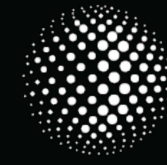


Generative AI Startups: Landscape & Trends



THINK TANK
BY LAB45

Gen AI is disrupting the global economy. This disruption creates opportunities for startups and prompts major IT players to adjust their strategies.

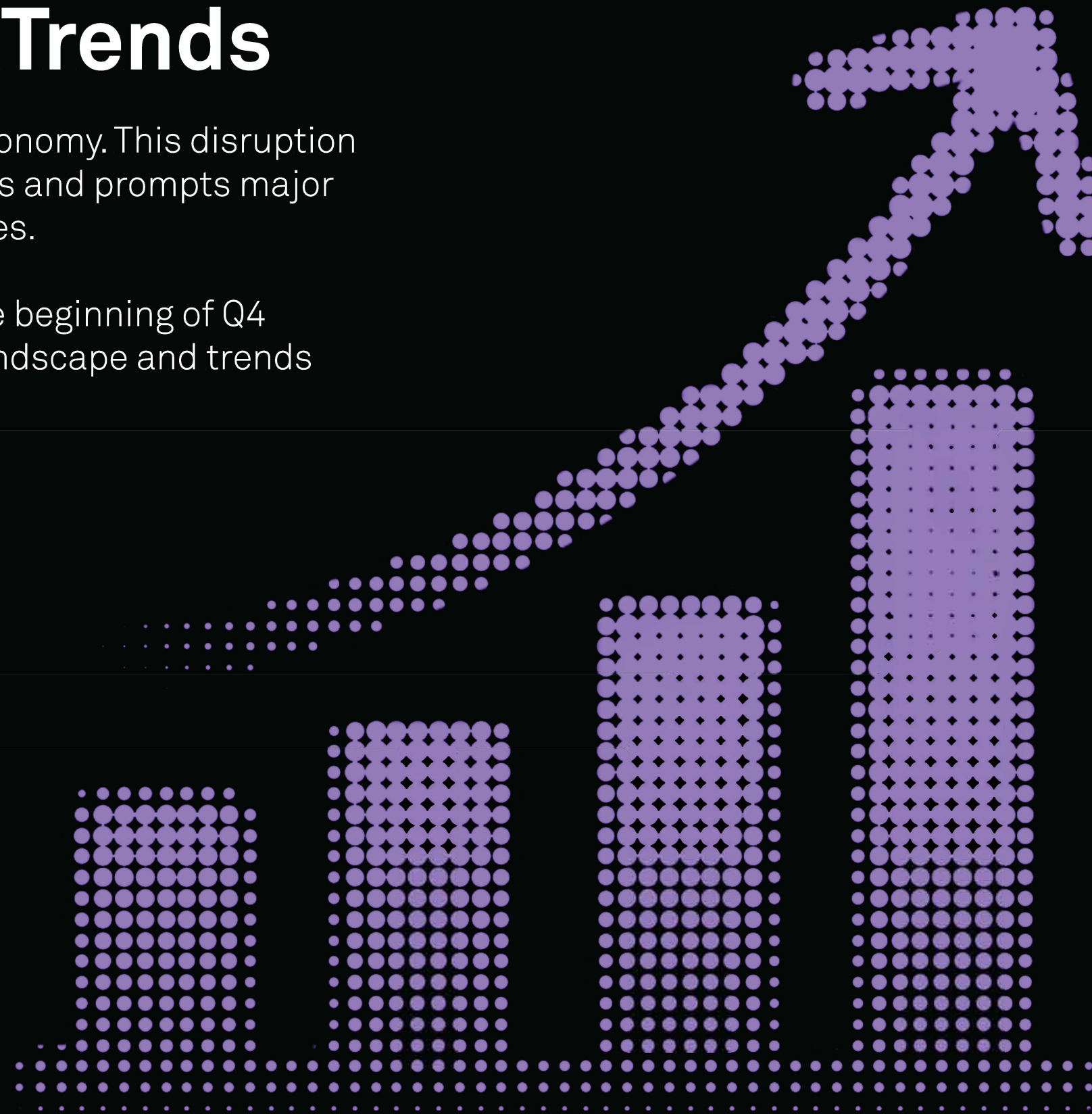
This primer was completed at the beginning of Q4 2023, documenting the Gen AI landscape and trends present at this pivotal moment.

The Rise and Impact of Generative AI

The Technology and Business Stack of AI

Business Niches

Future Trends in Gen AI



01
INTRODUCTION

The Rise & Impact of Generative AI

Artificial Intelligence (AI) has advanced significantly in recent decades, primarily focusing on specific quantitative tasks. That said, due to its high input cost, AI has remained largely niche and confined to a rather narrow spectrum of use.

The Rise of GenAI
The Rise of GenAI
The Rise of GenAI
The Rise of GenAI
The Rise of GenAI
The Rise of GenAI
The Rise of GenAI

Historically, AI model use has relied on three types of inputs: a skillful team that understands how to best develop and deploy the model, a correct and high-quality dataset that can inform the model, and a clear and quantitative question format that can extract value from the model.

■ Skillful Teams

Usually consist of data scientists, machine-learning (ML) experts, and cloud engineers, covering the model's data interpretation, learning, data access, and distribution capabilities.

■ Quality Datasets

Rely on vast amounts of structured numerical data from a high-trust source. This is more challenging to access than it may sound, as Gartner^[1], notes that 80% of enterprise data is unstructured, making it unsuitable for traditional AI.

■ Clear Questions with Measurable Outcomes

Precisely formulated and quantifiable questions equipped with specific metrics to evaluate the merit or excellence of expected responses are even scarcer outside specialized business domains.

Given the scarcity of these three inputs, the high cost of creating specialized ML models has mainly confined AI to number-crunching or otherwise highly quantitative roles, leaving most human activities largely untouched.

The rise of Generative AI (GenAI), stemming from the evolution of large language models (LLMs), has revolutionized AI's interaction with both data and questions.

GenAI enables AI to understand conversations, summarize texts, interpret requests, and produce content with human-like creativity. Essentially, AI can now handle tasks commonly done by corporate employees, such as content creation, emails, and chat.

Now, AI is capable of interpreting natural language rather than only numerical text, and questions that may not be as quantifiable or well-defined as was previously necessary.

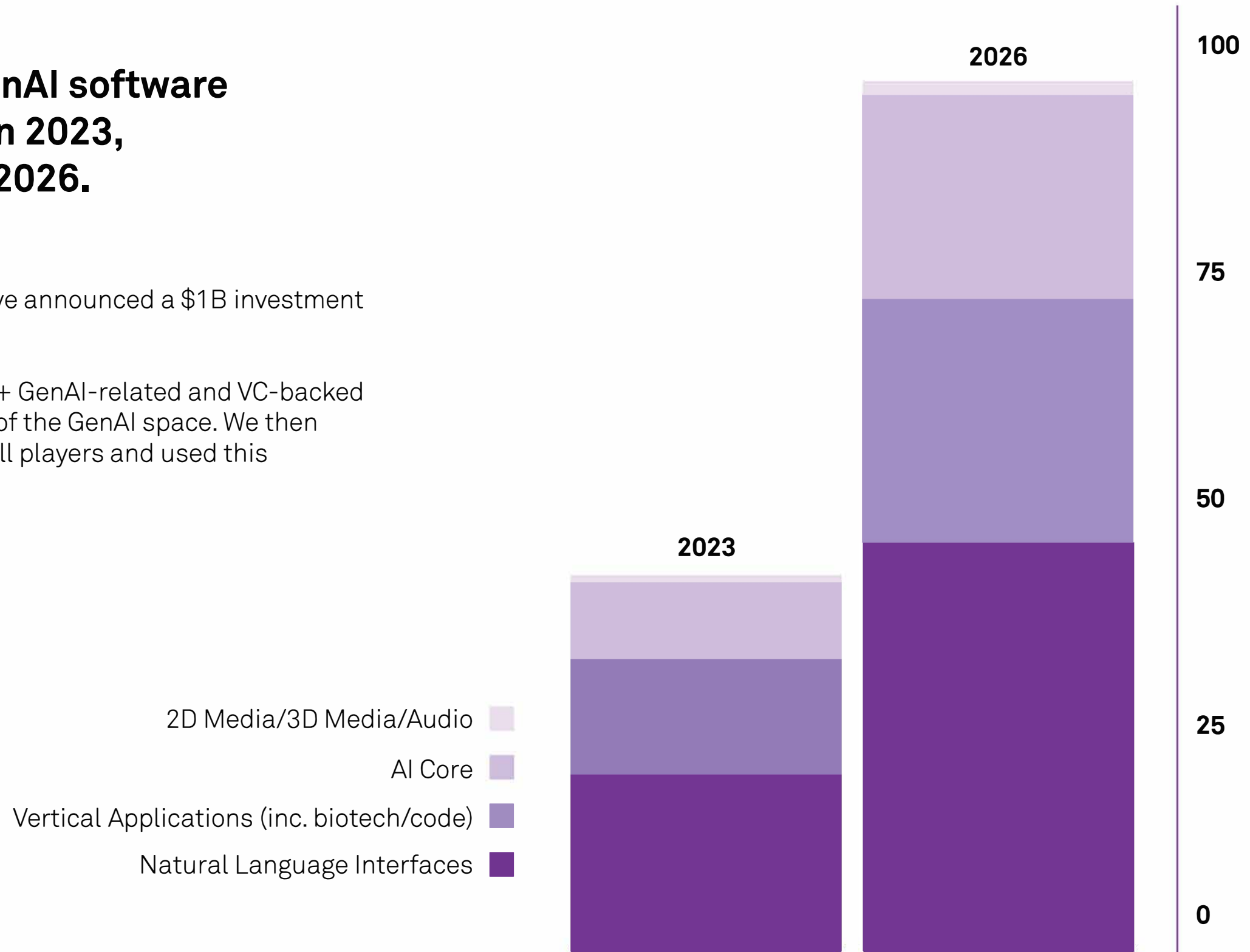
This shift is already disrupting areas previously untouched by AI. While some advancements empower workers with improved tools and enhance productivity, others threaten job security. This disruption indicates an economic shift, with capital and attention flowing into new markets. This will create opportunities for startups and prompt major IT players to adjust their strategies.

GenAI market size is estimated to more than double in the next 3 years

Reports suggest that the GenAI software market could exceed \$40B in 2023, growing to nearly \$100B by 2026.

This includes companies such as Wipro, who have announced a \$1B investment in AI capabilities over the next three years^[3].

In this primer we leveraged our knowledge of 50+ GenAI-related and VC-backed startups to reconstruct the technological stack of the GenAI space. We then noted the areas occupied by both large and small players and used this understanding to project future industry trends.



The GenAI Tech Stack

The GenAI Tech Stack

The GenAI Tech Stack

The GenAI Tech Stack

The GenAI Tech Stack

The GenAI Tech S

The GenAI Tech

02

The Technology & Business Stack of GenAI

Large tech companies are leveraging their existing technological and capital advantages to create the framework for the GenAI market landscape.

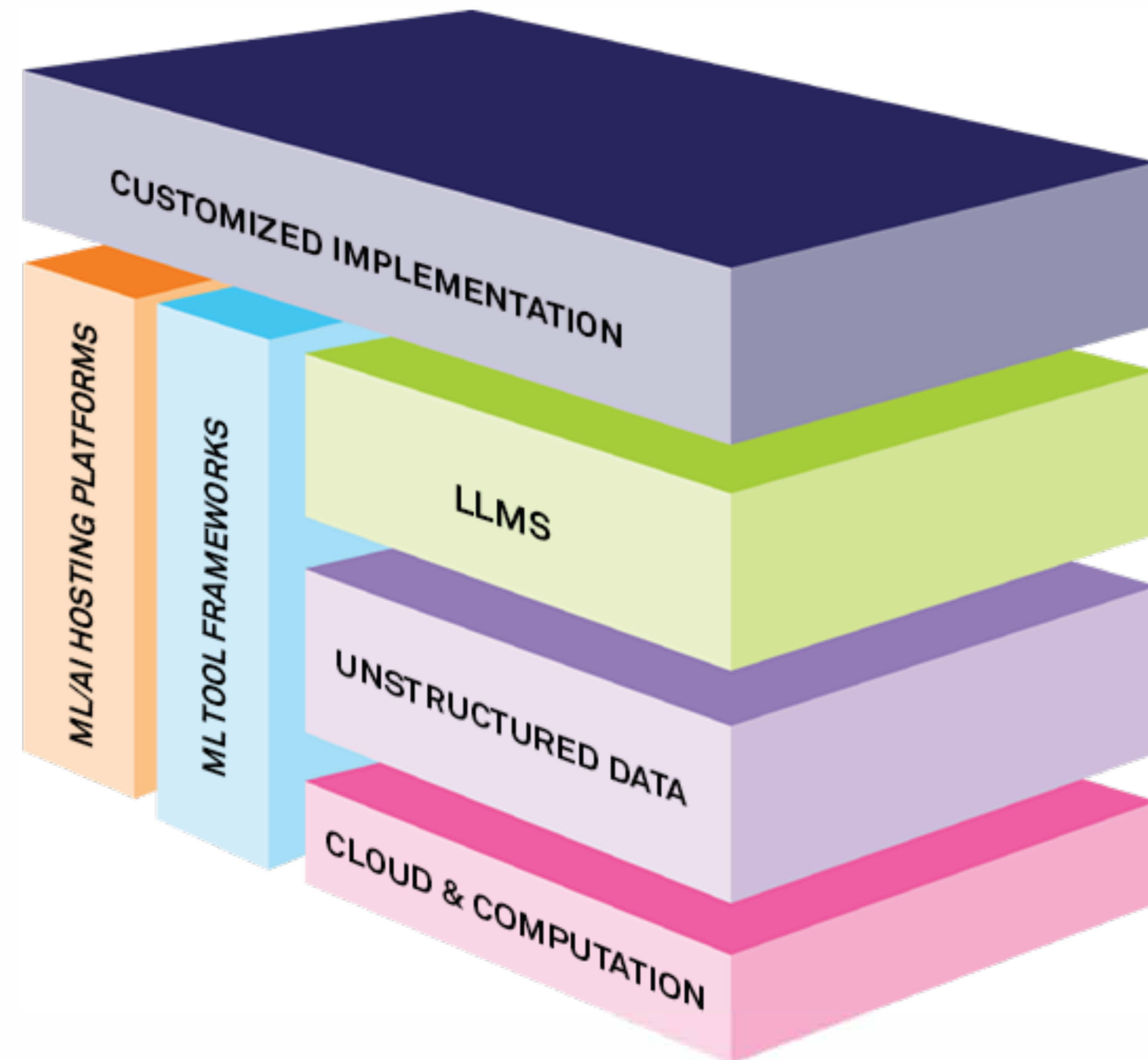
The GenAI Tech Industry Landscape: A High-Level View

While offering of the LLMs on the current scale and heavy focus on unstructured data are somewhat new, the other elements of the tech stack closely mirror those needed for any large computational modeling.

Established companies in the field of traditional AI are at an advantage, as they can expand and repurpose preexisting software, infrastructure, and services.

The figure represents the coarse technological and business landscape of GenAI

Technology Stack of GenAI



Cloud & Computation Infrastructure

Achieving optimal performance in GenAI necessitates the utilization of extensive computational resources. This demand has highlighted the pivotal role of cloud storage and GPU capabilities.

Example(s): OpenAI's GPT-4 was trained using 10,000 Nvidia GPUs, consuming approximately 7.5 megawatt-hours (MWh) of energy[4]. Dominant players like Azure, AWS, and Nvidia helped create the infrastructure necessary.

ML Tool Frameworks

While LLMs are a newer phenomenon, resource-intensive neural network ML/AI models aren't. Fortunately, tools designed for the development and deployment of neural networks are transferable to LLMs. These tools also manage extensive data and GPUs for training, as well as significant virtual memory and processing for hosting.

Example(s): TensorFlow and PyTorch are continually adapting and scaling their architectures and tools to accommodate the growing demand and complexity of Large Language Models (LLMs).

Unstructured Data

The training of Large Language Models (LLMs) necessitates vast volumes of textual data. Consequently, providers offering data storage solutions and open datasets apt for training emerge as key stakeholders in this space.

Example(s): GPT-4 was trained on a 1-petabyte total dataset. Datasets used included Common Crawl, RefinedWeb, Twitter, YouTube, Wikipedia, textbooks, and others[5].

ML/AI Hosting Platforms

For professionals seeking streamlined solutions reducing the needs for specialized skills there are platforms that simplify the training, deployment, and hosting of large ML models. As LLMs gain traction, many of these platforms are considering expanding their offerings to support them.

Example(s): Microsoft's cloud platform Azure hosts GPT-4 and provides a suite of AI-friendly services to enable AI models, including AI frameworks, cloud deployment, and data privacy tools.

LLMs

The financial commitment for training foundation models like GPT-4 is substantial, approximating \$100 million^[6]. Such an investment poses challenges for smaller startups, creating an advantage for well-capitalized entities.

Example(s): Household names like Google (PALM2), Meta (LLaMA), IBM (WatsonX) are making major investments in foundation LLMs. A bright constellation of Google spin-off major startups like OpenAI (GPT4), Anthropic (Claude 2), DeepMind (Gemini) are disrupting and reshaping the landscape.

Customized Implementation

Ultimately, the objective extends beyond simply training and hosting a model; it's about creating tangible business value.

This demands deep integration of AI technology into core business systems.

Example(s): Major IT service firms, such as Wipro, with experience in AI/ML integration, are branching out to offer tailored LLM solutions as well.

03

Business Niches For GenAI Startups

While large players are occupying a sizable portion of the GenAI tech stack, there remains more than enough room for GenAI startups to flourish.

GenAI S
GenAI Startups
GenAI Startu
GenAI Startups
GenAI Startup

GenAI startups may be placed into four categories: Add-on, Horizontal, Vertical, Service.

01

Add-On

Overview: Add-on startups work in synergy with existing platforms, extending or enhancing their functions.

Example: A startup might develop a plugin that allows developers to integrate LLMs into their own apps more easily. Another might make an LLM's output more structured or customized.

Opportunity: With major players focusing on core functionalities, these startups can address niche requirements and enhance usability or app interoperability.

03

Vertical

Overview: They aim for comprehensive solutions tailored for specific market segments or industries.

Example: An LLM trained specifically for legal documentation or medical diagnostics.

Opportunity: By understanding the unique needs and nuances of a segment, these startups can create highly specialized products that outperform generic solutions.

02

Horizontal

Overview: These focus on narrow functionality for a broad use, aiming to be best in class for a particular feature.

Example: A startup that creates managing software for prompts or focuses on fine-tuning LLM models with the user's feedback.

Opportunity: Becoming the go-to solution for a specific function can make these startups indispensable within a given ecosystem.

04

Service

Overview: These startups address the logistical, procedural, and process-oriented challenges of working with LLMs.

Example: Platforms that streamline the training process of LLMs or marketplaces where developers can find tools or experts.

Opportunity: As LLM adoption grows, so will the demand for services that simplify or optimize their implementation and maintenance.

The GenAI Tech Industry Landscape: A Detailed View

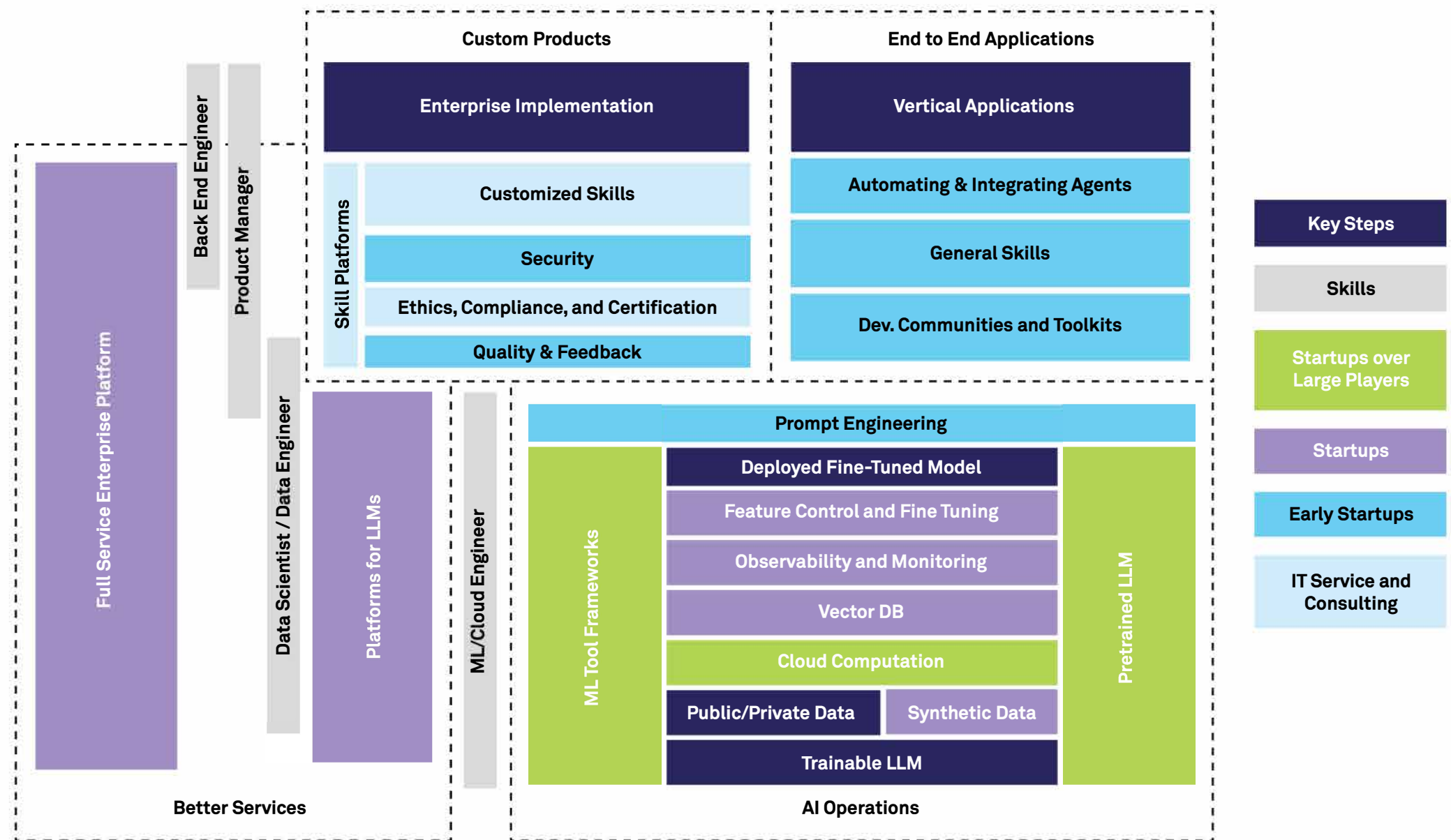
Bottom-Right (AIOps): Here, startups may offer tools for easier adoption of LLMs, facilitating the initial process of customizing and implementing these models.

Ascending (Integration): Moving upwards represents the process of integrating LLMs into various applications and business operations. Startups could offer integration services, templates, or frameworks to streamline this, or build an entire end-to-end app for a selected market niche.

Moving Left (Service Platforms): As we move leftwards, the focus shifts from core LLM functionality to auxiliary services. This could range from platforms offering specialized training data, to marketplaces for LLM apps, to optimization tools. These firms may automate the need for certain experts.

The landscape of AI and ML is continuously evolving, with new startups, technologies, and methodologies emerging regularly.

Visualizing the technological & market landscape



This taxonomy can serve as a foundational overview for anyone looking to understand the current state of the LLM ecosystem.

AI Operations

Pretrained LLM

Platforms that offer pre-trained models for users. Since training LLMs requires vast amount of the resources, the providers of pre-trained models are either established IT companies or major startups

Major players: Google, Meta, IBM, Amazon, etc.

Major startups: OpenAI, DeepMind, Anthropic, etc.

Trainable LLM

Providers of LLMs that allow users to access LLM parameters for fine-tuning. There are also a few startups and research teams that create and share open source and untrained LLMs.

Model examples: Meta, Hugging Face, Databricks.

Cloud Computation

Necessary for training large models and hosting them. Cloud storage stores model and training data and GPUs provide computational power. This area is mostly owned by large companies, but there do remain some niches for startups.

Cloud storage: AWS, GCP, MS Azure, etc.

GPU providers: NVIDIA, AMD GPUs, Google TPU, etc.

Niche startups: run:ai, OctoML, etc.

Public/Private Data

Essential for custom training of LLMs. Training custom LLMs requires substantial amounts of data. Depending on the specialization of the LLM, the data may be public, private, or both. The data must be collected, converted into a suitable input form, and labelled for quality control.

Providers: Snorkel AI, Scale AI, etc.

Vector DB

Training LLMs requires conversion of data into the vector format and storing the data in a searchable vector database.

Databases: Qdrant, Pinecone, Vald, etc.

Search Engine: Typesense.

Synthetic Data

If private data is too sensitive to be used for training or querying LLM, synthetic data can be generated and used in its place. Existing LLMs in a secure environment can be used to produce synthetic data.

Startups: Tonic, Mostly, Hazy, etc.

Observability & Monitoring

To control cloud/computational resources and oversee LLM's performance and resource usage.

Providers: Helicone, Arize, etc.

Feature Control & Fine Tuning

Tracking, systematizing, and fine-tuning model outputs based on user feedback and specialized datasets.

Providers: Autoblocks.

Prompt Engineering

The rise of the LLM has created a market for prompt engineering. Prompt engineering enables users to craft specialized prompts or prompt-chains that can extract specific outputs from LLMs. Startups and independent professionals are quite active in this space.

Startups for prompts: Promptbase, Promptlayer, etc.

ML Tool Frameworks

Already existing major portals, providing comprehensive set of libraries and ability to programmatically request and integrate all the required tools and resources for LLMs. New tooling open-source frameworks, specifically focused on LLMs, are also appearing, providing comprehensive toolsets for LLM management.

Major platforms: Hugging Face, Google AI, TensorFlow, etc.
New frameworks: LangChain.

Deployed Fine-Tuned Model

Hosting of controlled and customized environments that serve fine-tuned models.

Providers: Paperspace, Gradient.

Services

Platforms for LLMs

Service platforms that simplify the training, testing, tuning, hosting and deployment of LLMs. These services relieve the customer of the necessity of expertise in cloud, computational resources, and vector databases.

Providers: Qwak, Lamini.

Full-Service Enterprise Platforms

Large-scale service platforms allow customized guidance for an enterprise. Services can cover an entire stack, from an enterprise data lake to internal or external LLM-based functionalities, taking over the functions of data scientists and data engineers, also simplifying the further management and integration.

Providers: Scale AI, Nvidia NeMo, DataRobot.

Custom Products

Quality & Feedback

The performance of deployed LLM outputs can be tested and validated automatically and/or via user feedback.

Automatic Validation: Determined AI.
User validation: Humanloop, Honeyhive.

Skill Platforms

IT services companies are packaging customized functionality into platform repositories and offering them to a broader set of clients.

Providers: Wipro Lab45.

Enterprise Implementation

Large IT service providers are integrating newly received functionalities into the backend of existing businesses and enterprise software platforms.

Providers: Wipro.

Security

LLMs may expose security risks through exposure of sensitive data, either from custom training or via query-testing. Some startups are addressing this risk with data tracking and query handling.

Providers: Cadea, Opaque.

Customized Skills

The end-value of LLMs for business is custom LLM-based functionalities that end-customers can benefit from. Functionalities can be developed for the clients by highly trustworthy IT consultants.

Providers: Wipro, Topcoder

Ethics, Compliance & Certification

Regulators may introduce sets of requirements regarding LLM output and underlying data handling. Social pressure may also introduce ethical standards. As the power and impact of GenAI increases, corresponding regulations and ethical mores are expected to grow in tandem. This will make AI ethics, compliance, and certification a rapidly growing niche. The subject is currently highly debated. However, there are already some startups creating tools to enable more ethical AI models, and others that are certifying LLMs that generate synthetic data for certain verticals.

Ethical AI tool creators: Arthur, EthicsGrade.
Data certifiers: MDClone (healthcare-focused), Mostly AI, Getsubsalt.

End-to-End Apps

Dev. Communities and Toolkits

Developer communities and toolkits can help enable LLM-based app development through resources, guidance, tools and templates.

Providers: Metal, LangChain.

General Skills

Developers, startups, and large players provide specialized libraries and autonomous functionalities that can be used as plugins for broader applications. These include language translation, document summarization, specialized templates generation, and others.

Large providers: DocuSign, Google Template, OpenAI.

Small startups: Anarchy, Viva Translate, Dubverse, FollowFox.

Automating and Integrating Agents

A higher level of application layer provides tools that can use LLM capabilities to automate existing processes or better integrate them into existing apps and ecosystems.

Providers: AutoGPT, JARVIS, BabyAGI, AgentGPT, Steamship, KLU.

Vertical Applications

Developers can create interface and management layers to convert self-developed or available LLM-based skills into an end-to-end application, targeting a selected market niche. Due to the closest proximity to a consumer, this is the most diverse and attractive niche for startups, while the overhead of supporting full technological stack often becomes substantial. Below are some examples of startups in that niche.

Providers: Aikenist, PatentPal, Skeep, BHyve.

04

Future Trends in GenAI

Six trends will define the future of the AI landscape:

- Consolidation of Major Players
- Rise in Open-Source Adoption
- Surge in Service Platforms
- Expansion of Skill Marketplaces
- Segment-Specific Applications
- Regulatory Oversight and Standardization

The Future of GenAI
The Future of GenAI
The Future of GenAI
The Future of GenAI
The Future of GenAI
The Future of GenAI
The Future of GenAI

01

Consolidation of Major Players

As leading entities expand, they are likely to absorb many of the functionalities currently offered by startups in niche areas. These functionalities could either be acquired or independently developed, potentially rendering certain startups obsolete. Nevertheless, some niche offerings might find a home within the open-source community.

02

Rise in Open-Source Adoption

Given the transformative potential of GenAI in the global market, there will be a significant push towards democratizing access to LLMs, associated tools, and skills. Although independent entities and open-source communities may not have the resources to train models as expansive as ChatGPT4, they could potentially secure funding through philanthropic endeavors or fundraising campaigns.

03

Surge in Service Platforms

The evolution of LLM-driven capabilities will become more streamlined. Startups that currently specialize in singular tools or functions might either be assimilated into comprehensive platforms or risk becoming redundant.

04

Expansion of Skill Marketplace

Existing skills will increasingly be accessible through skill marketplaces and open-source platforms. Startups operating these marketplaces are poised to prosper initially and may eventually undergo consolidation.

05

Segment-Specific Applications

Applications will increasingly be tailored to cater to specific market segments, taking into consideration unique data needs and performance criteria.

06

Regulatory Oversight & Standards

As the GenAI sector burgeons, there will be concerted efforts from various stakeholders to institute regulations. These could encompass data management, response constraints, fact accuracy, partnership modalities, and more. While this regulatory space is in its infancy, its significance is projected to grow immensely, potentially becoming a cornerstone of the industry.

Appendix 1: Reality & Expectations of GenAI

While enhancing the users with great capabilities, the LLM-based service is neither a freebie, nor a cornucopia. Each implementation of LLMs carries its own advantages and downsides. In this section of the Appendix, we discuss what can and cannot be realistically expected from a GenAI model in each of the most popular use cases.

Pre-trained LLMs (GPT) in a basic configuration

The most popular use case is when LLM is being trained only once by the LLM provider, and then – accessed by multiple users via API services. An example could be the GPT-4 model in its primary configuration, without extended functionalities. Let's delve into the benefits and limitations of a one-time trained LLM from a business perspective.

The advantages of such case include:

- 1. Reduced Cost:** Running a single version of LLM with the preset parameters is highly beneficial from reducing the training and management costs. And reduced costs – allow the model provider to train a larger model with superior performance.
- 2. Ease of Use:** Pretrained LLMs offer users a streamlined experience due to their straightforward setup and deployment. The user interaction is primarily limited to providing input to the model and, if necessary, refining the output. What relieves the user from the need of specialized skills in order to use it to the full capacity.

The expected disadvantages of once-trained LLM include:

- 1. Loss of Facts:** While the most advanced LLMs might be trained on datasets as vast as 1 Petabyte of top-tier textual content, this doesn't imply the model can directly access or reproduce all that information upon request. Contrary to some assumptions, the model won't pinpoint or recall specific data sources. It responds based on a correlation with its 1 trillion parameters. Essentially, we're querying the model's parameters, not the underlying data.
- 2. Data Loss:** For perspective, GPT4's model parameters alone might occupy around 1 Terabyte of storage space. This is a thousand-fold smaller than the original dataset's size. Thus, it's implausible to assume that this 1 Terabyte of trained parameters can recreate the entirety of the original 1 Petabyte dataset without any data loss.
- 3. Lack of Adaptation.** Since these models aren't designed for retraining, users can't modify the model's parameters directly. Instead, any adaptation relies on reintroducing prior user inputs within a session to subtly influence subsequent responses and is usually getting lost in between the sessions.
- 4. Obsolescence:** As time goes by, the information loses its relevance to the current moment. And the same are the model insights drawn from parameters, determined at the time moment of the model training.

There are three common ways of using trained LLMs:

- 1. Informational:** Can the model still generate facts? Yes, if the facts are cited enough in the training data to be imprinted in the model parameters. Prompting for “common sense” facts is one of the most common methods of using ChatGPT4. It may not be reliable, but it is convenient.



- 2. Reflectional:** Users can provide an extended input to an LLM and specify a desired length and style for the textual output. If the model were human, this would be akin to seeking an opinion. However, since an LLM's response must align with the entirety of the input rather than select portions, the result is often a succinct summary that resonates most with conventional wisdom. Notably, these summaries are proving valuable and quickly growing in popularity. Another very practical use case here is when the input settings are selected to preserve at maximum the text in the input but return it with a primer for a different style or even a language, proving value in the areas of translation and reformulation. "Translate it to a language" or "Explain it to a fifth grader" - are among the most popular requests to ChatGPT.



- 3. Generational:** Finally, one can ask an LLM to produce a lengthy output from a short input. This translates an initial thought into a good-looking story, poem, or a speech template, already saving vast amounts of time to content creators.

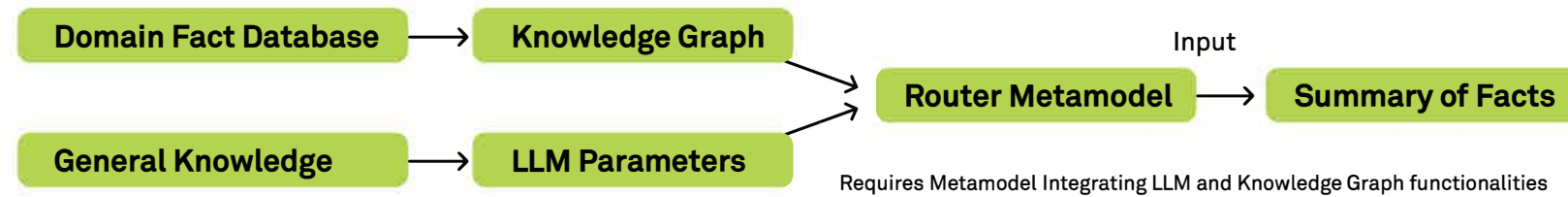


In essence, while pretrained LLMs offer significant advantages in terms of simplicity, it's crucial to understand their inherent limitations and the nuances of their data representation.

Integrating LLMs with a Knowledge Base

The data-related constraints of LLMs (1. Loss of Facts and 2. Data Loss) have prompted efforts to integrate them with genuine knowledge databases. When ChatGPT provides precise literature quotations or in-depth explanations of specific phenomena, it's not because such information is inherently stored in its parameters. Rather, certain queries are directed to structured databases, and the information from there is then seamlessly blended into ChatGPT's response. While this makes LLM-based engine much more powerful, it also increases its complexity, requiring managing structured and unstructured components of the engine in a cohesive way.

Multi-Domain Expert (Precise)

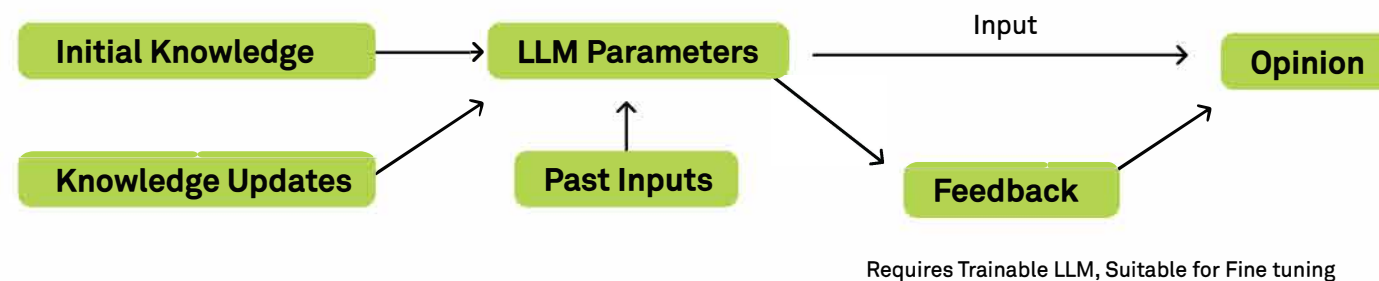


Sometimes, when we believe that we are testing the knowledge of GPT, we are in fact only searching for a database that GPT deemed to be most relevant - and then happened to formulate in a neat sentence.

Fine-tuning retrainable LLMs (LLaMA)

The inflexibility constraints of LLMs (3. Lack of Adaptation and 4. Obsolescence) can be addressed by the models that enable users to access and adjust their parameters. Such models can be trained on private or industry-specific data and be consistently updated as new data becomes available. They can also be refined based on previous inputs and user feedback, progressively adapting to the user's preferences. However, these benefits are accompanied by increased architectural complexity and the need to manage multiple tailored versions of the model.

Intelligent (Adopt)



Custom-training and fine-tuning of an LLM requires providing data, acquiring computational resources, managing vector databases, supporting quite complex architecture for model training and maintenance, and overall is quite an expensive and time-consuming process.

Fortunately, service platforms are rapidly evolving to support the fine-tuning needs, and advantages of fine-tuned models, such as being up to date, allowing domain expertise, or being highly personalized to the customer and trainable for any needs, are appealing.

Appendix 2: Startups across the GenAI Tech Stack

| | Primary Tech Niche | Startup Type | Short Description | Size |
|------------------------------|---------------------------------------|--------------|---|------|
| Steamship | Automating & Integrating Agents | Vertical | An agent/engine allowing production of private AI-bots for Web/Telegram. | 8 |
| KLU | Automating & Integrating Agents | Horizontal | Search engine across many apps. | 7 |
| AgentGPT (Reworkd AI) | Automating & Integrating Agents | Service | Autonomous Agents deployment tool. | 3 |
| AutoGPT | Automating & Integrating Agents | Vertical | ChatGPT adaptation, specialized on content creation. | 2 |
| BabyAGI | Automating & Integrating Agents | Vertical | Autonomous AI Agent, converting input into a list of tasks, and execute them. | 1 |
| run:ai | Cloud Computation | Add-on | Optimizes access to computational resources | 122 |
| OctoML | Cloud Computation | Add-on | Optimizes running LLMs in cloud | 109 |
| SuperU | Customized Skills | Vertical | Data Analytics. Customer Data insight engine. | 43 |
| Dust | Customized Skills | Vertical | Communication/Content Generation tool tuned for Enterprise. | 16 |
| LlamaIndex | Customized Skills | Service | Platform to connect the customer data with LLMs. | 5 |
| Meru | Customized Skills | Vertical | Intelligent Search/Communication Agent on Enterprise data. | 4 |
| Paperspace | Deployed Fine-Tuned Model | Service | Platform for development, deployment and hosting of AI models | 39 |
| Gradientj | Deployed Fine-Tuned Model | Service | LLM application builder. | 2 |
| Metal | Dev. Communities and Toolkits | Service | Self-service platform for Developers, focused on LLM applications. | 23 |
| Replicate | Dev. Communities and Toolkits | Service | Simplifies/unifies the LLMs coding for customers. | 17 |
| FlowGPT | Dev. Communities and Toolkits | Add-on | Improved interaction interface with ChatGPT. | 15 |
| Synativ | Dev. Communities and Toolkits | Vertical | Toolbox and LLM copilot for visual applications. | 5 |
| Berri | Dev. Communities and Toolkits | Add-on | Padding layer for OpenAI API with errors reduction. | 3 |
| Stack AI | Dev. Communities and Toolkits | Add-on | Self-service tool. Building a personal LLM app – like dragging Lego-pieces. | 3 |
| Baseplate | Dev. Communities and Toolkits | Add-on | Tool. Easy connection to LLMs, prompting with client's data. | 2 |
| Arakoo | Dev. Communities and Toolkits | Add-on | Code for easy deployment and fine-tuning of LLMs. | 2 |
| re:tune | Dev. Communities and Toolkits | Add-on | Self-served platform, allowing non-programmers to produce a customized chatbot. | 1 |
| AI Northstar Tech | Enterprise Implementation | Service | Custom implementation of LLMs – IT Consulting. | 4 |
| pyqai | Enterprise Implementation | Service | Customized deployment. Customer-centric. | 2 |
| Arthur | Ethics, Compliance, and Certification | Horizontal | Testing LLMs for risks and compliance | 63 |
| Autoblocks | Feature Control and Fine-Tuning | Horizontal | Feature observability and modification. | 4 |
| DataRobot | Full-Service Enterprise Platforms | Service | End-to-end Platform for enterprise GenAI applications | 1013 |
| Scale | Full-Service Enterprise Platforms | Service | Platform for Enterprise data. Tooling for Applications. | 600 |
| Dubverse | General Skills | Vertical | Translates language in video clip into 30+ different languages. | 19 |
| Viva Translate | General Skills | Vertical | Transcribes and Translates 3 languages into subtitles. Google meets plugin. | 17 |
| Fixie | General Skills | Horizontal | Platform. Connects front-end apps via tuned LLMs. | 13 |
| FollowFox | General Skills | Vertical | Text into an Art | 2 |
| GoCodeo AI | General Skills | Vertical | Testing copilot for programming. | 1 |
| Anarchy | General Skills | Service | Cloud platform for LLM tools with democratized access | |
| LangChain | ML Tool Frameworks | Horizontal | LLM-tuned tool and programming framework. | 13 |
| Arize | Observability and Monitoring | Horizontal | Observability platform. Monitors and troubleshoots. Allows fine-tuning. | 79 |
| Helicone | Observability and Monitoring | Horizontal | Performance/Users/Resource Monitoring. Dashboards. | 3 |

| | Primary Tech Niche | Startup Type | Short Description | Size |
|---------------------------|-----------------------|--------------|--|------|
| Qwak | Platforms for LLMs | Service | End to end LLM Platform | 56 |
| Automaton AI | Platforms for LLMs | Service | LLM Data management platform, from training to hosting. | 44 |
| Lamini | Platforms for LLMs | Service | Platform. Platform for developer to build customized LLM models | 6 |
| DeepMind | Pretrained LLM | Horizontal | Foundational LLM Models - Gemini | 2012 |
| Open AI | Pretrained LLM | Horizontal | Foundational LLM Models - GPT4 | 1234 |
| Anthropic | Pretrained LLM | Horizontal | Foundational LLM Models - Claude 2 | 198 |
| Vellum | Prompt Engineering | Horizontal | Integrates prompt engineering into LLM apps. | 6 |
| Dreamboat AI | Prompt Engineering | Horizontal | Tools for prompt-engineers – prompt templates and integration. | 5 |
| Promptlayer | Prompt Engineering | Service | Platform for prompt engineers, tracking prompt usage. | 3 |
| Prompthero | Prompt Engineering | Service | Prompt and Image marketplace for AI-generated images. | 2 |
| Promptable | Prompt Engineering | Horizontal | Organizing prompt personal storage and workflow. | 1 |
| Promptify | Prompt Engineering | Service | Marketplace for prompts. | 1 |
| Promptbase | Prompt Engineering | Service | Marketplace for prompts. "E-Bay for prompts". | |
| Promptstacks | Prompt Engineering | Service | Platform for engineers and amateurs for prompt generation. | |
| Snorkel AI | Public/Private Data | Horizontal | Curated data for LLMs | 124 |
| Determined AI | Quality & Feedback | Horizontal | Training and tracking performance of Deep Learning/LLMs | 43 |
| Humanloop | Quality & Feedback | Horizontal | Platform for storing and testing custom-trained LLMs. | 9 |
| Honeyhive | Quality & Feedback | Horizontal | Collects and visualizes human feedback to ChatGPT. | 4 |
| opaque | Security | Horizontal | Preserves data confidentiality while using external LLMs | 32 |
| Cadea | Security | Horizontal | Security of LLM endpoints for enterprises | 4 |
| MDCClone | Synthetic Data | Vertical | Synthetic data for research and operations. Healthcare Vertical Specialized. | 141 |
| Tonic | Synthetic Data | Horizontal | Solution. Synthetic data from Datasets/Enterprise data. Data-Sensitive Vertical. | 104 |
| Mostly | Synthetic Data | Horizontal | Platform. Uses Generative AI to produce Synthetic data. | 69 |
| Gretel | Synthetic Data | Horizontal | Per dataset tool. Synthetic dataset generation + PII labeling. | 65 |
| Hazy | Synthetic Data | Horizontal | Platform. Controllable Synthetic data from Enterprise data. Focus on metric. | 49 |
| Mindtech | Synthetic Data | Vertical | Image and Video synthetic data. | 44 |
| Aindo | Synthetic Data | Horizontal | Synthetic data platform. Consulting and personalized data transformation. | 20 |
| Syntho | Synthetic Data | Horizontal | Self-service Platform. Synthetic data-twin. | 14 |
| Getsubsalt | Synthetic Data | Vertical | Synthetic data with the focus on legal and compliance. | 8 |
| Pinecone | Vector DB | Horizontal | Managed Vector Database for Client Data. | 85 |
| Weaviate | Vector DB | Horizontal | Vector Open-Source Database with Search Engine. | 42 |
| Qdrant | Vector DB | Horizontal | Vector Open-Source Database with Search Engine. | 33 |
| Chroma | Vector DB | Horizontal | Vector Open-Source Database. | 18 |
| Typesense | Vector DB | Add-on | Vector Search Engine. | 8 |
| Milvus | Vector DB | Horizontal | Vector Open-Source Database with Fast Search Engine. | 2 |
| Vald | Vector DB | Add-on | Distributed vector fast search engine. | 1 |
| Jasper (Jarvis) AI | Vertical Applications | Vertical | Voice assistant for a broad variety of content. | 667 |
| BHyve | Vertical Applications | Horizontal | Business Productivity - Connects Information at the Workplace | 20 |
| Skeep | Vertical Applications | Vertical | LLM assistant in online retail | 13 |
| Aikenist | Vertical Applications | Vertical | LLM Assistant for Radiologists | 12 |
| PatentPal | Vertical Applications | Vertical | LLM Copilot for Patent Applications | 3 |

*The status of startups is reported as of August 2023

Conclusion

Conclusion

Conclusion

Conclusion

Conclusion

Conclusion

Conclusion

The landscape of GenAI and LLMs presents a myriad of opportunities for startups. Whether it's through enhancing existing platforms, perfecting a niche function, providing end-to-end solutions for industries, or facilitating the development process, innovative startups can find their niche and thrive.

References

- [1] Product Manager Insight: Top 5 Trends in Unstructured Data Storage. Gartner Report, 2021
- [2] Pitchbook Emerging Tech Research, 2022
- [3] Wipro Launches Wipro ai360, Commits to Investing \$1 Billion in AI Over the Next Three Years, 2023
- [4] Exploring the Environmental Footprint of GPT-4: Energy Consumption and Sustainability. TS2 Space, 2023
- [5] GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE. SemiAnalysis, 2023
- [6] OpenAI's CEO Says the Age of Giant AI Models Is Already Over. Wired, 2023

Author

Rinat A Sergeev
 Director, Open Innovation, PhD

Two decades of experience in multi-domain Academic Research – from Semiconductors to Bio and Social Sciences, ten years of hands-on Open Innovation Management and Thought Leadership at Harvard, what included introducing NASA to external problem solvers via innovation challenge mechanics. My passion lies in cutting edge technologies, enterprise transformation, and future of workforce and human societies. Worked as Research and Data Scientist prior to Wipro.



Lab45 is a visionary space developing ground-breaking solutions to foster and accelerate ideation throughout Wipro.

At Lab45, engineers, research analysts, and scientists come together to infuse creative ways of incubating solutions for customers that will transform the future. It is a space filled with ambition at the vanguard of far-reaching research across cutting-edge technologies.

Established with the Silicon Valley culture of free-flowing creativity, Lab45's goal is to make bold ideas a reality and to invent the future of enterprise. So come, collaborate, and see what happens when ideas are left unbound.

Feedback

Click to Know More

It will take less than a minute!

Disclaimer: This report was created using various sources such as expert interviews, internet reports, website research and media releases. This information is collated in good faith and used on an as is and as available basis by us. Our reports should only be construed as guidance. We assert that any business or investment decisions should not be based purely on the information presented in our reports. Information included in the reports. We do not guarantee or take responsibility for the accuracy, completeness, reliability and usefulness of any information. The opinion expressed in the reports is our current opinion based on the prevailing market trends and is subject to change.



Ambitions Realized.

Wipro Limited

Doddakannelli, Sarjapur Road
Bengaluru – 560 035, India
Tel: +91 (80) 2844 0011
Fax: +91 (80) 2844 0256
wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading technology services and consulting company focused on building innovative solutions that address clients' most complex digital transformation needs. Leveraging our holistic portfolio of capabilities in consulting, design, engineering, and operations, we help clients realize their boldest ambitions and build future-ready, sustainable businesses. With over 250,000 employees and business partners across 66 countries, we deliver, on the promise of helping our customers, colleagues, and communities thrive in an ever-changing world.

For more information, please write to us at info@wipro.com