Hewlett Packard Enterprise

HPE Compute For AI

About this playbook

Solution sales playbooks are available to all HPE and Partner sales teams. You can use this interactive document to access workload and solution information, sales plays, and enablement collateral all in one place, making it easy to find everything you need to prepare for sales calls.

To navigate through the document, just use the tabs across the top of the page. To download materials, simply click the links.

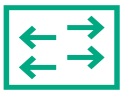
Why HPE Compute for AI

HPE Compute for AI Solutions enable our customers to deploy AI solutions on platforms they know and love - HPE ProLiant. With our focus on Inferencing and Generative AI, we are targeting the heat of the market and enabling our customers to take advantage of this next wave of computing to help move their business forward.

What's in it for me?

Learn a little, sell a lot.
You don't need to be a data scientist to sell HPE Compute for AI. Focus on the explosive growth in Inferencing and Generative AI with solutions featuring HPE ProLiant.

Unlock spending streams
With AI being the next big thing, some companies are unlocking strategic budgets to ensure they do not miss out and get put at a competitive disadvantage.

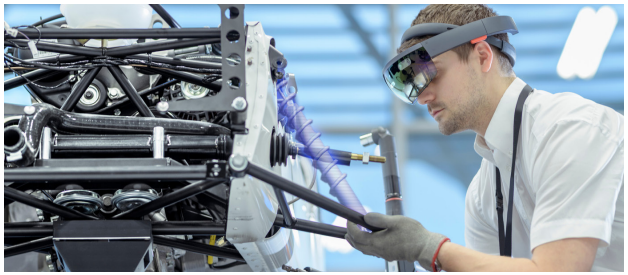


Capitalize on growth.
The AI market is forecast to continue it's growth at a fast rate. As customers look to adopt AI solutions, HPE ProLiant Compute platforms can help customers solve the unique challenges that AI adoption presents. See the Market Positioning section for more details.

The role of AI

AI defined

Artificial intelligence (AI) definition



AI represents a collection of software, logic, computing, and philosophical disciplines that aim to make computers perform functions once thought to be exclusively human, e.g. perceiving meaning in written or spoken language, learning, recognizing facial expressions, and so forth. The AI field has a long history, with many earlier breakthroughs, such as optical character recognition, now considered routine.

Why artificial intelligence?

AI is an attractive concept for many stakeholders in business, the sciences and government. On an economic basis, there is much appeal in having machines perform tasks that used to require human beings. An effective AI solution can "think" faster and process more information than any human brain. AI also has the potential to extend human abilities to places where people have trouble going.

Why HPE ProLiant Compute platforms?

The HPE ProLiant Gen11 / Gen10 Plus / Gen10 server portfolio offers flexible choices and versatile design, along with improved energy efficiencies, comprehensive management, and industry-leading performance, to support AI intensive workloads like generative AI tuning and inference, computer vision, speech AI, and more.

Defining AI terms

AI, Machine Learning, Deep Learning, Generative AI

Artificial Intelligence (AI)

A process where a computer solves a task, in a way that mimics human behavior. Today, narrow AI—when a machine is trained to do a specific task is becoming more widely used, from virtual assistants to self-driving cars to automatically tagging your friends in your photos on social media.

Machine Learning (ML)

Algorithms that allows computers to learn from examples without being explicitly programmed.

Deep Learning (DL)

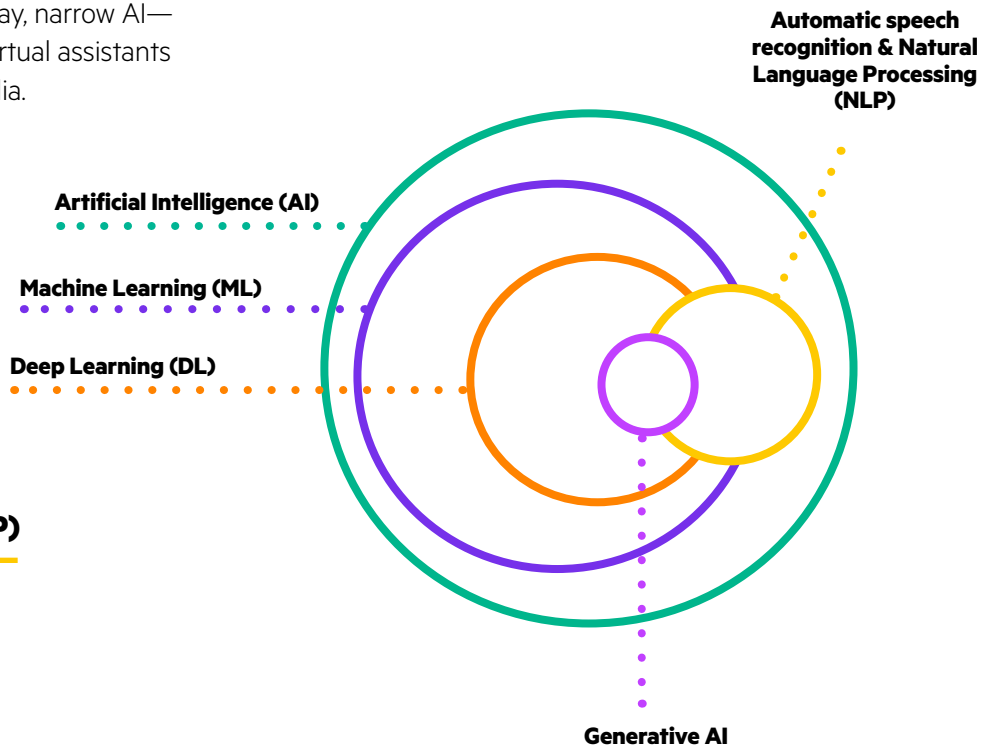
A subset of ML which uses deep artificial neural networks as models and does not require feature engineering.

Automatic speech recognition & Natural Language Processing (NLP)

Use of AI to help computers recognize and translate spoken language and to understand human language.

Generative AI

Use of AI to create something new, typically pieces of content like text, imagery, audio.



AI evolution

The evolution of Artificial Intelligence

Small data sets

Early artificial intelligence

- **ENIAC** Heralded the "Giant Brain"; used for WWII ballistics
- Industrial robots
- Expert Systems

Statistical and mathematical models applied to solve problems

1940-1980

Massive structured data sets

Machine learning

- **Deep Blue** Beat World Chess Champion Kasparov
- **DARPA Challenge** Autonomous vehicle drove 132 miles

Advanced Analytics and Heuristics

1990-2000s

Massive unstructured big data

Deep Learning

- Unsupervised training
- Generic code
- Pattern recognition

Systems can

- Observe
- Test
- Refine

Successes

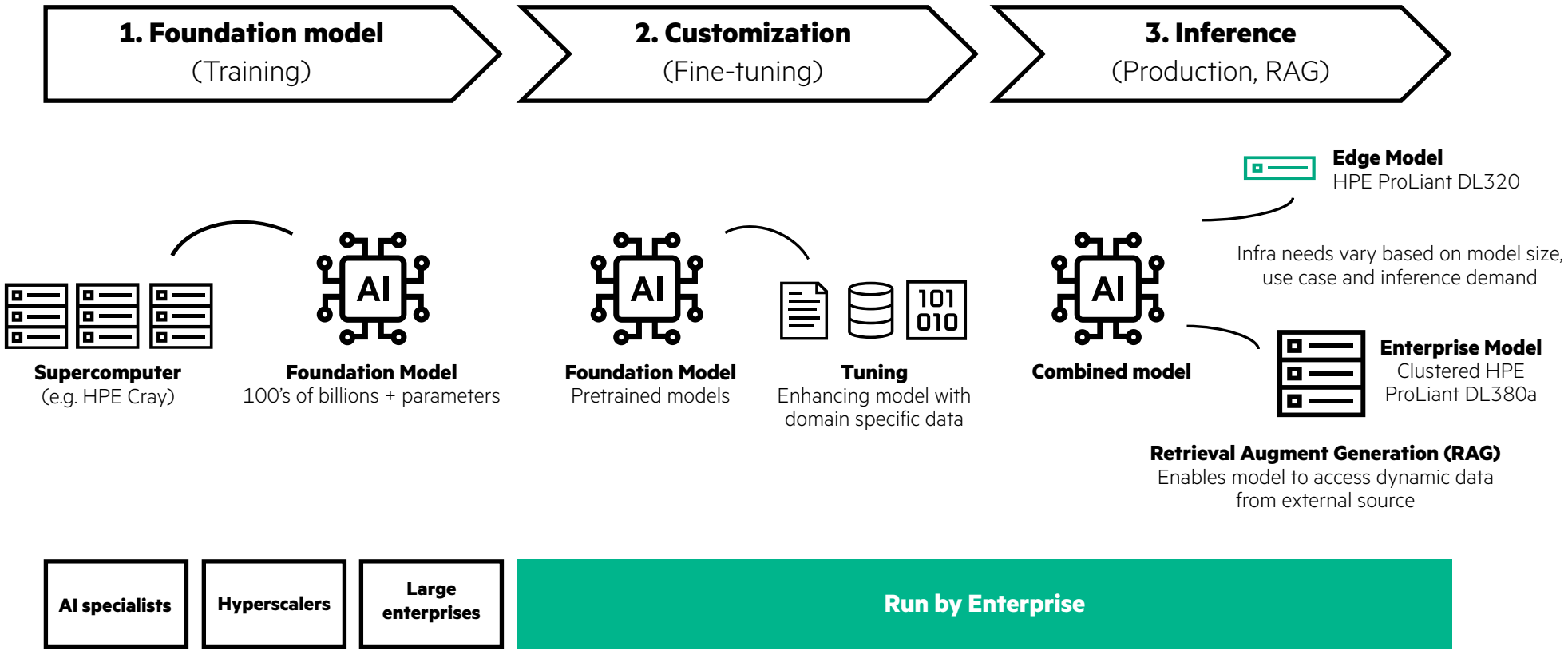
- AlphaGo first computer Go program to beat a human
- Deep Face facial verification
- Libratus AI Poker App
- Digital virtual assistants (Siri)
- Self-driving cars
- ChatGPT

Predictive models defined by machine-based neural networks

Today

AI evolution

Generative AI adoption follows three waves



Opportunity size

What is the HPE Compute for AI opportunity?

- By 2025, AI will represent a $35Bn global server market[1]
- By 2026, the total AI market will be worth $308B, and account for 30% of server sales[2]
- By 2027, IDC forecasts that GenAI will grow at 2X CAGR of overall AI spend to $143Bn[3]
- By 2034, IDC forecasts that GenAI will add nearly $10 Trillion to global GDP[4]

- GenAI spend will exceed $151billion worldwide by 2027, with a 2022-2027 total market CAGR of 86.1% (IDC #US51697124, January 2024)
- By 2027, 90% of organizations will augment operational roles with automation technology, elevating employee engagement and unlocking a 30% increase in worker efficiency. (IDC FutureScape: Worldwide Future of Operations 2024 Predictions)

"Generative AI has the potential to change the world in ways that we can't even imagine. It has the power to create new ideas, products, and services that will make our lives easier, more productive, and more creative. It also has the potential to solve some of the world's biggest problems, such as climate change, poverty, and disease."

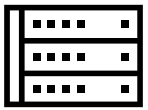
— Bill Gates, Microsoft Co-founder. Source: Forbes

[1] (IDC Worldwide Artificial Intelligence Spending Guide 2023)
[2] (IDC Worldwide Artificial Intelligence Spending Guide 2023)
[3] (IDC, GenAI Implementation Market Outlook: Worldwide Core IT Spending for GenAI Forecast 2023-2027)
[4] (IDC Infographic, sponsored by Microsoft, The Business Opportunity of AI, IDC #US51315823, November 2023. **Generative Artificial Intelligence: A New Chapter for Enterprise Business Applications, IDC Perspective #US50471523, March 2023.)

Market trends

Drivers, Restraints, Opportunities and Challenges

Drivers



- Need to stay competitive and profitable
- Increase in AI Applications and integrated development environment
- Improved computing power

Restraints

- The black box problem[5]
- Privacy concerns
- Existing legacy systems constraining new AI investments

Opportunities

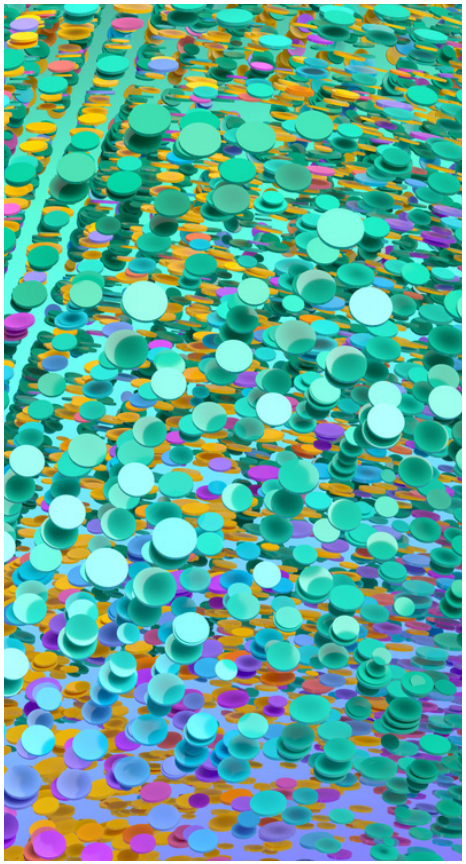
- Increase in investment from venture capitalists
- Integration of AI into every business process
- Emergence of edge technologies and devices amplify the usefulness of AI solutions

Challenges

- Limited skilled workforce
- Identifying the right use cases
- Data availability and quality

[5] In computing, a 'black box' is a device, system or program that allows you to see the input and output, but gives no view of the processes and workings between. The AI black box, then, refers to the fact that with most AI-based tools, we don't know how they do what they do.

Market trends



Banking is the largest of the industries with GenAI investments targeting tools to automate tasks and processes, enhanced financial services products, CX, and improving profitability with both cost reduction and revenue generation across business functions. (IDC #US51697124 (January 2024)

Retail is the second largest industry with GenAI moving retailers from being data rich to data driven, with initiatives focusing on new revenue and growth opportunities, and looking for more value from real-time, contextualized engagement (IDC #US51697124 (January 2024)

The most important business outcome that organizations are trying to achieve from AI initiatives is the increase of operational efficiency (Source: Global AI (including GenAI) Buyer Sentiment, Adoption and Business Value Survey IDC, October, 2023)

The top AI use case that has provided the greatest return on investment (ROI) for organizations is the automation of IT tasks.
(Source: Global AI (including GenAI) Buyer Sentiment, Adoption and Business Value Survey IDC, October, 2023)

Over the next 24 months, organizations plan to increase their AI spend by 23.4% (on average), to accomodate their usage of Generative AI
(Source: Global AI (including GenAI) Buyer Sentiment, Adoption and Business Value Survey IDC, October, 2023)

Elevator pitch

Hewlett Packard Enterprise is a trusted partner for the next era of AI. We offer proven technology and expertise, built from decades of experience as the global leader in supercomputing. Our AI strategy is hybrid by design, recognizing that data can live anywhere and that inferencing increasingly must happen at the edge.

Why GenAI solutions from HPE?

HPE's enterprise computing solution for generative AI is the ideal entry point for enterprises of all sizes. Each component is tuned for edge or data center deployments, powered by an ultra-scalable architecture that features cutting-edge AI software, industry-leading compute, pretrained foundation models, and comprehensive services to help you quickly develop production applications. GPU-enabled applications enable faster AI outcomes and greater inferencing performance to maximize the impact of GenAI.

The unique value of HPE

HPE ProLiant Compute platforms for AI are data-driven and results-oriented:

- Scalable and flexible solutions that are tested, proven, and secure to improve AI workload throughput and efficiency
- Designed to deliver higher flexibility, operational efficiency, and innovation potential while reducing cost
- Backed by decades of HPE expertise and extensive partner ecosystem to fast-track and speed your time to greater insights, control, and agility. HPE experts help design, implement and manage for customers with limited AI resources in-house to address skill and knowledge gaps
- From training to inferencing, HPE has a full portfolio of Compute solutions to address every sub-discipline in AI, including everything a customer would need: servers, storage, networking, and software
- HPE ProLiant Gen11, Gen10 Plus and Gen10 servers are the industry leader in performance and security
- HPE ProLiant Gen11, Gen10 Plus and Gen10 server performance is unmatched, setting world records for workload performance

Target verticals

- Financial services industry
- Manufacturing
- Healthcare and Life Sciences
- Retail

Target segments

- **Primary:**
  global enterprise
  commercial enterprise
- **Secondary:**
  midsize enterprises

Key decision makers

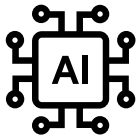
- CTOs and CIOs
- IT and datacenter managers
- LOB managers

Customer challenges

The challenges to AI adoption

No strategy:

Most business executives (80%) acknowledge the importance of AI for growth, but most haven't transitioned that belief into action, simply due to lack of strategy and experience. HPE ProLiant Compute servers give customers proven building blocks to accelerate digital transformation and validated solutions that match each customer's workload demands.



Lack of AI expertise:

Most organizations lack internal technical expertise to build, deploy, and manage AI solutions. With HPE ProLiant Compute platforms, users access common experiences, architecture, tools and options across server generations and product lines, resulting in a better user experience and faster time to value. Only HPE offers a workload-optimized portfolio where everything works together.

Costs:

HPE ProLiant Compute platform offers flexibility to meet a range of AI demands like model size and compute intensity. From the entry level server delivering 2P performance at 1P economics to powerful 2U2P servers.

Compute-intensive workloads:

AI workloads are compute-intensive, so powerful technology tools and modern compute infrastructure are a prerequisite. The HPE ProLiant Compute platform delivers right-sized, workload-matched server solutions built for the most demanding workloads.

Data scarcity:

AI models require enough data to properly train. Insufficient data makes models untrustworthy and unsuitable for production. HPE brings together hardware, software, and services to harness all relevant data. This combination delivers real-time insights that make AI adoption worthwhile.

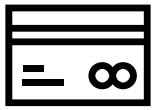
Benefits

HPE key message benefits

Improve workload performance

- Help customers improve AI workload performance with optimized and scalable compute solutions turning exponentially growing data into insights, action, and value faster.
- AI solutions need to be scalable as your business needs grow. The HPE ProLiant Compute Platforms provide record-breaking performance with its maximum core density and high memory capacity and energy efficiency, as well as support for up to 4 double-wide GPUs.

Reduce complexity and costs



- Organizations want AI to increase revenue, decrease support costs, and increase market competitiveness. HPE ProLiant Compute Platforms servers reduce costs and complexity, as well as offer flexible choices and versatile design.
- Fast, flexible compute infrastructure can be delivered on a consumption basis with HPE GreenLake through cost-optimized configuration options.
- HPE and NVIDIA have introduced a new enterprise computing solution for GenAI. It is an AI tuning and inferencing data center solution that provides the ideal entry point for enterprises of all sizes and is ready-out-of-the-box to jumpstart the AI journey, all available in a single SKU.

Rely on HPE expertise and partner ecosystem

- HPE AI expertise is based on hundreds of AI and data platform engagements, integrating our extensive partner ecosystem when needed for complete AI solutions.
- HPE experts integrate an evolving AI partner ecosystem, security, and key HPE products and move customers to the cloud with the HPE GreenLake consumption-based model.
- HPE Services experts help accelerate time to value through AI and Data strategy, design, implementation and ongoing support and management services.

Conversation starters

HPE key message benefits

AI customers run along a spectrum of needs. On one side are ISV-driven customers who need AI but don't want to manage it. (Most customers fall on this side.) On the other side of the spectrum are DIY customers who have AI experience, internal expertise, and the resources to apply it. Their goals are to improve MLOps process, control costs, and execute at scale.

HPE ProLiant Compute platforms deliver to customers all along this spectrum. These servers provide proven building blocks and validated solutions that can accelerate the digital transformation process, ideal for those on the ISV-driven side. They also provide secure industry standards, and a common user experience that delivers consistent operations and faster time to value for those on the DIY side of the spectrum.

No matter the persona of your customer, a conversation centered around the HPE ProLiant Compute platform covers the full spectrum of customer needs.

CTOs and CIOs

- What is your company's AI strategy?
- What would it mean to your business if you could employ AI for computer vision and NLP projects?
- How are you planning to implement GenAI in your business?

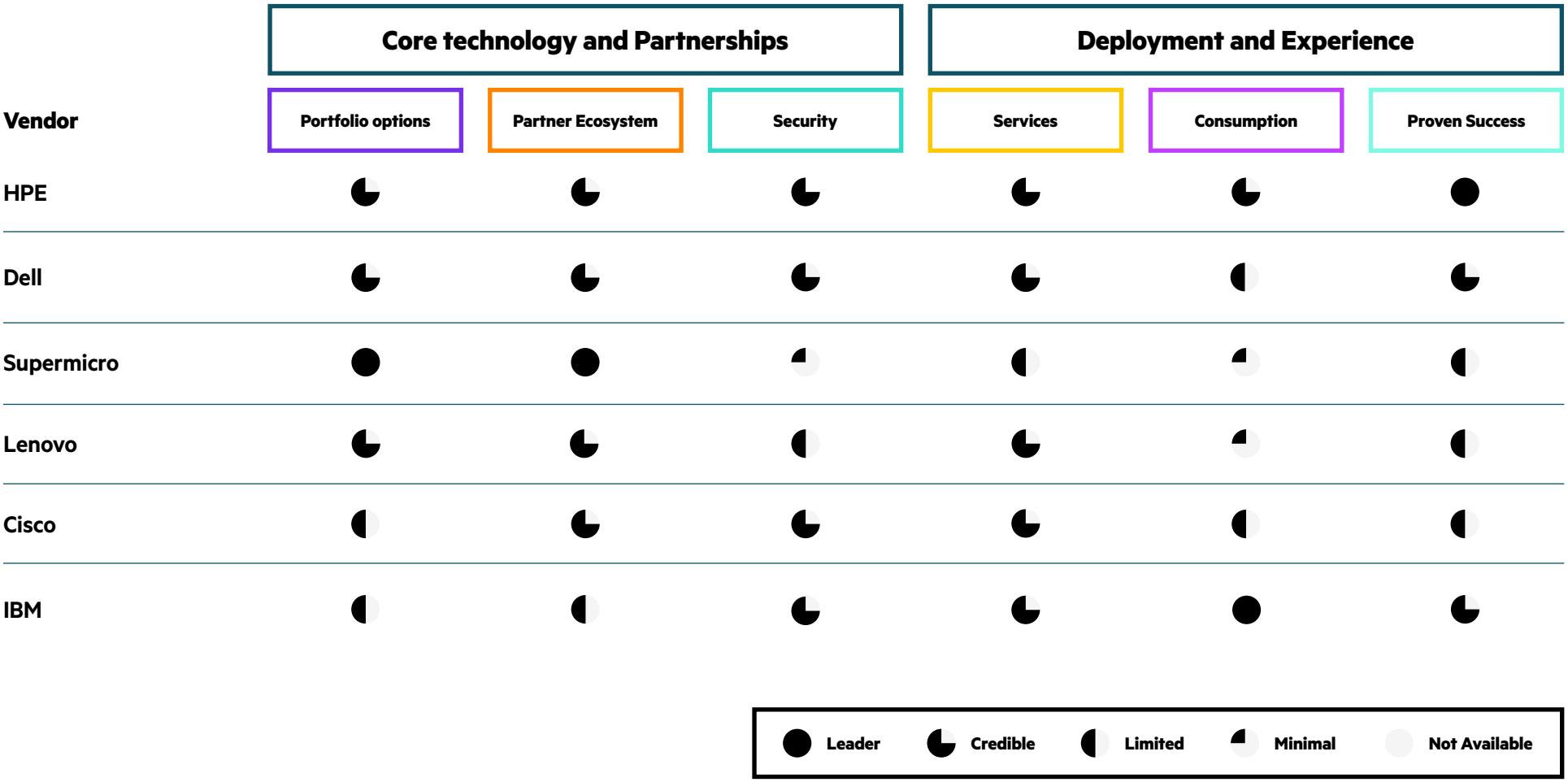
LOB managers

- How are you using AI to gain ground against your competitors?
- Have you considered AI for language, audio, or video applications as a differentiator between your business and your competition?
- Have you considered AI to improve operational efficiency?

IT and Datacenter managers

- What approaches are you taking to apply AI to data?
- How are you controlling AI costs and scaling AI initiatives to match business needs?

Competitive overview



Managing customer objections

Cost concern

Objection: "AI infrastructure sounds expensive. Can we justify the investment?"

Response: Highlight potential ROI - explain that GenAI can lead to operational efficiencies, reduced downtime, better resource usage and decision-making. Explain that long-term benefits will outweigh the initial investment. Additionally, through HPEFS, we can help customers fund, and even free up capital already invested in legacy systems to help accelerate new AI development.

Data security and privacy

Objection: "We're concerned about data security and compliance with AI solutions."

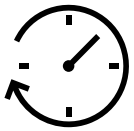
Response: Explain that security, such as data encryption, and access controls are built into the HPE GenAI solution, enabled by HPE software including MLDE and partner software including NVAIE from NVIDIA, all hosted on-premises under customer control.

Lack of experience

Objection: "Our team isn't familiar with AI. How can we manage and implement it?"

Response: Highlight that HPE AI experts available immediately to run a free 2-hour GenAI discovery workshop, to guide on AI benefits, together with complete services including training, professional guidance and deployment.

Implementation time



Objection: "We need quick results. Will implementing AI be time-consuming?"

Response: Explain how HPE and NVIDIA derisked and simplified GenAI with a preconfigured solution and enterprise support to speed deployment so you can fast track AI application development.

Green zone / Red zone

Green zone:

- Customer has a need to invest in scale out computing, inferencing and light data set training
- Customer has data sets that are not being used to their full potential, if at all.
- Customer has a safety focus and video/audio data sets could lead to predictive maintenance improvements, as well as greater safety and quality assurance.
- Customer has a need to improve customer experience, and wants to use data to improve customer interactions, inquiry responses, and improve product quality.
- Existing IT teams do not have the AI expertise to transition projects from POC to production.
- Customer is interested in Generative AI, but has not yet deployed.

Red zone:

For the following, please refer to an HPC/AI specialist:

- Scale up computing
- Heavy data set training
- Mission critical workloads

Engage these customers with caution. Consider engaging HPE Services to help develop a plan.

- Values price over performance, security, efficiency or other HPE strengths.
- Hasn't set a budget or timeline.
- Not willing to assess solution requirements.

Why HPE?

01

Expertise based on hundreds of AI and data platform engagements

02

Integration of partner ecosystem for each AI solution

03

Innovative, comprehensive computing portfolio to support edge to exascale AI implementations

04

Streamline Generative AI model deployment workflows integrated into a complete ML lifecycle with HPE Ezmeral Unified Analytics.

05

Fine tune models faster and more accurately with HPE Machine Learning Development Environment.

06

Economics and cloud-like experience from edge to core, and as-a-service

Use cases overview

Generative AI Use Cases
Enterprise search
Document intelligence
Chat bots
Code generation
Product design

Language and Audi Use Cases (AI Inference and Generative AI)
Natural Language Processing (NLP)
Natural Language Understanding (NLU)
Natural Language Generation (NLG)
Text Classification
Optical Character Recognition (OCR)

Computer Vision and Intelligent Video Analytics Use Cases (AI Inference)
Image Classification
Object Detection
Facial Recognition

Use cases

Generative AI use cases

Generative AI

Generates novel content, such as text, images or code, from user prompts, by using algorithms to learn from patterns in existing content.

Generative AI can help:

- Increase operational efficiency by automating tasks
- Realize cost savings by reducing manual labor
- Enhance employee productivity by automating routine tasks
- Innovate faster by accelerating idea generation, design, and prototype

Ideal for: Financial services, health and life sciences, customer service (cross industry)

Use cases

Enterprise Search: Enhance semantic search capabilities by grounding outputs in specific data sources to increase accuracy and relevancy

Why use

- Enhance employee productivity
- Simplify access to critical info
- Improve search accuracy

Possible applications

- Corporate search engines / Intranet

Document Intelligence: Automatically read, understand, and analyze business documents, facilitating the extraction of information and insights

Why use

- Enhance reliability of business information
- Accelerate workflow process
- Increase employee productivity

Possible applications

- Invoice analysis
- Mortgage workflow acceleration
- Risk mitigation (detection of counterfeits)

Use cases

Chatbots: using prompts, write code, compose emails, draft a report, generate art, or write Excel formulas

Why use

- Enhance customer experience
- Streamline operations
- Increase efficiencies

Possible applications

- Retail: Product recommendations
- Personalized assistance and support
- Healthcare virtual assistant

Code generation: create software code based on conversational prompts

Why use

- Reduce time to market
- Streamline processes
- Increase efficiencies

Possible applications

- Generate code snippets, complete code, or even write entire applications
- Bug detection
- Documentation

Use cases

Product design: Specifying design constraints and objectives, innovative designs can be generated that meet performance requirements while minimizing material usage, weight, or manufacturing costs

Why use

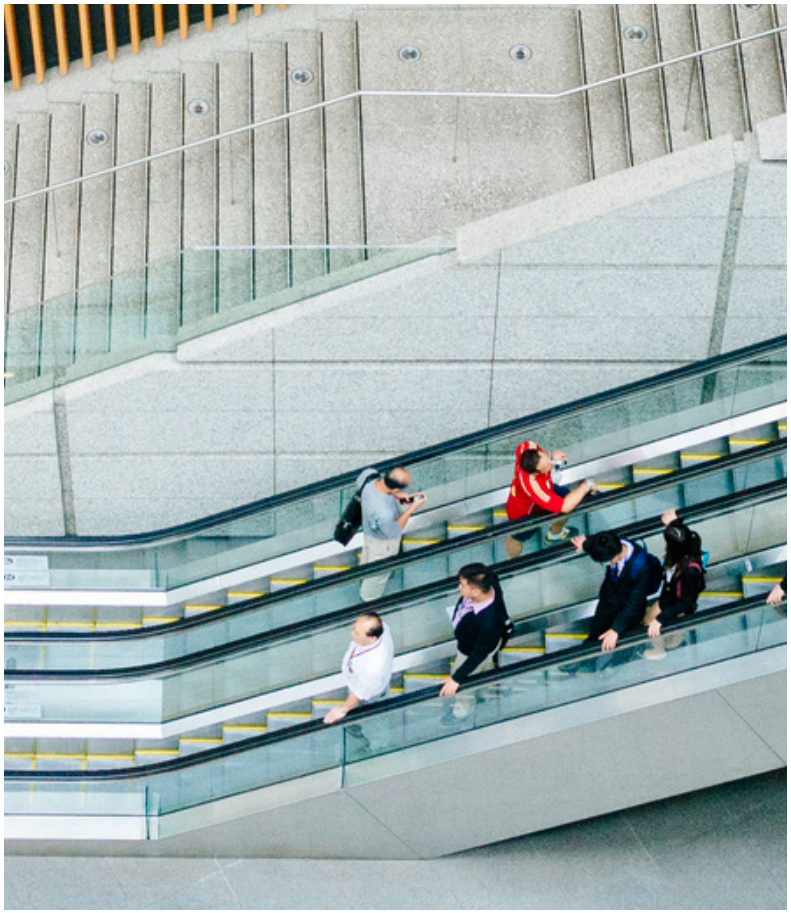
- Streamline tasks
- Accelerate time to market

Possible applications

- Concept development
- Concept testing and refinement

Use cases

Language and audio use cases



Speech and language processing

Automate the understanding of spoken and written language by applying natural language processing. This enables communications surveillance, speech-to-text analytics, biometric search, live call monitoring, and voice search.

These analytics may help:

- Automate fraud detection, improve compliance
- Improve customer service efficiency
- Speed up emergency response and law enforcement readiness

Ideal for: Healthcare, government, financial services, sports/entertainment, and publishing

Use cases

Natural Language Processing (NLP): Uses AI to help computers understand human language

Sentiment analysis (NLP sub-discipline):
Translates the sentiment and emotion behind text

Why use

- Improve customer experience
- Drive profitability

Possible applications

- Customer service call routing
- Boost branding
- Drive automated action based on voice input

Speech recognition (NLP sub-discipline):
Uses computers to recognize and translate spoken language

Why use

- Improve customer experience
- Improve employee performance

Possible applications

- Voice assistants
- Automated subtitles for audio/video files
- In-car automation systems
- Dictation apps

Natural Language Understanding (NLU): using machines to process text in the form of human language, to understand its meaning and intent and to process that information to drive commands without the need for coding syntax

Why use

- Improve customer experience
- Improve employee performance

Possible applications

- Chatbots
- Product recommendation systems
- Contract analysis for errors or fraud

Natural Language Generation (NLG): The opposite of NLU. Uses AI to turn data into natural language. Systems take data and must decide how to put that information into words

Why use

- Improve customer experience
- Improve employee performance

Possible applications

- Populate structured documents like contracts or applications
- Chatbots
- Financial analysis
- Compliance documentation

Text Classification: Assigns tags or categories to text based on its content. May be an underlying process for sentiment analysis and NLU

Why use

- Improve customer experience
- Improve employee performance
- Drive profitability

Possible applications

- Applied to data sets that require categorization or grouping
- Support ticket organization
- Content aggregation or clustering
- Fraud and spam detection
- Language detection
- Text analysis
- Lead generation
- Recruitment

Optical Character Recognition (OCR): Translates text within images to machine readable text. OCR is not new but AI has improved it

Why use

- Improve customer experience
- Improve employee performance
- Drive profitability

Possible applications

- Convert paper documents to digital format
- Assist with translation, sentiment analysis, pattern recognition, security
- Invoice capture
- Bank deposits
- Language translation mobile apps
- Document digitization

Computer Vision and Intelligent Video Analytics use cases

Video analytics and surveillance

Automatically analyze video for facial recognition, queue monitoring, and unattended items. This can help detect events, uncover identity, monitor environments, and obtain operational insights for safety and quality assurance.

For example, these analytics may:

- Increase customer insight (real-time video analytics)
- Shorten wait times (queue monitoring)
- Increase building or facility security—and safety

Ideal for: Manufacturing, retail, airports, schools, governments/smart cities

Image Classification: Groups images into classes to predict what an image represents

Why use

- Improve customer experience
- Improve employee performance
- Drive profitability

Possible applications

- Group image results found by a web search engine
- Diagnose diseases or conditions (e.g. MRI scan analysis)
- Classify large visual databases

Object Detection: Identify objects and their locations within an image or video

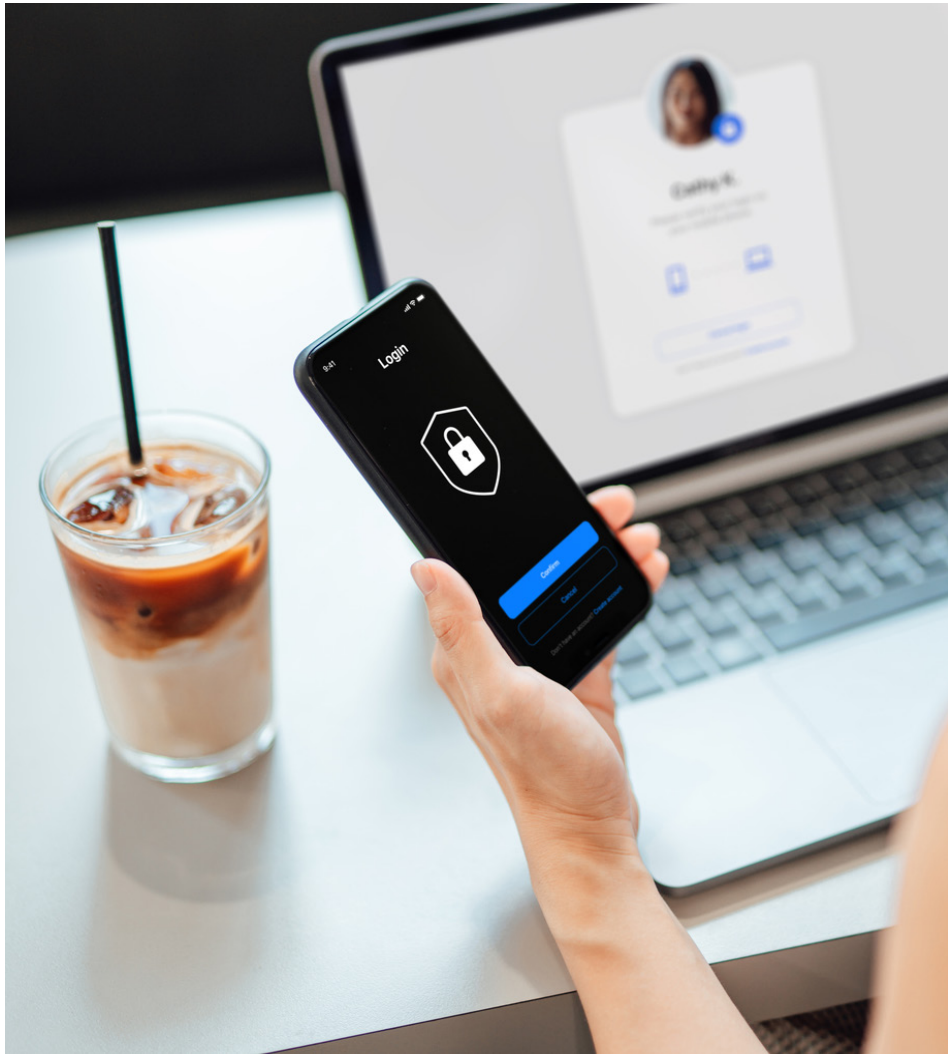
Why use

- Improve customer experience
- Drive profitability
- Improve safety / security

Possible applications

- Automated checkout
- Security surveillance
- Traffic regulation
- Patient monitoring

Facial Recognition: Using AI to identify or verify a person from a digital image or a video frame. Works generally by comparing selected facial features from given images with faces within a database



Why use

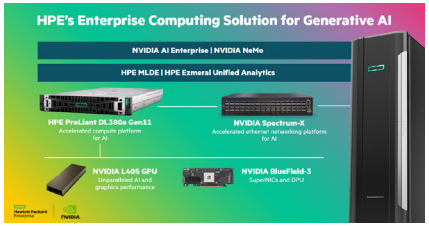
- Improve customer experience
- Drive profitability

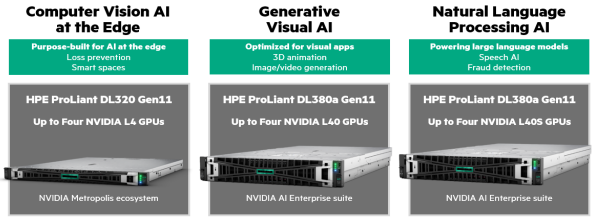
Possible applications

- Surveillance
- Criminal databases
- Marketing
- Social media
- Finance
- Retail
- Phone unlocking
- Auto-tagging videos in social media
- Automated check-in at events

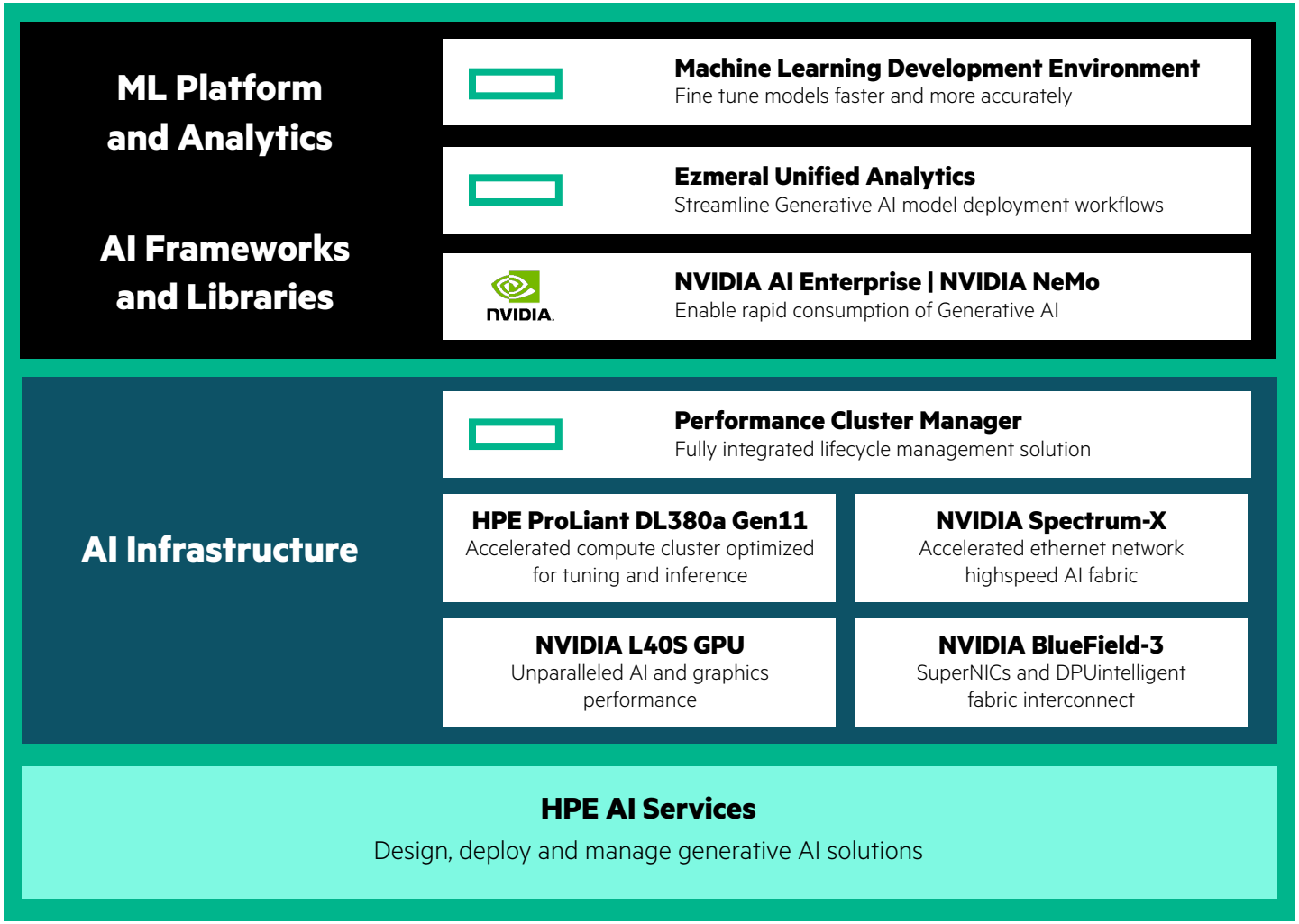
HPE Compute portfolio enables Generative AI adoption

1. Foundation model (Training)	2. Customization (Fine-tuning)	3. Inference (Production, RAG)
Time to Results: hours, days, weeks	**Time to Results:** minutes, hours, days	**Time to Results:** microseconds, milliseconds, seconds
$$$	$$	$
Time to Deploy: 6+ months	**Time to Deploy:** weeks to months	**Time to Deploy:** weeks
AI specialists | **Hyperscalers** | **Large enterprises**	**Typically performed by Enterprise**	**Run by Enterprise**
HPE Supercomputer e.g. HPE Cray	**HPE's Enterprise Computing Solution for Generative AI**	**HPE Compute Inferencing Solutions** (DL320, DL380a, DL385)





HPE's Enterprise Computing Solution For Generative AI



Accelerated compute cluster nodes for GenAI

16X

HPE ProLiant DL380a Gen11 Servers

Each Server

192GB
4x NVIDIA L40S GPUs

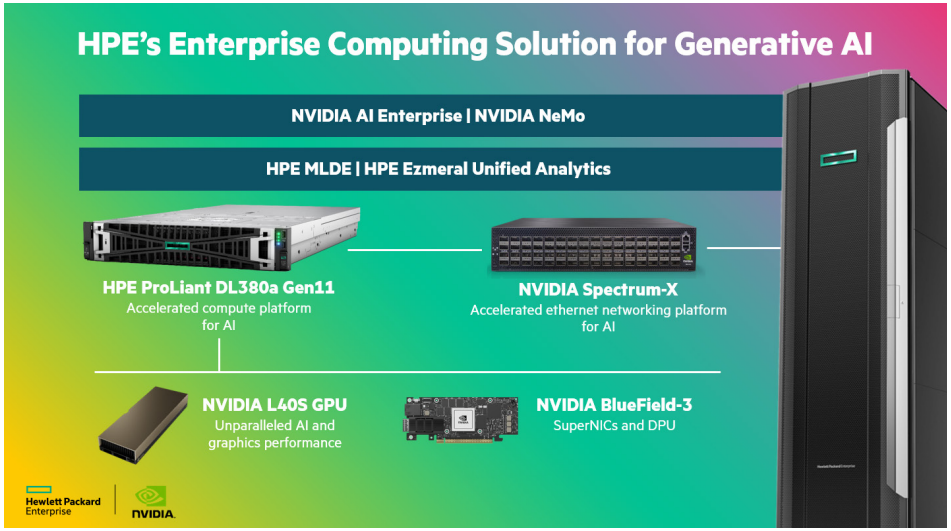
2 x 400G
NVIDIA BlueField-3 SuperNICs

2 x 200G
NVIDIA BlueField-3 DPUs

Dual
5th Gen Intel Xeon Scalable Processor

How does HPE's solution address customer Gen AI challenges?

- Easy to order in a single SKU - rapid time-to-value, factory built and delivered as an **integrated end-to-end** solution - removes guess work, **reduces time and cost**, limits risk.
- HPE & NVIDIA software work together in concert to ease connection to large language model (LLM), maintain data integrity and security, **accelerate deployment** and optimize management.
- HPE experts help design, implement and manage for customers with limited AI resources in-house to **address skill and knowledge** gaps



Key segment: Financial Services customers

Sample use cases:

- **Risk mitigation –** AI can help analyse vast datasets to identify patterns, anomalies and potential risks, improving decision making and predicting market fluctuations
- **Fraud detection –** AI can recognise unusual transaction patterns in real time based on historical data, and learn from new data to stay ahead of evolving fraud tactics
- **Customer service –** AI can streamline basic customer services processes, preserving personnel for complex cases

AI portfolio

	Computer vision AI at the edge	Generative visual AI	Natural language processing AI	HPE's enterprise computing solution for Generative AI
Smart template UCID / Tracking SKU	5134233671-03	5133904640-02	5133904763-02	S3P88A
Server	**HPE ProLiant DL320 Gen11**	**HPE ProLiant DL380a Gen11**	**HPE ProLiant DL380a Gen11**	**HPE ProLiant DL380a Gen11 (x16)**
GPU	Up to four NVIDIA L4 (Up to three today, four in October*)	Up to four NVIDIA L40	Up to four NVIDIA L40S	Four NVIDIA L40S per server
Use cases	• Computer vision & video analytics • Loss prevention • Smart spaces	• 3D animation • Image generation • Video generation	• Speech AI • Fraud detection • Predictive maintenance	• Generative AI • Tuning • Inference
Software	ISVs enabled by NVIDIA Metropolis	NVIDIA AI Enterprise	NVIDIA AI Enterprise	NVIDIA AI Enterprise NVIDIA NeMo
Description	**Purpose-built for AI at the edge** Ideal for computer vision and video analytics at the edge for customers in the retail, hospitality, and manufacturing sectors, the HPE ProLiant DL320 Gen11 server has a unique compact design purpose-built for edge computing. It can pack up to four NVIDIA L4 GPUs in a 1U form factor to power smart spaces and loss prevention solutions and deliver insights in near-real time. Customers can take advantage of offerings from NVIDIA's Metropolis ecosystem to deploy targeted solutions for their industry and use case.	**Optimized for visual apps** For generative visual AI to drive product design, 3D animation, or image and video generation, the HPE ProLiant DL380a Gen11 server with four NVIDIA L40 GPUs delivers the rendering and design performance needed by demanding visual applications used in the media and entertainment, healthcare, and manufacturing industries. The NVIDIA AI Enterprise suite offers pre-trained models to streamline development and deployment.	**Powering large language models** To drive enterprise implementations of natural language processing, such as speech AI and fraud detection, HPE designed the ultra-scalable HPE ProLiant DL380a Gen11 server. Together with NVIDIA L40S GPUs and the full suite of NVIDIA AI Enterprise tools, this is a great AI platform for running large language models used by the financial services and manufacturing sectors, as well as customer service across a range of industries.	**Simpler, faster path for the generative AI-powered Enterprise** HPE's enterprise computing solution for Generative AI delivers an AI-native architecture that is purpose-built to deliver the speed, scale, and control needed for AI production applications that streamline business processes and foster innovation. Optimized for tuning and inference, a rack-scale architecture features market leading HPE ProLiant DL380a Gen11 compute with NVIDIA L40S GPUs. High-performance cloud-scale networking purpose built for AI with NVIDIA BlueField-3 SuperNICs and DPUs, and the NVIDIA Spectrum-X Ethernet Networking Platform.
Key verticals & functions	Retail, hospitality, manufacturing	Media & entertainment, healthcare, manufacturing	Financial services, manufacturing, customer service	Financial Services, Health and Life Sciences, Cross Vertical (customer service and code generation)

Understand

Prepare yourself for the conversation

1. Prepare

- Understand the workload, particularly the value proposition
- Identify stakeholders and key users

2. Have discovery meeting

- Determine jobs to be done
- Identify technology environment
- Understand timeline for implementation
- Confirm decision criteria
- Assess competitive positioning

3. Assess opportunity after the meeting

Resources

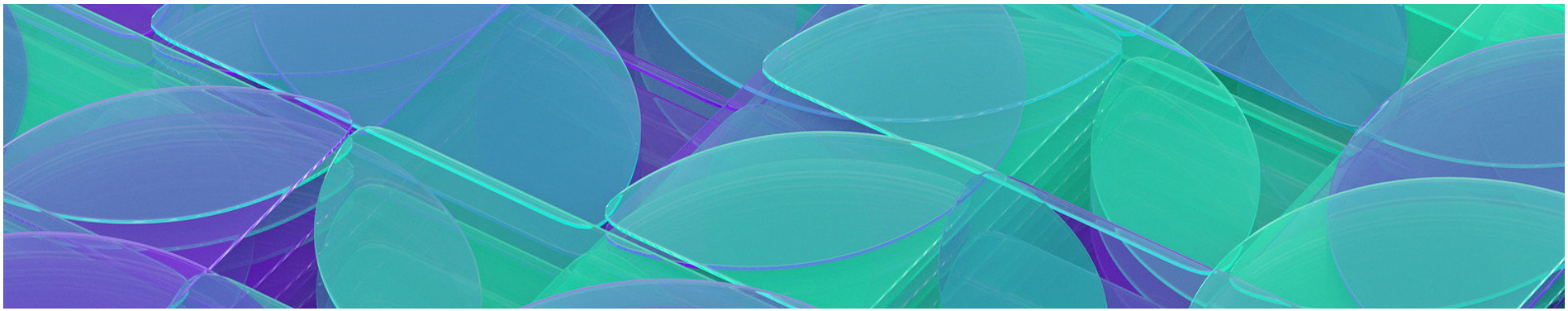
Toolbox building blocks

- HPE Core Compute - Briefcase
- HPE Services – Briefcase
- Perfect Attach from HPE Services Briefcase

Seismic

- HPE Compute AI Solutions Briefcase
- HPE AI Software – Briefcase
- Customer Presentation for HPE's enterprise computing solution for Generative AI
- Solution Overview for HPE's enterprise computing solution for Generative AI
- Technical guide for HPE's enterprise computing solution for Generative AI

Validate



Validate the customer opportunity

- Identify the customers and their challenges and triggers
- Prepare qualifying questions

Opportunity if:

- Interested in AI Inferencing or light modeling at the edge
- Wants to reduce costs in MLOps

Resources

- HPE Compute AI Solutions Briefcase
- Customer Presentation for HPE's enterprise computing solution for Generative AI
- Solution Overview for HPE's enterprise computing solution for Generative AI
- Technical guide for HPE's enterprise computing solution for Generative AI

Qualify

Qualify the customer

1. **Conduct a stakeholder meeting**
   - Determine business need and jobs to be done
   - Have a conversation using qualifying questions
2. **Align on what a proposal should include**
3. **Identify the decision maker**
3. **Determine a mutual plan to proceed**

Resources

- HPE Compute AI Solutions Briefcase
- Customer Presentation for HPE's enterprise computing solution for Generative AI
- Solution Overview for HPE's enterprise computing solution for Generative AI
- Technical guide for HPE's enterprise computing solution for Generative AI

Other Resources

- HPE Reference Architecture library

Propose

Develop a draft of the impact proposal

- Customize Business case—Desired business outcomes for customer
- Reference the value proposition
- Implementation plan / timeline—Time to initial impact
- Leverage unsolicited proposals

Collaborate and Engage with services

- HPE Services
- HPE Financial Services

Present crafted proposal

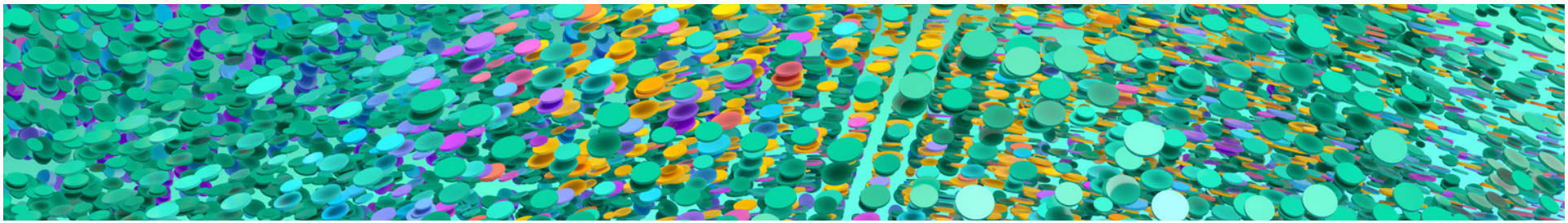
Resources

- HPE Compute AI Solutions Briefcase
- Customer Presentation for HPE's enterprise computing solution for Generative AI
- Solution Overview for HPE's enterprise computing solution for Generative AI
- Technical guide for HPE's enterprise computing solution for Generative AI

Other Resources

- HPE Reference Architecture library
- OCA
- iQuote

Close



Negotiate and close the deal

Deliver the proposal and handle objections

- Configure and quote with SSET (if supported solution)
- Use OCA configs (UCID#s and instructions)
- Be sure to include any promos
- Anticipate and handle objections including competitive information
- Attach HPE Financial Services offers

Resources

- HPE Compute AI Solutions Briefcase
- Customer Presentation for HPE's enterprise computing solution for Generative AI
- Solution Overview for HPE's enterprise computing solution for Generative AI
- Technical guide for HPE's enterprise computing solution for Generative AI

Resources

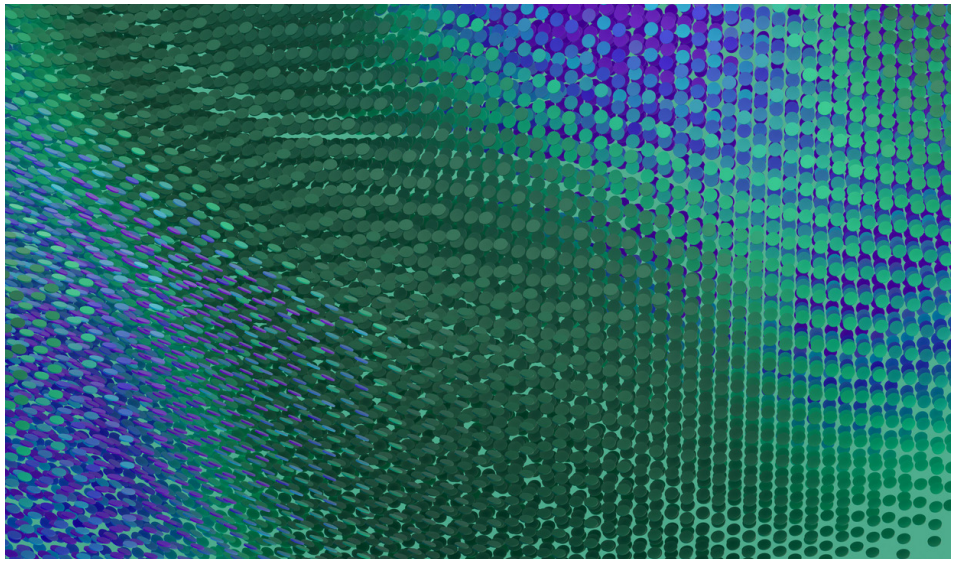
- HPE Reference Architecture library
- OCA
- iQuote

Win

Engage the customer to be a reference

Sales resources

- HPE Customer Success Stories
- Reference Success Center (HPE Internal Only)
- HPE Compute Customer References Briefcase

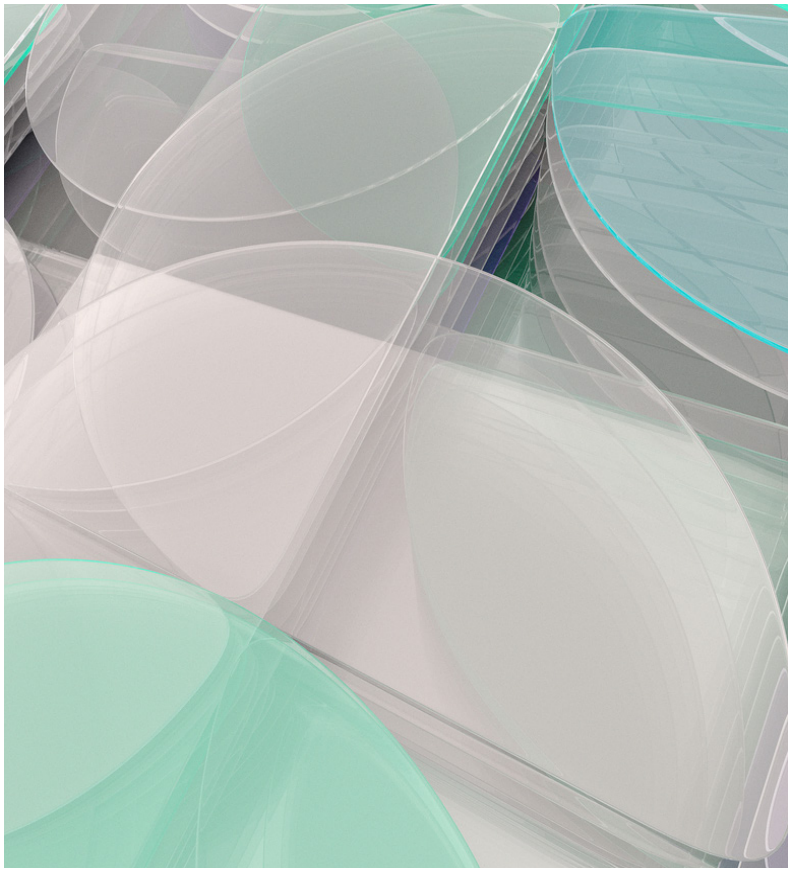


Key benefits

HPE GreenLake delivers a pay-per-use high-performance compute experience on-premises. You can design your infrastructure solutions, selecting from a broad range of HPE and partner technologies, as well as optional services that can span your infrastructure, apps, and workloads. HPE GreenLake gives your customer supercomputing access to power AI at scale.

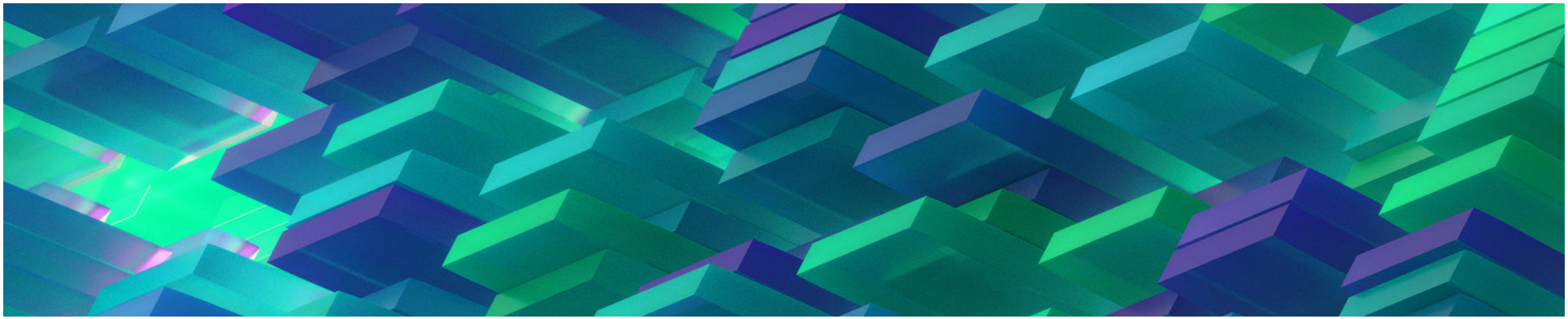
HPE Services helps your customer explore, experiment and evolve their data and AI projects, leveraging the right expertise, tools, technology and partnerships. HPE AI Services experts can help you plan, implement, deploy and scale your AI solutions, integrating and enhancing offerings from multiple vendors to build and deploy a unique solution to meet your desired outcome.

HPE Financial Services combines technology insights, financial expertise, and a deep-rooted focus on sustainability to create smarter IT lifecycles for customers and partners of all sizes. Working across the entire tech estate, from edge to cloud to end-user, our collaborative approach delivers asset management solutions that not only free up capital and maximize capacity, but also advance sustainable practices globally and consistently.



HPE GreenLake

HPE GreenLake provides an edge-to-cloud platform. That cloud platform comes to you, wherever your apps and data live.



Accelerate outcomes in four ways

Gain self-service agility
Easily deploy resources, view your costs, and forecast capacity – all from one intuitive platform: HPE GreenLake Central.

Flex with pay-per-use
Avoid heavy upfront costs and expensive overprovisioning and only pay for what you use.

Scale up and down
Reduce your worry and your costs with scalable capacity that's ready when you need it.

Managed for you
Offload the burden of operating IT and free up resources with fully-managed cloud services.

HPE Services

From planning to operating and beyond, our experts help you accelerate edge to cloud transformation, optimize operations, and maximize IT investments.

Strategy and design

- Lead customers to the development of an AI strategy when embarking on their AI journey

Integration and deployment

- Deploy custom data transcoding and visualization, software configuration, app store, and image development
- Deployment and integration services mitigate the deployment and project risks, as well as simplify operations of the solution and accelerate time to production

Operational services

- Manage and support the hardware and software environment including lifecycle management, and can be offered as-a-service
- Software support with 24/7 and 9 to 5 support options

How does HPE Services help?

Accelerate outcomes: Proven expertise to help expedite strategy, design, implementation and ongoing support of HPE Compute for AI solutions.

Remove project risks: Expert assistance to help with deployment and operations of new HPE Compute for AI environments environments

Simplify IT&O for data teams: Complement the IT team with globally available advisory and professional services assistance to validate the customer's HPE Compute for AI deployments so that they are fully commissioned and operational

Apply enterprise-grade capabilities: Fast and easy deployment and management of K8s clusters with enterprise security and authentication. Operates as hybrid from edge-to-cloud. Offers a flexible, multi-user, multi-tenant control plane to deploy on-premises in the cloud or at the edge.

Uses pre-integrated data fabric: Pre-integrated, scale out, edge ready persistent storage with a fully integrated data fabric that is 100% open-source K8s. HPE innovations like KubeDirector provide a controller to deploy non-cloud native apps.

Connectivity to external data: Connectivity to external data without copying data and use of HPE GreenLake.

HPE financial services

Aligning business goals with tech lifecycles to accelerate innovation.

Smarter economics for your project needs
Whether an AI project or transition to an as-a-service model, like HPE GreenLake, HPEFS can help your customers more easily align costs with deployment and ensure success with HPE Risk Mitigation Services.

More easily align costs with deployment
Meet short-term needs and ensure optimal capacity without a long holding period.

Lower risk associated with new projects
Have the flexibility to reduce tech investment based on pre-determined project milestones.

Simplify transition from proof of concept to full production
Establish a smarter IT lifecycle process that efficiently bridges infrastructures while controlling costs.

Learn more: HPE Risk Mitigation Services

Additional HPEFS solutions designed to create smarter economics, innovation and IT sustainability:

HPE Asset Upcycling Services
Recover value from workplace or data center tech securely with globally consistent and environmentally responsible services.

HPE Accelerated Migration Services
Convert existing, owned IT assets into an incremental cash source while converting use of your systems to an HPE GreenLake as-a-service model.

HPE Extended Deployment Services
Secure critical technology while aligning payments to deployment schedule.

Resources

Field enablement resources

- HPE ProLiant Servers Portfolio - Briefcase
- HPE Reference Architecture library
- HPE Services – Briefcase
- Perfect Attach from HPE Services Briefcase
- HPE Compute AI Solutions Briefcase
- Sales Summary Sheet - HPE's enterprise computing solution for Generative AI
- FAQ - HPE's enterprise computing solution for Generative AI – for sales and partners
- HPE Financial Services – IT investment - Briefcase

Customer resources

- HPE AI Solutions - Get Started with AI
- Customer Presentation for HPE's enterprise computing solution for Generative AI
- Solution Overview for HPE's enterprise computing solution for Generative AI
- Technical guide for HPE's enterprise computing solution for Generative AI
- HPE's enterprise computing solution for Generative AI
- Video: Streamline Generative AI production with HPE's enterprise computing solution for Generative AI
- HPE AI Services web page

a00118909enw

Hewlett Packard Enterprise