



The *Al*dea of India

**Generative AI's potential
to accelerate India's
digital transformation**

EY

Building a better
working world

Contents



PAGE
04

Foreword



PAGE
06

**Executive
Summary**



PAGE
12

Chapter 1
**Three Steps
to the Future**



PAGE
20

Chapter 2
Rethinking
the Enterprise
Agenda



PAGE
36

Chapter 4
The Economic
Opportunity
of Gen AI
in India



PAGE
42

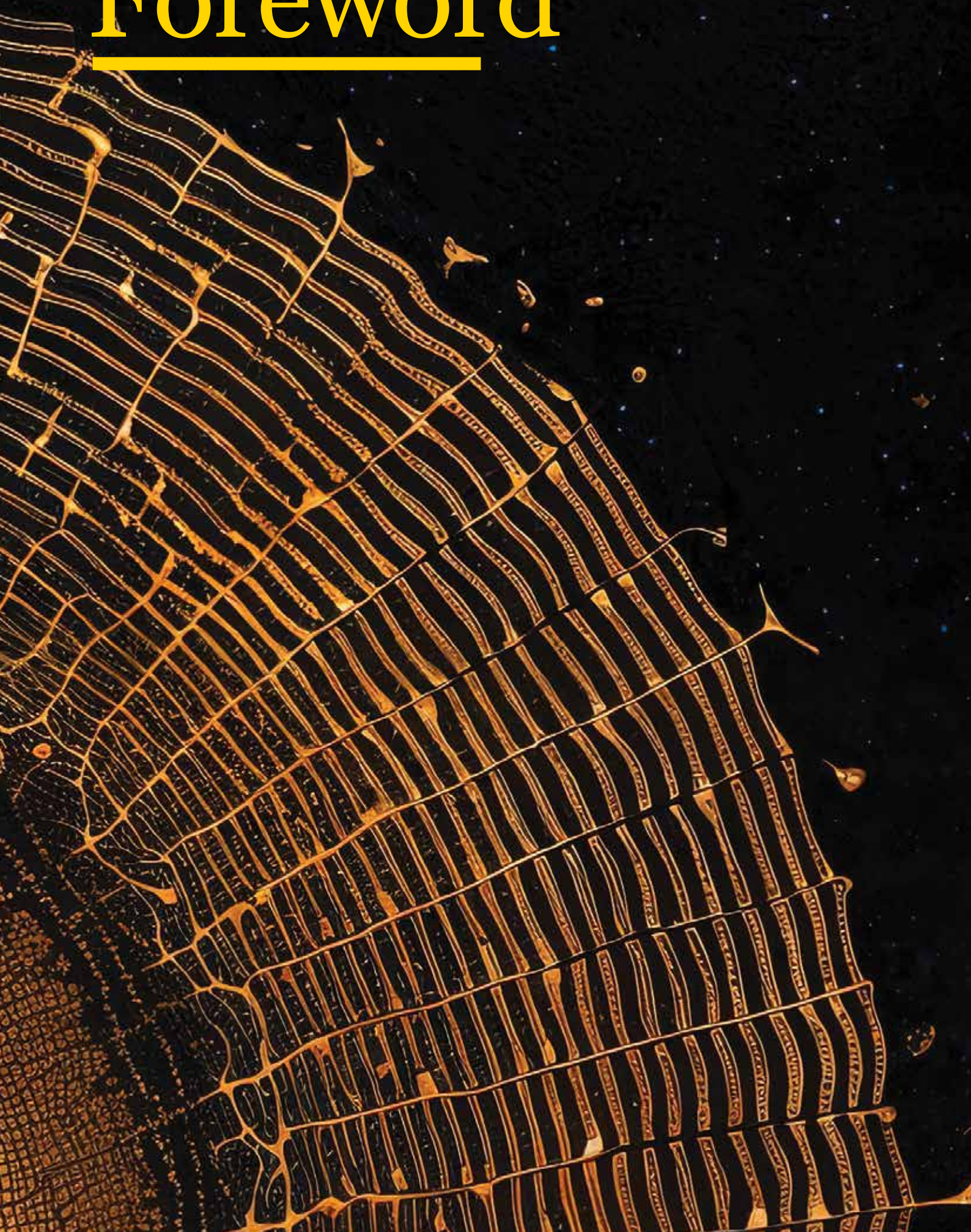
Chapter 5
A Gen AI
Policy Agenda
for India



PAGE
54

Annexures

Foreword



India, today, is in a powerful position to set the agenda and realize significant benefits as the world starts to leverage an explosion of AI capabilities. While 2023 was a year of rapid innovation, in the next decade, Generative AI (Gen AI) will break out from labs and Proofs of Concept (POCs) and into the open terrain of consumer and enterprise applications. Millions of Indian citizens stand to benefit from next-generation scaled AI applications across industries with the biggest impact on healthcare, drug discovery, financial services, education and entertainment.

A confluence of several factors can make India a big player in this transformation. Over the last decade, we have crafted a uniquely Indian approach to digital transformation. India has one of the most modern 'open' digital infrastructures in the world. With inclusion as a primary goal, today, nationwide scaled utilities aimed at digital identity, KYC, payments and e-commerce have become an integral part of the digital architecture of India. Breakthrough innovation in Gen AI can be accelerated by public-private collaboration and we have established models for success in this area.

India will itself soon be one of the largest markets in the world and hence a massive playground for AI applications to drive growth and productivity for enterprises. Investments in the next leg of consumer-facing technology, next-generation supply chains and intelligent automation platforms for straight through processing of entire processes have the potential to lead with an AI-first approach to build new tech, thus leapfrogging legacy paradigms.

Equally, India has emerged as the second largest generator of digital data, behind only China, and this gives it an advantage when it comes to training Gen AI models that need vast amounts of data to learn before they can perform their magic.

All this is buttressed by a large STEM talent base and a vibrant start-up ecosystem that has been nurtured to innovate faster using Gen AI. The COVID-19 pandemic has taught us that work can be delivered from anywhere and this positions India as the preeminent destination of choice as companies globalize their operations.

Companies across India are already starting to take an AI-first approach to digital. They are figuring out how to transform their customer experience, improve productivity and become more agile at the delivery of digital capabilities using new foundation models and other components of the emerging AI architecture. Big Tech companies are gearing up to launch their latest Gen AI tools for the large ecosystem of technologists in India.

While this is still very early, there is a tremendous sense of optimism in the air. To realize this potential, several pieces need to fall in place from a policy perspective. India will have to dramatically up its game to compete in this landscape in terms of increased government role in development and deployment, providing critical compute infrastructure for start-ups and corporates to innovate, enabling access to Indian datasets and data marketplaces and clarifying the regulatory regime around AI.

Similarly, enterprises need to also establish guardrails and governance for Responsible AI as this technology comes with its share of risks like bias, hallucinations, data security which if not addressed can lead to reputational and financial losses.

This report, '**The A/idea of India: Generative AI's potential to accelerate India's digital transformation**,' lays out the potential impact of Gen AI on India and dives deep into several facets of what companies and the government can do to enable leadership in this space.

Today, India has the economic and political heft to set the agenda from the inception of this new technological paradigm to its broad scale adoption. ●



Rajiv Memani
Chairman and CEO,
EY India



Executive Summary

Gen AI is poised to reshape industries by revolutionizing operating models, transforming value chains, and altering economic dynamics. Indian enterprises are optimistic about Gen AI's potential but acknowledge the need for better preparation, navigating a digital transformation journey to survive in the era of 'Digital Darwinism'.

Our report, '**The AIdea of India: Generative AI's potential to accelerate India's digital transformation,**' addresses key questions facing enterprises and policy makers. In Chapter 1, we delve into the key trends emerging in the AI/Gen AI ecosystem. Chapter 2 introduces an agenda for enterprises, outlining a comprehensive strategy for enterprise transformation to act as a guide in Gen AI adoption. Chapter 3 focuses on key industries that Gen AI is expected to impact and provides insights into potential use cases. Chapter 4 explores the economic opportunities presented by Gen AI in India throughout this decade. Chapter 5 outlines the Gen AI policy agenda for India, offering a strategic framework to navigate policy considerations in the Gen AI landscape.

As organizations navigate the early stages of their AI journey, some engage in successful Proofs of Concept (POCs) but lack a comprehensive strategy for scalable and sustainable business value delivery. Others are in the assessment phase, carefully evaluating risks and identifying suitable POCs tailored to their circumstances. This moment offers an opportunity to assess the technology stack fully and harness AI's potential. Foundation models, especially Large Language Models (LLMs), agent frameworks and knowledge bases, are becoming integral components of knowledge-intensive application architecture.

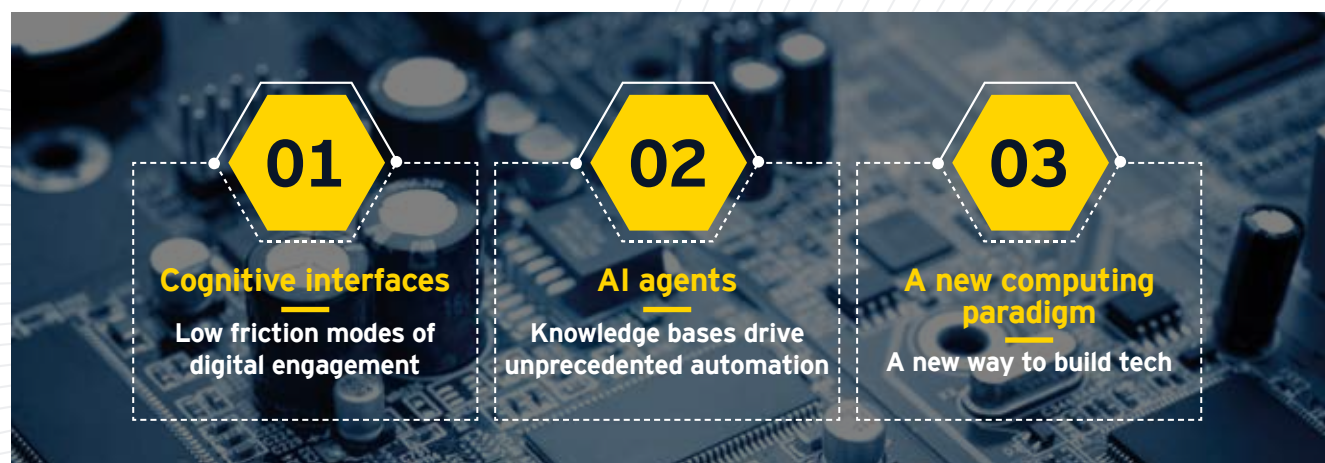
With consumer-facing applications progressively incorporating innovative cognitive interfaces, enterprises must enhance their existing technology stack with Gen AI tools, bolstering data and application layers to accommodate advancements. In navigating the dynamic business landscape, strategic decisions are imperative, focusing on data utilization, use case prioritization, and managing data security and compliance control related to LLMs, alongside a nuanced evaluation of cost, ROI and speed to value.

Indian enterprises are optimistic about Gen AI's potential but acknowledge the need for better preparation, navigating digital transformation to survive in the era of 'Digital Darwinism'

Despite the promises Gen AI holds, inherent risks accompany its complexity in model deployment and validation. While there is high awareness regarding data and model-related parameters for 'Classical'

Machine Learning (ML), the same is not true for Gen AI. Enterprises must be vigilant in managing Gen AI-related risks to prevent reputation and financial loss. C-suite executives and leaders could proactively comprehend and integrate processes for risk mitigation and governance. Currently, organizations identify data privacy as the single-most crucial risk of Gen AI.

In the short term, enterprises will grapple with a shortage of individuals possessing AI skills, a challenge that is expected to persist. Despite India's commendable standing in AI skill penetration and talent concentration, the advent of Gen AI exacerbates this shortage. The widening gap between skills that companies demand and the existing workforce necessitates strategic talent acquisition, particularly for successful initiation and scalable implementation of prioritized use cases. Gen AI implementation demands a diverse skill set within Gen AI Centers of Excellence or Pods, encompassing AI engineering, data science, and expertise in foundation models, LLMs, AI orchestration, prompt engineering, Retrieval Augmented Generation (RAG), fine-tuning and model deployment.



Amidst rapid evolution of Gen AI across various dimensions, it is by no means certain what specific path this ecosystem will take going forward. Despite this uncertainty, specific trends are emerging - cognitive interfaces are reshaping the application landscape, AI agents are transforming the nature of work, and a new general computing paradigm is emerging, signifying a fundamental shift in computational processes and capabilities.



Global institutions like the World Bank and the IMF recognize India's economy as the fastest-growing among major nations. Our study indicates that India could experience a substantial boost in its GDP over seven years (2023-24 to 2029-30). The cumulative impact on GDP may range from US\$1.2 trillion to US\$1.5 trillion, contributing an additional 0.9% to 1.1% in annual CAGR. Given the immense capability of Gen AI with respect to its productivity and efficiency enhancing effects, its adoption has the potential to accelerate India's growth trajectory. It is, therefore, necessary to increase investment in Gen AI, education and upskilling to fully capitalize on the demographic dividend.

While Gen AI's positive impact is anticipated across all sectors, its level of influence in each segment will depend on factors like feasibility, adoption rates, the organized sector's share and an industry segment's contribution to India's economic activity. Approximately 69% of the overall impact is expected to be derived from business services (including IT, legal, consulting, outsourcing, rental, etc.), financial services, transportation and logistics, education, retail trade and healthcare. The expected

impact encompasses improvements in employee productivity, enhanced operational efficiency and personalized customer experiences. These sectors, having rapidly embraced digitalization, are well-positioned to capitalize on the benefits of Gen AI. The IT sector also stands to gain significantly from the development of Gen AI platforms and tools both through productivity gains and through more revenues from its clients.

Amidst rapid evolution of Gen AI across various dimensions, it is by no means certain what specific path this ecosystem will take going forward

Recognizing the ability of Gen AI to serve as an engine of economic growth, governments all over the world are looking to actively promote and regulate AI. Strategies differ across countries and there is

no consensus yet on the approach to promoting or regulating Gen AI. The Indian government recognizes the economic potential of AI and some public figures have called for a sharper India strategy for AI development.

Feasibility, adoption rate, organized sector's share and the industry segment's share in India's economic activity will determine the level of Gen AI's impact on specific sectors

In line with the development of Digital Public Infrastructure such as the India Stack, Aadhaar, UPI, etc., the government can consider developing Gen AI systems as Public Goods. This approach is in line with the National Strategy for Artificial Intelligence (2018), which emphasizes an inclusive 'AI for ALL' lens. The development of Gen AI as Public Goods can be a game changer as it can also be deployed

across various sectors of impact such as education, healthcare, agriculture, smart cities, etc., where the government is a key player.

To promote the development of Gen AI, policy actions will have to ensure access to data. The development of indigenous training data sets (especially for local Indian languages) will be very important. The government may invest in creation of structured and unstructured datasets, which can be opened to the public. These datasets, compiled through various government surveys or generated through administrative processes, can be made available in a format that is easily usable for Gen AI development. The government may also consider setting up new data pipelines to capture digitized government data/documents (especially in Indian languages), opening up existing structured non-personal or anonymized government datasets for wider consumption.

Besides data, the government could ensure access to critical digital infrastructure (through roll-out of 5G, data center development, access to specialized chips and AI specific compute infrastructure), along with policies that cultivate and attract specialized talent.

The agenda for enterprises



Reimagining digital-AI first

Enable new business models, customer journeys, products and services



Rebuilding the Tech Stack

Rebuild the tech stack-apps, data, infrastructure



Equipping the AI workforce

Empower employees with the skills for the coming change

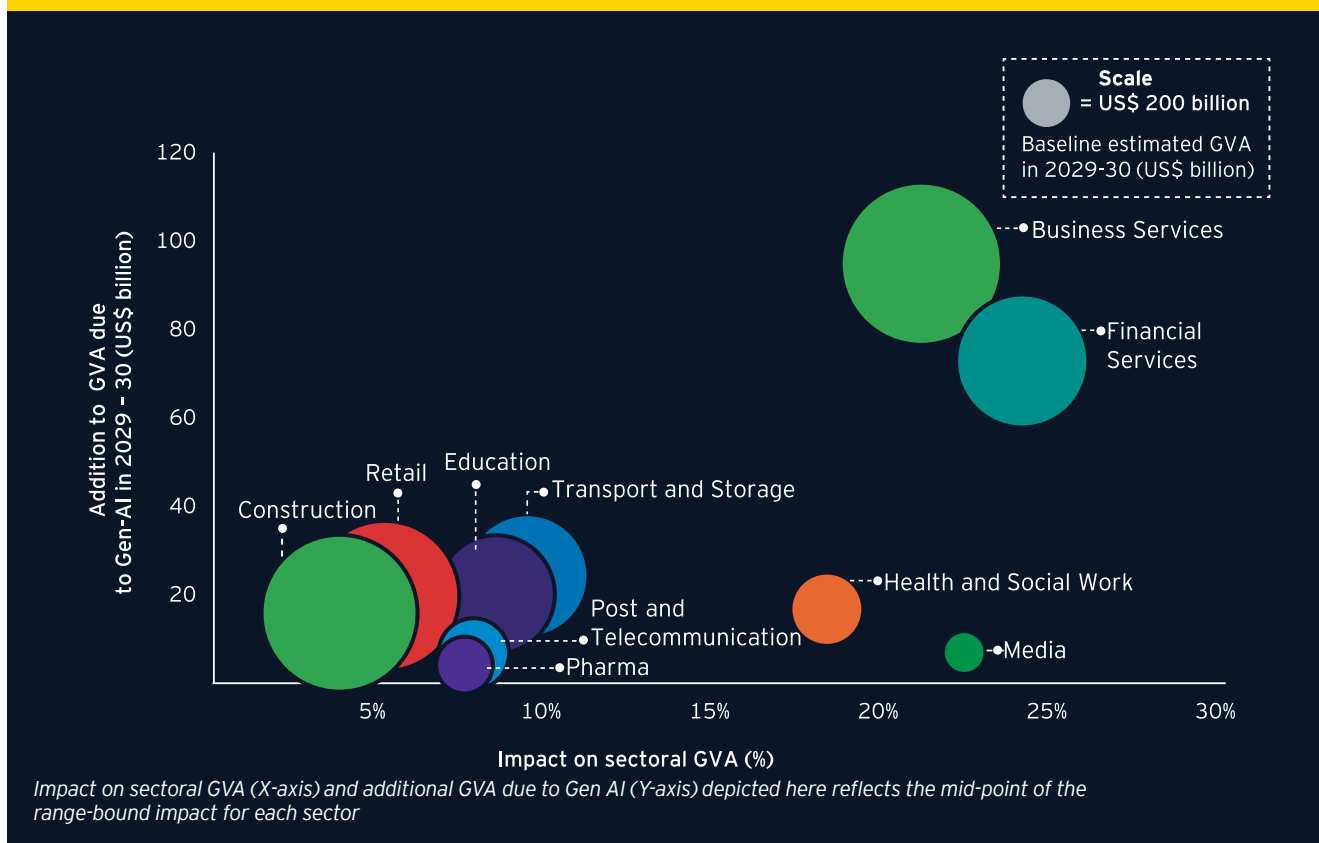


Shielding with Responsible AI

Put in place new governance-AI policies, data privacy, responsible approaches

The agenda for enterprises is clearly cut out. In the face of imminent digital transformation, the shift to an AI-first approach goes beyond implementing chatbots or experimenting with trendy tools. It demands a complete reimagination of the Digital Transformation strategy, leveraging AI alongside digital, cloud, and automation capabilities. This enables the emergence of new business models, widespread personalization and accelerated product and service innovation, in addition to the integration of Gen AI co-pilots and auto-pilots for intelligent automation and decision-making across processes and personas. Unlike other emerging technologies, Gen AI is now easily accessible on demand, simplifying the establishment of the technological foundation. Development of a clear enterprise AI and Gen AI strategy is paramount.

Sector-wise impact of Gen AI on GVA (2029-30)



To regulate Gen AI, a 'light touch' approach may help establish a responsive regulatory environment that effectively addresses the impact of this rapidly evolving technology.

The government can support innovation by facilitating interventions to improve access to data, chips, talent, computing resources, etc.

The approach will have to facilitate innovation while managing potential risks. This may be achieved through regulatory sandboxes, similar to what the government (RBI) had done to develop regulations for the FinTech industry. New algorithms can be tested in a 'controlled environment' to address risks and help develop new regulations that may be required to deploy the tested algorithms. Other measures could include the use of watermarking of Gen AI-created content and establishing technical standards. As the understanding of the technology

An agenda for the government

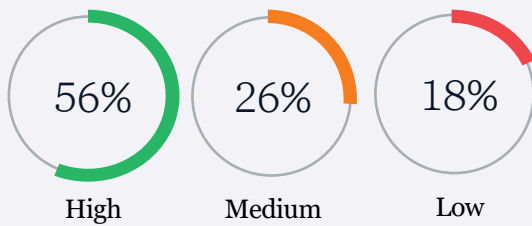
- Access to training data and marketplaces
- Deploying Gen AI systems as Public Goods
- Securing critical digital infrastructure
- Access to talent and public funding of R&D
- Clarity on regulatory framework
- Deployment of regulatory sandboxes
- Watermarking AI-generated products/media
- Accuracy, accountability and liability

and its implications improves, greater clarity may be developed on the regulatory framework required for safe adoption of Gen AI in India.

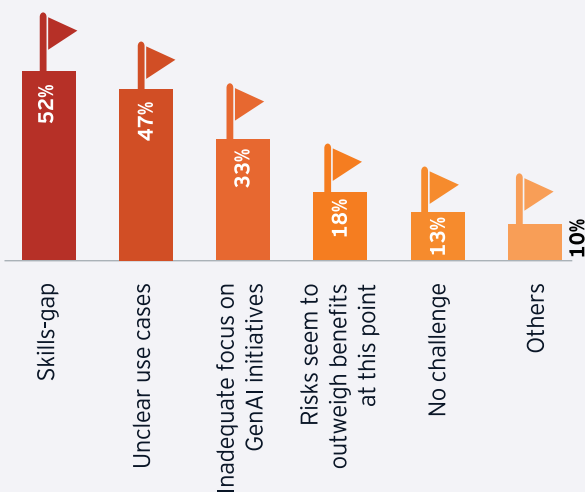
EY India's C-suite Gen AI survey

We conducted an in-depth Gen AI survey covering ~200 C-suite executives across India. They represent diverse sectors, including Technology; Media and Entertainment; Financial Services; Government; Health, Pharma and Life sciences; Retail; and Manufacturing.

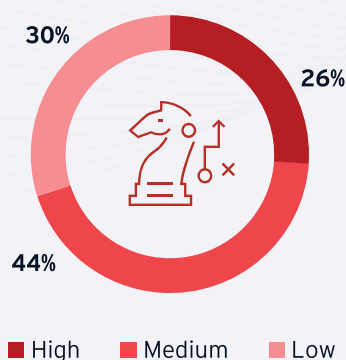
Three in five respondents see Gen AI having a significant impact on their business



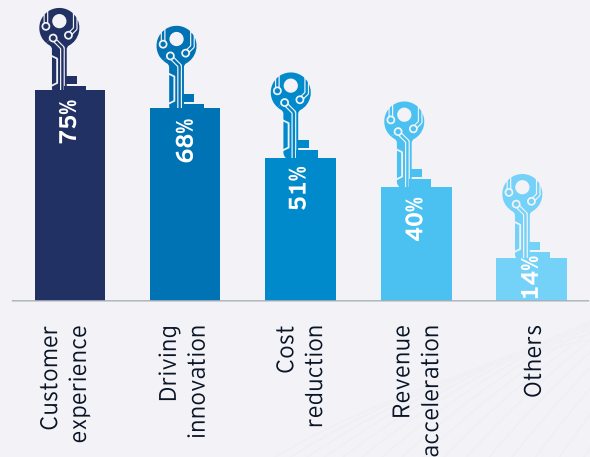
Skills gap and unclear use cases are two of the biggest challenges that organizations currently face



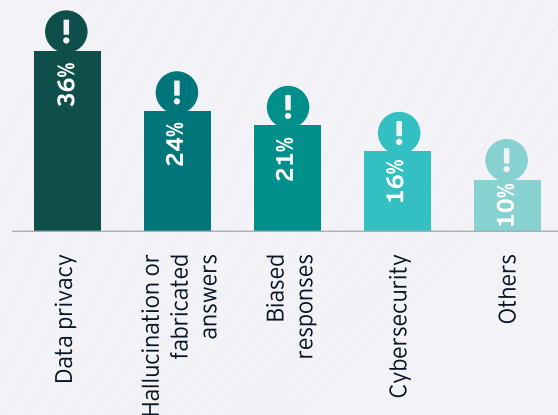
Three-fourth of the respondents indicate a low to moderate level of readiness in benefiting from Gen AI



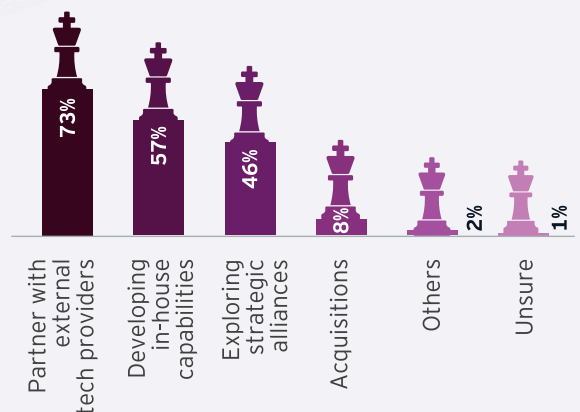
Customer experience is the single most important facet that Gen AI is expected to impact



Organizations see data privacy as the single most important risk of Gen AI



The preferred mode for Gen AI implementation is by partnering with external tech providers





Chapter 1

Three Steps to the Future



Three Steps to the Future



Foundation models will **fundamentally change the nature of digital interfaces** enabling richer textual, voice image and video interactions



Knowledge bases will allow the development of **intelligent autonomous agents** that will release massive productivity gains



Gen AI will be embedded into all applications enabling a move towards a **new computing architecture**

New battle lines on the digital front

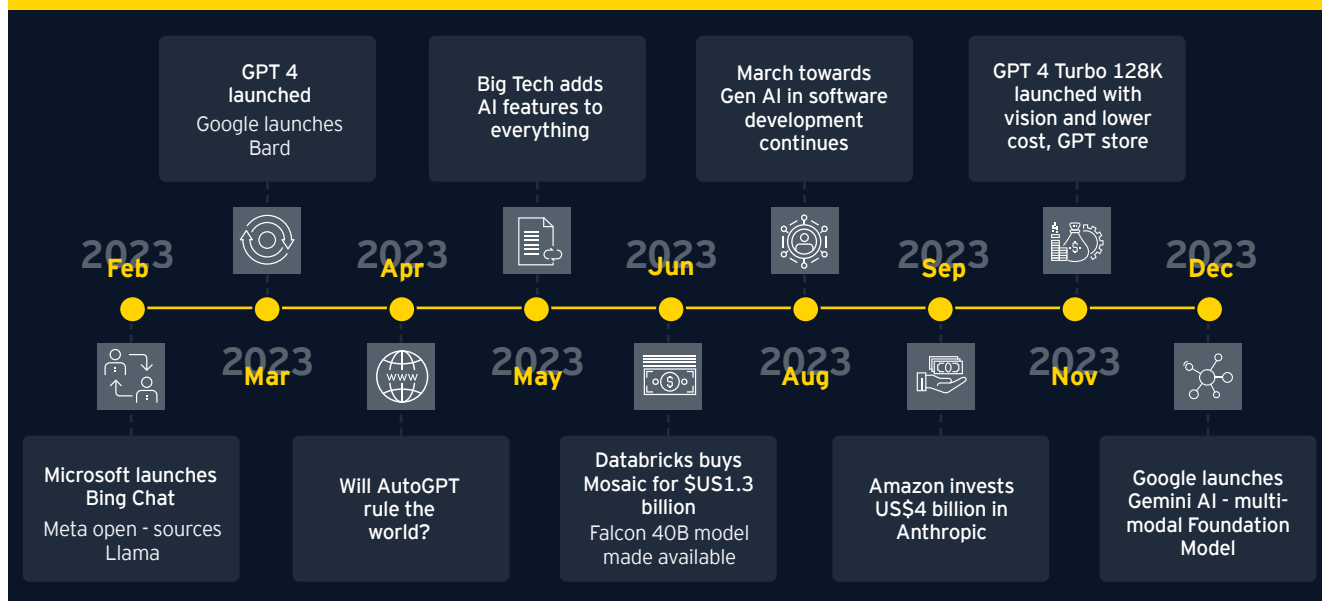
AI has captured our current zeitgeist. After a gradual maturation of deep learning technology in the last decade, stunning innovations in Gen AI and foundation models over the last few quarters have now significantly shifted the frontier of the possible. Supercharged with generative algorithms, massive datasets and extremely powerful computers, for the first time in history, we have AI models that seem to generalize well beyond the tasks on which they have been trained. Decades of research has broken out of labs, universities and big tech firms into the open terrain of the consumer and corporate world.

On the surface, there is a tremendous boost to tech optimism that is a sign of our times. As Marc Andreessen, co-creator of Mosaic browser and now General Partner of Andreessen Horowitz, one of Silicon Valley's eminent Venture Capital firms, puts it in his essay 'Why AI will save the world', "Anything that people do with their natural intelligence today can be done much better with AI."

Clearly, we can apply these rapidly emerging capabilities and use AI as a force for good across the planet. Foundation models seem poised to drive massive change across a range of domains - drug discovery, healthcare, education, science, citizen services, financial services, business, and more.

However, there are still many open questions regarding the seeming omnipotence and intelligence of these foundation models. Do they really understand anything about the world or are they just 'stochastic parrots' stuffed with data spitting out the next token? Can they ever make a coherent plan or reason about our world? Will the hallucination problem ever be solved with the current data oriented training paradigm? From this point of view,

A short history of the world since ChatGPT



foundation models are a necessary but not sufficient step forward.

There is no doubt that many more fundamental advances are needed till we can claim to achieve anything close to human-level intelligence. But the pace of innovation is rapid and the potential, even with the rudimentary tools in front of us, is immense. There is little doubt that we are at the precipice of fundamental change in how we use technology in our lives.

“

The future is already here – it’s just not evenly distributed.

”

William Gibson

In the last decade, ubiquitous mobile phones, large datasets and cloud computing enabled companies to reinvent their business models using digital-first approaches. Every business saw onslaughts from competitors finding new ways to reach and engage consumers, build more agile supply chains and compete on lower operational cost.

Today, we are poised for another wave of change, but the battle lines are different. A new class of apps powered by Gen AI with frictionless and ambient

interfaces will enable a whole new level of customer intimacy and challenge the hegemony of the ‘legacy’ mobile app ecosystem that we see around us today. Autonomous agents leveraging enterprise knowledge bases and powered by foundation models will work side by side with humans - not only to automate operations-intensive tasks like claims processing but also to enable cutting edge creative tasks like drug discovery and movie production. And along the way, a new technology architecture will be born. Spatial computing, augmented reality, vector knowledge bases, agent frameworks and foundation models will be the new picks and shovels of this gold rush.

A firehose of innovation

The pace of innovation over the last few quarters has been stunning. Every day we see the release of shiny new toys and capabilities that can accomplish heretofore impossible feats.

There has been a significant advancement in the capabilities of foundation models. Today we take the basic capabilities of foundation models for granted. These include conversations with long running context on abstract topics, summarizing web search results, condensing text and even querying structured datastores. We think it natural to point chat bots at PDF documents, websites and



even structured data stores and we expect answers with simple conversational queries. We have grown accustomed to these models being able to perform tasks on which they were not specifically trained - i.e., they can generalize well beyond their training data which is arguably one of the key measures of intelligence.

All this happened in a few months. It is worth taking a few minutes to record the advances we have seen in 2023.

A steady improvement in foundation model capabilities

Trained on more data and more compute, the release of the GPT4 model in March 2023 broke new ground. It was more creative than ChatGPT, demonstrated superior reasoning capabilities, performed better at the Uniform Bar Exam (90th percentile) and at the Biology Olympiad (99th percentile) and also had lower hallucination rates. It accomplished some amazing tasks - summarization, question answering and even some complex programming tasks. With GPT4 vision launched in September 2023 the model now incorporates image inputs - thus moving away from the language-only capabilities of earlier models - a key frontier in AI research and development.

GPT4 with vision can now glean information from complex images, recognize locations and thus enable a whole new set of use cases from healthcare to scientific analysis. It also enables a world of ambient

digital interfaces - imagine smart glasses that can see what you see by feeding a stream of vision data back to a foundation model. In November 2023, Open AI launched GPT 4 Turbo with 128K context window, which means you can pack 300 pages of a book in a single prompt. And that too at 2X to 3X lower cost of GPT 4. Open AI also launched GPTs that are no code tools to create Gen AI chat apps and also platforms to buy and sell these GPT apps on GPT store, with the aim of becoming the iOS or Android of the Gen AI ecosystem.

The open source movement to build foundation models has accelerated

Open source models are available for commercial use; they can be downloaded and used. This is distinct from the application programming interface (API) model initiated by Open AI. Companies can download and fine tune open source models for specific purposes. In fact, this is a key trigger for specialized industry or even company specific foundation models. They have improved significantly in performance over the last few months.

“

**Creativity is intelligence
having fun.**

”

Albert Einstein

Meta started this trend by open sourcing its Llama series of models, which it uses in all its AI capabilities. The UAE launched Falcon. Today, on the Hugging Face leaderboard - a key source of foundation model information - the top models are open source and it is becoming a viable contender, giving enterprises options.

Significant enhancement in generative art and creativity tools

Generative technologies have proven to be incredible enablers of creative tasks. All of us have experienced with awe the ability of ChatGPT to spin out a wall of text - original sonnets in the style of Shakespeare or very passable school essays created from a few basic

prompts. Midjourney, Stable Diffusion and many other generative art tools have matured significantly and can create amazing photorealistic images based on simple prompts. Runway can create 3D video from text and can zoom in and zoom out and generatively fill images, giving them a sense of motion. AI has been used to generate songs that went to the top of the charts. This trend has also come to enterprise software - Adobe has significantly revamped all of its offerings using its Firefly models that enable creative professionals to use Gen AI in their creative tasks.

Code generation becomes the first killer app for foundation models

Generating code is arguably one of the early killer apps for Gen AI. Code has an inherent structure that foundation models find easy to learn and predict.

GitHub copilot from Microsoft and CodeLlama from Meta has, in the space of the past few months, become standard tools used by developers. While foundation models can improve productivity across a wide range of repetitive programming tasks, they are also being used to document legacy code, refactoring legacy apps and help enforce coding standards.

The first industry tuned models are launched

Large language models (LLMs) will increasingly become domain tuned with industry data to make them more effective to use in the enterprise. The hardest part of this is of course the creation of benchmarks and datasets which are industry specific.

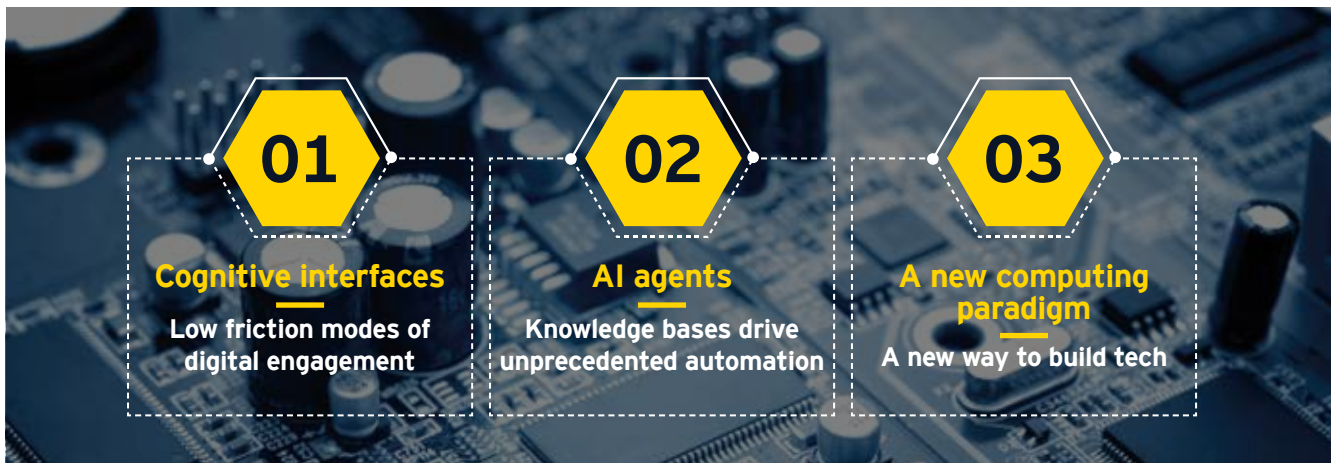
Google released MedPaLM 2 that focuses on the healthcare domain and we have BloombergGPT, a model that understands finance industry-related terminology and concepts as it is trained on vast amounts of financial data.

Big Tech takes AI into the enterprise

AI is no longer a preserve of large companies with terabytes of data and access to supercomputers. Big Tech is incorporating Gen AI into all their offerings and bringing AI enabled capabilities to enterprises of all sizes via their cloud offerings. Google, Microsoft and Amazon have all significantly revamped their cloud services with offerings to enable building AI apps.

Google is aiming to make Bard the personal assistant on your phone, handling tasks like planning a trip, sending texts, searching messages, etc. and is available through your workspace. Pixel phones with AI allow image editing to completely alter entire photos. Meta is adding Gen AI into its ad platform to help brands manage their creative campaigns - from editing images of products to writing copy. Microsoft is adding co-pilots into all its offerings - Bing, Edge, Windows and Azure. This allows image editing in all photo apps, chat with documents, assistants across all assets, and write emails like you. Google has Duet - an LLM-powered chatbot inside Google Cloud. Amazon Bedrock helps companies fine tune their own models. It has invested in Anthropic and made its powerful LLMs, including Claude and Claude 2, available on AWS.





The trouble in paradise

So, are we on the cusp of the launch of a new computing paradigm? Will this be as transformational as the internet? The new electricity?

If this is a revolution, it seems in its very early stages. As with every hype cycle, there is today a whole industry devoted to producing 'AI Snake Oil' - purely demo versions of capabilities, amplified by social media and nowhere close to enterprise-ready and will not survive beyond well-curated demo conditions. To become enterprise ready, native Gen AI capabilities need to be wrapped in layers of business context, integration, security compliance and control.

But there are bigger challenges. If intelligence is defined as the ability to learn new skills, to generalize, reason, plan and understand the world then LLMs are arguably bad at many of these things.

Start with the hallucination issue. Despite all the recent advances, foundation models still make stuff up. They do not give deterministic answers and cannot be relied upon to be predictable. For example, LLMs can fail at elementary math. Worse - they are 'confidently wrong' in many of their answers.

Foundation models are not databases - they do not have memory. They are also very bad at planning and managing long running tasks. This is inherently limiting in the context of enterprise use cases where an item, for example, a customer service request of a sales lead, needs to be taken to closure. Foundation models are also notoriously bad at reasoning, as they do not explicitly 'know' the world we live in, they can be fooled by simple misleading prompts.

Last but not the least, at our current stage of technology maturity, they are very difficult and expensive to train.

Three steps to the future

Given the breakneck speed of evolution in so many dimensions, it is by no means certain what specific path the AI / Gen AI ecosystem will take going forward. But certain trajectories of evolution are starting to emerge. One can discern three key trends:

01 Cognitive interfaces transform the app landscape

The way we engage with technology has evolved rapidly. Decades ago, we moved from punched cards to green screens on IBM mainframes. Then in the PC era we learned to use keyboards and mice on Windows. And now, since the birth of the mobile web, we have grown accustomed to compressing our digital world into small phone screens - six-inch rectangles that we fat finger every moment to navigate our lives.

Gen AI has the potential to liberate us from the tyranny of the mobile phone and the app ecosystem as the primary interface to our digital lives. With ChatGPT, we are already searching the web in new ways. Gone are the Google search terms and six blue links - Bard or Bing can understand what we are looking for and search the web for us. We are starting to use chatbots to search for vacations and to book restaurant tables. We can now upload images and ask for description of the image in 100 words. Many of us are comfortable enough speaking to computers. Welcome to a new low-friction world powered by cognitive agents who act on our behalf with few simple instructions. This will proliferate as cognitive agents lower friction in our digital interactions.

Foundation models trained on text, voice, image and video will fundamentally change the digital interfaces we will use.

02 AI agents transform work

Deep learning can be used to create meaningful representations of data. Francois Chollet, the author of the Keras library, likens this to 'a person trying to uncrumple a paper ball'. The uncrumpling process is a series of geometric transformations through which the network 'learns' to represent this data more meaningfully and in a simpler way. These 'embeddings' can then be stored, searched and analyzed in a knowledge base.

Autonomous agents built on this foundation of knowledge can dramatically change how we work. Imagine you are an insurance company. You create 'embeddings' of your policy documents. These are mathematical representations of your enterprises' business rules and are stored in a knowledge base which is no longer 'unstructured'. Today, these rules and flows are scattered across your enterprise - sitting as Word documents on employee laptops, dispersed across many applications or coded into your apps. Going forward, the knowledge base increasingly takes centerstage in your business architecture, and potentially becomes the single version of 'truth' of your enterprise know how.

On this foundation, one can envision an autonomous underwriting agent which enables the processing of a customer's medical and financial documents, simplifying search and understanding of multiple policies and helping improve real time communication and processing. The agent used leverages Gen AI to summarize customers' financial documents, medical reports and tests, then analyze applicant's medical history, and develop well-structured dashboards to present all the case findings, document summaries, recommendation on adjusted premium for the underwriter to read and make informed decisions.

The democratization of creativity will change the nature of our relationship with work.

Humans are innately creative. We constantly reimagine and shape our world using new tools that extend our physical and cognitive capabilities. We evolve new forms of art and new modes of thought and social organization.

Yet, ironically, many of us struggle with creativity in several aspects of our life. Writing and visual

communication is a challenge. So is synthesizing what we know to create the new. In the context of an enterprise this problem is magnified. How should we eliminate biases that have crept in after decades of an enterprise's existence? How can one enable employees to think outside the box? To collaborate on creative tasks? How to enable this at scale?

Now imagine these tools enabled at scale in an enterprise. A relationship manager gets a well structured talk track for her client based on analyses of client data and recent interactions. Indie artists create their own movies as generative production tools proliferate. Researchers at drug companies use protein sequencing data to create new drugs.

Here hallucination is a feature, not a bug. Generative technologies will unleash innovation at an unprecedented scale.

03 A new general computing paradigm is established

Foundation models are not apps by themselves. They are foundational components of the next generation of application architecture. Gen AI will fundamentally transform our approach to building apps. The front end moves from mobile apps to conversational interfaces. A significant part of the functionality of business rule engines will move to knowledge bases and vector stores. Process logic will be orchestrated via agent frameworks that integrate intelligent front ends with knowledge bases and enterprise applications. Integrations are accomplished via plug in architecture linked to foundation models. As we rethink our data architecture and move to more elegant data fabric- oriented architectures, we will need to integrate this with foundation models. This entire architecture will make use of fit-for-purpose cloud platforms which will increasingly be specialized to industries and functions. And then Gen AI also speeds up software coding by converting natural language instructions to complex code. This helps significant reduction in cost of digital apps and increase in velocity of new build & innovation.

As technology entrepreneurs create and tweak new foundation models - and then make them useful for everyday work and life by developing apps built on them - it will bring about a sea change in the functioning of enterprises. It is also likely to transform old business models while creating new ones. This will be an exciting, if tumultuous, journey as Gen AI becomes an integral part of business. ●





Chapter 2


Rethinking the Enterprise Agenda




Rethinking the Enterprise Agenda

 Organizations will benefit from putting in place **a Gen AI strategy that integrates** with their **overall approach** to customer engagement, digital operations and technology architecture

 The technology is nascent and fast evolving, hence keeping abreast of the change is critical. Proving success through agile experiments and then **scaling to full functional automation** is critical to realizing benefits

 **Key decisions relate to** a choice of LLMs and cloud vendors, integration with digital platforms and enterprise data ecosystems and the right security architecture

 The cost-benefit equation is also very dynamic. All along the journey, it is important to **keep a keen eye on costs** of training, inference and application integration of key components

The break of the AI dawn

Decades ago, enterprises embraced mainframes for centralized computing. This was followed by the rise of personal computers empowering individuals with greater work productivity. The end of the last century saw networking and connected computers taking collaboration to a new level all together with the web becoming a global nexus, transcending geographical boundaries. The 21st century has seen cloud computing reduce infrastructure costs, promoting collaboration with mobile-friendly work environments. Data analytics has empowered businesses to gain insights from big data, driving innovation. Automation, a cornerstone of the digital age, has revolutionized workflows, enhancing efficiency. In the current technology landscape, web, cloud computing, data analytics and automation have taken center stage. Their seamless integration and the relentless march of digital innovation continues to redefine the way we live and work.

Now, as the AI era dawns, specifically Gen AI, businesses are leveraging the technology to unlock unprecedented insights, optimize decision-making processes and usher in a new age of efficiency and innovation.

Gen AI holds the promise of fundamentally reshaping industries. In certain sectors, it is poised to revolutionize not just businesses but their entire operating models. This revolution manifests as an accelerated wave of new product releases, a transformation of the value chain and a fundamental shift in the underlying economic dynamics. In other sectors, Gen AI is set to trigger a new digital transformation wave - frictionless customer

EY CEO Outlook Pulse Survey

In October 2023, EY conducted a survey of 1,200 CEOs from large companies around the world. This CEO Outlook Pulse survey focussed on how they are continuing the journey into an AI-enabled future. The survey provides insights on capital allocation, investment and transformation strategies, as the economy reverts to a model with higher interest rates and inflation, more geopolitical headwinds but fewer economic tailwinds.

CEOs globally recognize the potential of AI, but most are encountering significant challenges in formulating and operationalizing related strategies. The outlook in India mirrors the global perspective.

Key survey highlights



While the vast majority (99%) are planning to invest in Gen AI, the investment landscape is complex. Many CEOs recognize AI's potential to disrupt their business models and are starting to initiate their response.

The survey clearly reflects that when it comes to AI, CEOs find themselves acting with urgency. 7 in 10 (70%) recognize that their organization must act now on Gen AI to avoid giving their competitors a strategic advantage.

At the same time, nearly the same percentage (68%) agree that the uncertainty around Gen AI makes it challenging to move quickly in developing and implementing an AI strategy.



In India, 100% of the respondents plan to make significant investments in Gen AI.

84% of Indian CEOs recognize the urgency to move quickly with Gen AI.

Majority of the Indian CEOs (80%) acknowledge the uncertainty surrounding Gen AI posing a challenge for a swift roll-out of their AI strategy.

experiences, more automation across various functions, empowered workforce with creative tools and data-driven decision-making at the last mile; all enabled by a seamless interaction with enterprise knowledge repositories.

Gen AI can revolutionize entire operating models, lead to a wave of new product releases, transform value chains and thus lead to a fundamental shift in underlying economic dynamics

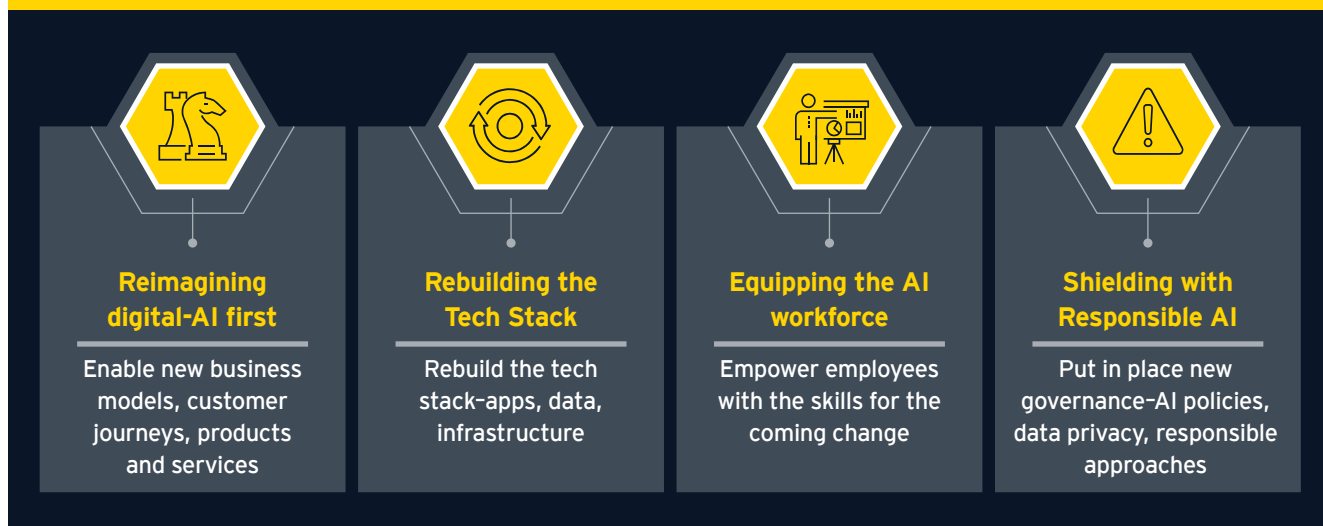
Indian enterprises have embarked on AI transformation journeys, harboring great optimism regarding the potential advantages of Gen AI. Nevertheless, many organizations recognize the need for better preparation to fully reap these benefits.

'Digital Darwinism' compels enterprises to adeptly navigate rapid technological shifts by investing in digital transformation. Survival hinges on the agility of those able to undertake this transformative journey swiftly and effectively.

Several fundamental questions have surfaced about how enterprises can effectively leverage this AI revolution to create value. Which use cases can be prioritized for implementation? Can open-source models deliver performance that meets expectations? How can AI be integrated into the broader context of enterprise digital strategy, data management and cloud adoption? How can risks associated with hallucination and data privacy be addressed? How can an organization navigate the journey of scaled AI adoption?

As we look to address these questions, we lay out a comprehensive enterprise transformation strategy to guide Gen AI adoption.

The agenda for enterprises



The agenda for enterprises

01 Reimagining digital AI-first

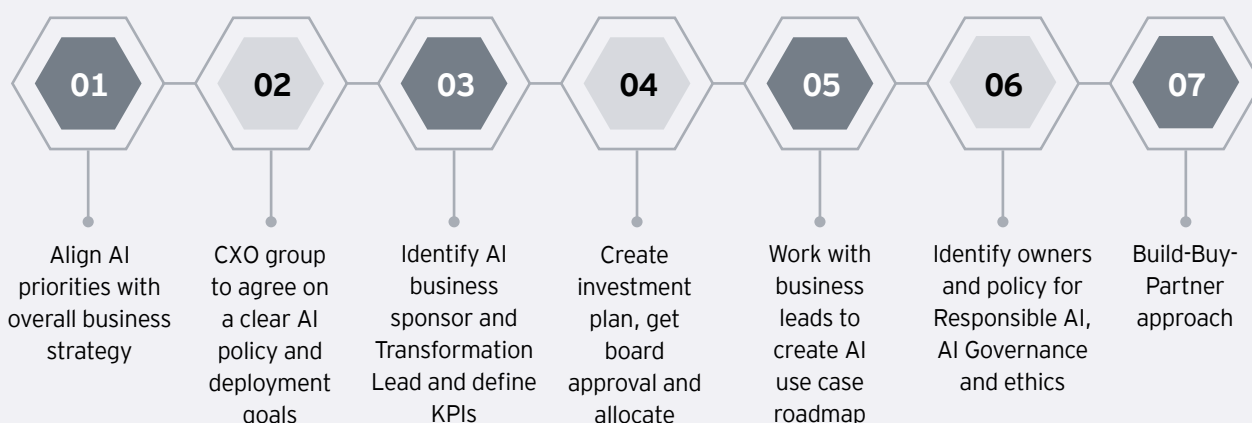
The imminent digital transformation is not merely a matter of implementing new chatbots or allowing the marketing team to experiment with the latest trendy tools. To adopt an AI first approach, a complete reimagination of the digital transformation strategy is needed to harness the potential of AI in conjunction with digital, cloud and automation capabilities.

This will facilitate the emergence of novel business models, enable widespread personalization, accelerate product and service innovation. It will also facilitate incorporation of Gen AI agents (co-pilots and auto-pilots) to facilitate intelligent automation and decision-making across processes and personas.

Unlike several other emerging technologies, Gen AI is now readily accessible on demand (through Application Programming Interfaces (APIs)) making the setting up of the technological foundation a relatively straightforward task. Of paramount importance is the development of a clear AI strategy.

Given the intense scrutiny from the market and board, in-depth planning and successful initiation of the AI first journey is crucial. This involves evangelizing AI with business leaders to create a shared understanding and adopting design thinking to finalize a use case roadmap. It is critical to prioritize these use cases basis potential value, complexity and associated risks. Gen AI champions among business leaders play a key role. Organizations can take a dual approach during initial Gen AI pilots. One is a Bottom-Up approach that empowers Gen AI champions for grassroot innovation

Elements of Enterprise AI Strategy



using nimble federated Gen AI Pods. The other approach, Top Down, entails utilizing a Gen AI Center of Excellence (CoE) to design, build and deploy priority, high-value and complex use cases and also centrally define the technology stack, governance, talent and risk processes.

Following the pilot launch, a clear ROI assessment using A/B testing is essential. Additionally, effective communication of the value created must be disseminated across the enterprise to gain buy-in, build support and drive organizational engagement for broader adoption of AI initiatives. This imparts the impetus for subsequent Gen AI use case iterations while constructing a compelling, self-funding business case for further investments in technology platforms, human resources and processes, thus ensuring the momentum is sustained.

Beyond business use cases, some enterprises are equipping a significant portion of their workforce with Gen AI tools to enhance day-to-day productivity at scale. These tools encompass an enterprise version of Chat GPT, which provides secure access to foundation models (FMs) and LLMs, which form the core of Gen AI, along with enterprise knowledge repositories. These Gen AI agents in the enterprise

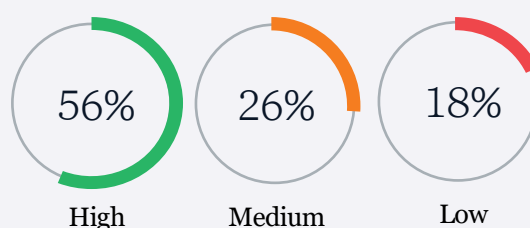
environment prioritize data security and adhere to robust AI risk governance protocols, unlike their consumer-oriented counterparts. Moreover, some enterprises are making Gen AI co-pilots from prominent software firms such as Microsoft and Adobe available to their employees.

EY India's C-suite Gen AI survey

EY India conducted an in-depth Gen AI survey covering ~200 C-suite executives across India. They represent diverse sectors, including Technology; Media and Entertainment; Financial Services; Government; Health, Pharma and Life sciences; Retail; and Manufacturing. Key findings from the survey:

Level of impact

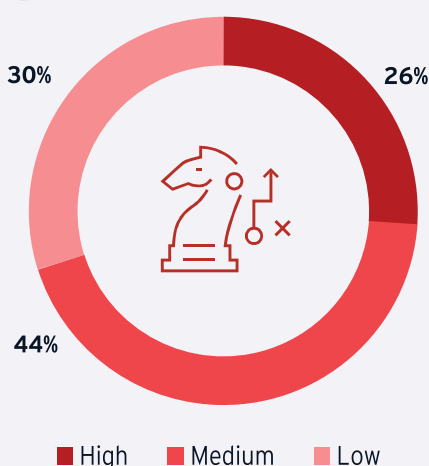
Three in five respondents see Gen AI having a significant impact on their business.



Level of readiness

Three-fourth of the respondents indicate a low to moderate level of readiness in benefiting from Gen AI. More effort is needed to develop specific use cases as well as de-risking to highlight the benefits.

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

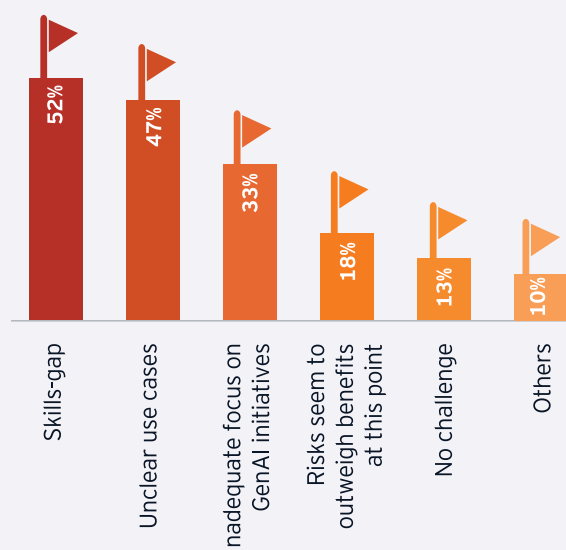


Survey data

Challenges

Skills gap and unclear use cases are two of the biggest challenges organizations face.

What are the challenges to Gen AI adoption in your organization?
(Mark all that apply)

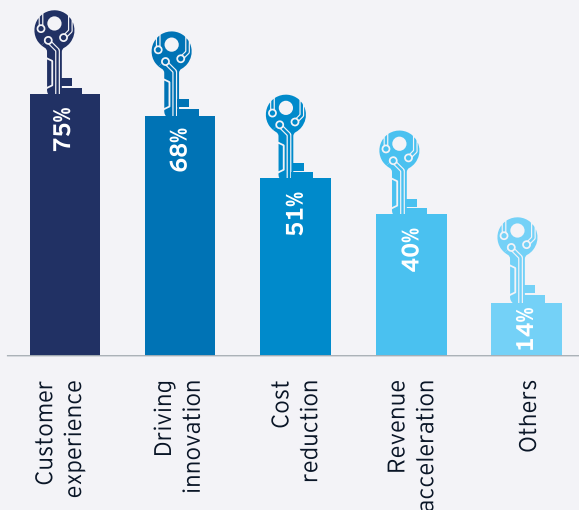


Survey data

Impact area

Customer experience is the single most important facet that Gen AI is expected to impact

What facets of your business would Gen AI impact? (Mark all that apply)



Survey data

Takeaways for enterprises

- Understand the potential of Gen AI across your **organization value chain**
- Identify relevant use cases for early implementation and **create a roadmap for sustainable scale**
- Focus on **functional transformations** aimed at revenue enhancement, productivity increase and agile digital capability release
- Identify the **AI sponsor and owner** for the program. Define where the capability will sit i.e., a Gen AI CoE at the hub in addition to spokes within business teams of early AI champions
- Assess a **combination of a top-down** strategy using enterprise-wide Gen AI tools to enhance productivity and deploy high-value yet complex use cases, along with **a bottoms-up approach** that empowers businesses to experiment and innovate.
- Ensure **accurate value measurement** and **communication** to drive change

02

Rebuilding the Tech Stack

Most organizations find themselves in the early stages of their AI journey. While some are currently engaged in pilots with reported initial successes, they are yet to formalize a comprehensive strategy for the scalable and sustainable delivery of business value. Others are currently in the assessment phase, diligently evaluating risks and identifying the most suitable pilots tailored to their specific circumstances. This is an opportune moment for companies to assess their technology stack and harness the full potential of AI while incorporating cognitive interfaces and knowledge stores in enterprise applications.

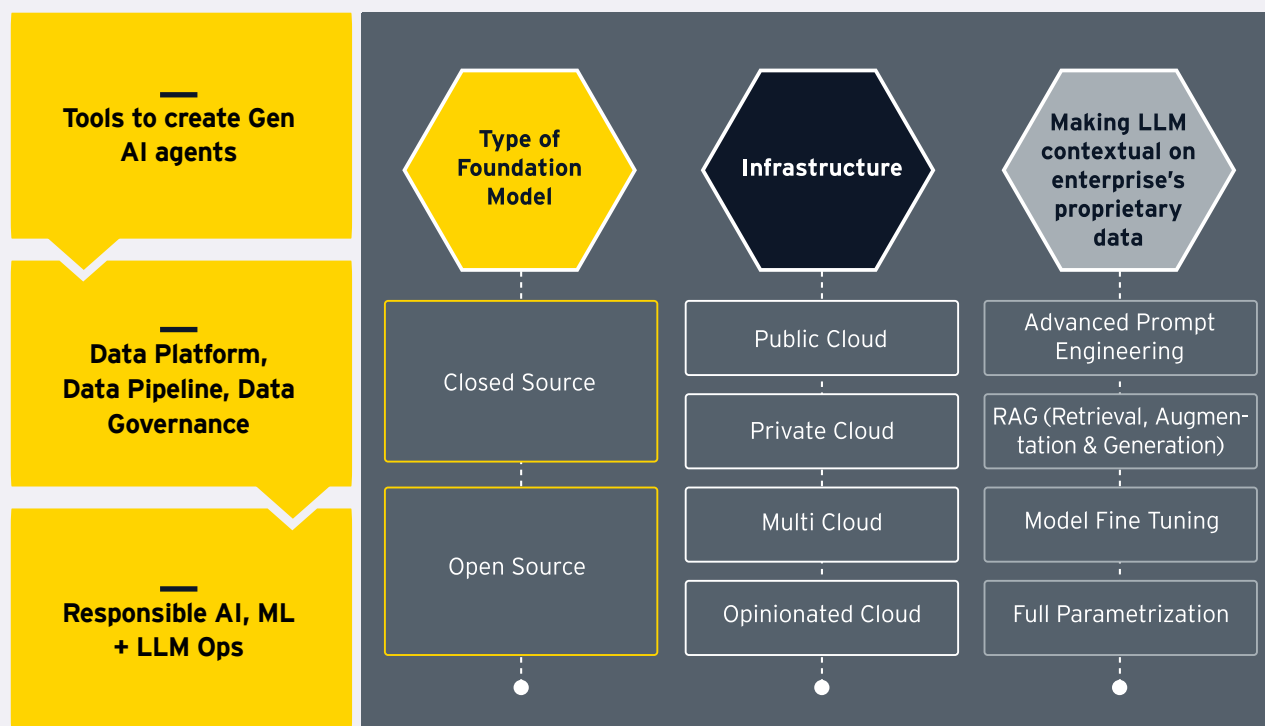
Enterprises need to make strategic decisions on the utilization of data, prioritization of use cases, management of data security and compliance and an in-depth evaluation of cost, ROI and speed to value.

Six key questions for enterprises to empower effective decision-making regarding Gen AI

- 01 Which FMs are best suited to the enterprise's use cases?
- 02 Which cloud platform can the enterprise invest in?
- 03 How can the LLMs be made more contextual with the enterprise's proprietary data?
- 04 What are the tools required to create Gen AI agents for enterprises and embed Gen AI in last mile business decisions?
- 05 What enhancements are needed in an enterprise's data platform, and governance processes to ensure adherence to Responsible AI?
- 06 What are the key considerations impacting Gen AI implementation cost?



Based on business priorities, enterprises need to design the right architecture across different technology components



Enterprises need to evaluate and design the right architecture across different technology components:

1. The first key decision is the choice of models for text, image, code etc. There are primarily two choices of Gen AI models - Closed Source and Open Source. Closed Source models are very large models enabling them to handle diverse, complex queries. These do not require infrastructure and are available immediately for enterprise use like Open AI's GPT 4, GPT 3.5 T, Google's PaLM2 and Anthropic's Claude 2. There is a cost attached for access and usage.

Open Source models need to be managed by enterprises, are more controllable and typically smaller in size like Meta's Llama 2, Falcon, Mistral. They are less versatile and tend to hallucinate (provide incorrect responses) more than the larger Closed Source LLMs.

2. The choice of models is closely intertwined with cloud platform providers. OpenAI, for instance, is available on Azure Enterprise and through OpenAI's own enterprise cloud, while other

Closed Source LLMs are exclusively available on other enterprise Cloud providers. Open Source LLMs, on the other hand, can be accessed through multiple platforms (for example, Hugging Face) and self-hosted on any cloud, including private cloud.

3. To improve accuracy of Gen AI agents; enterprises need to make models contextual to their proprietary data. Various tools and techniques are available for enterprises to explore (Prompt engineering, prompt optimization, Retrieval Augmented Generation (RAG), and Model fine tuning.)

See Annexure A for more details on RAG and fine tuning and also our suggested approach for making LLMs more contextual with enterprise's proprietary data.

4. The fourth critical step revolves around the seamless integration of models with various enterprise applications to drive effective decision-making. This entails the build and deployment of Gen AI agents tailored for high-priority use cases and their workflows.

These require adoption of specialized tools (for example, LangChain and LlamaIndex, Vector Databases like Pinecone, and model farms like Hugging Face). These agents can help automate end-end processes with minimal human intervention. Key differentiator is the intelligence and context of Gen AI to take a decision versus the existing business rules.

Gen AI agents must also be seamlessly integrated with existing enterprise applications, including SFDC, SAP, Adobe, ServiceNow. Enterprises can also assess existing enterprise software providers who are augmenting their platforms with Gen AI capabilities, such as Salesforce's Einstein and Adobe's Firefly for creative content generation. Furthermore, a plethora of specialist SaaS tools for Gen AI applications are now available, offering enterprises options for speed and cost-efficiency, but need to be evaluated for security.

5. The effectiveness of AI is inherently tied to the quality of the data it receives. Therefore, companies must prioritize augmentation of their data platform to cater to Gen AI use cases. This entails developing capabilities and implementing right tools for handling large volumes of unstructured data along with structured

data. The modern data platform needs added capabilities around data governance, data security and some new functionalities for Gen AI models.

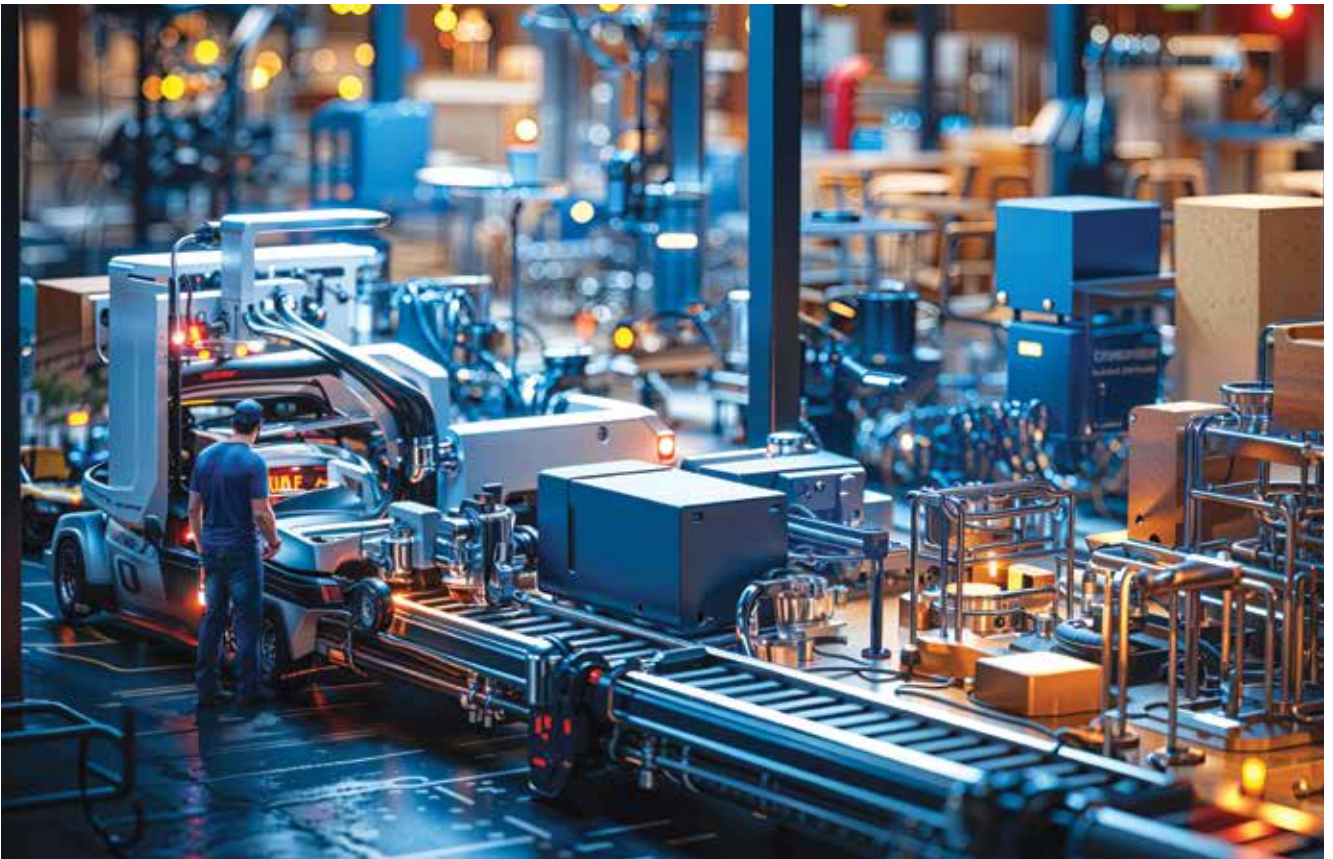
Similarly, AI governance process and tools need to be augmented. LLMOps (LLM Operations) is the primary tool to enable this via model monitoring, traceability etc. See Annexure B for key data considerations for Gen AI.

6. While there are multiple configurations of the Gen AI Tech Stack for enterprises, we see the emergence of two primary archetypes for Gen AI deployment:

- ▶ Domain specific Open Source LLMs, fine-tuned and self-hosted on enterprise's infrastructure
- ▶ Closed Source LLMs, accessed via commercial APIs and with access to enterprise data via RAG

See Annexure C for a detailed comparative analysis.

Open Source LLMs are best leveraged for use cases at scale where the 'token' volumes are high. A 'token' is the basic unit of text that a



model uses to process and generate a response. A typical use case of high token volume is customer service quality analysis which require reviewing all customer call transcripts for call quality and identifying reasons for escalation. Open Source Models are also suited for use cases that need domain specific data like clinical medical research data to answer user queries or ones which need access to sensitive data. Working with these models requires enterprises to plan and build a strong foundation for effective deployment.

Closed Source LLMs are ideal for getting started, pilot and rapid experimentation. These are better suited for deployment of use cases that require versatility and ability to handle complex conversations like financial planning and fraud investigation, with low to medium token volume.

Enterprises must further assess the performance, accuracy, cost, task type, token throughput, model size and processing requirements (GPU - Graphical Processing Unit) when choosing between Open Source and Closed Source LLMs.

Awareness of LLM costs is crucial to striking a balance between accuracy and cost optimization.

In the past six months, Closed Source LLM costs have seen a significant reduction, a trend expected to continue as GPUs becomes more readily available. Notably, costs vary considerably among LLMs, with OpenAI's GPT-4 being 8 to 10 times more expensive than OpenAI's GPT-3.5T and Google PaLM2 as of publishing date.

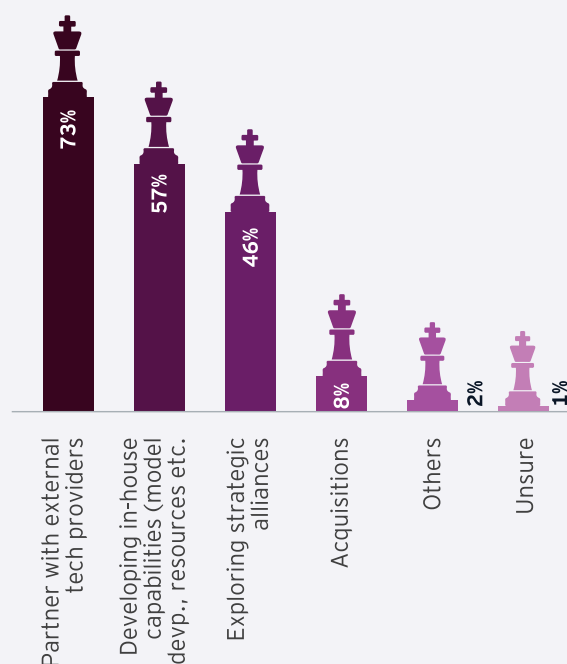
In the case of Open Source LLMs, costs primarily involve GPU usage cost per hour, as these models need to be self-hosted. However, there are additional costs related to managing infrastructure, optimization, and scaling which requires specialized and expensive resources. In both scenarios, there are additional expenses for necessary cloud services and third-party licensed software, contributing to the overall cost considerations.

Closed Source LLMs demonstrate lower costs when the volume of tokens is lower. Notably, the cost dynamics shift favorably for Open Source LLMs when Gen AI deployment scales up at the enterprise level, incorporating multiple use cases such as processing customer service voice calls

Executing Gen AI strategy

The preferred mode for Gen AI implementation is by partnering with external tech providers

How do you envision executing your organization's Gen AI strategy?
(Mark all that apply)



Survey data

or handling massive knowledge corpuses. This is expected to change in the next 12 months as Closed Source LLM costs can reduce significantly.

- Enterprises can consider taking a phased hybrid Gen AI deployment approach, tailored to their requirements.

Phase 1: Start

Enterprises can prioritize use cases and start development of high-value, internal-facing ones using Closed Source LLMs to create Minimum Viable Product (MVP) Gen AI agents and deploying them through A/B testing, evaluating their value, and subsequently scaling them. Enterprises need to be open to partnering with start-ups, consulting firms and academia for specialized skills in this phase.

Simultaneously, enterprises can finalize their enterprise Gen AI strategy, construct the enterprise Gen AI stack, cultivate a talent pool, and institute responsible AI practices.

Phase 2: Learn

Building upon the insights gained from Sprint 1 of use cases, enterprises can look to adapt and launch Sprint 2. This involves scaling existing and internal use cases and developing new ones, including those with customer-facing applications such as customer service or enhanced buying experiences, utilizing Closed Source LLMs.

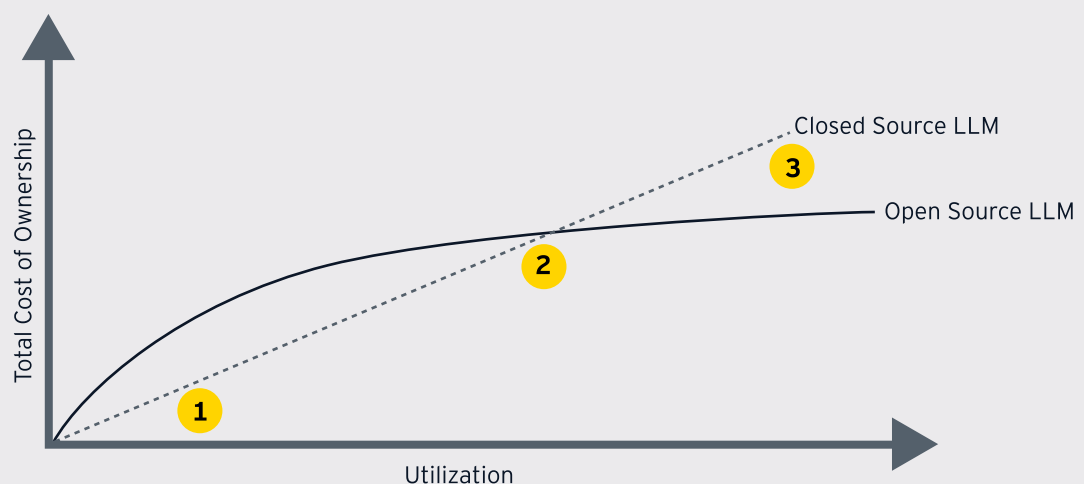
Additionally, enterprises can initiate experiments and construct select use cases that necessitate domain-specific proprietary data on Open Source LLMs. Rigorous assessment of accuracy, complexity, and deployment costs is crucial.

The implementation of the strategy involves constructing the technology platform, cultivating talent and the Gen AI CoE, and deploying responsible AI processes and tools.

Phase 3: Scale

Drawing on the learnings from Phase 2, enterprises can scale use cases to achieve 100% deployment following A/B Testing. This aims to drive both productivity and revenue growth. Rigorous monitoring of ROI for each use case is essential, guiding decisions on Open Source versus Closed Source LLM deployment based on token volumes and model accuracy.

LLM costs: Striking a balance between accuracy and cost optimization



Type	1. Start	2. Learn	3. Scale*
Closed Source LLM via APIs + RAG	Deploy lower risk internal customer use cases 1. Knowledge mgt. - SOPs, Risk, HR 2. Automation for UW, Fraud, HR process	<ul style="list-style-type: none"> Test accuracy and RoI of use cases Scale internal use cases Pilot customer use cases - customer service, buying experience 	<ul style="list-style-type: none"> For high token volume use cases compare cost vs accuracy with Open Source LLM Scale for complex and critical use cases which require versatility Use for new pilot use cases
Fine Tuned Open Source LLM Self Hosted		Pilot use cases which require contextualized responses from domain specific proprietary data	Scaled deployment of both internal and external use cases with high token volume. Fine Tuning LLMOps
Gen AI Cost	Closed Source LLM Cost < Open Source LLM Cost	Closed Source LLM Cost < Open Source LLM Cost but gap will narrow	Open Source LLM < Closed Source LLM

* Reassess as API costs can come down drastically in 12 months

Takeaways for enterprises

- Rethink architecture in the context of **new to firm components** – take key decisions on foundation models, orchestration stacks, plug in architectures
- Finalize **cloud and partner strategy**
- Rethink **digital and data platform roadmaps**

Similarly, enterprises must navigate choices among options such as Midjourney, Stable Diffusion, and DALL·E 3 for image and alternatives like OpenAI Codex, AWS Code Whisperer, and Code Llama for code. Notably, Open AI and Google are introducing Multi-Modal capabilities (ability to interpret images, text, audio and video inputs together) in their models - GPT-4o and Gemini respectively.

Enterprises need to closely monitor the emergence of domain-specific foundation models entering the market to make an informed decision for their organizations.

See Annexure D for a reference architecture for hybrid LLM deployment and Annexure E for best practices for optimizing LLM and Gen AI costs for FinOps.

03 Equipping the AI workforce

While Gen AI holds the promise of industry transformation and heightened efficiency, concerns persist about job displacement and evolving work dynamics. The true impact of Gen AI on employment remains uncertain, emphasizing the imperative for workers to proactively acquire new skills to stay pertinent in the age of Gen AI.

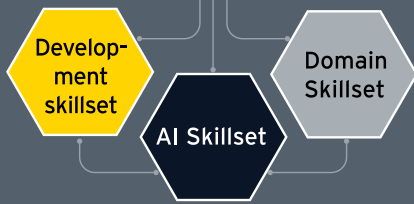
In the near term, businesses grapple with a shortage of individuals possessing AI skills, a challenge that is expected to continue. Despite India's commendable standing in AI skill penetration and talent concentration, the advent of Gen AI amplifies this shortage. The widening gap between the skills demanded by companies and the existing workforce underscores the urgent need for strategic talent acquisition, particularly for the successful initiation and scalable implementation of prioritized use cases.

Key questions to help enterprises formulate a comprehensive talent strategy

- 01 What is the preferred operating model for Gen AI – centralized or federated? What is the role of the central team – create consistent policies, a centralized platform and AI governance or is it also responsible for talent pool and delivery through Gen AI CoE?
- 02 Who from the C-suite can serve as the primary sponsor for driving AI transformation – CEO, CFO, CTO?
- 03 Where will the Gen AI CoE and AI Transformation Lead sit? Should the enterprise hire or identify a Chief AI Officer? Should the CAO report to the Chief Technology Officer, Chief Digital Officer, or the CFO?
- 04 What constitutes the optimal talent mix to achieve scalable Gen AI deployment success – AI engineers, data scientists, data engineers, full stack software engineers, prompt engineers, model testers, responsible AI experts, product managers?
- 05 How can we strategically acquire top Gen AI talent through hiring, training, or strategic partnerships?
- 06 How to motivate and retain top talent? How to design talent assessment and capability development processes for this skill set?
- 07 How to effectively train and upskill existing employees, ensuring a consistent and safe usage of Gen AI?
- 08 What strategies will motivate senior leaders to champion Gen AI adoption, and how can mid-level managers be effectively engaged in the program?
- 09 How can we engage external AI experts while promoting internal collaboration for holistic AI solutions?



Demand for AI Skills



Developers and architects should possess a comprehensive skillset encompassing full stack knowledge



Data scientists with a strong proficiency in Python for NLP pre-processing, chunking, and fine-tuning of generative AI models, as well as model fine-tuning



Sector and domain experts should prioritize enhancing their skills in prompt engineering

Existing resources across departments will continue to be an integral part of building and scaling enterprise AI but their roles will transform



Data Scientist

Keywords —
LLM, Deep Learning, FM, Model Fine Tuning, RAG



Software Developer

Keywords —
SDLC Methodology, Web Development, Software Testing



Solutions Architect

Keywords —
AI, Data, Cloud, Distributed Systems, Software Architecture



Data Engineer

Keywords —
ETL, Security, Database Systems, Data Processing, Cloud



Data Analyst

Keywords —
Statistics, Visualization, Modeling, Reporting



Legal, Risk & Compliance

Keywords —
Regulations: AI, Algorithms, Auto Decision Making, Data Protection



DevOps & Cloud Engineer

Keywords —
Containerization, Orchestration, CI/CD, Cloud, Networking



Business Analyst

Keywords —
Business Consulting, Industry Expertise, Process Expertise



Data Security & Protection

Keywords —
Security, Consumer Privacy, Vulnerability, Malware

Attracting new specialized AI talent is foundational for success



Gen AI Engineer

Keywords —
LLM, Fine Tuning, AI agents, AI apps, AI orchestration, LLMops



LLM Ops

Keywords —
Model monitoring and benchmarking, Prompt Store



Prompt Engineer

Keywords —
Text Inputs, Syntax, User Design, Prompt Quality



Responsible AI

Keywords —
AI Governance, AI Risk Assessment



Data Annotator, Model Tester

Keywords —
Labeling, Fine Tuning Dataset, Model Performance, Sentiment, Semantic, Reporting

Successful Gen AI implementation necessitates a diverse skill set embedded in either a Gen AI CoE or in Gen AI Pods. Building, deploying and governing Gen AI agents and apps demands a specialized skill set combining AI engineering, data science, and expertise in FMs, LLMs, AI orchestration, prompt engineering, RAG, fine-tuning, and model deployment. To address the scarcity of such experts,

enterprises may need to recruit individuals with NLP and traditional AI/ML backgrounds, providing them with targeted training to rapidly develop Gen AI-specific skills. This approach extends to cultivating talent within existing AI and analytics teams, leveraging their foundational expertise and fostering continuous learning.

Takeaways for enterprises

- Re-envision the **jobs of the future** in your organization
- Enable all employees with **AI assistants**
- Put in place a **skilling agenda** - from the boardroom downwards

04

Shielding with Responsible AI

The promises that Gen AI holds are accompanied by inherent risks. The intrinsic complexity of Gen AI presents new challenges for model deployment and validation. While there is high awareness regarding data and most model-related parameters for 'Classical' ML, that is not the case for Gen AI. See graph: Comparing Classical ML and AI products.

Enterprises must be vigilant in managing Gen AI-related risks to avoid any reputation or financial loss. C-suite executives and leaders need to proactively comprehend and integrate processes for risk mitigation and governance. At present, organizations see data privacy as the most important risk of Gen AI.

To address the concerns posed by Gen AI, enterprises can focus on addressing the following key risks:

- 1. Trust and performance risk:** Hallucinations in FMs and LLMs lead to erroneous responses and erode user trust. Enhancing performance involves maturation of LLMs, integration of enterprise data, and employing techniques such as Data Grounding, Dynamic Embeddings, and Reinforcement Learning from Human Feedback (RLHF). Transparency in Gen AI responses, such as sharing data sources and clarifying AI-generated content, fosters trust.
- 2. Bias and toxicity risk:** Bias in training data and models can lead to unfair outcomes and discrimination. Monitoring, identifying and removing biases through Bias Auditing and Data Fairness tools are crucial. Implementing measures such as meta prompts and content filtering, must be implemented to manage toxic prompts and responses.
- 3. Security and privacy risk:** Managing the risk of leaking proprietary and sensitive data to LLMs is a priority. Employee training and technical

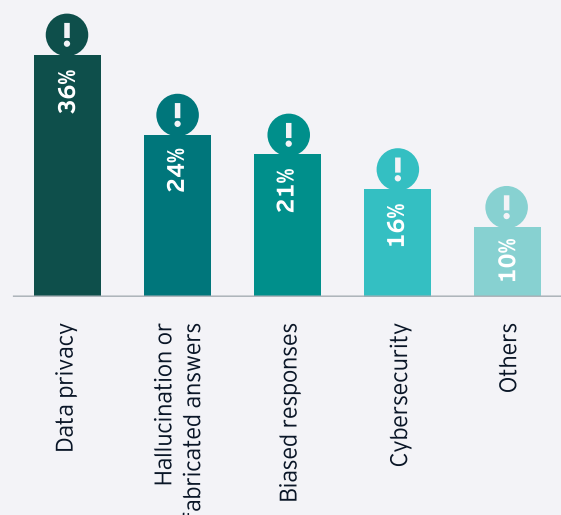
guardrails are essential to prevent the entry of proprietary data into non-enterprise Gen AI tools. Enhanced cybersecurity measures are necessary to counter external threats such as prompt injections and model theft. Compliance with GDPR (General Data Protection Regulation) and local data regulations is imperative which includes getting user permissions to use their prompt data, transparency on responses generated by AI and data sources used.

- 4. Regulatory, compliance and copyright risks:** Enterprises must stay informed about AI governance, ethics policies, and regulatory provisions. Compliance with evolving regulatory frameworks is essential. Awareness of copyright risks, especially in image and code domains, is crucial. Understanding existing lawsuits against FM companies is also necessary.
- 5. Ethical risks:** Enterprises must navigate ethical concerns related to job loss, technology misuse (for example, deep fakes), risks of super intelligence, and sustainability challenges. Establishing a clear AI governance framework, AI Ethics Board, and responsible AI practices is crucial for addressing ethical queries from internal and external stakeholders.

Adoption risk

At present, organizations see data privacy as the single most important risk of Gen AI

What risk worries you the most in Gen AI adoption? (Mark all that apply)



Survey data

Comparing classical ML and Gen AI products

Area	"Classical" ML Products	Knowledge of Risk	Gen AI Products	Knowledge of Risk
Logic	Explicitly coded		Driven by supplied data, prompts and queries	
Input	Requires large amount of data which are usually generated/made available in-house for training		Models are pre-trained using various sources. This data remains unavailable to public. Although domain specific data used for fine-tuning.	
Output	Input: Dataset + Hyperparameters Output: Stochastic, structured		Input: Prompts + Hyperparameters Output: Stochastic, unstructured	
Model management	Models are developed in house from scratch		Models are pre-trained by third-party organizations. These pre-trained models are known as Foundational Models. Only fine tuning of the same is done on need basis.	
Evaluation	Clear set metrics based upon the type of model. Evaluation on held-out from the input training data set		Evaluation is a challenge due to non-availability of original training data. Curated dataset and human feedback is required.	
Testing	Continuous testing/improving, real-time monitoring is desired		Real-time monitoring is essential	
Access to data	Access to pre-defined set of variables		Access to unstructured data that is difficult to control	

Known - Familiar - Unknown

To implement responsible AI effectively, it is imperative to establish a robust AI governance framework featuring explicit cross-functional ownership and accountability. This entails the formation of an AI Ethics Board dedicated to overseeing AI model risk management. Additionally, comprehensive training programs must be instituted for all employees, emphasizing responsible AI practices. The success of this deployment hinges on adept change management strategies to ensure a seamless and ethical integration of AI technologies within the organizational landscape. See Annexure F for EY's approach to Responsible AI.

Takeaways for enterprises

- Redesign **AI policies and design standards**
- Implement a new **risk and governance** framework
- Clearly define **ownership of risk mitigation** and controls including model risk management
- Put in place a new **data privacy and security architecture**
- Creation of **AI Ethics Board** and adoption of **AI Ethics Framework**

Planning ahead

Developing a strategic blueprint for Gen AI involves navigating a challenging landscape filled with multiple choices and continually evolving criteria. To facilitate and support enterprises on this challenging journey, we have developed a Gen AI strategy reckoner.

This strategic guide aims not only to distil the complexities inherent in planning for Gen AI but also to provide a nuanced understanding of the myriad factors that influence decision-making in this rapidly evolving technology. ●

Gen AI strategy reckoner

Archetype	Digital leadership	Advanced digital	Emerging digital
Characteristics	<ul style="list-style-type: none"> Digital First - mobile, web, social Cloud native Data and AI embedded in all business processes and decisions 	<ul style="list-style-type: none"> Significant investments in automation, digital apps for customer, employees and ecosystem Data and analytics Platform BI heavy, AI deployment in select BUs and functions 	<ul style="list-style-type: none"> Processes are not fully automated Digital apps not adopted across industry; few leading the way Data in silos; prioritization of data platforms in early stages Very limited application of analytics and AI
Typical industry	<ul style="list-style-type: none"> Ecommerce Fintech Media 	<ul style="list-style-type: none"> FS Telecom Retail Consumer Goods Auto Pharma 	<ul style="list-style-type: none"> Healthcare Industrial products Resources, energy, oil and gas Utilities
Posture for Gen AI	Aggressive: Embed across the value chain including end customer	Moderate: Prioritize use cases on value, complexity and risk. Initiate multiple pilots, mostly Internal Facing Chat (chat on policies, SOPs, contracts), Customer service, knowledge management, intelligent automation	Wait and watch / Experimentation
Roadmap	Enterprise-wide mobilization	Functional transformation of core processes	POC and scale
Organization model	Federated with BUs and functions building their own Gen AI Pods. Gen AI CoE to define overall AI governance, technology choices and central talent pool	AI CoE extended to cater to Gen AI to build and deploy Gen AI pilots working with business champions. CoE also mandated to drive strategy, tech and talent	Part of data and data science team
Tech Stack	Open Source LLM Fine Tuned + Closed Source LLM via API	Closed Source LLM now. To consider Open Source LLM in mid to long term as token volume grows	Closed Source LLM


An aerial, high-angle view of a large, modern industrial manufacturing facility. The floor is a polished, reflective blue-grey. Numerous orange robotic arms are positioned at various workstations. Several workers in orange safety vests and hard hats are visible, moving through the aisles. In the foreground, a yellow forklift is parked. To the right, a blue machine with 'OALAN' and 'CROWN' branding is visible. The background shows more complex machinery and structural elements of the factory.


Chapter 3


The Economic Opportunity of Gen AI in India




The Economic Opportunity of Gen AI in India

 India has the potential to add **US\$359 billion to US\$438 billion** to its GDP on account of Gen AI adoption in 2029-30 over and above its baseline estimates

 This represents an additional **5.9% to 7.2%** of GDP in 2029-30

 Over a period of seven years (2023-24 to 2029-30), Gen AI's contribution would translate to **US\$1.2 trillion to US\$1.5 trillion** cumulated GDP impact

 Achieving this potential would provide the Indian economy with an additional CAGR of **0.9% to 1.1%**

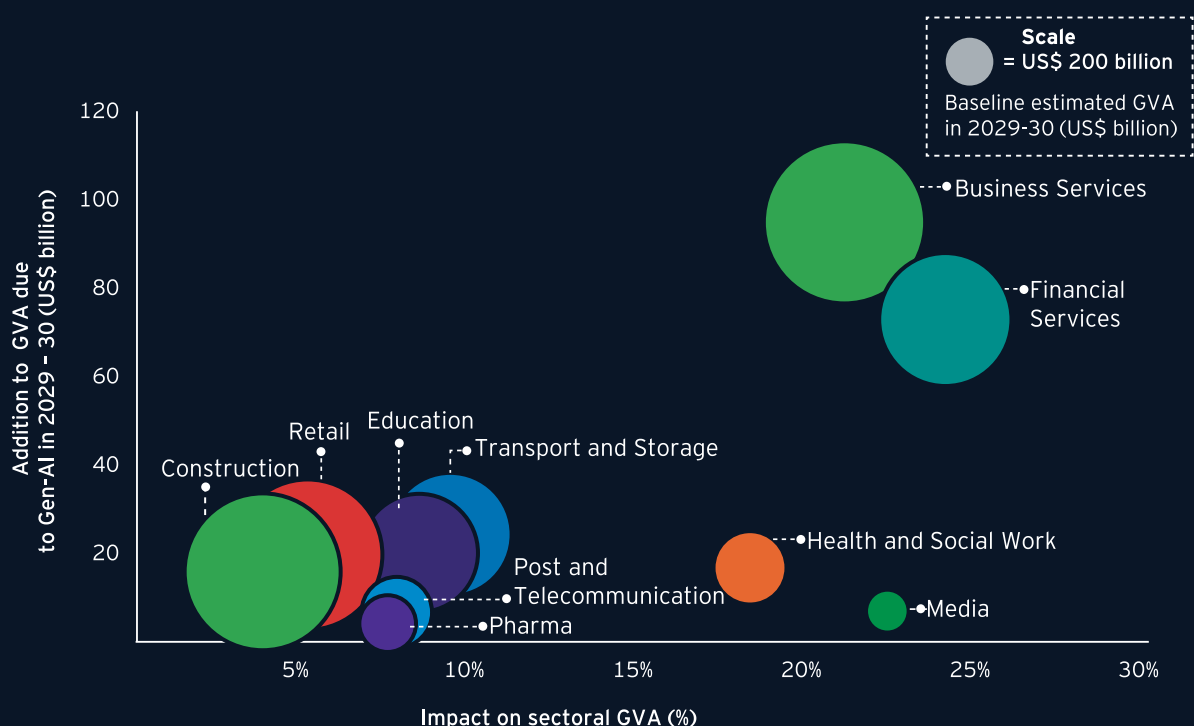
India is in a strong position to harness the potential of Gen AI

India's economy is recognized as the fastest-growing among major nations by global institutions like the World Bank and the IMF¹. Simultaneously, advanced economies grapple with issues such as economic slowdowns, supply shortages, high inflation, and aging populations. Currently, the fifth-largest economy in market exchange rate terms, India is projected to surpass Germany and Japan to become the third-largest by 2027, according to the IMF. In purchasing power parity (PPP) terms, India already holds the third spot.

A report from EY, titled 'India@100: Realizing the Potential of a US\$26 Trillion Economy,' forecasts that by sustaining a real GDP growth of 6-6.4% during the 'Amrit Kaal', India could become a US\$26 trillion economy by 2047-48 in market exchange terms, attaining a per capita income of US\$15,000, putting it in the ranks of developed countries. Given the immense capability of Gen AI with respect to its productivity and efficiency enhancing effects, its adoption has the potential to accelerate India's growth trajectory, enabling it to achieve these milestones sooner. This necessitates increased investment in Gen AI, education, and upskilling to fully capitalize on the demographic dividend.

¹ World Economic Outlook, April 2023 and World Bank Global Economic prospects, June 2023

Sector-wise impact of Gen AI on GVA (2029-30)



Impact on sectoral GVA (X-axis) and additional GVA due to Gen AI (Y-axis) depicted here reflects the mid-point of the range-bound impact for each sector

A large part of the value added will be from the service industries

Our methodology for assessing Gen AI's economic impact on India combines a macro framework with sector-specific insights across 27 sectors identified based on the KLEMS database (RBI). These insights were drawn from EY's sector leaders, based on their expertise and client interactions regarding the Gen AI's efficiency effects in terms of cost reduction and output expansion over the period from 2023-24 to 2029-30. For more details on the methodological framework, refer Annexure G.

The study indicates that India could experience a substantial boost in its GDP, with a potential addition of US\$359 billion to US\$438 billion in the fiscal year 2029-30, reflecting a 5.9% to 7.2% increase. Over seven years (2023-24 to 2029-30), the cumulative impact on GDP may range from US\$1.2 trillion to US\$1.5 trillion, contributing an additional 0.9% to 1.1% in annual CAGR.

While Gen AI's positive impact is anticipated across all sectors, its level of influence in each segment will depend on factors like feasibility, adoption rates, the organized sector's share, and its contribution to India's economic activity. Approximately 69% of the overall impact is expected to derive from business services (including IT, legal, consulting, outsourcing, rental, etc.) financial services, transportation and logistics, education, retail trade, and healthcare. The expected impact encompasses improvements in employee productivity, enhanced operational efficiency, and personalized customer experiences. These sectors, having rapidly embraced digitalization, are well-positioned to capitalize on the benefits of Gen AI. The IT sector also stands to gain significantly from the development of Gen AI platforms and tools both through productivity gains and through more revenues from its clients.


To realize the full potential of Gen AI, we need a proactive regulatory stance. However, much will depend on the policy actions taken to ensure safety of citizens (See Chapter 5: A Gen AI Policy Agenda for India).

Methodology employed to compute macro-economic impact

- ▶ Gen AI's economic impact has been estimated by utilizing a macro framework in the Indian context, i.e., using the current sectoral share in the overall economy and input and output ratios for industry segments.
- ▶ We have used a bottom-up approach wherein the additional gross value added (GVA) in each sector on account of Gen AI adoption is estimated. This is then aggregated to arrive at the economy wide additional GVA. Finally, by adding suitably estimated net indirect taxes (indirect taxes minus subsidies), we arrive at the additional GDP attributable to Gen AI.
- ▶ The business-as-usual case (which does not take into account the impact of Gen AI) is based on IMF projections for growth and exchange rates.
- ▶ The assessment of the economic impact of Gen AI is range bound instead of a point estimate. In an optimistic scenario (broad-based adoption), the impact may be closer to the upper end of the range. But if the adoption rates are less than envisaged, the lower limit may materialize.
- ▶ While estimating the impact of Gen AI, relative shares of organized and unorganized sectors for each industry segment have also been taken into consideration.

See Annexure G for technical analysis

Gen AI's boost to sectors

Sector	Impact on sectoral GVA	Addition to GVA due to Gen AI in 2029-30 (US\$ BN)	
 Business Services*	19%-23%	85-104	
 Financial Services	22%-26%	66-80	
 Transport and Storage	8%-10%	22-27	
 Education	8%-9%	18-22	
 Retail Trade	5%-6%	18-22	
 Health and Social Work	16%-20%	15-18	
 Construction	3%-4%	14-17	
 Media	20%-24%	6-8	
 Post and Telecommunication	7%-8%	6-7	
 Pharma	7%-8%	4-5	

*including IT, legal, consulting, outsourcing, rental, etc

Levers for realizing the potential of Gen AI

- ▶ Give productivity boost as large documents can be summarized within seconds, saving many working hours.
 - ▶ Can generate new documents from scratch (product manuals, proposals, etc.); write policies when instructed appropriately
 - ▶ Automates cumbersome data capture processes like KYC forms through an interactive conversation.
 - ▶ Aid in creating quick insights from large swathes of data like summarizing sales trends and customer segments by looking at multiple dashboards and transaction databases.
 - ▶ Gen AI bots can significantly automate customer care interface.
 - ▶ These levers will lead to increase in employee productivity as well as significant cost reduction.
- ▶ Gen AI-driven bots can act as conversational underwriting engines and help in real-time detection of frauds by keeping an eye on transaction data.
 - ▶ Can help guide financial services firms in risk management and portfolio optimization when risk policies are fed to it.
 - ▶ Can also be used in customer-facing interfaces to optimize lead conversions.
- ▶ Can aid in optimizing transport routes to save fuel and time and help utilize vehicular resources optimally by orchestrating intelligent fleet management.
 - ▶ Can be used as an intelligent scheduling and planning assistant that recommends the best use of transport vehicles matching them with appropriate tasks ordered by priority.
 - ▶ For example, e-commerce delivery fleets or middle mile logistics truck management using intelligent Gen AI bot.
- ▶ Can be a patient and hyper-personalized tutor who understands needs of every student.
 - ▶ Can explain concepts and design a tailored learning path based on student's ability, interest and efforts.
 - ▶ Help generate best-in-class curriculum content and aid educators in building teaching plans.
- ▶ Leverage Gen AI to create hyper-personalized advertisements, achieve targeted marketing through relevant content.
 - ▶ Write product summaries, product review insights and give personalized recommendations.
 - ▶ Specialized bots can also facilitate returns and warranty handling.
- ▶ Ease patient health record capture and management.
 - ▶ Can deliver personalized treatment plans.
 - ▶ Aid in diagnosing diseases by learning from millions of previous patient prognosis track records.
- ▶ Help in generating project reports, give multiple design options.
 - ▶ Help in project management-scheduling and optimization of supply chain and visualizing material delivery schedules.
- ▶ Aid in areas such as generating creative content (text, images, video, sounds).
 - ▶ Can design personalized content by catering to interests of an individual.
 - ▶ Automate tasks like news writing, carry out programmatic advertising and give personalized content recommendations/feeds.
- ▶ Gains are anticipated to stem largely from the heightened requisites for internet services and bandwidth, catering to the myriad offerings underpinned by Gen AI.
 - ▶ Can help monitor and optimize network operations and recommend predictive maintenance of base stations.
- ▶ Numerous applications explored to expedite drug development, deliver highly targeted therapies, streamline supply and demand planning, and augment operational efficiency.
 - ▶ Can also be used to optimize clinical trials of new drugs.

A vibrant peacock with its tail feathers fanned out, standing in a field of purple flowers. The background features a dense forest with green foliage and a tree with purple blossoms under a soft, hazy sky.

Chapter 4

A Gen AI Policy Agenda for India



A Gen AI Policy Agenda for India



In developing AI regulations, many countries are attempting to **balance innovation and risks**



Indian policy emphasis at present is on **collaborative effort by stakeholders**, with the government playing a central role



As subsequent measures bring clarity to the regulatory framework, the government can support innovation by facilitating interventions to improve **access to data, chips, talent, computing resources etc.**



Gen AI algorithms can be used to develop solutions that can be **deployed as Public Goods**

AI, especially Gen AI, has been attracting the attention of policymakers globally at the highest levels. It is seen as a technology that will drive the next level of scientific discovery and economic growth but carries risks that are yet to be fully understood. While all countries emphasize that AI regulation must strike a balance between fostering innovation while managing the risks, their approach and emphasis differs widely and has been evolving with time. Some countries are putting greater focus on promotion and development, while others on mitigating the risks from the implementation of the technology. The role of the government in developing AI algorithms also differs. Annexure H provides an overview of the approach taken by some countries.

For the purposes of this chapter, 'Artificial Intelligence systems' is defined to broadly comprise of a Data Component, the Software/Algorithm Component, the Hardware/Compute Platform Component and Integration/Real World Applications¹. Gen AI has been considered as a sub-set of AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video, computer code, 3D models, etc.

This chapter draws upon the global experience in promoting and regulating AI and the understanding of the Indian policy and business landscape, to provide a few recommendations.

¹ Building Blocks for Artificial Intelligence and Autonomy (publishing.service.gov.uk)



Key learnings from various countries and their policy approach towards Gen AI

United Kingdom (UK)

- ▶ Proposed pro-innovation regulatory outlook to be developed for Gen AI balancing risks and building public trust
- ▶ To set up regulatory sandboxes to overcome regulatory barriers and speed up product launches in Gen AI
- ▶ AI regulatory sandbox will prioritize sectors with substantial AI investment, strong industry demand, and a need for improved collaboration between regulators

United States of America (USA)

- ▶ Home to leading global enterprises that have led the development of foundational LLMs
- ▶ Success driven by the ecosystem with:
 - ▶ Decades of systematic investments in cyber infrastructure and research
 - ▶ Access to large and growing amounts of data (collected over many years) and high computational power
 - ▶ Highly specialized education and training programs together with collaboration between academic researchers and the private sector
- ▶ Current focus to maintain global leadership
- ▶ Leading AI companies have pledged to observe voluntary safeguards
- ▶ Issued an “Executive Order on Safe, Secure and Trustworthy Artificial Intelligence”
- ▶ Various US government agencies to formulate wide range of time-bound interventions relating to: safety, security, transparency, citizen/ consumer protections, effective government usage, innovation, competition, global leadership, etc.

United Arab Emirates (UAE)

- ▶ Launched a Gen AI guide to address the opportunities and challenges
- ▶ Funded the development of indigenous LLM called FALCON which was developed using rigorously audited data to overcome bias
- ▶ FALCON is expected to help Emirati enterprises drive efficiency, leveraging Gen AI tools for applications relating to language translations, virtual assistants etc.
- ▶ Launched AI71 - the commercial arm of the country's Technology Innovation Institute to market FALCON and other such AI related solutions

European Union (EU)

- ▶ Approved regulations mandate that AI systems launched for public must meet a set of risk-management, transparency, documentation, oversight, and quality requirements
- ▶ Proposed transparency requirements include publishing a summary of copyrighted material used in training and data integrity assessments to reduce the possibility of bias
- ▶ Protections and some exemptions to open-source community
- ▶ Promoting innovation through regulatory sandboxes and open sharing of data

China

- ▶ China has consistently been the second largest recipient of venture capital in AI behind only the USA, owing to strong policy signalling by Chinese government
- ▶ Large amount of data that is generated is readily accessible to the government and enterprises
- ▶ Investments in R&D and in development of quality human resource
- ▶ Home to internet giants that are investing in algorithmic innovation, chip development, and language data sets
- ▶ Proposed regulatory framework on Gen AI requires that:
- ▶ Algorithms should comply with existing regulatory framework and respect intellectual property rights
- ▶ Service providers must ensure the data accuracy, objectivity, and diversity, and avoid generating discriminatory content

While each country is looking at safeguards to address risks, many of the risks are global in nature. It is expected that as foundational models can perform a variety of tasks, the cost of access to these models will reduce as they become more widely accessible globally. Risks may arise from either intentional or unintentional misuse and could potentially arise anywhere. The recent Bletchley declaration², which 28 countries (including India, the US, the UK, Israel, and China) and the EU have signed, highlights the intention of countries to cooperate to both harness the benefits and to address the risks. This is therefore significant as it is possible that countries may look at common standards and regulatory principles in the future.

² The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 - GOV.UK (www.gov.uk)

AI and the Indian context: Government initiatives laying foundation of growth of AI

There have been several Government initiatives that have laid out the foundation for approach to promote and develop AI in India. Some of the key initiatives are as follows:

The National Strategy for AI (2018)

NITI Aayog released the National Strategy for Artificial Intelligence³ (NSAI 2018) which highlighted an approach to support “AI4ALL”. It emphasized a concerted collaborative effort by relevant stakeholders, with the government playing a leading role. A key element of the NSAI 2018 was the laying down of the sectoral priorities for AI in India, which include the use of AI in Healthcare (for diagnostics and personalised treatment); Agriculture (for demand prediction), Education (for improving access and quality), Smart cities and infrastructure (for enhancing quality of life) and Smart Mobility, which includes the use of AI in Transport and logistics.

Suggested in this strategy paper, AIRAWAT (AI Research, Analytics and Knowledge Assimilation platform) was also launched. This platform guides the research and development of AI and other emerging technologies in India.

The Principles of Responsible AI

NITI Aayog released the Principles of Responsible AI, which were finalized in 2021⁴. These principles were designed to provide a framework that would serve as an enabling environment for promoting a responsible AI ecosystem in India. These include the Principles of Safety and Reliability, Equality, Inclusivity and Non-discrimination, Privacy and Security, Transparency, Accountability, and Protection and Reinforcement of Positive Human Values.

National AI Mission under PM-STIAC

The Prime Minister’s Science, Technology and Innovation Advisory Council (PM-STIAC) launched the

National AI Mission in 2022 with a strong R&D focus. It aims to bring together academia and industry on developing core AI research capability at the national level and also encourage international collaborations.

- ▶ India’s National Artificial Intelligence Portal ‘INDIAai’ was also launched as a one-stop digital platform for AI-related developments in India⁵. This is a common infrastructure for developers to share tools, data and resources.

India AI 2023 Report by MEITY

The Vision of India AI Report: IndiaAI Mission⁶ ensures a precise and cohesive strategy to bridge the gaps in the existing AI ecosystem with regards to compute infrastructure, AI financing, research and innovation, skilling, and building institutional capacity for datasets. MEITY has identified the following seven AI Pillars in this report, which pertain to the focus areas of the government’s AI strategy:

- ▶ IndiaAI Centres of Excellence
- ▶ India Dataset Platform (IDP)
- ▶ Institutional capacity and design of National Data Management Office (NDMO)
- ▶ IndiaAI future design
- ▶ IndiaAI future skills
- ▶ IndiaAI future labs compute
- ▶ Semicon IndiaAI chipsets

MEITY recommends a “Design Linked Incentive (DLI) Scheme” to provide financial incentives for domestic companies and start-ups for development and deployment of AI related chips.

Some of the challenges to realize the full potential of AI as identified in the NSAI (in 2018) were:

- ▶ Low intensity of AI research
- ▶ Core research in fundamental technologies
- ▶ Transforming core research into market applications
- ▶ Inadequate availability of AI expertise, workforce and skilling opportunities
- ▶ High resource cost and low awareness of adopting AI in business processes

³ National Strategy for Artificial Intelligence (niti.gov.in)

⁴ Principles of Responsible AI, NITI Aayog

⁵ National AI Platform: <https://indiaai.gov.in/about-us>

⁶ MEITY: <https://www.meity.gov.in/content/indiaai-2023-expert-group-report-%E2%80%93-first-editionthe-ministry-electronics-and-information>

- Unclear privacy, security, and ethical regulations
- Unattractive intellectual property regime to incentivize research and AI

The relevance of the challenges remain despite all the foundational efforts made by the government. India's policy agenda pertaining to AI and Gen AI will have to account for these challenges, to effectively promote Gen AI adoption and to create an enabling regulatory environment.

The Digital Personal Data Protection Act (August 2023)

The Digital Personal Data Protection Act (DPDPA), August 2023⁷ provides the guidelines for processing digital personal data collected online:

- Processing has been defined as an automated operation or set of operations performed on digital personal data and it includes collection, storage, use, and sharing.
- The Act mandates the developers/data fiduciary to provide notices to the individuals if the personal data collected is used for the purpose of processing. Further, the private users of personal data may cease to retail such information once legal or business purposes are met.
- The state, however, is exempted for purposes such as archiving, research or statistical purposes.

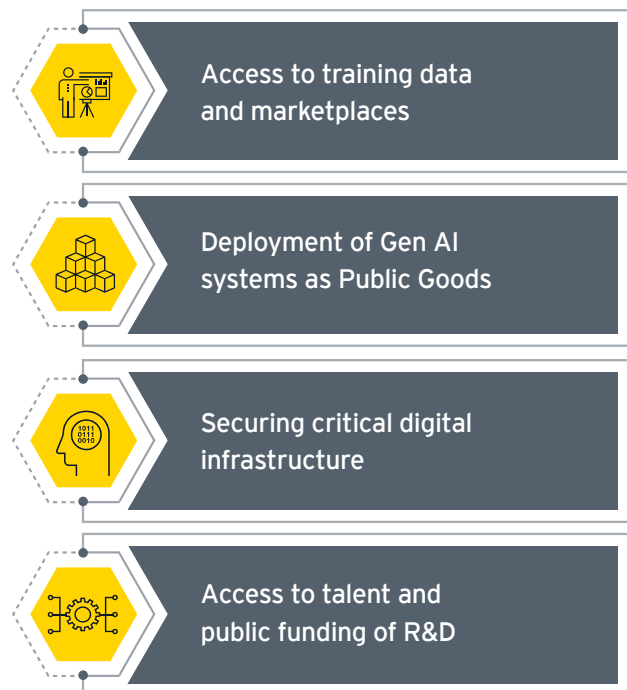
One implication of the DPDPA is that AI platforms would need to take consent to use personal data for use in LLMs. However, the government would also be making detailed rules under DPDPA and the full impact of DPDPA on Gen AI will become clearer once the rules are also framed and notified.

The government is also in the process of finalizing the Digital India Bill, which could impact the development of Gen AI in India. The enactment of this law is expected to facilitate AI development, including Gen AI as it aims to 'safeguard' innovation in AI and other emerging technologies⁸.

Recommendations: Enabling India

The following recommendations are centered around having a 'light touch' regulatory approach with a strong focus on promoting Gen AI systems as a Public Good. The recommendations will facilitate innovation and help develop a conducive regulatory environment for Gen AI in India:

Promoting AI in India



Access to Training Data and Marketplaces

Access to data is key for the development of AI systems. Government support would be needed to ensure that researchers, enterprises, and start-ups have access to structured and unstructured datasets. The support for the creation of data marketplaces would make it easier for developers to access both open-source training datasets as well as licensed private datasets.

⁷ National Strategy for Artificial Intelligence, NITI Aayog, 2018: National Strategy for Artificial Intelligence (niti.gov.in)

⁸ Principles of Responsible AI, NITI Aayog

- ▶ **Expedite the development of indigenous training datasets (especially for local Indian languages):** The government may invest in creation of structured and unstructured datasets (documents, media, etc.) which are made open to the public and contribute to Gen AI development. Setting up new data pipelines will help to capture digitization of government data/documents (especially in Indian languages) and open up existing structured non-personal or anonymized government datasets for wider consumption. Standards may appropriately be notified to facilitate integration of crowd-sourced data for faster creation of training datasets.
- ▶ **Public data commons and marketplaces:** Public data commons and marketplaces, along the lines of existing government efforts such as the Indian Urban Data Exchange (IUDX) Platform (under the Ministry of Housing and Urban Affairs), could be established to support AI and Gen AI development. The data commons can facilitate greater public access to open-source datasets, while the marketplaces would make it easier for private sector entities to reasonably license proprietary training data.
- ▶ **Bilateral arrangements:** This will allow reciprocal access to wider datasets for training.



Deployment of Gen AI Systems as Public Goods

India has a history of developing successful Digital Public Goods (DPGs) such as India Stack, Aadhaar, UPI, etc. Building on that success, the government may consider developing and deploying Gen AI algorithm(s) as 'Public Goods'.

- ▶ **Indian LLMs:** India can develop its own LLM (along the lines of UAE's Falcon LLM), and further develop local language LLMs. LLMs require large amounts of structured and unstructured datasets for training. India has a rich diversity in terms of languages and dialects which are spoken. A programmatic effort to collect and digitize the diverse written scripts and spoken languages

may be undertaken to help develop and localize Gen AI tools.

- ▶ **Usage and development of Gen AI for government services:** Interactive Gen AI tools may be deployed at various government portals, to help the public with new AI tools and improve service delivery/outreach. Such tools can enable the delivery of customized information and service provision requests. Government usage would also encourage adoption by private players. Gen AI use cases based on the priority sectors identified in NSAI may also be expedited.
- ▶ **Support for Open-Source ecosystem:** Development of an open-source ecosystem for basic algorithms and training datasets can help Indian entities and start-ups develop their own Gen AI products and fast-track indigenous innovation. A similar approach is being followed in jurisdictions such as the UAE and EU.



Securing Critical Digital Infrastructure

- ▶ **Access to chips:** Availability of computation infrastructure and securing the technology supply chain is imperative for the development and deployment of AI. Gen AI requires chips with high processing power such as graphics processing units (GPUs), field-programmable gate arrays (FPGAs), and application-specific integrated circuits (ASICs) that are specialized for AI. Presently, these chips are imported into India. Therefore, technology partnerships with countries like the US, Taiwan, etc. would be crucial in sourcing technology that enables the growth of AI systems in India and securing future supply chains. In addition, incentivizing the domestic manufacturing of the same may be expedited as recommended by MEITY's IndiaAI Expert Group Report 2023.
- ▶ **Access to enhanced computational capability:** Development of AI requires access to specialized computational capability. Cost of access to such computing power may be extremely challenging for smaller players, especially start-ups. Building on the suggestions of NITI Aayog in May 2023⁹,

⁹ National Artificial Intelligence Mission, PM-STIAC: <https://www.psa.gov.in/mission/artificial-intelligence/34>

the AIRAWAT supercomputer, focused on AI computing, was established at C-DAC¹⁰. Such efforts can democratize access to AI compute infrastructure. The technical requirements from a compute infrastructure/Gen AI point of view will still need regular review and additional infrastructure can be established. Like in Japan, domestic start-ups may be provided subsidized access to publicly funded compute resources¹¹.



Access to Talent and Public Funding of R&D

- ▶ **Access to talent:** The core development of AI systems requires specialized talent. To remain globally relevant and competitive, top-tier global talent may be attracted to help cultivate and develop the Indian talent pool and to improve technical proficiency. Cultivation of highly specialized talent will facilitate robust research programs, boost competitiveness, and help India become a global leader. The government would have to develop appropriate policies. Part of US's success in developing Gen AI, has been attributed to its ability to attract global talent. Further, the US Federal Trade Commission has identified access to specialized talent as a factor that can distort competition.
- ▶ **Boost public funding for R&D:** The US and Chinese governments have been earmarking public funds for investments in R&D, especially in areas where private funding may not be available such as for basic research. The Indian government could consider consistently allocating higher levels of public funds towards R&D.



Regulating Gen AI in India



Clarity on Regulatory Framework

- ▶ **Develop clarity on legal framework:** While DPDPA introduces a framework to protect data, in the absence of detailed rules, it is too early to comment upon its impact on development of Gen AI. The Digital India Bill is in the process of being finalized and is expected to address AI intermediaries. Speedier enactment of the same would provide greater regulatory certainty. Additionally, clarity on patenting algorithms, conferring inventorship and giving intellectual property rights to Gen AI products could help in encouraging/attracting small-scale players.
- ▶ **Self-regulation of Gen AI entities:** In the absence of detailed guidelines or a well-defined regulatory approach, self-regulation may be a practical interim solution. This is akin to the approach that the US is following, where leading AI players have signed pledges to observe certain rules pertaining to the development of AI. It will be easier and faster to roll out a self-regulation regime. This can help the government

¹⁰ National AI Platform: <https://indiaai.gov.in/about-us>

¹¹ MEITY: <https://www.meity.gov.in/content/indiaai-2023-expert-group-report-%E2%80%93-first-edition-the-ministry-electronics-and-information>

and private entities understand and manage emergent risks while preparing for the future. The government will have to issue standardized guidelines (for data use, processing etc.) for developers to mitigate instances of bias and other risks that arise out of deployment and use of Gen AI systems.



Deployment of Regulatory Sandboxes

Regulatory sandboxes could identify policy and regulatory gaps that are needed for deployment. The success of RBI's regulatory sandboxes (see Annexure I) for the fintech industry indicates that such a structure provides a safe place for innovators to test out their products. The government could consider setting up regulatory sandboxes, so that those algorithms that either require amendments to existing policies/ regulations or development of new policies can be tested in a controlled environment. For Gen AI development in India, regulatory sandboxes will provide regulatory flexibility for a limited time to test the usability and impact of emerging domestic Gen AI applications. Thus, furthering innovation in the Gen AI space and reducing the regulatory burden through the provision of safe testing spaces.

- ▶ **Roadmap with clear actions:** There should be a clear roadmap for developing regulatory capacity with specific regulators and technical capacity for security, testing and clearances. For instance, the UK has set up a single sector-multi-regulator sandbox.
- ▶ **Regulatory capacity enhancement:** Enforcement of regulations is likely to require a certain level of technical capacity, especially among governmental staff. Expert consultations may yield the technical requirements that will be necessary to augment and enhance the existing skillsets of officials. Assessments will also be necessary to determine the demand on potential regulators in terms of resources (monetary, personnel, technical infrastructure) so that required provisions can be made ahead of any regulatory roll-out. Regular technical capacity-building programs can help officials and technical personnel remain abreast with the changes. Tools and other capabilities needed to support regulatory capacity may also be examined.
- ▶ **Cybersecurity and testing infrastructure:** Cybersecurity frameworks are constantly evolving in response to emergent security threats. The sandboxes can also be used to determine the standards and approach to security testing. Gen AI systems will also have to address unforeseen vulnerabilities in the future. Robust security testing and auditing infrastructure will be needed for the protection



of all involved stakeholders (citizens, businesses, governments). Government capacity can be developed to support this regulatory function. Additionally, private-sector solution providers and experts may also be empanelled and engaged to support testing and security audit functions. The regulatory sandbox can help develop clarity around these aspects.

- ▶ **Protocols and standards for auditing algorithms/AI systems:** The approach to testing and evaluating algorithms and AI systems for bias, explainability, etc. will have to be standardized, similar to the algorithm testing framework that China has established and implemented. These standards may emerge from any sandbox initiative and will have to be regularly refreshed to ensure relevance. The standards may be notified and maintained by a government notified agency or standards body.
- ▶ **AI algorithms, tools and technologies to comply with existing laws and regulations:** The regulatory sandboxes may be made available for AI developers wherein the emerging legal implications may be studied in context of existing laws and regulations. Modifications or additional interventions may be required to ensure that the development and use of Gen AI are compliant with the existing regulatory/legal framework. Legal experts may also be included in the review and evaluation of the Gen AI sandboxes.



Watermarking AI Generated Products/ Media

- ▶ Gen AI systems have the capability to produce images, texts, and videos that are at par with content created by humans. Watermarking of AI products can safeguard human creativity and also address some of the concerns related to “derived” or “generated” content including fake news, deep fakes, etc. Since these are global tools used across different geographies and jurisdictions, a global framework/standard for watermarking technology could be considered. Different watermarking technologies can be piloted and tested.



Accuracy, Accountability and Liability

- ▶ To ensure that there is trust in the AI systems in the longer term, a clear liability framework for Gen AI may be required. Such a framework would require extensive stakeholder consultations as the technology is evolving and the full impact of the risks are not fully known/understood. ●





Annexures



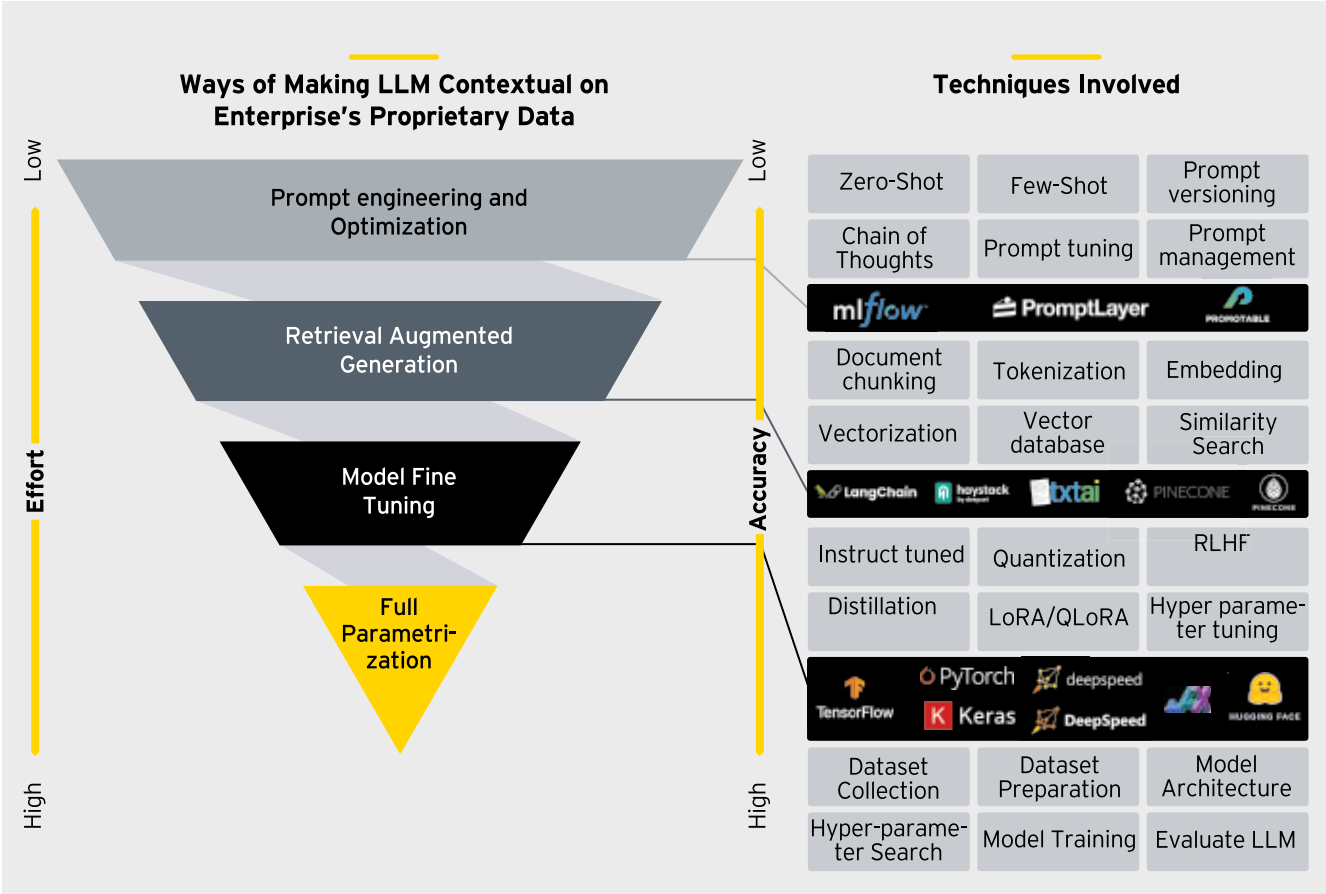
Annexures

Making LLMs More Contextual with Enterprise's Proprietary Data

Retrieval Augmented Generation (RAG): This is an approach where enterprise knowledge i.e., documents, policies, manuals, SOPs etc. is stored in a vector database. User query is converted into a semantic or contextual search to the Vector Database which extracts the most relevant content from

the knowledge base. This proprietary data is then used to generate the response ensuring a relevant response with no hallucinations. RAG helps overcome the context window token limit and adds more intelligence by providing the most relevant enterprise data across the knowledge base. It can also return the source data for AI transparency and can be made more accurate by knowledge graph, dynamic indexes and intelligent chunk retrieval.

Fine tuning: LLMs are built using data sourced from different sources such as websites, research papers, journals, forums etc. LLM model parameters that help generate LLM responses are trained on this massive repository. Fine tuning an LLM by adding



enterprise's own data means changing the underlying model parameters by training it on this new domain-specific data. The resultant model is a customized enterprise version of the LLM which understands and generates content more relevant to the domain and/or the enterprise.

Today, you can fine tune select Closed Source LLMs and most Open Source LLMs. As Open Source LLMs are smaller in size, resulting in lower compute cost

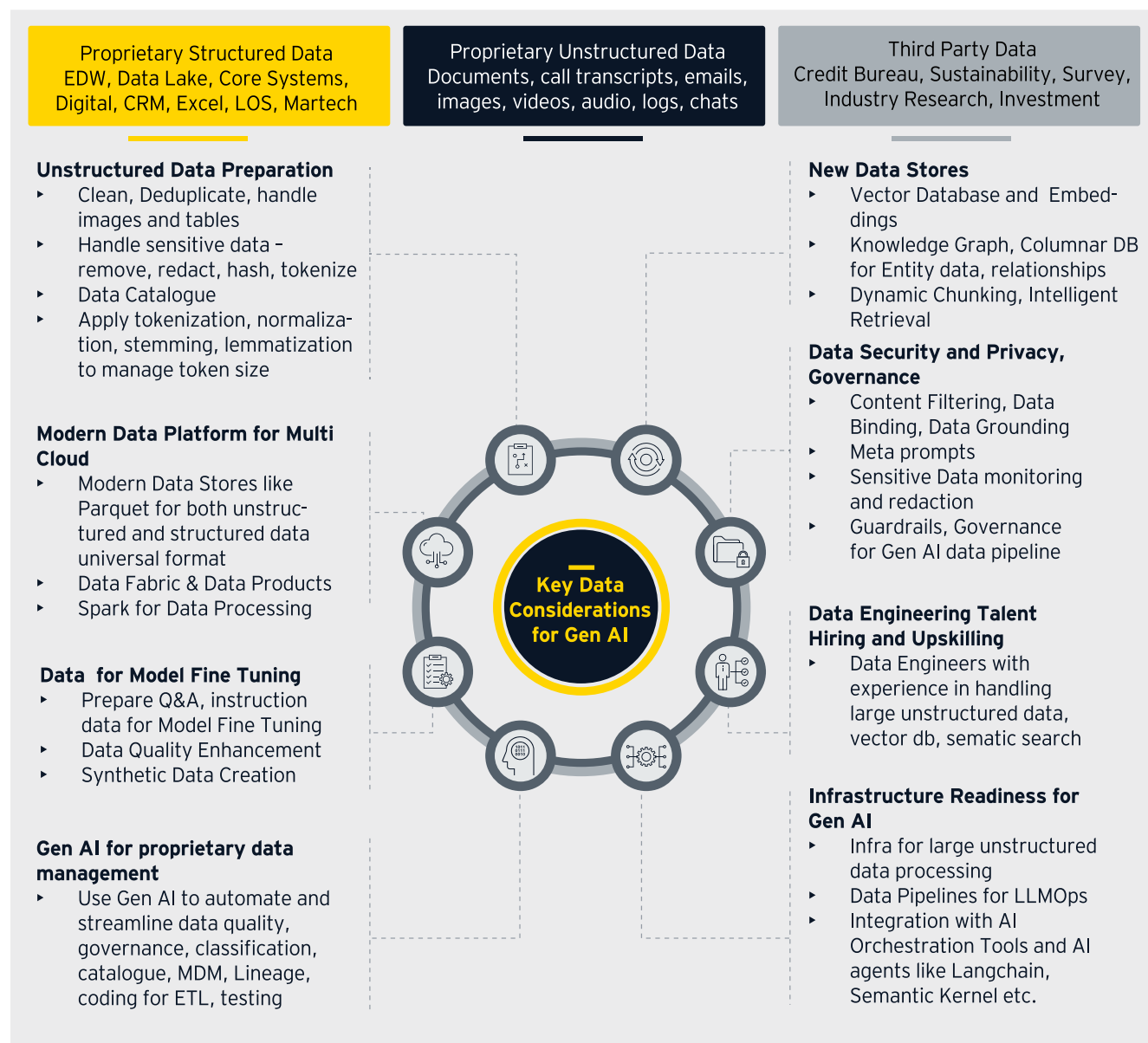
of fine tuning, and also as AI engineers have more control on the fine tuning process, many enterprises are exploring Fine Tuning of these models to create enterprise-specific variations.

Prompt: This is the input for LLMs. Techniques like Prompt Engineering and Optimization, including advanced methods like Prompt Chaining and Chain of Thought, enhance the model's ability to understand and use enterprise data contextually.

B Key Data Considerations for Gen AI

AI is only as good as the underlying data. So, for a successful AI-led transformation, it is critical to have

Data Foundation in place, which involves modernizing the Data Platform. See key pointers below.



Pros and Cons of Approach 1 : Fine Tuning Open Source LLM*

Pros

- Enterprise's own domain and instruction-specific LLM: It is easier to Fine Tune an Open Source LLM if there is a need for contextual relevant responses using very specific domain data
- Comparable performance in straight forward tasks like conversation and summarization versus Closed Source LLMs in benchmarks
- Lower cost of inferencing versus Closed Source LLM if the input token volume is very high
- High data security and control

Cons

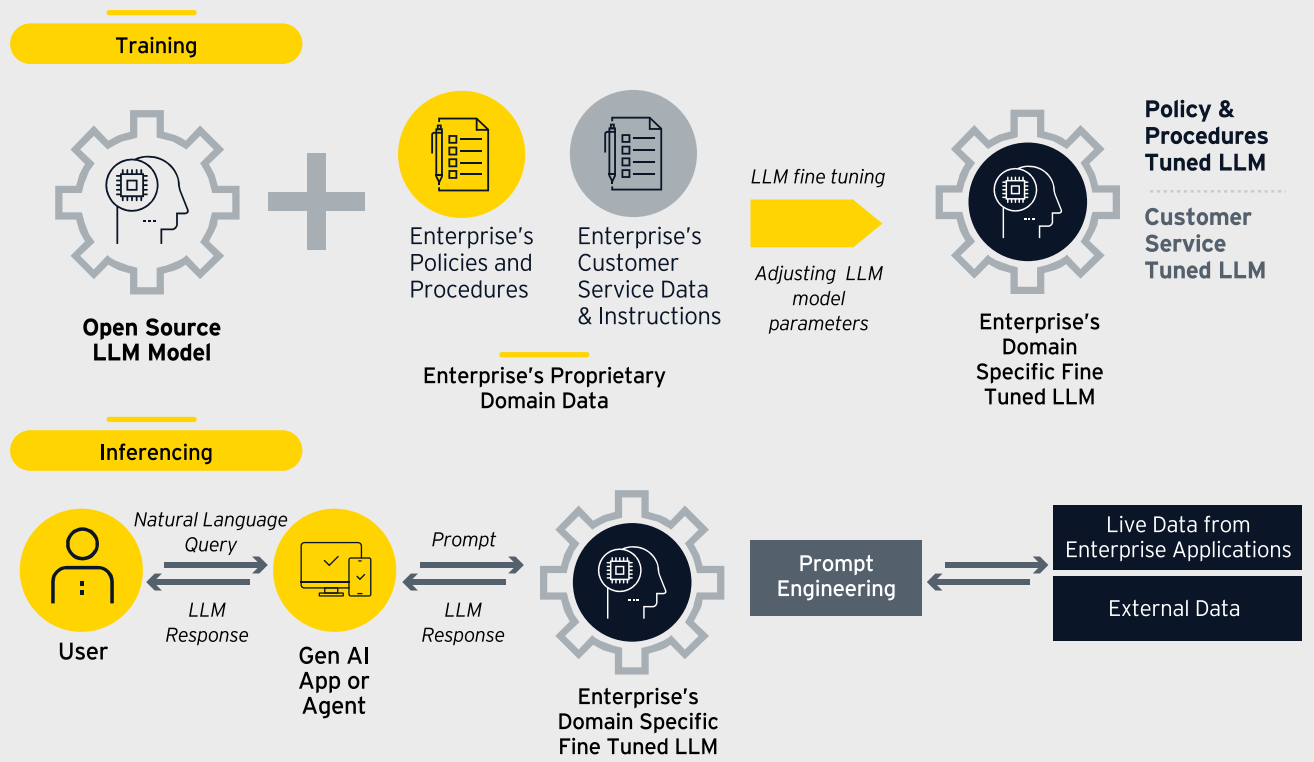
- High Fine Tuning cost: Can run into thousands of dollars if training dataset and
- LLMs are large, though there are techniques now for cost optimization with PEFT, Quantization, etc.

- Infrastructure can become a big bottleneck in Speed to Market given the shortage of talent required for self-hosting, fine tuning, GPU optimization, infra management. There is a big challenge around Training GPU availability, for example, Nvidia A100s
- Lock-in to a specific LLM model given rapid evolution
- LLM Ops gets complex with model versions
- Versatility of these Domain Open Source LLMs is low; higher hallucinations for new tasks, especially for smaller 7B parameter models

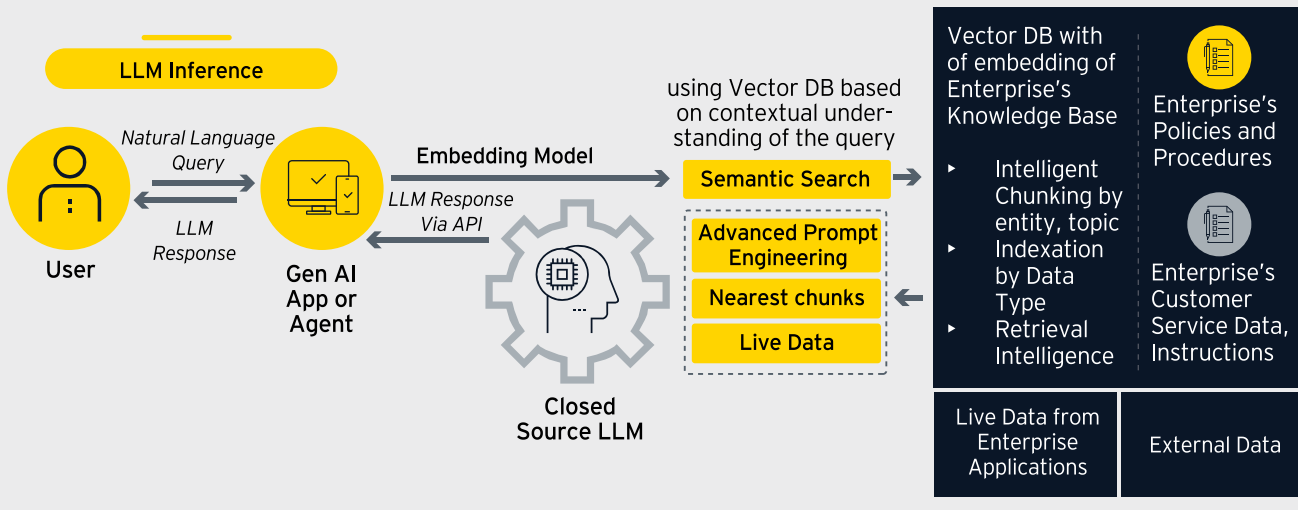
Typical use cases

- Use cases which require sensitive data like Personally Identifiable Information (PII)
- Chat and summarization use cases on enterprise's own proprietary data for employees like Risk SOPs, AML policies
- Lower complexity 'end customer' use cases, for example, queries on products and warranty. Not for complex interactions like grievance resolution, financial planning

Open source LLMs with fine tuning



Closed source LLM with enterprise's data embeddings



Pros and Cons of Approach 2 : Closed Source LLM with RAG*

Pros

- ▶ Quick deployment for a wide variety of use cases; no infrastructure management and talent required
- ▶ Embedding enterprise's own data in responses without upfront high fine tuning cost and GPU availability issues
- ▶ Creation of Easy to Query 'Enterprise's Knowledge Repository'
- ▶ Handle complex questions and prompts like grievance handling by CSR
- ▶ RAG is essential when prompt data exceeds context window limit, there is a need for summary of text chunks across multiple documents, there is a need for higher memory, i.e. factors where prompt engineering is not effective for data grounding
- ▶ Experiment with new LLMs while using same vector database for proprietary data embedding
- ▶ Show the data sources along with LLM response for transparency

Cons

- ▶ High inference cost if input tokens are very large
- ▶ Lower control on LLM, its optimization, size, tuning
- ▶ Some enterprises have reservations on PII or sensitive data sharing even over encrypted APIs via secure endpoints
- ▶ RAG requires right data engineering and AI orchestration talent

Use case

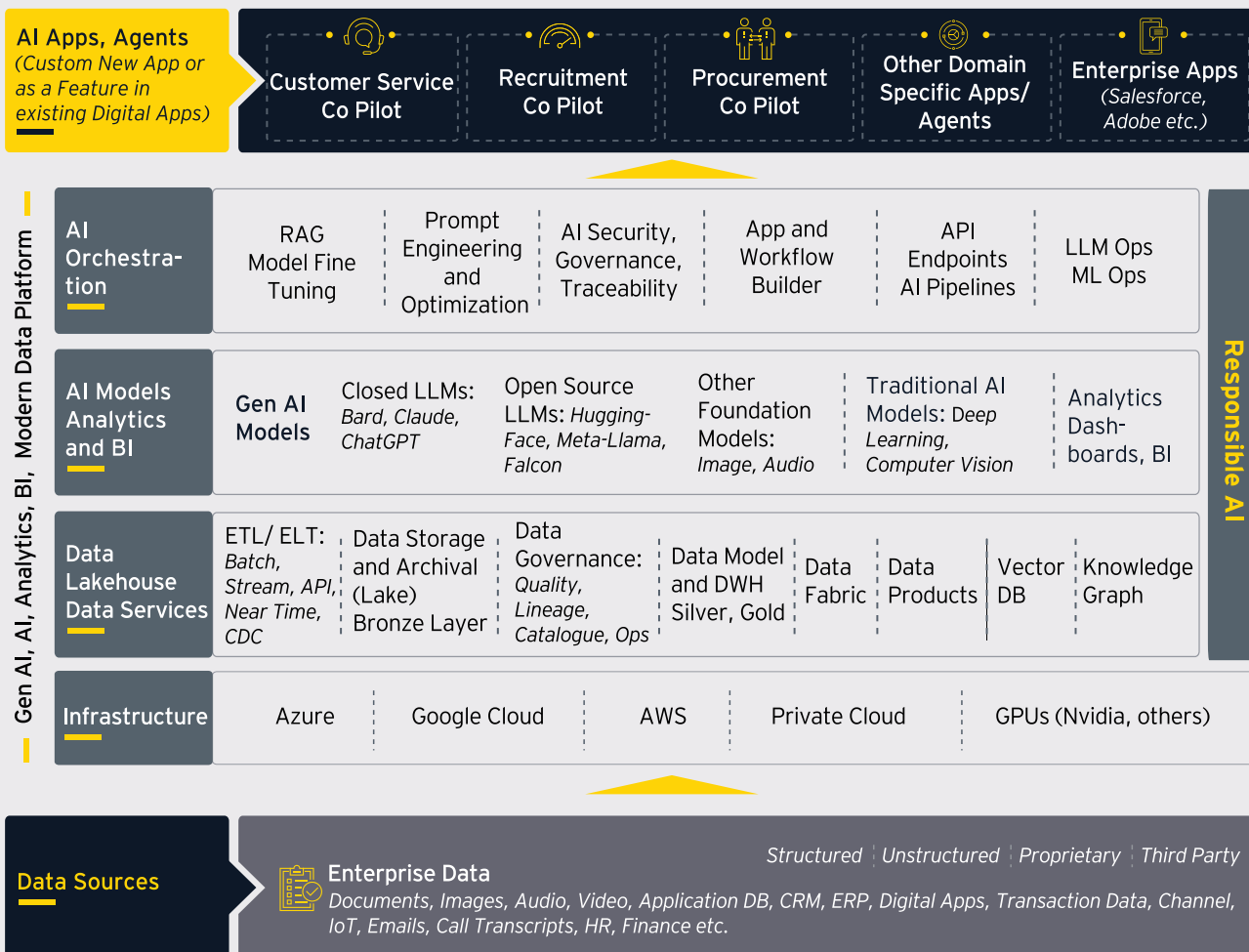
- ▶ Use cases with complex generation requirement, for example, using C360 data to create personalized 1-1 creative email copy
- ▶ Use cases with varied and versatile instructions, for example, converting call transcript to JSON and storing in database
- ▶ Use cases with complex automation and need with access to broad knowledge base, for example, analyzing vendor contracts to find deviations from regulations

*We are going into a detailed comparison of two key archetypes for grounding LLM in enterprise proprietary data. Please note that you can Fine Tune select Closed Source LLMs and use RAG with Open Source LLMs as well.

D

Reference Architecture

Reference Architecture for hybrid LLM deployment for an integrated Gen AI, AI, analytics and data platform









E

Best Practices for Optimizing LLM and Managing Gen AI Costs using FinOps

API cost management for Closed Source LLMs

- Optimize query length:** Tune prompt for output with fewer tokens; reduce redundant words for input
- Caching responses:** Store frequent responses locally to minimize redundant API calls
- Analyze response usage:** Monitor API responses to identify and optimize high-cost queries
- Use batch processing:** Combine multiple requests into a single batch can help improve response latency
- Rate limiting and throttling:** Managing requests and API volume for better user experience, latency and costs

Compute cost management for Open Source LLMs

	Optimize hardware resources: Match model's needs with CPU/GPU/TPU configurations to avoid overprovisioning and cut costs
	Model quantization: Use quantization to shrink model size and lower resource usage
	Caching responses: Cache and reuse common responses to save on repetitive computations
	Batch processing: Handle multiple requests simultaneously to optimize resources and reduce costs
	Monitoring and optimization: Regularly review resource usage and performance for ongoing improvements and maximized efficiency
	Use efficient model architectures: Choose balanced, lightweight models for performance and cost- efficiency

F Responsible AI

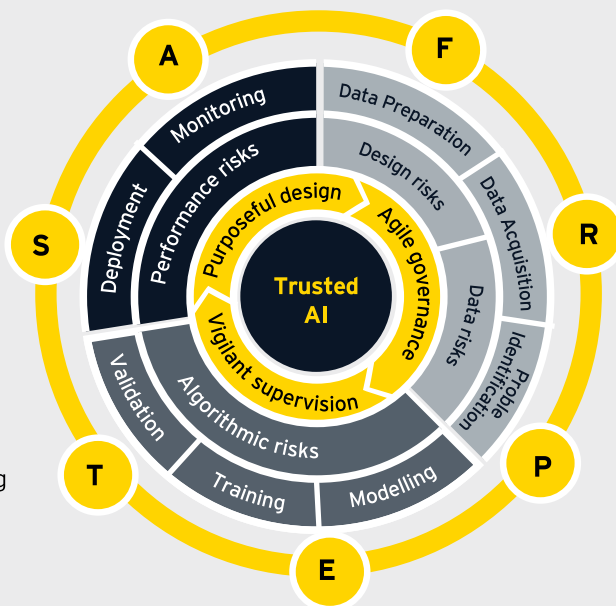
EY's approach to Responsible AI

The Responsible AI framework developed by EY enables clients to mitigate AI risks while complying with emerging AI regulations. It can evaluate AI risks and build controls across seven trust attributes, four risk categories and three governance domains.

Accountability: there is unambiguous ownership over AI systems and their impacts across the AI development lifecycle.

Sustainability: the design and deployment of AI systems are compatible with the goals of sustaining physical safety, social well-being, and planetary health.

Transparency: appropriate levels of openness regarding the purpose, design, and impact of AI systems is provided so that end users and system designers can understand, evaluate, and correctly employ AI outputs.



Fairness: AI systems are designed with consideration for the need of all impacted stakeholders and to promote inclusiveness and positive societal impact.

Reliability: outcomes of AI systems are aligned with stakeholder expectations and perform at a desired level of precision and consistency, whilst being secured from unauthorized access, corruption, and/or adversarial attack.

Privacy: AI systems are designed with consideration to data rights regarding how personal information is collected, stored, and used.

Explainability: appropriate levels of explanation are enabled so that the decision criteria of AI systems can be reasonably understood, challenged, and/or validated by human operators.



C Technical Appendix

Our methodology for estimating the economic impact of Gen AI on the Indian economy utilizes a macro framework in combination with survey-based inputs on the sectoral impact of Gen AI. The survey has been conducted across the 27 sectors of the economy as per the KLEMS database (RBI). The survey gathers inputs from EY's sector leaders whose assessments are based on their expertise and close interaction with clients. These clients include key players in their respective industry. The survey primarily captures the efficiency effects of Gen AI adoption in terms of expected cost reduction and output expansion¹ over a seven-year period starting 2023-24 to 2029-30.

These inputs have then been incorporated in a macro framework to estimate the potential additional contribution to India's GVA by 2029-30 on account of Gen AI adoption. The impact of Gen AI on India's GVA can be explained in three steps:

01 Estimation of sectoral impact of Gen AI on the organized segment of each sector

The following variables have been used in the estimation framework:

Table 1: List of variables

#	Variable	Description
1	V_i	GVA of the i th sector
2	O_i	Gross output of the i th sector
3	I_i	Intermediate consumption of the i th sector
4	U_i	share of the unorganized segment of i th sector
5	g_i	Gen AI impact in terms of percentage change in sector i 's GVA
6	ΔGVA_i^{FY30}	Additional nominal sectoral GVA in the terminal year 2029-30 on account of Gen AI
7	$\Delta BGVA_i^{FY30}$	Base GVA magnitudes for 2029-30 excluding the effect of Gen AI for each sector
8	$\Delta AGVA_i^{FY30}$	total additional GVA for the economy as a whole on account of Gen AI in the terminal year 2029-30

Using (a) cost reduction and output expansion effects (in percentage terms) captured through the survey and (b) ratio of output and intermediate inputs relative to value added estimated from KLEMS database, we have assessed the impact of Gen AI on the organized segment of sector i's GVA (in % terms) through the following equation (1)²

$$\frac{\Delta V_i}{V_i} = \left[\left\{ (1 + \% \text{ output expansion}) * \frac{O_i}{V_i} \right\} - \left\{ (1 - \% \text{ cost reduction}) * \frac{I_i}{V_i} \right\} - 1 \right] \quad (1)$$

Where V_i is the GVA of the ith sector (avg. of 2017-18 to 2019-20 from KLEMS); O_i is the gross output of the ith sector (avg. of 2017-18 to 2019-20 from KLEMS); I_i is the intermediate consumption of the ith sector (avg. of 2017-18 to 2019-20 from KLEMS)

02

Estimation of sectoral impact of Gen AI on each sector's total GVA

Impact of Gen AI on total GVA (organized + unorganized)³ of sector i (in % terms) is given by equation (2)

$$(1 - U_i) \frac{\Delta V_i}{V_i} = (1 - U_i) \left[\left\{ (1 + \% \text{ output expansion}) * \frac{O_i}{V_i} \right\} - \left\{ (1 - \% \text{ cost reduction}) * \frac{I_i}{V_i} \right\} - 1 \right] \quad (2)$$

Where U_i is the share of the unorganized segment of ith sector. The share of unorganized segment of each sector has been sourced from the IMF for the year 2017-18. The shares provided by the RBI for the unorganized segment have been used for the manufacturing sub sectors.

$$(1 - U_i) \frac{\Delta V_i}{V_i} \text{ may be referred to as the Gen AI impact } (g_i)$$

03

Estimation of the magnitude of sectoral impact of Gen AI

The magnitude of additional nominal sectoral GVA in the terminal year 2029-30 (ΔGVA_i^{FY30}) on account of Gen AI is estimated by applying the Gen AI impact (g_i) on the base GVA (BGVA) magnitudes (which does not include the effect of Gen AI) for each sector

$$\Delta GVA_i^{FY30} = (\Delta BGVA_i^{FY30} * g_i)$$

Hence the total additional GVA for the economy as a whole in the terminal year 2029-30 can be given by

$$\Delta AGVA^{FY30} = \sum_i^{27} (\Delta GVA_i^{FY30}) = \sum_i^{27} (\Delta BGVA_i^{FY30} * g_i)$$

The augmented overall GVA which includes the additional GVA due to Gen AI + Base GVA (ΔGVA_i^{FY30}) in the terminal year of 2029-30 can be written as

$$GVA^{FY30} = \Delta AGVA^{FY30} + BGVA^{FY30}$$

Further,

$$GDP^{FY30} = GVA^{FY30} + NIT^{FY30}$$

¹ It is assumed that the expansion in output on account of Gen AI adoption will be absorbed by the creation of corresponding demand in the economy.

² The derivation of equation (1) is given below:

Gross output (O) = Total input cost [energy + material + services] (I) + GVA (V)

$O = I + V$

$V = O - I$

$(V + \Delta V) = (O + \Delta O) - (I + \Delta I)$

The output expansion and cost reduction effects of the adoption of Gen AI can then be incorporated in the above equation as:

$$\frac{\Delta V_i}{V_i} = \left[\left\{ (1 + \% \text{ output expansion}) * \frac{O_i}{V_i} \right\} - \left\{ (1 - \% \text{ cost reduction}) * \frac{I_i}{V_i} \right\} - 1 \right] \quad (1)$$

³ We implicitly assume that the impact of Gen AI on the unorganized sector may be insignificant considering the requirement of relatively higher levels of investment in technological infrastructure and skilled workforce to adopt the new technology.

The base GVA in 2029-30 has been estimated using the projected nominal GDP magnitudes in INR terms available up to 2028-29 sourced from the IMF (WEO April 2023). Base GVA during the period from 2023-24 to 2028-29 has been derived from first projecting GDP and then subtracting from it net indirect taxes (NIT). An average nominal growth of 11% (avg. over 2023-24 to 2028-29) has been applied to project the nominal (Base) GDP for 2029-30. From this, the GVA in nominal terms has been estimated by netting out the net indirect taxes⁴ for 2029-30. The overall nominal GVA estimated in 2029-30 has been split across the 27 KLEMS sectors to derive the base GVA for each sector ($\Delta BGV A_i^{FY30}$) by utilizing the average share of these sectors in GVA during 2017-18 to 2019-20.

Using the sectoral estimates of Gen AI induced incremental GVAs, we have calculated the aggregate GVA including the impact of Gen AI and from this GDP including the impact of Gen AI has been derived by adding net indirect taxes for the terminal year 2029-30. These magnitudes in INR terms have been converted into US\$ market exchange terms by applying the estimated exchange rate for the terminal year⁵.

H Global Developments with Respect to Regulations and Gen AI

Countries are adopting differential approaches to promoting and regulating AI, and more specifically, Gen AI. While all countries emphasize that AI regulation must strike a balance between fostering innovation while managing the risks, however their approaches and emphasis differs widely. With some putting greater focus on promotion and development and others on mitigating the risks from the implementation of the technology. The role of the government in developing AI algorithms also differs.

⁴ Share of net indirect taxes averaged 9% during FY2016 to FY2020. This share has been used to derive the magnitude of net indirect taxes.

⁵ The exchange rate for FY2030 is estimated by applying on the derived exchange rate of FY2029 (IMF WEO April 2023) a depreciation rate of 2.2% (average over the five-year period of FY2025 to FY2029).

The US

Majority of the companies that are behind the development of the foundational Large Language Models (LLMs) which power Gen AI, are based out of the US. According to a recent report⁶ submitted to the US government, the breakthroughs have not happened by chance. The report acknowledges that they have emerged from an ecosystem characterized by:

- ▶ Decades of systematic investments in cyber infrastructure and research
- ▶ Highly specialized education and training programmes
- ▶ Large and growing amounts of data and computational power
- ▶ Collaborations between academic researchers and the private sector

Development of Gen AI has been underpinned by years of investments by US-domiciled corporations running into several billion dollars and building upon existing technical expertise⁷. The established players have benefitted from access to data collected over many years, access to computing power (including through ownership of data centers), billions of users and the ability to attract and retain talent including from outside the US.

Similar environment does not exist in most jurisdictions around the world and therefore governments in countries such as India may need to play a more active role in enabling the development of algorithms, like the development of the digital stack by the Government of India.

Given the lead, the current focus in the US is on maintaining global leadership. Some of the measures/proposals under consideration include:

- ▶ Setting up of a National AI Research Resource (NAIRR) that provides access to a federated mix of computational and data resources, testbeds, software and testing tools, and user support services via an integrated portal. These resources are proposed to be made broadly accessible to a range of users to increase the diversity of AI researchers.

⁶ Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource (ai.gov)

⁷ <https://www.economist.com/business/2023/03/26/big-tech-and-the-pursuit-of-ai-dominance>

- ▶ The US government issued an Executive Order (9 August 2023) on the ground of national security, that limits the ability of competing countries to get hold of software programs pertaining to the development of AI as well as at the hardware required to develop it⁸. Pursuant to that, the US would add or remove products/technologies under “covered national security technologies and products.” The order specifically mentions semiconductors and microelectronics, quantum information technologies, and artificial intelligence sectors.
- ▶ The US government issued another Executive Order on 30 October 2023, which signals efforts to introduce safeguards to address the perceived risks of AI. The “Executive Order on Safe, Secure and Trustworthy Artificial Intelligence” introduces a sweeping range of measures to not only cater to safety, security, transparency, various citizen protections, effective government usage, etc. but also targets the promotion of innovation, competition, and global leadership of the US in the AI space⁹.

The above measures/proposals have implications as policymakers in other countries including India formulate more detailed policies. While the US leads in terms of innovation, the process of creation of AI rules to address risks has commenced. Some of the notable points are:

- ▶ On 21 July 2023, the White House announced that seven leading AI companies (OpenAI, Amazon, Anthropic, Google, Inflection, Meta and Microsoft) have pledged to observe voluntary safeguards in terms of the development of the technology¹⁰. As a part of the pledge, the companies will ensure internal and external security testing before release, ensure people are able to spot AI watermarks (to know if something has been generated through AI), be transparent by regularly publicly reporting on the capabilities and limitations, and research any risks relating to discrimination, bias or violation or privacy¹¹.

- ▶ There have been a series of Senate hearings and press conferences, though policies are yet to be set¹². Further, some of the US agencies such as the Federal Trade Commission have raised both consumer protection and competition issues that need attention¹³.
- ▶ The US government issued an Executive Order on 30 October 2023, which signals efforts to introduce safeguards to address the perceived risks of AI. The “Executive Order on Safe, Secure and Trustworthy Artificial Intelligence”¹⁴ directs a wide range of executive actions to be taken in a time bound manner pertaining to:
 - ▶ **Establishing Standards for AI Security and Safety** - through safety testing/red teaming, standards development of standards/tools/testing protocols, testing tools, standards for biological synthesis screening, watermarking, standards and best practices for fraud/deception detection, authentication and verification, cyber security tools etc.
 - ▶ **Protecting the Privacy of American Citizens** - through need for bi-partisan data privacy legislation development of privacy preserving techniques, strengthening of privacy preserving technologies, privacy guidance for data collection by agencies, evaluation of effectiveness of privacy preserving techniques
 - ▶ **Advancing Equity and Civil Rights** - by providing guidance to prevent algorithmic discrimination, development of best practices for fairness for the use of AI in the criminal justice system
 - ▶ **Protections for Consumers, Patients, and Students** - through the promotion of responsible use of AI and leveraging AI in education
 - ▶ **Support for Workers** - through the development of best practices to mitigate harm and to maximize benefits and

⁸ Executive Order on Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern | The White House

⁹ Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence | The White House

¹⁰ Meta, Google and A.I. Firms Agree to Safety Measures in Biden Meeting - The New York Times (nytimes.com)

¹¹ Seven AI companies agree to safeguards in the US - BBC News

¹² A.I. Regulation Is in Its 'Early Days' - The New York Times (nytimes.com)

¹³ <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns>

¹⁴ Executive Order on the Safe, Secure, and Trustworthy Development and

Use of Artificial Intelligence | The White House

¹⁵ “Regulations of European Parliament and of the council Laying down harmonised rules on artificial intelligence and amending certain union legislative acts”, June 14, 2023 : <https://artificialintelligenceact.eu/the-act/>

¹⁶ “The EU AI act can get democratic control of artificial intelligence-but only if open source developers get the seat in the table”, July 17 2023, <https://fortune.com/2023/07/17/eu-ai-act-democratic-control-artificial-intelligence-open-source-developers-tech-politics-shelley-mckinley/>

¹⁷ Europe Takes aim at ChatGPT with what might soon be the west's first A.I. law”, June 14 2023, <https://www.cnn.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html>

commissioning of a report to evaluate the impact of AI on the labor market

- ▶ **Promoting Competition and Innovation** - by catalyzing AI research through provision of resources, providing small developers/entrepreneurs technical assistance and resources and attracting highly skilled global AI talent
- ▶ **Promoting American Leadership Abroad** - through the expansion of bilateral/multilateral/multistakeholder engagements to promote collaboration on AI, development of vital international standards, and promotion of responsible deployment of AI abroad
- ▶ **Enabling Effecting and Responsible use of AI in Government** - through the issuance of guidance of use of AI by government agencies, helping agencies with the acquisition of AI products and services, and accelerating rapid induction of AI professionals in the government

Europe's AI regulation approach remains "Risk Based"

The European Union (EU) is taking a risk-based approach to the regulation of AI, through which it seeks to strike a balance between protecting people from the risks of AI while promoting innovation. The EU parliament has recently approved a set of regulations for ratification/ consideration of member countries¹⁵.

Before an AI system can be launched for public, it must meet a set of risk-management, transparency, documentation, oversight, and quality requirements¹⁶. The Act defines four levels of risks that are unacceptable risk, high risk, limited risk and minimal to no risk. The models falling in the unacceptable risk category cannot be deployed in the

EU¹⁷ while the high-risk AI systems need to undergo third-party conformity assessment¹⁸.

The law proposes new transparency requirements for the developers of foundation models:

- ▶ They require the developers to publish a summary of the copyrighted material used in training such systems.
- ▶ The AI model developers will be required to carry out data integrity assessments to reduce the possibility of biases.
- ▶ The models will have to be consistent with the principles of democracy, rule of law, and mitigate risks to fundamental rights, health and safety of the citizens in the EU¹⁹.

It may be noted that EU was among the first to enact regulations relating to data privacy and consumer protection i.e., the GDPR. Since social media platforms and the internet ecosystem in general are global in nature, even companies domiciled outside the EU have complied with the GDPR regulations. Similarly, once implemented, this risk-based approach of EU may need to be complied with by companies domiciled outside the EU.

- ▶ **Support to open source:** The Act grants certain protections and some exemptions to open-source community from legal and compliance issues. While Open-source developers are encouraged to implement documentation of best practices, such as model and data cards, but the responsibility for compliance ultimately falls on the entities that incorporate open-source components into their AI applications²⁰.
- ▶ **Measures for innovation in EU:** The EU is promoting Gen AI/AI innovation through regulatory sandboxes, and measures to reduce the regulatory burden for SMEs and start-ups. In addition, they have also promulgated rules/directives such as Data Governance Act which facilitates the open sharing of data, Open Data Directive which mandates the release of public sector data, and EU strategy for data which enables the creation of a mechanism to create a single market for data. Further, it has also notified rules to access data to develop high-risk AI systems²¹.

¹⁵ Regulations of European Parliament and of the council Laying down harmonised rules on artificial intelligence and amending certain union legislative acts", June 14 2023

¹⁹ Europe Takes aim at ChatGPT with what might soon be the west's first A.I. law", June 14 2023, <https://www.cnn.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html>

²⁰ The EU AI act can get democratic control of artificial intelligence-but only if open source developers get the seat in the table", July 17 2023, <https://fortune.com/2023/07/17/eu-ai-act-democratic-control-artificial-intelligence-open-source-developers-tech-politics-shelley-mckinley/>

²¹ "The EU AI act": <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

UK proposes regulatory sandboxes to encourage innovation

The UK has published a policy paper that emphasises a “pragmatic” and “proportionate” approach. It proposes a pro-innovation regulatory environment while “responding to risk and building public trust”¹⁷. The activities in this respect would be undertaken over the next 12 months, including the development of AI regulation roadmap, stakeholder engagement/ consultations and the development of cross-sectoral principles for provisions²².

The UK government has proposed an AI sandbox, similar to what RBI has implemented in India for fintech²³ to help innovators overcome regulatory barriers and speed up product launches. It is also expected to highlight emerging technologies and market trends that may require regulatory adjustments.

Initially, the government intends to focus on a single sector, multiple regulator model while expanding to other sandbox models at a later stage. The AI regulatory sandbox will prioritize sectors with substantial AI investment, strong industry demand, and a need for improved collaboration between regulators. Through this sandbox, government expects to provide tailored advice to innovators to overcome regulatory barriers, with a focus on start-ups and small to medium-sized businesses.

China focuses on data accuracy for better Gen AI outcomes

The Chinese government, through a proclamation in 2017, made leadership in AI a strategic priority²⁴, thereby providing a strong policy signal to private players and government funded institutions. Much like US tech corporations, China has large tech giants with deep pockets and tech know-how, who have spearheaded AI development²⁵. In addition, as per

reports from OECD, Brooking Institution and Stanford University, China has consistently been the second largest recipient of venture capital in AI behind only the US.

China also benefits from large internet consumer base. Chinese consumers generate lots of data that can be accessed by government and provided to Chinese firms²⁶. Since 2017, the federal government has made concerted efforts through initiatives such as National New Gen AI plan, AI innovation action plan for institutions of higher education with stated targets for talent development, R&D and AI industrialization²⁷.

The Chinese government has stated that its goal is to enable China to emerge as a leader in the governance and regulation of AI²⁸, and to address the ethical, social, and economic impact of AI. As per the current draft policy which has been put out for consultation:

- ▶ Technology should comply with the existing regulatory framework
- ▶ Service providers are expected to “ensure the data’s veracity, accuracy, objectivity, and diversity”
- ▶ Service providers are expected to ensure that intellectual property rights are not infringed upon, nor are any other data-related laws violated (consent should be obtained as may be applicable)
- ▶ Providers are expected to ensure that the content that is generated is not discriminatory in any nature. Individuals and entities using Gen AI products to “provide services such as chat, text, image, or audio generation” are responsible for any content that is generated
- ▶ China’s approach to Gen AI governance has a particular focus on algorithms and the data used to train them. This is evident from the requirement to make a filing to China’s algorithm registry explaining how an algorithm works, trained and then pass security assessments²⁹

Chinese tech giants are set to invest over US\$5 billion by 2024 in buying chips³⁰. Also, local governments

²² A pro-innovation approach to AI regulation - GOV.UK (www.gov.uk) A pro-innovation approach to AI regulation - GOV.UK (www.gov.uk)

²³ “RBI Regulatory Sandbox”: <https://rbi.org.in/scripts/PublicationReportDetails.aspx?UrlPage=&ID=920> “RBI Regulatory Sandbox”: <https://rbi.org.in/scripts/PublicationReportDetails.aspx?UrlPage=&ID=920>

²⁴ “Next Generation Artificial Intelligence Development Plan Issued by

State Council”: <http://fi.china-embassy.gov.cn/eng/kxjs/201710/P020210628714286134479.pdf>

²⁵ <https://www.technologyreview.com/2019/01/22/137760/the-future-of-chinas-ai-industry-is-in-the-hands-of-just-three-companies/>

²⁶ <https://www.bbc.com/news/business-65034773>

²⁷ <https://oecd.ai/en/dashboards/policy-initiatives/http%2F%2Faiipo.oecd.org%2F2021-data-policyInitiatives-26851>

in China are taking steps to shore up local firms by providing state-sponsored computing resources to AI firms³¹, and investment in algorithmic innovation, chip development and language data sets³².

China had earlier rolled out detailed regulations pertaining to Artificial Intelligence and has been one of the first countries to come up with a draft document on the Measures for the Management of Generative Artificial Intelligence Services³³. The Cyberspace Administration of China (CAC) issued these draft measures on April 10, 2023 and were open for public consultation till May 10, 2023.

UAE has developed its own open-source Gen AI model "FALCON"

The UAE has launched a Gen AI guide to address the opportunities and challenges of this technology in the country. The guide not only details the economic potential of the Gen AI but also stresses the individual's data privacy and protections. It considers transparency to be essential in the use of data to build people's trust in the technology³⁴.

The UAE government through Technology Innovation Institute (TII) has taken a path much like India's UPI and has funded its indigenous technology. It has created LLMs called FALCON-40B and FALCON-7B as a Public Good and open-sourced it. A key feature of this LLM FALCON-40B is that it has 40 billion parameters while the FALCON-7B has 7 billion parameters. A greater number of parameters in FALCON-40B allows it to display a higher level of machine intelligence enabling the application to use it in relatively complex tasks when compared to FALCON-7B.

As per the UAE government, the data pipeline used for FALCON has undergone a rigorous audit. This is expected to address problems such as inherent bias

in the models. Further, this enhanced data pipeline allows it to match the performance of other models with only 75% of the compute budget while training and only a fifth during inference time³⁵. As per TII, this Gen AI model is expected to help Emirati companies be cost-efficient. One important public use of these domestic LLMs is that they may be extensively used for language translations, sentiment analysis and as virtual assistants³⁶.

Other countries

Several countries such as Japan, Australia, Singapore, Germany, Canada and France have put out broad outlooks towards the policy stance to be adopted for AI and Gen AI. Each country has a unique stance on AI and India can draw aspects that support AI innovation, and protection of Gen AI users without having any adverse impact.

Learning from RBI: use of Regulatory Sandboxes for Gen AI

A Regulatory Sandbox (RS) is an instrument that allows start-ups or companies to test new technologies and their impact where regulators may permit a certain level of relaxation to undertake a trial phase³⁷. This helps provide a safe platform for regulators as well as the innovators to collaborate and understand how new technologies can be developed and regulated in a responsible way. Presently, in India, the three main financial regulators - Reserve Bank of India, Securities and Exchange Board of India, and Insurance Regulatory and Development Authority of India - have an RS framework to test fintech products. RBI's RS has progressed to its fourth cohort of testing while the other two are in the nascent stage³⁸.

²⁸ "Measures of Management of generative Artificial intelligence Services draft Translation": <https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/>

²⁹ "Algorithms as a point of entry to AI Governance": <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>

³⁰ <https://www.yahoo.com/lifestyle/chinese-companies-spend-5-billion-153400526.html#:~:text=Chinese%20tech%20giants%20Baidu%2C%20TikTok,on%20Nvidia's%20professional%20AI%20GPUs.>

³¹ <https://www.scmp.com/tech/big-tech/article/3226808/china-asserts-ai-development-goals-shanghai-conference-beijing-braces-potential-new-us-tech>

³² <https://www.scmp.com/tech/policy/article/3220736/chinas-capital-beijing-provide-state-sponsored-computing-resources-ai-firms-amid-chatgpt-frenzy>

[beijing-provide-state-sponsored-computing-resources-ai-firms-amid-chatgpt-frenzy](https://www.scmp.com/tech/policy/article/3220736/chinas-capital-beijing-provide-state-sponsored-computing-resources-ai-firms-amid-chatgpt-frenzy)

³³ <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>

³⁴ <https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/>

³⁵ https://ai.gov.ae/wp-content/uploads/2023/04/406.-Generative-AI-Guide_ver1-EN.pdf https://ai.gov.ae/wp-content/uploads/2023/04/406.-Generative-AI-Guide_ver1-EN.pdf

³⁶ "Refineweb: dataset for Falcon": <https://arxiv.org/abs/2306.01116>

³⁷ <https://falconllm.tii.ae/> <https://falconllm.tii.ae/>

³⁸ Reserve Bank of India: <https://rbi.org.in/scripts/PublicationReportDetails.aspx?UrlPage=&ID=920>

RBI has noted several benefits of undertaking the RS approach for fintech rollout in India. These include reduced dependence on the regulator for stakeholder consultations, gaining empirical evidence on the merits and risks of new technologies, allowing for modifications based on user feedback before large-scale rollout and providing a structured environment for evidence-based regulatory decision-making³⁹. Thus, the RS approach holds value for Gen AI rollout as well, especially when domestic innovators would like to launch their products in India.

Mitigating the Issues of Liability and Bias in Context of Gen AI

Liability

Gen AI systems have become advanced enough that they undertake independent decisions using the knowledge learned by themselves, very much like human intelligence. Thus, there is a need to identify whether Gen AI systems may be held responsible for civil liability in the long term. In view of this, the question arises whether an AI system may be recognized as a legal person⁴⁰. Here, where liability is very clearly identified, the scenarios of risk may be much lower. However, there is much debate between stakeholders on determining the legal status of AI systems.

One side of the debate is conferring a separate legal status to AI systems determined by the level of autonomous decision-making and “intelligence” of that AI system⁴¹. The downside of bestowing AI systems with a separate identity may lead to a strict product liability regime, which may discourage innovation. An example of this conflict is the use-case of driverless cars, wherein users have been found to be met with accidents, leading to a question of liability. Another solution may be the distribution of this liability amidst the developer of technology, the owner of the car and any third-party user/ operator⁴².

It requires clear regulations that can address the issue of accountability of AI systems. This issue may

be resolved either by upgrading existing laws to encompass AI, or by providing a separate legislation to specifically address the legal aspects of the development and deployment of AI systems. Certain countries like China, believe in prior regulation, wherein algorithms are registered before being permitted for distribution in the country.

Mitigating risks of bias

Another pertinent issue existing in AI systems is the existence of bias in algorithms either arising out of the data provided in the AI system or the bias in the developers. This bias may lead to faulty results, especially if Gen AI is deployed as a public tool, or used in high-impact industries such as Finance or Healthcare. Generalized AI standards for developers may help reduce instances of bias by encouraging diversity in the sources of data used, scrutinising initial datasets, mandating continued testing and encouraging feedback, while also supporting research on practical techniques of promoting fairness⁴³. These methods are largely self-regulatory in approach and may not require new legislations.

In the absence of encoded laws, regulations, statutory rules, or guidelines pertaining to AI, there is some regulatory uncertainty. But based on the announcements made by the government and more recently the Indian Minister of State for IT, the government clearly indicated its intent to regulate AI to ensure AI protection⁴⁵. It was also indicated that a risk-based approach would be taken up to protect “Digital Citizens” from harm.

Clarity needed over IP rights

Regulatory infrastructure required, for AI systems to flourish in India and to be of public use, requires an emphasis on promotion of innovation in the country while balancing the risks that arise as Gen AI systems gain traction. One pertinent gap existing in the AI regulatory infrastructure needed for promoting innovation is the clarity on intellectual property (IP laws) which apply to patenting of algorithms and granting inventorship⁴⁶. This would significantly impact the ability of small players to develop applications and scale up. ●

³⁹ Regulatory Sandbox Explained: How RBI is moderating FinTechs' disruption in BFSI; Economic Times; <https://bfsi.economictimes.indiatimes.com/news/policy/regulatory-sandbox-explained-how-rbi-is-moderating-fintechs-disruption-in-bfsi/87098591>

⁴⁰ Reserve Bank of India: <https://rbi.org.in/scripts/PublicationReportDetails.aspx?UrlPage=&ID=920>

⁴¹ Report Of Committee - D On Cyber Security, Safety, Legal and Ethical Issues, July 2019, MEITY

⁴² <https://indiaai.gov.in/ai-standards/civil-liability-of-artificial-intelligence>

⁴³ Regulating Artificial Intelligence - New Technology - Worldwide (mondaq.com)

⁴⁴ Report Of Committee - D On Cyber Security, Safety, Legal and Ethical Issues, July 2019, MEITY

⁴⁵ India will regulate AI to ensure user protection | Mint (livemint.com)

⁴⁶ Regulating Artificial Intelligence - New Technology - Worldwide (mondaq.com)

Acknowledgements



Steering Committee

- ▶ Rajiv Memani
- ▶ Mahesh Makhija
- ▶ Dr DK Srivastava
- ▶ Rajnish Gupta



Core Group

- ▶ Gaurav Goyal
- ▶ Uday Gupta
- ▶ Aditi Gokhale
- ▶ Prosenjit Datta



Enterprise Agenda

- ▶ Gaurav Goyal
- ▶ Burgess S Cooper
- ▶ Abhishek Sen
- ▶ Rohit Pandharkar



Macroeconomic Insights and Projections

- ▶ Dr DK Srivastava
- ▶ Ragini Trehan
- ▶ Tarrung Kapur
- ▶ Dr Muralikrishna Bharadwaj



Sector Specialists

Financial Services

- ▶ Pratik Shah
- ▶ Bhavin Sejpal
- ▶ Bhargavi Sunkara
- ▶ Rajorshi Chanda
- ▶ Sheetal Disale

Retail

- ▶ Digvijay Ghosh
- ▶ Ram Deshpande

Healthcare

- ▶ Kaivaan Movdawalla
- ▶ Srimayee Chakraborty
- ▶ Ankur Dhandharia
- ▶ Srabati Nandy
- ▶ Eisha Anand
- ▶ Nisha Sharma

Life Sciences

- ▶ Suresh Subramanian
- ▶ Shobhna Mishra
- ▶ Rajni Sadana

Technology Services

- ▶ Nitin Bhatt
- ▶ Sachin Tyagi

Media and Entertainment

- ▶ Ashish Pherwani
- ▶ Shubh Mittal
- ▶ Devanshu Tiwari
- ▶ Aswathy John
- ▶ Kumar Kislai
- ▶ Lalit Verma
- ▶ Danush Dumasia

Government and Public Services

- ▶ Rahul Rishi
- ▶ Honnur Muralidhara
- ▶ Ashu Malik



Policy Agenda and Recommendation

- ▶ Rajnish Gupta
- ▶ Ankan De
- ▶ Shambhavi Sharan
- ▶ Vipul Gautam



Editorial

- ▶ Radhika KTP
- ▶ Kaveri Nandan
- ▶ Vikram D. Choudhury



Design, Layout and Infographics

- ▶ Ashish Kuttickal
- ▶ Ridhi Sharma Kapuria
- ▶ Ritika Saini



Generative AI Image Artist

- ▶ Cover and chapter images by Tapan Aslot

Proud to be part of
the face of the future

ey.ai unifying
platform

www.ey.com/en_in/ai/platform

Ernst & Young LLP

EY | Building a better working world

EY exists to build a better working world, helping to create long-term value for clients, people and society and build trust in the capital markets.

Enabled by data and technology, diverse EY teams in over 150 countries provide trust through assurance and help clients grow, transform and operate.

Working across assurance, consulting, law, strategy, tax and transactions, EY teams ask better questions to find new answers for the complex issues facing our world today.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via ey.com/privacy. EYG member firms do not practice law where prohibited by local laws. For more information about our organization, please visit ey.com.

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 - 122 003, Haryana, India.


© 2023 Ernst & Young LLP. Published in India.
All Rights Reserved.

EYIN2312-012
ED None

This publication contains information in summary form and is therefore intended for general guidance only. It is not intended to be a substitute for detailed research or the exercise of professional judgment. Neither EYGM Limited nor any other member of the global Ernst & Young organization can accept any responsibility for loss occasioned to any person acting or refraining from action as a result of any material in this publication. On any specific matter, reference should be made to the appropriate advisor.

RS1


ey.com/en_in

 @EY_India

 EY

 EY India

 EY Careers India

 @ey_indiacareers