



White Paper
by Jessica Hayes & Thomas Chauvin

GenAI in a bottle: Assessing real value amidst the hype

New York - London - Paris

www.axavp.com

[Introduction](#)

In a classic tale, a man uncovers a magic lamp and makes a grand wish to a genie, only to encounter complications as the wish brings unforeseen consequences. Generative AI, much like a modern-day genie, holds immense promise for creativity and productivity but also brings the risk of unforeseen effects such as hallucinations and new attack vectors.

Since the debut of ChatGPT just two years ago, expectations for GenAI have surged. Our 2023 report compared AI's potential to groundbreaking advancements like cloud computing and mobile phones - a comparison we still stand by today.

However, recent scrutiny has emerged regarding GenAI's enterprise value, highlighted by Goldman Sachs' June 2024 report, "Too Much Spend, Too Little Benefit" ¹. This report reveals GenAI has fallen short of ROI expectations, contrasting with last year's projection of a 7% global GDP boost. The focus has shifted to concerns about AI's long-term viability, scalability, ethics, and tangible returns.

Despite these concerns, GenAI continues to attract significant investment. In Q2 2024, VC funding rose 29% to \$42.9 billion, with AI representing nearly one-third of this capital, supported by major deals like xAI's \$6 billion Series B. Innovation is also thriving, with Microsoft investing \$19 billion in data centers and AI chips, and Meta launching its largest open-source model, Llama 3.1 405B.

Amidst the noise and excitement, one question stands out: **where will lasting value be realized?** This paper explores this in three parts:

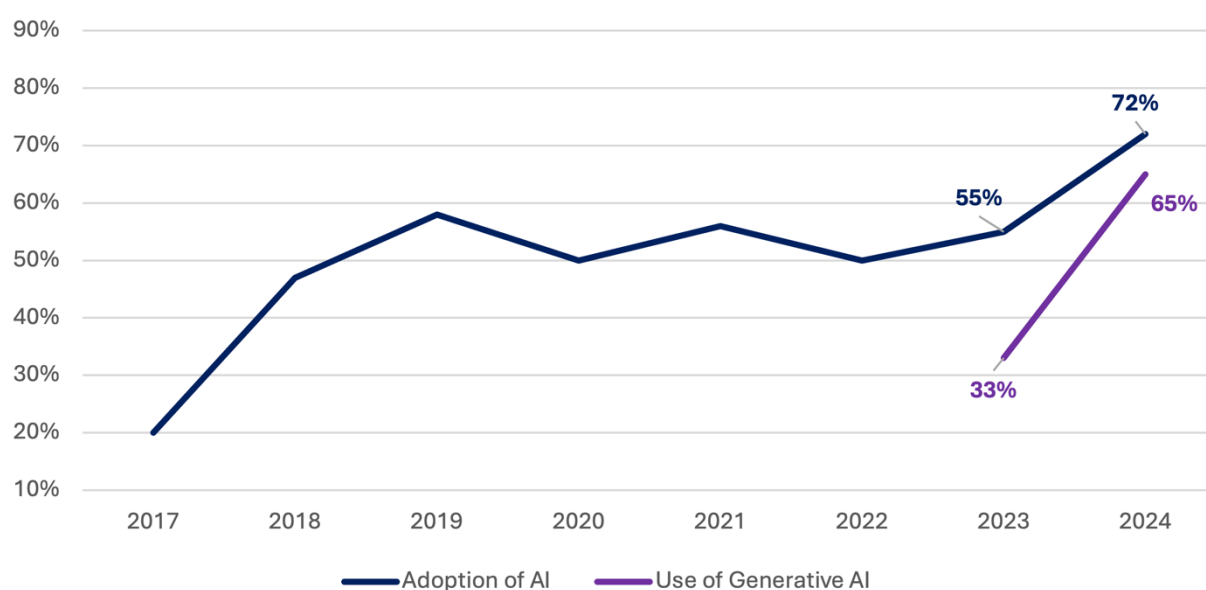
1. The Current State of GenAI
2. The Modern GenAI Tech Stack
3. Promising AI Applications

I. The Current State of GenAI

[AI is a top priority for executives, but homegrown deployment is lagging](#)

A recent BCG survey shows 85% of executives plan to boost their AI and GenAI spending², shifting from experimentation to production as expectations for value creation rise. This is mirrored by a rise in AI adoption, with organizations using AI in at least one function increasing from 55% in 2023 to 72% in 2024, and GenAI adoption nearly doubling in the same period.

Majority of organizations have adopted AI in at least one business function

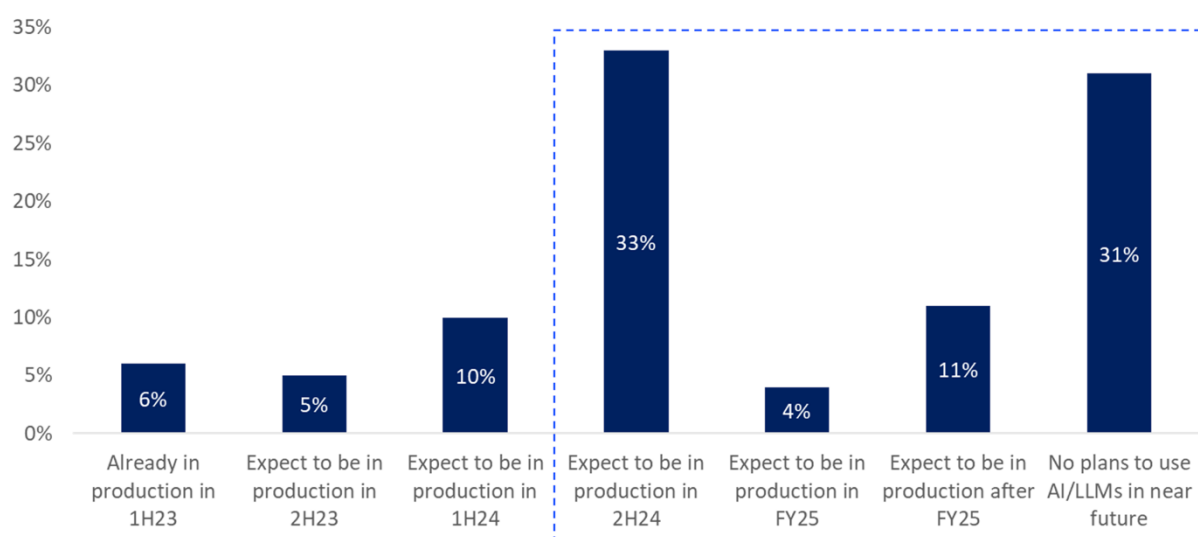


Source: McKinsey³

However, despite growing interest, in-house GenAI deployment remains nascent. Morgan Stanley's CIO survey shows 33% of CIOs anticipate their first AI/LLM projects going live in 2H24, with 15% targeting 2025 and 31% having no immediate plans. This is echoed by IBM's recent global study of 3,000 CEOs, which found that 71% of organizations are still in the pilot phase of GenAI projects⁴.

We at AVP have seen that most Global Fortune 2000 companies are opting for third-party GenAI solutions rather than developing their own models. Among the models in production, most are used for internal purposes by employees rather than for customer-facing applications.

Majority of CIOs expect initial projects to be in production in 2H24 and beyond



Source: Morgan Stanley, AlphaWise⁵

[The positive: concrete use cases have already emerged](#)

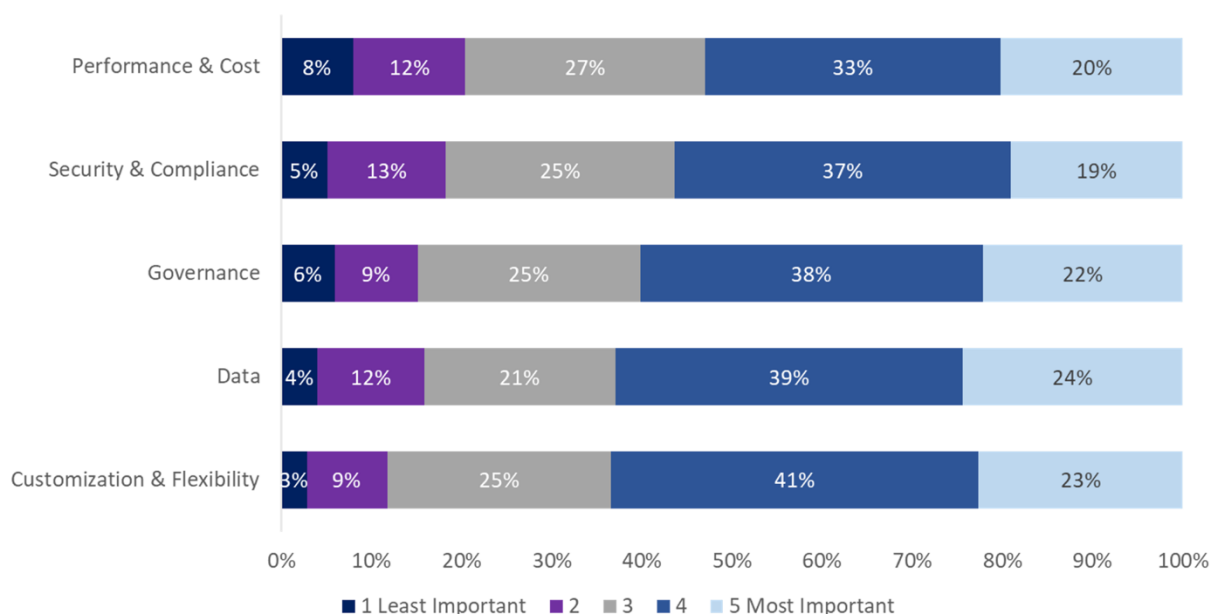
Internal chatbots: A growing trend in homegrown applications is the implementation of internal chatbots for employee search and support, aimed at boosting workplace efficiency. For example, a study from HBS shows⁶ that consultants at BCG who worked with a chatbot completed 12% more tasks and achieved a significant 40% improvement in quality than those who did not use a chatbot.

Coding: Since GitHub Copilot's debut in late 2021, AI code assistants have gained traction, with 63% of organizations either deploying or piloting these tools⁷. Code assistants are commonly used for time-saving efficiencies and cost reduction, though ROI on more qualitative tasks such as improving code quality and reducing bugs remains unclear.

Sales & marketing: AI sales tools have been rapidly adopted, with over 75% of sales leaders either using, implementing, or planning to use GenAI within the next year⁸. Tech giants are at the forefront, with Salesforce becoming an "AI-first company" in 2014 and deploying Einstein AI for personalized sales emails and customer journeys. Our latest sales automation report also highlights growing GenAI use cases, such as Cognism's Revenue AI, which enhances B2B data for lead generation to help teams surpass their revenue targets.

[The negative: enterprise challenges in adopting GenAI](#)

Key barriers to GenAI adoption



Source: ClearML⁹

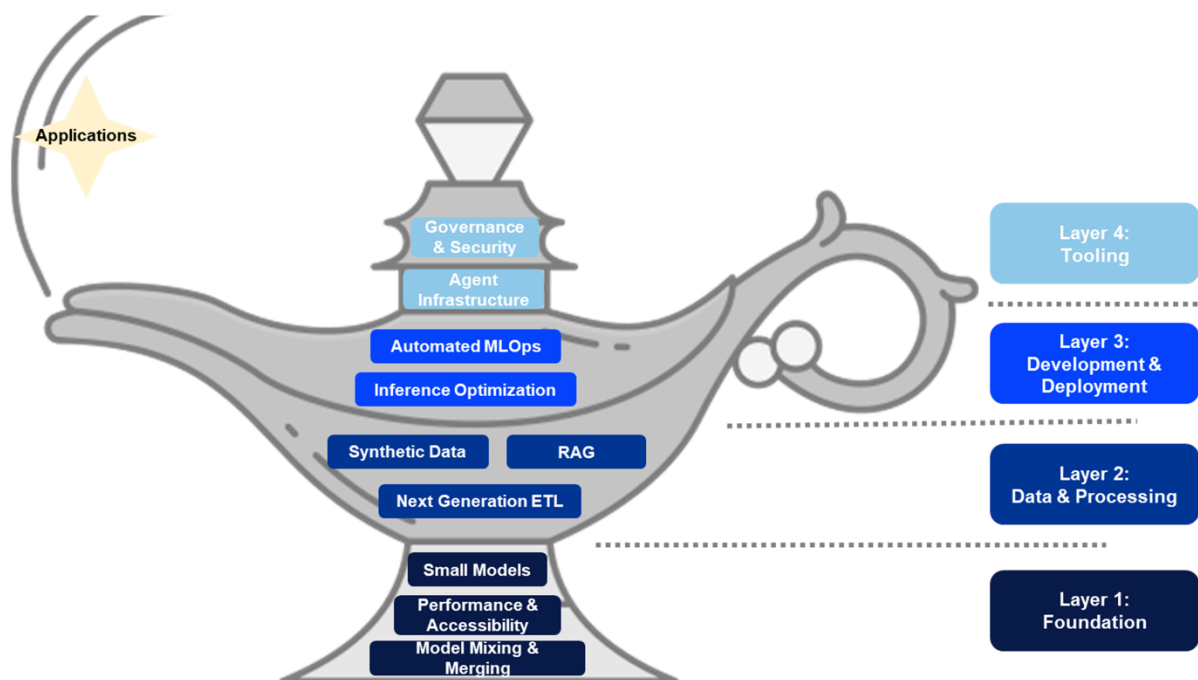
A study of 1,000 AI leaders and C-level executives conducted by ClearML identified five primary challenges in the adoption of GenAI, categorized into three key areas:

Cost, performance, and flexibility: CFOs are increasingly demanding clear returns and KPIs, but many companies struggle to demonstrate business use cases that justify high investment costs. Challenges include undefined best practices, difficulty of fine-tuning models with internal data, and the complexity of evaluating diverse models that are prone to drift.

Security concerns and data privacy issues: Security concerns remain a major barrier to AI adoption; a notable 30% of enterprises deploying AI have experienced security breaches¹⁰. AI introduces new vulnerabilities, including prompt injections, shadow AI, and sensitive data disclosure, creating additional risks and attack surfaces for CISOs to address.

Governance and reputational risks: GenAI models are inherently non-deterministic and unpredictable, which can lead to hallucinations—outputs that may annoy, mislead, or offend users. For example, evaluations indicate LLMs like GPT-4o are only 61% accurate on tax-related questions and often struggle with accuracy and relevance in specific enterprise contexts¹¹. Incorrect responses can pose significant operational risks for employee-facing applications and reputational risks for customer-facing applications.

II. The Modern GenAI Tech Stack



Foundation Models

Performance and accessibility: Foundation models are rapidly advancing in performance and accessibility. OpenAI's GPT-4o excels in multimodal functions, while Anthropic's Claude 3.5 Sonnet emphasizes ethical AI. Open-source models are also progressing, with

Meta's Llama 3.1 405B being the largest open-source LLM, and Mistral 8x22B offering strong multilingual support. As competition grows, AI models are becoming more affordable and accessible through APIs and platforms like Hugging Face. OpenAI's new GPT-4o mini, for instance, delivers better performance at 60% less cost than its previous models.

Smaller models: While large well-capitalized companies dominate the market, startups also have substantial opportunities, especially those developing small models that augment large models in a cost-effective way. For example, our portfolio company Synativ provides pre-trained, domain-specific vision models that enable companies to implement computer vision applications with minimal data. Similarly, emerging small language models (SLMs) are cheaper and faster to train, making them suitable for internal queries requiring less long-term context.

Model mixing and merging: New approaches like Mixture of Experts (MoEs) enhance scaling and reduce costs by integrating specialized sub-models within a larger framework, as seen in Mistral's Mixtral 8x22B, OpenAI's GPT-4, and Google's Gemini 1.5. Furthermore, emerging techniques such as Spherical Linear Interpolation (SLERP) combine multiple LLMs into a single model, improving performance across diverse tasks.

[Data & Processing](#)

Synthetic data: Access to clean, relevant data is essential throughout the development lifecycle. As traditional datasets become exhausted or difficult to acquire by 2026, synthetic datasets—generated through algorithms or simulations—are increasingly used. While concerns exist about its impact on model quality, blending synthetic data with real-world data, as seen in Google's Gemma models, can boost performance. Our recent deep dive on synthetic data highlights notable startups and trends in the space.

Next generation ETL: AI data processing tools are evolving beyond traditional ETL (Extract, Transform, Load) solutions, generating significant value. For example, data curation tools use AI to curate smaller, more performant datasets for model training, while feature engineering tools identify and modify features to boost model performance. Additionally, new tools process unstructured file types into LLM compatible formats such as JSON or CSV. This data is then converted into embeddings, indexed into vector databases that handle multi-dimensional arrays, and queried on demand.

RAG: Retrieval-Augmented Generation (RAG) enhances LLM responses by integrating real-time data from vector databases, reducing hallucinations and keeping responses current. Fine-tuning and RAG are most effective for different purposes – for example, a vertical specific model such as processing insurance claims would benefit from fine-tuning on static policies, while a conversational LLM providing financial advice would benefit from RAG to retrieve up-to-date data. RAG has exploded in popularity with accessible RAG-as-a-service solutions. Additionally, expanded context windows allow models to consider more information in a single interaction, reducing the need for extensive fine-tuning for additional context while enabling RAG to incorporate more retrieval processes.

[Model Development & Deployment](#)

Automated MLOps: Since 2023, the debate between using pre-trained models versus training from scratch has diminished, except in very niche areas. For example, GPT-4 outperforms BloombergGPT (developed from scratch) in most financial tasks¹². To fine-tune pre-trained models, data science teams are turning to automated machine learning operation tools to streamline the model development. This includes automating tasks such as model selection, version control, testing, and hyperparameter tuning. Alongside academic benchmarks like MMLU and HELM, enterprises are also adopting new tools such as evaluation foundation models and RAG metrics.

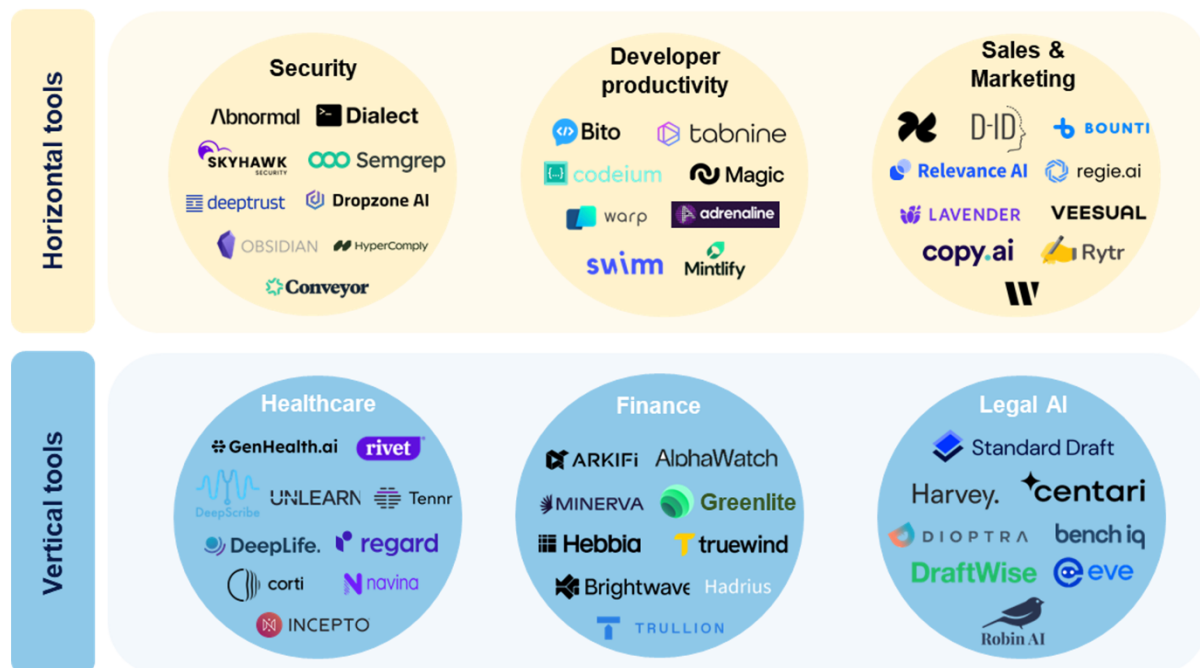
Inference optimization: Inference optimization platforms accelerate time to market by providing infrastructure for easy model deployment via API. Best-in-class platforms offer flexibility, enabling auto-scaling across multi-cloud and GPU datacenters to reduce costs, and ensuring fast cold starts and low latency for new model calls. Additionally, many optimization platforms enhance the inference speed and efficiency of GenAI models through techniques such as pruning, parallelism, quantization, knowledge distillation, and batching.

[Tools](#)

Agent Infrastructure: Emerging tools have made it easier for developers, data scientists, and non-technical users to leverage LLMs in enterprises. Popular frameworks such as LangChain allow developers to simplify AI app development with features such as chain building (whereby one output is used as the next input), prompt management (with prompt templates that provide output instructions), and workflow orchestration (with conditional logic). As agents look to handle more complex workflows, we're thrilled to see startups building infrastructure that empowers agents to perform tasks such as interacting with APIs and websites to gather data or processing payments. This emerging infrastructure will pave the way for the next generation of AI agents capable of executing autonomous workflows such as triaging security risks or conducting user experience research.

Governance and security: The EU AI Act, effective from August 2024, introduces the first comprehensive AI regulation framework. It requires high-risk AI systems - like those used for individual profiling - to adhere to strict data governance, including detailed record-keeping, dataset accuracy, and robust cybersecurity. Beyond mere compliance checklists, next generation governance tools prevent private customer data from flowing to models and automatically redact PII. With LLMs vulnerable to attacks like prompt injection, new cybersecurity solutions are emerging, including GenAI red-teaming and moderation systems designed to sit between queries and models.

III. Promising AI Applications



[Horizontal applications](#)

Every function characterized by repetitive and skill-based tasks is set to be transformed by GenAI. Success will hinge on solutions that are deeply integrated into business workflows and leverage unique data sets. We foresee substantial benefits emerging from these horizontal applications:

1. **Developer productivity:** GenAI coding automation tools promise to accelerate software development, delivery, and testing. Developers currently spend extensive hours on debugging and test / build cycles, which can be significantly streamlined with computer-generated code and code-review tools. Moreover, democratized access to development tools for non-coders offers exciting possibilities. For example, companies like Bito accelerate code review processes, while Tabnine's AI code assistant enhances code generation and automates repetitive tasks. Additionally, Swimm's AI-powered code documentation saves engineering teams valuable time and effort by quickly comprehending and managing any codebase in just minutes.

2. **Security:** While AI can amplify threats in the hands of cybercriminals, it also serves as a formidable defensive asset. For example, it can enhance application security by identifying and remediating code vulnerabilities. Traditional SAST frameworks can be generic and prone to false positives, but companies like Semgrep use AI to find false positives and propose tailored remedial actions. Additionally, GenAI can improve anomaly detection by analyzing event logs with specialized agents to identify unusual activities. For example, Abnormal Security uses AI to differentiate between attacks, spam, and legitimate emails - even identifying emails generated by AI.

3. Sales and marketing: GenAI is particularly well-suited for sales and marketing due to its iterative, creative, and personalization capabilities. Multi-modal models that convert text to images and videos are transforming outreach strategies. D-ID, an AVP portfolio company, generates highly personalized videos that achieve an 18% higher click-through rate compared to non-personalized marketing emails¹³. In e-commerce, our portfolio company Veesual has demonstrated substantial bottom-line impact with its AI-augmented virtual fitting shopping experience, resulting in a 77% higher conversion rate and a 22% increase in average order value (AOV).

Vertical applications

We're particularly excited about large verticals characterized by large amounts of complex data that is mission-critical to businesses, often compounded by labor shortages and regulatory demands. Below, we highlight several sectors poised to unlock significant value through AI:

1. Legal: Advancements in LLMs are transforming the legal industry by dramatically boosting efficiency and cost-effectiveness. Historically, the adoption of legal tech has been slow due to limited capabilities, a risk-averse culture, and a business model that didn't incentivize change. However, rising client expectations and a shift toward value-based pricing are accelerating the adoption of AI-powered tools in legal practices. We're especially excited about startups that seamlessly integrate AI into existing legal workflows, which requires a nuanced understanding of legal processes. Beyond drafting documents, AI can automate contract reviews, perform research, and track timekeeping. Ideal products should offer customization and appropriate permission levels for different personas.

2. Finance: The demand for AI solutions in finance is surging, as evidenced by notable developments like Bloomberg GPT, Morgan Stanley's partnership with OpenAI, and UBS's M&A AI co-pilot—a tool capable of scanning 300,000 companies in under 30 seconds to generate buy-side ideas and identify potential buyers in sell-side scenarios. In this complex field, professionals grapple with a mix of proprietary, confidential data and real-time market information, all while managing manual, repetitive tasks ripe for automation. Startups are making impressive strides, with companies such as Hebbia enhancing diligence processes and streamlining data room reviews for major financial institutions. Meanwhile, companies such as Portrait Analytics are developing chatbot AI analysts that leverage real-time market data. As general-purpose LLMs like ChatGPT are not yet fully equipped to handle complex financial tasks, founders with a financial background are uniquely positioned to innovate and excel in this space.

3. Healthcare: In healthcare, GenAI has potential to transform every facet, from biopharmaceutical R&D to diagnosis, treatment decisions, and patient experience. For example, AVP portfolio company Incepto provides AI tools that enhance medical imaging decisioning for radiologists. Another well-established use case includes AI-powered medical scribes, which apply human-like judgment to transform unstructured data into structured formats for electronic health records (EHRs). Emerging applications are making strides in improving the provider-payer relationship - for example, automating manual processes such as prior authorization and claims reimbursement that

involve synthesizing complex patient data. Looking ahead, GenAI may significantly accelerate drug discovery as advanced AI platforms generate and evaluate new materials and molecules, transforming the landscape of medical treatment.

Parting thoughts

At AVP, we are enthusiastic about exploring AI's vast potential while carefully minding its challenges, much like the man with the magic lamp navigating his wish. As high-conviction investors, we're thrilled to back founders who are shaping the future of AI infrastructure and applications, from early-stage startups to pre-IPO companies. We view AI not as a passing trend, but as the future of software. If you're at the forefront of this innovation, we look forward to connecting with you! You can reach us directly at jessica.hayes@axavp.com and thomas.chauvin@axavp.com.

-
1. <https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit>
 2. <https://www.bcg.com/publications/2024/from-potential-to-profit-with-genai#:~:text=71%25%20of%20the%20leaders%20we,three%20tech%20priority%20for%202024>
 3. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
 4. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/c-suite-study/ceo>
 5. <https://www.morganstanley.com/content/dam/msdotcom/what-we-do/wealth-management-images/uit/MSCo-Research-Report-Mapping-AIs-Diffusion.pdf>
 6. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321
 7. <https://www.gartner.com/en/documents/5298563#:~:text=Sixty%2Dthree%20percent%20of%20organizations,their%20impact%20on%20developer%20productivity.>
 8. <https://www.gartner.com/en/documents/5396163>
 9. <https://clear.ml/blog/new-industry-research-reveals-a-profound-gap-between-hyper-inflated-expectations-and-business-reality-when-it-comes-to-gen-ai>
 10. <https://www.preamble.com/blog/preamble-adds-new-ai-security-platform-offerings-including-free-tools-to-identify-shadow-ai-use-and-potential-breaches#:~:text=The%20risk%20associated%20with%20unchecked,that%20internal%20parties%20compromised%20data.>
 11. <https://www.vals.ai/taxeval>
 12. <https://arxiv.org/abs/2305.05862>
 13. <https://www.d-id.com/resources/case-study/video-campaigns-pilot-case-study/>

We invest in great entrepreneurs.
We support outstanding companies.



New York - London - Paris

www.axavp.com