THE STATE OF GENERATIVE AI
INSIDERS REPORT

From Innovation to Product Commercialization in 2024

R² ACCELERATION + SAMSONPR

We're embarking on a rewrite of application software driven not only by generative models themselves, but the use of private data and build out of incredible product experiences that are not in production today.

George Mathew

Managing Director, Insight Partners
Former President & COO Alteryx

TABLE OF CONTENTS

WHAT'S NEXT FOR THE GENERATIVE AI INDUSTRY

Beyond the initial hype of early 2023, beyond the “gold rush,” business and product owners of AI initiatives are now seeking an industry-validated framework to develop their generative AI capabilities in 2024.

Those closest to generative AI’s big bang know that we have only just begun to walk down a path where every choice can branch into a number of serious implications. To help answer these questions, we went out to assemble a group of generative AI leaders to begin to determine the steps necessary to bring us toward a generative AI industry that not only generates profit but also boosts human productivity.

For the experts that attended our roundtable, it’s no surprise that the rapid escalation of the generative AI, LLM and GPT market has introduced a new set of disparate challenges - determining best use cases in a quickly evolving space, regulation and data science resources, and costs of AI deployment. Naturally, we are led to the question - what is the right product commercialization approach for success? Industry leaders are looking for ways to climb on top of this complex situation sooner than later.

The goal of this report is to provide the results and analysis from The State of Generative AI Insiders Roundtable. At this event, high-profile generative AI VCs and CEOs deeply examined these market challenges. The mission was to discover the most likely paths to a solid market performance and identify the success benchmarks for generative AI.

THE STATE OF GENERATIVE AI INSIDERS ROUNDTABLE:

On one side of the table, we had a group of CEO founders of companies who are leading the generative AI space. On the other side, we had innovative VCs who have been ahead of the curve in investing in this sector.

All of our panelists came with decades of experience, deep product operations backgrounds in AI and data infrastructure, and have each built billion dollar businesses.

Lukas Biewald
CEO & Co-founder
Weights & Biases

Florian Douetteau
CEO & Co-founder
Dataiku

Arvind Jain
CEO & Co-founder
Glean

George Mathew
Managing Director
Insight Partners

Former President & COO
Alteryx

Max Schireson
Operating Partner
Battery Ventures

Former CEO
MongoDB

> *Generative AI products are themselves changing every few weeks. Whatever we build today will most likely have to be redone three weeks from now.*
>
> **Arvind Jain**
> **CEO & Co-founder**
> **Glean**

KEY TAKEAWAYS

1. The three key tailwinds that are driving the trajectory of generative AI's development
2. Why product commercialization has to iterate 3-4 times faster
3. Reasons to throw away product fit for now
4. The emergence of a "Generative AI tax"
5. What the generative AI team looks like
6. The VC forecast moving into 2024
7. Product Commercialization Scorecard: an expert rubric for 2024 generative AI readiness

THE THREE TAILWINDS THAT ARE CURRENTLY DRIVING THE TRAJECTORY OF GENERATIVE AI'S DEVELOPMENT

We started our roundtable by asking our experts to provide their thoughts on the key tailwinds that are driving the growth of generative AI.

1 **Opening of Robust Tools That are Easy to Code and Deploy** We're embarking on a rewrite of core software principles driven not only by generative models itself, but the use of private data to build a sort of incredible experience that's not in production today. We're seeing large numbers of companies that are starting off with creating these new narrow tasks with reasoned and compelling scaled out pilots, mostly starting with some of the foundation models themselves and fine-tuning them. The volume is effectively being set and there is a tremendous groundswell that's underway.

2 **Rapid Adoption by the Developer Community** We're seeing an astonishing number of new software developers building LLM, or traditional ML models. It's not just developers with formal ML training. At this point almost any software developer can suddenly build compelling applications. Across the board, we're seeing an incredible creation of companies with previously unseen rates of adoption.

TAILWINDS EXAMPLE

Tome is a 4-person company that auto-generates a presentation, story or outline from a prompt. The company has over 10 million users just 4 months after launch.

3 **Automation of Tasks that Humans Previously Had No Choice But to Do Themselves** Today. 0-2% of tasks are actually being automated by generative AI. That will change dramatically with a potential for a double digit portion of the economy to be performed via software. Once we have real momentum here, this will be the statistic to watch using three primary metrics – economics, policy and societal impact.

PRODUCT COMMERCIALIZATION HAS TO ITERATE 3-4X FASTER

Before generative AI was born, there already was a sense of urgency when building a new software capability. However, reaching product-market fit for these applications could take considerably more time than what we are seeing in the world of generative AI.

According to the experts on the panel, there is less time to iterate in generative AI because the opportunity is both huge and comes wit high competition. An iterative mentality is even more critical, with only 2-3 opportunities to achieve a release fit versus 7-8 opportunities in the previous software worlds. This makes it critical that the initial stab is better informed, and each subsequent iteration must be relatively crisp out of the gate.

GENERATIVE AI PRODUCT COMMERCIALIZATION FRAMEWORK

Together with our guest panelists, we were able to put together an outline of the key aspects of product commercialization that businesses need to continually iterate on for a defensible generative AI use case.

PROPRIETARY DATA FLYWHEEL

Large language models

Synthetic data

Vector databases

Data clean rooms

DOMAIN-SPECIFIC WORKFLOWS

Out-of-the-box generative AI use cases

Code generation co-pilots

Prompt engineering tools

INFRASTRUCTURE TOOLS

LLMOps

Observability platforms

Model tuning and training tools

This is an accelerated movement that's occurring across the entire software landscape. And it's the top priority for just about any CEO, any head of product and any head of go-to-market.

George Mathew

Managing Director, Insight Partners
Former President & COO Alteryx

PROPRIETARY DATA FLYWHEEL

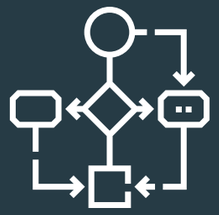
Companies should start by mapping their large language models around their proprietary data set. This data set would include user interactions, which correlations and suggestions that are surfaced by the LLM, user action from the suggestion and the resulting domain-specific business outcome.

This data is constantly leveraged to tune the LLM and creates business defensibility. This iterative process has a net positive impact on the quality of generative output - models improve over time, creates new opportunities, and enhances user conversations beyond the first interaction.

PROPRIETARY DATA EXAMPLE

BloombergGPT is Bloomberg's 50-billion parameter large language model, purpose-built from scratch for finance, using a diverse set of natural language processing (NLP) tasks within the financial services industry. This model assists Bloomberg in improving existing financial NLP tasks, such as sentiment analysis, named entity recognition, news classification, and free-form search.

Tools that can help companies with building their proprietary data flywheel include large language models, synthetic data, vector databases and data clean rooms.



DOMAIN-SPECIFIC WORKFLOWS

WITH IMMEDIATE INTEGRATIONS

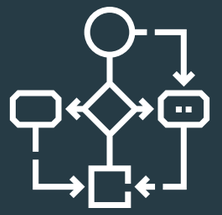
Domain-specific workflows are woven from different data clusters of user activity, resulting in a more unique human and readable experience. These new applications must provide a deep domain-specific experience and connect into the broader set of enterprise operations. It's key for these new generative AI use cases to integrate well with existing applications and other downstream systems so that they are not functional islands amongst themselves.

The workflows that offer the most opportunity in today's world include summarization, search and chat provided as a co-pilot user experience.

WORKFLOW EXAMPLE

Zoom Meeting hosts can now create a summary powered by Zoom's own large language models and share it via Zoom Team Chat and email without recording the conversation. Hosts receive automated summaries and can share them with attendees and those who didn't attend to improve team collaboration and speed up productivity.

Tools that can help companies with building their domain-specific workflows include out-of-the-box generative AI use cases, code generation co-pilots, and prompt engineering tools.



DOMAIN-SPECIFIC WORKFLOWS

WITH IMMEDIATE INTEGRATIONS

Emerging Workflows

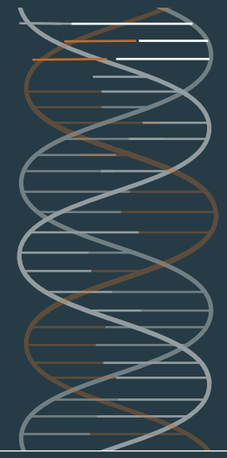
Today, workflows that offer the most opportunity include summarization, search and chat. Our panel has a close eye on two emerging workflows:

1 **Creative generative AI models integrated into existing enterprise products.** Adobe Firefly has integrated generative image fi l, text effects and 3D to image into their existing creative suite. Creators will be able to use content generated in Firefly commercially.

2 **New offerings born out of applying synthetic data to private data.** Microsoft and Epic have teamed up to leverage AI algorithms to automatically fill in missing information. Electronic health records can become more complete, accurate and easier to use, freeing up clinicians to focus on patient care.

> *A couple of years ago we wouldn't have known that there were even 500,000 people out there with the skills to build any ML model, which points to a real democratization of this technology.*
>
> **Lukas Biewald**
> **CEO & Co-founder**
> **Weights & Biases**

INFRASTRUCTURE TOOLS FOR FAST ITERATION



Our roundtable agreed that companies will need to build a new architecture around LLMs for developer use. It became clear in 2023 that companies do not even need a machine learning expert to extract the value of machine learning. Software developers can now build generative AI applications on their own by leveraging LLMs, thus bringing the cost of developing an AI application down.

LLMOps has emerged in the last few months as an evolution of existing machine learning operations (MLOps) as a fast iteration tool to enable software developers to build applications on top of LLMs. Key to LLMOps is ensuring that applications are reliable, auditable and reproducible. Developers can explore data and seamlessly experiment with machine learning building blocks.

INFRASTRUCTURE TOOLS EXAMPLE

For companies developing chat use cases, models are being iterated and deployed every week compared to previous applications being deployed every few months.

Tools that can help companies with building their generative AI infrastructure include LLMOps, observability platforms, and model tuning and training tools.

THROW AWAY PRODUCT -MARKET FIT FOR NOW

Our roundtable also agreed that businesses need to throw “product-market fit” out the window for now. We’re just at the beginning of the generative AI industry, and it’s too early to assign specific metrics for product-market fit. Especially, when businesses are still trying to add value to current workflows and finding themselves redoing their work every three weeks. Our experts cautioned that it’s a very fast moving environment and can appear to be unstable. The most interesting early indicator right now is the sheer number of active programs and initiatives that just about every company has underway.

Since it’s still not expected for applications or use cases to point to meaningful revenue at this stage, proxies (shown on the right) instead of product-market fit indicators are recommended as the measure of success.

PROXIES

Repeatable, Consistent Use Case

- Number of users accessing the use case
- Percentage of users accessing the use case for the same benefit
- Frequency of accessing the use case for a specific benefit
- Percentage of retained users accessing the use case for a specific benefit

Future Revenue

- Size of audience accessing similar use case and willingness to pay
- Size of audience that can be monetized

To What Extent Are Generative AI Features Being Used

- Size of audience accessing similar use case and willingness to pay
- Size of audience that can be monetized

Overall Experience and Productivity of People Using These New Features

- Net promoter scores
- Percentage of users that are renewing usage of application
- Percentage of users that are expanding usage of application

THE EMERGENCE OF A "GENERATIVE AI TAX"

Generative AI applications are built upon inferences, which are multiple calculations to determine a reasoned conclusion. An example of an inference would be when a user prompts ChatGPT to "write a poem about a cat." Given that this is an open-ended task, ChatGPT could approach this in multiple ways - from writing a silly, lighthearted poem about a cat playing with a ball of yarn to something more serious and introspective.

The cost of inferences needs to be scaled up in a way that will not accrue the same $4-5 million dollar per day cost as ChatGPT. All businesses are going to face this 'generative AI tax' associated with building models that are in production and serving up inference. LLMs processing costs across vast data sets, housed in both defined warehouses and data lakes, are not cheap. Even though the cost is declining, there will be larger inference precision requirements that will continue to drive the expense curve upward overall.

Simply put, software companies using LLMs will definitely see the ongoing effect of the Generative AI tax on project hurdle rates and company P&Ls. While not as immediately noticeable to date, this tax will actually increase significantly. **There are a number of strategies to keep this generative AI tax in check.** ⟶

> *One of the major inflection points was when everyone could see that there was a potential impact for their day-to-day job, or to their occupation.*
>
> **Florian Douetteau**
> **CEO & Co-founder**
> **Dataiku**

KEY STEPS TO GENERATIVE AI TAX:

1 COST OF USE CASE

First, ensure an understanding of the cloud cost required for any new generative AI use case or feature. There are three key pricing variables to monitor: the amount of data, the model being used (text, chat and summarization), and the output language (English, Spanish, Mandarin, etc.).

2 COST OF INFERENCE

Keep a careful watch on the cost of inference month over month, as pricing for some APIs will come down drastically, while other costs will be added. For example, API costs around OpenAI have reduced by 1/10 in just two months and costs will continue to decline. At the same time, businesses should expect that newer and more expensive LLMs will be introduced.

3 COST OF A FAST LAND

Businesses should carefully scrutinize all use cases to determine the investment tradeoff for a fast land. They need to understand the expected size of new user acquisition that comes with adding every new generative AI feature set. It's also important to understand how a future generative AI tax will impact a user expansion strategy.

THERE WILL BE REGULATION TO PROTECT BOTH THE CONSUMER & THE AI

1 **The first bucket is regulation for the consumer.** This regulation is similar to how we would regulate any other technology with potentially broad consumer and business impact, such as social media or medical devices. A framework needs to be in place to ensure a safe experience for consumers and in-line behaviors from the providers of these technologies. We'll be able to work through this side of generative AI regulation, even if some of them are novel around incorrect decisions, as seen with bias or hallucinations. Key pieces of regulation for the consumer has started to come together. Examples include watermarks, source generated IDs and the recent EU Artificial Intelligence Act.

2 **The second bucket is regulation for the AI itself.** The industry is now recognizing that we're close to AI being able to replicate itself, and we have to think about AI regulation very separately from existing consumer protection. This is because AI is already able to write code. When this happens, we have AI programs that self replicate code and direction. Any AI program that can create the offspring of new AI programs will be more successful evolutionarily if it can find more resources on which to run. One could imagine viruses created by first generation AI programs that find more places to self-replicate erroneous or malicious data protocols into second generation AI programs that they themselves create.

> *We have to think very carefully over the next few years about the regulation mechanism for AI being able to replicate itself.*
>
> **Max Schireson**
> **Operating Partner, Battery Ventures**
> **Former CEO, MongoDB**

THE GENERATIVE AI TEAM ALIGNED FOR FAST ITERATIONS

For generative AI to successfully iterate every few weeks, there needs to be full alignment across all business functions.

Functional Area	Primary Owner	Team	Business Stakeholder
Proprietary data wheel	VP Data Science	Data scientists, data engineers, machine learning engineers, software developers, Product Manager	Line of business owner
Workflows	Product Manager	Software developers, UX designer	Line of business owner
Infrastructure tools	VP Data Science	Data scientists, data engineers, machine learning engineers, software developers, Product Manager	
Product-market fit	Line of business owner	Sales, Marketing, Customer Success, Product Manager	
Generative AI tax	Line of business owner	Software developers, Product Manager	CIO, CFO
Go-to-market	Line of business owner	Sales, Marketing, Customer Success	CFO
Regulation	Legal	Product Manager	Line of business owner

THE VC FORECAST MOVING INTO 2024

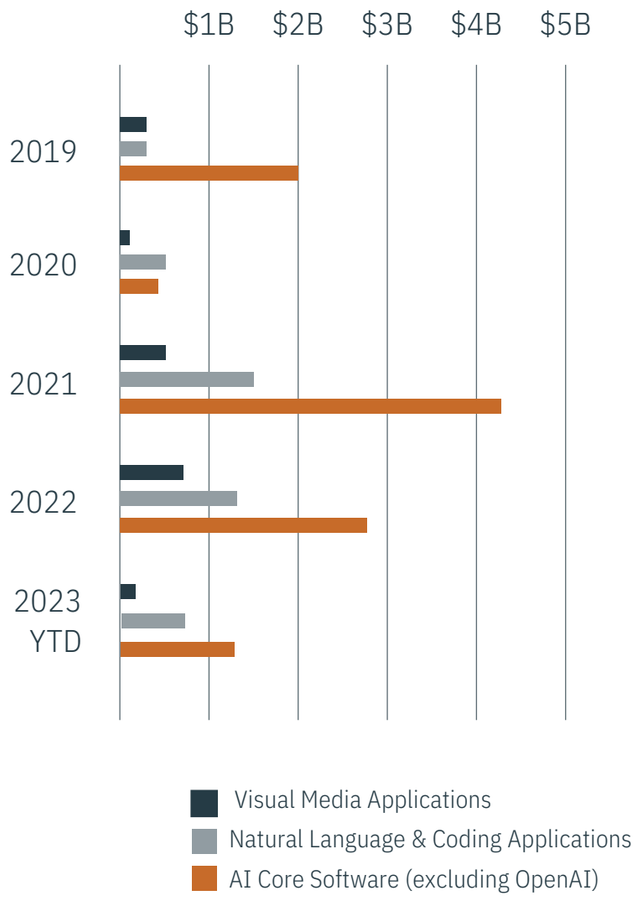
Never before has there been this size and scale of market opportunity, the speed at which it's happening, the amount of competition, the number of companies that are being funded and the amount of money companies are investing into generative AI.

Max Schireson

Operating Partner, Battery Ventures
Former CEO, MongoDB

CURRENT STATE OF GENERATIVE AI INVESTMENT:

GENERATIVE AI INVESTMENT BY SEGMENT, AS OF MAY 31, 2023



LANGUAGE APPS OVER VISUAL & CREATOR APPS

According to the latest generative AI investment numbers reported by Pitchbook in June[1], investors have been more heavily focused on language apps over visual and creator apps. Text and language are the most accessible economic value spaces where generative AI applications assist humans with tasks that they are doing today.

Given that the generative AI industry had a surprise birth in just Q1, a text-centric focus makes the most sense as the starting point of the journey. As large language architecture advances, the next generation set of applications will include more audio and visual based functionality.

DOUBLE-DOWN ON LLMOPS TO HOLD THE 85% FAILURE RATE DAM

On the infrastructure side, LLMOps has emerged as the backbone tool set to support applications developed with large language models. The good news is that this architecture is built on existing machine learning operations (MLOps). The industry is now using LLMOps as the catch-all term for MLOps plus any tools that are specific to large language models. We will continue to see this category to have tremendous investment given that 85% of machine learning projects fail[2]. Our experts do not see this number coming down anytime soon, especially as there will continue to be more and more generative AI applications added to a business' pipeline.

[1]PitchBook, "Generative AI Trends and Opportunities," June 14, 2023 [2]Forbes, "Why Most Machine Learning Applications Fail to Deploy," April 2023

NEXT STAGE OF GENERATIVE AI INVESTMENT:

DATA

VCs are focused on technologies that provide the ability for developers to access domain-specific information that is not available on the internet or in traditional databases, and to be able to update this data in real-time. This way, developers can provide better context and accuracy for generative AI models such as ChatGPT or GPT-4, which are often trained on outdated or incomplete data scraped from the web.

Key investment areas:

- Vector databases
- Clean room technologies
- Synthetic data generation

TOOLS

The rise of LLMs has created many gaps in today's data infrastructure stack. New tools will need to emerge for developers, data scientists and business users to leverage LLMs within an enterprise. VCs are excited by products that improve the accessibility and usability of LLMs. They are also looking for tools that make the lives of the "prompt engineer" or LLM-focused data scientist easier.

Key investment areas:

- Management of image and video data
- Dynamic LLM interfaces
- Integration of different LLM models
- Effective governance

HORIZONTAL APPS

Every function in an organization with repetitive work will be reshaped by generative AI, whether it is the democratization of coding, generating sales, supporting customers with virtual agents or creating content for marketing. Investors are looking for solutions to problems that are deeply embedded into business workflows with access to unique data sets.

Key investment areas:

- Developer productivity
- Testing of software code
- Tools for non-technical users to create use cases
- Security

VERTICAL APPS

Investors are interested in a new class of applications that encompass proprietary data sets, vertical go-to-market and the ability to embed deep into business workflows. The focus is on sizeable verticals with large and complex data that is mission-critical to businesses, combined with labor shortages and regulatory or compliance requirements.

Key investment areas:

- Life sciences: synthetic data for clinical trials, generative protein designs
- Supply chain and logistics: 3D modeling and blueprint generation

18-MONTH INDUSTRY PREDICTIONS:

George Mathew

Managing Director, Insight Partners

Former President & COO, Alteryx

1. We will run out of publicly available data to train next-generation LLMs.
2. Generative AI will move far beyond just text to visual and audio applications.
3. Immersive and interactive experiences, especially around images, videos and sound, will have emerged.
4. We will witness our first generatively-scaled cyberattack.

Max Schireson

Operating Partner, Battery Ventures

Former CEO, MongoDB

1. AI co-authors will be able to adapt to our writing styles.
2. Productivity growth will rise for knowledge workers.
3. Online engagement will be hyper-personalized.
4. Every service industry will be rethinking their value chains.

**These predictions are directly attributed to the featured panelists.*

PUTTING IT ALL TOGETHER: PRODUCT COMMERCIALIZATION SCORECARD

Businesses need the right approach to build out their short- and long-term generative AI strategy. Yet, they are without a consistent and industry-accepted commercialization framework. We have distilled the key recommendations from this report to create a scorecard to help executives establish generative AI readiness for their organization.

We believe that businesses should approach their generative AI development in these three phases:

Phase 1 – Create your data moat
Phase 2 – Strengthen your use cases with proxies
Phase 3 – Test your generative AI readiness

PRODUCT COMMERCIALIZATION SCORECARD:
PHASE 1 - CREATE YOUR DATA MOAT

Build Your Proprietary Data Flywheel

- ☐ Our data repository is accurate, complete and consistent
- ☐ We have mapped our data flywheel (interactions, correlations, resulting business outcome) to our domain-specific model
- ☐ We are able to articulate how our data flywheel provides us business defensibility
- ☐ We have an iterative roadmap in place for our data flywheel
- ☐ We have invested in tools, such as synthetic data or cleanrooms, to continually improve on the quality of our data repository as needed for our generative AI application

Develop Your Infrastructure Tools

- ☐ We have an LLMOps infrastructure stack in place with a 1-year expansion roadmap in place
- ☐ Continuous model improvement, LLM training and documentation is part of our business strategy
- ☐ Replicating a workflow experiment is easy
- ☐ Collaboration across data, operations and development teams is in real-time to support quick iterations
- ☐ We track datasets, metrics and hyperparameters to support our proprietary data flywheel

Calculate Your Generative AI Tax

- ☐ We are having ongoing discussions with our cloud provider to keep on top of pricing models
- ☐ We are tracking cost of inference on a monthly basis
- ☐ We have a LLM cost strategy to ensure we are aware of costs to deploy and train new model
- ☐ Our go-to-market strategy incorporates out generative AI tax

PRODUCT COMMERCIALIZATION SCORECARD:
PHASE 2 - STRENGTHEN YOUR USE CASES WITH PROXIES

Iterate On Key Workflows	☐ We have a UX designer on board constantly improving the user experience of our workflows	☐ Our workflows provide a deep domain-specific experience	☐ Our workflows connect into the rest of enterprise operations	☐ Our workflows integrate well into relevant existing applications and downstream systems	☐ We are able to launch a new workflow or enhancement every three weeks
Establish Proxies	☐ We have a framework in place for tracking key proxies	☐ We have identified potential repeatable, consistent use cases that we are tracking key proxies for	☐ We have identified potential audiences that we can monetize	☐ We are tracking to what extent are generative AI features being used with the goal of identifying key patterns	☐ We are tracking key proxies for the overall experience and productivity of people using generative AI features
Assemble Your Generative AI Team	☐ We have a data science head responsible for our data flywheel current state and roadmap	☐ There is a single product owner and line of business owner for each defined workflow	☐ Our LLM team is cross-functional, consisting of data scientists, data engineers, machine learning engineers, software engineers and product managers	☐ We have a central line of business owner responsible for product-market fit, generative AI tax and go-to-market	☐ Legal is ready with a process to help us stay on top of regulation and ensuring we are providing the most accurate and safe experience for all of our stakeholders

PRODUCT COMMERCIALIZATION SCORECARD:
PHASE 3 - TEST YOUR GENERATIVE AI READINESS

1. We have a proprietary data flywheel in place with data science head ownership.
2. We have an LLMOps stack and 1-year roadmap in place owned by a cross-functional LLM team.
3. We have a sustainable generative AI tax and a business owner responsible for this oversight.
4. Our product owner is able to launch deep-domain specific workflows every few weeks.
5. We have a proxies that validate a repeatable, consistent use case, future revenue, generative AI feature usage and productivity gains.
6. We have adhered to the most current AI regulation, such as EU Artificial Intelligence Act, articles 5&6, and have identified areas where AI can be replicated and have alerts in place of how inferences are made

Product commercialization for generative AI is an ongoing iterative process. There is less time to iterate in generative AI because the opportunity is both so huge and comes with intense competition. An iterative mentality has become even more critical, necessitating that each subsequent iteration is relatively crisp out of the gate.

Following these three phases, that are outlined in this section, will ensure that businesses have a solid framework and baseline to continually iterate on their generative AI use cases.

CONCLUSION AND LOOKING AHEAD

Six months into building the next generation of applications, we are seeing that the industry requires a framework for generative AI readiness. Our expert panel helped us to identify best practices to product commercialization for generative AI.

This report puts forth a Product Commercialization Scorecard as an expert rubric to help businesses become generative AI ready in 2024. The belief is that this report and scorecard will be a huge benefit for the industry in the next year.

DEVELOPED BY:

Rekha Ravindra
R² ACCELERATION
San Francisco Bay Area

The Next State of Generative AI Insiders Roundtable

In Q1 2024, we will develop the next State of Generative AI Insiders Roundtable and associated report. We'll pick up from where this report ended and start with these questions:

- What proprietary data flywheels have been created across key verticals?
- What new workflows are emerging from creative AI models or synthetic data?
- Did we move from proxies to indicators when describing product-market fit?
- Are there examples of revenue growth for generative AI applications?
- Are we able to start benchmarking generative AI tax and cost of inference?
- Has there been steps toward regulation for AI replicating itself?