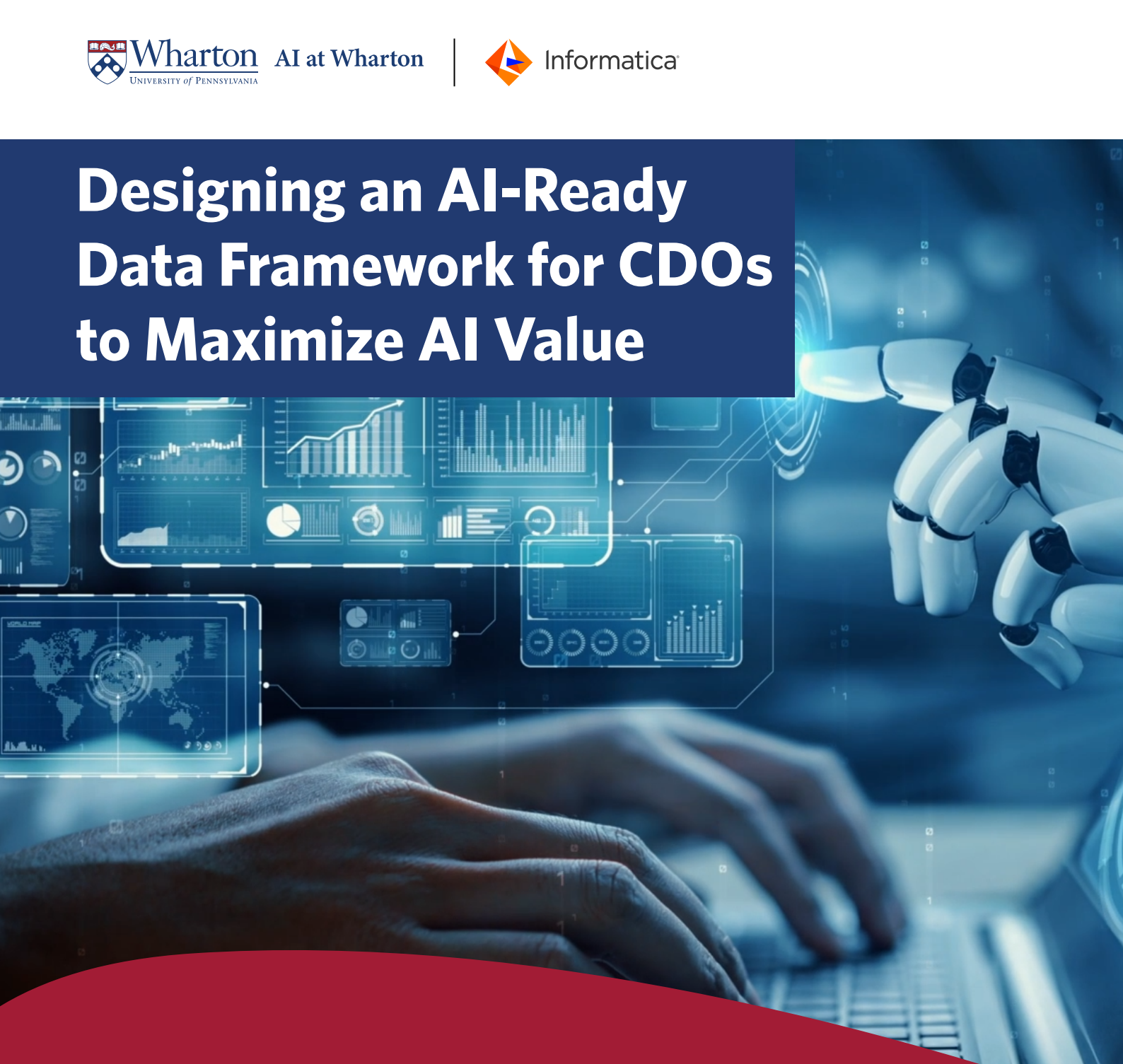# Designing an AI-Ready Data Framework for CDOs to Maximize AI Value

Organizations must keep up with Generative AI's (GenAI) explosive growth, shifting priorities around its adoption and making clairvoyant investments. In the Spring of 2024, Wharton experts and chief data officers gathered to discuss and elaborate on strategies and a framework to maximize value creation at an AI-ready organization. After a brief overview of the emerging trends in GenAI, they identified the major must-haves for an effective framework: Securing executive buy-in through convincing pilot projects that generate economic value swiftly; clearly delineating a path to ROI while steering clear of hype; fostering robust connections between business leaders and data solution developers, linking data directly to business value; investing in change management; recruiting the right talent; and establishing necessary governance structures. The resulting white paper expands on these concepts and other insights that emerged from the discussions.

Contents

How the AI Landscape is Evolving for CDOs

What can AI industry watchers learn from good old coal? A lot, according to Amit Gandhi, a Visiting Scholar with Analytics at Wharton and former Chief Economist at Microsoft Azure.

The invention of the Watt steam engine in the late 18th century greatly increased the efficiency of coal-fired engines, leading to expectations that higher productivity would lower coal consumption. But as the Watt steam engine made coal a cost-effective power source, more and more industries began using it, leading to greater demand for coal.

"More coal was burned and used than ever before in human history," Gandhi remarked, referencing the Jevons paradox — a concept introduced by British economist William Stanley Jevons in his 1865 book, *The Coal Question*. Gandhi was speaking at a data leaders' conference on Wharton's Philadelphia campus, co-hosted by AI at Wharton (AIW) and Informatica of Redwood City, CA, a software provider focused on enterprise cloud data management that can help organizations accelerate and deliver AI-ready data. "(Similarly), when the price of something falls, you use more of it. You will spend more on AI now than you can ever imagine, just from a P times Q (price multiplied by cost) perspective."

The Jevons paradox example brought clarity to the ongoing debate on the growth potential for AI as it delivers large productivity gains while threatening to replace many conventional job roles. For data and analytics

— 🗩 —

"The CDO's role is more of advising and proposing solutions; it's not about just waiting for the order to come in." – Jason Beyer, Chief Data Officer, Fortive

leaders, the growth of AI, especially GenAI, brings new demands; they must reorient their strategies with the right data, tools and governance structures, and be watchful of new risks. They must also be catalysts in organization-wide change management that must concurrently take root to maximize value creation with AI.

The Emerging Contours of AI

In the past decade, businesses have dramatically increased their use of computing power for training AI systems, overtaking academia and research in 2012, when advances in deep learning technology attracted large corporate investments, noted Stefano Puntoni, Co-director of AI at Wharton and Professor of Marketing at Wharton, citing a report from *The Economist*. Another milestone on that AI path was the "Transformer model" (a neural network that learns context and thus meaning by tracking relationships in data), the outcome of a Google paper in 2017, titled "Attention Is All You Need." This paper paved the way for so-called "attention mechanisms" to produce output informed by context and understanding, and thereon to GenAI.

Puntoni noted AI's evolution from its original moorings in computer science, engineering and statistics to increasingly become a social science. It impacts all kinds of processes, economic performance, and societal outcomes, including individual well-being.

"You need that analytic, data-oriented decision with 'What does it feel like?' And then have those tug at each other to arrive at a refined perspective." – Neil Bhandar, VP of Analytics & Insights, Generac

The efficiency leaps that GenAI has delivered promises to revolutionize how organizations perform tasks, marked by the November 2022 launch of OpenAI's ChatGPT chatbot, followed by ChatGPT-4 in March 2023. LLM chatbots bring the power of AI computing to more users, and aligned the technology to have human-like conversations so as to match human values and expectations.

Such leaps have engendered two overarching aspects of AI in the current context, noted Informatica Vice President, WW Field Chief Technology Officer Nick Dobbins at Informatica: (1) the tech shift in natural language processing has democratized the power of AI computing to users; and (2) how this "superpower of AI" can be harnessed reliably, responsibly and in ways that supplement, rather than replace, human intelligence.

In fact, "alignment" was pivotal in driving ChatGPT's popularity, noted Prasanna (Sonny) Tambe, Co-director of AI at Wharton and Professor of Operations, Information and Decisions at Wharton. To achieve that, OpenAI had invested in a number of human-centered feedback processes such as "red teaming," where external experts help identify possibly harmful capabilities such as hateful speech and risks, and find ways to mitigate them.

In addition to OpenAI's ChatGPT versions, the GenAI landscape now has multiple LLMs such as Meta's Llama, Google's Gemini, Claude from Anthropic, and a couple from the French firm Mistral. Over time, users will have a multitude of options as customized features are promoted alongside the LLMs.

> "Change management is successful when you have that executive sponsorship and accountability tied to that sponsorship."
> – Tarun Sood, Chief Data Officer, American Century Investments

CDO strategies for GenAI Challenges

CDOs confront multiple challenges in extracting value from GenAI. CDOs who attended the conference offered insights on how best to address them:

Secure executive buy-in: Establish the business case around the right priorities. To do this, CDOs will need to align "value" and "buy-in from leadership," which will further advance adoption and help in achieving scale. Ideas flowing through the pipeline must be quickly incubated and the unit economics must provide guidance in making the right choices.

Consider the factors that are at play in the dynamic between time-to-value, executive buy-in and pilot-to-scale idea. What drives the executive buy-in? Does it come from media buzz or from proof points that are adjacent to a firm's business? Piloting and prototyping are tried and tested ways to earn executive buy-in. Executive buy-in also allows room to fail. This is especially helpful in scenarios with a long tail, where a pilot runs for more than a couple of months.

Return on investment estimates are critical to establishing the value proposition in a given economic environment. It is important to convince business leaders to sign up for new P&L values and incrementalism in a constrained environment.

> "GenAI's potential is immense, but its realization hinges on the trifecta of robust data quality, stringent security practices, and proactive governance. The role of the CDO is pivotal in ensuring these elements are not only in place but continuously enhanced, facilitating secure and ethical AI implementations that drive real business value."
> – Sowgandhika Dusa, Chief Data and AI officer, Cadent

Find ways to create economic value quickly: Creating economic value is essentially building pilot-to-scale, which will inform how that value must be created. Related aspects are how to measure that value and bring it beyond the pilot stage as quickly as possible. Executive sponsorship hinges on identifying areas where the technologies will add the most meaningful value and achieve this quickly. Essentially, organizations should bulk up on use cases to increase adoption.

Foster a supportive ecosystem: To successfully integrate AI, organizations require a multifaceted approach that combines top-down leadership with bottom-up innovation. Key supportive mechanisms include:

- **AI Governance Committees:** Assemble cross-functional teams comprising business stakeholders, functional experts (e.g., privacy and security officers), and AI specialists. These committees evaluate AI opportunities, assess risks, and inform strategic decisions.

- **Innovation Incubators:** Encourage creativity through initiatives like hackathons, proofs of concept (POCs), and partnerships with external networks. This accelerates AI adoption and fosters a culture of experimentation.

- **Education and Awareness:** Implement training programs to educate stakeholders about the benefits and limitations of GenAI. Emphasize its role as a valuable tool, while also acknowledging its potential limitations in specific contexts and analytical settings. By establishing these

— 99 —

"The right way to think about this class of AI is not so much like trying to create a single, sentient, intelligent model, but lots of intelligent agents that kind of interact in a very decentralized and almost market-based way." – Amit Gandhi, Wharton Scholar

supportive mechanisms, organizations can cultivate a collaborative environment that promotes effective AI integration, mitigates risks, and maximizes benefits.

The need overall is for the right kind of thinking, both from within the organization and outside it. Such thinking can be enriched through AI councils and also by strengthening governance. Data management requires outside-in thinking because AI councils tend to become too focused on internal problems and often ignore the induction of broader ideas.

Avoid the hype cycle: Some organizations that suffer the hype cycle are marked by overenthusiasm and conversations that often start with "I need more AI" instead of starting with a specific use case or a business case that drives decisions on the type of AI that might be needed, why and where. The hype cycle is driving a herd mentality; companies weighing their AI options must be wary of this. "Get beyond the hype cycle by aligning with leaders on targeted uses of AI. By understanding the specific challenge you are trying to solve, you can avoid the assumption that AI is a hammer in your hand that will solve all the problems," said Jason Beyer, Chief Data Officer at Fortive.

Build a tight linkage between data and decision-making: Agile processes that link the builders of AI engines and the decision-makers also help in creating a level of symbiotic communication between them. The goal is to reduce friction between the data and the decision, keeping them tightly linked. A tight linkage also helps capture the speed of data-driven decisions in measuring value or ROI.

The organization must value the importance of a data-informed model output. "Data is not a substitute for thinking," Puntoni noted. "As we have smarter and smarter machines, we have to think harder and harder."

Invest in change management: Change management must focus on helping people be adaptable in order to be in sync with the GenAI opportunity. The way they think must accord with the way they work to achieve organizational alignment to support such change. Change management is directly tied to executive support, where top management designs, orchestrates and measures its progress to ensure that the intended change is taking hold:

- Acquire or nurture the skill sets required to ensure speed to production and scalability.
- Define jobs and create product teams along with organizational restructuring to match skills and team composition.
- Balance the technical skill sets with business acumen.
- Hire people with a background in multiple disciplines to help bring cross-functional visibility into business challenges.

Gather the right talent: In some firms, the absence of the right kind of talent and limitations in the available technology infrastructure constrain new technology, limiting the firm's ability to introduce and make technologies such as AI and analytics more pervasive.

Have clear delineations of the people playing the various roles — the technical expert, the enabler or implementer of the technology, the practitioner who will use the tool, and finally, the one who translates the insights and knowledge from the model. Such specificity will determine the granularity at which people will build and use models and the attendant governance of those models, and will cumulatively drive greater adoption of GenAI or even AI in general.

A federated model that coalesces inputs from multiple business units and functions must have data governance that checks several boxes: (1) Who has access to what type of data? (2) Are stakeholders actually accessing the right data and using it in a way that can drive business value? (3) How

does one ensure that people in an organization understand where to find the data they want, what that data means and when they should tap into it? All these aspects must be continually refreshed because of changes in the business environment; it's not a one-and-done effort.

Get the right data governance: Companies need stewards who understand and protect the data and can determine its appropriate use.

Data stewards must not just understand data, but also the context in which it must be used to drive a certain outcome. In that effort, data owners (or producers), data stewards and data consumers are three legs of a stool which must work together. Data users, who may not yet have a complete understanding of the data and the relevant context, often want the other participants out of their way.

Establish CoEs (Centers of Excellence) and governance bodies. Specify the methods of how the CoE and federated models orchestrate with one another, and how they connect into distributed teams. Specify the functional responsibilities of those teams and guidance for cross-functional fusion between them.

The data governance model must be designed to support the speed of the business. For instance, a firm may have too much data governance where it gets the right solution to a problem, but one that is two weeks too late. It is a challenge to find the right level of data governance that allows for timeliness and relevance to the business from the insights gained from analytics. Centers of Excellence could help design right-sized governance.

Focus on good quality data: Good quality data is a prerequisite, especially if the AI engine has to provide curated discovery processes where users can plot their individual journeys. Data product managers must ensure the availability of mature data, with the right policies being in place to inform

the right decisions. A tangible benefit in dependability and reliability of data is that it avoids the effects of biases, especially recency bias. Data governance is required also for unstructured data such as that in SharePoint or PDFs. Retrain people in an organization around content creation to ideally classify it and tag it at creation, with protocols in place on how to save the final version and where to save it so that it can be incorporated into a model to drive value.

Extend your watch to content. In designing data governance systems, organizations are beginning to delineate between data procured from external sources and user-generated content.

Tie the data to the business value: To achieve the maximum value from analytics, it is vital to understand the data and the desired business outcomes. Not everyone in an organization understands the what-and-why of data, and so investing in that effort is critical. But that approach may run into difficult territory in situations where those designing an AI model want to first have the data before deciding what they want to achieve with it.

Strike a balance between GenAI solutions from software providers and custom-made solutions being developed in-house, such as training or building your own language model for a specific use case, and the time-to-value between them, that is, the amount of time it takes to realize value.

Build fluency: There is a need for organization-wide literacy and fluency in data and analytics. AI fluency must also be a part of change management because it strengthens the knowledge base within an organization, and brings the ability to redefine processes.
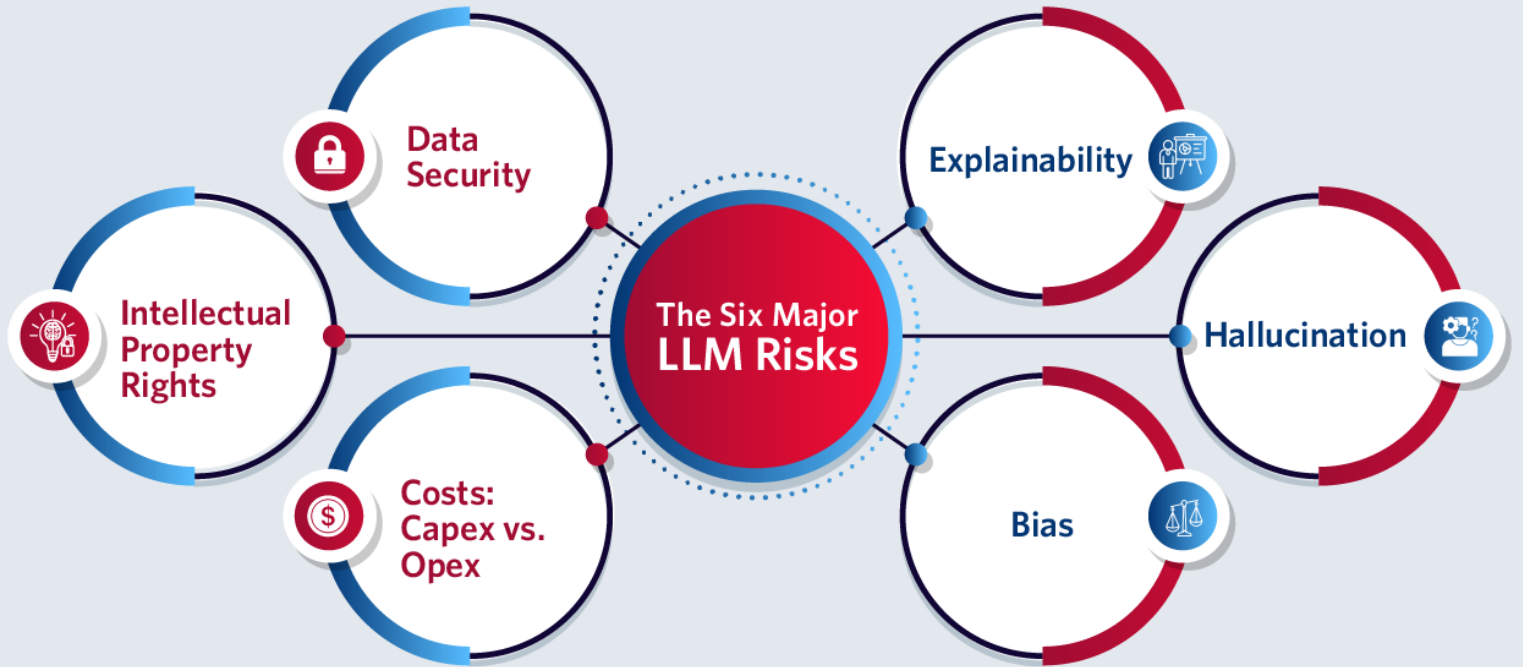
AI use cases might vary widely when companies operate silo-style, depending on which department — marketing, finance or supply chain — is looking for a solution to a problem. Because they are all individual use

cases, AI fluency must be tied to those functional areas as much as it is tied to the model itself.

Maintain objectivity: Analytics is most useful when it is objective, independent and transparent. Compromising on any one of the three will create a situation where people will not trust or depend on these analytics. That is why it is critical to have a tight linkage between data and the decision-making process, as noted above. That linkage must ensure robustness in both the quality of data and the business perspective on the data and the models.

Be alert to regulations: Decisions driven by data have implications from a regulatory perspective, too. This is important because sometimes the use of data — and not necessarily the data itself — can create certain risks, especially in cases involving consents and disclosures. Unreliable data can lead to AI hallucinations; other risks may lurk with chatbots directing users to legally enforceable contracts. Companies are often in scenarios involving secondary and tertiary use of data, which may be outside the scope of the contract within which the data was bought, collected and stored. All of those pieces need to be checked to avoid use scenarios that could put the firm at risk.

Confront challenges with dual-speed infrastructure: Many organizations find themselves juggling legacy IT architecture and new technology. Understanding the tradeoffs between the two is important in identifying the right operating model. It may often be necessary to outsource aspects, which could potentially induce new risks or points of failure. Here, the business leadership must find the right balance between guarding against those risks while appreciating the value of time-to-insight and time-to-market in building solutions to business challenges that are strategic and for the long term. All this calls for an eyes-wide-open approach, which also entails transparency when explaining decisions to the business side.

Diagram: The Six Major LLM Risks — Data Security, Explainability, Intellectual Property Rights, Hallucination, Costs: Capex vs. Opex, Bias

The Six Major LLM Risks

LLMs are still evolving and far from perfect. They come with a number of risks that Tambe puts in six buckets: data security, explainability, intellectual property, costs, bias and hallucination.

Data Security: Many organizational decisions rely heavily on analyzing sensitive data, such as those in the financial services industry and in health care. In addition to the requirement of robust data security measures, these industries also face tough restrictions on data access.

—— 99 ——

"Explainability is one of the biggest issues for LLMs, and it will be a hard one to crack. No silver bullet is in sight." – Prasanna (Sonny) Tambe, the Wharton School

——

Ensuring data security is a challenge also for organizations which use public clouds for managing their data. Some solutions are already available that allow organizations to retain full control of their data, including those being offered by cloud providers such as Amazon Web Services through its Hugging Face partnership.

Integrity of data is yet another area of concern for organizations and individuals. On that front, recent legislative efforts have sought to protect users; these include the Defiance Act in the U.S., which targets deepfakes, and the European Union's AI Act, which focuses on making AI trustworthy for users.

Explainability: Users are more comfortable and trusting of AI when insights are available into how specific GenAI tools work and deliver outcomes. Although academic research shows that users favor accuracy over explainability, explainability does matter, Puntoni said. "Explainability is one of the biggest issues for LLMs, and it will be a hard one to crack," Tambe added. "No silver bullet is in sight."

Intellectual Property: Broadly, fundamental questions about LLMs and IP issues will be around for decades, but in the foreseeable future, much depends on how U.S. courts adjudicate in some major lawsuits. *A Wall Street Journal* report highlighted that among three of the biggest copyright law cases awaiting rulings, two involve "human authorship" and the ability to copyright in works of art using GenAI, while the third contests the ability of an app that allows users to swap faces using deepfake technology. IP rights are critical for GenAI work to be commercialized.

One major question that persists for end-users is whether using content generated by some types of Generative AI violates copyright protections. That is a potential risk if companies face the fear of lawsuits from the use of such content.

Costs: GenAI is capital-intensive, unlike cloud computing, which reduced the need for organizations to make capital investments in IT hardware. Unlike cloud, AI entails significant amounts of fixed costs, such as in graphics processing units (GPUs), specialized data needs and specialized talent teams.

While the costs of training LLMs can run into tens of millions of dollars, and models are becoming increasingly complex, AI hardware has been advancing to make training more cost-efficient. Also, in question are "inference costs," or the costs per query, which can vary widely across models. The cost per 1,000 queries has a wide range, from 10 cents for the Mistral-Small model to $18 for GPT-4. One challenge with that massive price variation will be the ability to understand price/performance trade-offs.

Moving from an operating cost model to a fixed cost model brings constraints on the amortizing investments. "For medium-sized firms, it's a tough landscape to navigate. Instead of building their own solutions, they will have more dependency on companies such as OpenAI," Gandhi said.

Bias: Historical inequities and cultural biases that may exist in the training data of the LLMs could be amplified and lead to discriminatory and harmful outcomes. "We know this problem exists, but it is hard to assess and safeguard against," Tambe said. CDOs could ensure objectivity in their data sources, and thus prevent the risk of those biases causing distortions in the outcomes of the AI algorithms.

Hallucination: AI hallucination occurs when the output from a GenAI model is nonsensical or inaccurate, either because the algorithms are not based on training data or are incorrectly decoded, as an IBM note explained. Solutions to tackle misinformation or hate speech have emerged, such as RAG (retrieval augmented generation) models.

| How AI Governance Helps Value Creation

GenAI is an evolving technology where complete clarity is yet to emerge around many issues such as what data is good enough or which governance models are appropriate.

AI Centers of Excellence could help reduce uncertainty with data governance to try and ensure that only credible projects are put through the hoops of development, testing, deployment and scaling.

Tarun Sood, CDO of American Century Investments, said the company has a similar process under its AI CoE for new projects and the tools that will be used. Generac is adopting a different approach to manage AI uncertainty: using a PARC model (process, activities tied to processes, risks associated with each activity, and controls to ensure auditability and human in the loop) "to think about what could go wrong, [and estimate] the business impact if something were to go wrong," said Neil Bhandar, VP of Analytics & Insights at the company.

Even if it is difficult to measure value creation with AI, one useful metric to consider is how the model or technology enabled, and/or reduced, the time taken for decision-making. It is not a judgment on whether the decision was good or bad because that depends on multiple elements of execution that need to occur. "You need to perform those post-mortems to calculate ROI on a model's output: how quickly is it helping the business arrive at a decision?" said Bhandar.

—— 99 ——

"Let's not just think about how we can do what we do faster. Let's think about what we can do now that we couldn't do before. That's the hardest part." – Stefano Puntoni, the Wharton School

The new opportunities will provide access to new capabilities. Companies such as Microsoft, Google and OpenAI will position their LLMs as full-service stacks that solve problems and are tailored to provide specific business solutions. "The right way to think about this class of AI is not so much like trying to create a single, sentient, intelligent model, but lots of intelligent agents that interact in a very decentralized and almost market-based way," Gandhi said.

How those opportunities unfold depends on how our mindsets adapt to the new order. According to Puntoni, the human-or-AI mindset has a limited view of the potential of AI to bring societal benefits. "The human-and-AI (mindset) is better, for it opens up possibilities," he said. "We are engineering thinking machines that are deployed to optimally complement our skills and abilities rather than trying to emulate them. It seems to be a small distinction, but it could make a massive difference to the way we think about AI deployments."

> "It is important to bring the human context to whatever GenAI is telling you. One of the problems we will have to guard against is people saying, 'It said it's this, therefore it must be true.' Well, it may not be true."
> – Janelle Rolph, Chief Data and Analytics Officer, S&T Bank

Key takeaways for
AI-driven CDOs

Secure executive buy-in: Be a strategic consultant to the business. Build your case with quick prototyping and pilot-to-scale projects for ROI estimates. Avoid the hype cycle.

Find your cost-benefit equation: Build a custom LLM in-house (capex) or license an existing LLM where the cost per query is likely to fall.

Change mindset to decision-driven data analytics. Keep a tight linkage between the data and decision-making process. Balance gut feeling with data-informed decisions.

Establish governance structures including Centers of Excellence, data stewardship, and federated knowledge management. Focus on good quality data.

Embrace change management: Bring on board new skills, and new ways of working and measuring output.

Consider AI's productivity boosts won't necessarily replace humans, but bring new capabilities.

Think human *and* AI, not human *or* AI. AI will complement human skills and judgment, and not replace them.

Create supportive mechanisms such as bottom-up innovation, POCs, partner networks and organization-wide data/AI literacy.

Understand the business context before looking for an AI solution or your best LLM.

Prepare for LLM risks around data security, IP, hallucinations, costs, explainability and bias. Safeguard against analytics getting hijacked by a specific business leader.

Appendix

Data leaders at the conference were optimistic about AI's opportunity while being alert to potential roadblocks to success. The participants had specific expectations on how AI could bring efficiencies to their work and that of their organizations.

Jason Beyer is the Chief Data Officer at Fortive and formerly at Bridgestone, is leading Fortive's efforts to maximize opportunities that enhance data management in a complex risk environment. Shaping the data strategy across Fortive, Jason is focusing on strategic approaches to unlock new value and to accelerate new AI and product capabilities.

Neil Bhandar is the VP of Analytics & Insights for Generac, an energy technology solutions company based out of Wisconsin. He wants to use AI in data and analytics "to accelerate the learning curve and growth" at his firm, building upon his prior experience in financial services and pharmaceuticals.

Nick Dobbins is the Vice President, WW Field Chief Technology Officer at Informatica, captured the top questions in the minds of CDOs: "How do we make AI worthwhile? How do we make it a value enhancer? How do we measure it in a responsible way?"

Sowgandhika Dusa is the Senior Vice President, Chief of Data at Cadent TV, a New York City-based ad tech firm. Dusa, who is based in Philadelphia, PA, manages her firm's data services, data engineering and data science, using GenAI for customer service solutions only for the moment. She wants to use it to improve efficiencies across platforms, data modeling, data science and engineering processes.

Janelle Rolph is the Chief Data and Analytics Officer at S&T Bank, a 122-year-old midsized bank based in Indiana, PA. She wants to tap AI to explore opportunities for growth for her bank, which operates in a heavily regulated industry.

Tarun Sood is the Chief Data Officer at American Century Investments, an investment management firm headquartered in Kansas City, MO. As ChatGPT and natural language processing revolutionize asset management, portfolio construction, and portfolio management research, Tarun aims to collaborate with peers to harness AI solutions. He envisions immense potential in this domain, emphasizing that practitioners have only begun to explore its possibilities.

Contributors

Data Leaders:

Jason Beyer
Neil Bhandar
Nick Dobbins
Sowgandhika Dusa
Khoi Hoang
Priya Raman
Janelle Rolph
Tarun Sood

Wharton Experts:

Stefano Puntoni, Co-director, AI at Wharton, and Wharton Professor of Marketing

Prasanna (Sonny) Tambe, Co-director, AI at Wharton, and Wharton Professor of Operations, Information and Decisions

Mary Purk, Executive Director, AI at Wharton

Amit Gandhi, Visiting Scholar, Analytics at Wharton

Editorial: Shankar Parameshwaran, Global Media Space, Inc.
Design: Quantent