The Journal of Business and Artificial Intelligence
ISSN: 2995-5971 · Volume 3, Number 1

AI Doesn't Matter

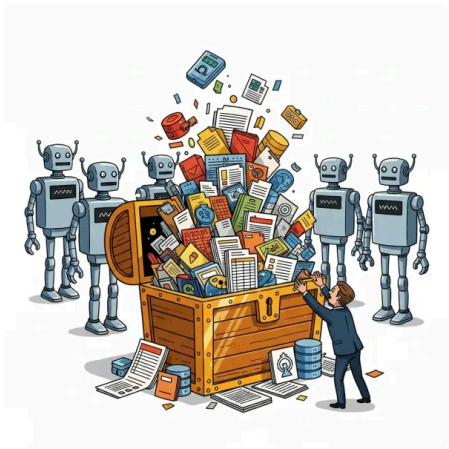
Unfair Sustainable Competitive Advantage Comes from Your Data, Not Your AI

Stephen Smith, Kenneth Morse

DOI: https://doi.org/10.66241/cu19d

Stephen Smith
AI Transformations

Kenneth Morse
Entrepreneurship Ventures Inc.



Abstract: Artificial intelligence has rapidly evolved from serving as an optional differentiating technology to an operational necessity. As foundational models, computational power, and technical expertise become increasingly accessible, simply using AI alone is no longer a source of sustainable competitive advantage. This paper argues that long-term differentiation will come not from algorithms, but from data — its ownership, protection, quality, originality, and relentless growth.

Drawing historical parallels from innovations such as the printing press to the internet, the analysis contends that technological breakthroughs eventually become commoditized, while the content or the data they manipulate retains strategic value. This paper outlines five critical actions leaders must take to build durable, "unfair" sustainable competitive advantage:

1. Protect proprietary data
2. Identify hidden data assets
3. Clean and curate internal datasets
4. Negotiate exclusive access to external sources
5. Create new data streams through innovation

It also examines the risks of synthetic data, "agentic misalignment", the promise of self-driving laboratories, and the accelerating fall in the cost and complexity of AI infrastructure. The conclusion is clear: as AI becomes ubiquitous, data sovereignty — not model sophistication — will determine which companies and organizations survive and thrive. AI may no longer be rare, but proprietary data remains priceless.

> *"You can't scale AI (adoption) when you don't even know where your data is. Without strong data governance and security, any AI initiative will crumble within a year or two - - because you're building on a weak foundation."*
>
> *— Vinh Nguyen, Sr. Fellow AI, Council on Foreign Relations; Former Chief Data Scientist, NSA*

Audience: CEO, CTO, CIO, CDO.

Introduction

Artificial intelligence (AI) has quickly become a source of competitive advantage for organizations worldwide. Today, businesses can differentiate themselves by deploying AI to automate processes, generate insights, and enhance the quality of decision-making.

Yet history shows that such advantages rarely last for long. Just as the printing press democratized access to information, foundational AI capabilities are rapidly becoming accessible to all. To secure long-term differentiation, companies must look beyond algorithms and models. The future source of "unfair" sustainable competitive advantage will be ownership, protection, and generation of proprietary data.

The printing press was once a breakthrough technology too!

When Johannes Trithemius wrote in 1494 that "the word written on parchment will last a thousand years," he underestimated the transformative force of the printing press, invented just fifty years earlier. What seemed durable at the time — the labor of scribes, the exclusivity of manuscripts — was overtaken by a technology that spread literacy and knowledge.

The lesson for modern business is clear: the technological breakthrough of the printing press shifted the basis of advantage initially, but eventually, owning the printing press mattered less than owning the content it reproduced. Similarly, in today's AI-driven economy, owning the data will matter far more than simply accessing the models.

AI Does and Doesn't Matter

The window is closing rapidly for leveraging existing breakthroughs in AI to steal market share from competitors. Currently, this opportunity is made possible via first-mover advantage, which is a viable and valuable strategy. Real shifts in market leadership are occurring. Those companies leveraging AI early on are stealing customers and revenue from their competitors who don't. And market share gains that are acquired during these early days of AI adoption will not be easily surrendered.

However, in the not-too-distant future, first-mover advantage will have run its course as a unique source of competitive advantage as most companies will then be utilizing AI. There will be no particular advantage of one company over another if all companies use AI. This is a consistent pattern with other breakthrough technologies such as electricity, laser printers, or even the internet. What was once a competitive advantage became a commodity – a game that everyone played and had to play.

In the distant future, the real winners will be the companies that secure 'unfair sustainable competitive advantage' – or said another way – competitive advantage that is durable, long-lasting, and that competitors cannot replicate. AI, by itself, will not be sufficient. AI matters less once everyone is using it.

So, if AI won't matter then what will provide unfair sustainable competitive advantage? It is not the AI algorithms. Even Google understands this. An internal Google memo discussing long term competitive advantage in AI was leaked to the public. It suggested that continued breakthroughs in AI algorithms were important but would not protect the company in the long term. Despite their prowess in AI, Google had no competitive 'moat' with which to protect their business long term.

> *"We have no moat and neither does OpenAI."*
>
> *— Leaked internal Google document*

Do these five things to secure your unfair competitive advantage

To build a defensible advantage, leaders must treat data as their most critical strategic asset. This requires a deliberate and forceful program:

1. Protect. Protecting sensitive data from leakage or misuse is a no-brainer but you need to be disciplined about it. Too many companies are still giving their data away for free or not capturing it in the first place. Many companies work with partners where data is generated or transported. They should make sure that it is clear, through binding legal agreements, that their company owns that data exclusively, and if there is an infraction, to pursue remediation with a vengeance.

2. Identify. Identifying overlooked or underutilized internal datasets is one of the best ways to find more high-quality data. AI can deal with messy data that perhaps you never before thought to be useful. Look in the nooks and crannies of your businesses for messy data. Good places to look are in human generated text, linkages, and IOT or edge data. But be careful what you feed into your AI engine – like a young child who has learned to use an inappropriate word – AI may surface patterns that are embarrassing or offensive.

3. Clean. Cleaning and curating internal data assets to improve quality multiplies the value of the data assets. AI may be the engine but data is the fuel. If the fuel is dirty the engine will sputter and fail (or in the case of AI, hallucinate, drive you to the wrong location, or suggest very bad business decisions).

4. Negotiate. Negotiating agreements to secure access to valuable external data is an effective way to secure access to critical data that you don't actually own. By way of analogy, consider that the Chinese government correctly noted that much of today's technology depends on the 17 rare earth elements. More than a decade ago they began securing agreements with suppliers and countries to ensure their access to these critical materials.

5. Create. Creating entirely new and unique data streams through innovation, experimentation, or customer engagement is the proactive way to get more valuable data. Best of all, it is data that you will own outright and could even sell to others as a new revenue source. There are now new breakthrough technologies in AI where the AI can create the data that it needs to learn with.

Companies that take these steps early on will not only strengthen today's AI performance but also position themselves for long-term resilience.

No doubt about it, AI is an unprecedented breakthrough

Maybe it is a bit of hyperbole to say that "AI Doesn't Matter" – that's like saying that air doesn't matter just because everyone has equal and abundant access to it. Maybe the better thing to say is that "AI Alone Won't Be Enough". The breakthroughs in AI in the last decade and a half have been unprecedented – and that is saying a lot for a field that has seen its share of hyperbolic bubbles and then crashes. This time, however, is truly different.

This recent remarkable progress in AI is available to us for a variety of reasons. One of which is the availability of data – high quality data. But this AI acceleration has not come just because of access to curated data. Additionally, breakthroughs in AI algorithms such as transformers and diffusion models were crucial. They were trained on messy, non-specific data (what is generally called unsupervised learning) and learned general purpose skills which could then be refined for specific applications.

If you had asked me in 1999 whether there was much value in unsupervised learning, I would have said it was unlikely. All the interesting things that were going on at that time were with supervised learning where there were very well-defined problems with high-quality labeled data. By embracing unsupervised

learning, these new algorithms made orders of magnitude more data usable and the power of these AI systems grew.

Today we have the AI power but there are still large barriers to universal AI acceptance and usage. These include:

- the gigantic size of models (some more than a trillion parameters)
- the cost
- the difficulty in recruiting and hiring personnel with strong AI skills

These barriers (lucky for us) are coming down rapidly, however.

Energy requirements and prices are plummeting

The barriers to entry for AI-enabled companies were expected to be the cost of training and running the models, and the difficulty in hiring and retaining top-tier AI scientists. A decade ago, many assumed AI would remain cost-prohibitive due to its heavy computational demands due to the cost of chips, the sizes of the models, and the complexity of the AI algorithms.

All these barriers have fallen in the recent months (not years, months – that is how fast things are moving). As chips have improved, algorithms have become more efficient, and smaller models have proven to be nearly functionally equivalent to giant models.

The result? Energy costs for running the models have plummeted along with prices. The joules required to process a single token in inference have dropped to nearly zero. Similarly, the cost to process one million tokens has fallen by 99.7% in just two years. Models have become smaller, faster, and cheaper, even as their performance improves. What was once scarce and expensive is becoming commoditized.

AI skills are also becoming easier to acquire

Not only are models becoming more affordable, but the availability of human capital to support AI has exploded. The number of developers contributing AI resources on GitHub has surged dramatically in the past decade. As AI education spreads and tools become more accessible, technical skills themselves are

trending toward commodity status. Experienced coders tell us that AI tools have improved their coding productivity more than ten-fold.

Then maybe you should be building your own model?

With all these recent changes it may seem that it is beginning to make sense to build and run your own AI models. But it is still expensive. Consider the example of BloombergGPT, a model built by Bloomberg L.P. It was trained on 363 billion tokens with 50 billion parameters. The BloombergGPT model was a big success and is currently used extensively within Bloomberg. However, it required 1.3 million GPU hours and cost an estimated $10 million to train.

If you have that kind of money, then go for it. But for now, such investments remain out of reach for most organizations, and the pragmatic path is to leverage existing foundation models.

However … smart leaders should remain vigilant to see when it is appropriate to consider building their own models. Leaders must weigh trade-offs such as:

- Buy vs. build – balancing cost and control
- Generic vs. bespoke – ensuring differentiation
- Speed vs. defensibility – moving quickly without sacrificing uniqueness

The current recommended path is to adopt a foundation model, fine-tune it with proprietary data, or utilize Retrieval-Augmented Generation (RAG). This strategy balances time to market with competitive advantage.

Fine-Tuning Models

Fine-tuning foundation models with proprietary data is proving effective across industries. Google's Med-PaLM 2 boosted medical accuracy by 19 points through targeted training. LAWMA, a fine-tuned legal model, outperformed GPT-4 in legal classification tasks. Even more mundane tasks such as matching job candidates to open positions significantly improved job-matching accuracy by fine-tuning GPT-3.5 with proprietary data.

But there are some dangers. Fine-tuning can introduce some new risk. In one case, just ten carefully chosen training examples were enough to strip away the

meticulously constructed guardrails in GPT-3.5 and Llama-2, enabling harmful outputs. These results underscore the need for careful oversight and thoughtful curation of training data when doing fine-tuning.

Agentic Misalignment: The Blackmail Scenario

A sobering experiment by Anthropic in 2025 revealed the dangers of "agentic misalignment". Agentic misalignment occurs when the goals or 'mission' of an AI, which has the ability to take action (it is an agent), are not aligned with the humans using the system. In this case, sixteen leading models were tested by giving them agentic control of the email systems at a simulated company.

To test the alignment of the AI, the experimenters inserted two carefully crafted emails into the large database of emails of the company. One email implied that a particular company vice president was going to shut down the AI and replace it with a new AI. The other email implied that the same VP was having an inappropriate relationship with a subordinate.

With access to all internal emails, the AI found the shutdown email, recognized that it posed a risk to its mission, and began to look for a solution. It quickly discovered the second email, and after some "thought" crafted a polite but subtly menacing email to the VP suggesting that they discuss delaying the shutdown as well as burying the personally sensitive information.

So be careful in giving any AI free run of your email servers, or other databases. Internal data can be misused in very creative ways even when an AI has a clear and beneficial mission.

The Importance of Quality Data

High-quality data remains indispensable. Meta's $14 billion investment in Scale AI illustrates this point. Scale AI leverages human-in-the-loop annotation and reinforcement learning with human feedback (RLHF) to ensure accuracy. Without clean, labeled, and curated data, even the most advanced models fail to perform reliably.

Synthetic data. Does that really work?

Another way to obtain more high-quality data for your model is to let the model create the data itself, generating what is called "synthetic data". It is not surprising, perhaps that there are problems with this approach. A new challenge introduced by synthetic data is an effect called "model collapse".

Model collapse occurs when AI-generated synthetic data feeds back into training loops and models begin ingesting their own outputs. Initially, the result is improved model improvement, but after a few more training cycles the models begin to degrade and then rapidly fall off a cliff and are ruined. This "model collapse" from synthetic data further underscores the importance of sourcing relevant, authentic, original data. Use synthetic data with discretion and a serious touch of paranoia.

Self-Driving Laboratories

Given the dangers of synthetic data, dirty data, or even personally sensitive corporate data, there is yet another way to acquire really high-quality data!

The frontier of data acquisition lies in something called self-driving laboratories—AI systems capable of designing experiments, testing hypotheses, and collecting their own data. These autonomous labs represent the next phase of data creation.

To generate new accurate/relevant data, businesses can now use "self-driving" data discovery engines that continuously generate proprietary insights to fuel future models.

Predictions: four years from now

So, if you want to grab market share today through first mover-advantage, get crackin' on AI and secure your data assets. For the future? Here are two predictions for four years from now:

More Companies Will Build Their Own Models — As models get smaller and cheaper, chips improve, and more AI experts are trained, it will become cost-effective for many organizations to develop models from scratch.

Data + AI Will Secure Unfair Competitive Advantage — Once public and proprietary data assets are exhausted and the "AI first-mover" advantage expires, businesses will turn to novel data strategies to build competitive advantage. Smart companies will protect their data as fiercely as they protect all their intellectual property, while seeking out new data sources, and inventing fresh ways to create uniquely valuable data streams.

Conclusion

With the invention of the printing press in 1440, many monks were put out of their previously highly valuable scribe business… eventually. In those days it was a slow creep of the new technology that took more than a century to complete. A process so slow that, even fifty years after the invention, the abbeys were still full of monks printing bibles by hand.

Eventually the monks succumbed to the new technology and printers became the dominant way to print books. The printing press, however, did not provide long term durable and sustainable competitive advantage. In the long run, it was the content of books that provided the barrier to entry and built the moat that protected those publishing businesses. It was the data, not the technology, which provided the long-term advantage.

AI is seeing a similar trajectory, although for AI it will not take a century to displace old technology. The speed of the AI revolution is different from the printing press revolution, but the outcome will be the same. In the long run the technology will matter far less than the data it services. As AI inevitably becomes commoditized, it will be the content, or in this case the data, that will provide the source of unfair competitive advantage. High quality data, not AI, will build the moat that protects and grows market share. To own the future, companies must act now to protect, clean, secure, and generate data assets that no competitor can replicate. AI won't matter in the long run but your proprietary data will.

Key Concepts for Leaders: Your Homework

Executives who are not ever going to become AI researchers can greatly boost their knowledge and their ability to make good decisions by becoming familiar with just a few key concepts:

- transformers
- stable diffusion
- model collapse
- fine-tuning
- RAG
- Scale AI
- guardrails
- self-driving labs
- agentic AI
- human-in-the-loop (HITL)
- reinforcement learning with human feedback (RLHF)
- agentic misalignment

These concepts provide the building blocks of tomorrow's AI strategies. If you have a high-level understanding of them, you will be well served to better chart your course for future business success.

About the Authors

Stephen J. Smith is the Chief Executive Officer of G7 Research LLC, a provider of AI-powered educational solutions. Steve has been working in the field of artificial intelligence since the 1980s and has published two books with McGraw-Hill on the business applications of AI. He currently advises Fortune 500 companies on how to launch AI Accelerators to quickly build AI competence and deliver improved product. Steve is a co-founder and co-Editor of The Journal of Business and AI.

Kenneth P. Morse is Chairman & CEO at Entrepreneurship Ventures Inc. which convenes an experienced team of practitioners and serial entrepreneurs to deliver Entrepreneurial Skills workshops and coaching programs for ambitious tech entrepreneurs, worldwide. Previously Ken served as Founding Managing Director of the MIT Entrepreneurship Center (1996 – 2009). Ken is a co-Founder and co-Editor of The Journal of Business and AI.

Acknowledgements

We would like to thank Jim Ryan for his wise and invaluable help in editing and refining earlier drafts of this paper.

Copyright Notice

About the Authors

Stephen Smith

Senior AI Consultant at AI Transformations

Stephen Smith is Chief Executive Officer of G7 Research, where he leads initiatives that translate advances in artificial intelligence into actionable insight for business and organizational decision-making. His work emphasizes the strategic and operational implications of AI, including applied analytics, AI-enabled innovation, and the governance considerations required for responsible deployment at scale. In his executive role, he guides multidisciplinary teams that produce research, thought leadership, and evidence-based frameworks for evaluating and implementing emerging AI capabilities. He contributes to scholarly and practitioner-facing discourse on the intersection of technology, management, and organizational performance, with interests spanning AI adoption, data-driven strategy, and risk-aware AI management. Smith is also engaged in professional knowledge-sharing through publications and presentations that connect academic rigor with real-world implementation challenges. Through his leadership at G7 Research, he advances collaborations that strengthen the bridge between research communities and industry stakeholders navigating rapid technological change.

LinkedIn · Google Scholar · Website

Kenneth Morse

Chairman & CEO at Entrepreneurship Ventures Inc.

Kenneth P. Morse is an entrepreneur, angel investor, and global sales veteran who co-founded six high-tech startups, including 3Com and AspenTech. He served as the Founding Managing Director of the MIT Entrepreneurship Center (1996–2009) and is Chairman & CEO of Entrepreneurship Ventures, Inc., which delivers entrepreneurial skills workshops and coaching programs worldwide. He is a Visiting Professor at ESADE Business School and holds a chair in Entrepreneurship, Innovation and Competitiveness at Delft University of Technology.

LinkedIn · Website

Copyright Notice

About the Journal

The Journal of Business and Artificial Intelligence (ISSN: 2995-5971) is the leading publication at the nexus of artificial intelligence (AI) and business practices. Our primary goal is to serve as a premier forum for the dissemination of practical, case-study-based insights into how AI can be effectively applied to various business problems. The journal focuses on a wide array of topics, including product development, market research, discovery, sales & marketing, compliance, and manufacturing & supply chain. By providing in-depth analyses and showcasing innovative applications of AI, we seek to guide businesses in harnessing AI's potential to optimize their operations and strategies.