Data and AI

Competitive Advantage and Applied AI

## Data Is the Fuel for the AI Era

Data is the critical input for AI. Combining big data and AI potentially enables a wealth of new profitable applications.

We see this as an era of "big data", where data has become the "new oil" driving artificial intelligence (AI) and its transformative potential. An estimated 181 zettabytes (ZB) of data are expected to be created in 2025 alone (Figure 1). To put that into perspective, 1 ZB is equal to 1,000,000,000,000,000,000,000 (10^{21}) bytes—or the equivalent of storing 250 billion DVDs.[1]

Figure 1.

More Data Has Been Generated in the Last Two Years Than All of Human History Combined

Zettabytes of data
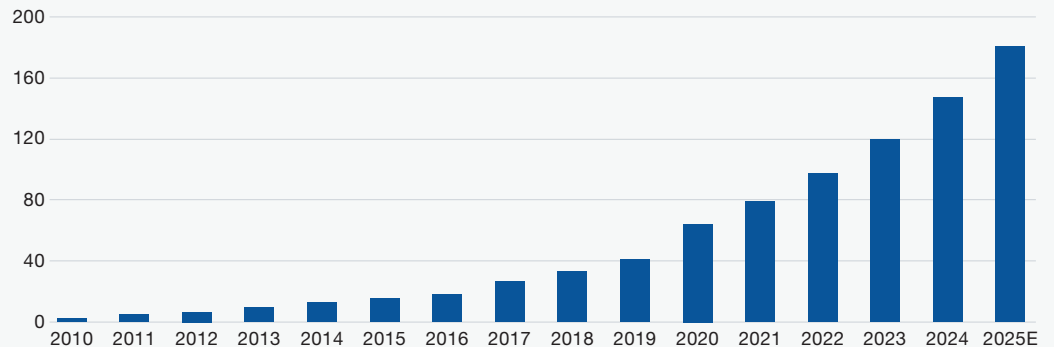
Source: Statista

This growth reflects the increasing digitization of global activities across all industries. It also includes social media and user-generated content, data from new connected devices, scientific research, streaming, and synthetic data being generated via AI. In our view the growth in data volumes creates new opportunities for training more accurate and adaptable AI and machine learning models via deep learning and self-supervised learning; analytics and research to reveal previously unreachable patterns, trends, and relationships; innovation via unprecedented cross-disciplinary collaboration; and improved risk management and security via real-time detection of patterns and anomalies before they escalate.

This theme is becoming increasingly relevant following rapid investment and developments related to large language models (LLMs)—a type of AI model trained on massive data sets that can recognize, translate, predict, or generate text or other content—and can be fine-tuned to leverage the learned knowledge to perform specific tasks more effectively. Major LLMs include OpenAI's ChatGPT, DeepSeek's R1, xAi's Grok, Google's Gemini, Anthropic's Claude, and Meta's LllaMA, among others. They are also at the heart of an estimated 750 million applications using LLMs in 2025, aiding in everything from smartphones to customer service chatbots to content creation and coding software.[2]

The combination of data and AI is also creating new economic opportunities via new industries, new job roles, and new revenue streams. The data and AI market opportunity could grow to $850 billion to $1 trillion by 2030, according to industry consensus estimates. As a result of these shifts, companies now recognize the potential from data and AI for both material opportunity and the risk of potential disruption—and thus, we believe, every CEO needs a data and AI strategy.

LAZARD
ASSET MANAGEMENT

Navigating the Data Maze

Accuracy, data privacy, security, and intellectual property rights are key challenges in implementation.

While proprietary data is an increasingly valuable resource, implementing data and AI solutions presents several key challenges that companies need to address to achieve success. These challenges include: data quality (ensuring data is accurate, consistent, and comprehensive); data privacy and security (proactively addressing data breaches and cybersecurity vulnerabilities—cybercrime costing $9.5 trillion in 2024[3] [Figure 2]); complex regulatory landscapes such as GDPR, CCPA, and HIPAA; and intellectual property or IP integrity (ownership of data, AI models and output, patentability of AI technologies, copyright and data licensing, status of AI-generated content).

Figure 2.

Data Breaches by the Numbers

$9.5 trillion	$4.8 million	1,636	42%
Cybercrime costs the global economy in 2024	Average cost of data breaches	Cyberattacks per organization each week	of the US population is impacted by data breaches

Source: Indusface, IBM, HIPAA

Many incumbent data services companies have strong competitive advantages in addressing these challenges. We believe their business-to-business (B2B) business models, where data exchange occurs under explicit contracts, are better placed than business-to-consumer (B2C) models, where publicly available data is indirectly monetized. These companies are also leaders in terms of policies and procedures, so AI solutions are developed with human oversight; keep data and user information secure via private, multi-modal LLMs that do not share user interactions with third parties; and secure deployment via commercial-grade cloud infrastructure.

The value add of proprietary, curated, secure data is an accelerator beyond the natural moats these businesses typically possess. Strong network effects and high switching costs provide valuable time for these businesses to offer and monetize tailored AI-based solutions and even entire new business models to customers with low risk of attrition.

Proprietary Data Is a Powerful Moat

90%

Company data that is unstructured

Proprietary, quality data sets and specialized domain expertise offer strong competitive advantages for companies able to leverage their data via AI.

As the competition in the AI market heats up, we believe there will be a particular focus on proprietary data—unique information that a company owns or has exclusive rights to use (and, hence, is not readily accessible to competitors). Examples include customer transaction histories, user behavior data from apps or websites, patient records, sensor data from IoT devices, privately conducted research, and industry-specific data sets collected through partnerships with other organizations. Leveraging and capitalizing on the nearly 90% of data currently generated by organizations that goes unused or underutilized is emerging as a major investment opportunity.

This marks a major shift in the world of AI where publicly available and synthetic data was once seen as the foundation of model development. That usefulness is hitting a ceiling, as most AI companies have largely exhausted the openly available internet. Proprietary, high-quality data sets are becoming a real value-add differentiator empowering companies to fine-tune AI models with domain-specific knowledge, creating smarter and more useful industry-specific and tailored applications for customers that outperform generic or generalist models trained on public data. In our view companies that own proprietary data sets and/or benefit from specialized domain expertise possess a major economic moat and growth engine.

This is especially true in highly specialized industries, where our conversations with companies show that every organization has its own unique "language" based on real-world environments and industry-specific constraints. For example:

- **Legal**: Companies are combining up to 161 billion legal documents (with 1.6 million new documents added daily) with law firms' internal data and IP to offer clients personalized Gen AI assistants, saving up to 11 hours of human legal work per week.[4]
- **Finance**: Companies are leveraging data on 4.6 billion active debit and credit cards (and 639 million transactions per day) to develop AI deep learning detection models to prevent $40 billion in fraud attempts at merchant clients annually (company information and data 2025). [5]

Overall we believe companies with large, high quality proprietary data are well positioned to capturing new market opportunities in the AI era, offering both data security and product innovation benefits to their customers.

Notes

1. Source: Rivery 2025
2. Source: Springs 2025
3. Source: Cybersecurity Ventures 2025
4. Source: RELX 2024
5. Source: Visa 2024

Important Information