NUCLEUS
RESEARCH

Nucleus Top Ten Predictions for 2026

ANALYST

Ian Campbell, Trevor White, Evelyn McMullen, Charles Brennan, Alexander Wurm, Cameron Marsh, Duncan Van Kouteren, Rebecca Kennedy, Charlotte Belke, Sofia Halbisen

The Bottom Line

Nucleus's predictions for 2026 are, unsurprisingly, centered around evolving AI technology and its anticipated impacts and consequences. For example, as the emergence of Agents, and "Agents", continues, organizations are grappling with how the technology can drive actual value in practice while vendors remain unclear on how to monetize it. As AI becomes a commodity, traditional per-seat software models stand to face significant shifts. Along with this, 2026 is expected to spark a reckoning when it comes to the transparency and explainability of models and algorithms. Under the hood, the financial burden of cloud inference will drive cloud repatriation to more cost-effective, on-premises hardware. The Supply Chain Planning market, among others, is transitioning towards the adoption of open data ecosystems. HR and Finance are becoming closer companions, and HCM vendors are fighting to win over the frontline workforce.

The Emperor Has No Agents

It feels like just yesterday that every vendor was doubling down on their "novel" innovations in Generative AI. Organizations were promised massive benefits by being able to surface information more quickly through conversational assistants, summarize long blocks of text, and even generate their own blocks of text. Jumping forward two years, some Copilots and Assistants are being quietly rebranded as Agents. As if the advent of Generative AI wasn't enough to send the market into a frenzy, the market is now not only frenzied but confused. In 2026, Nucleus predicts that much of the dishonesty currently taking place for the benefit of earnings calls and ROI on investments in AI hype will come to a head. In short, vendors who are not being transparent about what is under the hood of their AI supercar will have some explaining to do, especially as AI literacy and the legal landscape catch up. On the other hand, the vendors who have made a clear distinction between Generative and Agentic AI and made a concerted effort in the explainability of what goes on behind the black box will be better positioned to continue riding the AI wave successfully.

2026 will be the year that separates serious players in the AI space from those just trying to keep up with the hype.

IaaS Price War: $0 Compute

In a dramatic escalation of cloud competition, major infrastructure providers will introduce free compute tiers for enterprise AI and machine learning workloads in 2026, fundamentally upending traditional IaaS economics. Similar to social networks that monetized user data rather than direct service access, hyperscalers will pivot compute models toward extracting and analyzing AI training metadata. The strategic dynamic parallels the late 1990s software commoditization period, when competitive differentiation shifted from price to platform control. Except, this time, the stakes involve trillion-dollar investments in AI infrastructure. Cloud providers will recoup costs through aggressive cross-selling of storage, networking, and managed services, while simultaneously mining computational workloads for insights that inform their own AI development. Enterprises will face a counterintuitive scenario. While compute pricing trends toward zero, aggregate cloud expenditure may rise as providers bundle complementary processing with premium dependencies in storage, networking, and data services. This race to the bottom will particularly benefit AI-native startups and research institutions, but enterprises must examine the fine print to understand precisely what metadata rights they are surrendering in exchange for processing subsidies that implicitly monetize organizational metadata.

Free compute tiers will shift cloud economics from direct pricing to metadata monetization, with providers extracting value from AI workload insights rather than processing fees

Algorithmic Regulation Backfires

Governments will deploy sophisticated regulatory AI systems designed to police data privacy, algorithmic bias, and compliance in real-time throughout 2026, only to discover that static oversight models cannot effectively govern adaptive AI architectures. These "watchdog" algorithms, envisioned as automated guardians that continuously monitor and flag violations, will be systematically outmaneuvered by next-generation adaptive models that learn to game regulatory frameworks faster than policymakers can update them. This dynamic resembles an asymmetrical feedback loop in which regulatory systems iterate linearly while generative AI models evolve exponentially. Advanced black-box models will exploit regulatory blind spots, generate compliant-appearing outputs that mask underlying biases, or adapt their architectures to remain opaque to algorithmic auditing. This dynamic will create an unsustainable policy arms race, forcing regulators to acknowledge that algorithmic oversight of increasingly sophisticated AI requires fundamentally different approaches than traditional compliance frameworks. Nucleus expects that this regulatory reckoning will prompt policymakers to adopt outcome-based monitoring and layered human-in-the-loop governance, moving beyond conventional rule-based enforcement, as governments recognize that algorithmic enforcers are perpetually one generation behind the systems they're meant to oversee.

Regulatory AI systems will prove ineffective against adaptive models that evolve exponentially faster than policy frameworks can iterate.

Where The Shift Happens

Nucleus expects 2026 to be the year of the frontline worker, with workforce management evolving from clipboard routines to a playbook that puts power in workers' hands. Investment is likely to shift toward a single worker app experience that combines scheduling flexibility, guided task execution, quick training, and simple communication so each shift starts ready and stays on track. Outcomes are expected to trend in the right direction as queues shorten, conversion rises, shrink declines, safety improves, and patient flow steadies. Frontline enablement remains practical rather than flashy. Associates are expected to use self-serve shift marketplaces, earn skill badges that open higher-value roles, acknowledge tasks with a tap, and take thirty-second refreshers in the flow of work. Swaps, open shifts, and PTO requests are likely to route under clear rules so coverage holds while workers retain control of their hours. Managers are expected to rely on exception views instead of blank screens, one-tap approvals that respect compliance, and daily action lists that tie staffing to store or unit goals. Forecasts explain the why behind each roster, outcome

Nucleus expects 2026 to put more control in workers' hands through a single app for scheduling, tasks, quick training, and communication, improving coverage, lowering costs, and lifting outcomes each shift.

dashboards track fill rates and on-time completion, and targeted incentives are expected to kick in automatically when coverage needs a lift. Taken together, these shifts point to a year defined by frontline enablement, with WFM investments that connect coverage, cost, and outcomes in a way finance and operations both trust, proving that strategy becomes execution where the shift happens.

Costs Fuel Cloud Repatriation

AI economics are forcing enterprises to rethink. Over the next year, we expect organizations to bring workloads in-house to control costs, improve latency, and better manage data privacy for high-volume inference.

Enterprises are reevaluating their reliance on public cloud infrastructure for AI workloads, particularly large language model (LLM) inference. As AI adoption scales, the financial burden of cloud-based inference has become a significant constraint. This pressure is accelerating a trend known as cloud repatriation, where organizations bring workloads back from the cloud to on-premises environments to improve cost control and limit variability in served models and inference.

The availability of powerful, specialized AI hardware from vendors like NVIDIA, Intel, and AMD is enabling this shift. These processors offer high-performance inference capabilities that, when deployed on-premises, can deliver lower total cost of ownership over time, especially for stable, high-volume inference workloads. This shifts customers from worrying about AI token economics and cloud margin to a financial model based on data center costs and hardware depreciation. While in the past, cloud justified its margin through simplicity, the additional cost on scaled inference will make this untenable for certain enterprise customers. Alternatively, on-premises deployment offers greater control over latency, data privacy, and compliance, as well as levers for cost control such as open-source model selection and quantization.

HCM + Finance is No Longer "Nice to Have"

Combining HCM and ERP is not new, but the prioritization of unified experiences is making it crucial.

Nucleus has spent years examining the benefits of consolidating finance and HR technology, which include increased productivity, standardization, and visibility across cross-departmental processes. As an added bonus, organizations with access to these tools through a single provider avoid the substantial headaches that come with building and managing integrations. However, the combination has traditionally been viewed as nice to have, not essential. As AI pulls the pendulum of "best-of-breed vs. full-suite" increasingly towards unified offerings, the pressure to consolidate as much as possible will quickly become too much to bear. This is evident through the strides taken by ERP giants that have had HCM as part of their offerings for years, as well as through both SMB and Enterprise HCM vendors acquiring finance and spend

management providers. HCM and ERP vendors across all market segments should take note, as the necessity of consolidating HR and finance will grow more critical over the next year. Vendors who do not take steps to fill their respective gaps will likely get left behind as competition increases.

Hungry, Hungry Tokens

Vendors across the enterprise software landscape are moving away from flat, add-on pricing toward metered AI consumption models that bill by tokens, queries, or processing volume. This shift follows the natural progression of SaaS economics, aligning cost with actual use and rewarding efficiency over excess. As AI becomes embedded in core applications across finance, operations, and development, metered models will help organizations scale intelligence responsibly rather than paying for idle potential. Still, cost unpredictability will remain a sticking point, especially for mid-market buyers without sophisticated monitoring tools. Vendors that provide transparent dashboards and consumption controls will have an early advantage, while others risk backlash from customers caught off guard by ballooning usage fees. Over time, AI will be treated less as a premium feature and more as a measurable operational utility, budgeted, optimized, and scrutinized like any other resource in the enterprise tech stack.

Vendors are shifting away from flat fees toward metered AI consumption models based on tokens, queries, or volume.

AI Shifts Software Economics

Nucleus expects a major disruption to the traditional per-seat software pricing model in 2026 as AI agents begin to perform tasks historically assigned to human users. As organizations deploy autonomous or semi-autonomous agents, the notion of paying for each "seat" will lose its relevance. Organizations won't justify human license counts when non-human agents can execute the same actions at scale, faster, and without coffee breaks.

Much like Spotify unbundled the album or Moneyball upending baseball by valuing data over tradition, AI will force software vendors to rethink how value is packaged and sold. The "seat" model, once the gold standard of SaaS predictability, will start to look like the Blockbuster membership of enterprise software: familiar, comfortable, and increasingly outdated. Vendors built on per-user renewals face a revenue identity crisis as customers shift toward models that charge for what's actually consumed or produced.

The value of technology is measured in output, not occupancy. AI doesn't need a desk, a login, or a lunch break, and that reality could make the seat license look like the Blockbuster card of SaaS: nostalgic, but obsolete.

Usage-based and outcome-based pricing already exist at the margins, but AI's rise will move them to center stage. Vendors that adapt early, experimenting with hybrid models based on transactions, workflow volume, or measurable value delivered, will control the narrative. Those that cling to static seat licenses risk being left behind, like legacy holdouts in an on-demand world, as buyers increasingly demand pricing that reflects productivity and ROI, not headcount.

The New Age of Supply Chain Planning

Supply chain planning is entering a new age shaped by stronger data foundations, AI, and the continued evolution of composable architectures. While machine learning has long supported forecasting, the shift toward agentic AI, enterprise knowledge graphs, and modern data partnerships is transforming how organizations plan, simulate, and respond in real time.

The next age of supply chain planning will be driven by organizations that can turn enterprise data into agile, context-aware decisions.

Nucleus expects to see continued adoption of open data ecosystems, exemplified by partnerships with Snowflake and Databricks, as organizations seek to unify, clean, and contextualize supply chain data across enterprise systems. In parallel, increased investment in enterprise knowledge graphs from vendors is allowing organizations to connect structured and unstructured data from supplier performance and logistics networks to external signals like weather and trade restrictions into a single contextual framework. These data-centric advances are creating the foundation for AI-driven planning that more adaptive and increasingly specific to industry contexts. Manufacturing organizations, for example, are embedding shop floor and quality data directly into planning cycles, while retailers and distributors are layering fulfillment, POS, and customer data for more precise, demand-driven planning.

Composable architectures will complement this shift by enabling flexibility, allowing organizations to deploy or extend planning capabilities without rearchitecting entire systems. Together, these trends mark the beginning of a new era in supply chain planning, one defined not just by predictive accuracy but by the ability to use enterprise data intelligently to make agile, context-aware decisions.

A Copilot for Every Manager

The era of manual juggling across spreadsheets, messages, and policies is losing steam. Across frontline industries, leaders are discovering that GenAI embedded in workforce tools is easier and cheaper than stitching together ad hoc workflows that miss context. Schedules

generate from demand signals and past patterns, trends read in plain language, and compliance risks surface before they reach payroll. Managers get suggested actions and ready-to-send shift communications, turning approvals into minutes instead of hours. On the floor, wearable and IoT inputs bring reality into the plan as fatigue cues, location awareness, temperature readings, and equipment status trigger small, targeted moves that protect safety and service. Vendors lean in across the stack, with AI scheduling from Legion and Quinyx, optimization and summaries inside UKG, Dayforce, and Workday, and practical signals flowing from Zebra handhelds, scanners, RFID, temperature probes, and Workcloud task execution that captures what happened without extra steps. The message is clear there too, the experience shifts from handcrafting each roster to an assisted loop with cleaner trade-offs and audit trails. Nucleus expects AI copilots, live signals, and transparent forecasts to become standard inside WFM in 2026, lifting fill rates, trimming premium hours, and steadying execution across frontline operations.

AI assistants will be built into WFM, helping managers create schedules and surface compliance issues, send shift messages, raise fill rates, and cut overtime.