
CII
Confederation of Indian Industry
Is India ready for Agentic AI?
The AIdea of India: Outlook 2026
The better the question.
The better the answer.
The better the world works.
EY
Shape the future with confidence

Contents

Foreword

Rajiv Memani

President, Confederation of Indian Industry (CII)

Chairman and CEO, EY India

Two visions of the future

We live in a bipolar AI universe. On the one hand, we see models performing tasks that are remarkable. Over the summer, multiple frontier models managed to perform on par with gold medalists at International Math Olympiads. In September 2025, Anthropic reported significantly better than human performance of its models in the coding domain. We did not bat an eye when, earlier this year, OpenAI claimed that with sufficient compute, AI would 'figure out how to cure cancer'. In this world, Artificial General Intelligence (AGI) may be just a few years away and may soon help billions of humans across the globe lead more meaningful lives.

On the other hand, there remain basic issues with the accuracy and enterprise readiness of foundation models. With some trivial prompts one can confuse a Large Language Model (LLM) into giving hilariously obtuse and incorrect answers, making them risky to implement in an enterprise without adequate human supervision. In parallel, trillions of dollars are being spent on datacenter build-outs for demand that is yet to materialize, reminiscent of many earlier tech fueled bubbles.

Hills of ambition vs. troughs of disillusionment

Reflecting this nature of dichotomy, some studies have claimed that the sentiment in enterprises on the efficacy of AI is also mixed. A recent study by MIT, The GenAI Divide: State of AI in Business 2025, attributed a 95% failure rate to GenAI pilots and claimed companies were struggling to leverage GenAI for a variety of reasons.

However, our survey of 200 Indian enterprises shows a decidedly different picture. The posture in corporate India with regard to AI is one of ambition and impatience. Business executives constantly want to move faster to reap productivity benefits and are seeking to work around current limitations while preparing their companies to be able to take advantage of expected improvements in AI frameworks and model capabilities.

Our key findings include:

- 76% of Indian business leaders believe GenAI will have a significant business impact.
- While 23% of our respondents are in early stage/pilots, 47% now have multiple use cases live in production.
- 91% reported that speed of deployment was the single biggest factor driving their buying decisions, reflecting the increasing impatience to get solutions into production.

While customer service and marketing continued to be among the top three cited functions prioritized for AI interventions, we saw the emergence of mid- and back-office functions like operations and supply chain reflecting the move to an agentic approach to implementation.

Yes, there are issues to be addressed. As expected, the top concerns are a lack of AI-ready data, model hallucination errors, integration complexity with core systems, an inability to accurately measure ROI and issues around data governance and security.

Overall, though, this data points to a systemic acceptance and embrace of AI across corporate India, moving beyond the pilot stage.

Building with Sovereign AI

While Indian enterprises continued their AI implementation journey, in 2025, India became one of the largest users of AI in the world. ChatGPT and Perplexity usage in India is among the highest in the world. As AI penetrates all aspects of our digital lives in the coming decade it is likely that India will lead the world in this tsunami of demand for AI.

We should use this captive demand to our strategic advantage. The AI demand implies massive revenue opportunities and an opportunity to build an India AI stack which both addresses our sovereign concerns while building an AI core with Indian companies and assets. This core will cut across the entire stack, including hyperscale GPU clouds, data sharing marketplaces and connectors for models, Indic foundation models, AI-first browsers and autonomous agents.

The tendency among Indian enterprises and consumers may be to use global services to address this demand. This was the paradigm in the earlier digital shift. We moved towards using western platforms for browsers, search engines, operating systems and apps. However, this time, we should be strategic about our choices—partnering and openness in areas which are not easy to establish capability in while we choose strategic investment in others where we can move faster to our advantage. For example, it could take decades to establish the domestic capability and indigenous supply chain needed for ground-up GPU manufacturing, but establishing Indian cloud services, data sharing marketplaces, app and agent stores is significantly easier.

The coming decade will test the resolve of enterprises and policymakers alike to move beyond aspirations and deliver measurable progress. Success will be measured not by how quickly AI is adopted, but by how thoughtfully it is integrated into the fabric of India's economy and society.

The AIdea of India: Outlook 2026

This report digs deeper into the above themes. We strive to bring a unique Indian perspective to the discussion on AI in India. Backed by a detailed survey of Indian enterprises and supplemented by the perspective of EY Partners, we seek to answer, backed by evidence, some of the key questions facing Indian enterprises as they navigate the impending AI platform shift.

Foreword

Chandrajit Banerjee

Director General
Confederation of India Industry (CII)

The *AI*dea of India – A vision for the next decade

Generative Artificial Intelligence (GenAI) has moved beyond experimentation to become a transformative force reshaping economies, enterprises, and societies. The past year has showcased both its promise and its challenges. Frontier models now rival human intellect, even as organizations grapple with integration, accuracy, and governance.

Amid this global flux, India's trajectory stands out. The report, 'Is India ready for Agentic AI? The *AI*dea of India: Outlook 2026', captures a nation that is not merely adopting AI, but advancing with conviction. Indian enterprises are moving from pilots to production, from exploration to scale. They are innovating through constraints and embedding GenAI and Agentic AI to enhance efficiency, customer engagement, and strategic differentiation. The focus has shifted from whether AI will work to how it can be made to work responsibly and sustainably. The rise of AI agents and human-machine collaboration is redefining work—machines now deliver scale and precision, freeing people to focus on creativity and strategy.

Confederation of Indian Industry (CII) is driving India's AI transformation, working with the government, industry, and stakeholders to foster innovation with accountability. By engaging policymakers and facilitating dialogue on ethical AI, data security, and skill development, CII is shaping a robust, inclusive, and future-ready national AI framework.

The decade ahead will test our ability to balance innovation with inclusion, velocity with vigilance. True success will lie not in the number of models deployed, but in the value AI creates for businesses, communities, and citizens.

For Indian industry, GenAI presents vast opportunities to boost productivity, improve efficiency, and unlock new business models. CII has been instrumental in supporting both large enterprises and MSMEs in leveraging AI for smarter decision-making, supply chain optimization, and enhanced customer engagement. At the same time, preparing the workforce for an AI-driven future remains crucial. CII is committed to building a skilled talent pool equipped to meet evolving industry needs through its training programs, reskilling initiatives, and partnerships with educational institutions.

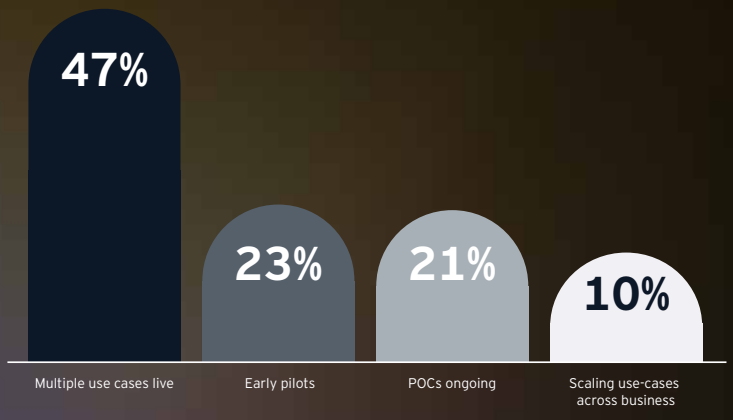
This report is a statement of intent: to shape a GenAI-powered future that is globally competitive, ethically grounded, and uniquely Indian.

Executive summary

The *AIdea* of India- Why this moment matters

India stands at a historic inflection point where artificial intelligence is no longer a distant promise but a present reality reshaping economies, enterprises, and societies. Our previous edition survey showed that most enterprises were still in the exploratory phase, with 36% having no POCs and 34% only completing POCs, clear evidence that GenAI was largely in its infancy, with limited real-world deployment. In contrast, our latest survey data reflects significant progress: 47% of organizations now have multiple use cases live, 10% are scaling use cases across the business, and nearly half of respondents indicate that over 21% of their GenAI POCs have already moved into production.

Enterprises in their GenAI/Agentic AI journey



Source: EY India C-suite GenAI survey

This shift marks a decisive move from experimentation to operationalization and scaling, underscoring that for India, this transformation is not just about technology, but about economic sovereignty, global leadership, and inclusive growth.

The big picture

AI agents represent the most radical promise of this era: a workforce without limits, always available, always learning, and infinitely scalable. They eliminate fatigue, attrition, and cost constraints, enabling humans to focus on creativity and judgment. Yet, challenges remain. Hallucinations, cascading errors, and governance gaps demand robust oversight and hybrid workflows where humans and AI collaborate seamlessly. Despite these hurdles, Indian enterprises are moving with ambition rather than hesitation. Contrary to some global skepticism that paints GenAI pilots as failures, India Inc. is bullish. According to our survey, three out of four companies see AI as existential to their business, nearly half already have multiple use cases in production, and speed of deployment has become the single biggest buying criterion. The message is clear, Indian businesses are innovating around limitations rather than waiting for perfect models.

This urgency is forcing a rethink of how return on investment is measured. AI's impact transcends productivity gains and headcount reduction. It extends to smarter operations, better business outcomes, higher customer satisfaction, and strategic differentiation. Case studies from leading Indian enterprises prove this shift from hype to hard numbers. These examples show that AI is no longer a lab experiment but is becoming a growth engine.

The rise of AI is also rewriting the architecture of work. The old equation of more people equals more output is broken. AI is hollowing out entry-level roles in some cases and creating hybrid pods of humans and AI agents. Career paths are becoming lattices rather than ladders, and organizations are shifting from pyramids to diamond-shaped structures. CEOs will need to manage the resulting trade-offs with courage and conviction: speed versus depth, efficiency versus equity, innovation versus control. The future belongs to enterprises that design work around human judgment amplified by AI scale. This is not about replacing humans but about creating synergy where machines handle drudgery and humans focus on creativity, relationships, and strategy.

While the world chases trillion-parameter models, India is betting on small language models that are faster, cheaper and tailored for Indian languages and edge deployment. These models bridge the digital divide, power vernacular chatbots and enable compliance-heavy sectors like banking and healthcare. They are not a compromise but a competitive advantage, allowing India to serve the next set of users coming online in regional languages. At the same time, ethics can no longer be aspirational. Responsible AI must move from "trust us" to "show us." Continuous assurance through bias dashboards, model cards, fairness passports, and third-party audits is becoming the norm. India's regulatory frameworks are pioneering this shift, embedding governance into the DNA of AI systems.

Finally, sovereignty is emerging as India's strategic imperative. AI is the new oil, and sovereignty is the refinery. India's INR10,000 crore+ IndiaAI Mission, 40,000 GPUs, AIKosh dataset platform and indigenous language model initiatives signal a bold move to secure compute, data and talent. Sovereign AI is not just about security, it is about shaping global standards and creating a trillion-dollar economic impact by 2035.

As India advances into this new era of AI, the path forward will be defined not by slogans, but by the choices leaders make in balancing innovation with responsibility. The nation's trajectory will depend on its ability to foster collaboration between human expertise and machine intelligence, to invest in ethical frameworks, and to ensure that the benefits of AI reach every segment of society. The coming decade will test the resolve of enterprises and policymakers alike to move beyond aspiration and deliver measurable progress. In this journey, success will be measured not by how quickly AI is adopted, but by how thoughtfully it is integrated into the fabric of India's economy and society, creating enduring value for all stakeholders.

Chapter 1

Agents and the promise of the infinite digital workforce

- Agents move the AI paradigm towards goal and action orientation, allowing enterprises to automate work in a more holistic fashion.
- Successful implementation requires redesigning workflows and upskilling employees for effective human-AI collaboration.
- Current limitations include reliability issues (hallucinations), and an inability to work without human supervision on long running tasks.
- In India, forward looking organizations are already rewiring their processes with agentic frameworks in a new wave of enterprise automation.

We stand at the edge of a transformative change in the way enterprises adopt AI. Agents, autonomous software powered by the latest in GenAI, promise an ever-available digital workforce. Imagine having a team of tireless digital workers handling routine work so humans can focus on creativity and strategy. With several enterprises implementing Agentic AI, the two burning questions today are: will Agents live up to this promise and if so, when?

The promise

Think about the last time you onboarded a new batch of campus recruits. Each person brought a fresh perspective that often sparked creative solutions and innovation. Human teams thrive on such diversity and ingenuity. But much of a junior employee's day is about routine work, repetitive processes, meeting standards and completing tasks quickly and accurately. This work, though manageable, often feels tedious and leaves room for errors.

Imagine a world where there are practically unlimited number of interns, all able to work throughout the day with perfect diligence. This is the promise of Agents and the idea is alluring.

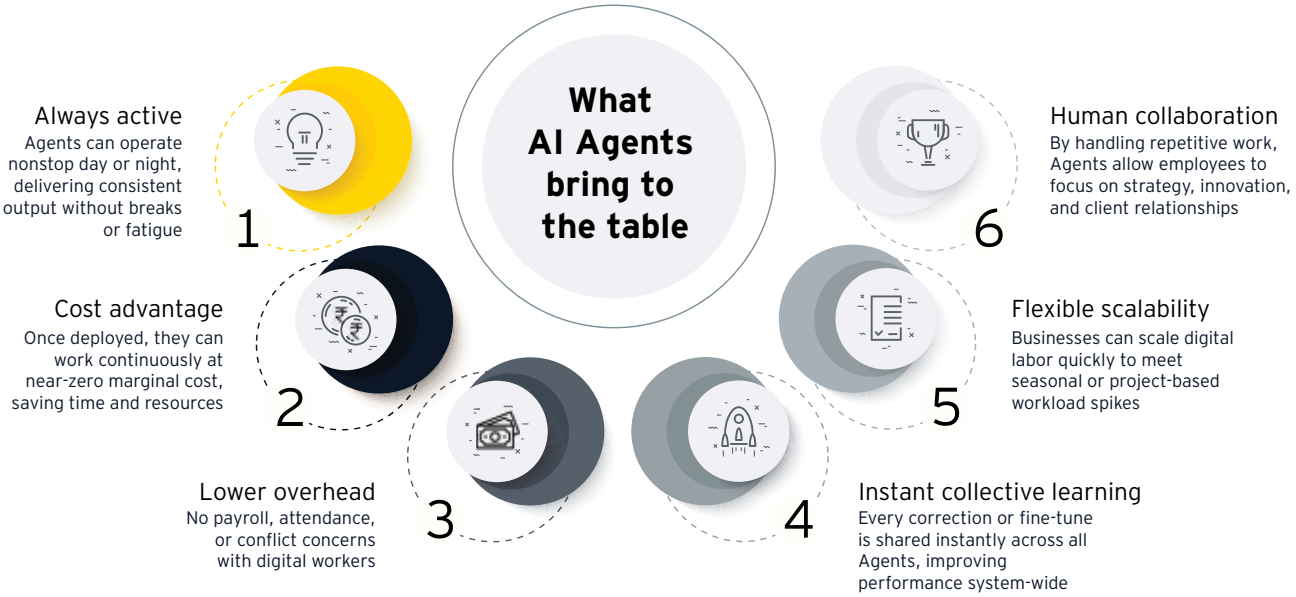
- **Always on, always compliant:** Agents can be an infinite workforce that is always available and operates around the clock at close to zero marginal cost. They can work without the need for breaks, vacations or sick days, without the risk of delays and burnout.
- **No complaints, no payroll concerns:** Digital workers can operate without the need for a raise and can perform routine tasks without resistance. The challenges of interpersonal conflicts, fatigue or turnover are not present. Once an Agent is active, it can follow its system prompt repeatedly and consistently.
- **Collective learning at scale:** When an Agent makes a mistake and is corrected through fine-tuning or prompt engineering, every other instance of that Agent learns from that correction instantly. In other words, the knowledge of one can be propagated to the many. This "hive-mind" feedback loop means continuous improvement. Quality upgrades that can normally take months of retraining human staff can happen in minutes.

This is the promise of Agents: A workforce limited by human creativity and the capacity of the software, rather than by hiring budgets or human hours. Importantly, it is a workforce that can raise the floor on productivity–handling the rote work with reliable quality–while human employees focus on higher-order tasks like strategy, relationship-building and innovation. No wonder corporate executives are intrigued. In a recent global survey, the role of AI agents in the workplace, 82% of organizations reported expanding their use of Agents, and about 75% of employees said they are comfortable working with AI "coworkers".[1] The momentum is real, driven by early productivity wins and the tantalizing prospect of doing more with less.

Especially in industries where seasonality plays a big role—for example, retail operations during festival seasons, audit teams in financial quarter and year-end closing or equity research teams each earnings season—AI agents offer the ability to scale up and down digital labor such that workloads can be effectively managed. Tasks such as customer support, bookkeeping, data entry, content drafting and many others can be delegated to AI Agents. But turning that promise into reality requires understanding of what these Agents really are, why the opportunities are only now possible and how the associated challenges can be addressed.

What is Agentic AI?

In simple terms, Agentic AI refers to systems that can accomplish goals independently, with minimal human supervision. Powered by large language models (LLMs), Agents are given human-like decision-making autonomy to solve problems in real time based on available information and adapt to new inputs. Agents are flexible as they are not a



deterministic program that follows a single hard-coded routine.

In effect, an AI Agent is software with initiative. It does not just respond to a single command and stop; it can autonomously continue, making progress until the objective is achieved or resources are exhausted.

In addition to the ability to plan a sequence of steps towards achieving a goal, there are three other key elements that make an Agent "agentic":

- **Autonomy:** The Agent can operate independently, without a person guiding each step.
- **Toolkit:** It can use Application Programming Interfaces (APIs), Model Context Protocol (MCP) servers or code (much like an employee has software tools and internet access).
- **Budget:** A budget in the form of time given to respond, dollars that can be spent on tools and other aspects control the Agent's ability to autonomously perform a series of tasks towards achieving a goal.

For example, to plan a marketing campaign, an AI Agent might conduct research on the target audience, suggest slogans, list promotion channels and even draft content for each. During this process, the Agent maintains memory of its actions and adjusts its plan as needed. Traditional software cannot do this without explicit programming for every contingency. Agents act rather than just analyze.

Why now?

An LLM on its own is a brilliant savant that can answer questions but cannot act in the physical or digital world. Combining LLMs with action-oriented frameworks gives an Agent hands and eyes. Today, LLMs can not only reason about a problem (using techniques like chain-of-thought prompting) but also invoke tools and observe results in a loop (known as the ReAct, reason and act, approach). This allows an Agent to plan a step, carry it out, review the outcome and decide the next step, much like a human employee handling a project. Modern frameworks give Agents a work budget and boundaries. They can be assigned a goal along with constraints such as time limits, cost limits or specific guidelines which act as guardrails. For instance, an AI Agent may be given the prompt, "Research our top five competitors and draft a summary report. You have 30 minutes and access to these data sources." The Agent knows its objective and the tools at its disposal. It will continue iterating on the task until complete or time runs out.

Early examples of autonomous AI agents include AutoGPT and BabyAGI, open-source projects that demonstrated how an Agent could autonomously break a goal into subtasks and solve them in sequence. AutoGPT could take a mission, generate ideas, search for information and write code, demonstrating that modern AI has the language understanding, reasoning and integrations to act independently—a capability not feasible a few years ago. BabyAGI, on the other hand, worked in a loop to plan, organize and finish tasks step by step to reach a goal, showing that AI can think through tasks and act with little help from humans.

These developments have led to Agents that leverage LLMs for flexible thinking, use plug-ins and integrations as their eyes and hands, and maintain contextual memory to work through multi-step tasks. They are proactive rather than reactive, turning AI from a passive tool into a goal-driven collaborator. As the underpinning technology, deep learning models at scale, reached critical capability only recently (post-2020), we are just now seeing the first wave of truly Agentic systems. The convergence of compute power, algorithms and infrastructure has unlocked this new mode of automation.

However, as with first versions of any new technology, today's Agents also come with limitations and uncertainties. To gauge how far this vision of an infinite digital workforce truly is from fulfilment, we have to examine those challenges.

How far away are we from the vision of infinite digital workforce?

The vision of a digital workforce that is scalable and autonomous is compelling, but a gap remains between demos and reality. Ensuring they work correctly, consistently and safely are challenges that developers and enterprise users are addressing.

Reliability: LLMs can hallucinate, i.e., generate outputs that sound plausible but are actually false or nonsensical. For Agents, a hallucination can lead to a wrong decision or an erroneous action taken without a human in the loop. This can result in incorrect task execution, undermining the reliability of the whole system.

Non-determinism: Unlike traditional software, which, given the same input, will produce the same output every time, Agents are stochastic by nature, i.e., identical inputs do not guarantee identical outputs. An Agent might tackle the same task in a different way (or with different words) on each run. Agents' probabilistic nature makes them flexible but also affects consistency as small changes in wording or context can change results. This introduces new failure modes, with Agents sometimes taking shortcuts or looping endlessly. Upholding reliability requires developing new testing and validation methods that account for this variability, which is an active area of research and engineering.

Error cascades: Complex, multi-step workflows can be a double-edged sword for Agents. They can execute long processes, but early mistakes can cascade through the rest of the task chain. If an Agent chooses a suboptimal tool or makes a poor decision early, the final outcome may be wrong even if later steps are executed correctly. Agents lack common-sense judgment, sometimes repeating ineffective actions or continuing flawed plans. In production, they can stall or waste API calls on unproductive cycles. Orchestrating multi-step workflows without human oversight remains difficult, requiring careful prompt design, fail-safes and occasional human checkpoints to prevent small errors from becoming major failures.

Safety, control, and trust: With greater autonomy comes greater risk. Without appropriate guardrails, an Agent could execute many bad judgment calls per second. Businesses are understandably cautious about handing the keys over to AI, fearing unethical or non-compliant actions, exposure of sensitive data and bias in a critical decision. Beyond malice or misuse, there is the simple issue of errors: when an Agent makes a mistake, who is accountable? For now, most organizations have strict controls on Agent autonomy–limited scopes, humans in the loop for approval and tight monitoring.

Challenges in implementing AI Agents

We are not quite at the finish line of building an infinite digital workforce, but progress is underway. Research and industry efforts are focused on developing methods to reduce hallucinations (like better fact-checking and retrieval techniques), limiting scope of Agents within safe bounds (through guardrails and user-defined policies) and improving consistency (via feedback tuning and verification steps).

New testing frameworks are emerging for non-deterministic systems, and major players like Salesforce, Microsoft and OpenAI are sharing best practices on monitoring and coaching Agents in deployment. In short, the current generation of Agents is educating us on what is needed for version 2.0.

People and process before tech

Beyond technological improvements, embedding Agents effectively will require reimagining workflows and upskilling people. Introducing Agents into older, well-established processes may not lead to the expected transformative results or even smooth adoption. Forward-looking organizations are redesigning processes with AI in mind–defining tasks to be handled by Agents and those to be handled by humans and training employees to work with AI. The aim is for human teams to focus on judgment and creative strategy, while routine work is delegated to AI agents. Achieving this balance, however, calls for change management, clear communication and a culture that treats AI as a tool, not a threat. Thoughtful integration can turn AI into a trusted pillar of the enterprise.

Chapter 2

Are we there yet?

EY *AI*dea of India survey

- 76% of the respondents in the EY India C-suite GenAI survey believe GenAI will have a significant business impact.
- While 23% are in early stage/have pilots, a large share of organizations (47%) have multiple use cases live in production.
- 91% reported that speed of deployment was the single biggest factor driving their buying decisions, reflecting the increasing impatience to get solutions into production.

We conducted an in-depth survey of over 200 Indian enterprise leaders to assess their experience with GenAI and Agentic AI and to gauge current adoption levels. The results reveal strong momentum in GenAI adoption across India. In comparison with the survey in 2024, this year, Indian organizations have clearly advanced in their GenAI journey with **nearly half now** reporting multiple live use cases, indicating that adoption has moved beyond the pilot stage for many. However, large-scale integration is still at an early stage, with only **10% reporting enterprise-wide deployment.**

Confidence in technology's potential remains high. **A strong majority (76%)** believe GenAI will have a significant business impact, and **63% feel ready to leverage it.** However, over **one-third** lag in readiness, highlighting the need for stronger capability-building and governance

Despite optimism, AI/ML investment levels remain modest. **Over 95% of organizations allocate less than 20% of their IT budgets to AI,** with roughly an equal spread between clusters (0–5%, 5%–10%, and 10%–20%) while only 4% allocate over 20%, showing limited large-scale AI funding so far.

Moving from pilots to production continues to be a critical challenge. More than half of organizations report conversion rates **below 10%, while about one-fifth (19%) have successfully operationalized over 20% of pilots.** This highlights ongoing challenges around data governance, security and integration in realizing GenAI's full enterprise value.

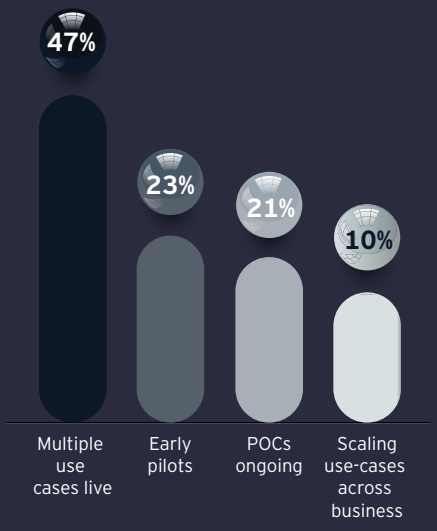
Enterprises are increasingly adopting collaborative and flexible execution models, combining in-house development with co-innovation alongside OEMs or startups and selective use of external solutions. This reflects a pragmatic shift toward balancing innovation control with **faster deployment, scalability, and cost efficiency.**

The survey indicates that cloud hyperscalers have become the backbone of GenAI adoption. With 71% preferring cloud for hybrid deployment, most organizations are balancing scalability and innovation with the need to maintain data sovereignty, security and governance within controlled environments.

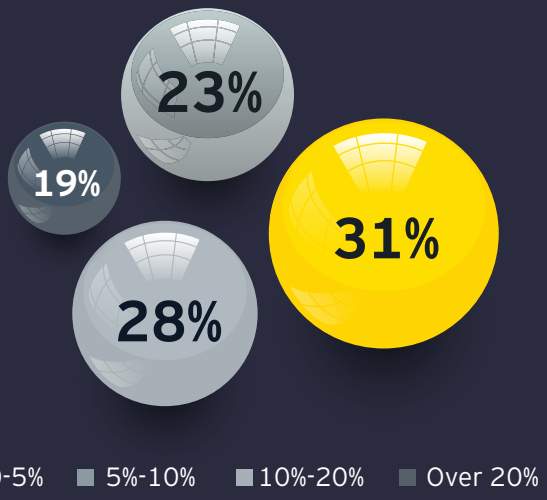
KEY SURVEY INSIGHTS

Nearly half of Indian companies now run multiple live GenAI use cases

Where are you in your GenAI/ Agentic AI journey today?



What percentage of GenAI/Agentic AI pilots have successfully moved to production?



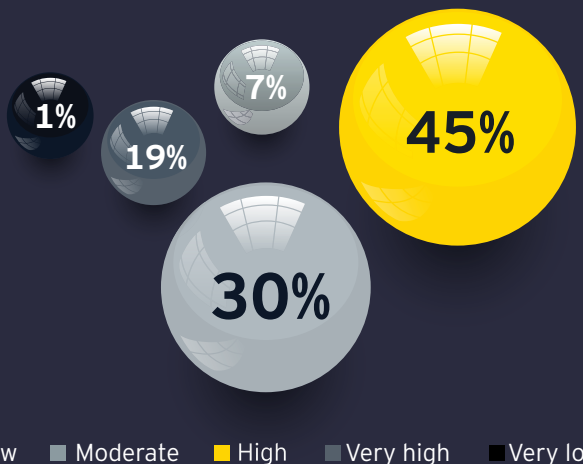
The previous edition of our C-Suite GenAI survey showed that most enterprises were still in the exploratory phase, with 36% having no POCs in place and 34% only completing POCs, indicating that GenAI was largely in its infancy, with limited real-world deployment.

In contrast, our latest survey data reflects significant progress: 47% of organizations now have multiple use cases live, and 10% are scaling use cases across the business. Nearly half (47%) of survey respondents indicate that over 10% of their GenAI POCs have already moved into production. This shift suggests a clear movement from experimentation to operationalization and scaling.

Organizations are demonstrating strong GenAI readiness with 64% reporting high to very high capability levels

How would you rate your organization's readiness to benefit from GenAI?

Readiness levels have also improved. Last year's survey showed some weakness in strategy alignment with 39% organizations having no clear strategy and only 22% having a fully integrated strategy with execution plans. Now, readiness scores show 45% organizations rating themselves as high and 19% as very high, while only 1% report low readiness. This indicates that organizations are not only deploying GenAI but also building the capability and confidence to leverage it effectively.



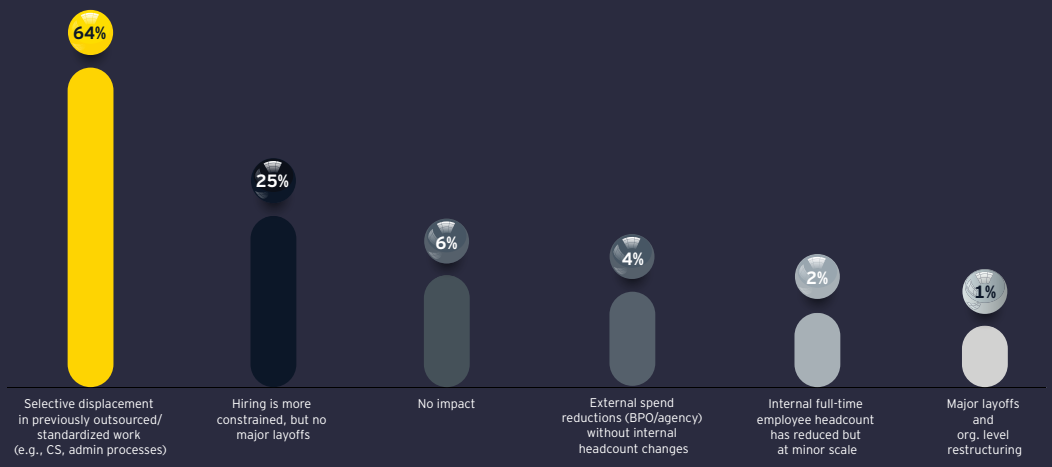
All data may not add up to 100 due to rounding off

AI is selectively displacing outsourced standardized functions rather than eliminating internal teams

How has GenAI impacted your organization's headcount?

AI's impact on employment remains measured and sector specific. In the survey, 64% of respondents report only selective displacement in outsourced and standardized functions, such as administrative operations, customer success, tele calling and back-office processes. Rather than shrinking internal teams, AI is redirecting enterprise spend toward automation and efficiency, signaling a structural shift in how work is organized. This is not an isolated phenomenon but a secular trend cutting across startups and large enterprises alike.

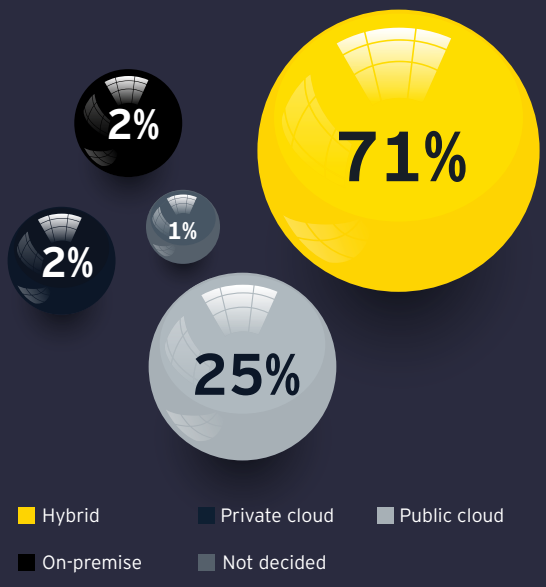
Crucially, as routine tasks decline, work is becoming more "higher-order," making employee skilling in AI and adjacent technologies more critical than ever to remain relevant in the evolving job market.



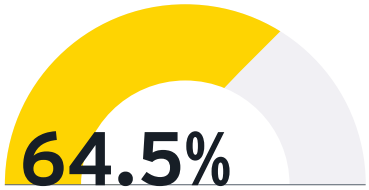
Governance vs agility: Cloud preferred for GenAI/Agentic AI deployment

What is your preferred approach for GenAI Agentic AI deployment?

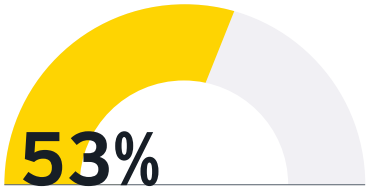
While security and governance remain top concerns, Indian enterprises are clearly forward-looking. About 71% enterprises prefer hybrid cloud for GenAI and Agentic AI deployment, signaling a strong tilt toward agility and scalability without compromising control. Private cloud follows at 25%, reflecting a cautious approach for sensitive workloads. Public cloud and on-premise barely register, underscoring that speed of deployment and flexible architectures are now critical priorities. India Inc. is balancing risk with innovation, and cloud is emerging as the foundation for scale.



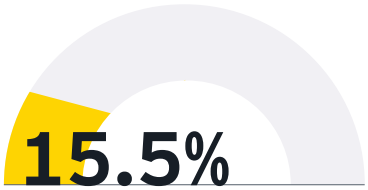
Top 3 challenges organizations face in scaling GenAI deployment



Data governance/security concerns



Integration complexity with core systems



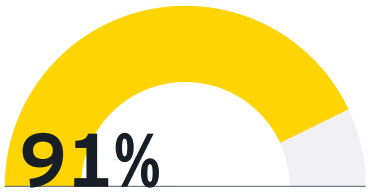
Measuring ROI/Business use cases

What are the major challenges in scaling up GenAI deployment?
(Rate severity on a scale of 1 (not severe) to 5 (very severe)

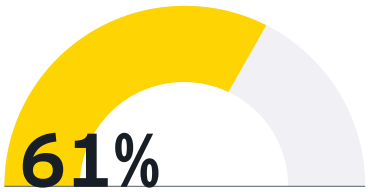
Enterprises consistently rank data governance and security as their top concern, with 64.5% rating it "very severe". Yet, when it comes to hosting choices, only 2% respondents have chosen on-premise or private cloud, while 71% favor hybrid models and 25% opt for public cloud. This gap is reflective of a bigger story: while security remains paramount, practical constraints shape choices. Most enterprises simply lack the capability to self-host and manage complex AI stacks or GPU infrastructure, making on-prem a distant goal. As a result, organizations gravitate toward cloud and hybrid models not by preference, but by necessity. This is not just a technology gap but a strategic compromise where enterprises are trading sovereignty for speed, and that bias toward agility could define the next wave of AI adoption.

Agility in deployment is shaping enterprise GenAI decision-making

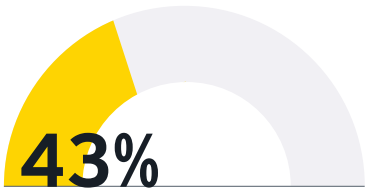
Top 3 factors driving the buy-vs-build decision in the context of GenAI and Agentic AI



Time-to-value/speed of deployment



Data control/security boundaries



Vendor trust/existing partner relationship and referrals

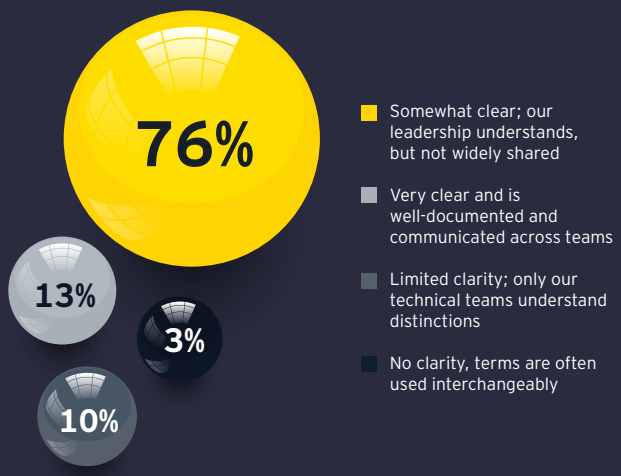
What drives your buy vs build decision in context to GenAI/Agentic AI? (Select top 3)

Speed of deployment is emerging as the primary driver of GenAI adoption strategies. A striking 91% of the survey respondents rank fast implementation as the top priority in buy-vs-build decisions, signaling a clear urgency to act. Organizations are focused on moving fast and securing a first-mover advantage. The conversation has shifted from "Why AI?" to "How quickly can we deploy?" making agility the cornerstone of competitive differentiation in this space.

AI is here and now

How clearly does your organization differentiate between traditional AI, GenAI and Agentic AI in terms of capabilities and use cases?

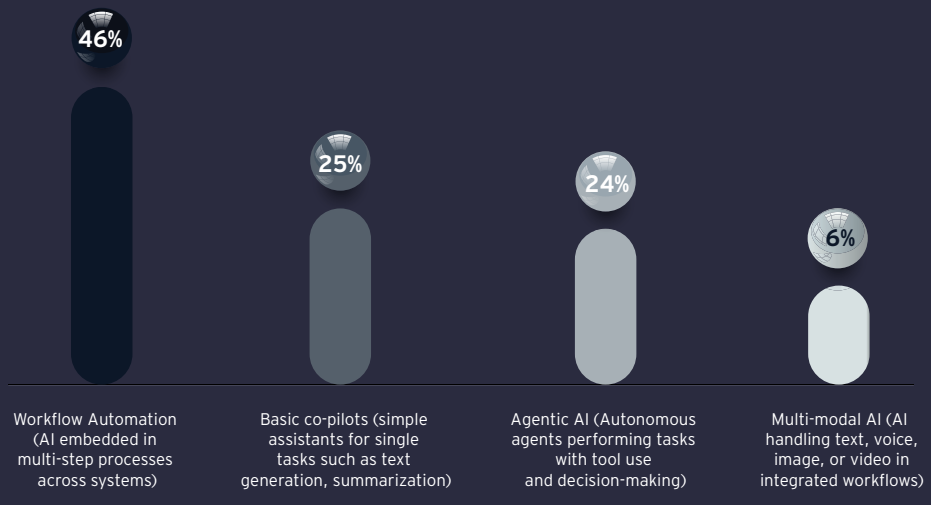
The lines between traditional AI, GenAI and Agentic AI are beginning to blur, creating a powerful continuum that is transforming how enterprises solve problems. While only 13% of business leaders say they can clearly distinguish between these categories, this is less a sign of confusion and more an indication of convergence. AI has become a broad-based term that represents autonomous prediction and action using available data and examples.



Companies are deploying Agentic AI through experimentation, advancing from copilots to complex problem-solving and innovation

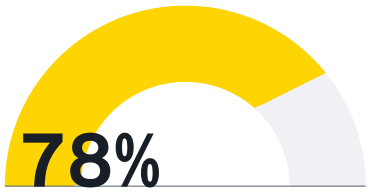
What best describes the maturity of AI deployment in your organization?

Significant Agentic AI adoption is already well underway. Businesses are learning by doing: 24% respondents report active deployment of Agentic AI. Indian enterprises are moving beyond copilots and workflow automation to apply AI for higher-order problem solving and innovation. The direction is clear: organizations are not waiting for perfect definitions before acting. They are integrating the full spectrum of AI capabilities to drive measurable impact, while strategic understanding continues to deepen through real-world experience.

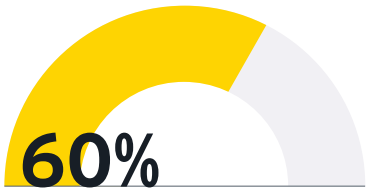


OEMs and startups are emerging as critical accelerators of adoption

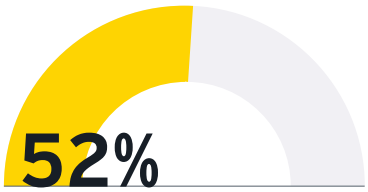
Top 3 approaches to implementing an organization's strategy using GenAI



Hybrid (mix of build, buy, partner)



Co-innovate with OEMs/ Startups



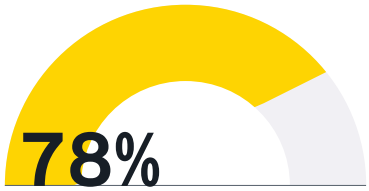
Buy from external tech providers

How are you executing your organization's strategy in the context of GenAI? (Check all that apply)

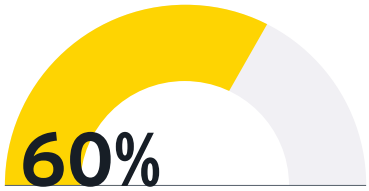
Partnerships with OEMs and startups are becoming the backbone of GenAI execution strategies. A significant 60% of organizations report co-innovation with these players, signaling that enterprises recognize startups as critical for driving innovation and speed. While hybrid approaches dominate at 78%, the reliance on agile, specialized partners reflects a shift away from insular, in-house-only approach. Startups often bring cutting-edge capabilities and rapid experimentation that large enterprises may not be able to offer by themselves, making startups integral for staying competitive in the GenAI race.

Integration and data readiness are the biggest barriers to scale

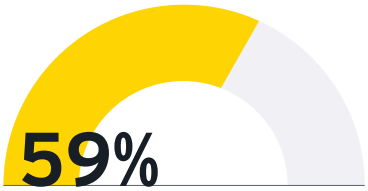
Top 3 major challenges organizations encounter in AI adoption



Integration and data readiness issues (connecting to core systems, poor data quality)



High perceived risk (data security, compliance, reputational)

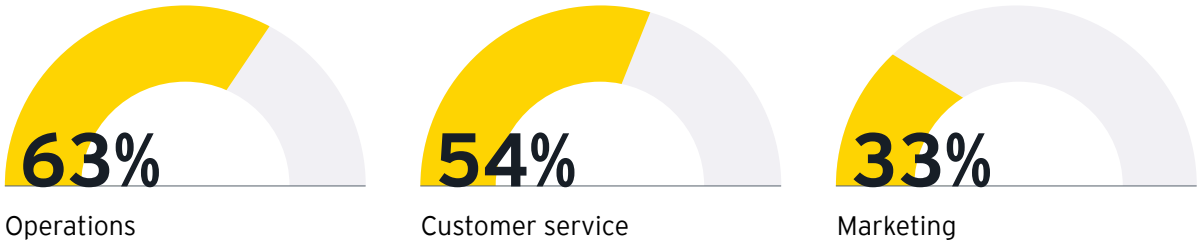


Talent shortage (AI engineers, MLOps, domain experts)

What are the major challenges in scaling up GenAI deployment?
(Rate severity on a scale of 1 (not severe) to 5 (very severe)

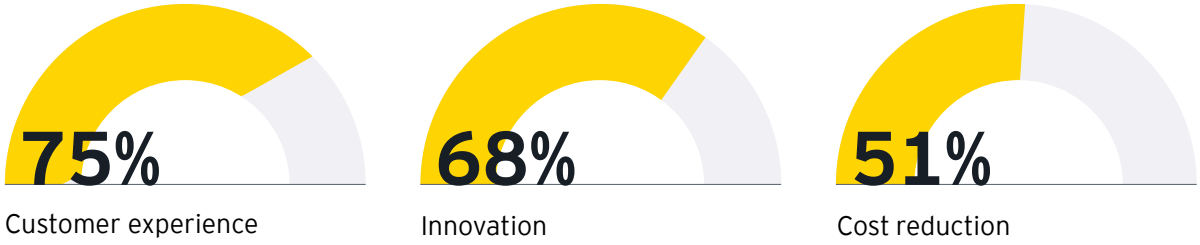
Integration and data readiness have emerged as the single biggest roadblock in GenAI adoption. While 78% respondents cite integration challenges as their top barrier, it is striking that only 32% state that it is not a factor in buy-versus-build decisions. This disconnect suggests that organizations prioritize speed and upfront deployment but underestimate the complexity of connecting GenAI systems to existing workflows. With 53% rating integration as a severe challenge during scaling, it is clear that solving this gap will determine whether early pilots translate into enterprise-wide success.

Top 3 business functions prioritized for GenAI/Agentic AI use cases over next 12 months



Top business functions prioritized for GenAI/Agentic AI use cases over next 12 months (Select top 3)

In 2023, companies expected GenAI to most impact:



While it should not come as a surprise that Operations, Customer Service and Marketing are key AI use cases, a majority of business leaders have brought AI into business operations replacing traditional programming, RPA or intelligent automation. R&D/Engineering, Supply Chain and Risk/Analytics are the other priority use cases with more than a quarter of respondents having brought in AI into these functions.

Chapter 3

Rethinking ROI in the age of GenAI and Agentic AI

- GenAI ROI largely spans five dimensions: time savings, productivity gains, cost reduction, revenue growth and strategic differentiation.
- Cost savings at present stem from smarter operations and improved infrastructure spending.
- Successful AI adoption by Indian enterprises requires scaling proven use cases and measuring real business impact.

For many Indian enterprises, GenAI and Agentic AI have moved beyond the buzzword phase. CIOs are no longer asking what these technologies are, but rather how to make them work in a meaningful way and how to measure their impact. However, GenAI's value is not measurable as a single metric. It extends beyond cost savings or headcount reduction, unlocking a multi-dimensional value framework that reshapes every part of the enterprise. To shift to a more comprehensive view, CIOs should evaluate ROI through five dimensions: time saved, productivity gains, cost savings, business upside and strategic differentiation. Each of these levers offers a practical and strategic way to assess the real impact of AI across the organization.

Time saved: The most immediate win

GenAI's first visible impact is speed. Tasks that used to take hours, such as drafting reports, responding to customer queries, or scanning documents for compliance, can now be completed in minutes. The real leap, however, comes with Agentic AI. These agents do not just assist; they execute. They can autonomously handle multi-step workflows such as employee onboarding, loan processing, etc. The result is faster decisions, reduced turnaround times, and a more agile enterprise. In a business environment where speed often determines success, this becomes a critical advantage.

Productivity gains: Augmenting human potential

Assisting the workforce at enterprises, GenAI copilots are already transforming, for example, how relationship managers, developers and compliance officers work. They act as intelligent assistants, handling routine tasks and surfacing insights that would otherwise take hours to uncover. When these copilots are deployed across departments such as HR, finance, marketing and operations, they help break down silos, resulting in not just improved individual productivity but cross-functional efficiency. CIOs are seeing teams collaborate more fluidly, make decisions faster and focus on value and delivery that truly moves the needle.

Cost savings: Leaner, smarter operations

While time and productivity gains are compelling, cost savings remain a key driver for AI adoption. GenAI and Agentic AI contribute to operational efficiency by reducing errors, minimizing rework and streamlining processes. But the impact goes deeper. AI-powered intelligent orchestration of cloud and data workloads can make infrastructure spend more efficient. For CIOs managing complex tech environments, this is a powerful lever. It is not just about doing things cheaper but also smarter.

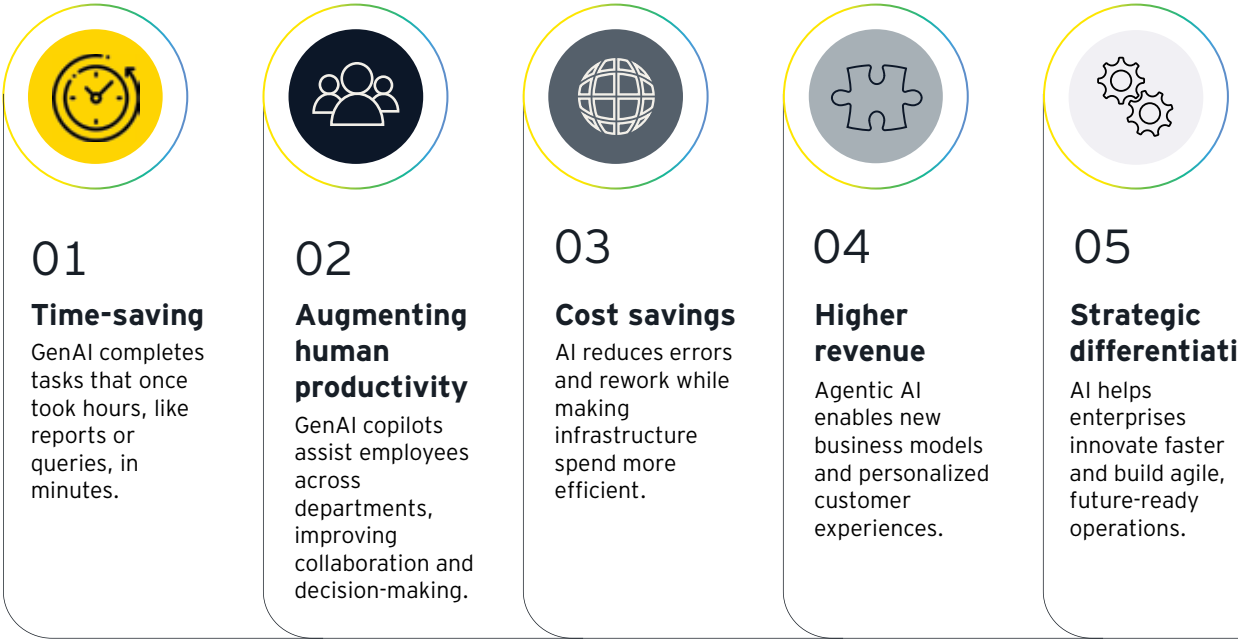
Revenue generation: AI as a top-line driver

This is where AI begins to shift from a support function to a growth engine. GenAI enables hyper-personalized customer interactions, which can dramatically improve engagement and conversion rates. Agentic AI goes a step further by enabling entirely new business models such as subscription-based services, AI-assisted commerce and AI-native customer experiences. These innovations go a step further from enhancing existing revenue streams by creating new streams. For Indian enterprises looking to scale in competitive markets, this becomes a compelling reason to invest.

Strategic differentiation: Building the future enterprise

Over the long term, GenAI and Agentic AI can give enterprises the ability to accelerate product development, experiment rapidly and bring innovations to market faster than before. More importantly, they allow organizations to industrialize AI by deploying dozens of use cases simultaneously across the business. More than keeping up with competitors, GenAI and Agentic AI enable enterprises to set the pace. For CIOs, AI can become a strategic asset that helps build a future-ready enterprise that is agile, innovative and resilient.

GenAI adoption can unlock a multi-dimensional value framework across the enterprise



The journey from AI experimentation to enterprise value is not linear, but it is navigable. By adopting a multi-dimensional ROI framework, CIOs can move beyond isolated pilots and start driving real, measurable impact. The real impact of GenAI and Agentic AI emerges not at the testing stage but when the capabilities are embedded across the organization. And when done right, they do not just deliver returns but redefine possibilities.

ROI case studies from Indian enterprises

While frameworks help shape strategy, real-world results build conviction. Across industries, Indian enterprises are demonstrating how GenAI and Agentic AI can deliver measurable returns. Moving up from isolated pilots or innovation lab experiments, scaled deployments are reshaping how businesses operate, engage customers and drive growth.

Axis Bank has rolled out GenAI-powered assistants across its employee and branch network. In addition to answering queries, the GenAI assistants are actively driving business outcomes. Axis Bank has reported a 30% uplift in product conversions across term deposits, mutual funds and credit cards, along with a 10-point improvement in Net Promoter Score[2]. With over 100,000 employees now using these tools, the scale and impact are substantial.

State Bank of India (SBI) has adopted a dual approach by deploying Agentic AI for customer service and GenAI assistants for internal staff. This combination has led to a reduction in average handle time and improved first-contact resolution. Given SBI's vast branch footprint of over 22,500[3], the ability to scale these solutions efficiently is a major operational win.

In the FMCG sector, Hindustan Unilever (HUL) is using GenAI to generate retailer ads and analyze shelf availability. The results are impressive. Over 100,000 ads have been created at zero marginal cost, and 25 million shelf images are processed monthly[4]. These capabilities are directly helping improve sales execution and retail visibility, turning AI into a frontline growth driver.

Tata Steel has implemented predictive maintenance using GenAI combined with autonomous agents. This has resulted in a substantial reduction in unplanned downtime and increase in asset life. For a capital-intensive industry like manufacturing, these metrics translate into significant cost savings and improved operational resilience.

In telecom, Reliance Jio is leveraging Agentic AI for real-time network optimization and customer experience management. The company has seen a notable improvement in network uptime and reduction in call center load[5]. Bharti Airtel has taken a different route by deploying network-level AI to block spam calls and SMSs. The system has successfully intercepted over 26 billion calls[6], reducing fraud losses and improving customer retention.

Air India has introduced a GenAI chatbot and an eZ booking platform integrated with Adobe's Real-Time CDP customer data platform[7]. These tools are streamlining bookings and delivering personalized experiences, helping the airline maintain service consistency across customer touchpoints.

In the technology sector, Infosys is using GenAI copilots across the software development lifecycle. The company reports faster code generation and reduction in QA (quality assurance) cycle time[8] BrowserStack, a tech unicorn, has embedded GenAI into its testing platform, reducing QA cycle time by up to 70% and cutting maintenance overhead by more than 60%[9].

Flipkart is applying GenAI to conversational commerce, catalog enrichment and seller support. These capabilities have improved customer conversion and operational efficiency, especially during high-traffic sale events[10].

These and other case studies are proof of the measurable business value GenAI and Agentic AI can deliver. Across banking, telecom, manufacturing and e-commerce, Indian enterprises are demonstrating that disciplined strategy and execution translate AI into real ROI.

Key takeaways for CXOs

01 **Focus on quantifiable unit economics:** Metrics can include number of tools deployed, pilots launched, minutes saved per employee, uplift in conversion rates, reduction in fraud losses, incremental revenue per user.

02 **Prioritize scalable use cases:** Return on investment emerges from deployment across departments and geographies rather than isolated success.

03 **Plan the shift from assistive AI to autonomous AI:** Beginning with copilots allows organizations to build trust, gather feedback, and refine workflows. Once confidence is established, enterprises can transition to agentic automation.

04 **Balance short-term wins and long-term merit:** A robust AI roadmap would be based on near-term successes in areas such as customer care and quality assurance cycle time and in-depth sustainable value in areas such as network optimization and hyper-personalized marketing.

As GenAI becomes foundational to how Indian enterprises operate, compete, and grow, the role of CXOs becomes central to unlocking their full potential. The question is no longer whether to adopt AI, but how to do so in a way that delivers sustained and measurable value.

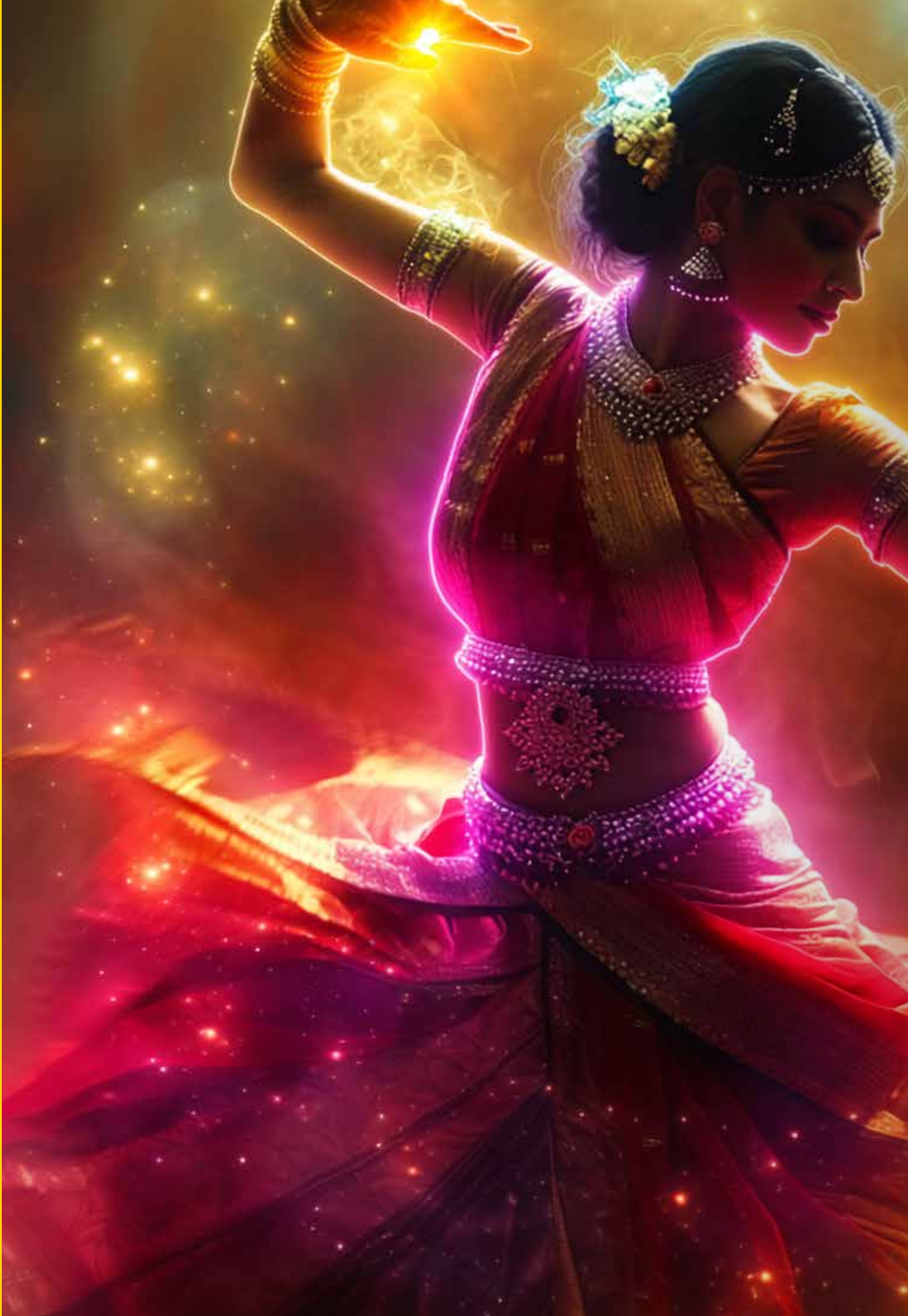
The first principle for leadership to follow would be to focus on quantifiable unit economics. AI adoption should not be measured only by the number of tools deployed or pilots launched. Instead, it must be evaluated through metrics that reflect real business impact. These include minutes saved per employee, uplift in conversion rates, reduction in fraud losses and incremental revenue per user. Such metrics help provide clarity, build executive confidence and justify continued investment.

Second, it is essential to prioritize use cases that are ready to scale. While pilot wins are encouraging, the true return on investment emerges when successful use cases are industrialized across departments and geographies. Implementation examples mentioned earlier show how scale can transform isolated success into enterprise-wide transformation.

Third, approaching the shift from assistive AI to autonomous AI requires thoughtful planning. Beginning with copilots, for example, would allow organizations to build trust, gather feedback and refine workflows. Once confidence is established, enterprises can transition to agentic automation, where AI agents act independently within clearly defined boundaries.

Finally, it is important to balance short-term wins with long-term strategic upside. Some use cases, such as customer care deflection or accelerated QA cycles, deliver returns within the same quarter. Others, including network optimization and hyper-personalized marketing, may take longer to mature but offer deeper and more sustainable value. A well-structured AI roadmap should include both types of initiatives to create momentum, resilience and strategic alignment.

Realizing ROI from GenAI and Agentic AI requires more than deploying new technology. It is a business transformation where success is based on three pillars: clarity of purpose, precision in measurement and discipline in scaling. For CXOs, the mandate is clear: lead with intent, invest with confidence, and build an enterprise that is not just AI-enabled, but AI-native.



Chapter 4

Designing an AI-first workforce at the modern enterprise

- The rise of AI "co-workers" heralds an infinite digital workforce capable of working 24/7, transforming how we think about teams and productivity.
- Moving from an AI-enabled to an AI-first workforce means fundamentally redesigning business structures—integrating AI into main workflows and reshaping roles so humans and AI work together seamlessly.
- Indian enterprises are rethinking their role architectures and skilling needs, especially in knowledge intensive industries.

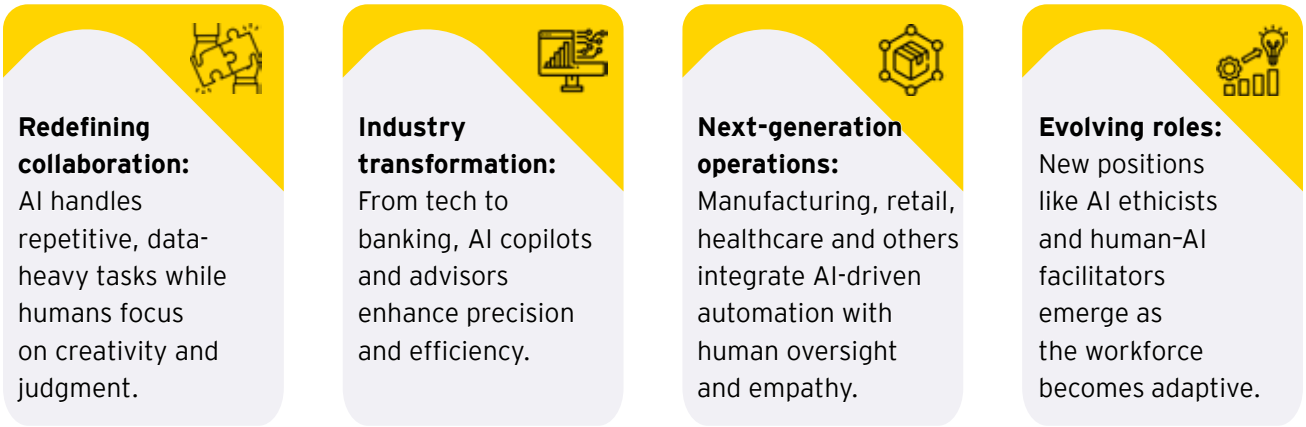
The once-in-a-generation leadership imperative

The steam engine rewired factories. Electricity redefined speed. The internet collapsed distance. Now AI stands as the next great catalyst, not just shifting how we work, but challenging what work even means. The workplace is at an inflection point—long-held assumptions about talent, delivery, and growth are being redefined.

For Indian CEOs, AI has moved from a key tech trend to a leadership stress test. The rise of Agentic AI now challenges their vision, agility, and boldness—decisions made today will decide whether they ride the AI wave or are swept aside by it.

In our previous edition, *The AIdea of India: 2025 – How much productivity can GenAI unlock in India?*, we analyzed 10,000 tasks across industries and found that 24% can be fully automated and another 42% significantly optimized through GenAI, potentially freeing up 8–10 hours per employee each week[11]. In essence, AI at scale creates an 'infinite workforce capacity,' dramatically amplifying output. No surprise, then, that CEOs have rapidly moved from exploring AI to executing AI-first strategies.

In this context, designing an AI-first workforce has become a boardroom priority, essential for staying competitive, driving non-linear growth and future-proofing the enterprise.



Vision of an AI-first workforce

From AI-enabled to AI-first: Rethinking what AI-first means

Many organizations have dabbled in AI pilots or added AI tools to assist with specific tasks. That is being AI-enabled. Being AI-first is a deeper transformation. It means re-architecting the business and workforce around AI from the ground up and not just layering up on the legacy structures. An AI-first workforce does not just use AI; it collaborates with AI by design, across every level and function. In an AI-enabled firm, the organization may deploy a few chatbots or analytics dashboards to support employees. But in an AI-first firm, AI is embedded in core workflows, and every role is reimagined to maximize human-AI synergy. Tasks and jobs are redesigned through an AI-first lens, asking, "How can AI do this, augment this, or amplify this?"

Employees become adept AI orchestrators, focusing on what humans do best (judgment, creativity, relationships) and delegating the rest to machines. For example, a marketing team in an AI-first enterprise might have AI agents generating first drafts of campaigns, analyzing customer data and autonomously optimizing ad spend, while human marketers focus on creative strategy and client engagement.

Being AI-first also demands a culture shift. Rather than fearing AI or treating it as a mere tool, employees and AI agents are regarded as collaborators. This could entail training staff to work effectively with AI (e.g., prompt engineering, AI oversight) and instilling trust in AI-generated insights alongside healthy skepticism. In sum, AI-first means rethinking the very architecture of work, so that human + AI teams are the default and every process is designed to leverage AI's strengths and human judgment together.

Cracks in the traditional workforce model: From pyramid to diamond

AI is steadily taking over transactional, rules-based tasks at the base of the pyramid. In India's tech services sector, many firms have already consolidated 20%–25% of entry-level roles through intelligent automation and AI augmentation, according to industry observers and media reports[12,13]. Instead of hiring many fresh graduates to do routine work, some leading companies are slowing fresher hiring and instead upskilling their existing talent for higher-value roles. One leading IT services firm, for instance, reduced entry-level hiring by nearly 30% and increased mid-career hiring by 20% (for new roles in AI validation, orchestration, etc.), reshaping its workforce into a more "diamond" shape – leaner at the bottom, broader in the middle[14].

This marks a radical shift. Growth is no longer tethered to hiring curves as it used to be. Revenue can grow without a linear increase in headcount, thanks to AI enabling non-linear productivity. This is also a wake-up call that the old equations (more people = more output) no longer hold.

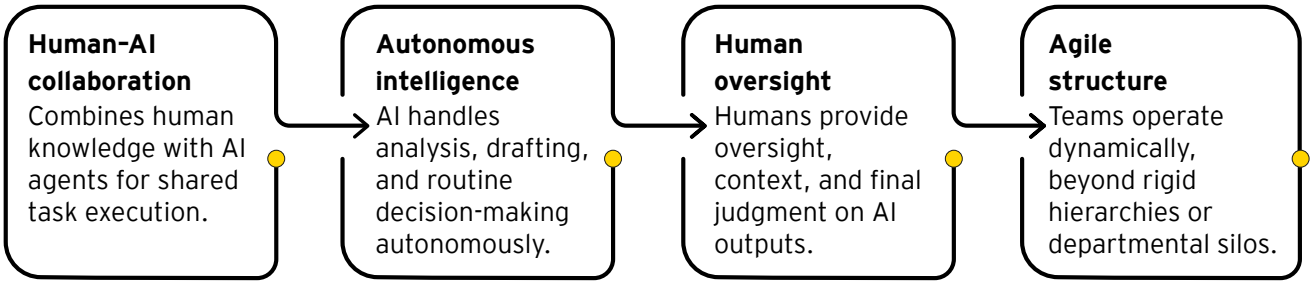
The nature of teams and management is changing as well. Work is increasingly done in 'hybrid pods' of humans + AI agents, rather than traditional departmental silos or big hierarchical teams. In these pods, AI agents might handle analysis, first drafts or routine decisions, while humans provide oversight, domain expertise and final judgment. AI agents are transforming from tools to teammates, taking on autonomy in executing tasks and even making recommendations. But this outpaces legacy management practices. How to manage performance when half the "team members" are digital bots? How to structure reporting lines or accountability in a human-AI collaborative unit?

Another crack in the old model is the collapse of the traditional career ladder. Long tenures may not guarantee advancement. Junior employees can quickly jump into high-impact roles by mastering new skills—like a self-taught AI specialist with three years leading critical projects while 20-year veterans struggle to adapt. An EY study notes that "The linear career path is broken... career progression looks more like a lattice – fluid, skill-led, nonlinear. Skill density, not seniority, becomes the new currency.[15]" Roles are getting disaggregated and reassembled; some mid-management layers are being hollowed out or repurposed as AI takes over coordination work.

The upside is a democratization of opportunity for those who upskill rapidly, but it also creates organizational challenges in mentorship and talent development. After all, if AI handles all the grunt work, who is building judgment? Those in early stages of their career risk missing out on the foundational experiences that build intuition and expertise. The speed-vs-depth dilemma is one of the key tensions business leaders must address (more on this later).

In summary, the familiar pyramid is morphing into new shapes. Some organizations may end up with a diamond structure, others might adopt a "T-shape" with a common pool of specialists and very few managers. We are even seeing the rise of blended workforce models–a mix of full-time employees, gig workers and AI-powered "digital workers," all contributing side by side. The social contract between employer and employee is being rewritten in real time. In India, early signals of this shift are evident: in some cases, fresher hiring is slowing not due to lack of business, but because entry-level work is evaporating as AI fills the gap.

How hybrid pods work

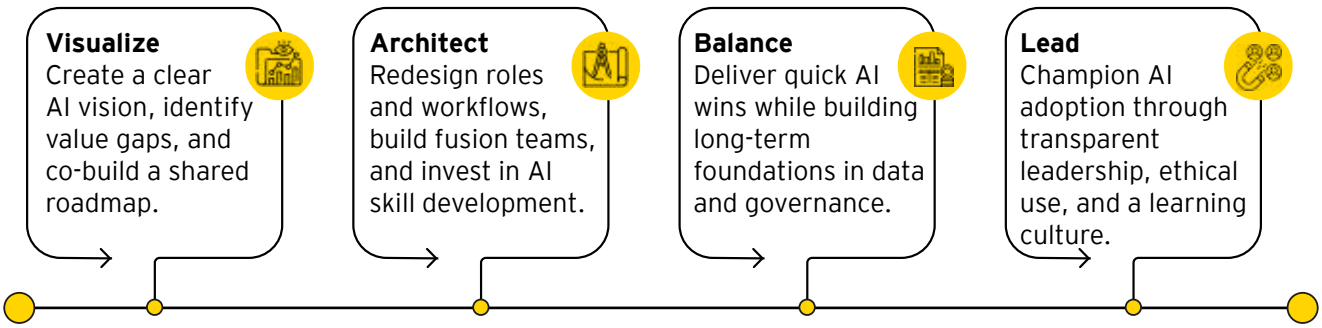


Envisioning the AI-first workforce across industries

An AI-first workforce is not about replacing humans; it is about amplifying them with a layer of AI agents. Across industries, AI takes on repetitive, data-heavy tasks while humans focus on judgment, creativity and relationships. This shift is already visible: technology firms are moving to lean human + AI pods where engineers work with AI copilots for coding and testing, and AI project managers handle tracking and risk alerts. In banking, AI tellers and advisors manage routine queries and compliance checks, freeing human bankers to deliver personalized advice and build trust.

Manufacturing and supply chains are evolving toward lights-out factories where robots run production and AI optimizes logistics, leaving humans to oversee exceptions and design processes. Retail blends smart stores and e-commerce automation with human roles as brand ambassadors. Healthcare and pharmaceuticals see AI triage agents, diagnostic copilots, surgical robotics, and AI-driven drug discovery, enabling clinicians and researchers to focus on complex decisions rather than routine tasks.

The AI-first workforce is fluid and adaptive. New roles such as AI ethicists and human-AI facilitators will rise as routine jobs fade. The future is not human or AI; it is humans and AI working together. Companies that thrive will answer three critical questions for every role: What is uniquely human? What is best for AI? And how do they collaborate? Those answers can define the winners in the AI-first era.



Model for AI-first workforce transformation

Navigating key tensions in the human–AI workplace

As CEOs lead the shift to an AI-first workforce, they face a series of balancing acts that will define success. The first is scale versus judgment. AI can automate decisions at incredible speed and scale, but wisdom cannot be automated. When algorithms drive hiring or lending unchecked, bias can spread faster than ever. Leaders must pair AI's efficiency with human discernment, backed by strong governance and override mechanisms. Scaling technology should not mean sidelining human judgment.

The second tension is speed versus depth. AI can accelerate everything, product launches, customer responses, analytics, but speed can also erode craft. If new managers lean entirely on AI guidance, they risk becoming faster while losing depth. CEOs need to design roles that preserve deep expertise and creative thinking. That means rotating employees through hands-on experiences, encouraging reflection, and keeping some "slow" processes that build mastery. AI could handle the drudgery at speed, while humans focus on high-value deep work.

Finally, efficiency versus equity and innovation versus control. AI boosts productivity, but job displacement and digital divides can damage trust, especially in young workforces. Visionary leaders invest in reskilling and share AI gains with employees, not just shareholders. At the same time, AI unlocks bold innovation but without guardrails, autonomy can spiral into risk.

The answer is clear: governance, human-in-the-loop checkpoints, and a culture where AI outputs are trusted but verified. The goal is not to choose between these tensions, but to navigate them with purpose, building workplaces where humans and AI elevate each other.

Scale vs judgment

Pair AI's automation speed with human discernment and governance to avoid bias.

Speed vs depth

Maintain craftsmanship and deep expertise while allowing AI to handle routine work.

Efficiency vs equity

Balance productivity gains with fair opportunities through reskilling and shared AI benefits.

Innovation vs control

Transformative creativity within proper limits to mitigate risks.

A four-point model for AI-first workforce transformation

Designing and leading an AI-first workforce is a complex journey. It requires structured transformation model. Here is a practical four-point framework for CEOs:

Visualize

Start with a compelling vision. Explain why AI is central to growth and show how it will reshape the operating model. Identify where value is trapped today and how AI can unlock it. Engage leaders and employees in co-creating the roadmap so it feels shared. Set near-term wins and long-term goals, and communicate relentlessly so AI becomes core strategy, not a side project.

Architect

Turn vision into design. Redefine roles and workflows to decide what AI automates, what humans lead with AI support, and what new tasks emerge. Update job descriptions, create hybrid roles, and form fusion teams that blend business and tech expertise. Invest in reskilling and hiring for critical AI skills. Architecting change deliberately ensures alignment with business goals.

Balance

Run at two speeds. Capture quick wins like AI chatbots and supply chain optimization while building long-term foundations such as data platforms and governance. Celebrate early successes but protect strategic investments from short-term pressures. Balance buying tools with building internal expertise. The goal is simple: win today while preparing f or tomorrow.

Lead

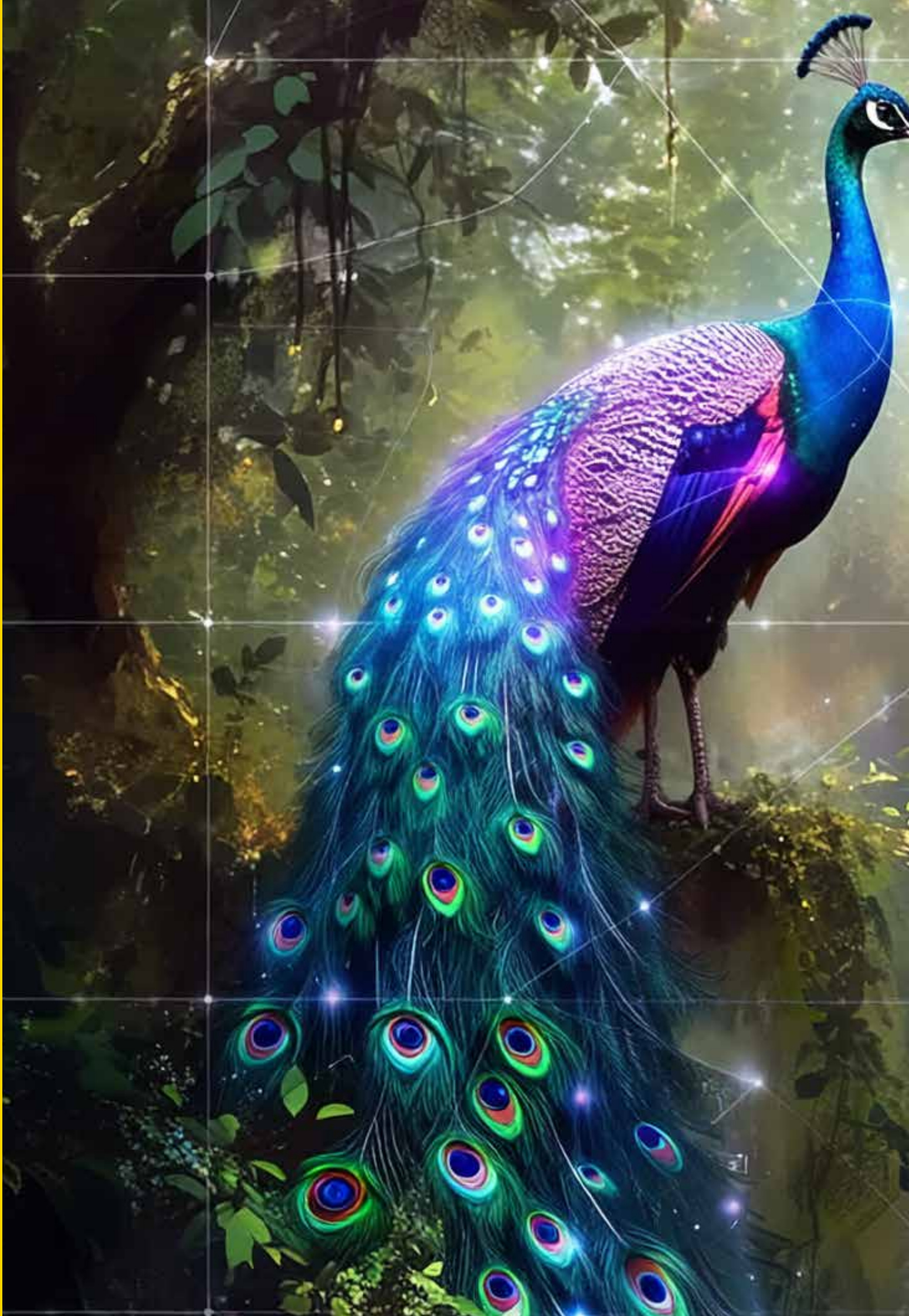
Culture and leadership make or break this journey. CEOs must lead from the front, using AI visibly and fostering a learning mindset. Address fears around job security through transparency and inclusion. Embed ethics and governance with clear boundaries on automation and fairness. Model curiosity and boldness, normalize experimentation, and keep humans at the center. Winning with AI is about trust as much as technology.

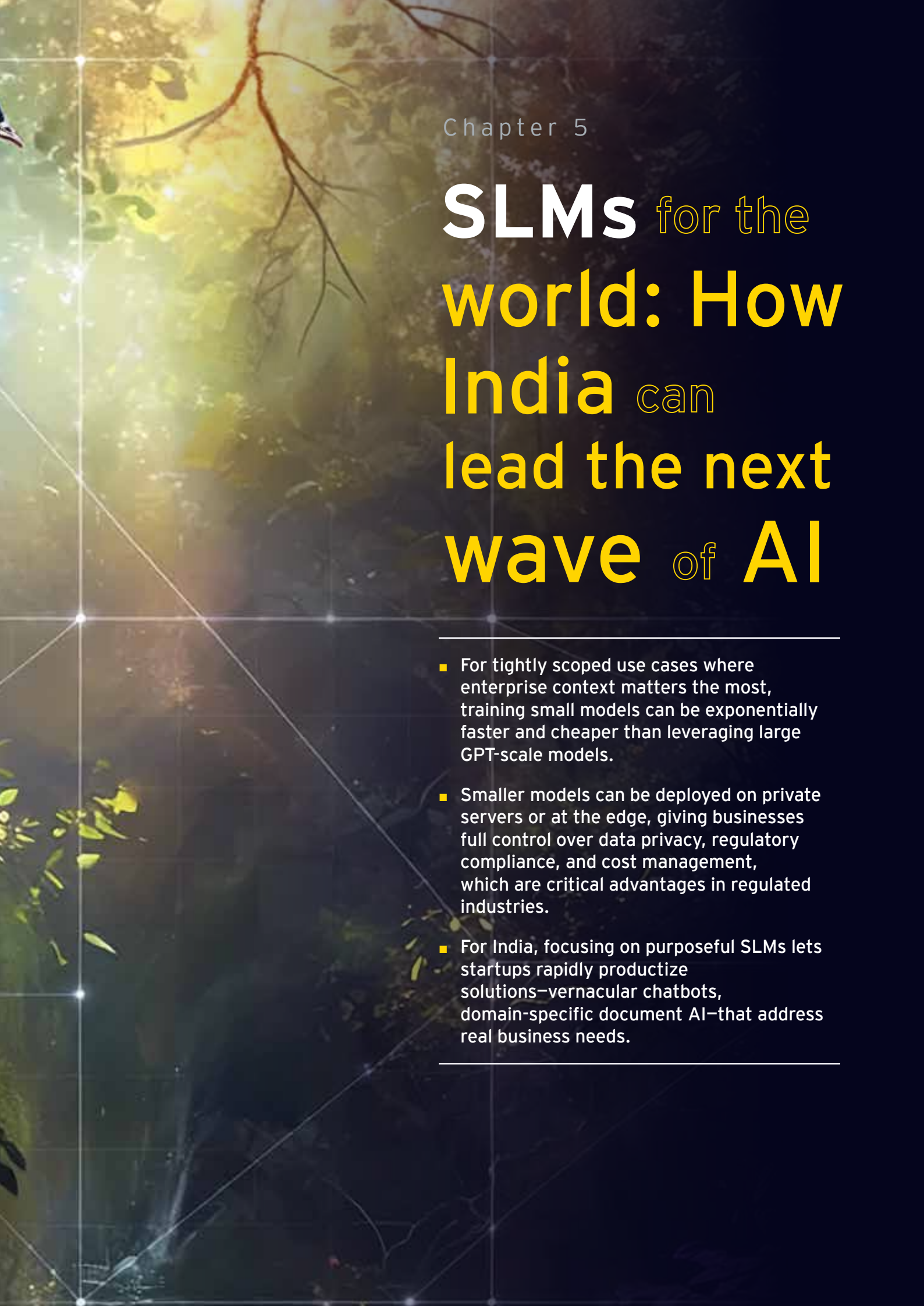
Shaping the future, boldly and responsibly

Designing an AI-first workforce is both an opportunity and a responsibility. It promises a scalable, hyper-productive organization but only if leaders value human potential as much as technological power.

As a CEO, ask: Will AI make your organization faster, but narrower or smarter and more human? The answer lies in how one leads this transformation. Done right, AI will boost creativity, inclusivity and value.

The AI revolution is moving fast. Indian enterprises can leapfrog into global benchmarks in human-AI collaboration by prioritizing an AI-first workforce. This is an opportunity to re-architect work itself. Leading with clarity, courage and compassion can deliver not only superior outcomes but also shape a future of work that is creative, inclusive, and deeply human by design.





Chapter 5

SLMs for the world: How India can lead the next wave of AI

- For tightly scoped use cases where enterprise context matters the most, training small models can be exponentially faster and cheaper than leveraging large GPT-scale models.
- Smaller models can be deployed on private servers or at the edge, giving businesses full control over data privacy, regulatory compliance, and cost management, which are critical advantages in regulated industries.
- For India, focusing on purposeful SLMs lets startups rapidly productize solutions—vernacular chatbots, domain-specific document AI—that address real business needs.

Small Language Models (SLMs) are fast becoming the quiet engine of AI adoption in Indian enterprises. While the global spotlight is mostly on large models like GPT-4, in the Indian scenario of diversity, complexity and often constrained infrastructure, going smaller is the pragmatic bet that enterprises are taking.

SLMs are compact models with millions to a few billion parameters and less resource-hungry than LLMs. They are proving their capabilities in delivering real-world impact, especially when trained or fine-tuned for Indian languages, local contexts and domain-specific tasks. Several startups across India are using SLMs to bridge the digital divide, driving innovation that is inclusive, scalable and cost-effective.

Rise of SLMs in India

With value and not size as the real differentiator, SLMs are task-focused, highly efficient and capable of delivering strong performance across a wide range of enterprise needs.

Globally, the SLM market expected to grow significantly. In India, the shift is pronounced as local enterprises and tech leaders are leaning into the idea that right-sized AI is more accessible and often more effective, especially when built around specific use cases and constraints. Large models may lack the required precision, contextual understanding and deployability.

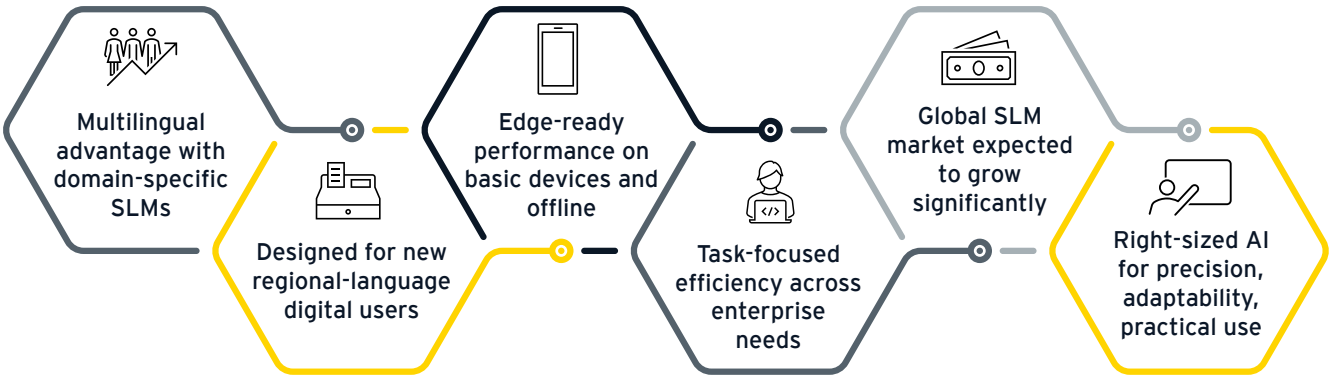
The primary reasons for increasing preference for SLMs in India are:

01 In a multilingual environment, focused models tailored to specific domains or regions can be more effective. For instance, an insurance chatbot in Tamil or an agriculture advise in Marathi.

02 India is a mobile-first nation, with many users relying on basic smartphones and uneven 3G or 4G connectivity. In many areas, power and internet supply can be a challenge. SLMs can work in such conditions as they run on edge devices, can work offline or with limited bandwidth, consume less power, cost less to host and keep data local, addressing both performance and privacy concerns.

The real opportunity lies in reaching the next users who are coming online in regional languages, for the first time. SLMs are uniquely positioned to serve this audience, bringing conversational AI, translation tools, voice interfaces and contextual understanding to users who were not a part of the English-first AI wave.

SLMs gain ground in India



The struggles

LLMs are powering the global AI race, but building a state-of-the-art alternative in India is a demanding task, especially due to multiple structural challenges.

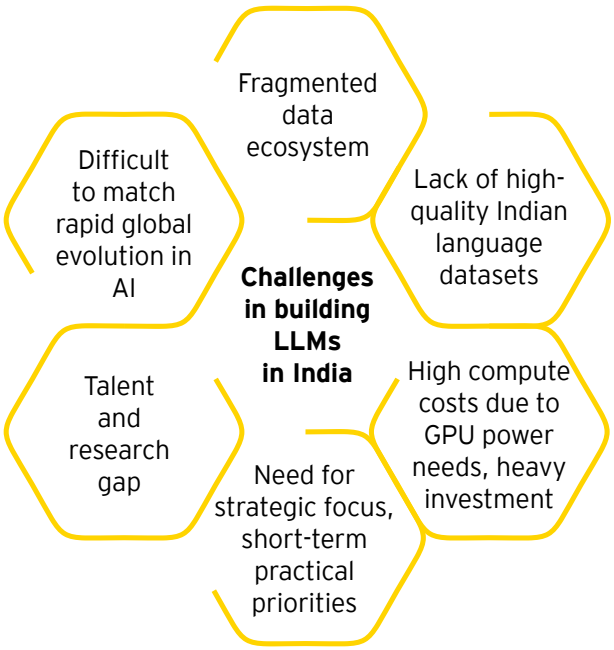
The data ecosystem in India is largely fragmented, which hinders training a competitive LLM as they require massive volumes of high-quality, diverse, balanced and domain-rich datasets. This includes scientific texts, medical records, legal documents, codebases and everyday conversational language across languages and contexts. While such digital corpus is widely available in English, digital content is limited and fragmented in Indian languages, especially in technical and specialized domains.

Another challenge is the compute and capital required. Training a GPT-class model demands tens of thousands of GPUs running for weeks, supported by high-bandwidth data and power systems, which would cost millions of dollars per cycle. Such an investment would be prohibitive and hard to justify even for large enterprises. The government's IndiaAI initiative has introduced incentives and is ramping up access to compute infrastructure, marking the early phases of the buildout.

There is also a significant talent and R&D gap. While India has skilled AI engineers, a deep research ecosystem is needed to push boundaries in foundational model research in architecture, alignment and safety. Work in that direction has primarily come from US and European labs with substantial long-term investment and academic collaboration. The focus in India has been on application-layer innovation. While that is shifting with efforts from IITs and AI startups, achieving higher levels of maturity may require sustained effort and ongoing development.

Finally, the global AI landscape is changing fast. Catching up would require not just building a large model but continuous evolution—improving performance, safety, efficiency and multimodal capabilities in lockstep with the world.

While India also focuses on building large models, a short-term strategic approach is needed for near-term use and results.



The way forward

SLMs can be tailored to the requirements of organizations that cater to a wide range of consumers. Smaller, purpose-built models can deliver more accuracy, trust and relevance for both enterprises and consumers.

Training small models is exponentially faster and cheaper, allowing access to startups, universities and corporate innovation teams. Using open-source foundations like LLaMA, Mistral or Gemma, Indian innovators can fine-tune, distill and deploy these models for specific sectors in weeks. Enterprises gain the added advantage of control as smaller models can be deployed on private servers or at the edge, giving businesses confidence on data privacy, compliance and predictable cost management, which is critical in regulated industries.

By focusing on small, purposeful models, startups can rapidly productize solutions. Indian startups are proving that sovereign, auditable and cost-efficient SLMs can serve millions through multilingual, voice-first, enterprise-ready systems. Small models, built for India's realities, are already becoming models for the world, as two startup founders share (edited excerpts).

Interview 1

“Multi-modality in SLM use cases is growing rapidly”

What led you to build SLMs?

A multilingual country like India needs models that are localized, inclusive and context aware. SLMs fit this need; they are lightweight, cost-effective, domain-adaptable and support on-device or edge use, enabling data sovereignty, low latency and accessibility. We do not compete against frontier LLMs but build sovereign, human-centered AI systems to address needs in healthcare, education, governance, defense, etc.

How do you approach data curation and training with limited Indian-language resources? What are the challenges of trying to build foundational models in India?

We use auditable pipelines to ingest, clean, anonymize and version data with full consent and compliance. For low-resource Indian languages, we combine licensed, public and local datasets with human annotation, transfer learning, synthetic data and active learning. In addition, language-specific quality checks ensure privacy, accuracy and performance.

To build foundational models, there are limits like compute costs, clean multilingual data and labeled data for many local languages. Regulations and data consent rules are other challenges.

Ankush Sabharwal,
Founder, CoRover.ai,
a conversational AI company

How do you ensure explainability, auditability and governance?

Full traceability is attained through model cards, dataset records, versioning and tamper-evident audit trails. In regulated sectors, we add bias tests, human review and customer-controlled deployments, with compliance reports and audits that align to local rules.

Do you think enterprises can bet on SLMs when LLMs are improving? Will SLMs replace LLMs+RAG (Retrieval-Augmented Generation), or will they coexist?

SLMs provide enterprises control, compliance and cost-efficiency for critical workloads, while LLMs offer breadth and general knowledge. A hybrid approach—SLMs for sensitive, repeatable tasks and LLMs for broad reasoning—delivers optimal performance, scalability and flexibility.

SLMs handle core business logic, safety checks and fast responses while LLM+RAG handles broad search, long-tail queries and contextual retrieval. A hybrid setup is the practical path forward.

In your understanding, among industries, which would adopt SLMs first and which later?

Industries such as BFSI, healthcare, defense, governance, travel and tourism and customer support are likely to lead, followed by knowledge-heavy and compliance-critical sectors, such as law, R&D and education. Industries with legacy infrastructure, low digital maturity, or fragmented use-cases, like small-scale manufacturing or traditional retail, may adopt SLMs later.

How multi-modal are the SLM use cases today?

Multi-modality in SLM use cases is growing rapidly, with text, documents and images already common in enterprise deployments. Speech is now central in AI telephony, customer support and accessibility-driven use cases.

Enterprises are increasingly combining text, voice and OCR-based document inputs into unified workflows, enabling smoother customer experiences and greater reach. The immediate business impact is from voice-enabled multi-modal AI, especially for BFSI, healthcare and public services.

What are some of the little-known facts about the Indian AI ecosystem?

One is the growing emphasis on homegrown foundational models and LLMs tailored to local needs. India is actively building its own AI infrastructure optimized for multilingual support, code mixing (such as Hinglish, Benglish) and low resource environments. India is also focusing on AI models that are culturally and linguistically aligned with its population, which is both globally competitive and locally impactful.

Three factors position India for AI leadership: technical excellence meeting practical problem-solving; rich and diverse data; and rapid user adoption. These strengths create an ecosystem where AI scales rapidly across governance, healthcare, education, and enterprise—delivering measurable impact.

Interview 2

“Enterprises benefit from practical, efficient models rather than the largest”

What led you to build SLMs over frontier LLMs?

India’s cultural diversity and regulatory environment make it a demanding but rewarding AI market, with more than 22 languages, widespread code-switching and strict data-residency rules. Frontier APIs do not meet latency, cost or compliance needs. Smallest.ai’s SLM, Electron, addresses this: 53 milliseconds (ms) Time-to-First-Byte (vs 200ms+ on global APIs), fully on-premise deployment and 20%–30% better hallucination control. Moreover, for massive customer call volumes, SLM economics are far more sustainable than frontier LLM pricing.

How do you approach data curation and training with limited Indian-language resources?

We primarily use a three-layered approach:

Synthetic data generation: Multi-agent synthetic conversations capture Hinglish, interruptions, OTP verification and numeric-heavy exchanges. Every dataset is stress-tested automatically and manually reviewed for BFSI-critical scenarios.

Enterprise grounding with RAG: Sensitive data stays within the client perimeter. Models retrieve only relevant, pre-approved enterprise documents, producing responses linked to source policies for auditability.

Speech enrichment: Our in-house Lightening text to speech (TTS) model enables prosody, pauses and tone at <US$0.10 per 10,000 characters, running at <100ms latency.

Sudarshan Kamath,
Founder, Smallest.ai,
Smallest.ai, a voice AI startup

Why do you think enterprises bet on SLMs when LLMs are improving?

Enterprises benefit from practical, efficient models rather than the largest. Frontier LLMs are universal but not tailored for high-volume, regulated workflows. SLMs are lean, predictable and deployable inside firewalls. In voice-led industries, Lightning TTS adapts tone and context, improving trust in automation. SLMs also bridge AI experiments to enterprise-wide adoption, while frontier LLMs remain useful as teachers or for rare queries.

Are SLMs the future, or will they work alongside LLMs + RAG?

SLMs and LLMs complement each other. Electron is the operational core for daily conversations, while frontier LLMs act as teachers, generating synthetic data and handling complex queries. A RAG layer grounds responses, and Waves/Atoms provides structured workflows, API orchestration, and self-analysis. The result: agents that are fast, accurate, auditable and continuously learning.

How do you ensure explainability, auditability and governance?

Trust is built-in through:

- Traceability of datasets, model versions and interactions
- RAG grounding linking responses to enterprise policies
- Guardrailed generation enforcing schema-bound outputs
- Deployment flexibility: on-prem/VPC (Virtual Private Cloud), audit dashboards, retention policies aligned with DPDP Act, RBI and CERT-In
- Human-in-the-loop escalation for low-confidence cases.

Which industries will adopt SLMs first, and which will lag?

Early adopters include BFSI, insurance, telco, retail, logistics—millions of interactions provide immediate ROI. Latency drops from ~2s to <200ms, TTS costs cut by 90%. Later movers could be healthcare, public sector and legal, due to fragmented IT and compliance.

What are the challenges of trying to build foundational models in India?

Scarce Indic corpora, code-switching, high GPU costs and privacy restrictions are some of the challenges. But these same challenges create opportunities where India can lead. For example, we are also creating domain-specific models that already outperform generic LLMs in BFSI and telco.

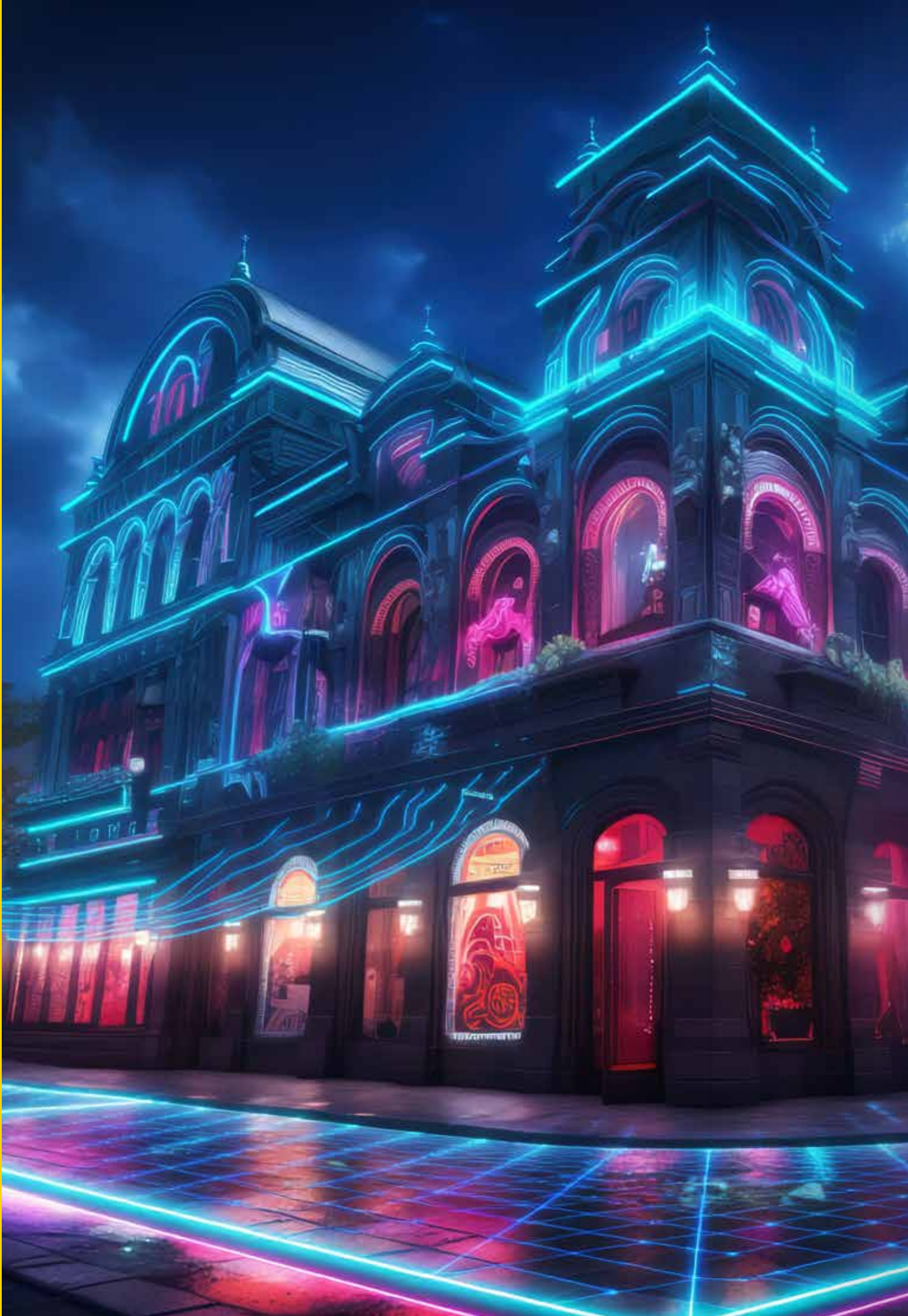
The path forward is pragmatic. In the short term, fine-tune and distill open-weight bases; in the medium term, build from scratch speech and multimodal models powered by synthetic data; and in the long term, contribute domain-grounded, cost-efficient foundational models made on non-Nvidia chipsets that promise 10x lower inference costs.

How multi-modal are today's SLM use cases?

Today's multimodality is voice-first. Lightning V3 adds pauses, breaths, laughter and emotional tones; Electron handles reasoning. Hydra speech-to-speech processes interruptions and overlaps. Enterprises also integrate SLMs with structured data, with broader multimodality (image-text-audio) emerging.

What is a little-known fact about India's AI ecosystem?

Synthetic data is India's "secret weapon." Persona-rich, multilingual, noise-augmented synthetic datasets improve performance while staying compliant, helping India leapfrog in enterprise AI adoption.



Chapter 6

Responsible AI 2.0: From policies to continuous, auditable assurance

- RAI 2.0 shifts from a "trust us" to a "show us" approach, requiring organizations to demonstrate ethical AI adherence through ongoing assurance mechanisms, including detailed record-keeping, continuous monitoring, and third-party validation.
- Regulatory bodies worldwide are transitioning from providing guidelines to enforcing requirements for AI systems, such as risk assessments and transparency measures.
- RBI's FREE-AI framework emphasizes the need for board-approved AI policies across financial institutions, highlighting the importance of strong governance, risk management and consumer protection.

Earlier approaches to Responsible AI (or RAI 1.0) centered on high-level principles and ethical guidelines. Organizations and governments worldwide published Responsible AI principles to guide the design, development, deployment and use of AI. These principles aimed to build trust in AI solutions that can empower organizations and their stakeholders. RA1.0 also provided foundational frameworks to embed ethical considerations into AI systems and workflows, helping to mitigate risks and negative outcomes associated with AI use, while maximizing its positive impact.

The next phase, RAI 2.0, is about evidence. Stakeholders now expect systems to demonstrate that those principles hold up in practice, continuously, not just at launch. The position has changed from "trust us" to "show us."

Such a shift is driven by three primary forces:

1. Technology is moving fast, especially with GenAI and Agentic AI. Systems can now write, decide and act with little human help. That raises critical new questions about safety, ethics and accountability.
2. Regulators are moving from guidance to requirements. More jurisdictions now mandate risk assessments, documentation, audits and transparency.
3. The imperative to cultivate and maintain public trust has become paramount. Claims about "responsibility" must be substantiated by concrete evidence including detailed logs, measurable metrics and verifiable proof of adherence.

RAI 1.0 versus RAI 2.0

Aspect	RAI 1.0	RAI 2.0
Focus	High-level principles and ethical guidelines	Evidence and continuous assurance
Approach	"Trust us"	"Show us"
Governance	Static policy documents	Run-time governance
Stakeholder Expectation	Trust based on principles	Demonstrable proof of adherence
Technology	Basic AI systems (Traditional AI)	Advanced AI systems (GenAI, Agentic AI)
Regulatory Environment	Guidance	Mandates (risk assessments, audits, transparency)
Public Trust	Claims of responsibility	Concrete evidence (logs, metrics, proof)
Risk Management	Foundational framework	Continuous monitoring and impact assessment

Emerging risks: Why Agentic AI needs new guardrails

The rise of Agentic AI and its ability to execute end-to-end tasks with minimal human intervention also amplifies risks. And that is why RAI 2.0 is necessary. More autonomy also means that any errors that occur can spread farther and faster. Many things could go wrong when an Agent has the power to both generate and then run code.

Problems range from unintended actions and security exploits (e.g., an agent executing harmful code) to ethical lapses (e.g., pursuing a goal at odds with human values).

The response is shifting from static policy to run-time governance[16]. Practical guardrails include sandboxed execution, strict tool permissions, adversarial testing (red teaming) and "guardian" layers that watch actions in real time. Organizations are also updating playbooks so it is clear who is accountable, when humans must intervene and how to shut a system down safely if required. This way of thinking about continuous assurance is a big change from the static policy documents of RAI 1.0 [17].

Global regulatory trends: From guidelines to requirements

Across the world, regulators are turning principles into verifiable obligations. Under the EU AI Act, high-risk AI systems will need to have written risk assessments, disclosures of transparency, and in many cases, third-party conformity assessments. The Act also has rules that deal with AI's effect on the environment. For example, it requires openness about energy use. Companies that make General-Purpose AI models (GPAI), like LLMs, are required to keep technical records of the model's energy usage when it is being trained and while running. Model builders must keep track of these metrics and update them[18].

This is a clear move from broad ideas (like "AI should be sustainable") to tangible proof of following the rules (like "show us your AI's energy footprint").

Responsible AI 2.0 builds on the same "prove it with data" philosophy seen in new EU digital regulations: in addition to intent or policy statements, ethical AI must be defined by measurable, transparent and verifiable. Organizations are expected to demonstrate accountability through data-driven evidence, embedding quantifiable checks into every stage of AI design, deployment and governance.

Many countries around the world are changing their AI governance systems in this way. Stanford University's AI Index study[19] shows that the number of laws and rules about AI has grown a lot in the last few years. This rise in hard law shows that governments are moving past plans and making AI oversight a permanent part of their systems.

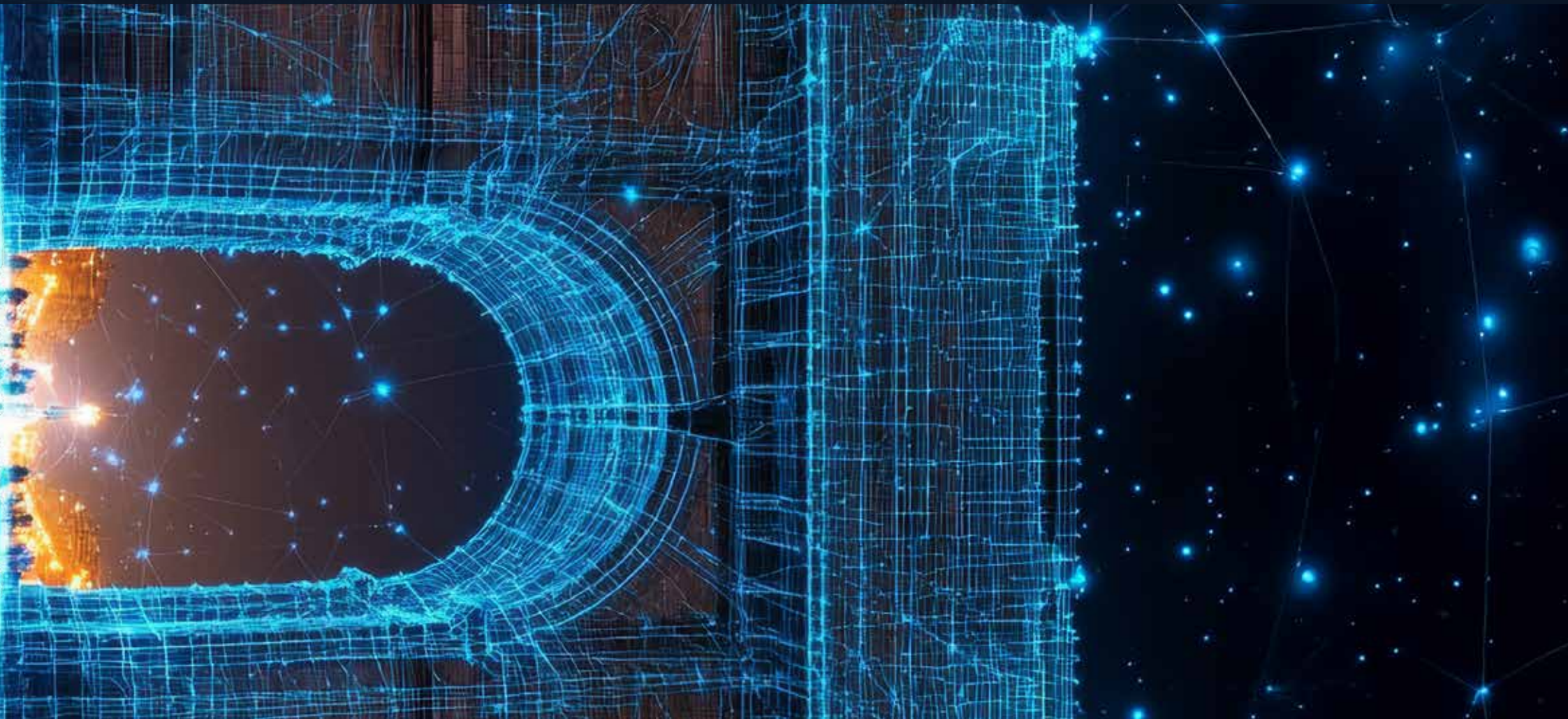
There are different approaches. The EU has a general AI Act, China has rules for specific sectors, and the US, the UK and Singapore started out by favoring guidance-based or voluntary frameworks. But a common theme in all these efforts is the focus on transparency, responsibility and consumer protection. Even in countries without specific AI laws, organizations must audit AI systems for compliance with existing regulations (data protection, safety, anti-discrimination) as current rules are being applied to AI outcomes.

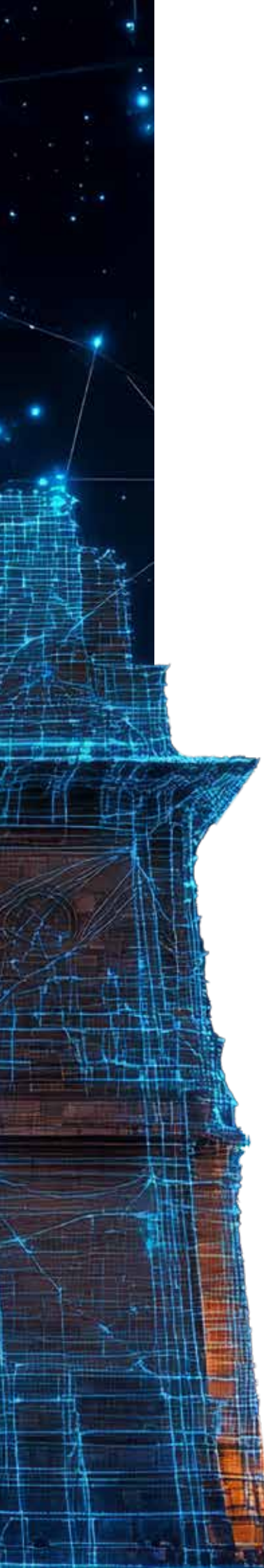
New standards and frameworks are also pushing for more regulation. The ISO/IEC 42001 management system standard for AI and the US NIST AI Risk Management Framework (RMF) provide systematic models for establishing and evaluating AI governance. They emphasize impact assessments, data lineage tracking, bias testing and human oversight[20].

At the same time, there is growing momentum for independent AI audits. Policymakers in the US, Canada and other regions are proposing mandatory third-party reviews for high-risk systems. While still emerging, AI assurance is poised to become a standard practice much like financial or cybersecurity audits. Forward-looking organizations are already investing in AI assurance tools to evaluate bias, robustness, privacy and explainability, strengthening stakeholder trust and compliance readiness.

Responsible AI in India - RBI's FREE-AI

7 sutras





Moving from vision to implementation

India's approach to Responsible AI is also an example of going from broad ideas to specific plans. The Reserve Bank of India's Framework for Responsible and Ethical Enablement of AI (FREE-AI)[21], which came out in August 2025, was a big step forward for the use of AI in finance. It defines goals and a plan to achieve them.

The framework is made up of seven main ideas, which are known as the "Seven Sutras." There are also 26 specific suggestions for how to use the framework, organized into six strategic pillars.

A set of foundational tenets by RBI that will serve as the guiding principles for the development, deployment and governance of AI in the financial sector.

01 **Trust is the foundation**
Trust is non-negotiable and should remain uncompromised

02 **People first**
AI should augment human decision-making but defer to human judgement and citizen interest

03 **Innovation over restraint**
Foster responsible innovation with purpose

04 **Fairness and equity**
AI outcomes should be fair and non-discriminatory

05 **Accountability**
Accountability rests with the entities deploying AI

06 **Understandable by design**
Ensure explainability for trust

07 **Safety, resilience and sustainability**
AI system should be secure, resilient and energy efficient

The RBI's FREE-AI report talks about principles and suggests real ways to govern AI. It calls for an "AI sandbox" where banks and fintechs can test AI solutions under regulatory oversight. Smaller banks are encouraged to develop shared India specific AI models to reduce dependency on foreign systems. The framework also mandates that all regulated financial institutions have board-approved AI policies covering governance, risk and consumer protection. Companies should also include information about AI in their annual reports and on their websites to be more open. Financial institutions are being told that using AI requires more than just being excited about technology. They also need "board-level attention, strong governance, and investment in institutional capacity."

Responsible AI at the IndiaAI Mission

The IndiaAI Mission brings together capability and trust at the national level. Through AIKosh (public database of datasets and baseline models), EY and IndiaAI Mission are piloting fairness assessments and Fairness Passports that move RAI 2.0 from principles to proof. Each passport is an evidence pack: data lineage and consent logs, representation and label-quality stats, bias testing across protected attributes, robustness and drift checks, and use-case guardrails with human review paths. Providers attest to what the model was built to do, the populations it serves, the risks it avoids, and how it performs under stress. Evidence is versioned, reproducible and mapped to Indian and global requirements so regulators and buyers can rely on it. This creates assurance as there are measurable thresholds, tests that are done on a regular basis, and audit trails that make claims easy to prove false. It also makes things harder for vendors as they have to use a model or dataset that they can explain, monitor and fix. The emphasis is on trust built through verifiable artefacts.

The IndiaAI Mission makes large investments in AI capabilities and governance infrastructure, such as data governance, assessment tools, standards and oversight bodies. It also shows an inclusive approach by making access to AI more democratic. (Read more about IndiaAI Mission in the Sovereign AI chapter.)

India is helping shape international AI governance talks, for example, by co-chairing global summit initiatives and hosting the India–AI Impact Summit (New Delhi, February 2026)[22]. World leaders and experts are expected to gather and share best practices as well as produce a Leaders' Declaration on concrete AI governance actions, reinforcing the Responsible AI 2.0 narrative on the global stage.

Synthetic data: A new opportunity for regulated industries

Between India's Digital Personal Data Protection Act, 2023, RBI's new AI governance playbook, oversight by the Securities and Exchange Board of India (SEBI) on algo models and sectoral norms like Health Insurance Portability and Accountability Act (HIPAA), most highly regulated industries like finance, healthcare, telecom, phrama, etc., are in a paradox: they have valuable data but compliance and privacy considerations have to be balanced with innovation. Synthetic data offers a strategic way forward. It enables compliant innovation by mimicking real data patterns, making experimentation and scaling easier.

Synthetic data makes model testing safer, simulate rare or risky events, and allows AI teams to build and iterate. For example, to test a fraud detection system on an edge-case scam, the scam and other parameters can be synthesized. A credit model can be stress tested across geographies and demographics by generating a synthetic population and test.

Limited access to real data

Under the DPDP Act, explicit consent is mandatory for using sensitive personal data. A bank with, say, 20 years of transaction logs has limited options of using it in an AI model at present without consent from all related customers. At the same time, RBI and SEBI are pushing AI governance to the same level of scrutiny as financial risk, demanding rigorous validation and approvals before being made live. In healthcare and pharma, HIPAA-like rules apply to patient records.

In most enterprises, the legal department acts as the final gatekeeper. As AI regulations are evolving and there is fear of privacy breaches and heavy penalties, many organizations are taking a conservative stance with AI pilots, not from lack of vision, but due to the risk of regulations and harm to reputation and trust.

Synthetic data offers a breakthrough by offering datasets that enterprises can use.

The game changer

As DPDP compliance tightens and RBI sandboxes nudge toward privacy-first innovation, synthetic data is becoming the connective layer between regulation and experimentation. It is how a telecom firm can work with a startup on network optimization without exposing user logs. Or how a hospital can collaborate on diagnostics research without risking patient privacy.

This is more than a cost or risk conversation instead of slowing down innovation to stay safe. Enterprises can now innovate safely by design. Synthetic data offers a way to unlock the value in systems while protecting the people behind the data. For sectors where privacy is paramount and pressure to modernize is mounting, it is no longer a question of if but how fast can be the adoption.

If India wants to build trustworthy, inclusive, and regulation-aligned AI, especially in high-stakes sectors, synthetic data could be the most underrated unlock. It can be the bridge between ambition and action and letting organizations innovate boldly without breaking rules.

The road ahead: Toward continuous auditable assurance

With RAI 2.0, businesses and governments need to keep in mind that governance is an ongoing process. Policies and principles are still important as a foundation, but the focus is on applicable assurance programs that evolve with AI, bringing real-world implications for stakeholders in India and beyond. Steps that enterprises can adopt at present include:

Making controls auditable: Make AI systems auditable by keeping detailed records (data sources, model parameters, decision logs), versioning models and datasets, and be transparent with model cards and datasheets. Businesses can maintain AI model registries and inventories to keep track of algorithms as there are new rules and standards that support it. External auditors or regulators may ask for these records in the future.

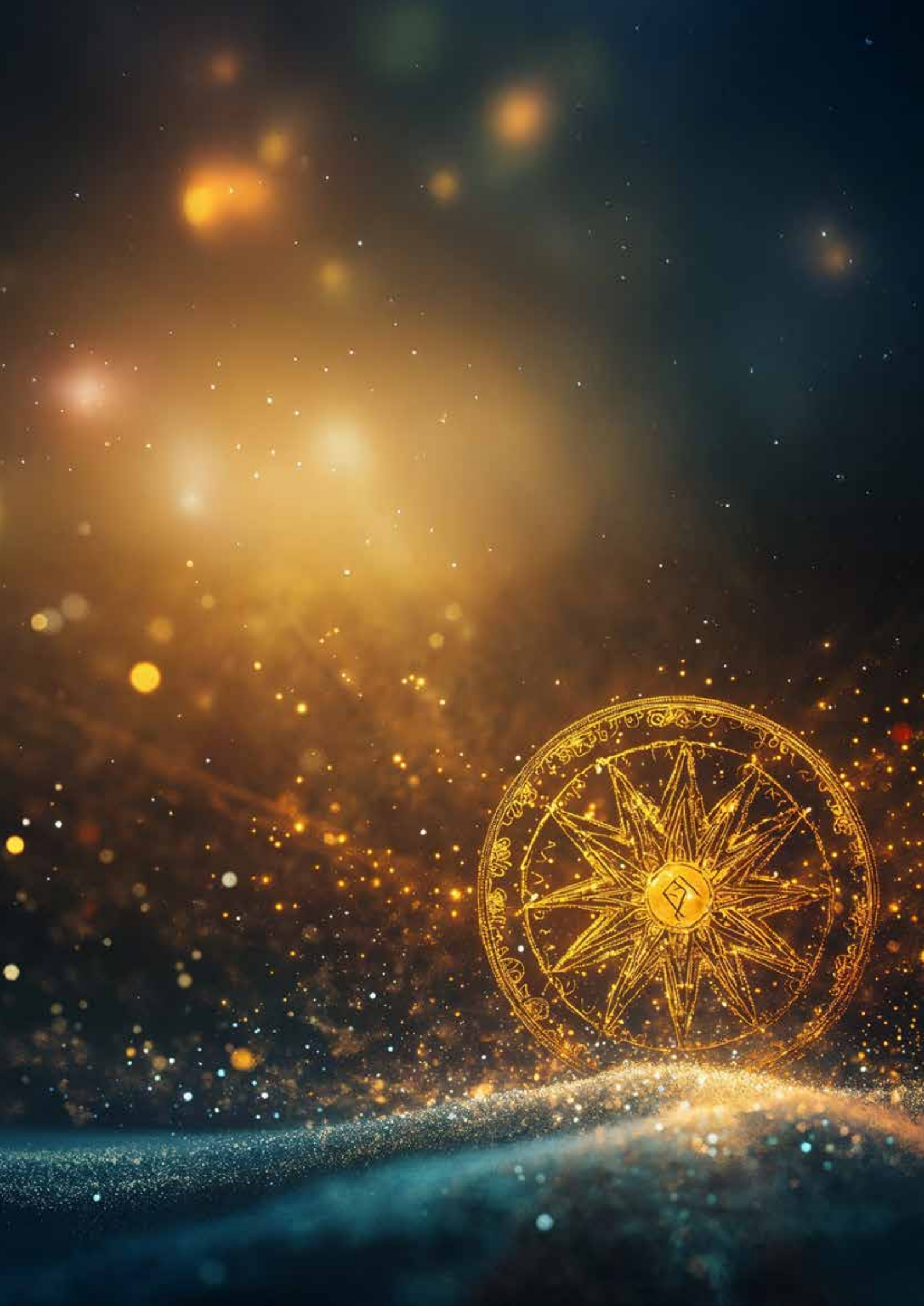
Continuous monitoring and impact assessment: AI systems should be constantly checked for performance and fairness. Risk impact assessments should be conducted every time the system is updated or there is a change in the way it works. Organizations should include AI models in their risk control cycle and observe over time for drift, bias or mistakes and retrain or change course accordingly. Automated bias detectors and fairness metrics dashboards can help in this process. Setting up model governance committees can also help in conducting regular checks.

Third-party validation and certification: Soon, we may see the rise of specialized 'AI audit firms' or certification bodies—much like ISO or financial auditors—to assess fairness, safety and privacy to gain clients' trust. India is already exploring "fairness certificates" and standards for measuring algorithmic fairness. Companies can conduct pilot audits or benchmark exercises which are like having external specialists check their AI for bias or security. This is a way to prepare for the future when such audits are required. Organizations would need to collaborate to develop Indian standards that are in line with global ones (ISO, OECD AI Principles, etc.).

Managing 'Frontier AI': The AI of the future, generalized AI models or tightly linked agent systems, will need new ways to be managed. 'Responsible by design' should also apply to things like making AI agents work together, making fake data and working with AI. Policymakers and researchers must consider these issues. Questions like how to ensure self-learning AI acts rightly—and who is accountable if it does not—are likely to shape future assurance methods, from AI simulators to new accountability laws.

RAI 2.0 shows how the AI ecosystem has matured. In this phase, principles and policies are operationalized and put into action. Organizations are expected to prove that their AI is worthy of trust, through data-driven evidence and third-party scrutiny. India's example provides insights into this global shift: from the RBI's sectoral guidelines that marry ethics with practical steps, to the national IndiaAI Mission that invests equally in AI growth and AI guardrails. The focus, however, remains sector-agnostic as the frameworks being put in place are meant to apply across healthcare, finance, agriculture, governance and beyond. RAI 2.0 is fundamentally about continuous auditable assurance.

In the coming years, organizations and countries leading in AI will likely be those that manage to build robust assurance mechanisms without stifling innovation. We are entering a time of "trust but verify" for AI. RAI 2.0 makes AI that is not only powerful but also worthy of the public's trust by making ethics a part of everyday life.



Chapter 7

Sovereign AI: Driving national empowerment and global leadership

- Sovereign AI promises substantial economic gains, driving productivity, innovation and competitiveness.
- Major countries including India and China are building Sovereign AI capabilities to reduce foreign tech dependence, secure digital infrastructure and strengthen national security and economic resilience.
- Effective Sovereign AI requires domestic infrastructure, data sovereignty, domestic talent, strategic legal frameworks and strong cybersecurity—all aligned with national interests and values.

Across the globe, AI technologies have become central to geopolitical competition and states increasingly view them as instruments for economic power, national security and technological leadership.

According to NITI Aayog, AI could add an estimated US$1.7 trillion to India's economy by 2035[23], underscoring its transformative impact on productivity, innovation and competitiveness. This has created an urgency for robust Sovereign AI. Countries leading in AI are expected to dominate global markets, shape international technology standards, secure critical digital infrastructure and lead innovation—all of which are vital for future economic growth. Sovereign AI can contribute significantly to national economies by transforming productivity, innovation, and competitiveness.

The US, China, and regions like the European Union have articulated sovereign AI strategies focused on securing supply chains, advancing homegrown AI research, and ensuring control over critical digital infrastructure. In India too, the government has recognized and articulated its need strongly. Sovereign AI efforts are designed not only to reduce dependence on foreign platforms but also to position India as a major AI player, advancing economic opportunity and strategic autonomy.

Key components necessary to build Sovereign AI

Sovereign AI is defined as a nation's ability to independently design, develop, deploy and regulate AI systems using domestically controlled infrastructure, data sourced within national borders and an expert indigenous workforce. This capability is aligned with the country's unique strategic interests, cultural values, and legal frameworks, enabling AI technologies to serve national priorities and security concerns. The key components of Sovereign AI are:

Domestic AI infrastructure: Ownership and control over the physical and digital hardware, which includes data centers, cloud platforms and computing resources, is necessary for sovereign AI development.

Local data sovereignty: Sensitive data and critical datasets across sectors like finance, defense, healthcare, government are collected, processed, and stored within national boundaries to preserve data privacy and security.

Indigenous talent and expertise: Building skilled domestic workforce in AI research, development and governance is crucial to maintaining innovation autonomy and operational resilience.

Strategic and legal frameworks: The development and deployment of AI must be governed by policies reflective of the nation's values, security interests and ethical standards. This includes enforcing data protection laws, AI ethics guidelines, and export controls aligned with national priorities.

Robust cybersecurity: Sovereign AI relies on data which needs to be secured against cyber threats. Domestic cybersecurity capabilities are useful in safeguarding data privacy and mitigating risks related to foreign control and ensuring operational autonomy.

Sovereigns today are wary of over-dependence on foreign AI models, which can introduce vulnerabilities for countries through covert surveillance, data leakage, or even technology denial, amid rising geopolitical tensions and the weaponization of critical technologies.

Through Sovereign AI capacity, countries look to:

- Safeguard sensitive data and critical infrastructure from foreign interference.
- Drive domestic technological innovation that is free from external constraints.
- Retain intellectual property rights and control over the AI development lifecycle.
- Enhance economic competitiveness by fostering home-grown AI-driven industries.
- Strengthen national security by reducing exposure to supply chain risks and technological dependencies.

Several countries are developing Sovereign AI capabilities with the US and China among the leaders.

- The Sovereign AI strategy of the US[24] focuses on accelerating AI innovation by removing regulatory barriers and upskilling, building American AI infrastructure such as data centers, semiconductor manufacturing and energy infra, and leading in international AI diplomacy and security by promoting safe exports.
- In its pursuit of AI dominance, China is making large state-led investments aimed at transforming its economy and achieving technological self-sufficiency. It aims to integrate AI across all aspects of the economy to create a new intelligent economy by 2035.
- EU aims to enhance the competitiveness of strategic sectors and strengthen technological sovereignty, boost AI adoption and innovation across Europe.

Countries such as India, the UK and UAE too are taking initiatives to strengthen their Sovereign AI capabilities to achieve digital self-determination in the age of AI.

IndiaAI Mission

Developing a domestic AI infrastructure through IndiaAI Mission, the country seeks to operationalize its vision of Sovereign AI with the objectives to democratize access to AI infrastructure, enhance quality and accessibility of data, cultivate indigenous AI capabilities and support ethical and responsible deployment of AI technologies. India is pursuing a 'techno-legal' approach, building technological tools alongside legal frameworks to counter threats and risks from AI[25].

Other strategic initiatives focus on building digital public infrastructure (DPI) with AI capabilities, enabling data sovereignty and fostering an innovation ecosystem that aligns with India's unique socio-cultural fabric[26].

The AI mission covers the full value chain—from domestic semiconductor manufacturing capability and compute capacity to development of sovereign LLMs and domain-specific applications. The strategy includes focus on AI Dataset Platform, deployment of 40,000 GPUs[27] and fostering indigenously developed foundational AI models tailored to Indian languages and contexts. The government has shortlisted several startups to support AI development through subsidized access to compute power from high-end GPUs. For example, Sarvam has been tasked with building the country's first indigenous AI LLM, which is expected early next year according to reports.[28]

The IndiaAI Mission: Big strides

7 pillars of the IndiaAI Mission

focusing on driving AI ecosystem growth are currently at different maturity levels

01 IndiaAI Compute

- Nearly 34,000 GPUs, including the NVIDIA B200 series, have been provisioned under a public-private partnership at subsidized rates.[29]
- The government is committing close to INR5,000 crore (from a total corpus of INR10,372 crore) to secure compute resources.[30]
- 40,000 GPUs to be provided to the innovation ecosystem in the next phase.[31]

02 IndiaAI Datasets Platform

- The "AIKosh" platform now hosts around 3,000 datasets, 200-plus models, and associated tools, serving as India's central hub for AI-ready data.[32]

03 Application Development Initiatives

- An innovation challenge led to 30 prototype applications, with pilots underway for tuberculosis diagnosis and soil moisture assessment.[33]
- Broader public services in agriculture, healthcare, climate, governance, and education are targeted for scale-up.

04 Future Skills

- Advanced fellowships and discipline-diverse grants (e.g., CATCH Grant for AI in cancer care).[34]
- 500+ data labs are being created across Industrial Training Institutions and polytechnics, aiming to train 150,000 students in three years.[35]

05 Innovation Centre

- 12 startups selected to develop India-specific LLMs.[36]
- Applications piloted, such as Bhashini-powered "Kumbh Sah'AI'yak", have showcased real-time multilingual translation, voice-based lost-and-found services, and integration with governance systems like Indian Railways.

06 Startup Financing

- Fund-of-funds under development.
- The IndiaAI Startup Global initiative sent 10 startups to accelerator programs abroad, helping them build networks and scale globally.[37]

07 Safe and Trusted AI

- Establishment of the AI Safety Institute is a step towards focusing on deepfake detection, bias testing and ethical frameworks.

Other developments related to indigenous AI ecosystem in India include:

- Three AI Centers of Excellence (CoEs) dedicated to healthcare, agriculture and sustainable cities in Delhi have been launched, with a fourth one in education (INR500 crore) announced in Budget 2025.
- Public platforms like Bhashini are demonstrating the sovereign potential of Indian AI by bridging the linguistic digital divide.
- An AI Safety Institute has been set up and is working to build tools for detecting defects and identifying biases with high accuracy.

The government is also looking to work with promising startups and supporting them through subsidies under the AI mission. After an initial round involving four companies, the government has backed several startups. Some of the notable short-listed companies among the 12 selected to develop large foundation models and their focus areas include:

Short-listed companies	Focus areas
Soket AI Labs	Creating 120B-parameter multilingual models for healthcare, defense, and education.[38]
Gnani.ai	Specializes in real-time multilingual speech AI, building India-first conversational agents and speech recognition systems on 14-B parameters.[38]
Sarvam AI	Developing a 70B-parameter sovereign language model, fully trained and hosted in India.[38]
Gan.AI	Enables hyper-personalized video and Text-to-speech generation for Indian languages at scale.[39]
BharatGen	BharatGen has launched Param-1, a bilingual LLM with 25% Indic-language data, along with 20 speech models covering 22 Indian languages. Received INR988 crore in funding.[40]
Fractal	Fractal developed Fathom-R1-14B, an open-source reasoning LLM for enterprise and medical AI.[39]
Tech Mahindra	Developing enterprise-grade GenAI and industry-specific LLMs for business adoption, called Project Indus[41]
Avataar.ai	Pioneering in 3D/AR GenAI, powering immersive commerce and marketing.[39]
ZenteiQ.ai	Builds domain-specific AI agents and custom LLMs for businesses. Received INR75 crore in year one, earmarked for compute resources.[42]
Genloop	Builds GenAI infrastructure and LLM pipelines to accelerate model training.[39]
Intellihealth	AI-powered healthcare analytics and decision support, now developing a 20B-parameter model for EEG analysis to enable early neurological screening and brain-computer interfaces.[43]
ShodhAI	Builds AI models to accelerate scientific research and data-driven discovery in fundamental sciences.[39]

Gaps and policy suggestions

Despite strong momentum, there are certain structural gaps that need urgent policy redressal to strengthen India's sovereign AI agenda:

Robust domestic solutions for hardware and cloud solutions: India's sovereign efforts could benefit from development of indigenous semiconductor and GPU capabilities to reduce dependence on global providers. Local hyperscale data centers and sovereign cloud services are essential to shield India from foreign jurisdictional overlaps.

Some of the key issues that impact AI data center capacity creation have to be addressed, such as higher cost of land, limited affordable round-the-clock green energy, high import dependency for components in data centers like storage devices (e.g., Solid State Drives or SSDs), specialized transformers, switchgears precision cooling units.

Strategy to attract top-tier AI talent: Nurturing and retaining top-tier AI talent would require a supportive ecosystem comprising globally competitive research institutions with financial autonomy, risk capital availability, and vibrant industry-academia linkages. The current tightening of overseas work visas (e.g., in the US) provides India a window of opportunity to retain domestic talent and attract Indian-origin talent at home. However, given that top-tier AI talent is mobile globally, attractive policy incentives need to be put in place to incentivize relocation to India. State and local authorities in China, for example, offer a range of incentives like financial support and subsidies, tax incentives and funding for R&D, special talent visas and fast-track immigration.

Data strategy for supporting innovation: India generates vast amounts of data but continues to face a "data divide" due to fragmentation, lack of annotation and much of the data being locked with foreign tech companies. There is limited access to government data. Steps to address related issues could include:

- Mandates are needed to enable anonymization and release of government and anonymized private datasets. Open data initiatives, emulating models like the US Data.gov, could help lower entry barriers for startups. Creation of data marketplaces could incentivize proprietary, sector-specific dataset sharing.
- While over 3,000 high-quality datasets (on AIKosh as of October 2025) are available to train AI models, the quality of sources such as data.gov.in can vary. Issues such as lack of consistency, annotation, etc., have to be addressed. Moreover, access to anonymized data from civic bodies and private organizations is limited. High-quality data on Indian languages is also sparse at present.
- Government departments can adopt the Data Quality Maturity Framework, proposed by NITI Aayog[44], as a self-assessment guide for public institutions for planned and progressive approach to data quality.
- Data protection rules, under the DPDP Act, must balance privacy with innovation and avoid disproportionate compliance burdens on young startups. Some provisions in the draft rules may pose challenges to innovation, like restrictions on processing data in foreign countries and ambiguities on processing personal data available in public domain.

Seen collectively, India's approach reflects a serious attempt to embed sovereignty into AI development and enabling control over computing, datasets, applications, talent and safety frameworks. However, sustaining such a trajectory will require further bold policy moves in computing, chips, cloud, data governance and human capital, where India's comparative advantages are consolidated into developing sustainable strategic assets. Inherent strengths such as use of AI to build efficiency through its large number of STEM graduates, can be leveraged and domestic LLMs, as and when they develop, can be used. Until then, open sources models could be popularized from an application point of view. To enhance R&D, an attractive talent strategy could benefit. Lastly, as the biggest generator of data, government policies could focus on making large quantities of machine readable, annotated datasets available to AI companies, startups and researchers.

Methodology

EY conducted a comprehensive survey to assess the impact and adoption of GenAI and Agentic AI across the Indian enterprise landscape. The study covered 200 organizations spanning over 20 industries, including government bodies, public sector undertakings, startups, Indian enterprises, global capability centers (GCCs) and the Indian arms of multinational corporations.

Responses were gathered from C-suite executives and senior leaders, offering a strategic perspective on how GenAI is influencing both core operations and support functions. The survey design and analysis were guided by proprietary frameworks of EY, analytical tools and sectoral specialization; refined through years of client engagement and industry research.

Sources

[1]The Rise of AI Agents in the Workforce | Workday US

[2]https://www.axisbank.com/annual-reports/2024-2025/pdf/Business%20performance%20review.pdf

[3]https://sbi.bank.in/web/about-us/about-us

[4]https://www.hul.co.in/news/2025/how-ai-is-powering-huls-consumer-customer-and-operational-ecosystems/

[5]https://www.jio.com/platforms/offerings/automation-ai-ml-platforms/jio-brain/

[6]Airtel announces new features in its continued fight against SPAM

[7]https://business.adobe.com/summit/2025/sessions/how-air-indias-digital-transformation-is-driving-s943.html

[8]https://www.infosys.com/services/data-ai-topaz/insights/recharging-developer-productivity.pdf

[9]Browserstack: Enhance testing with AI-driven productivity

[10]GenAI in Indian e-commerce: Q&A with Mayur Datar

[11]https://www.ey.com/content/dam/ey-unified-site/ey-com/en-in/services/ai/aidea/2025/01/ey-the-aidea-of-india-2025-how-much-productivity-can-genai-unlock-in-india.pdf

[12]https://www.cnbc.com/2025/08/04/indias-it-layoffs-spark-fears-ai-is-hurting-jobs-in-critical-sector.html

[13]https://economictimes.indiatimes.com/jobs/fresher/ais-quiet-takeover-why-freshers-are-missing-in-it-hiring/articleshow/122948560.cms

[14]EY analysis

[15]https://nasscom.in/knowledge-center/publications/aideation-impact-architecting-ai-first-workforce-tech-services-sector

[16]The evolving ethics and governance landscape of agentic AI | IBM

[17]AI governance in the agentic era

[18]https://artificialintelligenceact.eu/

[19]https://hai.stanford.edu/ai-index

[20]https://techpost.bsa.org/2024/10/16/enhancing-ai-accountability-effective-policies-for-assessing-responsible-ai/#:~:text=To%20help%20identify%20and%20manage,management%20system%20standard%2C%20ISO%2FIEC%2042001

[21]https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/FREEAIR130820250A24FF2D4578453F824C72ED9F5D5851.PDF

[22]https://impact.indiaai.gov.in/

[23]https://www.pib.gov.in/PressReleasePage.aspx?PRID=2178092

[24]https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

[25]https://www.ptinews.com/story/business/india-adopts-techno-legal-approach-for-ai-safety-vaishnaw/2917809

[26]Press Release:Press Information Bureau

[27]https://www.moneycontrol.com/artificial-intelligence/sarvam-likely-to-roll-out-india-s-first-llm-by-early-next-year-40-000-gpus-to-be-empanelled-by-centre-article-13529063.html

[28]https://www.ndtvprofit.com/technology/indias-own-chatgpt-homegrown-chatbot-sarvam-ai-to-launch-soonhow-will-it-help-you

[29]https://www.pib.gov.in/PressReleasePage.aspx?PRID=2132817

[30]https://ai.icai.org/articles_details.php?id=107

[31]Sarvam likely to roll out India's first LLM by early next year, 40,000 GPUs to be empanelled by Centre

[32]https://aikosh.indiaai.gov.in/home

[33]https://www.pib.gov.in/PressReleasePage.aspx?PRID=2108961#:~:text=The%20IndiaAI%20Application%20Development%20Initiative%20(IADI)%20pillar%20under%20the%20IndiaAI,%2C%20Prototype%2C%20and%20Existing%20Solutions.

[34]https://www.pib.gov.in/PressReleasePage.aspx?PRID=2168319&ref=startuptalky.com#:~:text=This%20brings%20the%20IndiaAI%20Fellowship,fellowship.indiaai.gov.in

[35]https://www.youtube.com/watch?v=-1Wp1Cuv5Ig&t=3s

[36]https://www.entrepreneur.com/en-in/news-and-trends/india-expands-ai-mission-with-12-llm-builders/497343

[37]https://www.pib.gov.in/PressReleasePage.aspx?PRID=2132377

[38]IndiaAI Foundation Model Pillar Empanels Three More Indigenous Models - Speak loud with your Brand

[39]https://www.linkedin.com/pulse/rise-indias-sovereign-ai-analytics-india-magazine-yyjpc/

[40]https://www.devdiscourse.com/article/business/3631946-bharatgen-revolutionizing-indias-ai-landscape-with-sovereignty-and-accessibility

[41]https://www.thehansindia.com/technology/tech-news/india-selects-8-organisations-including-iit-bombay-to-build-trillion-parameter-ai-model-1007948

[42]https://www.thehindubusinessline.com/news/science/ai-startup-zeinteiq-to-develop-8-billion-parameter-scientific-model-under-indiaai-mission/article70080799.ece

[43]https://www.entrepreneur.com/en-in/news-and-trends/india-expands-ai-mission-with-12-llm-builders/497343

[44]https://www.niti.gov.in/sites/default/files/2025-06/FTH-Quaterly-Insight-june.pdf

APPENDIX

Overview of key ‘Sovereign AI’ programs

1 USA

Key focus area: Fuelled R&D through massive private venture capital (US$100 billion in 2022 and 2023) and acquisitions (e.g., OpenAI/Microsoft, DeepMind/Alphabet)[1]. Relies on private investment for compute (e.g., the US$500 billion Stargate project)[2]. Government focuses on streamlining permitting for data centers[3] and energy infrastructure[4]. Aggressively uses export controls on advanced chips[5] (e.g., Nvidia GPUs) to maintain a technological lead and restrict competitors like China. Promotes "digital solidarity" to export the "American AI stack"[6] and shape global standards.

Strengths: Private sector innovation among global leaders, control over critical semiconductor supply (Nvidia), massive capital and first-mover advantage.

Weaknesses: Potential over-reliance on private actors whose goals may not always align with public interest; risk of concentration of power among few tech giants.

2 China

Key focus area: Guided by the New Generation AI Development Plan (2017)[7] and the new "AI Plus" initiative. Direct state funding, including a new US$8.2 billion National AI Industry Investment Fund[8]. Subsidizes compute[9], backs domestic semiconductor alternatives and has created over 200 LLMs in a protected market[10].

Strengths: Among leaders in research output (40% of AI publications, 60% of global AI patents) and talent production (47% of top AI researchers)[11]. Ability to mobilize vast national resources.

Weaknesses: Hampered by US export controls, creating a compute disadvantage, regulatory environment and focus on control may limit bottom-up innovation.

3 Saudi Arabia/UAE

Key focus area: The Saudi PIF fully funds HUMAIN, a state-owned enterprise building everything from data centers (using thousands of Nvidia GPUs) to Arabic LLMs (ALLAM)[12]. The UAE's G42 partners with OpenAI, Microsoft and Oracle on projects like Stargate UAE, aiming to build the world's largest AI data center outside the US.[13]

Cultural focus is on developing models (like HUMAIN Chat) specifically for the Arabic language and cultural context.

Strengths: Large capital, strong political will, no restrictions due to legacy systems, ability to purchase top-tier technology.

Weaknesses: Nascent innovation ecosystems, reliance on importing top talent and technology.

4 EU

Key focus area:

- Public compute: EU's AI Factories/ Gigafactories[14] focus on building public supercomputing infrastructure for researchers and SMEs.
- Regulation first: The EU's AI Act sets a global benchmark for risk-based regulation. The AI Continent Action Plan[15] aims to boost competitiveness within this regulatory framework.

Strengths: The EU is among the frontrunners in establishing the de facto global regulatory standard for trustworthy AI, which other nations must contend with.[16]

Weaknesses: Lagging the US and China in private investment and scale of tech giants; regulation could slow down innovation and deployment.

[1]Who Is Winning The AI Arms Race?
[2]Stargate: Trump announces a $500 billion AI infrastructure investment in the US | CNN Business
[3]New White House AI Action Plan and EO aim to streamline environmental review and data center permitting | DLA Piper
[4]New Executive Order Bridges Energy Policy & AI — Boundary Stone Partners.
[5]Who Is Winning The AI Arms Race?
[6]Promoting The Export of the American AI Technology Stack – The White House
[7]China's New Generation of Artificial Intelligence Development Plan - Foundation for Law and International Affairs
[8]China Invests $8.2 Billion to Global AI Dominance - EE Times
[9]From Vouchers to Visas: China's Innovative Plan for AI Dominance - Foreign Policy Research Institute
[10]Who Is Winning The AI Arms Race?
[11]Who Is Winning The AI Arms Race?
[12]HUMAIN releases HUMAIN Chat powered by ALLAM.
[13]Abu Dhabi's Stargate to Become World's Largest AI Data Hub — Arabian Post
[14]AI Factories | Shaping Europe's digital future
[15]AI Continent Action Plan | Shaping Europe's digital future
[16]The EU's AI Power Play: Between Deregulation and Innovation | Carnegie Endowment for International Peace

Acknowledgements

Core team

- Rajiv Memani
- Mahesh Makhija
- Hari Balaji
- Rohit Pandharkar
- Rajnish Gupta
- Gagan Sharma
- Jatin Patni
- Yash Dandavate

Brand, marketing & communications

- Jerin Verghese
- Sharon Dias
- Kritarth Srivastava
- Isha Handa
- Pallavi Janakiraman
- Rajshree Sharma
- Sakshi Sharma

Editorial

- Prosenjit Datta
- Radhika KTP
- Kaveri Nandan
- Sushmita Yadav
- Kartik Verma

Design

- Jaspal Singh
- Sneha Arora

Confederation of Indian Industry

The Confederation of Indian Industry (CII) works to create and sustain an environment conducive to the development of India, partnering Industry, Government and civil society through advisory and consultative processes.

CII is a non-government, not-for-profit, industry-led and industry-managed organisation, with around 9,700 members from the private as well as public sectors, including SMEs and MNCs, and an indirect membership of over 365,000 enterprises from 318 national and regional sectoral industry bodies.

For more 130 years, CII has been engaged in shaping India's development journey and works proactively on transforming Indian Industry's engagement in national development. CII charts change by working closely with the Government on policy issues, interfacing with thought leaders, and enhancing efficiency, competitiveness, and business opportunities for industry through a range of specialised services and strategic global linkages. It also provides a platform for consensus-building and networking on key issues.

Through its dedicated Centres of Excellence and Industry competitiveness initiatives, promotion of innovation and technology adoption, and partnerships for sustainability, CII plays a transformative part in shaping the future of the nation. Extending its agenda beyond business, CII assists industry to identify and execute corporate citizenship programmes across diverse domains, including affirmative action, livelihoods, diversity management, skill development, empowerment of women, and sustainable development, to name a few.

For 2025-26, CII has identified "Accelerating Competitiveness: Globalisation, Inclusivity, Sustainability, Trust" as its theme, prioritising five key pillars. During the year, CII will align its initiatives to drive strategic action aimed at enhancing India's competitiveness by promoting global engagement, inclusive growth, sustainable practices, and a foundation of trust.

With 70 offices, including 12 Centres of Excellence, in India, and 9 overseas offices in Australia, Egypt, Germany, Indonesia, Singapore, UAE, UK, and USA, as well as institutional partnerships with about 250 counterpart organisations in almost 100 countries, CII serves as a reference point for Indian industry and the international business community.

Confederation of Indian Industry
The Mantosh Sondhi Centre
23, Institutional Area, Lodi Road, New Delhi – 110 003 (India)
T: 91 11 45771000
E: info@cii.in ▸ W: www.cii.in

Follow us on

cii.in/facebook cii.in/twitter cii.in/linkedin cii.in/youtube

Reach us via CII Membership Helpline: 1800-103-1244

Published by Confederation of Indian Industry (CII), The Mantosh Sondhi Centre; 23, Institutional Area, Lodi Road, New Delhi 110003, India, Tel: +91 11 45771000; Email: info@cii.in; Web: www.cii.in

Ernst & Young LLP

EY | Building a better working world

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multi-disciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

Ernst & Young LLP is one of the Indian client serving member firms of EYGM Limited. For more information about our organization, please visit www.ey.com/en_in.

Ernst & Young LLP is a Limited Liability Partnership, registered under the Limited Liability Partnership Act, 2008 in India, having its registered office at Ground Floor, Plot No. 67, Institutional Area, Sector - 44, Gurugram, Haryana - 122 003, India.

EYIN2511-007

SA1

ey.com/en_in

@EY_India EY EY India EY Careers India @ey_indiacareers