FORRESTER®

The Total Economic Impact™ Of Intel AI Spotlight: Interoperability Efficiency

According to Forrester Research: “AI is quickly becoming an elemental core capability that every enterprise must leverage to be an industry leader. It takes leadership, talent, and tools. But it also requires AI infrastructure to expand use cases and infuse AI-powered intelligence into production applications.”[1]

Intel provides both hardware and software solutions for companies to use in building and deploying their AI and machine learning (ML) models. AI/ML workloads demand high power and infrastructure costs, inhibiting organizations from optimizing costs and speeding up inferencing. Intel provides AI chips and optimized solutions to scale and unlock AI insights.

Intel commissioned Forrester Consulting to conduct a Total Economic Impact™ (TEI) study and examine the potential benefits enterprises may realize by deploying Intel AI.[2] The purpose of this study is to provide readers with a framework to evaluate the potential financial impact of Intel AI on their organizations.

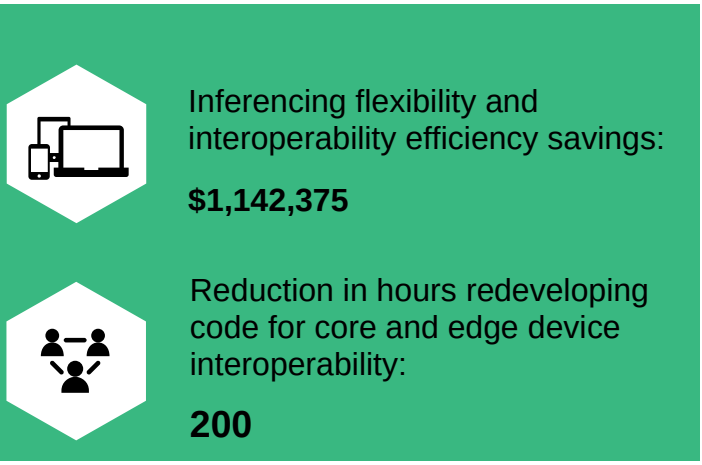
To better understand the benefits, costs, and risks associated with Intel AI, Intel commissioned Forrester Consulting to interview seven customers with experience using Intel AI chips and software for their AI/ML inferencing workloads.

This is a spotlight of the full Total Economic Impact™ study. Please read the full study for more detailed information.

CUSTOMER JOURNEY

Interviewees came from a variety of previous legacy environments, i.e., from a reliance on cloud service providers (CSPs), to running GPUs across training and inferencing workloads. Interviewees struggled with latency, performance, and cost limitations when trying to use GPUs for inferencing workloads. One healthcare customer, that was utilizing computer vision and deep learning technologies for medical image scanning, told Forrester that the size of their medical images were too large and expensive for GPUs. In addition, the organization was unable to add GPUs to their edge medical devices that were responsible for inferencing.

A professional services organization that was modernizing their data analytics, by offering robotic and intelligent automation, found success training models in the cloud. However, the organization required a model optimizer solution that could convert those training models to the required hardware format, especially across edge and internet-of-things (IoT) devices. Customers required a cost-effective solution that could support their AI/ML initiatives of improving inferencing speed, latency, and flexibility to run inferencing workloads across various infrastructure from data centers to the cloud to the edge.



KEY FEATURES

Ultimately, interviewees noted several reasons for investing in Intel AI chips, including:

- **Size, weight, and power.** Intel AI chips were smaller, weight less, consumed less power, and produced less heat than alternatives when running AI/ML inference workloads. This was especially important when trying to move AI compute to edge devices to speed up inferencing tasks, as opposed to sending data to the cloud or back to the data center for processing.
- **Ecosystem and breadth of portfolio.** Intel's chipset covers the breadth of infrastructures and AI use cases for companies, making it simpler to deploy across ecosystems. This is especially important when considering interoperability and compatibility of AI/ML workloads across a company's IT infrastructure (e.g., from edge to core).
- **Ability to optimize the most important metric.** Intel's variety of processor chips including CPUs and field-programmable gate arrays (FPGAs) afforded them the ability to balance data size, latency, and overall performance.
- **The extension of existing infrastructure.** Interviewees noted that investing in Intel AI allows them to scale their AI/ML usage by extending existing infrastructure and edge devices rather than purchasing new point-solutions and hardware. One interviewee noted that upgrading existing CPUs is much easier and cheaper than building GPUs.

> **"[For AI inferencing] you can get GPU speeds on Intel CPUs."**
>
> *Managing director, professional services industry*

KEY RESULTS

Based on the customer interviews and TEI analysis, Forrester found the following risk-adjusted present value (PV) benefits from using Intel AI:

- **Interoperability efficiencies totaling more than $1.1 million.** Interviewees reported using Intel AI chips to deploy inferencing workloads across a broad range of infrastructure. Deployment flexibility might be needed between edge and data center devices if an edge device can process a subset of computer vision inferencing workloads but then needs to send more complex data back to the data center for processing.

  One customer noted that their organization expected ten times reduction in developer resources by developing once in OpenVINO and Intel and porting the code across data center and edge devices, as opposed to developing on another chipset and platform and then requiring a separate x86 edge team to redevelop the code. Another customer reported that up to 40% of their AI/ML projects require interoperability between inferencing devices.

Modeling and assumptions. Based on the customer interviews, Forrester modeled the financial impact for the composite organization with the following estimates:

- In Year 1, 75 total AI/ML models are developed, growing to 150 annual AI/ML models by Year 3 as the composite expands its data scientist headcount.
- Thirty percent of all AI/ML models developed require interoperability between edge and data center compute resources.
- Each model requiring interoperability saves 200 hours since the team can port the code across edge and data center devices, as opposed to redeveloping the code on a different platform

Inferencing Flexibility					
Ref.	**Metric**	**Source**	**Year 1**	**Year 2**	**Year 3**
B1	Total AI/ML models per year	A1*A2	75	100	150
B2	AI/ML models requiring interoperability between core and edge devices	Interviews	30%	30%	30%
B3	Additional effort per model avoided in redeveloping code (hours)	Interviews	200	200	200
B4	Average data scientist fully burdened salary (hourly)	Composite	$85	$85	$85
Bt	Inferencing flexibility	B1*B2*B3*B4	$382,500	$510,000	$765,000
	Risk adjustment	↓15%			
Btr	Interoperability efficiencies (risk-adjusted)		$325,125	$433,500	$650,250
	Three-year total: $1,408,875		**Three-year present value: $1,142,375**		

UNQUANTIFIED BENEFITS

Additional benefits that customers experienced but were not quantified include:

- **Improved inference performance.** Customers noted that Intel AI chips improved inference performance compared to alternatives. With this solution, inferencing workloads ran quickly. Additionally, edge devices allowed inferencing to run locally on the device as opposed to sending the data to the cloud and back, saving more time.
- **Less power required for FPGAs vs. GPUs.** Customers also noted that edge workloads required special considerations, all of which Intel AI chips addressed, noting that FPGA chips are a much more power-considerate device. Intel AI provides size/weight/power considerations, the ability to power the chip and edge device from a battery, and heat generation considerations.
- **Software adoption.** Customers noted that the simple developer interface for OpenVINO and other software associated with Intel AI chips was key in driving adoption for their company and data scientists.

> **"With Intel, we're pushing into the smart world by integrating IoT devices and other operational technology to create smart infrastructure."**
>
> *Chief technology advisory, technology industry*

READ THE FULL STUDY HERE

TOTAL ECONOMIC IMPACT ANALYSIS

For more information, download the full study: "The Total Economic Impact™ Of Intel AI," a commissioned study conducted by Forrester Consulting on behalf of Intel, June 2021.

STUDY FINDINGS

Forrester interviewed seven organizations with experience using Intel AI and combined the results into a three-year composite organization financial analysis. Risk-adjusted present value (PV) quantified benefits include:

- Development time savings with OpenVINO of **$2,150,353**.
- Inferencing flexibility and interoperability efficiencies totaling more than **$1,142,375**.
- Hardware savings totaling over **$1,623,155**.

Appendix A: Endnotes

[1] Source: "Now Tech: AI Infrastructure, Q1 2020," Forrester Research, Inc., February 7, 2020.
[2] Total Economic Impact is a methodology developed by Forrester Research that enhances a company's technology decision-making processes and assists vendors in communicating the value proposition of their products and services to clients. The TEI methodology helps companies demonstrate, justify, and realize the tangible value of IT initiatives to both senior management and other key business stakeholders.


FORRESTER®