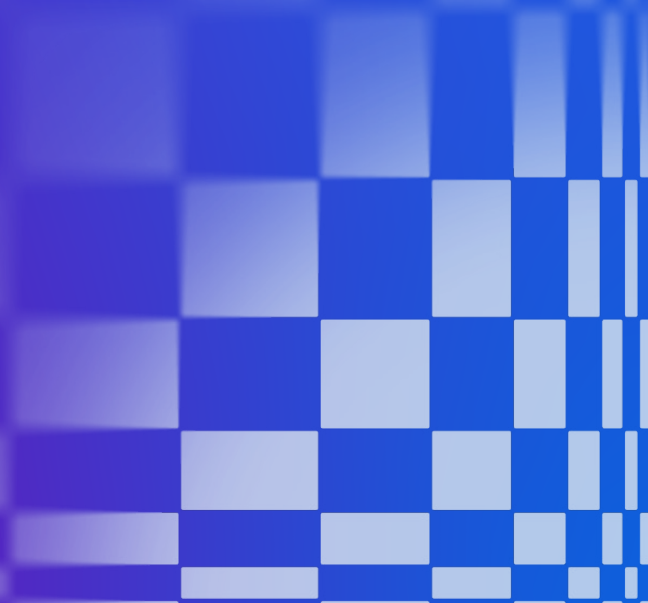


HCLTech | intel.

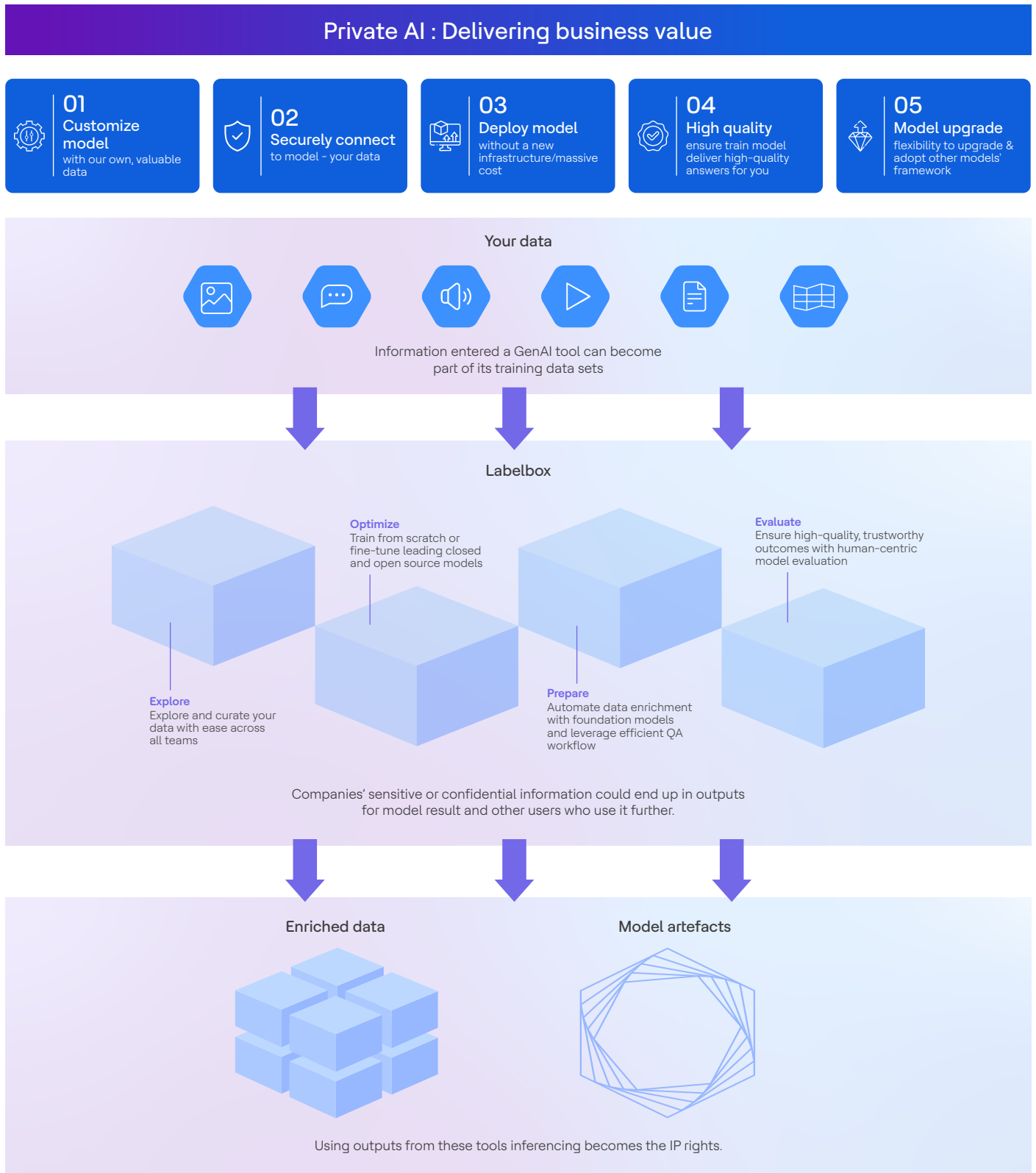
VMware Private AI solution powered by Intel and HCLTech

For enhanced security and performance



Overview

VMware Private AI as a Service on Intel is designed to deliver business value by enabling enterprises to securely connect and deploy AI models without requiring massive infrastructure investments. With a strong focus on customization, organizations can leverage their own valuable data while ensuring that high-quality AI models deliver accurate results. This solution provides flexibility to upgrade and adopt various AI frameworks while addressing critical concerns such as security, compliance and responsible AI governance. As enterprises navigate the rising demand for GenAI, this approach optimizes AI models for CPUs, enhances interoperability across VM and container platforms and supports a scalable, cost-efficient infrastructure.



Market trends

82%

Organizations seek to leverage their proprietary data for AI-driven applications, highlighting the importance of private AI solutions that ensure data security and compliance

\$280 billion



Rising demand for GenAI products is projected to contribute approximately \$280 billion to IT revenue, with the market expected to grow at a CAGR of 42% over the next decade

\$4.4 trillion



GenAI is forecasted to deliver up to \$4.4 trillion in annual economic value across industries, showcasing its transformative impact on global enterprises

42%

The GenAI market is anticipated to grow at an annual CAGR of 42% over the next decade, reinforcing its rapid adoption across industries

36.2%

The AI market is projected to grow at a CAGR of 36.2%, with businesses prioritizing the adoption of AI-enhanced technologies to drive innovation and efficiency

\$407.0 billion



The global AI market is expected to reach \$407.0 billion by 2027, driven by increasing investments in AI applications across multiple sectors

Adoption challenges



Responsible AI and governance:

Ensuring ethical AI governance (HIPAA, GDPR, CCPA) remains a challenge, requiring transparent, accountable AI practices



Infrastructure complexity:

AI deployment demands a hardware-agnostic environment capable of supporting diverse training and production models, increasing the need for flexible IT infrastructure



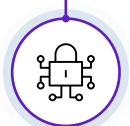
Scalability and interoperability:

Organizations must design AI systems that can scale seamlessly while maintaining interoperability across VM and container platforms to support diverse workloads



Cost optimization:

Enterprises face challenges in efficiently allocating resources, requiring phase-wise AI investments and infrastructure planning to balance performance with costs



Security and compliance:

Securing AI implementations involves robust encryption, access controls and auditing to mitigate risks associated with data privacy and compliance

Private AI Service offering journey

Need analysis

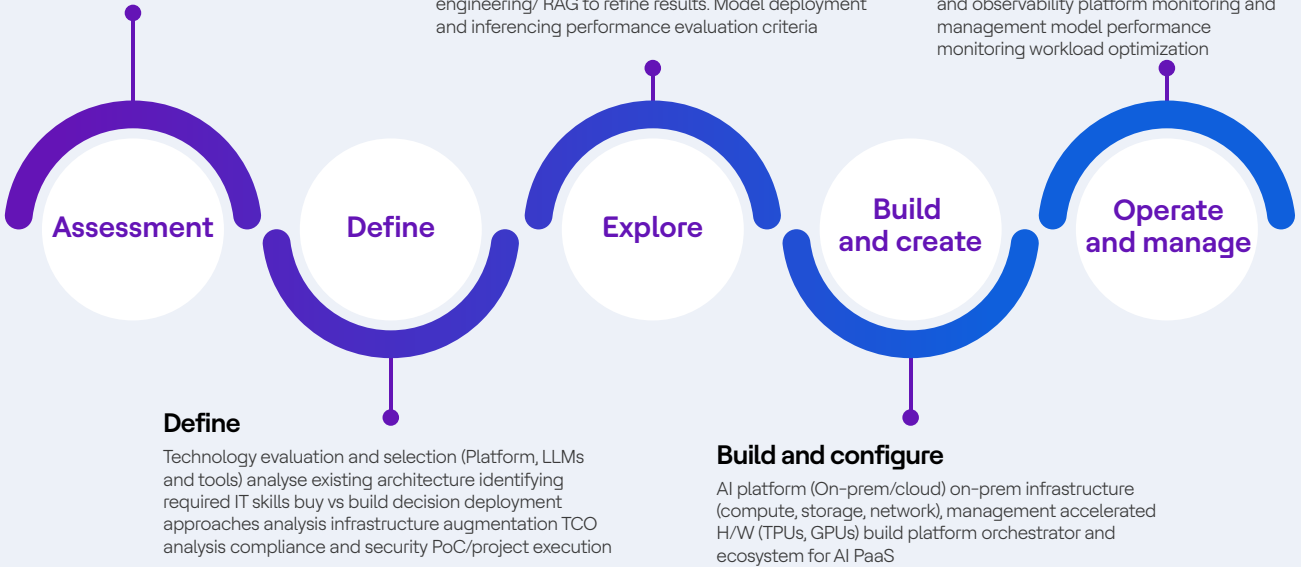
Use case identification feasibility evaluation As-Is infrastructure assessment Ideal future state assessment

Explore

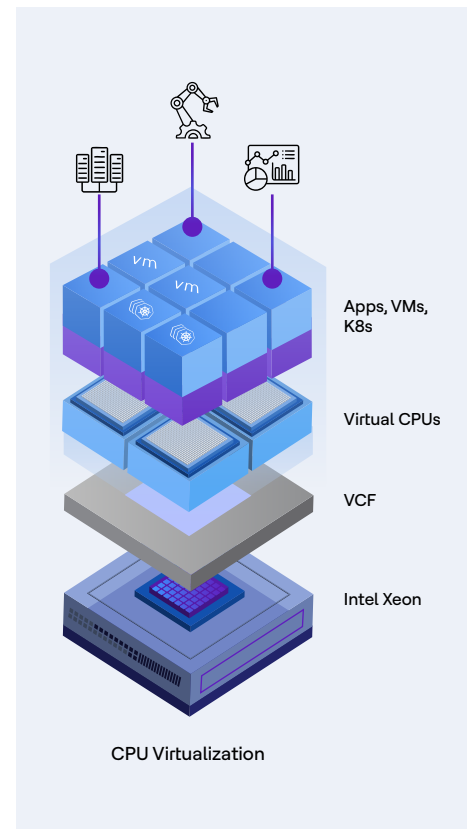
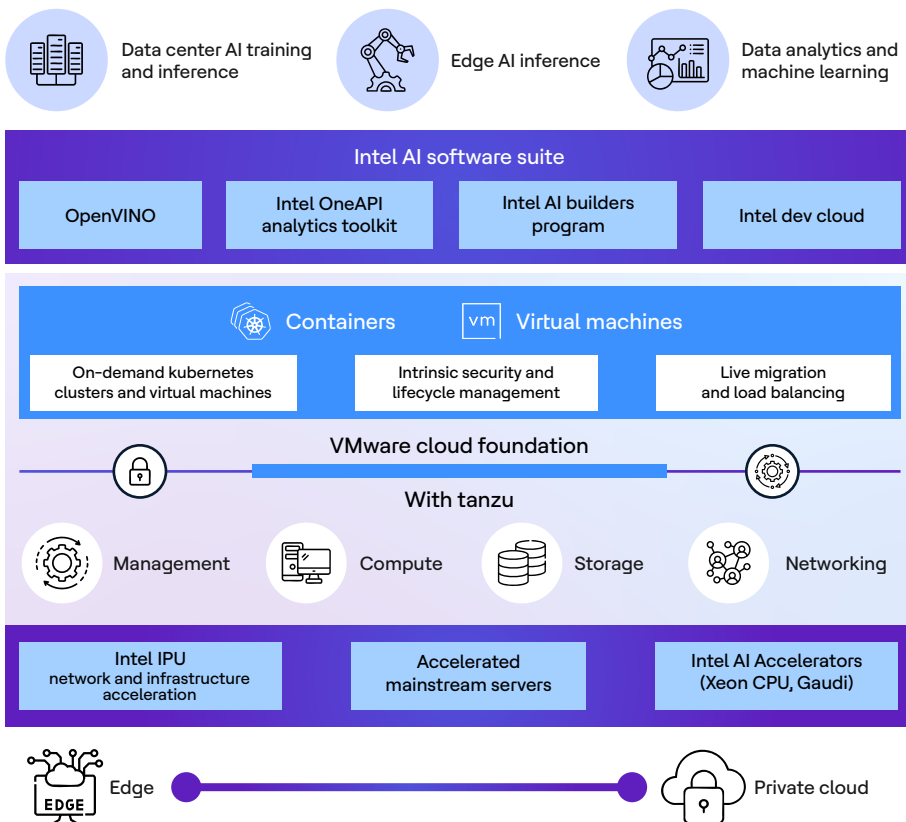
Infra and platform identifications architecture design. provisioning of platform services customized day-2 operations design. Need of fine tuning/ prompt engineering/ RAG to refine results. Model deployment and inferencing performance evaluation criteria

Operate and manage

Infra life cycle management HA and performance monitoring MLOPs observability unified IT infrastructure and cloud monitoring and observability platform monitoring and management model performance monitoring workload optimization



VMware Private AI as a Service integrates SAP AI Agents (Joule Agents) with Intel-optimized AI frameworks to enable secure, scalable AI deployment. Leveraging VMware's multicloud infrastructure, it ensures seamless interoperability across VMs and containers, while Intel's OpenVINO and hardware accelerations optimize AI workloads for edge, cloud and on-prem environments.



Model-based use cases



Video generation

AI-driven tools automate video creation, from scripting to editing, optimizing resources for media and marketing industries



Conversational AI

LLMs power virtual assistants and chatbots, enabling human-like interactions across customer service and enterprise applications



Intelligent process automation

Automates repetitive tasks like data entry and invoicing, enhancing operational efficiency



Code and app builders

AI accelerates software development by automating coding, testing and deployment processes



Synthetic data and research

LLMs generate insights from large datasets while preserving data privacy through synthetic data creation



Personalization

AI tailors content and recommendations based on user preferences, improving customer engagement



Speech synthesis

Convert text-to-speech (TTS) for voice assistants and interactive AI services



Image generation

AI models like GANs create realistic images and artwork for various applications



Generative music

AI assists musicians by composing and enhancing music creation



Generative design

AI automates 3D design, simulation and visualization, optimizing design constraints

Ready-to-use demo: Industry applications of Private AI



Medical imaging

AI-driven 3D imaging enables accurate segmentation analysis across anatomies using transformer-based models



Patient assistance

AI-powered virtual assistants help patients navigate healthcare centers and provide first-aid guidance using BioBERT



Route planning

Optimizes logistics with dynamic rerouting, load balancing and robotic simulations, improving efficiency and reducing costs



Customer behavior analysis

AI models analyze videos to recognize human activity, product interest and recommendations



Content personalization

AI transforms images and videos into high-resolution, photorealistic outputs for enhanced media customization



Code generation

Assists developers by automating code completion, bug detection and resolution with foundation models



Character development

Multi-sensor AI processes video, image and audio data to generate actionable insights for media applications



Contact center support

AI-powered automation helps L1 agents resolve queries efficiently without escalating to L2 resources

Benefits



Privacy and security

Prevents data leakage, maintains IP rights and ensures compliance with Intel TDX and SGX for secure data usage



Ethical and responsible AI

Designed for fairness, transparency and accountability in AI practices



Flexible deployment

Intel-optimized AI models run on diverse computing infrastructures with L3 and L4 support



Edge AI optimization

OpenVINO enables efficient deep learning on edge devices, reducing reliance on GPUs



Ease of operation:

Pre-integrated Intel and VMware stack with automated management for simplified AI deployment



Multi-model and distributed training

Supports fine-tuned foundation models, third-party AI models and open-source AI frameworks



Leverage existing investments

Utilizes VMware software, Intel AI infrastructure and optimized compute resources for cost-effective scaling



Open standards in programming

The programming model is open and it promotes the use of open standards and frameworks

Why HCLTech



AI transformation and deployment

- Demo-ready use cases, proofs of concept and phased AI enablement
- Integration of traditional and generative AI with responsible, licensed AI adoption
- Best industry practices for AI transformation



Robust infrastructure and security

- Pre-integrated, validated stack ensuring production-ready deployment
- Scalable infrastructure, cost optimization and HCLTech benchmarking
- Strong focus on data privacy and security



Expertise and industry collaboration

- A dedicated center of excellence and innovation lab for GenAI adoption
- Partnerships with leading technology providers
- Skilled, industry-certified professionals

For more information, visit our [page](#)

HCLTech | Supercharging
Progress™