

2H 2025 AI Platforms Market Sizing & Five-Year Forecast

A Global Study of the Key Market Segments Within AI Platforms

Analyst(s): Nick Patience, Varun Shitole

Published: October 2025

Document #: 2H25MSFRAIPNPVS202510

This written report serves as a “highlights reel” of findings from our study of the key product segments of AI Platforms.

This is a “companion” report to the entire dataset within Futurum’s interactive visualization dashboard, where users can filter and drill down further into the study details - see app.futurumgroup.com.

Key Findings

- **Explosive Market Expansion:** The AI Platforms market scales from \$24.9B in 2024 to \$292B by 2030 (approximately 50.8% CAGR), reflecting enterprise mainstreaming of model training, inference, and agentic application enablement across all verticals.
- **Broad-Based Yet Concentrated Growth:** Hyperscalers (Microsoft, AWS, Google Cloud) capture most early revenue, but independent vendors (Databricks, Snowflake, Anthropic, OpenAI, Vector DB specialists) expand rapidly through mid-decade before consolidation tempers share diversity.
- **Inference Overtakes Training:** By 2026, inference workloads such as copilots, agents, and real-time assistants drive the majority of revenue growth, while training growth moderates as cost curves flatten and model reuse rises.
- **Hybrid and Edge Shift:** Privacy, latency, and efficiency considerations fuel a hybrid and edge resurgence between 2026 and 2028, enabled by small language model deployment and localized inference at or near data sources.
- **Regulation, Talent, and Power Constraints:** Divergent regional regulation (EU AI Act versus US Executive Orders), AI skills scarcity, and data center and power bottlenecks emerge as systemic constraints that will shape late decade growth trajectories.

Recommendations

1. **Design for Cost Elasticity and Performance Transparency:** Publish clear cost per token or parameter benchmarks and optimize both training and inference economics to win enterprise workloads facing strict ROI scrutiny.
2. **Embed Governance and Compliance by Default:** Integrate model monitoring, bias detection, and auditability features as table stakes. Early compliance alignment with the US, EU, and China rules will become a decisive differentiator in regulated verticals.
3. **Build Toward Hybrid and Edge AI:** Extend offerings beyond cloud APIs to support on device and federated deployments. Vendors that provide flexible Public, Private, and Hybrid architectures will capture the next wave of enterprise demand.
4. **Leverage Data Gravity and Ecosystem Lock-in:** Strengthen integration with proprietary enterprise data such as feature stores, vector databases, and CRM or ERP systems. Owning the data to model pipeline is the most durable moat against commoditized APIs.
5. **Cultivate Developer and Partner Ecosystems:** Provide open SDKs, revenue-sharing marketplaces, and agent framework APIs that let third parties build and monetize on your stack, converting platform usage into sustained network effects.

TABLE OF CONTENTS

Key Findings.....	1
Recommendations.....	2
Market Model.....	4
Vendor Landscape.....	4
Market Segments.....	5
Industries.....	5
Deployment Models.....	6
Use Cases.....	6
Regions.....	8
Five-Year Forecast for AI Platforms Market (2025–2030).....	8
Forecast Details, by Market Segment.....	9
Forecast Details, by Vendor.....	9
Forecast Details, by Segment.....	12
Forecast Details, by Industry.....	14
Forecast Details, by Deployment.....	17
Forecast Details, by Region.....	19
Forecast Details, by Use Case.....	20
Research Background and Methodology.....	23
About Us.....	25
About the Authors.....	25
About The Futurum Group.....	25
Copyright & Use License.....	25
Copyright Notice.....	25
License Notice.....	25
Limitation of Liability Notice.....	25

Market Model

Futurum views the AI Platforms market model through the following lens:

Software and managed services that help organizations build, train, deploy, and run AI systems. Not included: hardware (e.g. GPUs/servers) or finished business apps that just happen to use AI.

How the Data Is Organized: We view the market from four angles. Each year, all four add up to the same locked total in our Base forecast (so numbers stay consistent).

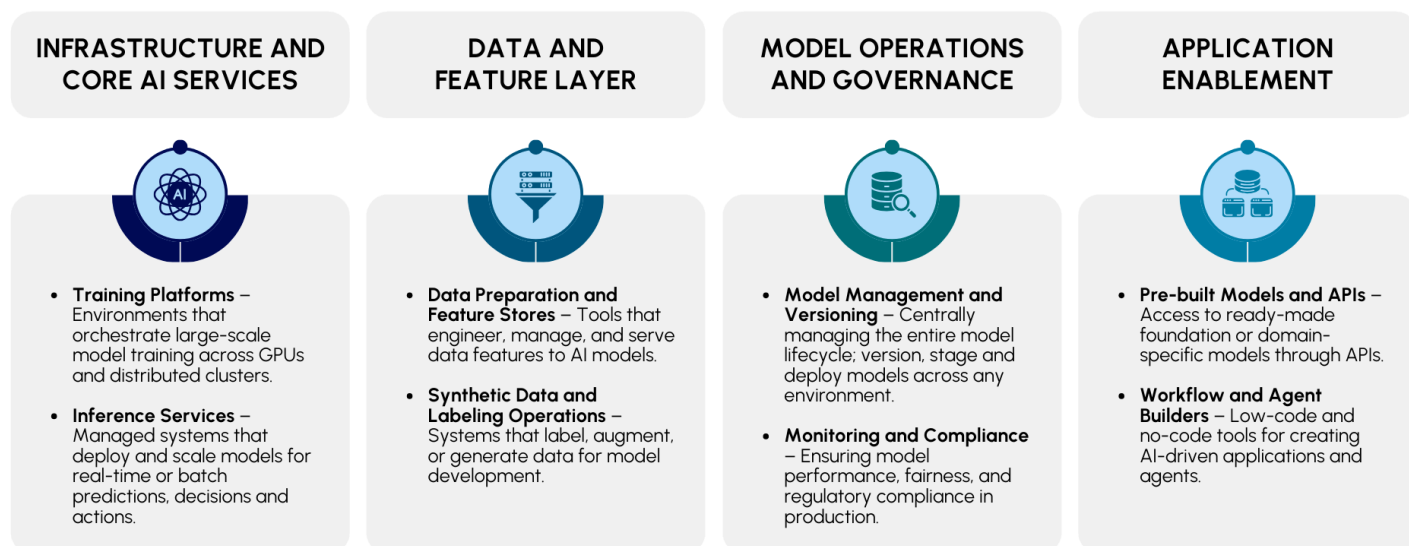
- By Vendor (which companies earn the revenue)
- By Segment (what part of the AI stack the spending is for)
- By Industry (which customer industries buy it)
- By Deployment (where it runs—cloud, private, hybrid, or colo)

We also show three overall scenarios (Bull, Base, Bear) as separate top-line paths—not to be added together.

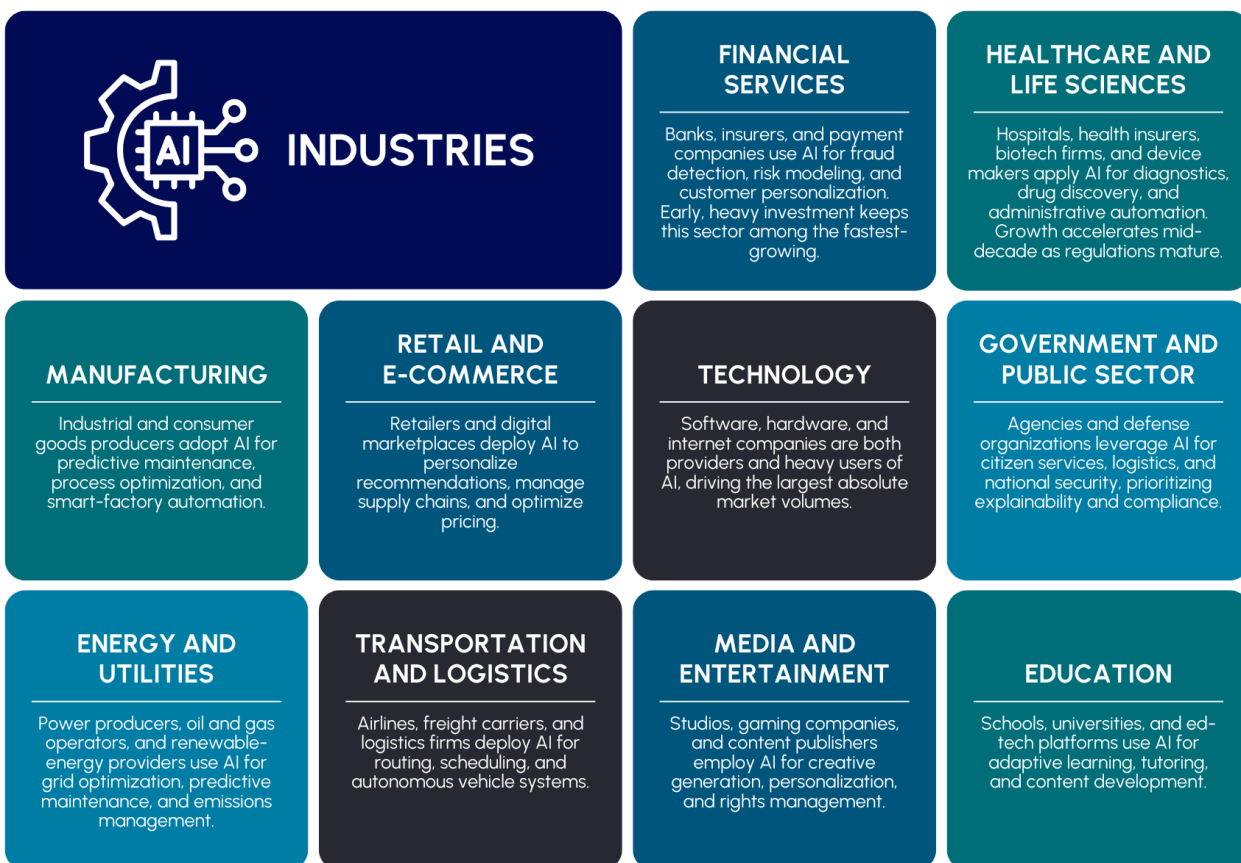
Vendor Landscape

Category	Example Vendors	Description
Hyperscale Cloud Providers	Amazon Web Services, Microsoft Azure, Google Cloud	Offer end-to-end AI infrastructure and APIs
Enterprise AI Platforms	IBM watsonx, Oracle Cloud, SAP AI Core, Salesforce Einstein, ServiceNow Now Platform	Deliver integrated AI capabilities within enterprise ecosystems
Data and Feature Platforms	Databricks, Snowflake, Palantir	Provide unified environments for analytics, features, and model deployment
ModelOps and Governance Specialists	C3 AI, Fiddler, Arize, Weights & Biases	Focus on model lifecycle management, monitoring, and compliance
Foundation and Model API Providers	OpenAI, Anthropic, Cohere, Mistral, AI21 Labs	Supply general and domain-specific models via API
Vector Database and Retrieval Platforms	Pinecone, Weaviate, Zilliz, Elastic, MongoDB	Enable retrieval-augmented generation and vector search
Emerging Builders and Orchestration Tools	LangChain, LlamaIndex, Modal, Together AI	Provide frameworks and infrastructure for custom AI applications

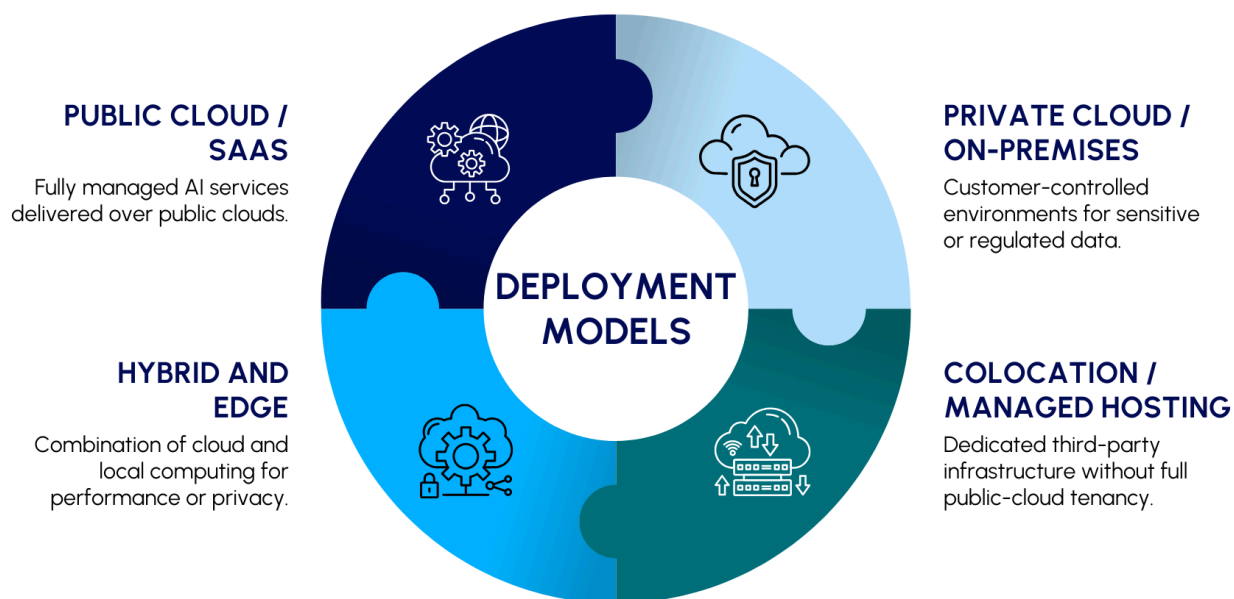
Market Segments



Industries



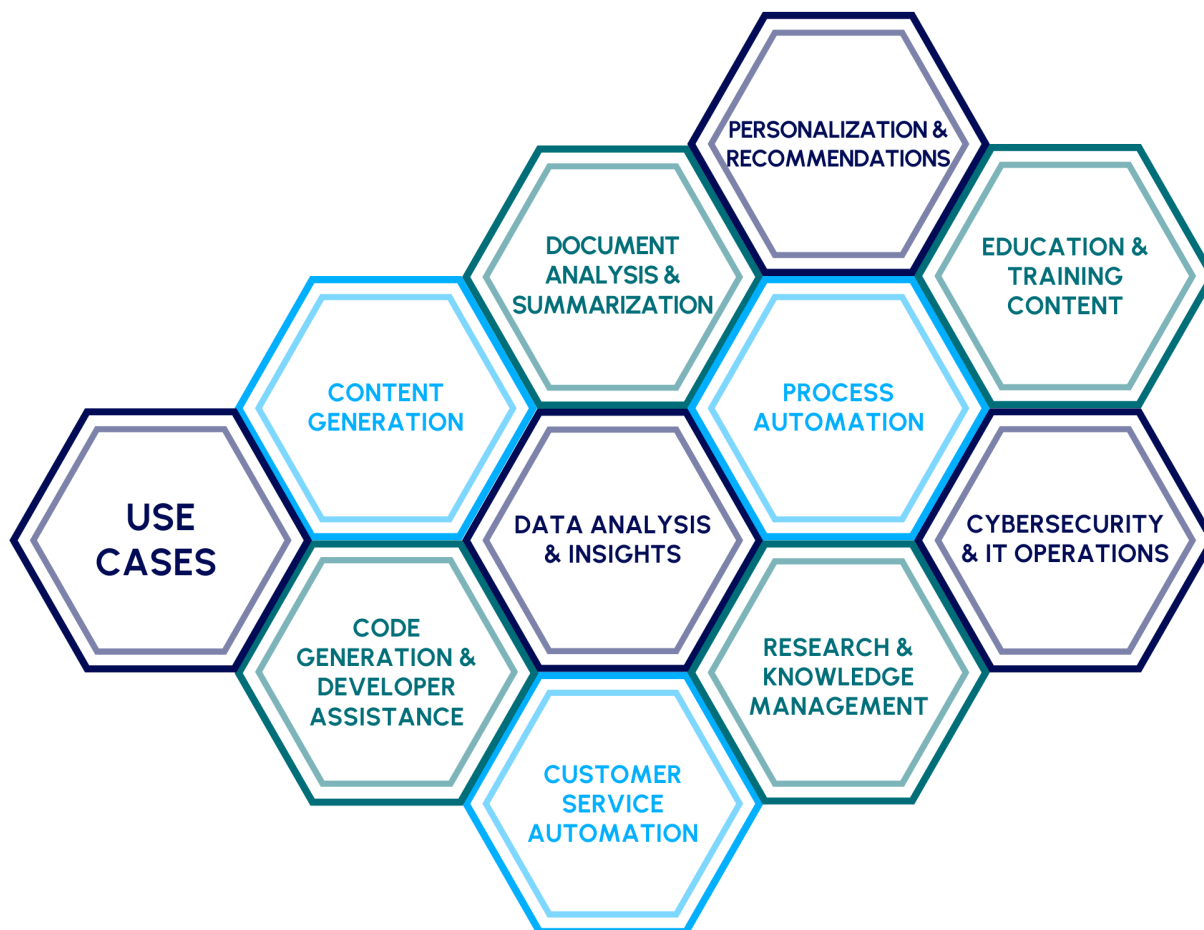
Deployment Models



Different organizations choose deployment options based on security, latency, and control requirements:

Deployment Model	Description	Example Scenarios
Public Cloud/SaaS	Fully managed AI services delivered over public clouds	Most new enterprise workloads through 2025
Private Cloud/On-premises	Customer-controlled environments for sensitive or regulated data	Finance, healthcare, and government sectors
Hybrid and Edge	Combination of cloud and local computing for performance or privacy	On-device inference, industrial IoT, and 5G use cases
Colocation/Managed Hosting	Dedicated third-party infrastructure without full public-cloud tenancy	High-performance computing or specialized GPU hosting

Use Cases



- Content Generation – Creating and transforming text, images, audio, or video for marketing, media, and design.
- Code Generation and Developer Assistance – Helping engineers write, debug, and modernize code within integrated development environments.
- Customer Service Automation – Powering chatbots, virtual agents, and contact-center automation.
- Document Analysis and Summarization – Extracting, classifying, and summarizing information from contracts, forms, or reports.
- Data Analysis and Insights – Discovering patterns, forecasting outcomes, and detecting anomalies in structured and unstructured data.
- Personalization and Recommendations – Tailoring content, offers, or experiences for consumers in real time.
- Process Automation – Streamlining workflows across finance, HR, and supply-chain operations.
- Research and Knowledge Management – Enabling semantic search, discovery, and knowledge graph construction within organizations.

- Education and Training Content – Generating or adapting instructional materials and personalized learning paths.
- Cybersecurity and IT Operations – Detecting threats, responding automatically to incidents, and optimizing infrastructure performance.

Regions

AI platform adoption varies by geography:

- North America (NA)
- Latin America (LATAM)
- Europe, Middle East, and Africa (EMEA)
- Asia-Pacific (APAC)

Five-Year Forecast for AI Platforms Market (2025–2030)

Top-line Outlook: The Base forecast grows from \$24.9B in 2024 to \$292.0B in 2030 ($\approx 50.8\%$ CAGR). Alternative paths bound the range by 2030: Bear at \$155.1B and Bull at \$819.4B. Scenarios are independent trajectories (not additive), and Base totals anchor all normalizations across views.

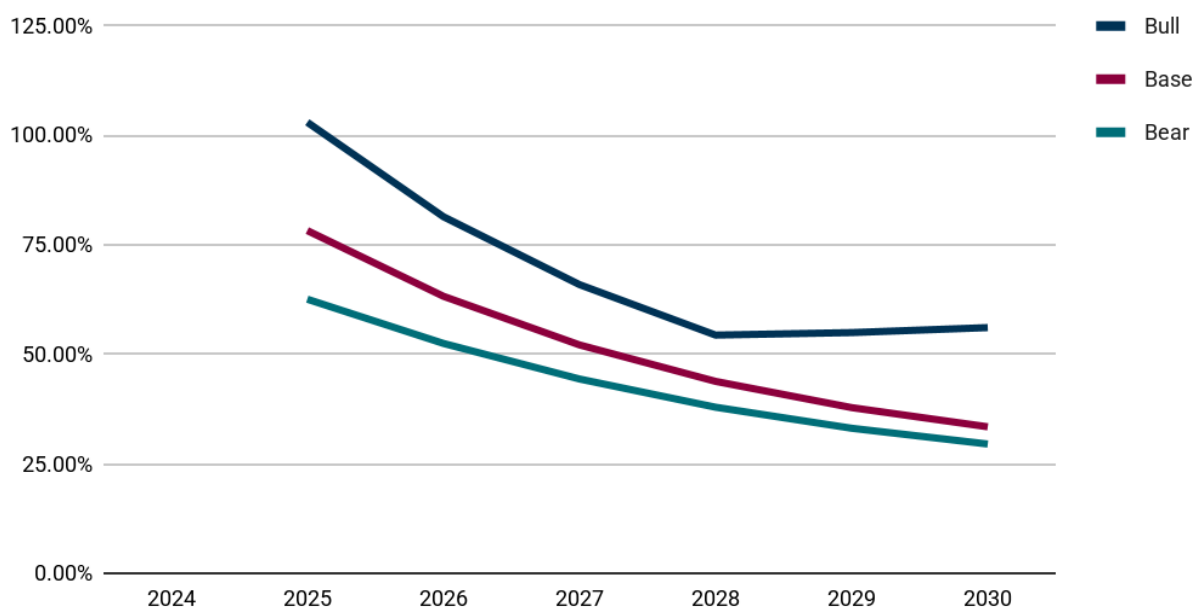
What moves the market between scenarios (key assumptions)

- Training & Inference Cost Curves: Faster unit-cost declines (hardware + software efficiency) accelerate adoption; slower declines compress ROI and delay deployments.
- Enterprise Adoption & Talent: Availability of skilled teams and mature partner ecosystems raises the number of production apps per enterprise.
- Regulation & Governance: Predictable, harmonized rules and built-in compliance tooling unlock regulated sectors; fragmented regimes slow rollouts.
- Data Center & Power Capacity: Adequate GPU supply, grid capacity, and regional cloud coverage enable scale; persistent shortages constrain growth.

Scenario narratives

- Bull: Hyper-adoption across industries, rapid cost/performance gains, strong API/platform monetization, limited regulatory friction ($\sim 2.8\times$ Base by 2030; implied CAGR $\approx 55\text{--}56\%$).
- Base: Balanced adoption and pricing normalization; steady enterprise scaling anchored to locked totals (CAGR $\approx 50.8\%$).
- Bear: Slower enterprise uptake; stronger commoditization and margin pressure; spending concentrated in a few vendors (implied CAGR $\approx 32\text{--}33\%$).

Figure 1: H2 2025 AI Platforms Market Scenarios: 2024–2030 (Annual Growth Rate in %)



Source: Futurum Research, October 2025

Table 1: Five-Year Market Forecast, by Scenario (in US\$ Million)

Year	2024	2025	2026	2027	2028	2029	2030
Bull	35,843.3	72,768.9	132,051.0	219,170.6	338,499.8	524,733.5	819,403.6
Base	24,872.6	44,347.2	72,423.1	110,231.8	158,608.6	218,719.4	292,027.6
Bear	18,167.2	29,546.4	45,073.9	65,102.2	89,831.6	119,646.7	155,051.7

Source: Futurum Research, October 2025

Forecast Details, by Market Segment

The following sections provide forecasts and commentary on the individual market segments – Vendor, Segment, Industry, and Deployment

Forecast Details, by Vendor

Vendor trajectories reflect differing starting positions, product breadth, data gravity, go-to-market reach, and pricing dynamics. Early growth is led by hyperscalers on the strength of integrated clouds and model access; independent platforms accelerate as enterprise architectures diversify and governance needs mature. Late-decade outcomes hinge on cost curves, compliance clarity, and partner ecosystems.

- Microsoft (Azure AI Platform): Growth is anchored in deep enterprise penetration and tight integration across Microsoft 365, Dynamics, and Azure. Early outperformance

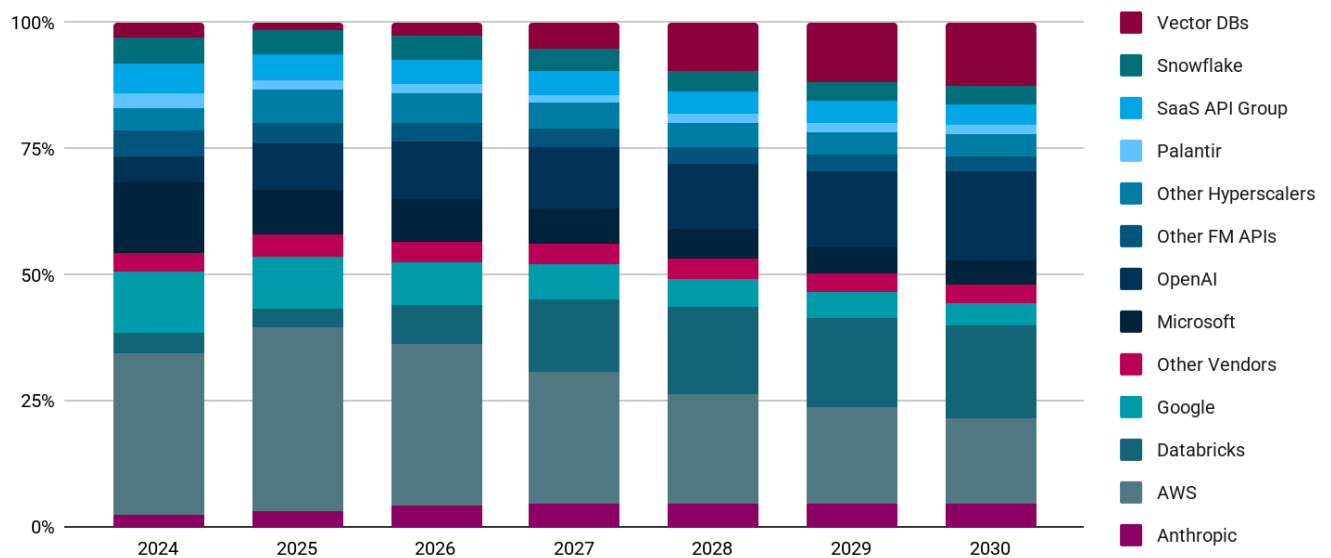
stems from rapid Copilot adoption and the Azure OpenAI stack; normalization appears later in the period as saturation rises and price transparency increases. A developing third-party Copilot/extension ecosystem supports durable workload expansion.

- Amazon Web Services (Bedrock, SageMaker): Strong infrastructure-led monetization through broad model catalogs and managed training/inference. Share stabilizes mid-period as buyers diversify across clouds and consider hybrid/edge deployment for latency and cost control. Consumption-based pricing and SI/channel depth sustain absolute growth.
- Google Cloud (Vertex AI): Early usage converts to revenue more sharply mid-period as multimodal capabilities mature and data analytics tie-ins (e.g., BigQuery) streamline productionization. Strength in developer tooling and retrieval workflows supports expanding seats beyond initial pilots.
- OpenAI: Rapid growth from a smaller base driven by model leadership, API adoption, and enterprise licensing. As larger accounts scale, emphasis shifts to customization, reliability, and governance features. Execution in enterprise sales and support is the key swing factor.
- Anthropic: Safety and reliability positioning resonates with regulated industries. Strategic cloud partnerships improve distribution, while iterative model releases attract developer loyalty. Growth moderates late period as procurement standardizes and price competition intensifies.
- Databricks (AI Slice): Expansion follows the convergence of analytics and AI: feature stores, retrieval-augmented generation, and governed data pipelines. Competitive pressure from hyperscaler-native services tempers share, but multi-cloud posture and open tooling preserve momentum.
- Snowflake (AI Slice): The AI Data Cloud and model access within familiar data workflows drive steady adoption, particularly where data governance is paramount. Competition from embedded cloud-native features compresses margins over time, but data gravity maintains stickiness.
- SaaS Platform APIs (Salesforce, ServiceNow, SAP, Adobe): Growth is paced by large installed bases and strong data moats. As customers seek low-code agent builders and domain-specific models, these platforms monetize via extensibility and embedded governance, though they generally trail hyperscalers in raw infrastructure velocity.
- Vector Database Specialists (e.g., Pinecone, Weaviate, Zilliz, Elastic, MongoDB, Neo4j, Redis, Qdrant): High growth as retrieval becomes standard in enterprise AI. Over the horizon, native vector capabilities in major data platforms and clouds increase

commoditization risk, shifting competition to performance, total cost of ownership, and ecosystem fit.

- Palantir, C3 AI, SAS, and Similar Mid-caps: Durable demand in government and regulated sectors where program delivery, security posture, and domain templates matter as much as raw model access. Absolute revenue grows while share drifts modestly as broader platforms expand.
- Other Foundation Model/API Vendors (e.g., AI21 Labs, Cohere, Mistral, xAI): Rapid early ramp from differentiated models and licensing approaches (including open-weight options). Distribution via cloud marketplaces and strong SI alliances is decisive for sustained scale.
- Other Hyperscalers (Oracle, Alibaba, Baidu, Tencent, Huawei): Solid regional momentum supported by domestic ecosystems and regulatory alignment. Global share is bounded by cross-border compliance and ecosystem fragmentation.
- Long Tail/Net Remainder: A wide field of orchestration, evaluation, guardrail, and vertical-specific vendors benefits from early experimentation. Consolidation and platform absorption increase over time; absolute dollars rise, but share declines as standards converge.

Figure 2: H2 2025 AI Platforms Market, by Vendor, 2024–2030 (Revenue Share, %)



Source: Futurum Research, October 2025

Table 2: Forecasted 5-Year CAGR, by Vendor (in %)

Vendor	CAGR in %(2025–2030)
Vector DBs	129.2
Databricks	103.0
OpenAI	70.9
Anthropic	59.7
Palantir	48.2
Other Vendors	44.2
SaaS API Group	43.1
Snowflake	41.1
Other Foundation Model APIs	39.3
Other Hyperscalers	36.4
Microsoft	31.5
AWS	28.4
Google	25.9

Source: Futurum Research, October 2025

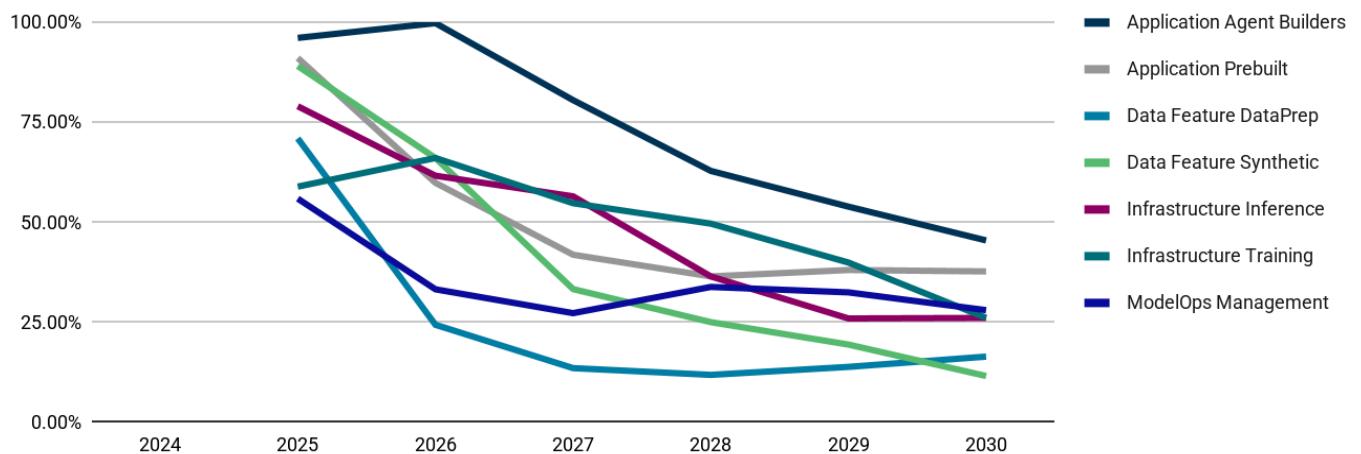
Forecast Details, by Segment

Spending patterns reflect where enterprises are in the AI build-out: training and core infrastructure lead early, inference and application enablement dominate mid-period, and governance becomes mandatory as deployments scale.

- Infrastructure – Training Platforms: Rapid near-term growth as organizations train and fine-tune models; pace moderates later as model reuse, transfer learning, and efficiency gains reduce retraining intensity.
- Infrastructure – Inference Services: Becomes the primary engine of revenue by mid-decade as copilots, agents, and real-time assistants move into production; second lift comes from latency-sensitive use at edge locations.
- Data & Feature – Data Preparation & Feature Stores: Strong adoption with the rise of retrieval-augmented generation and governed feature pipelines; growth tapers as these capabilities become standard parts of the stack.
- Data & Feature – Synthetic Data & Labeling Operations: Starts smaller, then accelerates when safety testing, agent behaviors, and long-context applications demand larger, more diverse training corpora; automation tempers intensity later.
- ModelOps & Governance – Model Management & Versioning: Steady adoption as teams formalize experimentation, registries, and lineage; procurement and audit requirements make these table-stakes in the enterprise.
- ModelOps & Governance – Monitoring & Compliance: Faster late-period growth as regulations and internal risk policies require drift, bias, provenance, and auditability controls across industries.

- Application Enablement – Prebuilt Models & APIs: Hypergrowth in the early years as enterprises integrate general-purpose model APIs; growth normalizes later as open-weight and fine-tuned, self-hosted options expand buyer choice.
- Application Enablement – Workflow & Agent Builders: Gradual early ramp during pilots; sharp acceleration once businesses standardize multi-agent patterns for customer service, operations, and software delivery, and then a more measured pace as frameworks mature.

Figure 3: H2 2025 AI Platforms Market, by Segment, 2024–2030 ((Annual Growth Rate in %)



Source: Futurum Research, October 2025

Table 3: Five-Year Market Forecast, by Segment (in US\$ Million)

Year	2024	2025	2026	2027	2028	2029	2030
Application Agent Builders	3,233.4	6,342.8	12,676.9	22,892.8	37,288.7	57,399.2	83,505.5
Infrastructure Inference	4,974.5	8,907.5	14,400.4	22,537.8	30,757.5	38,737.2	48,841.6
Infrastructure Training	4,477.1	7,116.0	11,820.2	18,296.3	27,385.4	38,322.5	48,318.7
Application Prebuilt	4,228.3	8,081.8	12,921.9	18,334.9	25,021.7	34,548.0	47,580.5
ModelOps Monitoring	1,989.8	3,803.2	6,756.6	11,039.4	17,009.5	23,802.7	32,781.8
ModelOps Management	2,238.5	3,490.4	4,650.7	5,919.0	7,921.3	10,493.7	13,434.5
Data Feature DataPrep	2,487.3	4,253.6	5,289.7	6,004.4	6,714.0	7,641.6	8,893.7
Data Feature Synthetic	1,243.6	2,352.0	3,906.8	5,207.4	6,510.5	7,774.5	8,671.3

Source: Futurum Research, October 2025

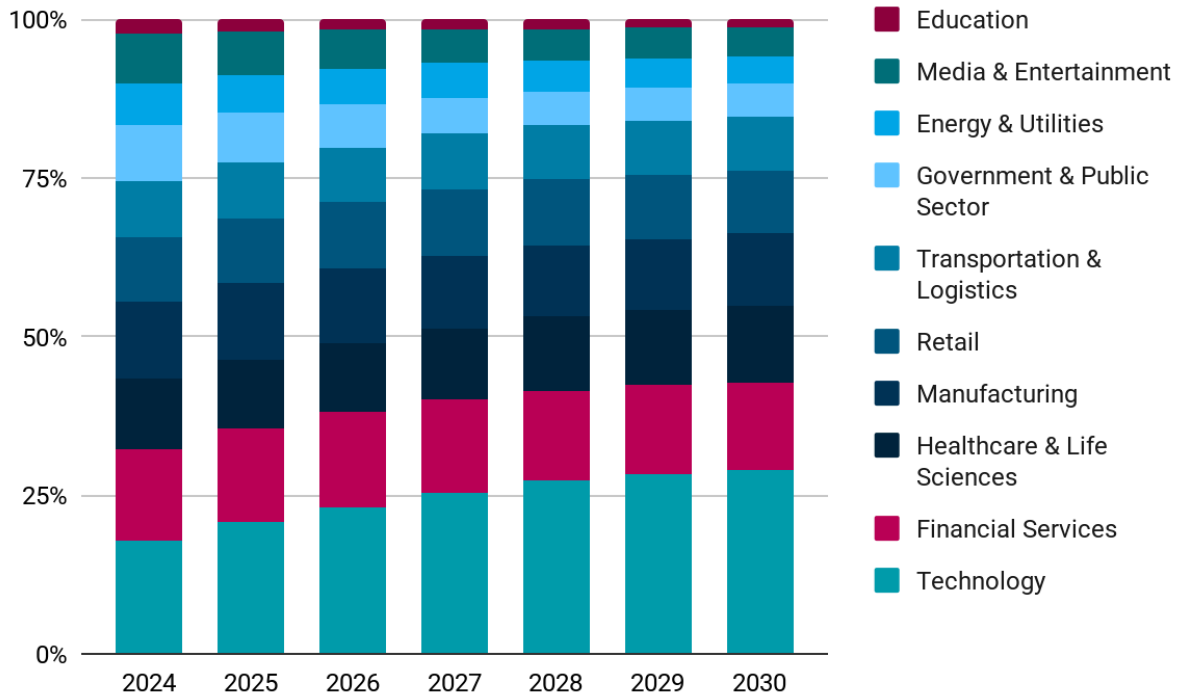
Forecast Details, by Industry

Industry trajectories reflect the mix of data sensitivity, regulatory burden, ROI clarity, and digital maturity. Early leaders pair clear business cases with abundant, well-governed data; later lift comes from scaled operations, workflow redesign, and standardized controls.

- Financial Services
 - Fast adoption in fraud, risk, personalization, and compliance operations; strong, early ROI sustains above-market growth.
 - Preference for governed platforms with auditability; private/hybrid deployments common for sensitive workloads.
 - Late-period focus on model risk management and explainability as model counts scale.
- Healthcare & Life Sciences
 - Demand split between clinical/diagnostic support and administrative automation (coding, documentation, revenue cycle).
 - Early progress tempered by validation and privacy requirements; growth accelerates as approved workflows and vendor attestations mature.
 - Partnerships with EHR/clinical systems and secure data-sharing frameworks drive adoption.
- Manufacturing
 - Durable growth from quality inspection, predictive maintenance, and supply-chain optimization; many deployments at the edge.
 - Longer rollout cycles tied to OT/IT integration and plant modernization.
 - Late-period uplift from multi-agent workflows coordinating procurement, maintenance, and scheduling.
- Retail & E-commerce
 - Early surge from personalization, dynamic pricing, and merchandising; strong data exhaust supports rapid iteration.
 - Growth normalizes as core use cases saturate and privacy rules tighten; winners integrate store, web, and ad channels.
 - Edge inference for in-store experiences and loss prevention adds incremental lift.
- Technology
 - Largest absolute growth as ISVs, device makers, and consumer platforms embed AI across products and developer tools.
 - Continuous reinvestment in model families and agent frameworks; rapid cadence drives sustained platform consumption.
 - Vendor differentiation shifts from raw model access to ecosystems, telemetry, and governance-by-default.

- Government & Public Sector
 - Measured early adoption; procurement, security accreditation, and explainability set the pace.
 - Defense and national security programs drive step-changes; civilian services focus on citizen support and document workflows.
 - Preference for private/hybrid deployments and vendors with attestable compliance.
- Energy & Utilities
 - Grid optimization, demand forecasting, and asset monitoring provide steady, regulated growth.
 - Site and safety constraints slow net-new deployment; operational reliability and audit trails are mandatory.
 - Late-period gains from integrating AI with distributed energy resources and predictive maintenance at scale.
- Transportation & Logistics
 - Route optimization, network planning, and terminal operations mature first; autonomous/assisted operations progress under stringent safety regimes.
 - Capital intensity and certification requirements lengthen timelines; multimodal (text/vision) AI expands impact.
 - Partnerships with OEMs, carriers, and platform providers are decisive.
- Media & Entertainment
 - Early spike in generative content and localization; pricing compression and commoditization moderate growth later.
 - Rights management, provenance, and watermarking become table-stakes for scaled production.
 - Differentiation shifts to workflow integration and audience analytics.
- Education
 - Gradual adoption given funding cycles and procurement complexity; strongest traction in tutoring, content generation, and grading support.
 - Institutional policies on privacy and academic integrity shape vendor selection.
 - Late-period improvement as validated outcomes and accessible pricing expand deployment.

Figure 4: H2 2025 AI Platforms Market, by Industry, 2024–2030 (Revenue Share, %)



Source: Futurum Research, October 2025

Table 4: Five-Year Market Forecast, by Industry (in US\$ Million)

Year	2024	2025	2026	2027	2028	2029	2030
Technology (TEC)	4,421.8	9,215.7	16,822.2	27,926.0	43,080.9	61,685.7	84,471.6
Financial Services (FIN)	3,592.7	6,571.7	10,747.7	16,300.1	22,829.8	30,794.0	40,571.6
Healthcare & Life Sciences (HLS)	2,763.7	4,779.4	7,816.5	12,175.0	18,287.9	25,806.2	35,338.6
Manufacturing (MFG)	3,040.0	5,392.1	8,648.0	12,849.7	17,997.2	24,649.0	32,731.2
Retail (RET)	2,487.2	4,466.9	7,446.8	11,446.6	16,264.3	22,275.7	29,579.6
Transportation & Logistics (TLG)	2,210.9	3,970.6	6,368.1	9,527.3	13,537.3	18,540.7	24,620.0
Government & Public Sector (GOV)	2,210.9	3,357.8	4,783.0	6,469.8	8,536.2	11,336.9	14,701.3
Energy & Utilities (ENU)	1,658.2	2,720.6	4,133.6	5,845.5	7,890.6	10,397.6	13,159.7
Media & Entertainment (MED)	1,934.6	3,002.4	4,371.8	5,913.6	7,802.3	10,119.4	12,912.6
Education (EDU)	552.8	870.1	1,285.3	1,778.1	2,382.1	3,114.2	3,941.5

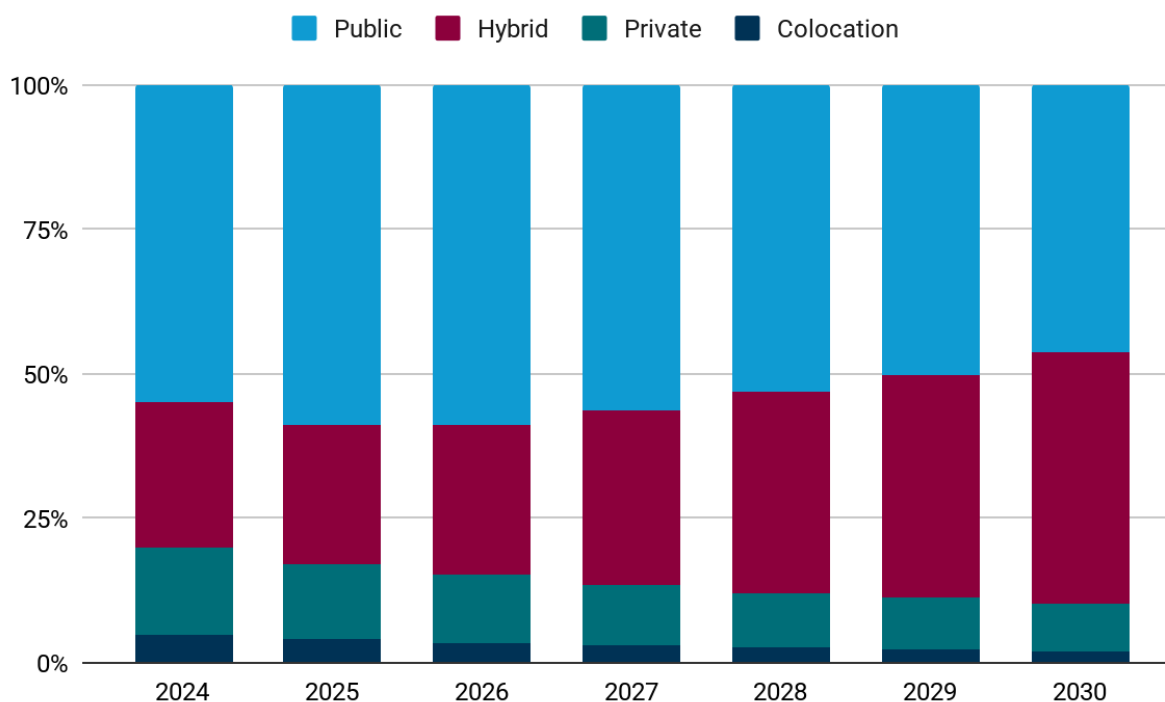
Source: Futurum Research, October 2025

Forecast Details, by Deployment

Deployment choices follow data sensitivity, latency needs, cost transparency, and regulatory posture. Most programs start in the public cloud for speed, and then rebalance toward hybrid and private patterns as workloads scale and compliance tightens.

- **Public Cloud/SaaS (PUB)**
 - Role: Primary on-ramp for pilots and early scale due to managed services, elastic capacity, and broad model catalogs.
 - Growth Drivers: Time-to-value, integrated security/compliance features, marketplace distribution, and close adjacency to existing data/analytics stacks.
 - Watchouts: Unit-economics scrutiny at higher utilization and data-egress costs; some enterprises shift steady-state inference to hybrid/edge for latency and cost control late-period.
- **Hybrid/Edge (HYB)**
 - Role: Fastest structural share gainer mid-decade as enterprises place inference closer to data and users.
 - Growth Drivers: Privacy/sovereignty requirements, latency-sensitive apps (operations, retail, manufacturing), and the rise of smaller, efficient models suitable for on-device/edge.
 - Watchouts: Operational complexity (lifecycle management across sites), skills for fleet-scale model updates, and fragmented hardware footprints.
- **Private/On-premises (PRV)**
 - Role: Steady option where data residency, control, or cost predictability outweigh public-cloud convenience—especially in finance, government, and healthcare.
 - Growth Drivers: Compliance mandates, dedicated security enclaves, integration with existing data estates, and predictable budgeting for sustained, high-utilization workloads.
 - Watchouts: Longer implementation cycles; requires stronger ModelOps discipline to match cloud agility.
- **Colocation/Managed Hosting (COL)**
 - Role: Niche but durable path for dedicated clusters when organizations want third-party facilities without public-cloud tenancy.
 - Growth Drivers: Hardware supply availability, power density, and cost arbitrage versus self-built data centers; useful as a bridge during capacity constraints.
 - Watchouts: Smaller ecosystem breadth than hyperscalers; requires careful TCO analysis once workloads mature.

Figure 5: H2 2025 AI Platforms Market, by Deployment, 2024–2030 (Revenue Share, %)



Source: Futurum Research, October 2025

Table 5: Five-Year Market Forecast, by Deployment (in US\$ Million)

Year	2024	2025	2026	2027	2028	2029	2030
Public	13,679.9	26,024.8	42,616.9	62,207.4	84,223.2	109,504.4	135,301.1
Hybrid	6,218.2	10,712.2	18,857.4	33,105.6	55,067.0	84,855.0	126,960.6
Private	3,730.9	5,796.4	8,424.1	11,631.9	15,186.1	19,305.7	23,667.3
Colocation	1,243.6	1,813.8	2,524.7	3,286.9	4,132.3	5,054.3	6,098.6

Source: Futurum Research, October 2025

Forecast Details, by Region

Regional trajectories are shaped by enterprise digital maturity, government investment, regulatory posture, and macroeconomic stability. While North America establishes an early lead and APAC drives the fastest initial growth, the market will normalize toward a more globally balanced landscape as regulatory frameworks in EMEA mature and other economies scale.

North America

- Maintains early lead via high digital maturity; regulatory clarity (NIST RMF) boosts 2024–26 growth.
- Growth slows post-2027 due to market saturation, shifting enterprise focus to cost optimization and ROI.
- Late-decade lift expected from the adoption of edge and on-device AI for real-time applications.

EMEA

- Initial deployment pace is restrained by the regulatory complexities of GDPR and the EU AI Act.
- Post-2026, mandated compliance for "Trustworthy AI" is forecast to flip from a restraint to become a key growth driver.
- Steady, durable late-decade growth is sustained by public-sector modernization and strong adoption in the industrial automation sector.

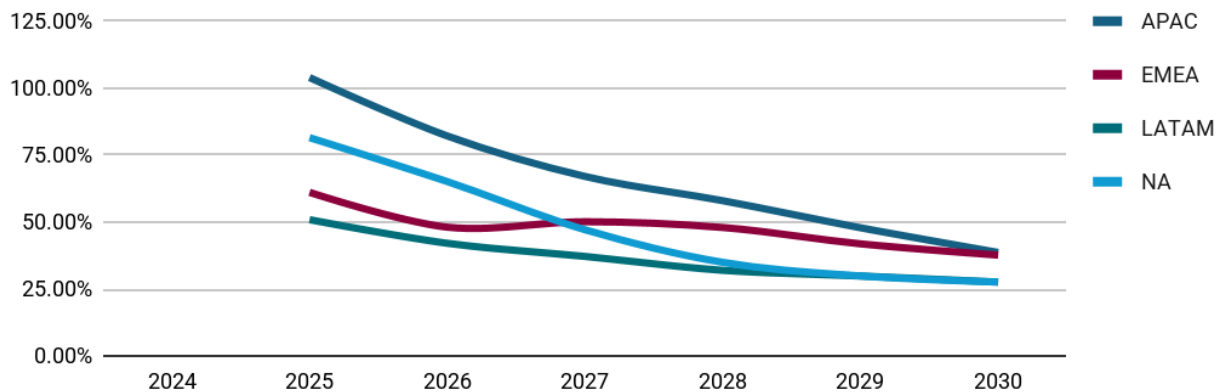
APAC

- Experiences the fastest early growth, overwhelmingly driven by large-scale, government-led investment programs.
- 2025 is marked by major national initiatives in China (AI+ drive), India ("IndiaAI Mission"), and Japan (AI Promotion Act) to build sovereign compute and models.
- Growth normalizes toward 2030 as key economies mature and new regulations tighten data sovereignty and export controls.

LATAM

- Market growth is fundamentally constrained by macroeconomic volatility, currency risk, and regional talent shortages.
- Adoption is highly clustered, with steady demand driven by the modernization of the financial services and digital government sectors.
- National-level scaling is limited by patchy digital connectivity outside of major urban hubs.

Figure 6: H2 2025 AI Platforms Market, by Region, 2024–2030 (Annual Growth Rate in % – 2025–2030)



Source: Futurum Research, October 2025

Table 6: Five-Year Market Forecast, by Region (in US\$ Million)

Year	2024	2025	2026	2027	2028	2029	2030
North America	11,989	21,767	35,952	52,920	71,458	92,852	118,539
Asia Pacific	4,477	9,133	16,640	27,803	43,938	64,998	90,167
EMEA	7,412	11,945	17,697	26,578	39,344	55,842	76,904
Latin America	995	1,502	2,135	2,930	3,869	5,027	6,418

Source: Futurum Research, October 2025

Forecast Details, by Use Case

Use case adoption is determined by ROI clarity, potential for productivity gains, data accessibility, and the complexity of safety or compliance requirements. Initial growth is dominated by high-impact generative tasks like content and code generation, but the market's center of gravity will shift toward more deeply embedded, complex workflows like multi-agent process automation and enterprise knowledge management.

Content Generation

- Experiences an early spike in adoption across 2024–25, driven by marketing and media use cases.
- Growth moderates significantly post-2026 due to rapid price compression and model commoditization.

Code Generation/Dev Assist

- Drives the strongest sustained CAGR of all use cases due to clear, measurable productivity gains for enterprise developers.
- Growth plateaus after 2028 as the primary market of active developers becomes saturated.

- Significant bull-case upside from the largely untapped market of legacy system and code modernization.

Customer Service Automation

- Sees fast early uptake for text-based chatbots; clear ROI from reduced operational costs.
- A second pulse of growth is expected around 2026, driven by more capable multimodal (voice, text, vision) agents.
- Growth flattens late-decade as the primary use cases for Level 1 support become fully saturated.

Document Analysis/Summarization

- Mid-decade growth accelerates, driven by the cross-industry adoption of Retrieval-Augmented Generation (RAG).
- Becomes the standard for building internal knowledge bases and enterprise-specific AI assistants.

Data Analysis/Insights

- Sees strong early growth from natural language query tools; adoption dips as data quality issues create complexity.
- A late-decade rebound is expected as more advanced financial modeling and predictive forecasting use cases mature.

Personalization/Recommendations

- Early growth is slowed by new privacy regulations and the deprecation of third-party data sources.
- Growth revives and re-accelerates post-2026, driven by the adoption of privacy-preserving techniques like federated learning.

Process Automation

- Drives the second-fastest CAGR as generative AI is integrated with existing RPA platforms.
- A significant growth spike is anticipated in 2026–27, driven by the emergence of multi-agent orchestration frameworks.

Research/Knowledge Management

- Shows steady, durable growth as RAG becomes the core of the enterprise knowledge stack after 2026.
- Adoption is driven by the critical need to unlock and query unstructured internal data.

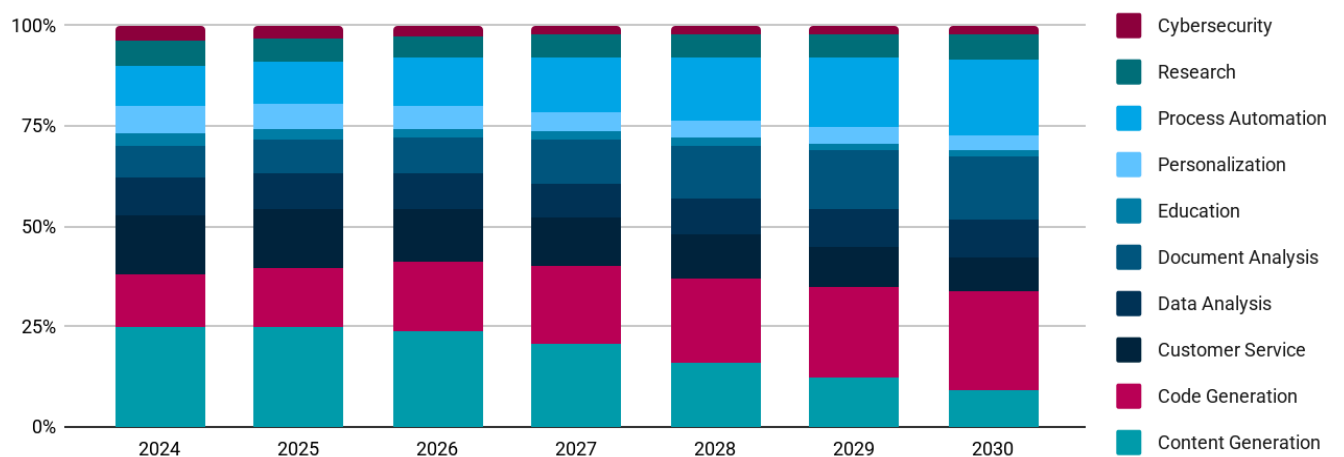
Education/Training Content

- Adoption is gradual, paced by public funding cycles and long institutional procurement processes.
- Growth accelerates late-decade as validated solutions with proven outcomes become available at more accessible price points.

Cybersecurity/IT Ops

- Shows strong, non-discretionary growth due to the rising complexity of the threat landscape.
- The market is characterized by breach-driven spikes in spending, with major events in 2025 and 2028 expected to accelerate adoption.

Figure 7: H2 2025 AI Platforms Market, by Use Case, 2024–2030 (Annual Growth Rate in % – 2025–2030)



Source: Futurum Research, October 2025

Table 7: Five-Year Market Forecast, by Use Case (in USD Million)

Year	2024	2025	2026	2027	2028	2029	2030
Code Generation	3,233.0	6,563.0	12,440.0	21,080.0	32,914.0	49,515.0	71,794.0
Process Automation	2,487.0	4,824.0	8,608.0	14,588.0	24,404.0	37,923.0	55,622.0
Document Analysis	1,990.0	3,680.0	6,600.0	12,494.0	21,223.0	32,103.0	45,740.0
Content Generation	6,218.0	11,106.0	17,252.0	23,015.0	25,668.0	26,945.0	27,551.0
Data Analysis	2,239.0	3,897.0	6,270.0	9,043.0	13,576.0	19,975.0	26,451.0
Customer Service	3,731.0	6,496.0	9,790.0	13,413.0	17,835.0	21,819.0	24,687.0
Research	1,492.0	2,531.0	3,978.0	6,311.0	9,367.0	13,007.0	17,660.0
Personalization	1,741.0	2,701.0	3,996.0	5,295.0	6,814.0	8,562.0	10,620.0
Cybersecurity	995.0	1,391.0	1,904.0	2,608.0	3,535.0	4,676.0	6,348.0
Education	746.0	1,158.0	1,584.0	2,385.0	3,274.0	4,194.0	5,554.0

Source: Futurum Research, October 2025

Research Background and Methodology

Purpose and Scope: This study measures software and managed-services revenue that enables enterprises to build, train, deploy, and operate AI/ML systems. Hardware (e.g., GPUs/servers) and finished business applications that merely embed AI are excluded. Figures are reported in USD, nominal, on a calendar-year basis.

1. Data Assembly and Vendor Mapping

- **Primary Sources:** SEC and equivalent filings, earnings transcripts, investor presentations, product/pricing pages, and published customer counts where available.
- **Secondary Inputs:** Partner/channel checks, ecosystem benchmarks (e.g., marketplace listings, model catalogs), and public product telemetry where disclosed.
- **Vendor Taxonomy Alignment:** Each product family is mapped to a single platform segment (e.g., training, inference, data/feature, ModelOps, application enablement). Revenue that spans multiple segments is allocated using documented keys (e.g., SKU mix, usage ratios, disclosed attach rates).

2. Baseline Construction and Locked Totals

- The model reconciles annually to a locked Base forecast for total AI Platforms spend (2024–2030).
- Historical context years 2022–2023 are maintained as zero baselines for consistency with prior versions.
- All views—Vendor, Segment, Industry, Deployment—are normalized each year so their totals match the locked Base figure.

3. Estimation Techniques

- **Top-down Triangulation:** When vendors disclose AI/platform revenue (or adjacent metrics such as “GenAI contribution”), we use those disclosures and reconcile them to product scope.
- **Bottom-up Modeling:** Where disclosure is limited, we estimate with unit-economics and adoption curves (e.g., API consumption × price, seat counts × ARPU, workload volumes × \$/1M tokens), adjusted for discounting and utilization.
- **Allocation Rules:** Mixed SKUs are split by observable usage, product packaging, or management commentary; services revenue is counted only when bundled as recurring platform capability (not standalone consulting).

4. Scenario Framework (Bull/Base/Bear)

- **Base** is the locked outlook used for annual normalization.

- Bull/Bear totals are derived by applying year-specific multipliers to the Base path, reflecting upside/downside in: cost curves (training/inference), enterprise adoption velocity and talent, regulatory clarity, and data-center/power capacity.
- Scenarios are independent trajectories and are not additive.

5. Normalization and Calculated Fields

- For each group, totals equal the locked Base figure; shares sum to 100.0%.
- Year-over-year growth is computed from the prior year within each view.
- Values are rounded to one decimal place; reconciliation tolerance is USD ± 0.1 million at the view-year level.

6. Quality Assurance

- Cross-view Checks: Vendor, Segment, Industry, and Deployment totals match the Base path annually.
- Outlier Reviews: Abrupt share or growth jumps are flagged and re-evaluated against disclosures and product milestones.
- Reproducibility: All transformations (allocation keys, multipliers, rounding) are deterministic and documented to enable repeatable rebuilds.

7. Assumptions and Boundaries

- Only platform/API revenue is counted; embedded AI inside end-user applications is excluded unless sold as a platform capability.
- Regional reporting uses Tier-1 regions (North America; Europe, Middle East & Africa; Asia-Pacific; Latin America).
- The model emphasizes enterprise consumption; consumer-only use is included only where monetized as a platform/API.

8. Revision Policy

- The dataset is refreshed when vendors materially update disclosures, pricing, or packaging, or when market conditions shift the scenario drivers. Revisions preserve the documented scope and normalization rules to maintain comparability over time.

About Us

About the Authors

Nick Patience, Vice President & AI Practice Lead is a thought leader on the development, deployment, and adoption of AI, an area he has been researching for 25 years. Prior to Futurum, Nick was a Managing Analyst with S&P Global Market Intelligence, with responsibility for 451 Research's coverage of Data, AI, Analytics, Information Security, and Risk. Nick became part of S&P Global through its 2019 acquisition of 451 Research, a pioneering analyst firm Nick co-founded in 1999. He is a sought-after speaker and advisor, known for his expertise in the drivers of AI adoption, industry use cases, and the infrastructure behind its development and deployment. Nick also spent three years as a product marketing lead at Recomind (now part of OpenText), a machine learning-driven eDiscovery software company. Nick is based in London and can be reached at npatience@futurumgroup.com.

Varun Shitole is a seasoned research analyst who has been a research professional for over 7 years – having spent a major tenure in market research for the technology industry, specializing in market sizing and forecast, go-to-market strategy, competitive benchmarking, and opportunity assessment.

About The Futurum Group

The Futurum Group's analysts, researchers, and advisors help business leaders anticipate tectonic shifts in their industries and leverage disruptive innovation. Unlike traditional analysts, The Futurum Group works in analysis and research and takes that insight and knowledge even further, engaging through the go-to-market process.

Futurum Research provides in-depth research and insights on global technology markets using advisory services, custom research reports, strategic consulting engagements, digital events, go-to-market planning, and message testing. It also creates, distributes, and amplifies rich media content that all stakeholders read, watch, and listen to.

See more details on The Futurum Group at futurumgroup.com.

Copyright & Use License

Copyright Notice

Copyright ©2025 by The Futurum Group, LLC. All rights, including translation into other languages, are specifically reserved. No part of this publication may be reproduced in any form, stored in a retrieval system, or transmitted by any method or means, electrical, mechanical, photographic, or otherwise, without the express written permission of The Futurum Group futurumgroup.com. United States copyright laws and international treaties protect this publication. Unauthorized distribution or reproduction of this publication, or any portion of it, may result in severe civil and criminal penalties and will be prosecuted to the maximum extent necessary to protect the publisher's rights.

License Notice

This document may be distributed only to the licensed organization.

The following acts are prohibited:

Transmittal to others outside your immediate organization, including partners, resellers, external consultants, etc., in any media format

Posting on a website that is accessible to others outside your immediate organization

The possession or use within an unlicensed organization

Limitation of Liability Notice

The information contained herein has been obtained from sources believed to be reliable. The Futurum Group shall have no liability for errors, omissions, or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for selecting these materials to achieve their intended results. The opinions expressed herein are subject to change without notice.