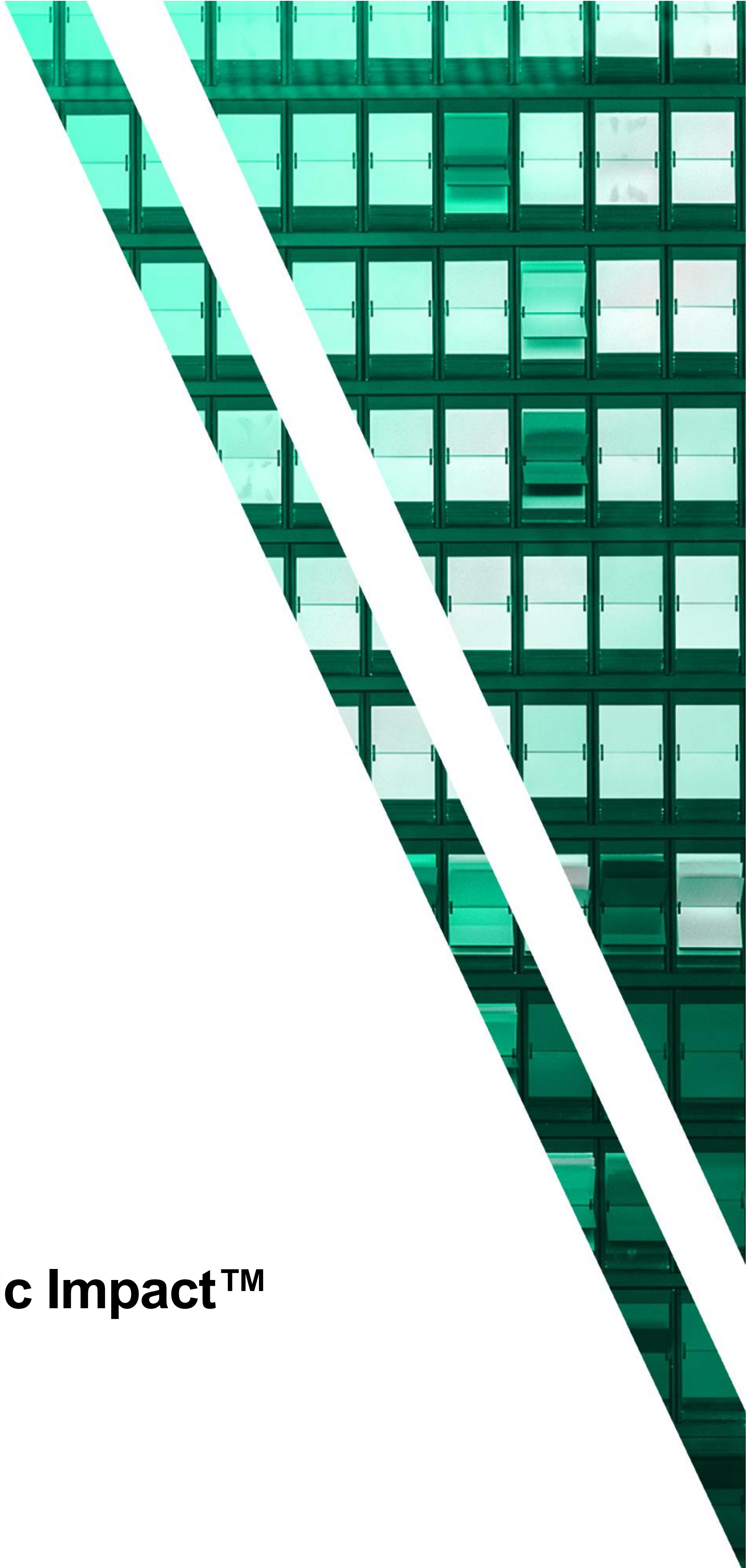
FORRESTER®



The Total Economic Impact™ Of Intel AI

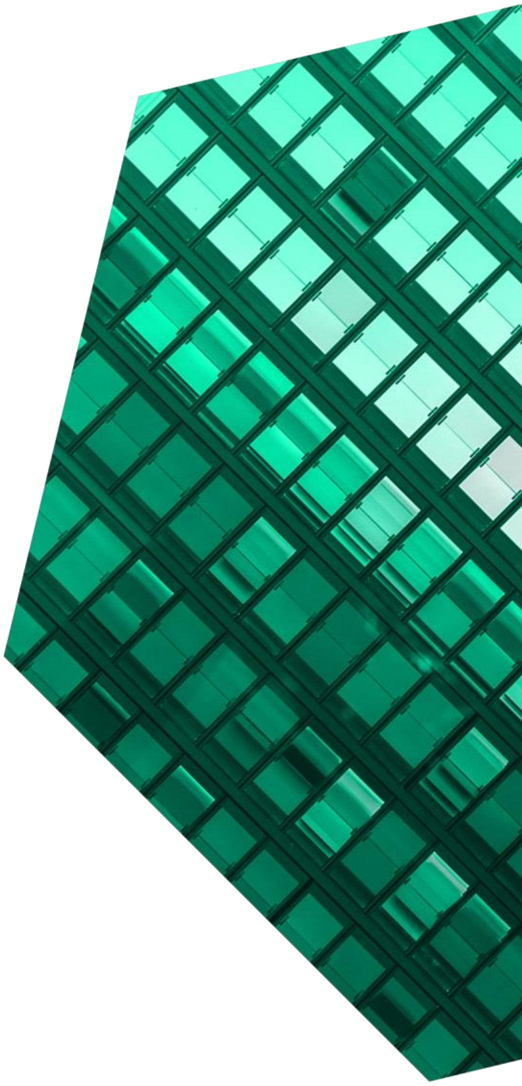
A Benefits Analysis Of Intel AI

JUNE 2021

A FORRESTER TOTAL ECONOMIC IMPACT™ STUDY COMMISSIONED BY INTEL

Table Of Contents

Consulting Team: Steve Odell
Luca Son
Jasper Narvil



ABOUT FORRESTER CONSULTING

Forrester Consulting provides independent and objective research-based consulting to help leaders succeed in their organizations. For more information, visit forrester.com/consulting.

Executive Summary

> As enterprises invest in AI and deep learning solutions to drive digital transformation, rapidly generate new business insights, and stay highly competitive, they need cohesive hardware and software solutions. Hardware coupled with optimized AI software allows enterprises to seamlessly build and deploy AI applications faster and at scale.

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.[1] 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 and risks associated with this investment, Forrester interviewed seven customers with experience using Intel AI chips and software for their AI/ML inferencing workloads.[2] For the purposes of this study, Forrester aggregated the experiences of the interviewed customers and combined the results into a single composite organization.

The interviewed organizations decided to deploy Intel AI due to the size, weight, and power of Intel AI chips, Intel's ecosystem and breadth of portfolio, and the ability of Intel's chips to process workloads they couldn't run on their graphics processing units (GPUs). With Intel AI, organizations saw development time savings with OpenVINO, interoperability efficiencies, and hardware savings.

KEY STATISTICS

Total benefits PV
$4.9 million

Hardware savings PV
$1.6 million

KEY FINDINGS

Quantified benefits. Risk-adjusted present value (PV) quantified benefits include:

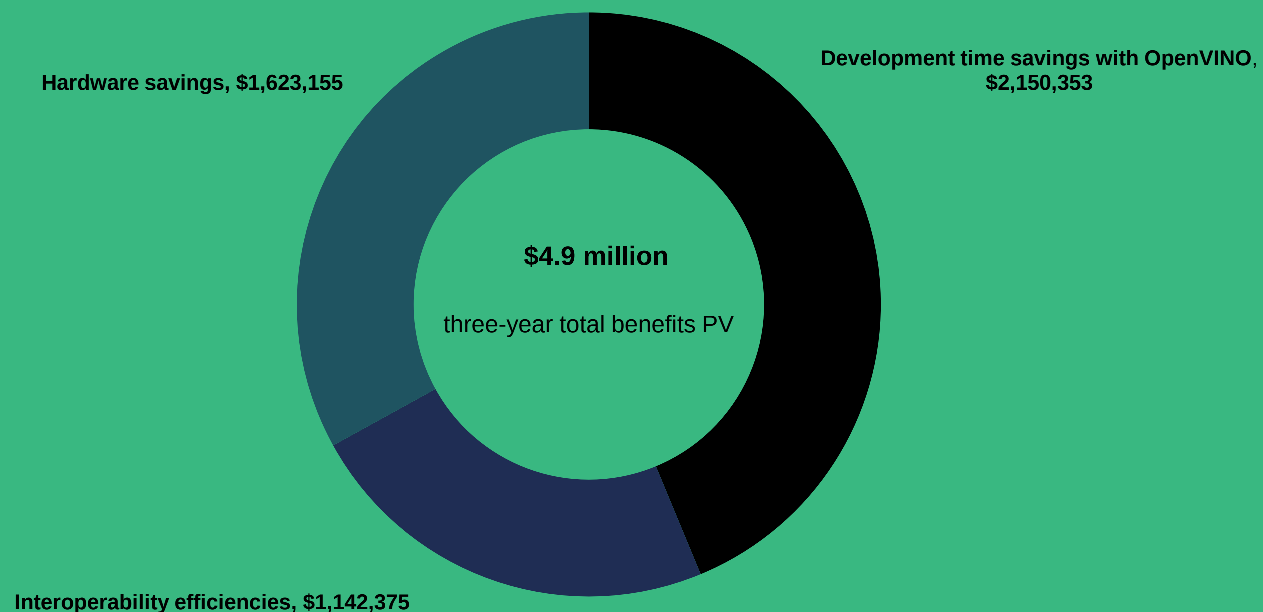
- **Development time savings with OpenVINO totaling nearly $2.2 million.** Interviewees noted that using Intel's OpenVINO toolkit saved their organizations' data scientists time when deploying their inferencing models to Intel AI chips. This significantly reduced coding and deployment time for their inference models.

- **Interoperability efficiencies totaling more than $1.1 million.** Interviewees stated the ability to use a consistent Intel infrastructure and ecosystem across their AI/ML inferencing devices is an interoperability benefit of using Intel AI chips for their AI/ML workloads. Interoperability 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.

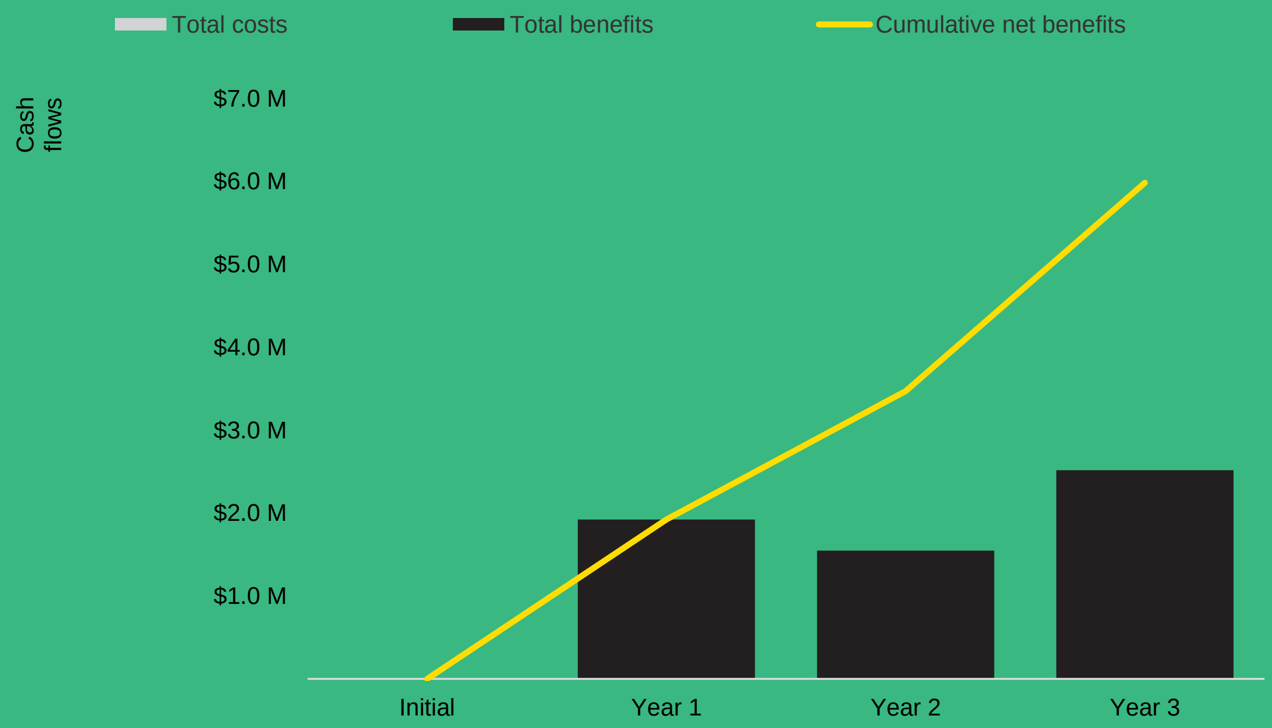
- **Hardware savings totaling over $1.6 million.** Interviewees reported using Intel chips for their organizations' AI workloads resulted in significant cost savings. The organizations used their existing infrastructure for inferencing workloads run in the data center. Upgrading their edge devices to run inferencing workloads cost less with Intel chips compared to alternatives.

Unquantified benefits. Benefits that are not quantified for this study 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 field-programmable gate array (FPGA) 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.
  - 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.



Cash Flow Chart (Risk-Adjusted)



TEI FRAMEWORK AND METHODOLOGY

From the information provided in the interviews, Forrester constructed a Total Economic Impact™ framework for those organizations considering an investment in Intel AI.

The objective of the framework is to identify the cost, benefit, flexibility, and risk factors that affect the investment decision. Forrester took a multistep approach to evaluate the impact that Intel AI can have on an organization.

DUE DILIGENCE

Interviewed Intel stakeholders and Forrester analysts to gather data relative to Intel AI.

CUSTOMER INTERVIEWS

Interviewed seven decision-makers at organizations using Intel AI to obtain data with respect to costs, benefits, and risks.

COMPOSITE ORGANIZATION

Designed a composite organization based on characteristics of the interviewed organizations.

FINANCIAL MODEL FRAMEWORK

Constructed a financial model representative of the interviews using the TEI methodology and risk-adjusted the financial model based on issues and concerns of the interviewed organizations.

CASE STUDY

Employed four fundamental elements of TEI in modeling the investment impact: benefits, costs, flexibility, and risks. Given the increasing sophistication of ROI analyses related to IT investments, Forrester's TEI methodology provides a complete picture of the total economic impact of purchase decisions. Please see Appendix A for additional information on the TEI methodology.

DISCLOSURES

Readers should be aware of the following:

This study is commissioned by Intel and delivered by Forrester Consulting. It is not meant to be used as a competitive analysis.

Forrester makes no assumptions as to the potential ROI that other organizations will receive. Forrester strongly advises that readers use their own estimates within the framework provided in the study to determine the appropriateness of an investment in Intel AI.

Intel reviewed and provided feedback to Forrester, but Forrester maintains editorial control over the study and its findings and does not accept changes to the study that contradict Forrester's findings or obscure the meaning of the study.

Intel provided the customer names for the interviews but did not participate in the interviews.

The Intel AI Customer Journey

Drivers leading to the Intel AI investment

Interviewed Organizations

Industry	Region	Interviewee	Annual Revenue
Technology industry	Global HQ in North America	Chief R&D scientist	$100M+
Technology industry	Global HQ in North America	Chief AI architect	$10B+
Professional services	Global HQ in North America	Managing director	$10B+
Technology industry	Global HQ in North America	Technical fellow	$10B+
Technology industry	Global HQ in North America	Chief technology advisor	$10B+
Healthcare industry	Global HQ in Asia	Chief executive officer (CEO)	Private
Technology industry	Global HQ in North America	Director	$1B+

WHY INTEL AI?

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).
- **Inability of GPUs to process very large images.** One customer noted that the size of images they work with are too large for GPUs to process. Intel's variety of processor chips including central processing units (CPUs) and FPGAs afforded them the ability to balance data size, latency, and overall performance.

"[For AI inferencing] you can get GPU speeds on Intel CPUs."

Managing director, professional services industry

"Intel has solved the challenge of categorizing and cutting huge image files, which couldn't be done with our GPUs."

CEO, healthcare industry

COMPOSITE ORGANIZATION

Based on the interviews, Forrester constructed a TEI framework, a composite company, and a ROI analysis that illustrates the areas financially affected. The composite organization is representative of the seven companies that Forrester interviewed and is used to present the aggregate financial analysis in the next section. The composite organization has the following characteristics:

Description of composite. The composite organization is a global organization headquartered in North America with $10 billion in annual revenue. It employs a growing data scientist team of 15 full-time equivalents (FTEs) in Year 1, 20 FTEs in Year 2, and 30 FTEs in Year 3. The composite uses Intel AI chips and software across the organization for inferencing workloads. AI/ML models are built for use cases where inferencing workloads are run in the data center and in edge devices.

Key assumptions

- **Global organization**
- **$10 billion in annual revenue**
- **15 data scientists in Year 1**
- **Intel AI chips and software for inferencing**

Analysis Of Benefits

Quantified benefit data as applied to the composite

Total Benefits

Ref.	Benefit	Year 1	Year 2	Year 3	Total	Present Value
Atr	Development time savings with OpenVINO	$612,000	$816,000	$1,224,000	$2,652,000	$2,150,353
Btr	Interoperability efficiencies	$325,125	$433,500	$650,250	$1,408,875	$1,142,375
Ctr	Hardware savings	$985,625	$296,875	$641,250	$1,923,750	$1,623,155
	Total benefits (risk-adjusted)	$1,922,750	$1,546,375	$2,515,500	$5,984,625	$4,915,883

DEVELOPMENT TIME SAVINGS WITH OPENVINO

Evidence and data. Interviewees noted that their organizations' data scientists used Intel's OpenVINO toolkit to deploy their inferencing models to Intel AI chips, optimize Pytorch Models, and save development time. Customers reported that their organizations used Intel's pre-trained deep learning encoders, and one customer gave the example that their organization used OpenVINO's eyeglass detection module instead of building that from scratch. This significantly reduced coding and deployment time for their inference models.

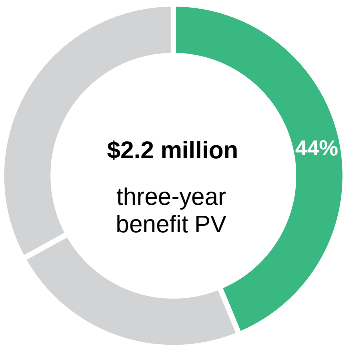
Modeling and assumptions. Based on the customer interviews, Forrester modeled the financial impact for the composite organization with the following estimates:

- The composite employs 15 data scientists in Year 1, growing to 30 data scientists in Year 3.
- Each data scientist develops five AI/ML models per year.
- Before OpenVINO, each AI/ML model took 160 hours to develop the inferencing model to be deployed to an Intel AI chip.
- With OpenVINO, this time is reduced to 40 hours.
- Data scientists have an hourly-equivalent fully burdened salary of $85.

Risks. This benefit can vary due to uncertainty related to:

- The number of data scientists.
- The number of models each data scientist can develop.
- The time saved by using OpenVINO.
- Data scientist salary.

To account for these risks, Forrester adjusted this benefit downward by 20%, yielding a three-year, risk-adjusted total PV (discounted at 10%) of nearly $2.2 million.



Development Time Savings With OpenVINO					
Ref.	**Metric**	**Source**	**Year 1**	**Year 2**	**Year 3**
A1	Number of data scientists	Composite	15	20	30
A2	Number of AI/ML models developed per data scientist (per year)	Composite	5	5	5
A3	Average development time per model before OpenVINO (hours)	Interviews	160	160	160
A4	Average development time per model with OpenVINO (hours)	Interviews	40	40	40
A5	Average data scientist fully burdened salary (hourly)	Composite	$85	$85	$85
At	Development time savings with OpenVINO	A1*A2*(A3-A4)*A5	$765,000	$1,020,000	$1,530,000
	Risk adjustment	↓20%			
Atr	Development time savings with OpenVINO (risk-adjusted)		$612,000	$816,000	$1,224,000
	Three-year total: $2,652,000		**Three-year present value: $2,150,353**		

INFERENCING FLEXIBILITY AND INTEROPERABILITY EFFICIENCIES

Evidence and data. Interviewees reported using Intel AI chips to deploy inferencing workloads across a broad range of infrastructure from data centers to cloud to edge. 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.

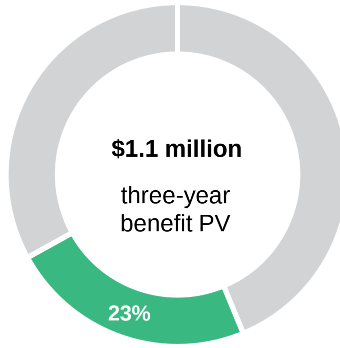
Modeling and assumptions. Based on the customer interviews, Forrester modeled the financial impact for the composite organization with the following estimates:

- 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.
- Data scientists have an hourly equivalent fully burdened salary of $85.

Risks. This benefit can vary due to uncertainty related to:

- The number of AI/ML models developed that require interoperability.
- The time required to redevelop code on another platform.
- Data scientist salary.

To account for these risks, Forrester adjusted this benefit downward by 15%, yielding a three-year, risk-adjusted total PV of more than $1.1 million.



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**			

HARDWARE SAVINGS

Evidence and data. Interviewees reported that using Intel chips for their AI workloads resulted in significant cost savings. Their organizations used their existing infrastructure for inferencing workloads run in the data center and upgrading their edge devices to run inferencing workloads cost less with Intel chips compared to alternatives. A customer told Forrester that their organization ran up to 70% of their AI/ML workloads on their existing data center infrastructure, and another customer reported that they saved up to $5,000 upgrading edge devices with Intel CPUs.

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

- The composite requires three server racks to run AI/ML workloads in their data center in Year 1, another rack in Year 2, and another two racks in Year 3.
  - Note that the number of servers required to run AI/ML workloads can vary significantly and depends more on how much data is being processed and how often/frequently its being accessed (as

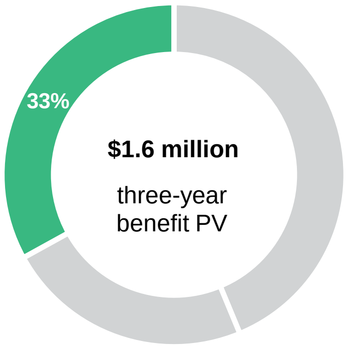
opposed to scaling based solely on number of AI/ML models developed and in production).

- Of those, two existing infrastructure server racks are repurposed for AI/ML workloads in Year 1, and another existing infrastructure server rack is repurposed in Year 3. Each of these existing racks save $50,000 that is otherwise required to build out a new rack to serve AI/ML workloads.
- In addition to data center workloads, the composite organization upgrades 375 edge devices to handle AI/ML inferencing workloads, another 125 devices in Year 2, and another 250 devices in Year 3.
- The composite saves $2,500 per edge device upgrade by using Intel CPUs.

Risks. This benefit can vary due to uncertainty related to:

- The number of data center servers avoided for running AI/ML workloads.
- The number of edge devices requiring upgrades for processing AI/ML workloads.
- The cost savings for repurposed data center servers and edge device upgrades.

To account for these risks, Forrester adjusted this benefit downward by 5%, yielding a three-year, risk-adjusted total PV of more than $1.6 million.



Hardware Savings

Ref.	Metric	Source	Year 1	Year 2	Year 3
C1	Number of new server racks required each year for AI/ML data center workload processing	Composite	3	1	2
C2	Number of existing server racks that can be used for AI/ML workloads	Interviews	2	0	1
C3	Avoided costs by using an existing server rack	Interviews	$50,000	$50,000	$50,000
C4	Subtotal: Savings on data center infrastructure	C2*C3	$100,000	$0	$50,000
C5	Number of edge devices needing upgrade to run AI inferencing	Composite	375	125	250
C6	Reduced upgrade costs per device with Intel CPUs	Interviews	$2,500	$2,500	$2,500
C7	Subtotal: Savings on edge devices	C5*C6	$937,500	$312,500	$625,000
Ct	Hardware savings	C4+C7	$1,037,500	$312,500	$675,000
	Risk adjustment	↓5%			
Ctr	Hardware savings (risk-adjusted)		$985,625	$296,875	$641,250
	Three-year total: $1,923,750	**Three-year present value: $1,623,155**			

UNQUANTIFIED BENEFITS

Additional benefits that customers experienced but were not able to quantify include:

- **Improved inference performance.** Customers told Forrester that Intel AI chips improved inference performance compared to alternatives. With this solution, inferencing workloads ran quickly. Additionally, with edge devices this allowed inferencing to be run locally on the device as opposed to sending the data to the cloud and back, saving more time.
- **Less power required for FPGA vs. GPUs.** Customers also stated 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.
  - 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.

Appendix A: Total Economic Impact

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.

TOTAL ECONOMIC IMPACT APPROACH

Benefits represent the value delivered to the business by the product. The TEI methodology places equal weight on the measure of benefits and the measure of costs, allowing for a full examination of the effect of the technology on the entire organization.

Costs consider all expenses necessary to deliver the proposed value, or benefits, of the product. The cost category within TEI captures incremental costs over the existing environment for ongoing costs associated with the solution.

Flexibility represents the strategic value that can be obtained for some future additional investment building on top of the initial investment already made. Having the ability to capture that benefit has a PV that can be estimated.

Risks measure the uncertainty of benefit and cost estimates given: 1) the likelihood that estimates will meet original projections and 2) the likelihood that estimates will be tracked over time. TEI risk factors are based on "triangular distribution."

The initial investment column contains costs incurred at "time 0" or at the beginning of Year 1 that are not discounted. All other cash flows are discounted using the discount rate at the end of the year. PV calculations are calculated for each total cost and benefit estimate. NPV calculations in the summary tables are the sum of the initial investment and the discounted cash flows in each year. Sums and present value calculations of the Total Benefits, Total Costs, and Cash Flow tables may not exactly add up, as some rounding may occur.

PRESENT VALUE (PV)

The present or current value of (discounted) cost and benefit estimates given at an interest rate (the discount rate). The PV of costs and benefits feed into the total NPV of cash flows.

NET PRESENT VALUE (NPV)

The present or current value of (discounted) future net cash flows given an interest rate (the discount rate). A positive project NPV normally indicates that the investment should be made, unless other projects have higher NPVs.

RETURN ON INVESTMENT (ROI)

A project's expected return in percentage terms. ROI is calculated by dividing net benefits (benefits less costs) by costs.

DISCOUNT RATE

The interest rate used in cash flow analysis to take into account the time value of money. Organizations typically use discount rates between 8% and 16%.

PAYBACK PERIOD

The breakeven point for an investment. This is the point in time at which net benefits (benefits minus costs) equal initial investment or cost.

Appendix B: Endnotes

[1] 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

[2] This study is focused on the benefits of using Intel chips and software for AI inferencing workloads. While this analysis was ongoing, Intel announced their Habana Gaudi AI processors/accelerators specifically focused on AI training workloads; these are outside the scope of the current study.


FORRESTER®