Qlik

Trends

Bridging the Trust Gap in Generative AI

The Big to Better Data Imperative

2024

We are squarely in the middle of an AI boom, with generative AI promising to take us into a new era of productivity and prosperity.

But there are rising concerns that could limit that promise or even worse, reverse it altogether. Until now, data access for generative AI has been a free-for-all, with no traceability of data origin or quality control. This has allowed faulty data to pollute the reliability of generative AI's output with hallucinations, misdirection, or even outright untruths. If left as is, false data — supercharged by generative AI — will pose an exponential danger to business and society.

We need to find a new model that promotes better* — and trusted — data. This means data that has been proven to be valid and valuable. Trusted data, in combination with analytics and automation, will be the foundation for helping humans and organizations make efficient and better decisions while providing the fuel for reliable and responsible AI.

Eventually, trusted data will accelerate the use of data as capital, putting it in the same standing as human capital and financial capital. Imagine a product fundamental to AI that can be traded and becomes more valuable the more it's used.

Exciting, yes — but how do we get to this promised future? Right now, there's a lot of noise around that very question. We've sifted through the buzz to identify 10 trends in AI, data, analytics, and automation that will help organizations like yours get onto the path of ensuring that all data is trusted data with value in the AI economy.

Dan Sommer, Senior Director, Market Intelligence Lead

* "Better data" definition: Data that expands the typical characteristics of Big Data to be AI-ready, through handling **Volume, Velocity, Variety, Validity,** and **Value.**

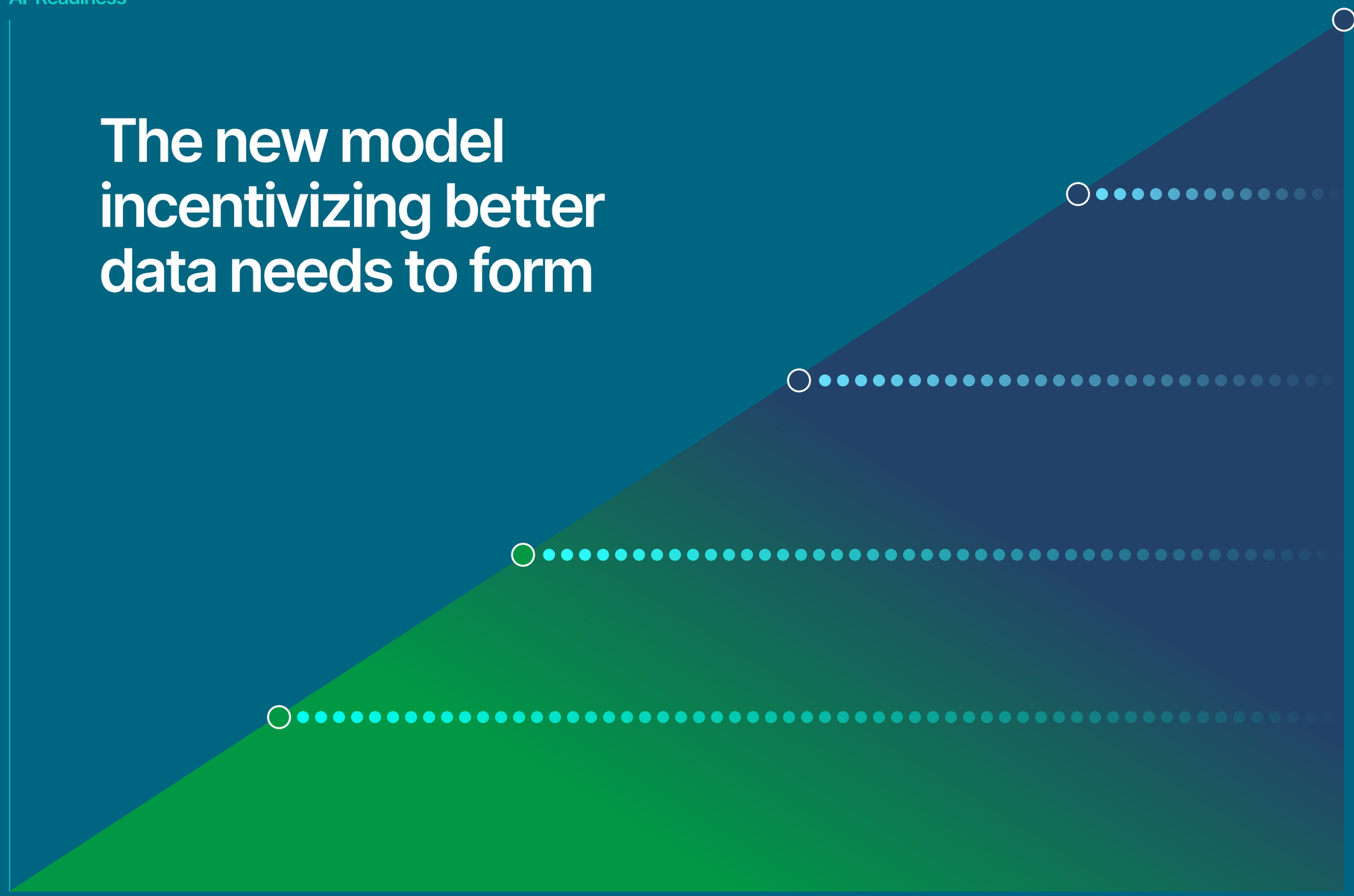
From big to better data: It all comes down to the 5 Vs

Moving from big to better data can be described as moving up a value chain of 5 Vs: **Volume, Velocity, Variety, Validity,** and **Value.**

'Big Data' has often been attributed with the three V's of **Volume**, **Velocity** and **Variety**. This is because big data is significant in quantity and often accumulated rapidly from a variety of sources (the latter being something that many companies are still trying to achieve).

'Trusted' data or 'better data' goes further. It encapsulates the other two – and arguably most important V's — **Validity** and **Value**. Trusted data has been gathered from various sources, checked for credibility, and features the right lineage and traceability. This data can then be refined, packaged, and governed, i.e., treated as a known, quality product that can be shared and traded.

The new model incentivizing better data needs to form



"Big Data"

"Better Data"

Value

- Data as a product that can be traded
- Automation and AI create a virtuous cycle
- Last-mile AI customization becomes critical for business

Validity

- Data origin matters: Understanding your data's DNA
- The rise of novice developers demands AI literacy
- Data engineering, analytics and data science are merging

Variety

- From BI to AI and back again, business analysis is changing
- The age of unstructured data is now

Velocity

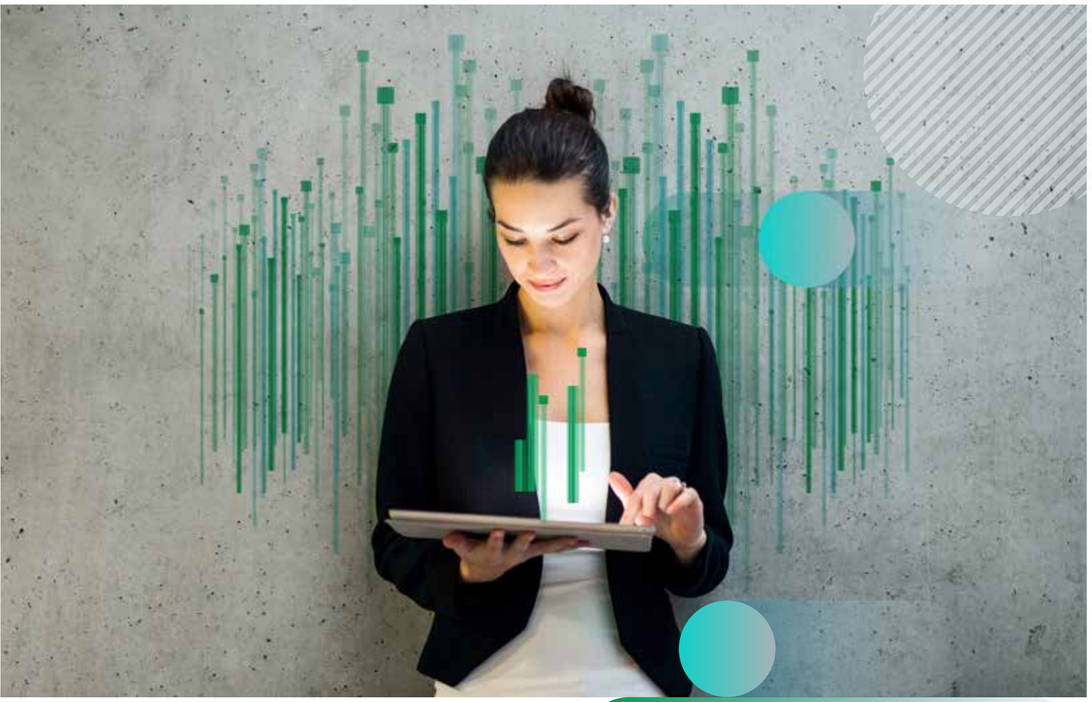
- Hybrid AI bridges the maturity gap
- Generative AI for insights: Supercharging the data consumer experience

Volume

With Volume being further along than the other Vs, there are no pressing trends of note.

The Promise of Generative AI

Gartner® places generative AI at the peak of the Hype Cycle™. It has exploded onto the market in less than a year and is already transforming the world.



McKinsey

$2.6 to $4.4 trillion[1]

Generative AI could add the equivalent of $2.6 trillion to $4.4 trillion USD annually to the global economy based on 63 use cases in which the technology can address specific business challenges across 16 business functions.

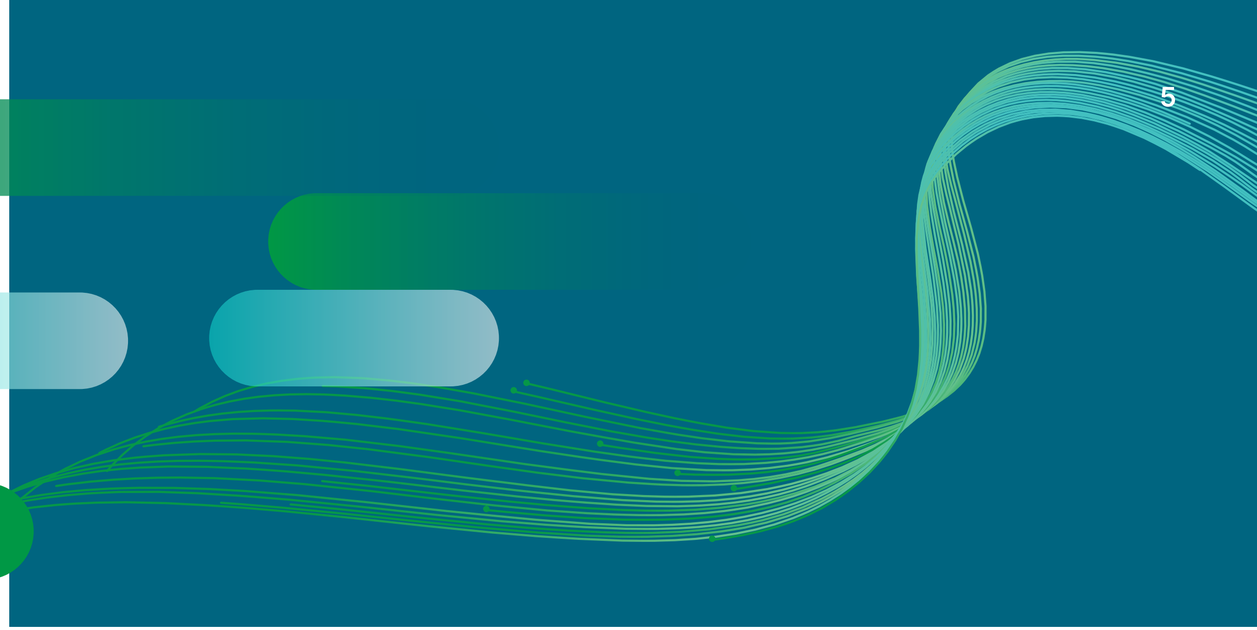
Will the promise ever become a reality?

But in its current state, can it really reach the heights it is expected to? Until now, it has been a “free for all”. Large Language Models (LLMs) have been able to crawl enormous amounts of information, without guard rails, to train its models. This has given rise to a backlash, with many making the argument that generative AI is only going to get worse, not better.

Why? Because AI is only as good as the data that feeds it. In the era of exponentially increasing big data, we’ve solved for **Volume** and **Velocity** and are working on **Variety**, but not yet solved for **Validity** and **Value**. The notion of “garbage in, garbage out” will evolve from an adage to a stark warning with critical consequences as AI takes center stage. We all need to pay careful attention, because the path to trust is littered with the following challenges.

[1] McKinsey: The economic potential of generative AI: The next productivity frontier, Jun 14, 2023

The Trusted Data Dilemma



Too MUCH data

Generative AI is currently trained on the entirety of human expression — the useful and the junk — making it noisy and biased. Good data can get polluted with bad data, sometimes intentionally so. This can lead to hallucinations and factual errors. By 2025, **experts predict that up to 90% of online content could be generated by AI.**[2] This may be great for productivity, but it is alarming for accuracy. If the ratio of synthetic (false) data to real, trustworthy data becomes too skewed, we will get "reality drift" — where information becomes unhinged from reality.

Too LITTLE data

At the same time, organizations are withholding data, reducing the vastness of knowledge available to LLMs. Social media giants and media conglomerates, including **Reuters and CNN**, have started a data rebellion, curtailing the data on their systems to restrict access. This move reduces the amount of quality information freely available for generative AI — directly affecting output for the worse — and could mark the beginning of a "Data for AI" economy.

The regulatory landscape

Because trust is key to any technological adoption, regulators are looking to curtail the spread of misinformation and enforce privacy controls. Some countries, such as Italy, have taken an extreme approach and **banned generative AI tools such as ChatGPT** (Italy has since reversed course). The push for better governance, origin, and lineage has also led to legislation. The EU AI Act also deems LLMs to be "high risk" and subject to multiple restrictions. All of these efforts could slow or even wind down existing models — hindering the effectiveness of generative AI.

Sustainable compute

Generative AI's requirement for compute power is high, leading to escalating costs and undesirable environmental impact. In fact, it's estimated that **AI data centers could use more electricity than the whole of the Netherlands by 2027.**[3] If left unaddressed, both the global economy and the drive towards sustainability will be negatively affected.

[2] Nina Schick interview with Yahoo Finance Live, January 7, 2023

[3] Alex De Vries, a PhD candidate at the VU Amsterdam School of Business and Economics, first published in Joule, October 10, 2023

Qlik

Human dependency

Human judgement and manual labelling are still needed for generative AI. But as innovators try to balance oversight and efficiency, the lack of suitable skills in the workforce could slow down progress.

Massive corporate failures

We haven't seen any widely publicized failures from generative AI that have landed corporations in hot water, but they are likely to emerge as usage increases. With all eyes on those who are early adopters of this technology, when they do make a misstep, it could carry grave consequences.

The trust crisis

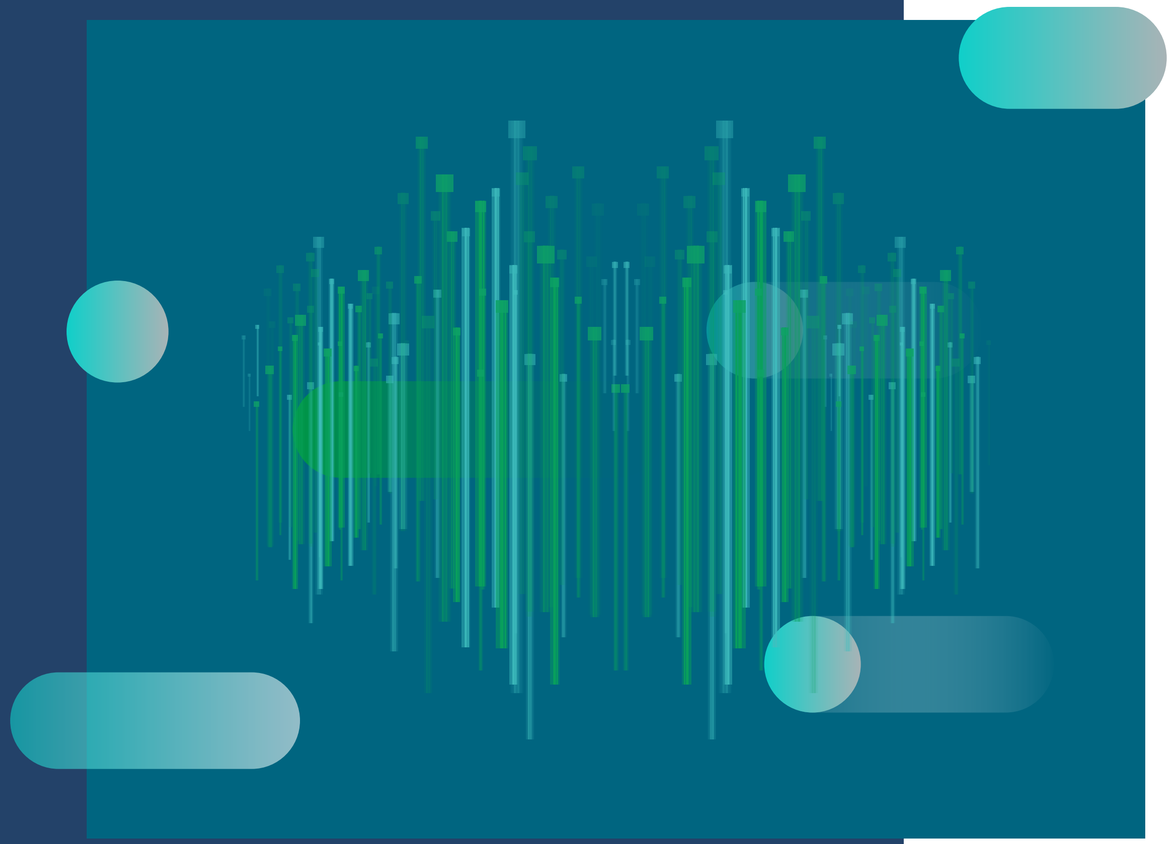
Corporations still have a lot of mistrust in generative AI. Qlik data shows that **only 39% of organizations currently**[4] have a formalized AI strategy in place. We're seeing this reflected in corporate policy. Major tech companies are hesitant for their employees to use ChatGPT, for instance, as these tools still have some way to go before they have been adjusted enough to suit corporate needs. But crucially, corporate data also needs to be adjusted to AI. This is done through addressing the Vs of Validity and ultimately, Value.

The imperative for a new model now

Unleashing generative AI without good, governed, and trusted data is dangerous. A new mandate needs to form — inside and outside organizations — that demands more accountability for the quality, lineage, and transparency of data. It needs to happen sooner rather than later — but it shouldn't just be driven by negative consequences. It should also use incentives, where creating quality data products is rewarded, to democratize use.

[4] Qlik Generative AI Benchmark report, September 2023

Qlik



Listen to the Music

In previous decades, music became a free-for-all, accessible on platforms like Napster and Pirate Bay. But this model was unsustainable. The rise of streaming platforms changed the economic model, enabling governed redistribution that is aligned to rights management. This approach, in turn, also benefited the songs — and the writers and creators of the songs — that were listened to the most.

That's the kind of evolution that needs to happen in the data economy if AI is to be properly fed. Creators of quality data products must have more exchanges on which to trade them. When that happens, providing valid, trusted information with traceable origins will have an underlying financial incentive. An important byproduct is that the more it is used and valued, the better that data will become.

Data management, analytics, and automation will all be key enablers that bring efficiencies and **Value** to AI. **But what does the roadmap to this model look like in action? Here are 10 AI trends that are going to impact organizations in 2024.**

1

Hybrid AI bridges the maturity gap

Will traditional AI finally mature, or will GenAI overtake it?

All the talk today is about generative AI, which, as described, holds vast potential. But there have been several other AI efforts underway for years, some of which have started bearing fruit. This could be Machine Learning (ML), which is becoming more democratized, or general augmentations throughout the pipeline. One of the biggest misconceptions is that generative AI will replace all of this. That would be a big mistake. BARC announced last year, that "playtime is over" with "traditional" AI efforts, meaning that AI has matured enough to be put into production and scaled. This is done especially in well-established use cases such as fraud analytics and churn analysis. So, while generative AI is finding its footing, Machine Learning and other AI have already shown boundless potential — and may even be used to bridge this maturity gap.

Analyst Prediction

~30%

Generative AI is expected to achieve ~30% share of the overall AI market by 2025.[5]

[5] Boston Consulting Group

2

Generative AI for insights: Supercharging the data consumer experience

How is AI empowering less-technical information workers?

Not everyone wants to build apps. In fact, most of us belong to "**the other 75%**"[6] who may not even realize when we've been touched by analytical tools. All these "day-to-day information workers" really want is an answer — ideally now — and they don't have the time, desire, or skills to perform an analysis. Consumers also tend to trust people more than data, so collaboration and data sharing is key. This user base appreciates auto-generated visualizations and insights, enhanced with explanations in natural language. There are bonus points if this can happen within the systems where they operate.

Analyst Prediction

66%

By 2025, 66% of G2000 will adopt AI-driven headless BI and analytics with chat, Q&A, and proactive notification functionality, quadrupling the number of users with access to contextual information.[7]

[6] BARC (Business Application Research Center) and Eckerson Group survey "Strategies for Driving Adoption and Usage with BI and Analytics", March 2022

[7] IDC FutureScape: Worldwide Data and Analytics 2024 Predictions, IDC #US51295223, Oct 2023

Qlik

3

The age of unstructured data is now

Is GenAI the key to making unstructured data easier to analyze?

The majority (Forrester says 80%) of the world's data is unstructured.[8] In other words, it is not neatly sorted in rows and columns. An example of this is emails and documents on your intranet.

Many have previously tried and failed to analyze unstructured data, but with new metadata and semantic techniques, we can unlock it. By using knowledge graphs and vector databases, complemented with RAG (Retrieval, Augmentation, Generation), the opportunities for combining structured and unstructured data in a trusted way are endless. Combined with an answer management layer, you can re-use verified and trusted questions and answers, allowing you to scan your entire data estate and use private LLMs created internally through data analysis.

Analyst Prediction

2x

Unstructured data managed by enterprises will double in 2024[8]

[8] Forrester, Predictions 2024: Data And Analytics

4

From BI to AI and back again, business analysis is changing

How is generative AI supporting business analysis?

The generative AI revolution is moving along at warp speed, enabling new ways of interacting with data — including multi-modal analytics. Now you can just drag a file onto a simple chat interface and start talking to it. It can generate queries and code, help build content, and expedite automated processes. Increasingly, individuals might start their analytical journey in these generative AI tools, using them for simple data visualization and business projections. This is BI coming to AI.

As a next step, they may want to tap into enterprise-grade tooling for further analysis, bringing the benefits of generative AI to their trusted tools. That's AI coming to BI. In other words, we'll toggle between these two different modes — enabled by embeddability, connectivity, and APIs — to get maximum benefits from each platform.

Analyst Prediction

70%

According to Gartner®, by 2026, generative AI will significantly alter 70% of the design and development efforts for new web applications and mobile apps.[9]

[9] Gartner, Top Strategic Technology Trends for 2024, 16 October 2023. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

Qlik

5

Data origin matters: Understanding your data's DNA

If you don't know where data comes from, how can you trust it?

As we've established, if data quality and lineage were important before, they've become non-negotiable in a world with AI. This is critical for the data that runs your business as well as for training AI models. The need for identifiable, understood data origins is particularly acute with public LLMs, where the origin is not currently traceable. Without this knowledge, it's hard for the best generative AI models to differentiate fact from fiction. This can lead to symptoms such as hallucinations, fake facts, and deepfakes. For businesses, trusting outputs like this can have serious consequences. That's why organizations must make this a priority now.

We need a mechanism to clearly label and signpost data, using techniques of provenance and cryptography alongside techniques we haven't invented yet to create the equivalent of a "DNA test for your data."

There are already several efforts under way, such as **The Coalition for Content Provenance and Authenticity** which has members including Intel, BBC, and Sony; **Google Watermarking** (SynthID), which identifies AI- generated images, and **Hugging Face ModelCard** which creates simple Markdown files with additional metadata.

When there is trust in data origin and traceability, it puts in motion a self-perpetuating cycle where people take responsibility for data. It's also one of the most important pieces of the puzzle to turn proprietary corporate data into products that can be traded and watermarked.

Analyst Prediction

90%

"90% of online content could be generated by AI by 2025"[10]

[10] Nina Schick interview with Yahoo Finance Live, January 7, 2023

Qlik

6

The rise of novice developers demands AI literacy

How can you safely put great power in the hands of citizen developers?

In a short space of time, we've seen an evolution from low code to plain English (or any other language) as the new dominating programming language. When coding is simplified, it facilitates more advanced tasks like creating apps. This will spawn an explosion of apps built by the 'everyday developer' — resulting in a flurry of innovation. But it can also lead to governance chaos and application glut. As this process puts very strong powers into the hands of the many, organizations must take steps to educate their workforce in the benefits and pitfalls of generative AI. If the last five years have all been about teaching your teams data literacy, now we need to pivot to AI literacy. At the same time, application lifecycle management and promoting the right data and apps will take on a new importance.

Analyst Prediction

60%

To mitigate new risks created by the pervasive use of generative AI by their workforces, by the end of 2025, 60% of large enterprises will mandate formal data literacy and responsible AI training.[11]

[11] IDC FutureScape: Worldwide Future of Enterprise Intelligence 2024 Predictions, IDC #US51293423, Oct 2023

7

Data engineering, analytics, and data science are merging

Will data be democratized by removing the need for advanced knowledge and tools?

According to IDC[12], enterprises prefer to work across the data pipeline with best-in-class capabilities, with fewer or even just one vendor. New platforms, combined with the evolution of data fabrics, will "consumerize" data engineering to a new breed of users — especially if augmented with powerful AI, automation, and data science. This will empower business analysts, for instance, who can now go back earlier in the pipeline to do data management and preparation tasks. That same analyst can also apply advanced statistical models to the data and tools they work with every day without needing to export it to an advanced workbench.

Making the hard stuff easier and merging the roles and capabilities of data engineering, data science, and analysis will enable organizations to solve tougher problems. We'll move from asking how much profit was made this quarter to asking, "What customers should we target going forward?" and "Which high-value employees are at risk of leaving and what are the factors driving that decision?" Adding more glue between previously siloed functions will help businesses move data and outputs from big to better.

Analyst Prediction

50%

Gartner® says that by 2026, 50% of organizations will have to evaluate ABI and DSML platforms as an all-in-one, composable platform due to market convergence.[13]

[12] IDC Data Management Survey, 2023

[13] Gartner, Predicts 2023: Analytics, BI and Data Science Composability and Consolidation.

Qlik

8

Automation and AI create a virtuous cycle

Why is the focus of AI changing from analysis to execution?

So far, LLMs and generative AI have mainly been used to support reasoning and conduct analysis, rather than for iPaaS and actioning. But now, there are several exciting efforts underway to support the latter, including an approach to LLMs that involves synergizing reasoning and action. Of course, this requires transformed data in near-real time and in the right place. Across organizations, we will start to see new ways of using generative AI with application automation, such as using sentiment analysis to automate and generate different responses depending on mood. Generative AI, connected with automation, will mean less manual work for humans to connect and build workflows and instead take on the role of decision stewards.

However, alongside the opportunities, we need to realize that connecting automation with AI will further increase AI's power and agency. We need to prepare for this and put guardrails around it.

Analyst Prediction

20%

According to Gartner®, by 2027, outlier detection and other augmented analytics capabilities will evolve into autonomous analytics platforms that fully manage and execute 20% of business processes.[14]

[14] Gartner, Predicts 2023: Analytics, BI and Data Science Composability and Consolidation.
GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

9

Last-mile AI customization becomes critical for business

How do we develop generative AI tools specifically for business needs?

Right now, early applications of generative AI are massively scalable — but currently generic — projects that can use LLMs. These are typically done in a business-to-consumer (B2C) context. Over time, we'll increasingly see AI customized to industry, and more specific business-to-business (B2B) use cases. This will take the shape of private LLMs and applications where the foundation may be common, but with layers of customization that better serve the "long tail". An example of this is an **AI cluster that Mark Zuckerberg is building for medical research**.

Extrapolating this trend, we can also see that with less effort and fewer consultancy hours, sophisticated applications can be built that address a specific industry or issue at hand.

Your organizational proprietary data will be a valuable raw material here and "solution fabrics" will emerge where domain-specific data and apps can be shared and traded. However, the question of which AI will be the foundation for building this remains unanswered. While there are currently 2-4 dominant LLMs forming, **others think this will increasingly be built on open source**.

Analyst Prediction

80%

By 2026, more than 80% of generative AI use cases in enterprises will leverage customized, specialized AI models rather than generic foundation models offered via public APIs.[15]

[15] IDC FutureScape: Worldwide Future of Enterprise Intelligence 2024 Predictions, IDC #US51293423, Oct 2023

Qlik

10

Data as a product that can be traded

How has AI become the tipping point for monetizing your organization's data?

Architectural approaches to harmonizing distributed and varied data, such as data fabrics and data meshes, have moved from hype to reality over the last year due to AI and technological breakthroughs. A key component of these approaches — one that resonates with customers — is "data as a product." This is about applying principles of product management to data, asking questions about what problem(s) we are solving, what it is going to be used for, and by whom. It emphasizes the importance of data quality, governance, and usability for end users. Data as a product is evolving to become the foundation for consumability for all forms of analytics and AI.

The concept of treating data as a valuable asset or product means it can be surfaced in a catalogue, used for various purposes internally, and even evolve into a tradable good. The aim is to monetize data as a product outside of your organization.

We are starting to see more platforms where validated data can be refined, bought, sold, and traded — with those who own it remunerated. OpenAI's recent launch of "GPTs" is a significant milestone and definitive tipping point, as it has an app-store approach to contextualized AI apps, with a revenue-sharing model.

A further evolution of this will be to enrich it with additional data. This should encourage organizations to use their own data to further train ChatGPT models, which can then be monetized. In the future, similar exchanges will serve as vetted sources upon which LLMs can crawl sanctioned data and distribute compensation for the access, similar to how the music industry did with streaming services. The more the data product is used, the more valuable it is.

Analyst Prediction

60%

By 2026, 60% of leading enterprise intelligence companies will have identified data products, and 15% will have attributed business value to the products with a data valuation methodology.[16]

[16] IDC FutureScape: Worldwide Future of Enterprise Intelligence 2024 Predictions, IDC #US51293423, Oct 2023

Qlik

Unlocking the

4th & 5th Vs

Validity and Value are vital to the future of generative AI

Qlik

Conclusion

If data quality was important before, it's exponentially more important in a world with generative AI. This is where the 5 Vs of data come into play. While we've solved for Volume and Velocity, we're still working on Variety. And if we truly want to have our data ready to fulfill the promise of GenAI, we must move from big data to better (trusted) data — which will require solving for Validity and Value as well. Your company's data and metadata are a unique asset. Taking the other two Vs — **Validity**, and **Value** — into account will ensure that you can use and act on it effectively to enable AI.

Our journey to the promised future of generative AI relies solely on one thing: the quality of the data used for this technology. If data is consistently and thoroughly checked for origin and quality, it can be turned into a product. Then, the more your data is used for AI, the more valuable it will be — both internally and externally. We will see an evolution of better data becoming the raw material feeding trusted LLMs and becoming a tradable good. Data capital will gain prominence and underpin all innovation using generative AI.

Of course, you need to figure out how to turn your data into valuable, better data. But what's valuable in the era of generative AI might surprise you. It's the classic gold rush analogy — those who built the shovels, jeans, boots, transportation, and railroads benefited massively. Similarly, being a trusted platform and enabler of mining better data in the AI economy, enhanced with analytics and automation, will be very valuable.

Finally, it's now or never. This is the last moment — the calm before the storm — before AI gets embedded into all aspects of knowledge work. This is not the time to be complacent, or you'll get left behind. Generative AI will change the world to the same extent as the internet did. There are challenges, but by taking the right steps and channeling the outputs correctly, overcoming the obstacles will usher in an era of unprecedented innovation and prosperity.

If this feels overwhelming, remember that you don't have to go it alone — in fact, you shouldn't. Work with knowledgeable partners who can turn your big data to better, trusted data so you, too, can realize the value of generative AI.