

WAVESTONE

2025 DATA & AI RADAR

10 challenges
to master your Data
& AI transformation
in 2025

Summary

CHAPTER 1

The rise of AI: time to scale up and industrialize governance **4**

1. Industrializing AI 5
2. GenAI at scale: moving beyond PoC to realize the opportunities offered by generative AI 7
3. AI governance: increased complexity to address all the underlying issues 9
4. Data/AI acculturation on a large scale, to accelerate innovation and prepare for the future 12

CHAPTER 2

Unlock the full potential of Data **13**

5. A federated organization to unify Data-related roles, standards, and practices 14
6. Democratize the use of Data for as many people as possible 17
7. Data storytelling, or the art of making Data speak for itself 20

CHAPTER 3

Data governance and quality remain key concerns **21**

8. Successfully combining Data Quality and Data Observability 22
9. Governance of unstructured Data: a growing problem for organizations 24
10. Systematize the measurement of the value generated by Data 25

... and what's next? Preparing for the future by putting people at the heart of transformation **27**

The rise of AI: time to scale up and industrialize governance





1. Industrializing AI

Despite the rise of artificial intelligence in the business world, around 85% of AI projects still fail to reach production. This figure decreases slightly from year to year, but remains relatively high, and only the most mature organizations manage to bring it down drastically.

To take AI projects all the way to production and integrate them at the heart of business processes, companies are rethinking their operating model. The stakes are twofold:

- Define an **organization**, roles and responsibilities that provide all the skills needed to autonomously deliver a product throughout its lifecycle;
- Define **practices** to ensure proper implementation of AI, from design to maintenance in operational conditions.

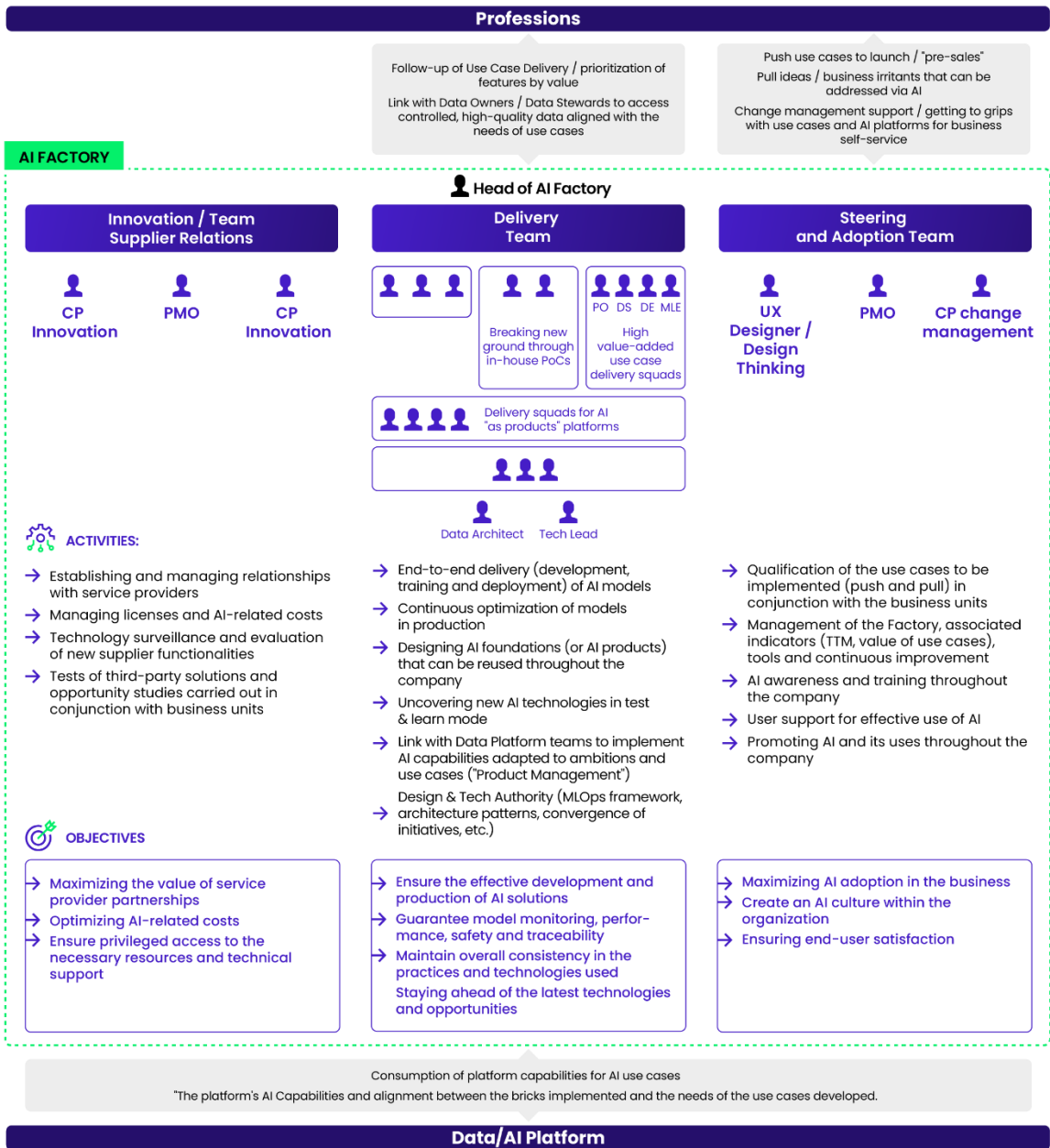
To date, the first challenge has mostly been addressed by setting up **AI Factory-type** organizations. These are specific to each organization, but we can outline a state-of-the-art AI Factory in 3 main functions:

- The first aims to **address** the Data/AI ecosystems, build relationships with publishers and derive maximum value from the opportunities offered by the market;
- The second is to **deliver** use cases, both for experimentation and industrialization;
- The third aims to **steer** the Factory's activities, both upstream (managing business demand, driving proposals to address new use cases) and downstream (facilitating adoption, driving change and promoting AI and its potential within the organization).

The second challenge lies in **delivery practices**. An ML model is not enough to deliver value; it needs to be designed for deployment in production. All too often, model creation is decoupled from deployment, and designed for experimentation. There is no design authority, no development framework, no automated tasks, no versioning. Moving towards AI industrialization means deploying partially automated model and infrastructure management practices that are designed for production use. **MLOps** aims to address these issues.



Illustration générique d'une AI Factory à l'état de l'Art en 2025



MLOps: the key to managing AI models in production

Following on from DevOps, MLOps is a set of practices aimed at unifying development activities (Dev) and operations (Ops), to continuously deliver and manage AI models, from development to monitoring, via acceptance and deployment. While certain teams are now increasingly proficient in initial model production, the next challenges to be addressed will be advanced, automatic performance monitoring, drift detection and orchestration of continuous retraining to maintain model relevance.

The challenge for Data Scientists is no longer to confine themselves to model development, but to acquire ML Engineering skills to ensure the implementation of these practices. And for organizations, the challenge is to define a coherent, shared framework across the organization, based of course on best practices shared by the market, and to drive change to ensure its application.





2. GenAI at scale: moving beyond PoC to realize the opportunities offered by generative AI

2024 saw the rise of GenAI. Our study showed in May 2024 that 74% of organizations had already begun work on implementing generative AI. PoCs have been launched since 2023 and have led to convincing results. Nevertheless, we note that many companies are still stuck in the PoC phase and have not yet been able to move on to the industrialization stage. Some projects are still PoCs more than a year after their launch!

“Generative AI has made decision-makers PoC-phobic.”

To avoid this situation, we recommend:

- Define simple rules for all PoCs launched within an organization. For example, PoCs must be limited in time and budget, undergo an initial arbitration at the end of the first milestone, and can potentially be extended once if the initial results are convincing. If the expected results take a long time to be demonstrated, it's likely that the project will have to be terminated, and the focus shifted to other use cases;
- Manage a single portfolio of GenAI PoCs (even if delivery is decentralized), and collectively arbitrate which PoCs deserve to go from delivery to scale (demonstrated value, optimal return on investment, etc.). These must be limited in number, and a collective effort must be made to deliver these projects as effectively as possible.

A technological strategy to be carefully thought out

It's tempting to want to move fast to take advantage of GenAI and gain a competitive edge quickly.

Some companies have entrusted data scientists with the keys to the house, in order to promote innovation and time-to-value. However, this strategy can be fraught with risks: the fact that use cases are delivered locally, conceived as a single unit, with relatively little flexibility and scalability. All this at a time when the entire GenAI ecosystem is evolving very rapidly, and innovations abound (new models, new

capabilities, etc.). How, for example, can we ensure that we will be able to make the most of a new model available on the market that offers improved performance, without having to rebuild everything?

Worse still, some companies, out of a sense of having fallen behind their competitors, quickly forged partnerships with certain tech giants to reassure decision-makers. These choices have sometimes led to **vendor locking** situations, in which an organization finds itself limited in its ability to innovate, explore alternatives and optimize its orientations for specific use cases. In addition to flexibility, the financial equation can also be degraded in the long run.

Companies that have thought through their medium- to long-term strategy have thus taken their desire for robustness and technology diversification into their own hands. For example, many are building **model-agnostic GenAI platforms** and **architectures**. Such a platform makes it possible to host, train and monitor a variety of LLMs, enabling the LLM best suited to each situation to be used to derive optimum value. Flexibility is guaranteed, as is adaptability and the ability to accommodate a freshly published model.

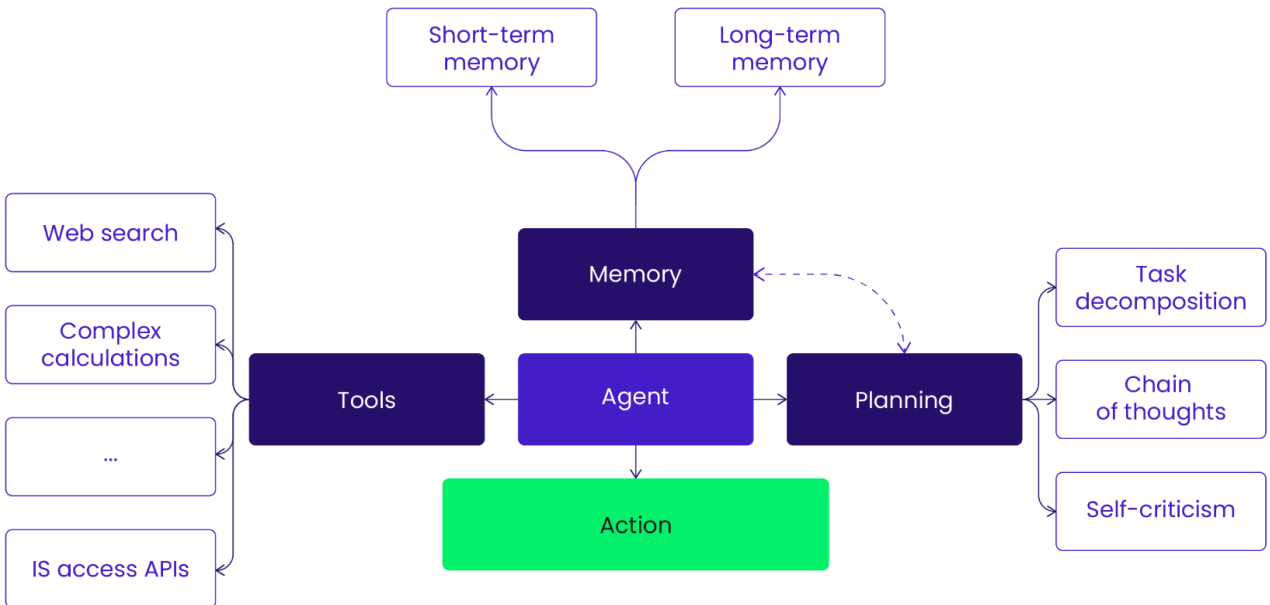


Materialization in the form of Agents

An AI Agent is a system designed to reason with complex problems, create action plans, and execute these plans using a series of tools. Unlike traditional computer applications, these agents have advanced reasoning, memorization, and task execution capabilities. These agents can, for example, solve complex problems (e.g., generate project plans, write code...), perform self-criticism by analyzing their own outputs, use existing tools and IS, and even perform inter-agent collaboration.

These Agents are made up of several elements:

1. **A Core Agent:** the central element integrating all processing functionalities.
2. **A memory module:** stores and retrieves information to maintain context and continuity over time.
3. **A set of tools:** external resources and APIs that the agent can use to perform specific tasks.
4. **A planning module:** analyzes problems and develops strategies to solve them.





3. AI governance: increased complexity to address all the underlying issues

The rise of AI is also prompting decision-makers to strengthen governance, by grasping all the issues at stake. These are many: risk management, compliance, sovereignty, ethics, carbon footprint, are just some of the issues to be addressed through this governance. A number of issues need to be tackled at the same time, in order to establish a sustainable approach to implementing artificial intelligence.

An organization adapted to address AI-related challenges holistically

Organizations have begun to set up an organization and roles to govern AI-related topics. All of them are facing a challenge: managing to strike the right balance between having a global, centrally-controlled view, while not bridling local initiatives and innovative business line initiatives. The key concern is therefore to "manage decentralization" of the IA initiative portfolio. To achieve this, we need to:

- Define clear roles and responsibilities between central and local levels;
- For cross-functional functions (CDO, DPO, AI Factory, etc.), define each person's responsibilities and specify boundaries and interactions.
- Implement multi-level governance:
 - A strategic body, involving members of the executive committee, to apprehend all the subjects inherent in AI, particularly the impacts on HR, partners, customers...;
 - An operational body, to manage the portfolio, highlight local initiatives and encourage the dissemination of successful initiatives throughout the organization. This body enables us to maintain an exhaustive view of the portfolio, so as to better manage risks and ensure compliance with the AI Act (see below).

A growing need for trusted AI

Trusted AI refers to artificial intelligence designed and deployed in such a way as to guarantee high levels of transparency, safety, fairness, and respect for human rights and ethical values. This implies that AI is developed in compliance with rigorous standards to avoid uncontrolled bias, protect user privacy, and ensure robustness in the face of errors or cyber-attacks. Trusted AI is also explainable, enabling users to understand how and why it makes certain decisions. Finally, it includes responsible governance, where designers and operators assume responsibility for its impacts, while integrating control and supervision mechanisms to prevent abuse or malicious use.

Setting up trusted AIs requires the involvement of a variety of players:

- Ethics and CSR teams, to define an AI policy that embodies clear principles and is aligned with the organization's values;
- Engineers developing the models, in order to provide explainable and documented models
- HR teams, to help set up training modules so that teams apply these instructions;
- Businesses and Data scientists, to analyze biases and monitor results;
- CISO and DPO, to guarantee security, test vulnerabilities and carry out controls;
- Chief Data & AI Officers, to steer the overall approach.



The AI Act: Getting your organization started on compliance

The AI Act was enacted and came into force in August 2024. The Act aims to ensure that artificial intelligence systems and models marketed within the European Union are used ethically, safely and in a way that respects EU fundamental rights.

The AI Act thus creates regulations applicable to artificial intelligence systems and models being commercialized and marketed. Research activities with no commercial objective are not affected. All suppliers, distributors or deployers of AI systems and models, legal entities (companies, foundations, associations, research laboratories, etc.), headquartered in the European Union, or when headquartered outside the European Union, who market their AI system or model in the European Union.

The level of regulation and associated obligations depend on the level of risk presented by the AI system or model. There are 4 levels of risk, and 4 levels of compliance:

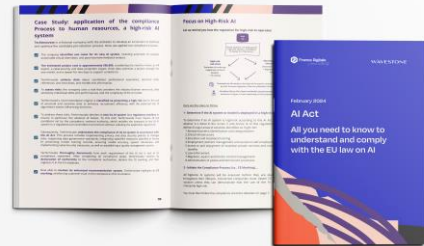
- **AI with unacceptable risk:** AI systems and models with unacceptable risk are prohibited and may not be marketed in the European Union or used for export;
- **High-risk AI:** high-risk AI systems and models must be CE marked to be marketed;
- **Low-risk AI:** low-risk AI systems and models must be subject to information and transparency obligations vis-à-vis users;
- **Minimal-risk AI:** minimal-risk AI systems and models can comply with conduct measures.

Special obligations apply to generative AI and to the development of general-purpose AI models* (e.g. LLMs), with different regulations depending on whether the basic model is accessible or not, and on other subsidiary criteria (computing power, number of users, etc.).

AI will be implemented gradually over the coming years:

- **February 2, 2025:** Article 5 takes effect, banning AI systems with unacceptable risks;
- **August 2, 2025:** regulations on general-purpose AI models will begin. The EU AI Office, already established, will oversee governance and regulatory procedures. Sanctions for non-compliance will also begin to apply;
- **August 2, 2026:** general application of the AI Act begins, with the exception of article 6 paragraph 1 on high-risk systems;
- **August 2, 2027:** applied rules will be extended to high-risk systems. The European Commission will publish practical guidelines and examples of high-risk AI systems by February 1, 2026.

How to get started on compliance? Start with an assessment of the AI systems in place and in the process of being set up.



As a reminder, Wavestone has published, with France Digitale and Gide, [a practical guide to enable companies to understand and apply European AI law.](#)

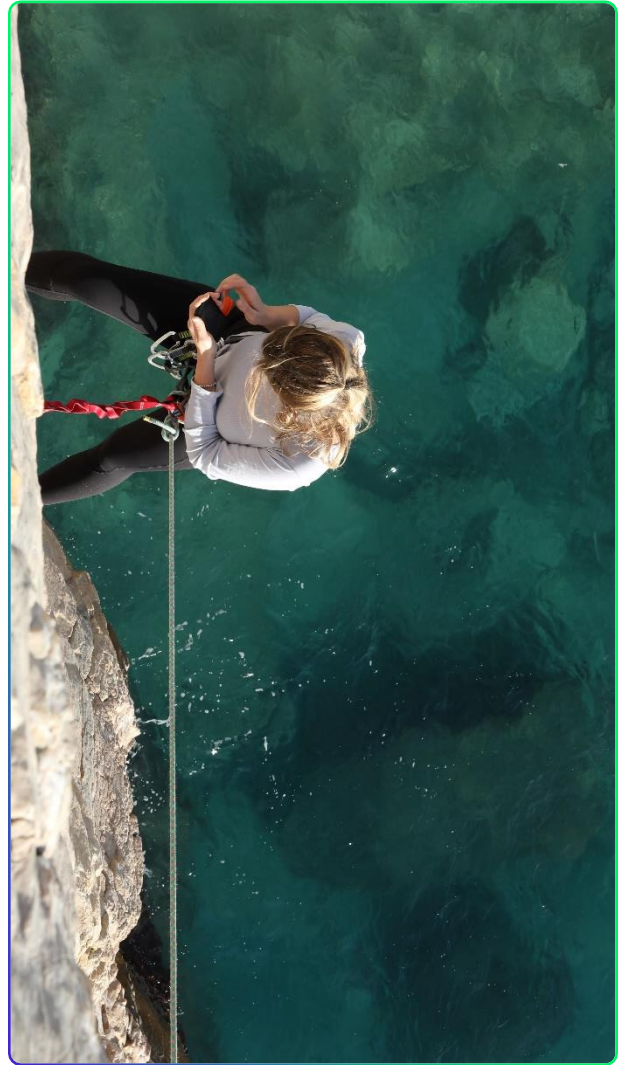


Risks and cybersecurity, all too often forgotten in projects

With all the buzz around artificial intelligence (AI), organizations are facing unprecedented threats that go to the very heart of these models. New attacks are taking shape, such as poisoning (modifying training data to trick it), oracle (hijacking AIs to make them reveal things they shouldn't), or illusion (making AIs believe things that are false but invisible to humans). New risk assessment and protection measures need to be put in place.

In the short term, the priority is therefore to secure business projects using AI, particularly in the following stages:

- **Classification of AI use cases** according to regulatory criteria (refer to the future European AI Act) or the NIST (National Institute of Standards and Technology) AI risk management framework;
- **Definition of the responsibility matrix and governance** for validating use cases, taking into account cybersecurity, transparency, privacy, bias and ethics;
- **Implement specific measures when necessary**, either by integrating them directly into the project design, or by implementing new AI security products that are starting to appear.





4. Data/AI acculturation on a large scale, to accelerate innovation and prepare for the future

One of the main obstacles to the adoption of innovations remains human resistance. AI is no exception, and its adoption remains a major challenge, not least because of the fears and misunderstandings it arouses. It is therefore becoming crucial for companies to acculturate their teams to AI, by demystifying this technology and explaining its applications in concrete terms.

Demystifying AI: Reassuring and informing about what it can (and cannot) do

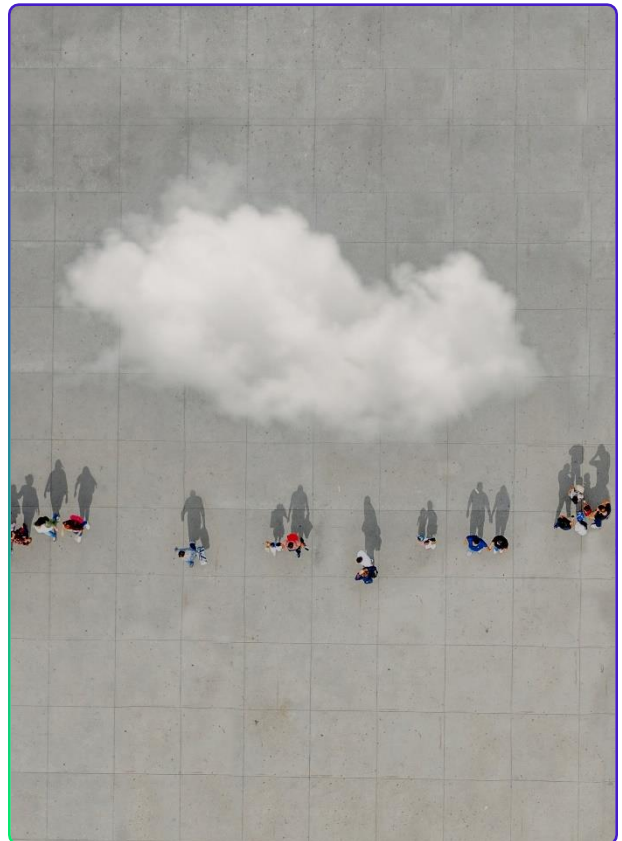
The first challenge of acculturation to AI is to dispel the misunderstandings and fears that surround it. For many, AI is still perceived as a mysterious, even threatening technology, capable of replacing humans or making uncontrollable decisions. It is therefore essential to demystify AI by explaining what it can do, but also its limitations. For example, AI is extremely powerful when it comes to processing large quantities of data and automating repetitive tasks, but it lacks human awareness and intuition. Reassuring teams on these points helps prepare them to collaborate with these technologies, rather than fearing them.

Daring to explore technical concepts: the challenge of making applications concrete

Beyond this demystification, it's important to dare to go further in pedagogy by explaining to business teams what the technologies underlying AI are in concrete terms. AI managers: dare to explain to your business teams and executives what OCR, LLM and clustering are. By repeating these explanations and taking a step back, you'll be able to make the application of AI in business processes tangible. Employees will then be able to understand how these technologies can be integrated into their day-to-day work, identify the processes that could be transformed and the pockets of value to be exploited for the company.

"No Data, no AI!"

Finally, acculturation must insist that magical AI doesn't exist, and that it requires training models on quality Data. And to involve business units in Data management.



Unlock the full potential of Data





5. A federated organization to unify Data-related roles, standards, and practices

Everyone now agrees that Data is a key asset for the success of organizations. Yet its potential often remains under-exploited, due to organizational silos, knowledge, data control, accessibility, and interoperability. Heterogeneous practices also make collaboration between different teams difficult.

To meet these challenges, Data governance remains the watchword. An operational model covering organization, roles, and responsibilities, as well as operating modes, remains a must-have for all organizations. And beyond theory, thinking about implementation through concrete practices that can be understood by all players is a strong differentiator, where many organizations still confine themselves to description without tangible application of the elements defined.

This operational model covers several key themes:

1. A new organization, federated around Data domains, in which businesses regain control of their Data

For a long time, IT and Data management have been the mainstays of Data management. They are still doing so in the least Data-mature organizations. This poses several problems:

- Business units lack autonomy in controlling their Data, and become relatively dependent on IT or Data departments for access to their Data, delaying the time-to-market of analyses and projects;
- On the other hand, IT Departments are sometimes held responsible for the quality of Data over which they have no control in order to develop the processes used to generate and process it. Attempts to remedy the situation often involve reprocessing the stock, without addressing the flow problems at the source of data creation.

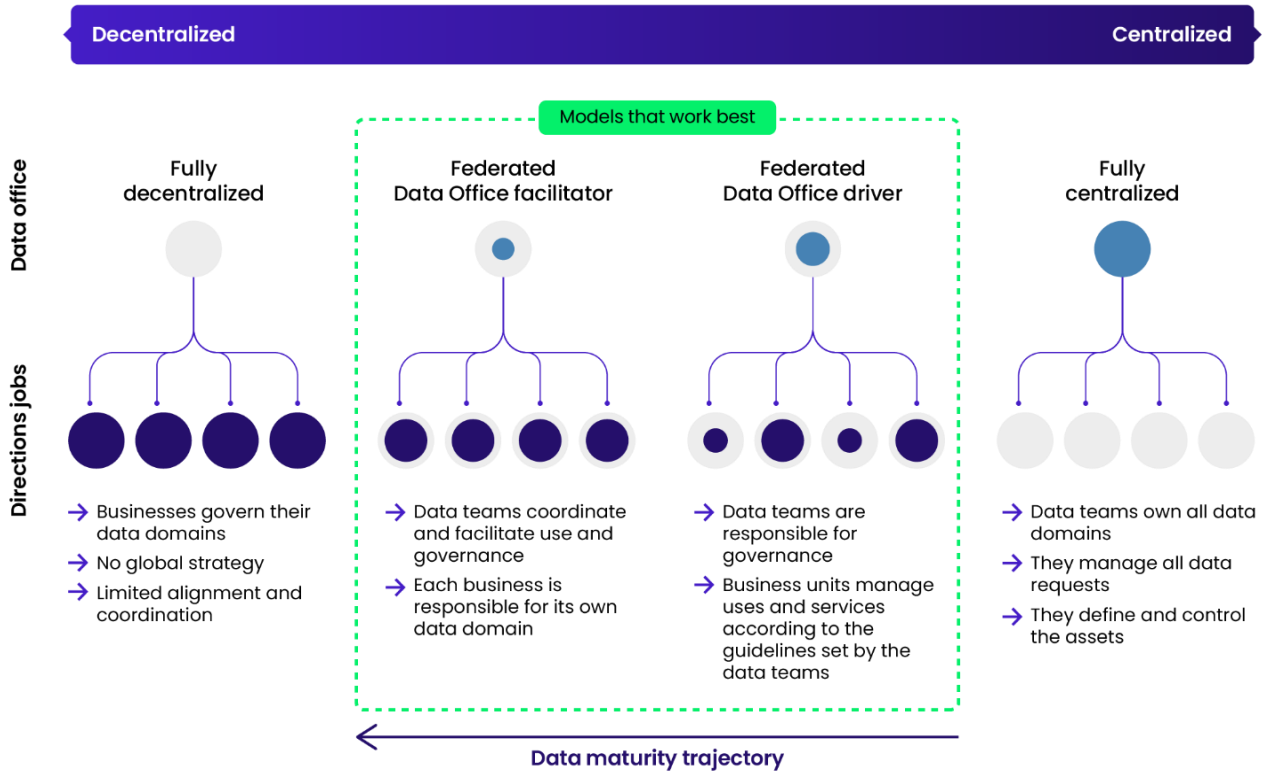
The most mature organizations have taken steps to give the keys to their Data assets back to the business. This means involving business functions in Data management and making them accountable for these activities. For example:

- Businesses (re)become *accountable* for Data management within their scope (Data mapping, documentation, quality monitoring and implementation of corrective actions, etc.);
- The Data Office plays the role of orchestra conductor, defining the framework, tools, policies, practices, and standards. It is also responsible for supporting the business units in their development, through training and coaching.



The challenge is to give the business units back the keys to their Data assets.





2. A unified role repository to facilitate know-how deduplication, recruitment and career management

It's not uncommon to find organizations that can't tell you how many "Data employees" they have. Data's unique position at the crossroads of business and technology is a major contributing factor. The consequences are manifold:

- Difficulty in managing the workforce and strategic workforce planning;
- Difficulty in managing skills and maintaining them at the state of the art for these profiles, against a backdrop of rapidly evolving technologies and associated know-how;
- Recruitment and internal mobility can be laborious, with candidates unable to understand the activities behind these profiles and career paths unclear;
- Data-related operating methods and responsibilities that need to be finely defined on a perimeter-by-perimeter basis, and whose understanding by the rest of the organization remains opaque.

So, beyond a coherent organization, Data/AI roles are tending to be streamlined towards a common, shared repository across the organization. HR teams are fully integrated into this approach.

3. Common standards and practices throughout the organization

For this federated organization to function effectively, it is essential to define and unify Data management standards and practices. This includes common rules for data governance, cataloguing procedures, documentation standards, as well as security and compliance policies. In addition, an organization-wide Data catalog is defined, which describes business concepts via a *Data glossary*, documents Data in the Data dictionary, and records the transformations undergone by key Data (*Data lineage*).

Unified standards ensure Data consistency and quality throughout the organization. They also facilitate system integration and interoperability, making Data more accessible and usable by all. For example, a standardized Data model enables teams from different areas to collaborate more easily, share insights, and multiply the value created.

What's more, these standards facilitate Data sharing, not only within the organization, but also via other subsidiaries, and even with third parties and external partners.



Focus on FiDA, the new European regulation that will make a major contribution to the growing maturity of Data sharing in financial services:



The Financial Data Access (FiDA) regulation proposed by the European Commission in June 2023 aims to create a legal framework for the access and use of consumer financial Data. It is part of the broader Open Finance strategy, extending the rules already introduced by the Payment Services Directive (PSD2), which only concerned payment accounts.

Essentially, FiDA will make it possible to:

- Greater Data transparency, with clear and transparent communication on how customer Data is used and shared between financial institutions;
- Fine-grained customer consent, giving customers the ability to grant, manage and withdraw consent for data sharing;
- Enhanced security, by implementing

strict security measures for the protection and processing of financial Data;

- Standardization of user Data and technical interfaces to accelerate Data-sharing capabilities.

FiDA compliance will therefore require the implementation of rules for accessing and sharing very specific customer Data, and will contribute to a sometimes forced rise in the maturity of financial services players in their Data sharing capabilities.



6. Democratize the use of Data for as many people as possible

The framework is thus set, with an organization, roles/responsibilities and operating modes. All that remains is to prepare the Data and make it available to the greatest possible number of people.

a. Data Products to build up a wealth of Data ready to be exploited

With the explosion of Data and the acceleration in its use, a new paradigm has emerged over the years to facilitate its access and use. The concept of the Data product has thus emerged, and is now being adopted everywhere.

A Data product can be generically defined as "a product that facilitates an end-goal through the use of Data". The term thus encompasses various items:

- Technical foundation" type products, i.e. technological capabilities that represent the foundations for managing and leveraging Data, and whose construction and evolution are carried out in product mode (e.g., a RAG platform, the Data catalog...);
- Data products, where the Data itself is offered as a ready-to-consume product (e.g., customer repository, sales per perimeter, etc.);
- Analytics products (e.g., BI dashboard, recommendation engine, scoring model, etc.).

In particular, and by misuse of language, the Data product is assimilated to the concept of "Data as product". It is in this sense that the Data product is defined through the *Data Mesh* as defined by Zhamak Dehghani (see below).

This Data product has several basic features:

- **Discoverable:** Data is "sorted" by business domain (then by sub-domains, families, business objects, etc.) and stored in a *marketplace*. Potential consumers are thus able to see what Data is available, read its description, and request access to it if necessary;
- **Self-describing:** products are documented and users can independently understand what they contain (via product definition, metadata description, Data freshness, scope, etc.);
- **Trustworthy:** the quality of this Data is monitored, and metrics are shared so that consumers can judge whether it is in their interest to consume this Data. In addition, responsibilities are specified: an owner is identified, as are the teams in charge of Data stewardship activities;
- **Easily accessible:** mechanisms are in place to easily manage access, and consumption is facilitated (via APIs, for example).

The aim of these Data products is to build up a wealth of Data, the value of which no longer needs to be demonstrated, and to factualize this value creation through multiplied use cases.



Data Mesh: A new approach to Data-driven business

The Data Mesh concept is emerging as a solution to the challenges posed by centralized data architectures.

Traditionally, organizations have tended to aggregate all data in centralized data lakes, but this approach has shown its limitations, particularly in terms of scalability, governance and agility.

Data Mesh is an approach designed to **decentralize Data, while accelerating its sharing and consumption by empowering producers and consumers.**

The Data Mesh proposes a federated organization where each domain or team takes charge of its own data as a product, while respecting common standards to guarantee interoperability and quality. This decentralized approach enables teams to work autonomously, while ensuring overall cohesion thanks to a federated governance framework.

Data Mesh is thus based on 4 pillars, the concepts of which have been detailed above:

- Organizing Data assets into Data domains;
- Data management as a product;
- A self-service platform for accessing and managing Data;
- Federated governance to govern it all.

What are the aims of the professions?

- Reduce time-to-Data by eliminating bottlenecks (everyone can share their Data easily) while making Data more reliable (e.g. by avoiding Data recopies);
- Move from a "Data on demand" logic to a "Data as a product" logic, where Data is designed to be easily consumable.

These are the kind of results that will enable us to move towards a true democratization of Data.

b. Self-servicing and the data marketplace

Beyond the creation of data products, it is essential to make them easily accessible to maximize their value. For this, a *data marketplace* plays a central role. It enables these data products to be centralized, organized and distributed intuitively and securely. By providing a platform where users can search, explore and acquire relevant datasets, the data marketplace promotes transparency, efficiency and optimal use of data. It thus becomes a strategic lever for democratizing access to data and accelerating innovation within companies.

The market is very dense in terms of available data marketplaces. These tools are rich in functionality, with the ability to address complex issues. The downside of these tools is that they can also be seen as complicated to grasp by the Business Units, who are the stakeholders who need to be convinced of the benefits of this marketplace.

In particular, we recommend implementing such a marketplace in product mode, involving a panel of end-users in the process of choosing a solution, designing and implementing the product roadmap. This approach is the cornerstone of the change management process, and will strongly encourage the support of the business units.



c. Citizen Data Science, or Data science made accessible to business, with fewer technical prerequisites

The profile of Citizen Data Scientist is emerging within organizations. Without being a position or even a formal role, it lies at the intersection between business expertise and Data science. Without being a specialist in advanced Data science techniques, this professional has a solid understanding of business issues and accessible analytical tools. Thanks to Data products made accessible via the marketplace, and via platforms that democratize the use of Data science via

no/low code (Alteryx, Dataiku, Knime, etc.), the Citizen Data Scientist can manipulate, analyze, and interpret Data to generate insights and make informed decisions. They thus play a key role in democratizing the use of Data, facilitating the independence of business units in their operational Data analysis activities, and ultimately contributing to the adoption of Data-driven practices within organizations.





7. Data storytelling, or the art of making Data speak for itself



Many companies have recently launched ambitious Data training plans, notably by deploying Business Intelligence (BI) tools among their teams. While this training has enabled employees to familiarize themselves with the "what" - i.e. the technical use of the tools and access to Data - it has often left out the "how": how to translate this Data into usable, impactful information. Today, the aim is to go one step further by training teams in Data storytelling. It's no longer just a question of manipulating Data, but of learning how to tell a relevant and convincing story based on insights. This skill is becoming crucial to reinforce teams' ability to extract the full strategic value from data, by transforming figures into informed decision-making levers.

How to train your teams in Data storytelling:

- **Understanding the basics of storytelling:** Before diving into the Data, it's important to train participants in the fundamentals of storytelling. This includes narrative structure (the key message to be emphasized, the analysis and methodology that led to that conclusion, and the elements of analysis), the construction of powerful messages, and how to captivate an audience. These skills will help make sense of the Data;

- **Focusing training on the target audience:** Data storytelling focuses not only on core Data populations (Data analysts, Data scientists), but also on key functions that are frequently called upon to add value to Data (strategy, marketing, operations...). Training teams to adapt their discourse to the stakeholders they are addressing (management, customers, etc.) also helps maximize the message's impact;
- **Going beyond the numbers and choosing the right visualization:** Teams need to be supported in interpreting Data, extracting relevant insights, and translating them into concrete recommendations. They also need to be supported beyond that, particularly in choosing the right graphics or the right way to present the Data. This point is often overlooked and tends to water down the message we want to put across.

This notion of Data storytelling must therefore become an integral part of the training courses already in place for the target populations.

Data governance and quality remain key concerns



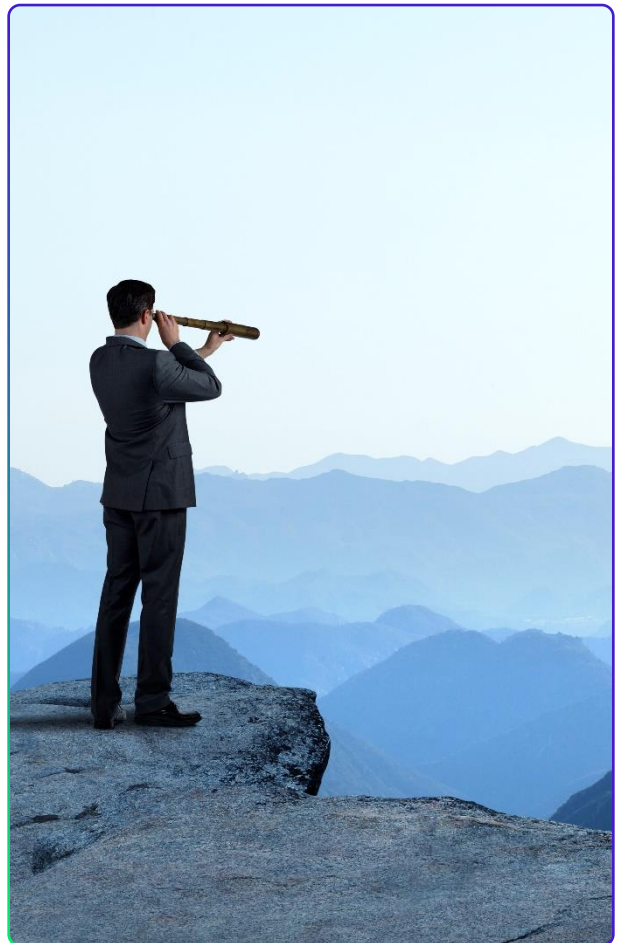


8. Successfully combining Data quality and Data observability

If there's one theme that's invariably at the top of the list of Data issues to be addressed within an organization, it's Data quality. To hear project teams, Data scientists and Data analysts tell it, all their problems stem from the unreliability of the Data they are supposed to be handling. These Data quality issues remain a major concern, with only 37% of our customers claiming in 2024 that their efforts to improve Data quality had been successful.

Data observability is on the rise for a number of reasons. The proliferation of Data sources, combined with increasingly complex IT infrastructures, makes continuous monitoring of Data flows indispensable. With so many interconnections between different Data bricks, it is difficult to maintain manual, non-automated supervision, as this would risk leaving room for reactive action in the face of potential incidents affecting Data pipelines. Data observability takes a proactive approach to quality monitoring, quickly and accurately detecting anomalies in Data flows. It aims to provide complete visibility over the health of Data and systems, enabling teams to be the first to know of any problems, their origin, and how to correct them.

In contrast to a static approach to Data quality, this discipline focuses on real-time monitoring of flow performance and the state of Data throughout its lifecycle, from source systems to downstream analysis. Using these tools, teams can anticipate Data problems, an essential element in ensuring the reliability of information. These tools enable automated monitoring, problem alerting, Data comparison, root cause analysis, and SLA tracking, giving teams a better understanding of end-to-end data quality. For Data-driven organizations, that leverage their Data to generate insights, build machine learning models, or drive innovation, a Data observability solution is crucial to ensure that Data remains a valuable asset rather than a risk.





How can we distinguish between Data observability and Data quality?

Data observability supports Data quality, but the two concepts address different aspects of Data management. An organization can have very good Data quality without an observability tool, and conversely, implementing this tool does not guarantee excellence in quality management either.

DATA QUALITY	CRITERIA	DATA OBSERVABILITY
Measuring data accuracy, completeness and reliability	Definition	Proactive monitoring and analysis of systems to detect performance and reliability problems
Manual or automated quality control according to predefined validation rules and standards	Approach	Automatic pipeline monitoring with metrics and alerting to detect anomalies
Precision, completeness, consistency, uniqueness, etc.	Main metrics	Freshness, integrity, latency, pattern, etc.
Focus on error resolution, correction and process evolution in the workflow	Incident management	Focus on rapid identification to prevent the propagation of errors in systems
Professions, data managers, data stewards	Main users	Data engineering, DevOps, IT teams
Can be slow, as it focuses on audits and controls	Reaction time	In near real time, for rapid response

Tools for Data observability

- A real-time monitoring tool is essential for rapid, detailed analysis of the performance of Data ingestion flows. Today's observability solutions are numerous and offer increasingly powerful and comprehensive functionalities;
- However, it is essential to complement this monitoring with a system of automatic, personalized alerts. These alerts will enable immediate detection of incidents, and rapid intervention by operational teams. They need to be parameterized according to different levels of criticality to guarantee an appropriate response to each type of problem;
- Setting up an observability system requires precise definition of the steps to be followed, as well as roles and responsibilities at each stage of the process. This includes the teams in charge of monitoring indicators, those who will receive alerts in the event of a problem, the actions to be taken, and the people to mobilize depending on the seriousness of the situation. It is essential to co-construct this process with all stakeholders to ensure effective implementation and buy-in from all teams.

Pure players are entering the market to address this issue (Sifflet, Montecarlo, IBM Databand, Anomalo, etc.). The traditional players in Data Quality Management (Informatica, Talent, etc.) are also gradually addressing this issue.

It should also be remembered that, although this tool will help to address pain points, it alone cannot remedy all Data quality issues. And it's up to Chief Data Officers to remind all their organizations of the famous adage: **"Data quality is everyone's business"**.



9. Governance of unstructured Data: a growing problem for organizations

The rise of AI, and Generative AI in particular, is putting the spotlight back on the governance of unstructured Data, which has always been the poor relation within organizations. Indeed, Data lakes and other EDMs have accumulated increasingly varied Data over time, and RAG use cases require the ability to manipulate reliable, usable documents.

On the whole, organizations are still in their infancy in this area. In particular, 2 topics are currently being addressed by Data Management teams:

1. Operationalize unstructured Data practices

Managing unstructured data requires specific practices. Unlike structured Data, which takes the form of organized tables, unstructured Data (documents, images, videos, etc.) requires particular attention to metadata, which provides a minimum of standardization for its content.

Good metadata management not only makes it possible to classify and retrieve information more efficiently, but also ensures that documents are up to date and that obsolete information does not interfere with decision-making. For example, it is good practice to systematically date documents, both in the metadata and in the document content, so as to deal with conflicting information and consider the most recent as the authoritative information.

2. Using AI to strengthen Data governance

We know that Data governance is key to obtaining high-performance AI algorithms, but on the contrary, AI also helps improve Data governance. They automate Data categorization, cleansing and anomaly detection, reducing manual intervention and the risk of human error. Examples include:

- The use of OCR capabilities, placed at the front end of operational systems such as EDMs, to recognize, categorize and transform documents into structured Data, which is easier to govern;
- The use of NLP models (or even LLMs) to extract structured information from unstructured documents.



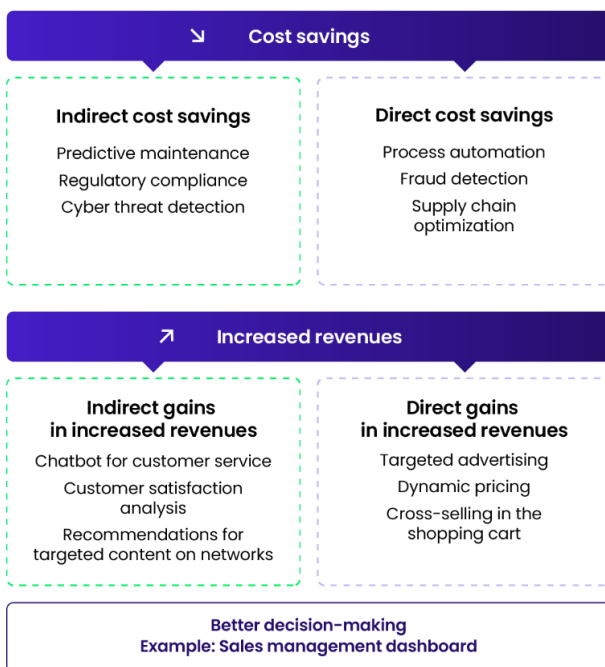
10. Systematize the measurement of the value generated by Data

Finally, once the projects have been launched, it's now a matter of factualizing the value they generate. Indeed, very substantial investments in Data and AI have been decided this year by executive committees. This subject is mechanically highly visible to ExCom members, who want to "get their money's worth" and measure returns on investment (ROI). The subject is nonetheless ambivalent, since Data is still often perceived as a cost center by decision-makers.

It is therefore necessary to distinguish between Data and AI use cases, and the "enablers" that enable these use cases to be implemented.

On uses:

Use cases make it easier to factualize – at least for some of them – the return on investment of these initiatives. Financial gain can be achieved in two ways: through increased revenues, or through cost avoidance.



We recommend:

- **On direct gains:** to be rigorous in drawing up business cases and monitoring gains, by involving the Finance and Controlling teams in this process so as to be able to transcribe these elements into the P&L;
- **On indirect gains:** to define KPIs, financial or otherwise (e.g. Customer NPS) to factualize the added value of a use case, without necessarily having to transcribe this added value in accounting terms, as it cannot be translated into a P&L. And to assume that some of them (e.g. BI dashboards) cannot be transcribed into ROI, despite their key importance for business activities.

Note: Even for direct gains, some are trivial to transpose into € (e.g. cost avoidance due to fraud detection, or FTE gains due to automation). However, others are more complex to measure (e.g. targeted marketing use cases), and require the implementation of AB testing mechanisms, the features of which must be thought through right from the design stage.



On enablers:

Finally, the value of Data and AI "enablers" is ever more complicated to factualize. Indeed, it's difficult to directly measure the contribution of actions to improve quality or modernize organizational and technological foundations.

For example, the implementation of good Data governance improves information processing, sharing, and management, enabling the identification of critical organizational processes, improved decision-making, and overall optimization of operational efficiency. The same goes for the technological foundation: a Data platform on the scale required constitutes the cross-functional asset needed to deliver use cases handling increasingly varied, voluminous and complex Data, while ensuring greater speed of execution, deployment or maintenance.

The paradox is that these enablers now account for the majority of long-term Data/AI investments, but decision-makers expect their intrinsic value to be factualized by short-term ROI.



We recommend that CIOs, CDOs and CTOs:

- **Factualize** the value of enablers by proposing other, more qualitative management indicators (adoption KPIs, compliance with recommended usage, data quality targets, etc.);
- **Over the long term, acculturate** decision-makers up to COMEX level, and business units, to the need to invest in enablers;
- **Rapidly capitalize** on early successes, so as to convey the value generated through concrete, measurable examples.

Note: To show their credentials and win (and keep) their trust, they must also succeed in factualizing the ROIs of the use cases they address, in order to be heard by decision-makers.



... and what's next?

Preparing for the future
by putting people at the
heart of transformation





A shortage of talent in Data and AI

Our analyses show that the **number of Data/AI jobs** posted online **increased by 102%** in 2024.

Faced with this imbalance of supply and demand, which is tending to increase as talent is also attracted by prestigious, high-paying positions abroad, companies are getting their act together on several fronts to attract and retain talent:

- Contribution to **school relations** and student training pathways within institutions training future Data and AI talent
- **Reskilling** and **upskilling** existing resources
- Definition of **career paths** to offer attractive career paths to technical profiles in particular, who are not all destined to occupy managerial positions but to support technical expertise and technological strategy

The impact of AI on employment: a long-term strategy to think about now

AI will also have an impact on employment in general. Its potential for automation will profoundly transform many sectors of activity. Some professions will be heavily impacted, requiring a re-evaluation of skills and associated tasks. Against this backdrop, it is crucial for companies to adopt a proactive stance, and to start **planning their workforce** now. In concrete terms, this means:

- **Evaluate the impact** of AI on the various businesses, particularly in relation to the first industrial AI projects planned and the associated roadmaps
- Implement **forward-looking employment and skills management**, and plan, upstream, the reskilling of impacted employees to new, more sustainable functions that are less impacted by AI
- **To manage change**, with the involvement of social bodies and in a spirit of transparency

CDO Radar methodology

The CDO Radar presents a selection of the leading topics for Data and AI professionals, which they are likely to handle in their business.

The visual is organized into dials that delimit key themes: A, B, C, D, E.

Each of these themes is divided into three maturity levels: "Mature", "Trending" and "Emerging".

- **Mature** level topics can and should be mastered by any CDO;
- Topics categorized as **"Trending"** are beginning to be handled operationally; initial feedback can be shared;
- The **"Emerging"** level covers subjects that are still little-known, in movement or for which there are no obvious solutions. Identifying them enables you to anticipate future developments and prepare for their arrival in your organization.

Authors



**Chadi
Hantouche**

Partner,
Wavestone



**Ghislain
de Pierrefeu**

Partner,
Wavestone



**Stéphan
Mir**

Associate Partner,
Wavestone



**Romain
Levan**

Manager,
Wavestone



Contributors: Guillaume Le Floch, Sina Movaghar, Ibrahim Sail.



ABOUT WAVESTONE

Wavestone is a consulting powerhouse, dedicated to supporting strategic transformations of businesses and organizations in a world that is undergoing unprecedented change, with the ambition to create positive and long-lasting impacts for all its stakeholders.

Drawing on more than 5,500 employees in 17 countries across Europe, North America and Asia, the firm offers a 360° portfolio of high-value consulting services, combining seamlessly first-class sector expertise with a wide range of cross-industry capabilities.

Wavestone is listed on Euronext Paris and recognized as a Great Place to Work®.

www.wavestone.com