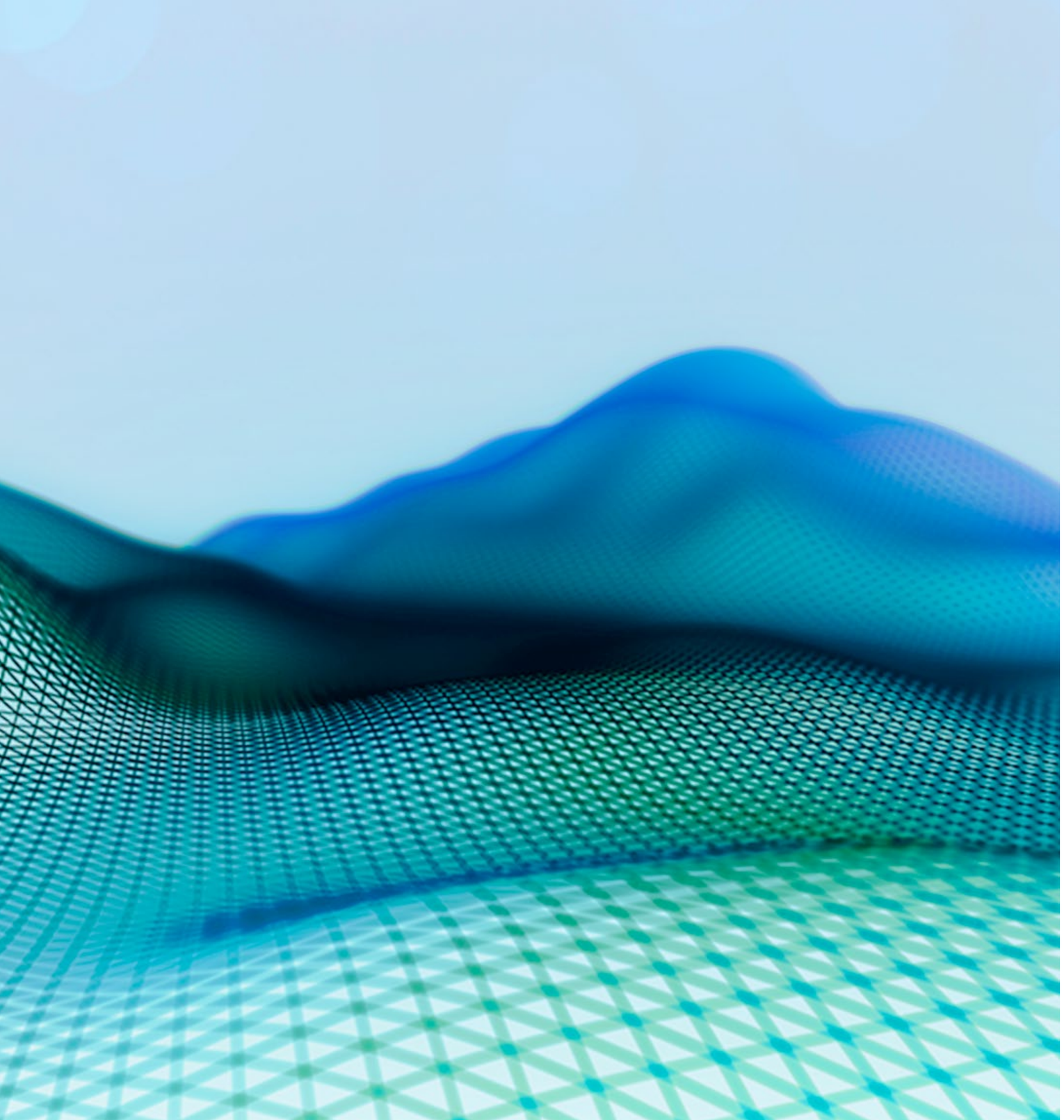
- Intel® Movidius™ Vision Processing Units (VPU) push the boundaries of what's possible with AI at the edge with extreme low-power deep neural network (DNN) inferencing on the device.

FIND OUT MORE >

By their nature, many AI use cases require systems to perform inference in real-time, rather than offline or in batch mode. In addition, models may need to be re-trained and updated over time. However, memory, power and data movement can all create bottlenecks that can drive down utilization and create more cost.

Intel offers hardware and storage solutions for enterprises that are navigating these challenges:

- Many inference workloads in the world's data centers today already run on **Intel® Xeon®Scalable processors**. The Intel Xeon Scalable processor is enhanced specifically to run high-performance AI applications alongside the data center and cloud applications they already run.
- The latest 2nd Generation Intel Xeon Scalable processors deliver greater AI acceleration through new **Intel® Deep Learning Boost**, which includes embedded instructions (Vector Neural Network Instructions, or VNNIs) that speed up dense computations characteristic of convolutional neural networks (CNNs) and other deep neural networks (DNNs). The result is more efficient inference acceleration for deep learning use cases such as image classification, speech recognition, language translation and object detection. In tests, 2nd Generation Intel® Xeon® Platinum 8280 processors with Intel® Deep Learning Boost improved inference throughput by an incredible 14x[2] compared to two-socket Intel® Xeon® Platinum 8180 processors. The Intel® Xeon® Platinum 9282 processor delivered further improvement, with inference performance on image classification boosted by an outstanding 30x[3] compared to the same processor.
- With support for Intel® Optane™ DC persistent memory built into the 2nd Generation Intel Xeon Scalable processor, more memory is closer to the CPU, allowing data to be sustained even throughout power cycles or system maintenance. As it is persistent, this memory also supports fast restart times, helping organizations to patch, upgrade and secure their infrastructures while maximizing uptime and service delivery. Unplanned outages can also be less impactful on the business, as time-to-recovery is reduced. Low latency allows enterprises to activate larger working data sets in-memory, meaning it's possible to extract more value from significantly larger data sets and to reach powerful business decisions quickly.
- **Intel® Optane™ SSDs** help enterprises break through storage bottlenecks by allowing data centers to deploy bigger data sets more affordably, accelerate applications and take advantage of the enterprise-level insights that come from working with larger memory pools. This means Intel® Optane™ technology can add value to both the training and inference aspects of deep learning. It optimizes batch training, enabling data sets to massively expand with the cost and capacity benefits of storage. At the inference stage, accelerated real-time inference boosts streaming AI capabilities, while batch inference is streamlined for more efficient analysis and insight gathering.



- **Intel® Select Solutions for BigDL on Apache Spark*** optimize hardware components into an integrated platform to simplify deep learning development and deployment in a verified infrastructure based on Apache Spark. By enabling analytics to run on your data, wherever it is stored, you can eliminate data transfer and duplication, and accelerate AI innovation. It also reduces TCO by enabling you to use your existing Intel Xeon processor-based architecture to run new AI workloads, helping boost utilization and minimize IT costs. Rich toolsets and optimized libraries help further accelerate time to value.
- **Intel® Select Solutions for AI Inferencing** provide a “turnkey platform” solution for low-latency, high-throughput inference performed on a CPU, not a separate accelerator card. It provides a jumpstart to deploying efficient AI inferencing algorithms on a solution composed of validated Intel® architecture building blocks. It uses the Intel DL Boost feature of 2nd Generation Intel Xeon Scalable processors to accelerate AI inference by performing in one instruction inferencing calculations that previously took multiple instructions. Accelerate inferencing further with the Intel® Distribution of OpenVINO™ Toolkit.

Run demanding AI inference workloads on 2nd Generation Intel® Xeon® Scalable processor-based infrastructure – vastly simplifying your production environment, saving cost and increasing utilization.

5. THE MAGNIFYING IMPACT OF OPTIMIZED SOFTWARE

Hardware is nothing without the right software to make it perform at its peak. Each AI use case will require a software architecture that selects the best tools for the job in hand, potentially taking into account downstream systems, customization, optimization and other modifications – a selection of these are noted in the table below.

When examining these options, consider if the toolkits, libraries and frameworks are:

- **Optimized for existing environments and capabilities.** Many open source software libraries and frameworks like TensorFlow*, MXNet*, PaddlePaddle*, and PyTorch*, are optimized for Intel Xeon Scalable processors. Intel has worked with Google * on TensorFlow, with Apache on MXNet, with Baidu* on PaddlePaddle, and on Caffe* to enhance deep learning performance using software optimizations for Intel Xeon Scalable processors in the data center, and is continuing to add frameworks from Microsoft and other industry leaders.
- **Aligned with the needs of the enterprise's use case.** Whether looking to run a PoC or scale an existing solution, IT leaders need to define what specific bottlenecks or challenges they would like software optimizations to help them overcome.
- **Preferred by in-house talent.** Data scientists and developers in the enterprise will have to use these tools. If they aren't familiar with the optimizations chosen by IT, it's important they understand how they can reduce development time and improve efficiency.

TOOLKITS	WHAT DOES IT DO?
Intel® Distribution of OpenVINO™ toolkit	A software toolkit that helps computer vision teams **speed the development and deployment of neural network applications** on gateways and devices across multiple Intel® platforms (CPU, GPU, FPGA, VPU). Quickly optimize pretrained models and deploy them across a wide range of Intel® hardware and accelerators, often with significant performance improvements over using deep learning frameworks as is, with no large changes to how you deploy today.
Intel® Movidius™ Neural Compute SDK (NCSDK)	A software development kit which allows for the **rapid prototyping and deployment of deep neural networks (DNNs)** on compatible neural compute devices such as the Intel Movidius Neural Compute SDK (NCSDK). It includes a set of software tools to compile, profile, and validate DNNs.
LIBRARIES	
Intel® Math Kernel Library for Deep Learning Networks (Intel® MKL-DNN)	Intel MKL-DNN is an open source, performance-enhancing library for **accelerating deep learning frameworks on Intel® architecture**. Software developers who are interested in the subject of deep learning may have heard of Intel MKL-DNN, but perhaps haven't yet had the opportunity to explore it first-hand.
Intel® Distribution for Python*	This tool advances the performance of the most popular and fastest growing programming language closer to native speeds. It includes out-of-the-box access to high-performance Python* with a drop-in replacement, multiple optimization techniques, and fast access to Intel® architecture optimizations.

Bone-Age-Prediction Model

188X INCREASE

in Images per Second[4]

Lung-Segmentation Model

38X INCREASE

in Images per Second[4]

CASE STUDY:

PHILIPS *: EFFICIENT AI-DRIVEN MEDICAL IMAGING

Intel teamed up with Philips to show that servers powered by Intel Xeon Scalable processors could be used to efficiently perform deep learning inference on patients' X-rays and CT scans, without the need for accelerators. The ultimate goal for Philips is to offer AI to its end customers without significantly increasing the cost of the customers' systems and without requiring modifications to the hardware deployed in the field. The companies tested two healthcare use cases for deep learning inference models: one on X-rays of bones for bone age–prediction modeling, and the other on CT scans of lungs for lung segmentation.

Discover the results Philips achieved

CASE STUDY:

ZIVA DYNAMICS*: PUSHING THE BOUNDARIES OF SIMULATION WITH AI

CGI visual effects (VFX) generally require a complex orchestration of expertise, technology, and often time-consuming and exhaustive creative iterations. Ziva Dynamics tackles this using artificial intelligence-based simulation software that allows VFX artists to build creatures that look and move correctly based on the laws of physics. Ziva runs its software in an Intel Xeon Scalable processor-based environment. Much of its software was also written using Intel® Math Kernel Library (Intel® MKL) PARDISO* and Intel MKL Linear Algebra Package* (LAPACK*). As a result, the company can create realistic effects quickly.

Read the Ziva Dynamics story in full

6. NEXT STEPS: BREAKING BARRIERS BETWEEN MODEL AND REALITY

While AI's impact on technology and society is still in its infancy, the momentum is palpable. Some leading companies and markets are adopting AI now, but those just getting started should look to establish a path to deep learning inference at scale – and that starts with AI readiness: exploring business cases, getting data in order, and finding the right blend of people and technology that can translate the AI hype into a reality for their business.

As you think about the first or next steps in your own AI journey, try thinking about where your organization sits on the AI readiness model. Depending on your level of readiness, use this checklist to ensure that your enterprise can scale with the business objectives, tools, talent and security considerations necessary to succeed.

Foundational readiness: Using AI for the first time

- Is the scenario, use case or problem to be solved with AI clearly defined?
- Are priorities set around where AI will deliver the most business value?
- Is the planned infrastructure architecture clear and appropriate?
- Are all necessary data sources clearly understood and accessible?
- Can your chosen software packages deliver the AI solution end-to-end?
- Are sufficient skills and resources available (either in-house or externally)?
- Have expectations been set around training and learning times?
- Is the total cost of ownership (TCO) of the end-to-end solution clear and signed off?

Operational readiness: Scaling up use of AI

- Can the planned solution scale beyond initial testing and evaluation?
- Is a clearly defined business case confirmed with a business unit?
- Is sufficient direct resourcing available, with time allocated and reserved?
- Is network bandwidth sufficient to ensure timely data delivery at scale?
- Are operational management processes in place which cover AI delivery?
- Does the architecture align with industry standards and best practices?
- Has a cybersecurity risk assessment been undertaken and acted upon?
- Have realistic deployment plans been set and communicated?

Transformational readiness: Broadening out use of AI

- Is a team in place to oversee AI-based continuous improvement?
- Are the broader AI opportunities for the organization researched and clear?
- Are AI solutions developed and deployed following agile best practice?
- Are measures in place to monitor business effectiveness of AI solutions?
- Is the architecture for AI provided as a platform, rather than as one-off solutions?
- Are lines of business fully engaged in how AI will affect their processes?
- Are the governance needs of the AI solution clearly understood?
- Is AI seen as a central pillar of an IT-enabled business strategy?

FURTHER READING

- **Find out more: Intel® Deep Learning Boost >**
- **White paper: Lower Numerical Precision Deep Learning Inference and Training >**
- **Solution Brief: Intel Select Solutions for Big DL Apache Spark >**
- **Solution Brief: Intel Select Solutions for AI Inferencing >**

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

Performance results are based on testing as of the date set forth in the configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

[1] https://www.nextplatform.com/2018/10/18/deep-learning-is-coming-of-age/

[2] 14x inference throughput improvement on Intel® Xeon® Platinum 8280 processor with Intel® DL Boost: Tested by Intel as of 2/20/2019. 2 socket Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x200004d), Ubuntu 18.04.1 LTS, kernel 4.15.0-45-generic, SSD 1x sda INTEL SSDSC2BA80 SSD 745.2 GB, nvme1n1 INTEL SSDPE2KX040T7 SSD 3.7TB, Deep Learning Framework: Intel® Optimization for Caffe version: 1.1.3 (commit hash: 7010334f159da247db3fe3 a9d96a3116ca06b09a), ICC version 18.0.1, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d94195140cf2d8790a75a, model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/resnet50_int8_full_conv.prototxt, BS=64, synthetic Data, 4 instance/2 socket, Datatype: INT8 vs. Tested by Intel as of July 11th 2017: 2S Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384GB DDR4-2666 ECC RAM. CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7. x86_64. SSD: Intel® SSD DC S3700 Series (800GB, 2.5in SATA 6Gb/s, 25nm, MLC).

Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact', OMP_NUM_THREADS=56, CPU Freq set with cpupower frequency-set -d 2.5G -u 3.8G -g performance. Caffe: (http://github.com/intel/caffe/), revision f96b759f71b2281835f690 af267158b82b150b5c. Inference measured with "caffe time -- forward_only" command, training measured with "caffe time" command. For "ConvNet" topologies, synthetic dataset was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models (ResNet-50). Intel C++ compiler ver. 17.0.2 20170213, Intel MKL small libraries version 2018.0.20170425. Caffe run with "numactl -l".

[3] 30x inference throughput improvement on Intel® Xeon® Platinum 9282 processor with Intel® DL Boost: Tested by Intel as of 2/26/2019. Platform: Dragon rock 2 socket Intel® Xeon® Platinum 9282(56 cores per socket), HT ON, turbo ON, Total Memory 768 GB (24 slots/ 32 GB/2933 MHz), BIOS: SE5C620.86B.0D.01.0241.112020180249, Centos 7 Kernel 3.10.0-957.5.1.el7.x86_64, Deep Learning Framework: Intel® Optimization for Caffe version: https://github.com/intel/caffe d554cbf1, ICC 2019.2.187, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d9 4195140cf2d8790a75a), model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/resnet50_int8_full_conv.prototxt, BS=64, No datalayer syntheticData: 3x224x224, 56 instance/2 socket, Datatype: INT8 vs. Tested by Intel as of July 11th 2017: 2S Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384GB DDR4-2666 ECC RAM. CentOS Linux relèase 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7. x86_64. SSD: Intel® SSD DC S3700 Series (800GB, 2.5in SATA 6Gb/s, 25nm, MLC).

Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact', OMP_NUM_THREADS=5 6, CPU Freq set with cpupower frequency-set -d 2.5G -u 3.8G -g performance. Caffe: (http://github.com/intel/caffe/), revision f96b759f71b2281835f690 af267158b82b150b5c. Inference measured with "caffe time -- forward_only" command, training measured with "caffe time" command. For "ConvNet" topologies, synthetic dataset was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models (ResNet-50). Intel C++ compiler ver. 17.0.2 20170213, Intel MKL small libraries version 2018.0.20170425. Caffe run with "numactl -l".

[4] Configuration details: Hardware: Intel® Xeon® Platinum 8168 processor at 2.70 GHz, Intel® Hyper-Threading Technology (Intel®HT Technology) disabled. BIOS Version: SE5C620.86B.OD.01.0010.072020182008. System Memory: 192 GB, 2,666 MHz. Intel® Turbo Boost Technology enabled. SSD: ATA device, with non-removable media, model number: INTEL SSDSCS2CW240A3. Software: Ubuntu 18.04.1 LTS (GNU/Linux 4.15.0-29-generic x86_64*. Keras 2.1.1. TensorFlow 1.2.1. OpenVINO Toolkit 2018 R2. Intel® Math Kernal Library for Deep Neural Networks v0.14. Datasets: Bone-age prediction model: 299x299x3 .png images. Lung-segmentation model: 512x512 .dcm images.

0519/MMK/CAT/PDF

338678-002EN