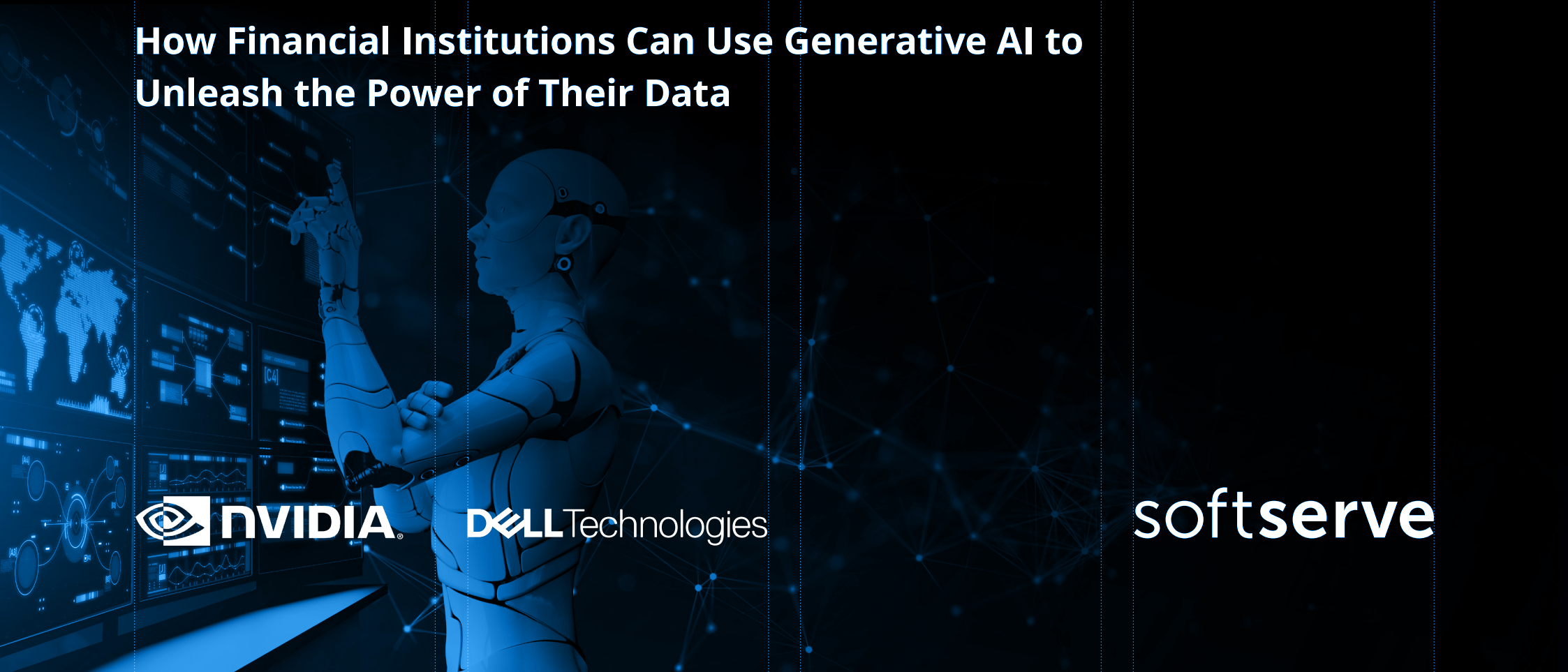
AI OR NOT TO BE?

E Book

THAT IS THE QUESTION FINANCE FIRMS SHOULD BE ASKING

How Financial Institutions Can Use Generative AI to Unleash the Power of Their Data



NVIDIA

DELL Technologies

softserve

INDEX

OVERVIEW

Financial institutions – like Shakespeare's Hamlet - face an existential dilemma.

> *Do they accept the life-saving opportunities offered by Generative AI (Gen AI) that are being dangled tantalisingly in front of them, or look away and face potential oblivion if they don't?*
>
> *Perhaps they need to consider how Hamlet rationalised his decision: "Whether 'tis nobler in the mind to suffer the slings and arrows of outrageous fortune, or to take arms against a sea of troubles and, by opposing, end them."*

It is a debate that should be taking place urgently across banking and insurance boardrooms. Do we suffer in silence and maintain the status quo amid the rising forces of competition and more sophisticated customers? Or do we arm ourselves with the powerful antidotes offered by artificial intelligence (AI), Machine Learning (ML), and Large Language Models (LLMs) to better navigate the potential "sea of troubles" that lies ahead?

This paper examines where Generative AI can make a real difference to finance firms. We discuss how it could be deployed to maximise those opportunities. Apart from the right hardware and software, strategic implementation will be crucial. It will ensure that Gen AI incorporation is safe and customised, but sophisticated and flexible so any organisation can proceed at its own pace.

PROVIDE CONFIDENCE

This paper debunks some of the myths surrounding AI. It also puts in context the benefits of planning a professional Enterprise AI platform, versus the more consumer orientated ChatGPT or its commercialised twin, Co-Pilot. By explaining the rigorous safeguards that can be implemented, it will also provide the confidence and security expected by shareholders, customers, and regulators.

This paper aims to empower executives in financial institutions with the knowledge to ask the right questions and lead better informed debate on the use of AI in their organisations. It will show why the adoption of AI and its array of applications offers businesses a critical competitive advantage in the areas that are priorities for them.

But, most importantly, this paper showcases how exciting the future of Gen AI-embedded businesses can be for customers and employees. To paraphrase Hamlet at the conclusion of his "To be, or not to be" soliloquy when he realises what he must do: **the time to act is now.**

That is why SoftServe has combined with Dell Technologies and NVIDIA to show how three leading technology providers have pooled their AI resources and expertise to deliver optimum outcomes for those financial institutions that want to embrace the future of Gen AI. Much of the narrative in this paper is leveraged from presentations by experts from the three companies.

INTRODUCTION

Financial institutions are no strangers to technology. They have been and remain among the biggest spenders on tech and continue to grow as industry estimates suggest global banking and investment IT budgets will be around $760 billion this year, up around 20% in the past two years. This has been driven largely by a shift from in-house builds towards a mixture of AI spend and other third-party solutions that generate value more quickly. But while capital spend has declined, spending on services and operations that provide more flexibility and agility have seen sustained gains.

Clearly there is an appetite for change in the industry. But this situation and the dynamics now driving new decisions also largely reflect the industry's accumulated or legacy problems. Many banks have built significant IT estates, multiple data centres and portfolios of software and applications that are now cumbersome and have become unwieldy to operate and maintain.

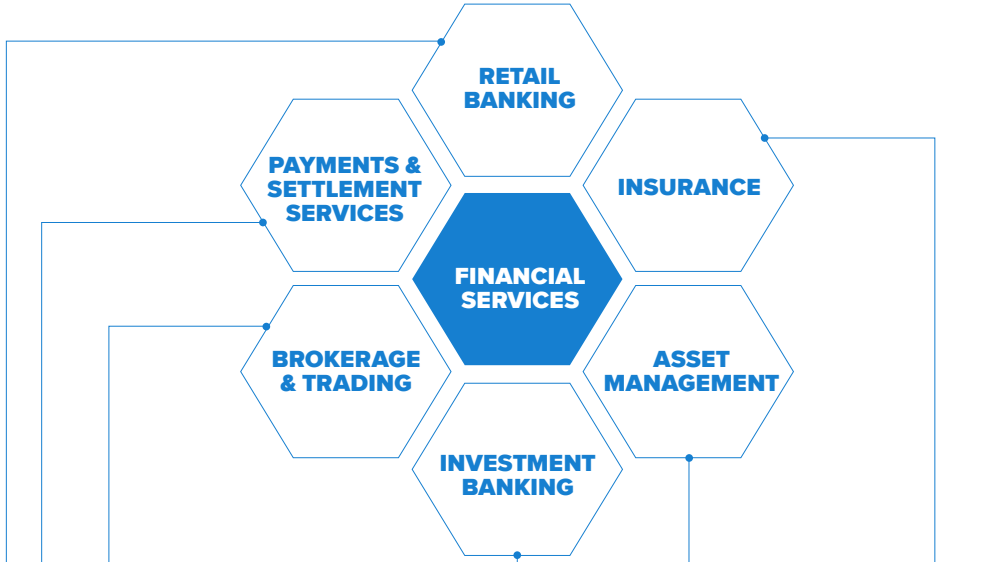
This legacy sprawl also makes it more difficult for incumbents to compete with newer, more nimble competitors who are better able to offer the services customers want, when they need them, at lower cost.

Banks therefore now urgently need a way to keep up. This can come from the raw power and sophisticated analytics Gen AI offers to turbocharge their extensive data resources as a competitive response to the newcomers.

Banks and insurers have historically not been "first movers", or even "early adopters" when it comes to new IT capabilities. However, AI is not new and has been around in various forms for decades.

What has changed recently has been the ability to deploy AI more economically and at scale. It accelerates data processing capabilities that transform data analysis into actionable insights. It should therefore be an open door for banks to walk through, despite historic reservations and a perceived shortage of qualified skills.

GENERATIVE AI BUSINESS SOLUTIONS IN FSI



Algorithmic Trading Strategies Marked
Prediction and Trend Analysis Automated
Trading Signal Generation

Underwriting and Risk Analysis
Claims Processing
Customer Service and Chatbots

AML Monitoring
Transaction Monitoring
Consistency Checking

Portfolio Generation and
OptimizationPredictive AnalyticsAlgorithmic
Trading

Risk Assessment and Fraud Detection
Customer Experience and Personalization
Portfolio Optimization

Mergers & Acquisitions (M&A) Deal Analysis
Financial Modeling & Valuation Automated
Investment Research Reports

WHAT IS GENERATIVE AI?

At its most basic, Gen AI enables users to quickly generate new content based on a variety of inputs. These inputs can include text, images, sounds, animation, code or other types of data. It uses neural networks to identify patterns and structures within the existing data sets to generate that new content.

In essence it automates different learning approaches. These enable organizations to quickly deploy large amounts of data to create foundation models that can be used as a base for AI systems.

Ultimately, Gen AI becomes a powerful tool streamlining the workflow of creatives, engineers, researchers, scientists and more, with use cases that span all industries and individuals.

In financial services, where there is an abundance of data, Gen AI deployment can be at the heart of automating many manual processes. Areas including regulation and compliance, through balance sheet analysis and risk management, to pre-investment decision-making and smarter customer engagement all stand to benefit from it.

In fact, there are few financial firms processes that cannot be improved by Gen AI. These range from back-office clearing and settlement through to middle and front office activities. That is why it is imperative finance firms need to consider Gen AI.

KEY ADVANTAGES OF GENERATIVE AI

Delivers Significant Business Benefits

Makes Data an Asset not Liability

Works OnPrem, Cloud, or Hybrid

Strengthens Security & Cuts Fraud

Boosts Customer Experience

Transforms Data into Actionable Intelligence

Partners Have the Skills to Implement

BUILDING BLOCKS

There are three main building blocks that are essential to provide a platform for the successful deployment of Gen AI in a financial institution.

The right combinations of hardware, software, and systems integration

All being employed in tandem

Via a mixture of On-prem, Hybrid or Cloud services

Done correctly, this will transform the way finance firms operate by liberating data which should be businesses' primary asset, not a millstone. At the heart of that smarter way of doing business will then be the various ways Gen AI and its accompanying tools can be leveraged.

But there are some important rules to remember, including that there can be no Gen AI without data. The deployment on fine-tuned and/or customer trained LLM models means there first needs to be a deep evaluation of customer data ecosystems. This helps understand how to master value extraction and create models with enough accuracy and personalization to guarantee ROI and a positive impact on the business with proper use cases.

The second is the "garbage-in-garbage-out" rule. Data quality and proper modern governance are critical, so understanding your data and what you want to do with it is a pre-requisite foundation.

For finance firms, Gen AI will strengthen performance and latency, reduce costs, improve customer experience, and satisfy regulatory compliance. And all within a trustworthy AI environment that builds confidence in the technology.

SoftServe understands the stack, technologies, FSI use cases, technical jargon, and specific industry challenges to ensure smooth implementation. Working with Dell Technologies and NVIDIA, SoftServe experts can utilize the right hardware for whichever option of in-house or hybrid cloud is selected to then enable the smartest Gen AI software and drivers to do their job.

HARDWARE TO ARCHITECT THE FUTURE

When re-architecting existing hardware infrastructure it is important to prioritize two main factors.

First, it is necessary to integrate design flexibility into hybrid infrastructure platforms, with an Open Data Ecosystem. This allows for the full scope of the Gen AI potential to emerge and deliver the required scale. Second, don't bite off more than you can chew. Instead, focus on a few specific use cases where Gen AI assets can be deployed. This will deliver the required trusted business outcomes without overwhelming implementation.

Gen AI offers financial institutions numerous deployment opportunities where it can improve outcomes. However, many projects fail because they try to "conquer too much too quickly." By maintaining a focus on where there can be measurable and tangible outcomes the benefits of Gen AI will be more likely to emerge.

Research from Dell Technologies shows that 69% of the 5,000 plus firms polled are struggling to turn data into real-time insights, and 67% say they do not have a holistic end-to-end security strategy. This combination of vulnerabilities suggests businesses are neither fully equipped to compete, nor are they secure.

A further 56% said they are battling with complexity at "the edge", which can offer a huge new source of valuable data, while some 41% are already concerned about the rising costs of cloud computing. But these can both be addressed with correct planning and implementation.

IDEAS TO INNOVATION

Firms must consider certain key factors if they want to accelerate from ideas to innovation. At the core should be an integrated strategy from desktops to data centres that ensures a shift away from previously siloed environments and liberates access to the requisite data.

Historic configurations and business working practises often saw data scientists, subject matter experts, data engineers and AI/ML talent all operating autonomously in different silos. Those configurations hindered efficiency and resource utilisation.

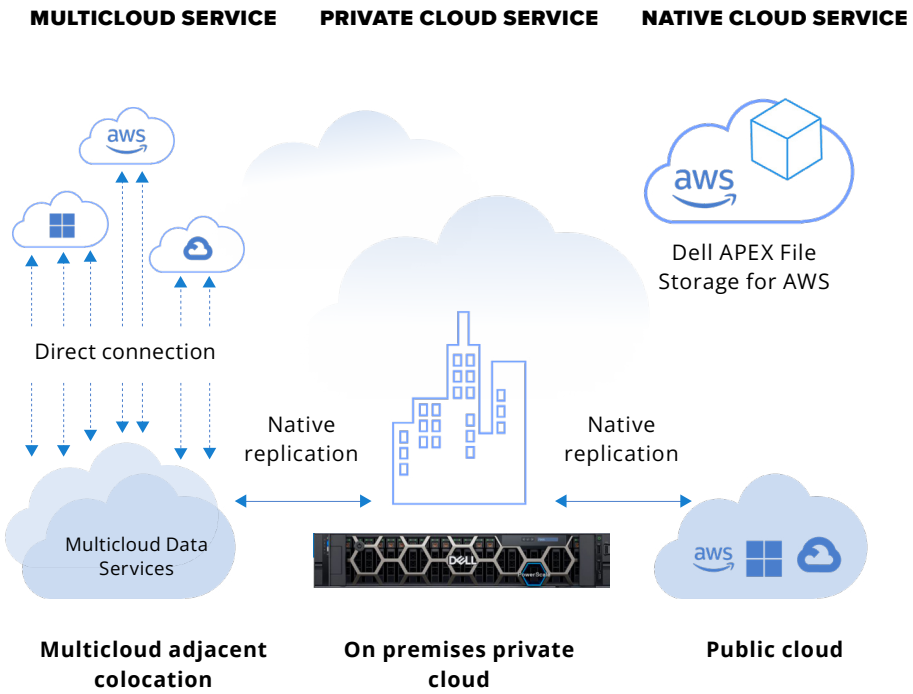
These silos can only be dismantled by the application of a unified, cloud-native, hardware stack that brings everyone into the same environment. This ensures a flexible catalogue for data processing and machine learning, alongside the required security and privacy considerations.

The platform can then deliver valuable innovation through data utilisation, transforming it into actionable intelligence for trusted outcomes.

Previously this has often failed as firms have either tried to do too much too quickly or have invested in the equipment before deciding on their goals. To avoid these mistakes, firms should start by deciding what IT environment works best. Is it private, public, hybrid or edge cloud, or a combination of those? Preparation will also consider what latency and throughput standards are required.

DEPLOY YOUR PLATFORM ANYWHERE YOUR DATA IS....

NVIDIA ENTERPRISE



TRAINED MODELS

Once firms have decided the right IT environment to meet their operational needs, they will then have to determine how the trained models will be brought into production and via what process. They will need to consider which model approaches work best, and which tools the data scientists will use to build, train, and prove those models.

Then, more fundamentally, firms need to ask what data they are going to use to create these new benefits? Where is this data stored and what needs to be done to prepare it and make it useful? These might sound like obvious considerations, but many firms have overlooked them to their detriment.

Lastly, firms need to make sure a specific use case is identified and what the desired strategic impact should be. Define its strategic impact, feasibility and, critically, the measurement criteria.

Once those basics are addressed, firms will be able to consider how the enterprise can effectively be “re-thought” through the application of Gen AI. The deployment of LLMs can then synthesize billions of data points to transform that information into actionable intelligence enabling rapid and scalable insights.

ACCELERATE OUTCOMES

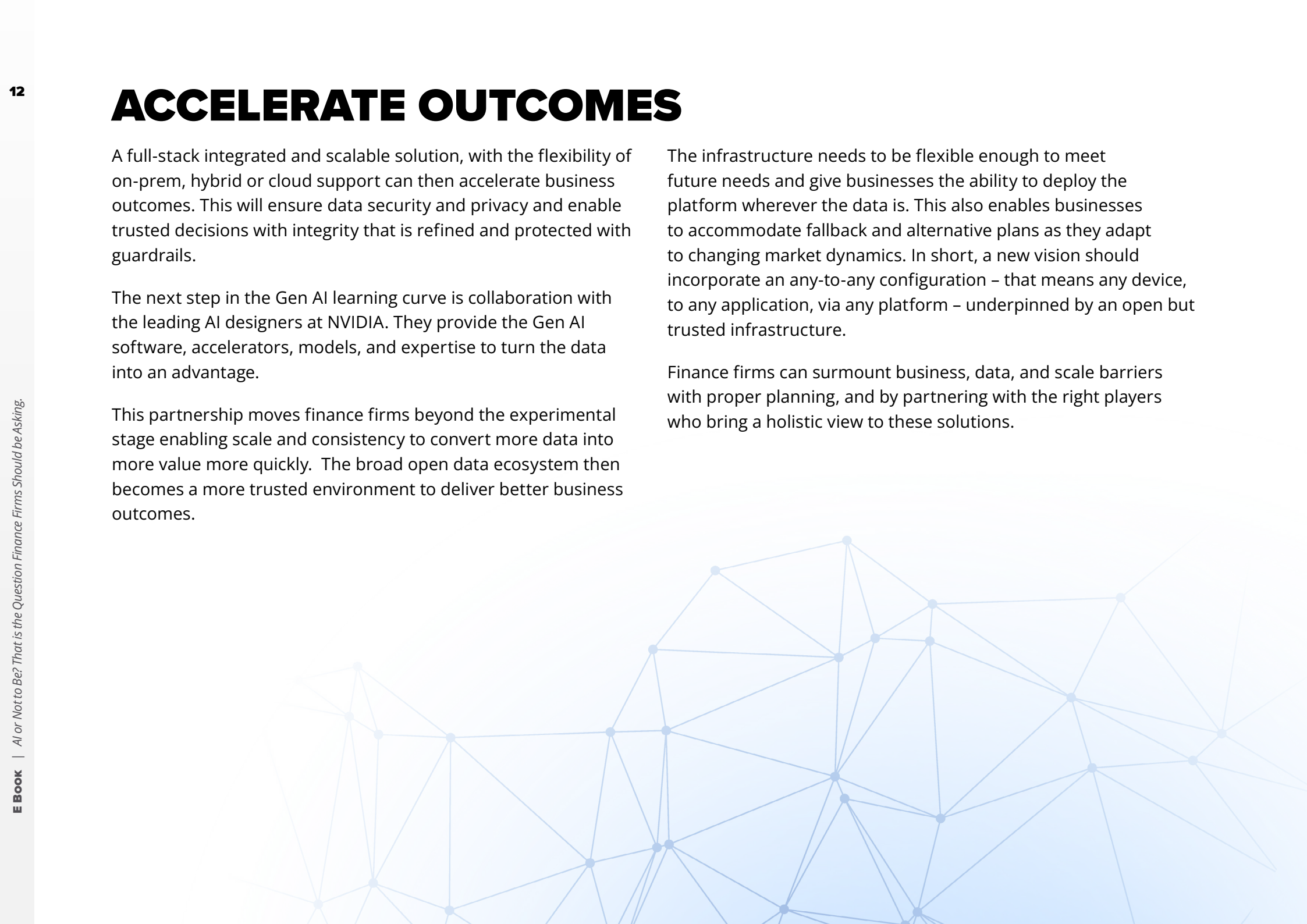
A full-stack integrated and scalable solution, with the flexibility of on-prem, hybrid or cloud support can then accelerate business outcomes. This will ensure data security and privacy and enable trusted decisions with integrity that is refined and protected with guardrails.

The next step in the Gen AI learning curve is collaboration with the leading AI designers at NVIDIA. They provide the Gen AI software, accelerators, models, and expertise to turn the data into an advantage.

This partnership moves finance firms beyond the experimental stage enabling scale and consistency to convert more data into more value more quickly. The broad open data ecosystem then becomes a more trusted environment to deliver better business outcomes.

The infrastructure needs to be flexible enough to meet future needs and give businesses the ability to deploy the platform wherever the data is. This also enables businesses to accommodate fallback and alternative plans as they adapt to changing market dynamics. In short, a new vision should incorporate an any-to-any configuration – that means any device, to any application, via any platform – underpinned by an open but trusted infrastructure.

Finance firms can surmount business, data, and scale barriers with proper planning, and by partnering with the right players who bring a holistic view to these solutions.



SECRET SAUCE SOFTWARE

Once the hardware foundations are established, the next step is to consider how and where the secret sauce of Gen AI is applied to deliver accelerated data science for financial institutions.

A survey by NVIDIA of over 500 financial services professionals found that the future for AI is bright. It showed that executive support for AI is “at a new high” and that progress is already being made across many dimensions. The top priority of 26% of respondents was to explore the opportunities offered by Natural Language Processing (NLP) and LLMs.

However, several specific business use cases also featured high on their list. This included additional priorities such as:

23%

Portfolio optimization

22%

Fraud prevention and detection

21%

Algorithmic trading

20%

Conversational AI

20%

Marketing optimization tools

There was a further long list of areas that were expected to be improved by AI investments.

HEADWINDS SURMOUNTABLE

Headwinds obstructing the rush to Gen AI adoption certainly exist, including difficult macroeconomic environments, challenges in identifying the right strategic approach to AI infrastructure, and potential changes to AI implementation legislation.

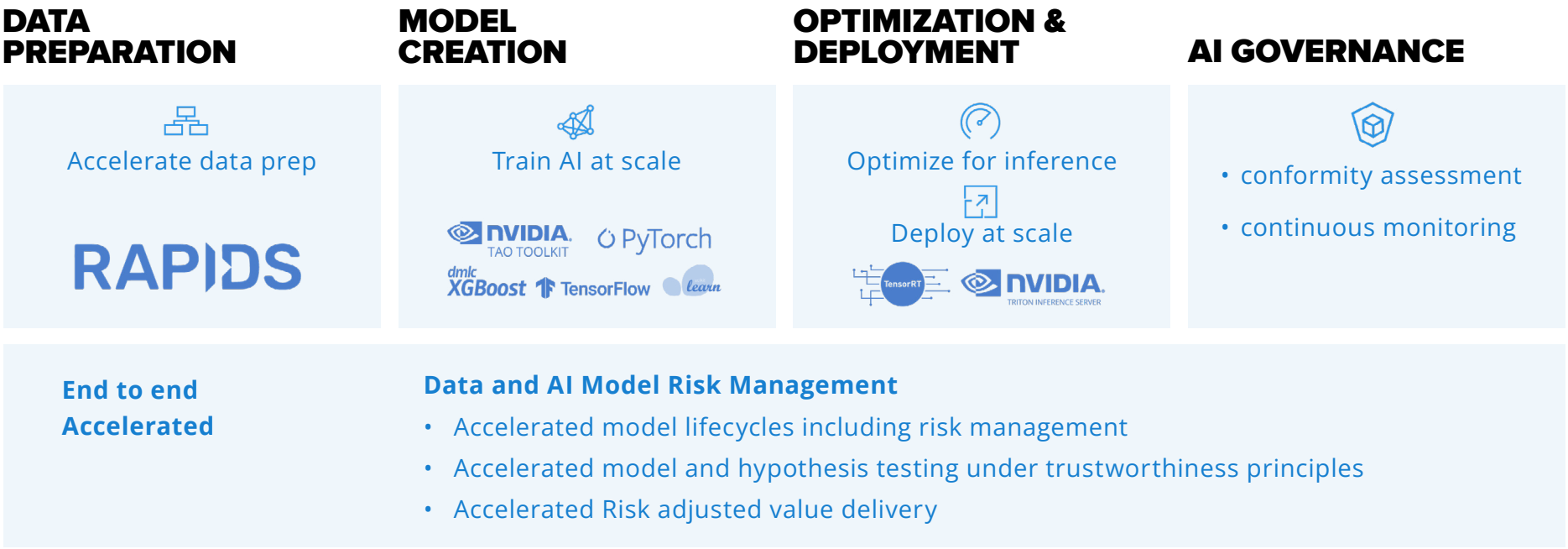
But an integrated and cohesive approach to AI data and an AI enterprise platform enables an approach that is trustworthy, purpose-driven, sustainable, safe, business-aligned and energy-efficient.

Accelerated computing and enterprise AI benefits derive from moving the compute-intensive functions of the business from traditional Central Processing Units (CPU) to Graphics Processing Units (GPU). These advanced microchips can process vast amounts of data simultaneously to accelerate results at significantly lower energy costs. This enables productivity at scale that not only produces greater returns of investment (ROI), but significantly lowers total cost of ownership (TCO).

MAXIMISE RESOURCES

GPU technology empowers businesses to accelerate workloads facilitating easier testing and training of hypotheses and models. This translates into cutting the time taken to execute specific actions from days and hours to minutes and seconds.

ACCELERATED ENTERPRISE AI PLATFORMS EMPOWER TRUSTWORTHY AI AT SCALE



It also offers a much higher probability of finding the right model that delivers the desired benefits and maximises data science resources. Using end-to-end GPU-accelerated data science, these actions become the bedrock of sustainable computing.

For this to flourish the business will require an AI enterprise platform. That should have at its core the key elements of accelerated infrastructure, GPUs, an orchestration layer, plus AI and data science deployment tools and rounded off with the enterprise MLOps platform.

Together, this combination of hardware and software enables the financial business to deploy pre-trained models and newly created AI models to realise the true value of enterprise data. The business will then have access to AI-streamlined and accelerated computing from concept to production at scale.

“AI’S iPHONE MOMENT”

In 2023, AI created a lot of headlines when ChatGPT attracted 100 million users within two months of launch, comfortably creating a record for a new technology. Some referred to this as the “iPhone moment for AI.” It certainly captured the world’s attention and signalled to many that AI and accelerated computing had finally arrived.

However, an enterprise requires much more than just buzz, particularly those which have data centres already processing mountains of data. That extensive, continuous data output makes them prime candidates for the deployment of AI software. In fact, those businesses can now become their own giant AI factories, using their own data to manufacture intelligence for their own benefit.

The pivotal challenge to making a success of AI within the enterprise is building those models and being able to use the models. That necessitates significant compute infrastructure, along with deep expertise and complex algorithms.

STARTING POINTS

To overcome these challenges, firms can use customization. There is a starting point for everyone, whether that's using Gen AI-as-a-service, moving towards greater fine-tuning of pre-trained models, or developing custom foundation models that have significant fine-tuning.

To get started and leverage customised LLMs to differentiate the business, firms should take these initial actions:

- Identify the business opportunity.
- Build out domain and IT teams.
- Analyse the data required for training/customization.
- Invest in accelerated infrastructure.
- Develop plans for responsible AI.

With those plans in place, firms will be ready to develop the customization techniques that will enable enterprise ready LLMs. This will enhance a range of business capabilities from supply chain forecasting, financial modelling, and sales pipeline analysis to legal contract discovery and more. Lastly, the addition of guardrails will enhance security and ensure the exclusion of everything unwanted outside the desired functional domain.

These guardrails bolster LLM infrastructure and are broadly split into three areas:

Topical guardrails that focus interactions within a specific domain

Safety guardrails that prevent hallucinations or other toxic or misinformative content

Security guardrails that thwart malicious action or power transfers to third-party applications.

Once financial institutions reach this level of Gen AI competency, they will be more competitive. This infrastructure investment creates a force that will transform traditional computers into GPU supercomputing and quantum computing resources. Though some may be content to walk with Gen AI before they run.

IMPLEMENTATION

The third key pillar of a successful Gen AI project is implementation and integration. This will bring together the hardware, software, and processing capabilities to deliver the desired business outcomes. It is effectively the glue that binds the process together.

Implementation enables the hardware and software experts to focus on their skills. This leaves the third element of the partnership to work with the client to ensure successful integration and application. It bridges communications between various teams, fills skills gaps, maintains industry compliance commitments and safety standards, and ensures delivery commitments are met.

Most importantly, it ensures a team with intimate knowledge and experience of working with financial institutions to produce smarter technology outcomes. This is crucial in the developing world of Gen AI and LLMs, where the appetite for the results often exceeds the skill sets to enable it. That talent pool widens by engaging with partners.

For finance firms to harness Gen AI's potential to automate tasks and scale processes, reduce costs, enhance user experience, and accelerate innovation, they need that expert partner. This will help them identify appropriate first use cases and to work at a pace of their choice.

The first step is to help firms ask questions in alignment with their goals, to understand where the resource gaps are and where they want to be. Next is pinpointing what insights they expect to derive from that knowledge, looking at its reasoning, classification, topic recognition and key value extraction. It means working together to establish how to generate new data based on those insights using conversational tools, text and code generation and language translation.

CHALLENGES

Gen AI projects naturally entail many key business challenges. These broadly fall under: people, processes, and technology. In the first the partner complements any skill gaps, in the second it ensures coordination and oversight to monitor and mitigate for potential misuse and hallucinations. Lastly, it means simplifying technical complexity, coordinating internal and external data sources, tackling and integrating legacy systems, and maintaining data and IT security.

INDUSTRY EXPERTISE

The implementation partner must grasp the specific business opportunities for financial institutions, where Gen AI can make the biggest difference in desired outcomes. These include:

:For insurance	**:For retail banking**	**:For capital markets**
Underwriting	Fraud detection	Portfolio optimization
Risk analysis	Customer experience	Algorithmic trading
Claims processing	Personalization	M&A analysis

Another promising area for Gen AI in financial services is efficient Environment, Sustainability & Governance (ESG) standards observance and analysis. Here considerable volumes of different types of data from multiple sources are required to produce comprehensive assessments. These can now be accumulated, stored, analysed, and turned into actionable knowledge at speed and scale.

From an implementation perspective there are several key pillars to consider when embarking on a Gen AI project. These are: Data Accessibility; The AI Engine; Trustworthy AI; Lifecycle Management.

The partner can then ensure that the most relevant data (both structured and unstructured) is deployed and that the right people have access to it. Steps are also taken to apply the appropriate data governance standards.

MODEL ADAPTION

For the AI Engine, the integrated Large Language Models, processes are initially established to deploy pre-trained foundation models, then developed alongside proprietary and open-source models. These building blocks then pave the way for greater customization and model adaption.

In tandem are steps to ensure trust and confidence in data use and security, particularly with an eye on pending regulations. This involves the application of Explainable AI, where actions can embrace the required fairness, transparency, and accountability.

Finally, lifecycle management processes are put in place to monitor and prevent potential concept drift and maintain active learning for the models. This includes frequent user feedback, to reflect the data's constantly changing environment.

To manage these outcomes to the optimum standards the integration team must also demonstrate multi-disciplinary expertise. It will embrace design, innovation, and business analysis across what has become an intelligent enterprise. This also includes close coordination with in-house IT resources to maintain critical services for all platforms and solutions.

The outcome that embraces Gen AI, advanced analytics and the MLOps and infrastructure maintenance ensures the successful deployment and ongoing development of the new capabilities.

CONCLUSION

Returning to our original dilemma, later in the play Hamlet continues his self-examination, asking what he could do if he had “the motive and the cue” to act. He questions what reason (motive) he has to resolve his problem and, if so, whether he has the passion (cue) to actually fulfil it.

Shakespeare is often credited with being ahead of his time. But he might well have foreshadowed the dilemmas of modern-day finance with his version of a Call To Action (CTA). He knows what needs to be done, the reasons appear obvious, but does he have the confidence to commit to it?

There seems little doubt that the adoption and implementation of Generative AI across large parts of banking and insurance enterprises will make those businesses faster, leaner, smarter, and easier for customers and staff. It can drive growth and competitiveness.

We also see how to avoid potential pitfalls with a staged introduction of Gen AI aimed at specific, pre-identified business challenges in ways that can be measured, enhance ROI and lower TCO by using partners with experience across the discipline.

Importantly, Gen AI can make a business more secure, not less so, with automation, predictive analytics, early warning signals and guardrails. These all combine to provide confidence to the business and its customers that their most important asset – data – is being treated with respect.

So once the situation and the opportunity has been fully assessed it should no longer be a question of “AI or not to be.” But rather how did we exist without Generative AI? Those businesses should therefore certainly have the motive and now, hopefully, also the passion (cue) to embrace it.

The future for finance looks increasingly interesting.

Please contact SoftServe, Dell Technologies or NVIDIA for more detailed information.

softserveinc dell nvidia

About SoftServe

SoftServe is a premier IT consulting and digital services provider. We expand the horizon of new technologies to solve today's complex business challenges and achieve meaningful outcomes for our clients. Our boundless curiosity drives us to explore and reimagine the art of the possible. Clients confidently rely on SoftServe to architect and execute mature and innovative capabilities, such as digital engineering, data and analytics, cloud, and AI/ML.

Our global reputation is gained from more than 30 years of experience delivering superior digital solutions at exceptional speed by top-tier engineering talent to enterprise industries, including high tech, financial services, healthcare, life sciences, retail, energy, and manufacturing. Visit our **website**, **blog**, **LinkedIn**, **Facebook**, and **X (Twitter)** pages for more information.

Contacts

AUSTIN HQ

201 W 5th Street, Suite 1550
Austin, TX 78701 USA
+1-866-687-3588 (USA)
+1-647-948-7638 (Canada)

LONDON

30 Cannon Street
London EC4M 6XH
+44-203-807-01-41
Toll free: 0-800-05-13-820

FRANKFURT - OFFICE 2

Taunusanlange 9-10, Marienturm
60329 Frankfurt am Main
+49-0-69-505-060-4281
Toll free: 0-800-18-90-559

DUBAI OFFICE

Building No 16,
Dubai Internet City
Dubai
+97180003111269

softserve