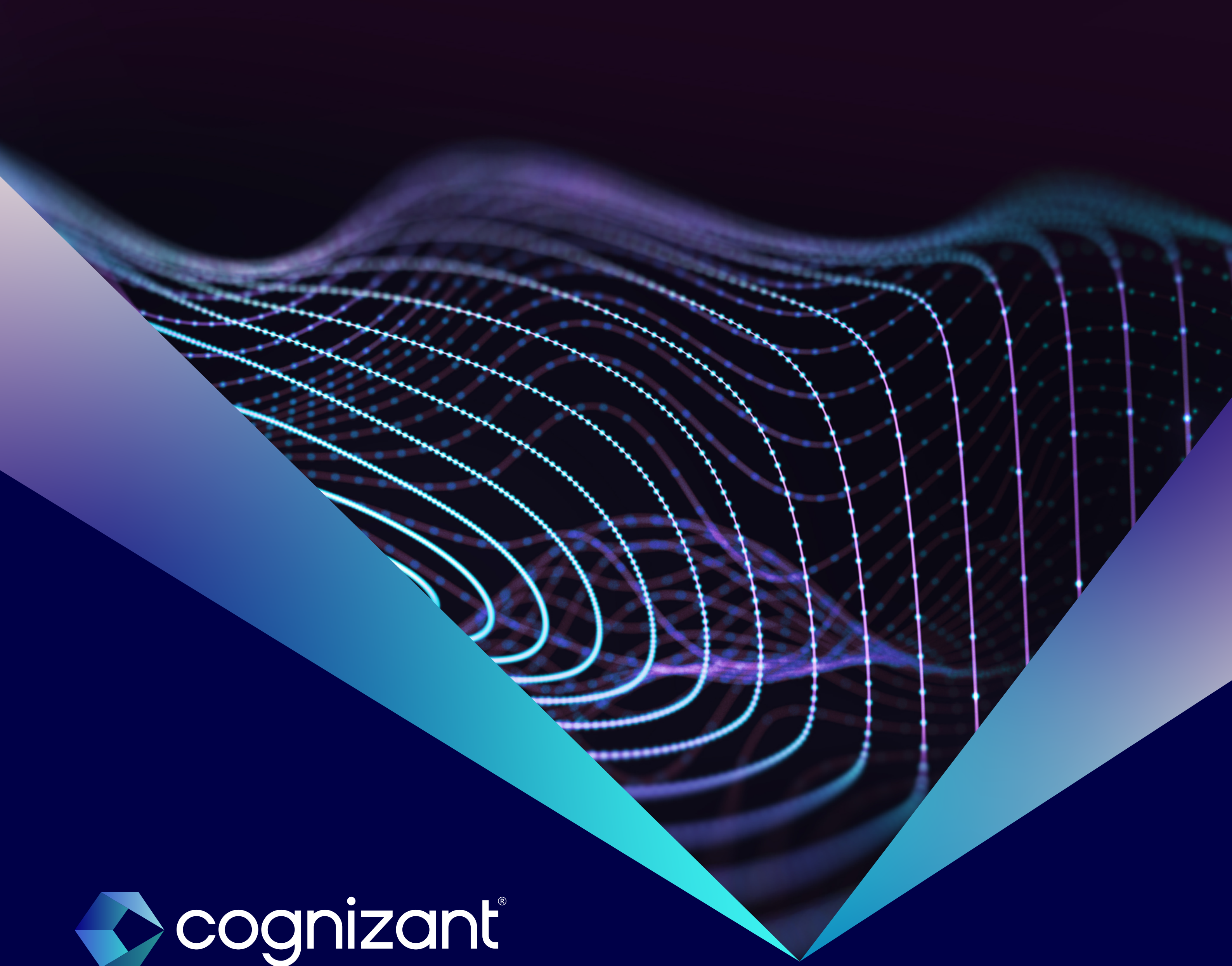


How financial firms can maximize value, minimize risk with gen AI

Part 2 of 3

By addressing 10 big risks of generative AI during the systems design phase, financial institutions and insurers can capitalize on this fast-evolving technology.

By Ed Merchant and Babak Hodjat

What's inside

Introduction

Scores of financial services and insurance institutions are investigating generative AI and how it can boost customer service, adviser proficiency, systems development capabilities and process efficiency and effectiveness. They are also looking to automate onerous manual activities, which could reduce their costs dramatically.

Nonetheless, they must manage 10 big risks of generative AI, especially in the design phase of systems development. In this report, we explore these risks and how to reduce them.

Executive summary

In March 2023, ChatGPT inventor OpenAI announced it was working with investment firm Morgan Stanley to help its wealth advisers summon valuable insights in seconds buried in the 100,000 documents it has amassed over the decades. "You essentially have the knowledge of the most knowledgeable person in wealth management—instantly," said a Morgan Stanley executive in a press release. "We believe that is a transformative capability for our company."[i] By September, its advisers had full access to the system, called the AI @ Morgan Stanley Assistant.[ii]

Morgan Stanley is by no means alone. Generative AI excitement has spread to all corners of financial services and insurance. For example:

- At its 2023 investor day, JPMorgan Chase estimated its own tech spending to grow by $1 billion this year to $15.3 billion, including salaries for software engineers, cybersecurity and AI. The big, diversified bank expected AI to generate $1.5 billion in business value by year-end 2023.[iii]

- A significant (but unspecified) portion of The Travelers' $1.5 billion 2023 IT budget was devoted to generative AI. Alan Schnitzer, chairman and CEO, told investors on the firm's second-quarter 2023 earnings call that ongoing IT expenditures include "a meaningful increase in investments to develop or require cutting-edge AI capabilities built on modern cloud technology."[iv]

- Visa has invested $500 million in artificial intelligence and data to fight fraud. The cards and payments behemoth uses over 60 different AI capabilities to automate these time-consuming, often manual processes.[v]

- Wells Fargo, Nationwide Building Society (a diversified UK financial institution owned by its customers) and Intesa Sanpaolo (a diversified financial institution based in Italy) participated in a $9 million Series A funding round for startup Hazy, whose technology is aimed at accelerating the development of generative AI systems.[vi]

Announcements like these have put incredible pressure on financial services and insurance firms to invest in generative AI. McKinsey & Co. estimates the economic benefits to banks at $200 billion to $340 billion annually if they pursue particular use cases.[vii] A recent Gartner survey found 68% of executives believe the benefits of generative AI outweigh the risks, compared with just 5% who believe the opposite.[viii]

While we similarly believe the economic potential of generative AI is enormous in financial services and insurance, we also believe the risks (financial, reputational and regulatory) loom just as large—and perhaps larger.

We are not alone. A survey of company board directors by cybersecurity firm Proofpoint found 59% believe generative AI presents a high security risk because it can help bad actors infuse malware into IT systems.[ix] It's no surprise then that many financial institutions are using generative AI in controlled experiments. Some are limiting which teams can experiment with what forms of generative AI. And a few have altogether banned ChatGPT, the leading generative AI large language model (LLM), fearing employees will unwittingly put confidential information into a system that will make it publicly available.[x]

Capitalizing on the upsides and reducing the downsides begins in system design, long before a stitch of code is written or generated

Financial services and insurance firms must deeply understand those risks—i.e., what can go wrong if they don't put guardrails on the ways they tap LLMs, the data those models use and the tools that bring forth that data.

Why conventional design approaches fall short

With evolving specs, fast-evolving AI technology and proliferating use cases, software designers need to work quickly and iteratively. But the 20-plus-year-old Agile software design and development methodology must be updated to keep pace with the technology's swift advance.

What's needed is Agile on steroids, fueled in part by generative AI. Generative AI agents can create a detailed system design plan that spells out each step of the design and development process. It can also speed and enhance the Agile process. For instance, software architects can use it to automate requirements gathering, story/epic creation and code documentation.

But using generative AI in this way doesn't reduce the technology risks that can be mitigated in systems design. These include: misplaced trust, IP infringement, IP loss, orphan code, regulatory challenges, tool/vendor partnerships, unsustainable advantage, audacious overreach, malicious behavior (enabled by exploiting security vulnerabilities) and organ (i.e., system) rejection (see Figure 1).

Core risks of generative AI		Phase 1: System design
		Mitigation recommendations
Unintended consequences	**Misplaced trust** Inaccurate output	• Build in feedback mechanisms (human in the loop) • Employ ensemble models • Enable "explainability"
	IP infringement Unauthorized use of copyrighted or patented content (including software code)	• Conduct regular reviews to ensure patent and copyright protections are not violated • Maintain detailed documentation of design decisions
	IP loss Unwittingly giving away proprietary information	• Conduct regular reviews to ensure patent and copyright ownership is identified early • Strictly implement and enforce third-party and non-disclosure agreements • Develop clear exit procedures for designers and developers
	Orphan code Code developed by people who no longer maintain it	• Establish coding and documentation standards, guidelines and templates • Leverage design copilots (i.e., automated assistants) • Incorporate feature flagging to modify system behavior without changing code

Market evolution	**Regulatory reflux** New government rules and laws on the use of generative AI	• Implement an LLM that can compare existing and emerging regulations with existing rules and suggest modifications/additions
	Tool/vendor roulette Choosing tech vendors that will stay in business	• Employ existing policies that have been successfully used to mitigate vendor stability and longevity risks, such as contractual provisions to own source code in the case of vendor dissolution • Make sure the "right to hire" clause in vendor contracts can be enabled
	Unsustainable advantage Creating generative AI systems that competitors can easily copy	• Optimize total cost of ownership based on the expected evolution of solution value by architecting for incremental implementation of operational controls in design
Human nature	**Audacious overreach** Creating unrealistic goals that lead to unfulfilled promises and skepticism	• Develop estimation tools that clarify potential costs of achieving stated goals • Develop LLMs that educate non-tech executives on implementation considerations
	Malicious behavior Leaving your systems open to bad actors who exploit security vulnerabilities	• Address through standard security policies, processes and infrastructure, such as access and anomaly controls • Perform audits
	Organ rejection Employees, customers and other system users who don't take advantage of powerful new generative AI capabilities	• Invest upfront in usability design focused on human augmentation rather than replacement

Source: Cognizant
Figure 1

Unintended consequences

Misplaced trust

Financial institutions are built on trust. Without it, consumers and corporations will not do business with a bank, wealth manager, insurer or card/payment provider.

Providing inaccurate information to customers is a big risk of generative AI systems, and thus a major potential source of losing their trust. Just because a system can comb through millions more documents in seconds, it doesn't mean the answers it generates to a user's questions are always true. In fact, the frequency of wrong answers is a reality in these early days of the technology.

In addition, even if some of the answers aren't wrong, the algorithms that drive those answers may be biased. An inability to explain results generated by a generative AI system will surely get a financial institution in trouble if regulators want to know why, for example, certain customer segments are getting much higher rejection rates or higher prices than other segments. This will undermine consumer and government trust.

Bias can enter generative AI systems from flawed foundation models (i.e., models that contain prejudiced training data). Bias is typically addressed through fairness and ethics policies as well as regulatory compliance. Bias, ethics, fairness, misrepresentation and misinformation on products and services must be central to the design (see sidebar).

How to reduce this risk

During system design, limit the answers that a generative AI system can provide. This means only permitting questions—so-called "prompts"[xi]—specific to that financial institution. The system's response to any non-compliant prompt should be designed to politely say, "Sorry, that's out of scope. I can't answer you there."

Generative AI software companies don't like to reveal how they built their foundational LLMs. It's a source of their competitive advantage. However, that can make it difficult for them to explain the factors that generated output to a particular prompt. (Generative AI vendors keep that insight to themselves—if they totally understand how their models work at all.)

Trust, however, can be maintained by controlling the structures and techniques used to develop the prompts.

Prompt design strategies that can increase trust do the following:

- Provide relevant data or specify sources
- Specify factors that should not be considered when generating the response
- Limit possible outcomes (e.g., indicate the answer must be on only X choices)
- Provide data templates for prompt design that will lead to greater predictability in interpretation/output
- Specify the perspective from which the answer should be generated (i.e., the prompt asks the LLM for a response that reflects the interests/priorities/general knowledge of a particular persona)

Designing privacy into the system

Trust is pivotal to the acceptance of any new technology. If system users can't put faith in the answers that a software program delivers, they'll likely stop using it.

But there's another source of trust that a generative AI system can violate: handing over private data. With generative AI (and any information system), keeping sensitive data private is essential. This is especially the case with consumer information and in particular, their personally identifiable information, known in privacy circles as PII.

As generative AI vendors address data loss, gaining consumer confidence will be an uphill climb. In fact, over half (53%) of US adults believe AI of all types "hurts more than it helps" people keep their personal information private, according to Pew Research.[xii] This finding builds on previous Pew Research studies, which found most people feel they have lost control over their personal information in the online world.[xiii]

But given that monitoring and control strategies for generative AI usage are only beginning to evolve, instilling trust will be a challenge. Closely monitoring and complying with shifts and disparities in the global regulatory environment are crucial first steps. (Europe's General Data Protection Regulation, or GDPR,[xiv] and the EU AI Act[xv] are among the best defined and rigorous regulations.) From there, financial institutions must thoroughly deconstruct and reconstruct their data privacy policies to ensure they protect customer data.

Organizations must assess whether their data privacy policies are in line with their new generative AI strategy. Not only do they need to ensure that access controls and anonymization approaches are adequate to protect individual privacy when requests to feed the data into a generative AI engine are initially made, but they also need to keep track of how the data travels during its use. For example, if data is used in a prompt, how can the firm ensure PII is cleared from caches and not inadvertently left exposed in embeddings? If their policies need to change, they must share them across the organization. Ultimately, if the vision is "do no evil," how do you create a design that ensures no harm will happen? That's a tall order.

Tahir Latif, Cognizant's Global Practice Lead for Data Privacy & Responsible AI, contributed to this sidebar.

Data privacy must be designed into the system

Once the policy issues are settled, financial institutions need to create a clear and compliant set of generative AI privacy-conscious design principles. As suggested above,

> The most important design consideration pivots around data traceability (i.e., clearly understanding where the data resides throughout its travels across systems inside and outside of the four walls).

A closely related issue is creation of mechanisms to enable appropriate retrieval of data upon request (e.g., per data retention regulations) and deletion of data (e.g., as mandated by GDPR). In today's interlinked cloud-enabled, digital world, it can be a serious challenge to demonstrate data is purged from all locations where it has been stored if not properly controlled.

A third key design consideration is telling customers how the financial institution will use their data. A generative AI system must unequivocally explain this.

Data privacy is everyone's concern

> System designers must realize that privacy is their responsibility—not just the concern of the data privacy team

The design team will need the input of a much wider group amid the data free-for-all of today's digital world. This means creating a data privacy governance committee within the generative AI group. Some financial institutions already have generative AI centers of excellence (CoE) to centralize knowledge on building and maintaining these systems.

Beyond data scientists and algorithmic modelers, this group should include machine learning DevOps (MLOps) experts, the firm's lawyers and company C-suite leaders. The CoE must have enough people to look into the proliferation of generative AI use cases. The group's charter should start with the core principles of data privacy: lawfulness, fairness and transparency.[xvi]

Lawfulness means that the consumer has given written or oral consent to use their data, their image (i.e., to improve the firm's facial recognition system) or location (i.e., to improve their online experience.

Fairness is about how the processing of a person's financial data could impact them. This means handling such data in a way most people could reasonably expect.

Transparency is about how much a financial institution discloses to customers how it uses their data and who has access to it.

> Addressing data privacy at the design stage may lengthen the generative AI development process. However, it could save money in the long run

For instance, non-compliance penalties can reach more than the 4% maximum of an institution's revenue set by GDPR when generative AI is involved. In fact, fines could reach 6% of revenue due to the vast amount of data crunched in LLMs.[xvii]

IP infringement

A generative AI system that relies on third-party content to train public LLMs could contain information susceptible to copyright infringement. It's not just copyrighted prose and work processes to be concerned about—financial institutions can unknowingly use patented programs to write code or use such programs and the objects they contain, which require a license to legally use.

Banks and insurance companies can also infringe on protected intellectual property if the LLMs they use draw on information that isn't appropriately tagged as protected.

How to reduce this risk:

Proper system design begins and ends with safeguards and quality checks. While both AI and humans should be involved with checking for IP infringement, this is something many system designers are not experienced at doing. Moreover, the scale of this endeavor may be too difficult for human beings to undertake; it's even hard for AI.[xviii]

The best recourse may be to rely on the guarantees of the LLM providers. For example, Microsoft says it will back up organizations if an infringement claim is made on a generative AI system built using its GitHub Copilot development suite.[xix]

Another approach is more basic: Have designers work with the legal department to make sure an LLM they plan to use isn't trained on copyrighted code.

Unfortunately, mitigating the risk of using someone else's IP remains a work in progress. Legal precedents have not been set for determining under what circumstances parties using generative AI can be held accountable for patent infringement.

IP loss

Generative AI systems that use public models trained on sensitive or confidential data could give competitors your proprietary information. Before designing a generative AI system, inventory all data (sourced internally and externally) to ascertain whether third-party content could be accessed through an application programming interface (API) and be infringed upon.

How to reduce this risk:

IP loss requires designers to undertake substantial training. In many cases, designers envision systems that can be broadly queried. And as noted above, a query might require a response that contains organizational IP or other sensitive but non-confidential material.

If the LLM is hosted outside a company firewall, for example, it's usually powered by a public/commercial model. As noted previously, certain queries shouldn't be allowed because they require going outside the organization's security perimeter.

You can use AI to check on your generative AI system to reduce the risk that a query produces an answer with proprietary (to your company) data.

If a financial institution wants to use a powerful public/commercial model, it will have no choice but to allow queries to go outside the firewall.

An in-house open-source model allows the system to be designed with an initial check to say "yay" or "nay" regarding whether the query can tap certain internal data. Again, this should be decided in system design.

LLM providers are also stepping up to the IP loss challenge. Open AI, for instance, claims the enterprise edition of GPT-4 doesn't mix client data with publicly available data. And it guarantees the security of the client's data.[xx]

Although designers should be diligent about the data they specify for use in generative AI systems, the best protection against IP loss is to specify a system development environment that either automates controls for IP leakage or isolates sensitive IP from other systems. In fact, designers should mandate that software developers work in environments that do one or both of the following:

- **Automate the tagging and filtering of protected IP to ensure it cannot be included in data sets exposed to public LLMs.** For example, LLM Shield[xxi] scans company devices for the LLM's input box text (i.e., prompt) before the prompt is sent. It also uses advanced encryption and filtering techniques to secure sensitive data before it can be intercepted, analyzed or stored by LLMs.

- **Implement and train LLMs that are on-premises or accessed via private endpoints.** One way to do this is setting up hardened bastion hosts[xxii] with private connections to cloud LLMs.

Orphan code

Generative AI is expected to provide tools and capabilities that will enable non-techies to become programmers without any software education. You could euphemistically call it "the democratization of software engineering." Or, pessimistically, you could call it "a recipe for orphan code."

By orphan code, we mean code created by business managers and their non-IT staff members that is later abandoned when they leave the organization or lose interest in the system they developed. (It won't likely be in their job descriptions to maintain it.) If that code is still in use, it must be maintained by the corporate IT function and when necessary, connected to core operational systems.

How to reduce this risk:
Standardize how generative AI systems are designed and built. This could include:

- Having prompt templates to standardize what users can query
- Offering libraries of reusable prompts and embeddings
- Providing an AI-enabled tool to recommend library items relevant to developers' needs
- Creating a simulation environment (see sidebar, page XX)

In addition, Implement tools to assess similarities between the code developers are writing and the code in your libraries. These measures should minimize the amount of code that is created and later abandoned.

Using generative AI to simulate innovation and reduce risk

Financial institutions have used simulations for decades to develop and validate strategies, especially in risk management.[xxiii] Central banks such as the US Federal Reserve and the Bank of England use simulations to model the potential impacts of changes in monetary policy. Commercial banks model risk scenarios, assess the impact of market functions and test the resilience of their financial systems.

However, financial institutions rarely use simulations to test innovative new products and processes. Reasons include perceived minimal benefit to being a first mover, fear of simulation unpredictability and complexity, the steep learning curve to apply findings, and insufficient budget and talent.

Generative AI may provide a breakthrough for affordable simulation experimentation. And when used with other machine and deep learning techniques, the technology could significantly reduce design cost and time in the following areas:

Ideation and analysis:

Generative AI can analyze large volumes of text data, including research papers, articles and user feedback, to identify emerging trends and innovative ideas. By understanding the meaning and context of language, the technology can identify patterns, connections and relationships between different concepts, aiding in the generation of innovative ideas.

Modeling competitor and business partner behavior:

Financial institutions can use AI algorithms in simulations to model the behavior of competitors. That, in turn, should make those institutions better prepared to respond to competitor pricing, new product introductions, market entries and other moves. AI can simulate the decision-making processes of different entities, accounting for factors such as intelligence, strategy and resource allocation. This could also reduce the risks of choosing generative AI tools and vendor partners.

Dynamic simulation adjustment:

Simulations often involve dynamic and evolving scenarios. Generative AI can adjust the simulation parameters in real-time based on the actions and decisions of the participants. This ensures that a simulation is relevant throughout the exercise. This is another way to avoid audacious overreach.

Scenario generation:

Generative AI can create diverse and dynamic scenarios for simulation exercises. It can analyze historical data, market trends, political trends and emerging factors to generate realistic and challenging scenarios, including best- and worst-case scenarios to help financial institutions make more informed decisions in a more dynamic world.

Market dynamics:

Banks, wealth managers and insurers can use generative AI to simulate market dynamics, consumer behaviors and the impact of various factors on the success of innovative products or services. In doing so, they can better identify potential risks with innovative ideas. They can simulate scenarios and develop strategies to mitigate risk related to “audacious overreach,” another key risk of generative AI (see page X). They can also simulate the impact of regulatory changes on innovation projects, anticipate compliance issues and adjust strategies accordingly. As noted above, several financial institutions are using deep learning algorithms to assess whether they would be non-compliant with existing regulations.

Learning and adaptation:

Generative AI can learn from the outcomes of simulation exercises and fine-tune its models over time. For example, generative AI can simulate resource allocation of many potential innovation initiatives and recommend optimal funding.

Automated data analysis:

Simulations generate vast amounts of data. Generative AI can automate the analysis of this data and extract meaningful insights. With the clear summaries generative AI can produce, decision-makers can more clearly and quickly understand the potential implications of different strategies.

Virtual prototyping:

Generative AI can simulate the performance of prototypes in virtual environments, which can speed testing and evaluation of innovative ideas before they are implemented. This can reduce audacious overreach by more precisely determining investments in tools, modeling refinements, database servers, cloud computing and other key areas.

While they experiment with and implement generative AI systems, financial institutions should also use the technology to simulate potential new products and services.

Market evolution

Regulatory reflux

Regulations on data privacy, generative AI use and related issues are in flux globally. This places financial institutions that operate across borders at risk.

How to reduce this risk

Design with a close eye on established or emerging regulations in the geographies in which your firm operates.

> Realize that the regulatory arena will be in constant motion and that governments are struggling to keep pace with generative AI's technical advances

For example, the US has not issued regulations—only guidelines via an executive order[xxiv]—on the safe and secure use of generative AI.[xxv]

Using AI to check on AI is critical here. In fact, many financial institutions have used deep learning algorithms for some time to check for regulatory compliance.[xxvi] For example, prior to designing a new generative AI system that gives wealth managers advice on certain investments, a financial services firm could prompt the LLM with a related regulation on equity trading information and ask, "Does this interaction or output meet this regulation?" Large public/commercial models excel at checks like this and can provide a "yes" or "no" answer. As new regulations emerge, they should be added to the model. The algorithms should then be trained on them to ensure any generative system in design remains in compliance.

One European bank with whom we work took a hub-and-spoke approach to regulatory compliance. A central team used AI to check generative AI systems designs for regulatory compliance. They then asked generative AI teams in various country units to double-check any flagged findings to make sure they complied with local regulations. By keeping humans in the loop who are more attuned to local regulations, this bank is proactively scoping out potential regulatory problems before launching new generative AI systems in those markets.

Tool/vendor roulette

Choosing the right generative AI toolset and vendors with staying power is a risky proposition given the technology's embryonic state. A generative AI platform that files for bankruptcy in three years is not likely to be as easy to maintain as one whose owner has a thriving business.

How to reduce this risk:

As in previous technology waves, financial services organizations need to assess trade-offs in the design stage between using multiple vendors whose products may result in lock-in (and thus heavy costs of moving to better alternatives at a later date) and the near-term benefits of a single technology partner. This partner could be a "hyperscaler" (i.e., large cloud-based vendors that provide compute, storage and now generative AI services at scale), a stand-alone commercial LLM player or a best-of-breed tools provider.

> In general, the best defense against overreliance on a single vendor is to design a loosely coupled, modular architecture that separates various functions such as data preprocessing, feature extraction and the actual generative model

Design areas to focus on include:

- **Interface definition:** Modularization via APIs is a given. However, you must evaluate the API choice at the beginning. For example, financial institutions are experimenting with GraphQL (a query language for an API created by a neutral foundation)[xxvii] as the API standard instead of REST (a communications protocol that allows different web-based systems to communicate).[xviii] The reason: the former can support prompts that return data limited to a specific query, single queries that draw responses from multiple resources and real-time updates.

- **Model orchestration and integration:** The design should support the ability to assemble and modify service "chains" to enable technological advancements to be incorporated without an architecture overhaul.

- **Data pipelining:** The architecture must provide a mechanism to manage and orchestrate the flow of multiple data sets.

- **Hyperparameter tuning:** Designs should include a utility that simplifies tuning, making it easier to experiment with multiple generative AI systems configurations.

Unsustainable advantage

As hyperscalers expand their generative AI offerings and make them more affordable, a growing number of providers are now offering solutions specifically tailored to the needs of financial institutions. The net result: banks, wealth managers and insurers of any size have access to similar generative AI capabilities.

As with other waves of technology-driven change, first-mover advantage can quickly deteriorate, undermining the business case that originally inspired adoption.

How to reduce this risk:

Sustainable advantage is derived from things that cannot be readily copied by competitors. Solution features that maintain sustainable advantage include exclusivity (e.g., patents, sole source of distribution), innovation (which implies high degrees of adaptability to market, policy and infrastructure changes, scale and network effects).

Current experiments and pilots across the industry that utilize readily available solutions from hyperscalers and other product vendors are focused on functions and capabilities that will become rapidly commoditized (e.g., chatbots, document summary tools, etc.). This approach is understandable given concerns of potential financial, operational and regulatory risks associated with the use of generative AI in financial services. However, if nearly every company is using the same tools and infrastructures, sustainable advantage can rapidly shrink.

Consequently, financial services institutions must determine whether they have the assets and risk appetite to design solutions capable of maintaining sustained advantage.

Institutions that are prepared to accept this challenge need to make upfront investments to design (and later build) AI infrastructures and operating models that provide the adaptability, vendor independence and scale required to stay ahead of the competition. These architectures also need to support short-term solutions that can establish advantage and be easily discarded without incurring technical debt.

This perspective is supported by our interactions with several large financial institutions, where conversations have shifted from "show me use cases" to "I don't care about specific use cases. I care about the architecture."

For organizations that can't afford to make this type of investment, their design focus should be in two areas:

- **Developing "ecosystem assembly" strategies to affordably maintain pace with market expectations as services become commoditized/productized.** These strategies should seek to avoid platform lock-in as much as practical within total cost of ownership constraints.
- **Increasing focus on safely leveraging proprietary data not readily available in the marketplace.** As the amount of supporting data that can be added to prompts (i.e., allowable tokens[xxix]) increases, developers can use the combination of this data with creative prompt design to generate unique outputs even when designing for commercially available LLMs.

Human nature

Audacious overreach

CEOs, CFOs, CMOs, CIOs and company board members can fall in love with the potential of generative AI, and for good reason. The current flood of pronouncements from futurists, consulting organizations and product vendors inflate these leaders' expectations in two ways:

- Overestimating the capability of the technology
- Overestimating the speed and magnitude of the business impact

These inflated expectations, in turn, create governance challenges as organizations prioritize an ever-growing list of use cases that may (or may not) be based on realistic assessments of implementation complexity or ROI.

Ultimately, overly ambitious objectives can lead to huge, and hugely speculative, investments. If the early returns on those investments fall far short of expectations, initial excitement can quickly turn into skepticism. And great skepticism can too quickly result in leaders pulling the plug on viable experiments.

How to reduce this risk

Two types of design-time controls are needed to guard against audacious overreach:

- **A formal systems design philosophy** and supporting methodologies/playbooks that embrace experimentation and innovation simulation (see sidebar, page 12)
- **Explainable estimation models** that help communicate the cost implications (both build and run) of requested solutions

The design philosophy should be developed by a cross-disciplinary group that spans tech, business functions and strategists (e.g., “the AI council”). This group should be responsible for guiding the overall approach to designing generative AI applications. Representative design principles should include:

- Prioritization of practicality and ease of use.
- Limited system functionality (at least initially) to accelerate development and testing. This will enable rapid learning, adaptation and refinement over time.
- Agreement to not promise anything more than any generative AI system can deliver at this early stage of the technology’s evolution. Time and budget should be allocated to validate assumptions about the capabilities of the technologies used.

A great way to avoid audacious overreach during the design phase is to reality-test the exceptionally high projections on near- to medium-term ROI of generative AI solutions.

For example, many financial institutions are focusing their experiments on optimizing generative AI-powered chatbots. The step change in interaction quality enabled through generative AI can be quite impressive. However, as seductive as the initial pilots may be, a lack of transparency regarding the investment required to scale and operate these solutions, relative to the real business benefit, and the sustainability of that benefit, can lead to either over- or under-investment that ultimately undermines the initiative.

To combat the ROI hype, designers need to create models that help non-technologists understand the implementation and operational costs of generative AI systems, using reality-based metrics that an institution cares most about (i.e., lowering transaction costs, elevating customer satisfaction scores, faster time to market on new products/services, etc.).

By doing so, the top management team could greenlight designs with a much higher probability of success in moving the needle on business impact. Designers will then feel greater accountability and commitment given that the promises they make will be judged against more realistic expectations—not the hype of external generative AI proponents.

To avoid estimation models that resemble “science projects,” designers must pick simpler use cases to show relationships between implementation costs and post-production ROI. However, these use cases still need to be bold enough to demonstrate an impact that is greater than can be achieved using traditional approaches.

A great use case to pursue would be a generative AI system that advises the strategy team on how much to spend (if anything) on a particular acquisition target and then accelerate the due diligence by orders-of-magnitude. At the outset, it may seem audacious, but if properly designed and built, it could generate enormous returns. The reason: the implementation costs of a generative AI-enabled platform capable of summoning all the relevant knowledge a firm has (internally and through public sources) to close a deal would readily be offset by the reduced spend on an army of strategy, finance, legal and compliance experts.

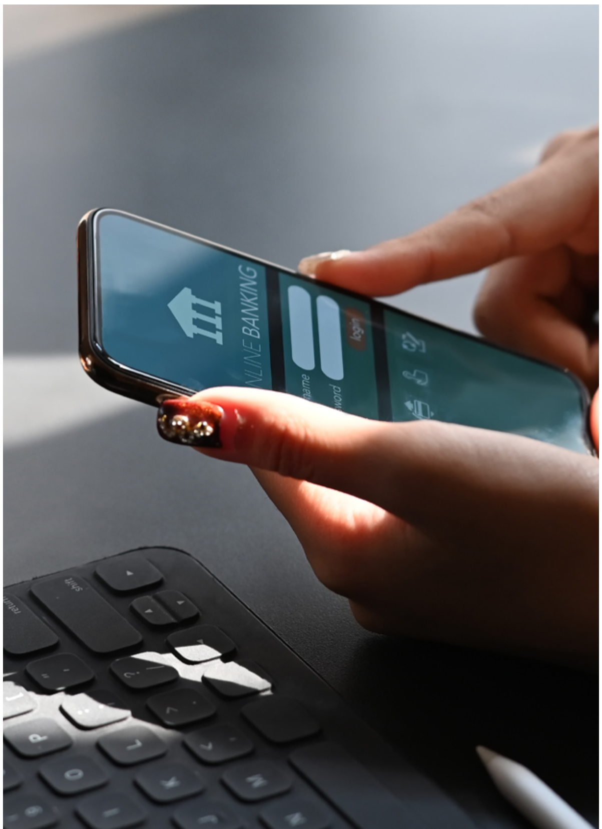
Malicious behavior

Every time a new technology tool emerges, cyber criminals (often working for rogue nation-states or crime syndicates) figure out how to abuse it—sometimes much faster than the good actors. Take a generative AI system designed to reveal how to skim a hundredth of a penny off every financial transaction. A human being may not be smart enough to figure this out. But the generative AI system could be designed to do so. Perhaps the system could reveal the last person to skim a hundredth of a penny from each transaction—and how or why they were caught. But it can also reveal that if the bad actor only skims a hundredth of a penny half of the time in an irregular pattern, they may never catch him.

How to reduce this risk:

If generative AI has real visibility into system configurations and timings of configuration changes, it could reveal when the windows of vulnerability are open. Designers, therefore, should design systems to track logins for a change request.

The design should allow humans to see the precise time of each login and whether the change request was implemented. This would allow IT operations to see which files were changed and what was changed within them. If a bad actor embedded two lines of code somewhere that enabled skimming, it would be detected and handled before money is lost.



Jail breaks and other security shortcomings

Generative AI technology remains a work in progress. Buyers must beware, especially of the security issues. Bad actors have historically exploited shortcomings in immature technologies (see e-commerce, IoT, cloud and social media). Although improving, guardrail protections around public and open source LLMs are porous. That could open Pandora's box and put sensitive data at risk.

In fact, a recent study by researchers from Princeton, Virginia Tech, Stanford and IBM found that existing generative AI guardrails aren't as foolproof as developers contend when they put datasets in pre-trained models like Open AI's GPT 3.5 and 4 and fine-tune them for specific use cases.[xxx] The unintended consequence is what the industry calls a "jail break." It's a form of hacking in which bad actors use creative prompts (known as prompt injection[xxxi]) to trick the system into releasing confidential or sensitive data that an organization would otherwise safeguard.

> One way to stymie potential jail breaks is to understand the implications of prompt injections, and to design in ways that prevent them

Designers also need to consider obvious security loopholes such as open ports[xxxii] in existing cloud architectures.

The design of generative AI systems gives developers the keys to the API. Therefore, designers must determine how to discern regular use from misuse. For instance, if the system is hacked, how do designers prevent it from generating or revealing something that could harm the company. Once again, this means using AI to check AI for security vulnerabilities—a cyber form of white-hat hacking that scales faster and better than human hackers.

Another approach is to create a design where open-ended queries are not permitted. From there, designers must only permit non-ambiguous prompts to ensure responses don't break the firm's ethics code.

Organ rejection

Employees, customers or business partners could be slow to adopt, or reject altogether, generative AI-based solutions for a myriad of reasons. These include:

- A lack of clarity about company and regulatory policies
- A lack of confidence in system outputs due to inflated expectations and/or transparency on why a given output was generated
- A lack of suitable education/training on how to best leverage the new capabilities
- A lack of trust in employer intentions (i.e., "driving productivity improvements" really means "facilitating staff reductions")

How to reduce this risk:

At a high level, any strategy for reducing the chance of organ rejection needs to emphasize usability design.

However, designing generative AI applications (e.g., writing polished prose or generating images and audio) requires a different design mindset from creating core business applications that drive operational efficiencies and organizational productivity (e.g., automating routine tasks, searching/finding/organizing data pulled from a variety of formats). With the latter, usability means ease of navigation and functional predictability (e.g., inputting the right data into a given field will deliver a predictable result). The goal of these applications is typically to automate and eliminate manual effort.

Conversely, generative AI applications need to be designed to augment a knowledge worker's experience and judgement. A core assumption of the design is that humans need to be in the loop and be empowered to override machine-generated output.

Solution designs that make users feel as though the system is an assistant rather than a replacement will overcome potential user reticence.

This is particularly true for a knowledge worker who is highly educated and compensated but fears being displaced by such systems.

A good illustration of this can be found in the adoption of generative AI tools by software developers. Generative AI tools can help save time for development of limited scope components by providing services such as code generation from natural language descriptions, code completion, code review and test case generation. However, the integration of end-to-end solutions still requires the skill of an experienced developer.

For example, a Stanford University study found that coders who used AI tools generated less secure code than those who did not.[xxiii] Therefore, designers must remain mindful of the productivity benefits that such tools provide while also making sure not to alienate developers whose role is paramount in avoiding potential security problems.

Treating insurance claims leakage in design

By Susan Rickard

Claims leakage costs the insurance industry billions of dollars annually, or between 20% and 30% of all claims paid, according to conventional wisdom. Plugging this gaping hole is a big target of many insurers.

That's why some insurers are exploring generative AI systems to help eliminate unnecessary claims outlays. Their objective: give all claims professionals the knowledge of the most effective claims professionals to approve valid claims. Such insights would allow less experienced claims processors to minimize unapproved claims.

Much is at stake. Say a customer submits a claim for an injured foot, the pain of which then migrates to their elbow and to the other side of their body. It's unlikely to be related. But without having access to the right information across a variety of sources (i.e., International Classification of Disease codes, mortality statistics, claims histories and relationship status, etc.), it's hard to establish the claim's veracity. (Note: Some customers may believe such an injury is related, even when it isn't, but they just don't know.) If the insurer pays a claim that includes unrelated maladies, the carrier could be on the hook for related claims the rest of the claimant's life, which would be extremely costly.

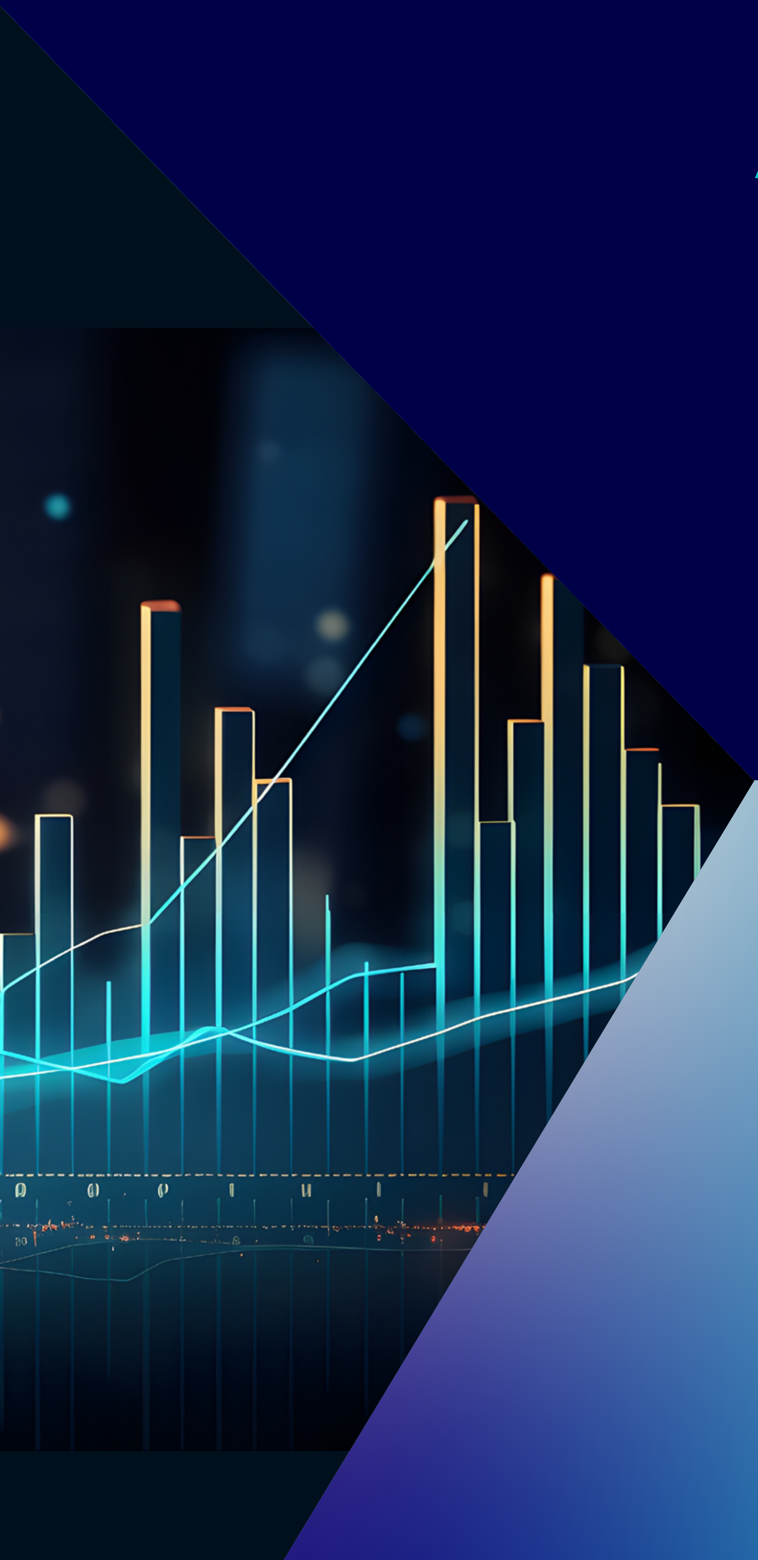
Optimizing claims approvals requires claims professionals to be properly trained. However, it can take carriers up to nine months to train new claims professionals on the vicissitudes of claims processing. And then, it can take years for them to become a top performer.

Some of our insurance clients have asked us to lay out the contours of a generative AI system that acts as a virtual assistant to their claims professionals. Our design focuses on the use of generative AI to seek and find all the information a claims professional needs to effectively process a claim, regardless of where it resides. Importantly, our design approach requires a claims professional to apply human judgment before they approve the claim.

In our experience, as much as 20% of a carrier's claims professionals are its highest performers; and they often receive the most complex claims. The use of a well-designed generative AI virtual assistant could elevate the average claims professional to high-performer status by giving them timely access to the accurate information needed to approve more complex claims. One client told us a small increase in accuracy by claims professionals could significantly reduce leakage, resulting in significant savings.

Amid rising claims volumes, carriers will need all the help they can get. Supplementing human claims professionals with generative AI assistants will reduce stress levels and help carriers retain their most experienced claims professionals. Moreover, a well-designed generative AI system could attract a new generation of claims professional comfortable with the latest technologies and ready to enhance their value contribution by tackling more complex claims.

Conclusion: Getting and keeping generative AI design on-track



As the list of potential applications of generative AI expands and experiments take shape, the time is ripe for financial firms to create a generative AI design guide. This guide should highlight the organization's generative AI vision, ethical code and business objectives, and the numerous risks that can undermine them. We recommend:

- **Quickly size up the financial, operational and reputational risks, and reduce them in design.** Generative AI amplifies some established concerns (e.g., bias) but presents a host of new ones (hallucinations, explanability, etc.). Consider the maturity of your firm's AI usage and governance controls before designing generative AI systems.

  At a minimum, address the explanability of models and results in your design where possible. Design in the ability to display links to third-party copyrighted data used to generate results. Specify that watermarks be placed on all copyrighted imagery to reflect its origin (i.e., content created by generative AI). The sooner you do this, the better you can tackle high-value use cases.

- **Keep humans at the forefront of design.** Since generative AI systems will initially act as virtual assistants and advisors to human beings, design them to keep humans in control. Humans cannot consider generative AI results as veritable truth (as noted above). They must be encouraged to countermand machine output or, at a minimum, provide feedback on everything from prompt utility and data purity through result accuracy.

  Use generative AI to validate both human decisions and decisions made by machines. Organizations have been trying to get employees and computers to check each other's work since the dawn of data processing. But with generative AI, it is now much easier to do.

- **ASAP, explore how to use generative AI to design (and later develop) generative AI systems.** Emerging tools can accelerate design by conceiving new ways of interacting with users or generating synthetic data. This will help build momentum for generative AI.

Generative AI can also speed requirements gathering. It can capture and sum up discussions in ideation sessions and relevant third-party research findings, all of which can stimulate new design approaches. Use generative AI to create design artifacts such as interaction flows. Developers can use them to guide code generation and/or translation across various programming languages.

About the authors

Ed Merchant

Ed Merchant is Head of Consulting, Americas within Cognizant's Banking and Financial Services (BFS) business unit. His group is responsible for advising and assisting CxOs and other senior leaders on strategy execution for technology driven operational improvement, transformation, and innovation initiatives. He has 40-plus years of experience as an engineer and technologist focused on the implementation of mature and emerging technologies—the last 27 years exclusively in banking and financial services. He has deep expertise in helping financial institutions generate tangible business value from leading-edge technologies, including big data, advanced analytics, cloud computing, and now generative AI.

Prior to joining Cognizant, Ed was Global Solution Leader for the BFS division at another major service provider. Previously he held various regional and divisional CIO roles at a top-15 global bank, during which time he also served as Global Head of Architecture for its wholesale banking group. He has also held the position of Principal at a Big 4 consulting firm, leading a large-scale systems architecture, and engineering practice focused on trading and payments platforms.

Ed has an MS degree in mechanical engineering from Fairleigh Dickinson University and a BS in industrial education and technology from Montclair State University.

Babak Hodjat

Babak Hodjat is Cognizant's Chief Technology Officer for Artificial Intelligence. In this role, he leads a team of developers and researchers bringing advanced AI solutions to businesses. Previously, he was co-founder and CEO of Sentient Technologies, where he was responsible for the core technology behind the world's largest distributed artificial intelligence system. Babak was also the founder of the world's first AI-driven hedge-fund, Sentient Investment Management.

Prior to co-founding Sentient Technologies and Investment Management, Babak was senior director of engineering at Sybase iAnywhere, where he led mobile solutions engineering. Before Sybase, Babak was co-founder, CTO and board member of Dejima Inc. Babak is the primary inventor of Dejima's patented, agent-oriented technology applied to intelligent interfaces for mobile and enterprise computing – the technology behind Apple's Siri. Babak has published numerous papers on artificial life, agent-oriented software engineering, and distributed artificial intelligence. He has 32 granted or pending patents to his name. He is considered an expert in numerous fields of AI, including natural language processing, machine learning, genetic algorithms and distributed AI. Babak is also the author of The Konar and the Apple, a collection of short stories based on his youth.

Babak holds a PhD in machine intelligence from Kyushu University, in Fukuoka, Japan.

Endnotes

i. https://openai.com/customer-stories/morgan-stanley

ii. https://www.wealthmanagement.com/technology/wealthstack-roundup-ai-morgan-stanley-assistant-now-live

iii. https://www.americanbanker.com/news/jpmorgan-chase-aims-to-create-1-5-billion-in-value-with-ai-by-yearend#:~:text=One%20of%20the%20pillars%20to,goal%20due%20to%20recent%20results.

iv. https://s26.q4cdn.com/410417801/files/doc_financials/2023/q1/q1/1Q23-TRV-Transcript.pdf

v. https://usa.visa.com/visa-everywhere/blog/bdp/2022/04/18/post-pandemic-economies-demand-1650310496845.html

vi. https://www.finextra.com/newsarticle/42044/big-banks-invest-in-generative-ai-startup-hazy

vii. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/The-economic-potential-of-generative-AI-The-next-productivity-frontier

viii. https://www.gartner.com/en/newsroom/press-releases/2023-05-03-gartner-poll-finds-45-percent-of-executives-say-chatgpt-has-prompted-an-increase-in-ai-investment

ix. https://www.csoonline.com/article/651237/cxos-and-directors-are-growing-wary-of-generative-ai-report.html

x. https://jaxon.ai/list-of-companies-that-have-banned-chatgpt/#:~:text=More%20and%20more%20companies%20are,that%20have%20banned%20ChatGPT%20internally.

xi. By “prompts,” we mean the words that system users type in to get the output they want: answers to research questions, prose or outlines of prose, images, etc.

xii. https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/

xiii. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

xiv. https://gdpr.eu

xv. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

xvi. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/the-principles/lawfulness-fairness-and-transparency/

xvii. https://gdpr.eu/fines

xviii. The likelihood of the large language model verbatim spitting out a copyrighted algorithm is low. But it may share some of the concepts with a modeler who could craft a new algorithm based on the former’s IP. The modeler may change the way the algorithm is described or the variables within. The refined algorithm works, but because of the modifications, it’s hard to detect whether it infringes on any trademarked or copyrighted code.

xix. https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/

xx. https://openai.com/blog/introducing-chatgpt-enterprise

xxi. https://llmshield.com

xxii. A bastion host is a specialized server exposed to the public internet that acts as a secure gateway to a private network. Aside from software essential for rigorous implementation of access controls, it runs minimal services and a software stack to reduce the attack surface (i.e., decreasing risk of vulnerabilities).

xxiii. Pinpointing the “first” use can be challenging due to the evolving nature of strategic thinking. However, one notable early example is the use of war games by the Prussian military strategist Carl von Clausewitz in the 19th century.

xxiv. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

xxv. https://www.nytimes.com/2023/03/03/business/dealbook/lawmakers-ai-regulations.html

xxvi. https://www.kdnuggets.com/2022/05/deep-learning-compliance-checks-new.html

xxvii. https://graphql.org

xxviii. https://www.codecademy.com/article/what-is-rest

xxix. A token is a portion of a word, or characters, used in a prompt. Different generative AI models have different token limits for expressing system requests through prompts; https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them

xxx. https://oecd.ai/en/incidents/45787

xxxi. https://learnprompting.org/docs/prompt_hacking/injection

xxxii. https://www.beyondtrust.com/blog/entry/what-is-an-open-port-what-are-the-security-implications

xxxiii. https://arxiv.org/pdf/2211.03622.pdf cognizant®

Cognizant (Nasdaq-100: CTSH) engineers modern businesses. We help our clients modernize technology, reimagine processes and transform experiences so they can stay ahead in our fast-changing world. Together, we're improving everyday life. See how at **www.cognizant.com** or **@Cognizant**.

World Headquarters

300 Frank W. Burr Blvd.
Suite 36, 6th Floor
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277

European Headquarters

280 Bishopsgate
London
EC2M 4RB, England
Tel: +44 (01) 020 7297 7600

India Operations Headquarters

5/535, Okkiam Thoraipakkam,
Old Mahabalipuram Road,
Chennai, 600 096 India
Tel: 1-800-208-6999
Fax: +91 (01) 44 4209 6060

APAC Headquarters

1 Fusionopolis Link,
Level 5 NEXUS@One-North,
North Tower Singapore 138542
Phone: +65 6812 4000