APR 18, 2023

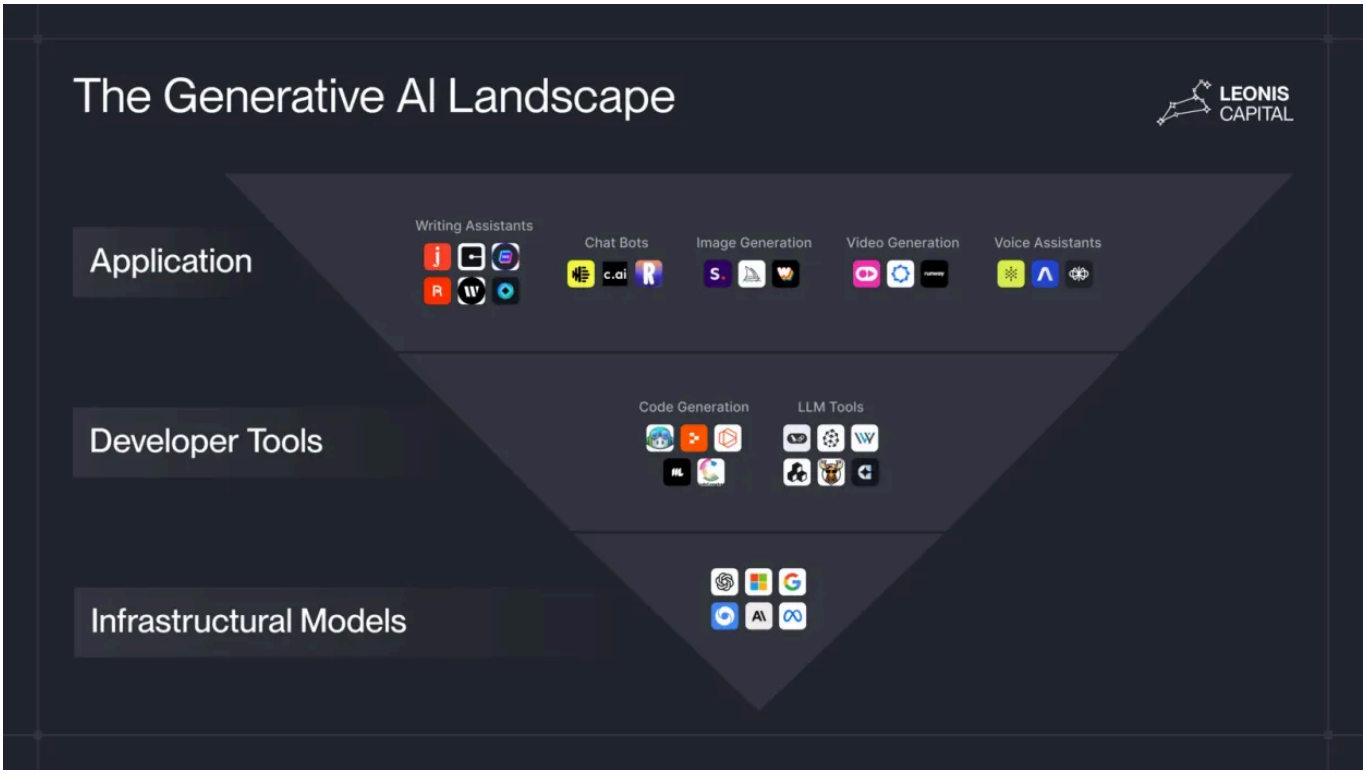
Recent Advancements in Generative AI

An introduction to generative AI.



By Jenny Xiao and Jay Zhao

Leonis Capital · leoniscap.com/research/recent-advancements-in-generative-ai/



[Full webinar deck available here.]

The tech world today is abuzz with excitement over generative AI. The release of ChatGPT in late 2022 had catapulted LLMs into the mainstream consciousness, and suddenly everyone from Silicon Valley VCs to Wall Street analysts are talking about the transformative potential of generative AI.

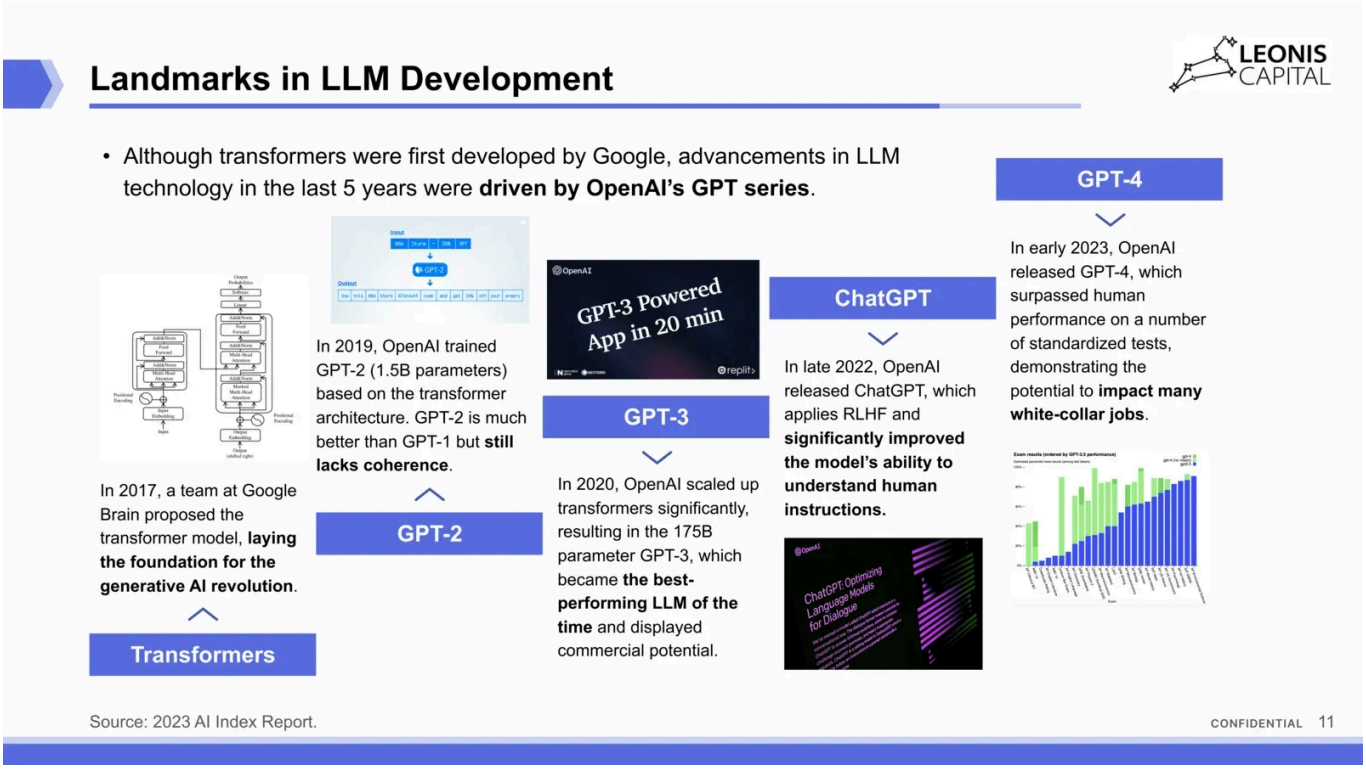
Amidst this frenzy, we at Leonis Capital have been receiving a wave of questions from the Leonis community about this new technology: **What are generative AI models capable of? What are their limitations? And where are the opportunities for startups and early-stage investors?**

As early-stage investors focused on AI-powered companies, we've been closely tracking developments in this space for years. While the sudden explosion of interest in generative AI might have seemed to come out of nowhere for many, we saw it as the culmination of years of steady progress in the field. Our recent webinar aimed to provide context for these recent breakthroughs and offer a framework for thinking about the future of AI innovation.

In this blog post, we'll share the key insights from our webinar, exploring everything from the fundamental technology behind LLMs to our investment thesis for the industry.

The Transformer Revolution

At the heart of recent AI breakthroughs is the transformer architecture, first proposed by Google Brain researchers in 2017. This innovation addressed a key limitation of traditional deep learning models – their inability to process large amounts of data efficiently. The transformer's attention mechanism allows it to capture relationships between data points in very long sequences, enabling the generation of much more coherent and contextually appropriate content. What's particularly interesting is how the field has evolved since then. While many expected the next AI revolution to come from a new architectural breakthrough, it turns out that scaling existing models has been the primary driver of progress. OpenAI's research on scaling laws showed that model performance correlates strongly with parameter count and dataset size. This insight, though initially met with skepticism, has proven remarkably accurate.



The Current State of LLMs

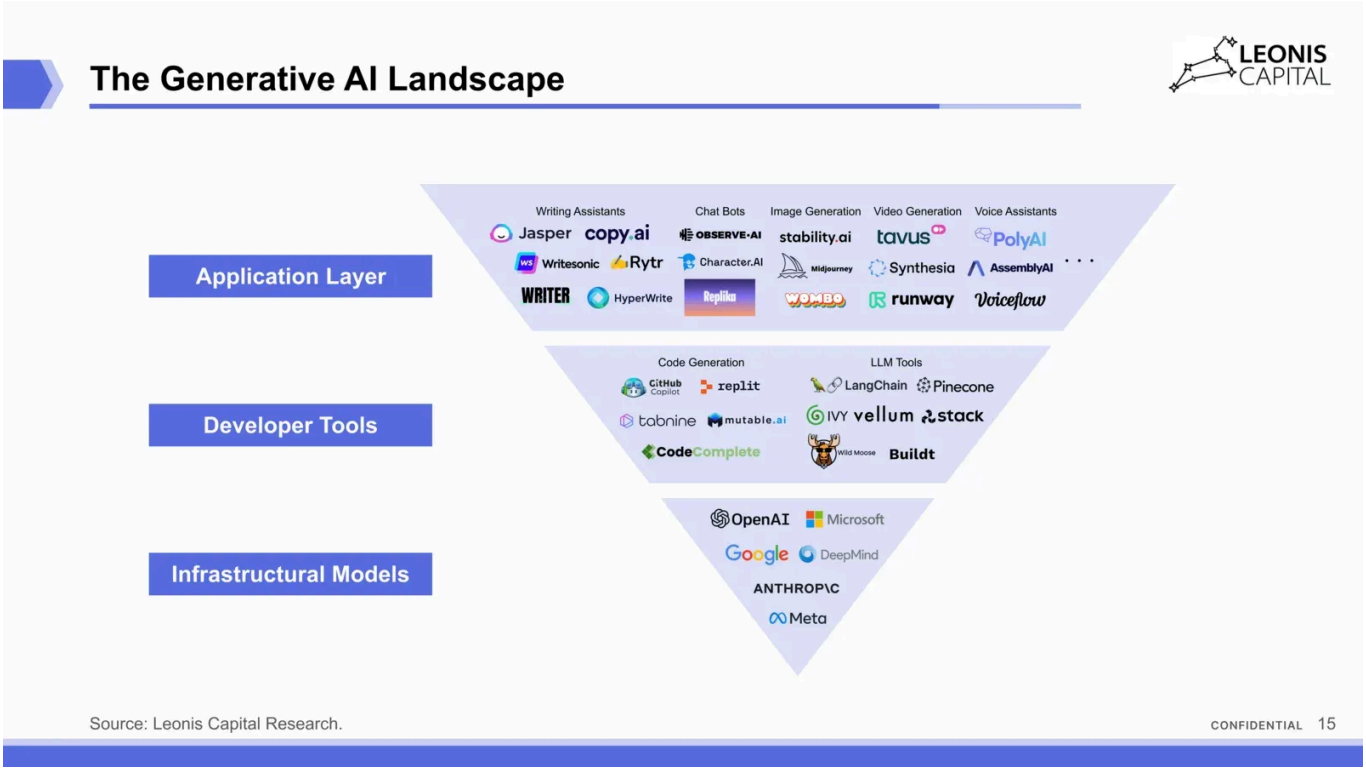
As investors, we need to be clear-eyed about both the immense potential and the current limitations of LLMs. Despite their promise, widespread adoption of LLMs is hindered by several key limitations:

1. **Hallucination:** LLMs can confidently generate false or nonsensical information. This isn't just a bug, but a feature of how these models work. While there are ways to mitigate this issue, it remains a significant hurdle for high-stakes applications in fields like medicine or law.
2. **Limited memory:** Most LLMs have a fixed context window, limiting their ability to work with very long documents or maintain coherence across extended conversations. Solutions like vector databases and models with larger context windows (like Anthropic's) are emerging to address this.
3. **Enterprise data security:** Many large organizations are hesitant to use public APIs due to data privacy concerns, leading to a growing market for self-hosted and fine-tuned models.

The AI Industry Landscape

We see the generative AI landscape as consisting of three main layers:

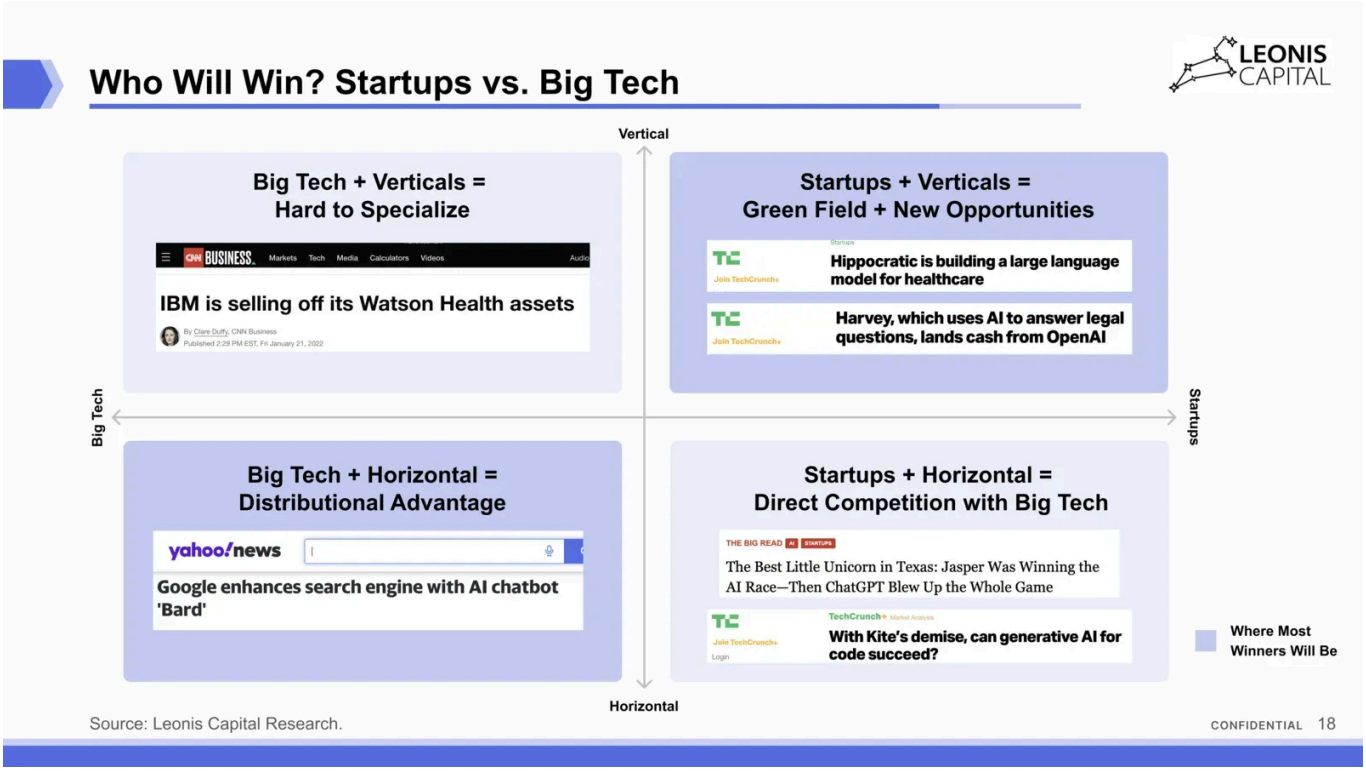
1. **Infrastructure/Model Layer:** Dominated by a few key players like OpenAI, Google (DeepMind), and Anthropic.
2. **Developer Tool Layer:** A rapidly growing ecosystem of tools for working with and deploying AI models.
3. **Application Layer:** The most diverse area, with both horizontal and vertical applications targeting various industries and use cases.



Interestingly, this structure in the U.S. resembles an inverted pyramid, with the largest number of companies in the application layer and the fewest in the infrastructure layer.

Who Will Win: Startups or Big Tech?

This is a hotly debated question in tech circles right now. Our view is more nuanced – we believe the answer depends largely on whether we're talking about horizontal or vertical applications.

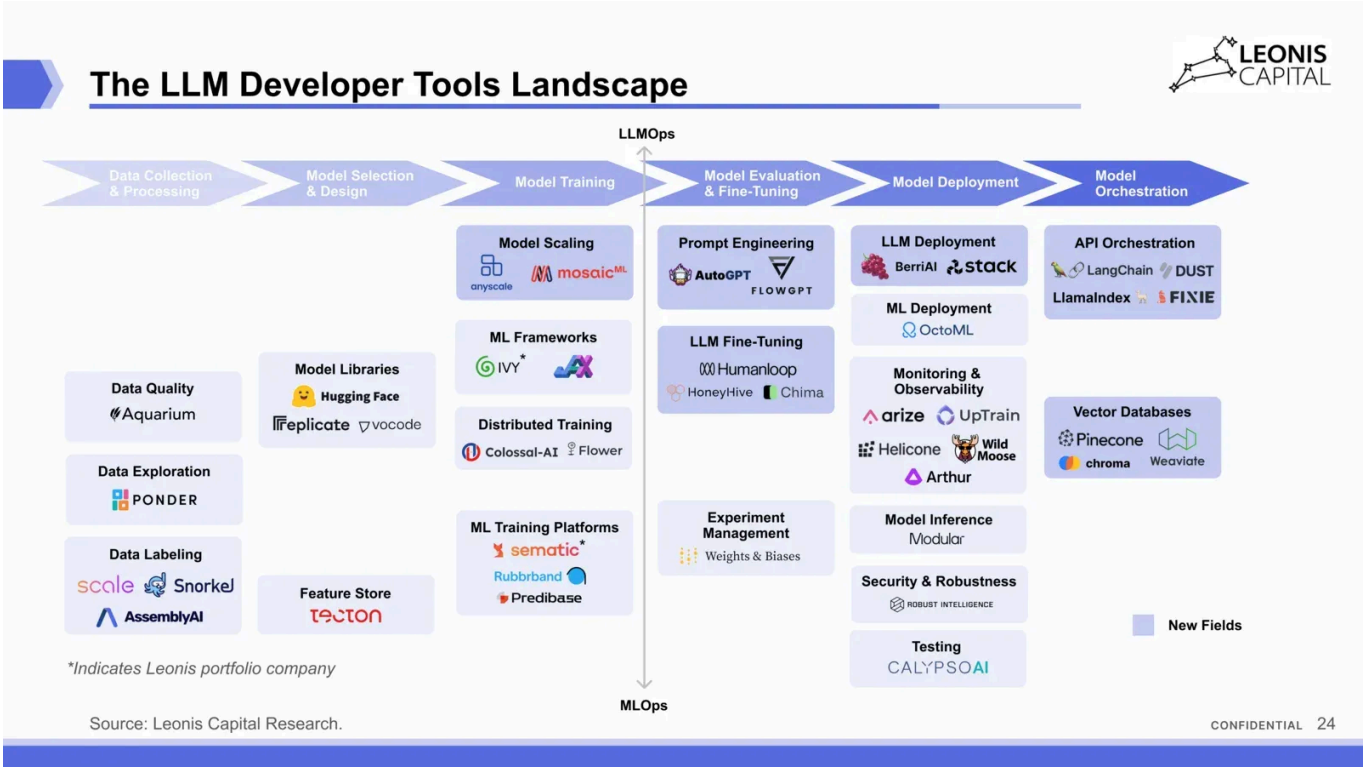


In horizontal applications (like general-purpose chatbots or search), big tech companies have a significant distributional advantage. It's incredibly challenging for startups to compete head-to-head with giants like Google or Microsoft in these spaces.

However, we see tremendous opportunities for startups in vertical applications. Areas like healthcare, law, and finance require deep domain expertise and specialized solutions that big tech companies often struggle to provide. This is where we believe the next generation of transformative AI companies will emerge.

The Developer Tool Opportunity

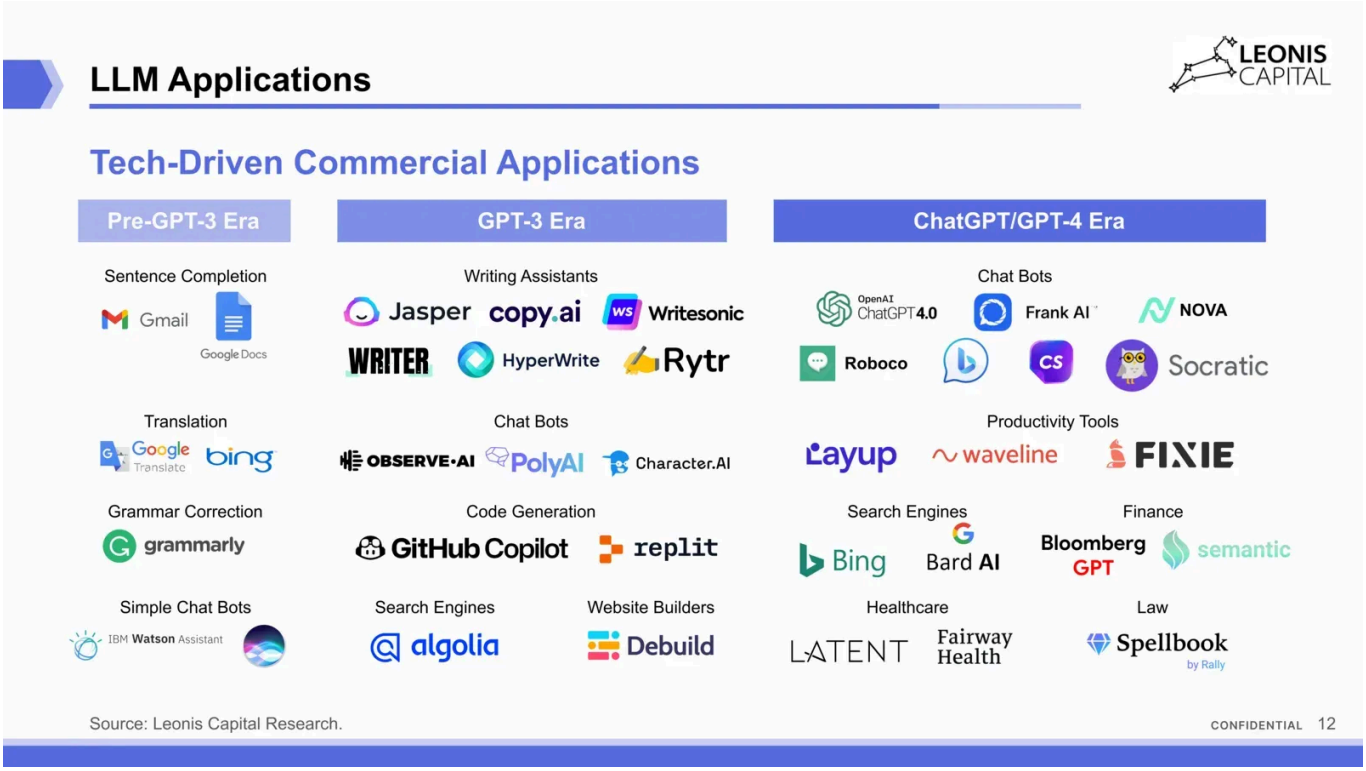
The explosion of AI development has created a booming market for developer tools. However, we urge caution when evaluating investment opportunities in this space. While there are genuine needs being addressed, many of these tools lack strong defensibility. It's often too easy for larger players to simply incorporate similar functionality into their existing offerings. That's why we're pretty bearish about some of the "new fields" in the LLM developer tool space including prompt engineering and API orchestration.



When we do invest in developer tools, we look for companies solving fundamental, sticky problems that are unlikely to be easily replicated. A great example is Ivy, a unified machine-learning framework we recently backed. Ivy addresses the long-standing issue of fragmentation in ML frameworks, providing a universal solution that integrates with all major backends and hardware. This type of product requires significant engineering work to build, becomes deeply integrated into workflows, and benefits from strong network effects – all characteristics we love to see.

The Application Layer: Where the Magic Happens

While infrastructure and developer tools are crucial, we believe the application layer is where the most value will be created in the coming years. When evaluating application-layer companies, we focus on three key factors: 1) **user experience**: In an era where startups have access to similar underlying AI capabilities, a superior UX can be a significant differentiator; 2) **unique edge**: We look for companies that have some special insight, dataset, or industry knowledge that gives them an advantage; and 3) **avoiding direct competition with big tech**: As mentioned earlier, we're wary of startups going head-to-head with tech giants in horizontal applications.



A recent investment that exemplifies these principles is Layup, an AI-powered workflow automation tool. Layup stands out by offering a much more user-friendly experience than traditional integration tools, targeting a clear user base (non-technical employees in tech companies), and creating a sticky product by integrating deeply into existing workflows.

The AI Hype Cycle and Long-Term Transformation

It's impossible to ignore the current hype surrounding AI. We're seeing echoes of the 1990s internet boom, with grandiose predictions about AI's immediate impact. While we're incredibly bullish on the long-term potential of AI, we also recognize that meaningful transformation takes time. The real excitement for us lies in the AI-first companies being founded now. These startups are being built from the ground up to leverage AI's capabilities, often allowing small teams to accomplish what would have previously required much larger organizations. Over the next few years, we expect to see truly transformational companies emerge in verticals like finance, healthcare, climate tech, and global logistics.

Looking Ahead: Challenges and Opportunities

As we navigate this rapidly evolving landscape, several key areas stand out as both challenges and opportunities:

1. **Content moderation and AI safety:** As AI systems become more powerful and widely deployed, ensuring they behave safely and ethically is crucial. While there are opportunities here, it's a complex space with both technical and political considerations.
2. **AI security:** As more companies integrate AI into their core processes, securing these systems against attacks and unintended behaviors will be critical. We see this as a promising area for developer tool companies.
3. **Vertical AI applications:** We remain extremely excited about the potential for AI to transform specific industries. Companies that can combine deep domain expertise with cutting-edge AI capabilities have the potential to create enormous value.
4. **AI-powered productivity tools:** As demonstrated by companies like Layup, there's significant potential in tools that use AI to streamline and automate knowledge work.

We're living through an extraordinary period in the development of artificial intelligence. The pace of progress is breathtaking, and the potential applications seem limitless. However, as investors, we must balance our excitement with a clear-eyed view of the challenges and limitations.

By focusing on early-stage companies with strong fundamentals – superior user experiences, unique edges, and defensible positions – we believe we can identify the truly transformative AI companies of the future. These are the companies that will not just ride the current wave of hype, but will fundamentally reshape industries and create lasting value.

The AI revolution is just beginning, and we couldn't be more excited to play a part in shaping its future. Stay tuned for more insights as we continue to explore this fascinating and rapidly evolving field