4 Congrés d'Economia i Empresa de Catalunya 2025
Cap a una economia més digital i sostenible

EIX6 | El sistema de ciència i innovació i el repte de la transformació digital

Col·legi d'Economistes de Catalunya
Al servei dels professionals de l'economia i de l'empresa

Chilling out data chaos: leveraging cold chain principles and the triple helix framework

Giulio Toscani,
ESADE

ABSTRACT

As big data increasingly influences decision-making, ensuring data consistency and reliability has become essential. This paper explores how the cold chain's singular focus on temperature can serve as a model for building unified, high-quality, and sustainable data systems. Despite the primary emphasis on machine learning of literature, this grounded in theory research, investigates how big data quality impacts technological execution and decision-making across industries. Drawing on 65 expert interviews, the study's final discussion highlights the potential of the Triple Helix model, collaboration between government, industry, and society, as a framework for creating unified data systems. The cold chain, with its focus on temperature as the sole variable, exemplifies how such collaboration can enhance data reliability. Key challenges identified include data governance, operational inconsistencies, and barriers to collaboration. The paper proposes solutions such as professional training and the development of collaborative infrastructures to improve data dependability and support sustainable decision-making across industries.

Keywords: Outcome, Data expert, Big Data, Value deterioration, Cold Chain, Triple Helix

1. Introduction

In recent times, the proliferation of digital technologies has led to a surge in big data accumulation, sparking interest among professionals and scholars. Organizations leverage big data for insights and a competitive edge, but its size and complexity present challenges (Fan, Fang & Han, 2014). Concerns include data quality, emphasizing the role of subjective choices by researchers (Naeem et al., 2022). This research contrasts human design with unbroken technological execution in the cold-chain sector, analyzing value deterioration over time (Singh et al., 2018) and its impact on sustainability (Farrukh & Holgado, 2020). Miller and Mork's (2013) data value-chain model guides the study, examining decision-making processes from a value-rational approach. The focus is on how data experts navigate decision-making processes and behaviors within this context (Snihur & Zott, 2020). More research is needed on how Data

experts' thinking patterns and behaviors relate to data deterioration, aiming to provide insights for optimal data use (Rahwan et al., 2019).

The paper aims to help organizations understand how data use may drive or impede data projects by exploring data experts' decision-making patterns and behaviors. The research question is:

RQ: How do Data experts' decision-making patterns and behaviors influence data value?

2. Literature review

In the present era, the widespread assimilation and advancement of digital technologies have led to a significant amplification in the accumulation of big data, the extensive volumes of data due to their efficacy in enhancing organizational efficacy (Al-Dmour et al., 2023). Firms utilizing AI exhibit distinctive traits: a unique operational framework prioritizing data, integration of AI into core processes, real-time decision-making through experimentation, and granular forecasting. They proactively learn from customer feedback via real-time experiments, enhancing offerings through continual data evaluation. (Iansiti & Lakhani, 2020). The extensive attention and scholarly curiosity from experts and researchers across diverse academic fields can be attributed to the rationale behind the significant interest evoked by big data (Abbasi, Sarker, & Chiang, 2016). The variability of human conditions will undermine the outcome of human decisions, instead, an approach to a machine learning decision-making tool would be less prone to variation, but still affected by human biases (Obermeyer & Lee, 2017; Allen & Choudury, 2022). However, regrettably, the findings indicate that the degree to which companies have implemented big data analytics applications is moderately and diversely observed among them due to inadequacies and disparities in their organizational and technological capabilities (Al-Dmour et al., 2023). Big data poses challenges due to its high dimensionality and large sample size. These challenges include noise accumulation, spurious correlations, and incidental homogeneity caused by high dimensionality. Additionally, the combination of high dimensionality and large sample size leads to issues like heavy computational costs and algorithmic instability. Furthermore, the aggregation of massive samples from multiple sources at different time points using various technologies adds complexity to the analysis (Fan, Fang and Han, 2014). This is important because when classification algorithms use human-generated input data that suffer from human biases (Sayogo et al., 2014), the predictions they generate may exacerbate the errors stemming from such biases and technical debt increase (Park, Jang & Lee, 2018). In the presence of variability in the bias-induced error, the impacts of bias can be mitigated, but not eliminated, even if the algorithmic design is adjusted to account for the bias (Ahsen et al. 2017). The moderation of resource capital influences the collection and storage of consumer activity records as Big Data, the extraction of insights from Big Data, and the utilization of insights to enhance dynamic/adaptive capability. Inadequate organizational alignment and member education on proactive utilization of insights can hinder effective utilization of consumer insights and impede a firm's adaptive capability (Erevelles, Fukawa, & Swayne, 2016). This may hinder the effective utilization of data may lead to bias-conscious algorithms that have the potential to greatly enhance anticipated results, although the extent of improvement

relies heavily on the discriminatory capacities of information, encompassing accurate long-term strategizing; however, despite the awareness of big data's advantages, top-level management fails to prioritize it, ranking as the secondary factor contributing to unsuccessful data outcomes, consequently leading to a scarcity of financial resources as the primary cause (Ahsenet al., 2017; Al-Dmour et al., 2023). Miller and Mork (2013) recommended a collaborative partnership for data collection from diverse stakeholders, aiming to optimize service delivery and quality decisions. They proposed streamlining data management activities to benefit all stakeholders and adopting a portfolio-management approach to invest in people, processes, and technology, thus maximizing the value of integrated data and improving organizational performance (Toscani, 2023).

Furthermore, despite significant advancements in safeguarding healthcare data in the digital age, vulnerabilities remain, allowing attackers to compromise even highly secure and sensitive information stored on cloud servers, potentially modifying it for their personal gain, despite ongoing efforts by researchers to enhance security measures (Gupta, Gaurav, & Panigrahi, 2023). The study conducted by Lebovitz et al. (2022) substantiates the costly nature of Big data outcomes, prompting a focused examination on the direction of investments and the prioritization of key factors by companies to achieve superior data quality. With this research, we tried to shed a light on uncovering latent consumer insights significantly enhances adaptive capacity, enabling firms to anticipate future trends and behaviors, as exemplified by Amazon's anticipatory shipping and Target's pregnancy detection, thereby emphasizing the importance of an ignorance-based perspective and inductive reasoning for eliciting novel inquiries that can uncover hidden consumer insights beyond existing knowledge (Erevelles, Fukawa, & Swayne, 2016). When considering the human-led brokerage process in an algorithm, we have dug into the collective long-term output vs. the single decision-maker short-term brokering in her/his favour (Waanderburg et al., 2022). This study highlights the absence of a cohesive framework for data collection and sharing among organizations, resulting in security vulnerabilities exploited by attackers. Consequently, the establishment of a resilient and standardized protocol for processing business-to-business healthcare data becomes imperative. (Gupta, Gaurav, & Panigrahi, 2023). Our collaboration involving Data experts has the potential to intricately analyze the procedural aspects of data generation, as opposed to its utilization. Our focus is not directed towards the methodology of chefs preparing dishes, but rather akin to the techniques employed by chemists in formulating the constituent elements that are subsequently manipulated by the chefs. Concurrently, our endeavor seeks to demystify the exalted status often accorded to technology, while advocating for the primacy of human agency by advocating for decision-making processes rooted in human cognition. However, it is imperative to contemplate mechanisms whereby collaborative efforts between the public and private sectors are leveraged to foster innovation (Toscani, 2024) via the facilitation of platformization processes (Woods, Bunnel & Kong, 2023). The analysis of these extensive datasets is crucial for assessing both firm-specific risks and systematic risks. It necessitates the expertise of professionals well-versed in advanced statistical techniques applicable to portfolio management, securities regulation, proprietary trading, financial consulting, and risk management (Fan, Han and

Liu, 2014). The primary focus of this paper is not centered on individual skills in isolation; rather, it delves into a comprehensive capability that surpasses the confines of human expertise. From an academic standpoint, when analyzing the data cold-chain process, it becomes evident that numerous obstacles arise due to the lack of uniform practices across diverse domains, leading to the deterioration of data value. Indeed, it is unlikely to portray a comprehensive workflow that encompasses the entire data lifecycle, spanning from the initial acquisition phase to the subsequent stages of data processing and storage (Belcore et al., 2021). In attaining this objective, it is imperative to engage key stakeholders and implement a techno-bureaucratic framework, amalgamating strategies targeting urban social, economic, and environmental complexities, prioritizing citizen engagement and ecological stewardship (Palumbo et al., 2021). Therefore, while it would be an act of reductionism to conflate the complexity inherent in data metrics with temperature, focusing on the singular metric characterizing the cold-chain, the proposition of establishing reduced and collectively agreed-upon metrics presents noteworthy advantages, as previously underscored by Fu et al. (2022) through their formulation of objective performance benchmarks. This would be similar to how the internet operates seamlessly today, a remarkable illustration of the advantageous outcomes resulting from a consensus of various stakeholders, including government, industry, and academia, have collaborated as partners in the progressive advancement and implementation of the internet (Leiner et al., 2009). Furthermore, the OCI, (Open Container Initiative), introduced under the patronage of the Linux Foundation in the middle of 2015, endeavors to institute inclusive benchmarks within the domain of container runtime and configuration. In light of the pervasive adoption of containerization, OCI has swiftly garnered patronage from an excess of 30 corporations and entities, encompassing eminent purveyors of cloud computing and application frameworks (Fu et al., 2016). Furthermore, bureaucratic structures uphold fundamental principles while dynamically influencing and responding to surroundings through lower-level agents, crafting regulations amid societal shifts (Lekkas & Souitaris, 2023).

Data science endeavors confront obstacles pertaining to data accessibility, data integrity, and dependence on external repositories, culminating in escalated intricacy. In response to these issues, one proposed resolution has been to embrace interconnectivity (Gunther et al., 2017), which refers to the ability to integrate data from heterogeneous big data repositories. Evidently, the advent of sophisticated technologies is progressively empowering users to assimilate disparate data sources and extract valuable insights from their combination. Hence, this scholarly investigation endeavors to bridge the lacuna in contemporary data research, which predominantly centers on the domain of Big Data. It does so by pivoting the attention towards a comprehensive end-to-end data cold-chain paradigm.

3. Methodology

The grounded theory method (Corbin and Strauss, 2015) facilitates a rigorous analysis of data, exploring the interplay between good data and big data. Interviews with 65 Data experts from 21 countries across various fields aimed to understand their relation to data value deterioration. The broad geographical scope was strategically chosen for diverse

insights, maintaining proximity to unorthodox paradigms (Bamberger, 2018). Respondents were contacted via Linkedin.com, resulting in a 5% response rate from invite to interview. The interviews were conducted in English through Zoom, recorded, transcribed by otter.ai, and analyzed using Atlas.ti.

The primary theoretical construct derived from interviews is the "data cold-chain," emphasizing coherence in inputs and processes for data tasks (Von Krogh, 2018). An iterative approach to data analysis led to the identification and categorization of distinct features within data technology. Decision-making processes of executives and managers in creating data systems were explored through recorded interviews, averaging 30–45 minutes in length. A four-step, grounded, iterative analysis process (Corbin & Strauss, 2015) involved open coding, analytical filters, axial coding, and selective coding. The resulting conceptual model illustrates Data experts' thinking patterns and behaviors regarding data value.

Semi-structured interviews covered respondents' backgrounds, the data cold-chain, thinking patterns, and behaviors in data deployment. Open-ended "grand tour" queries and exploratory inquiries explored their experiences, learning trajectories, skill sets, challenges, and decision-making processes. The iterative refinement of the interview protocol accommodated emerging thematic elements, ensuring a comprehensive understanding of data experts' experiences. The sampling process continued until data saturation was achieved, indicating no new information emerged from supplementary responses.

4. Data analysis

We established a theoretical foundation in innovation and technology management literature, guiding interviews with 65 diverse data experts (refer to figures 1, 2, and 3).

Figure 1 Machine learning data (Source. BCG primary and secondary research; AI implementation projects including multiple interviews)

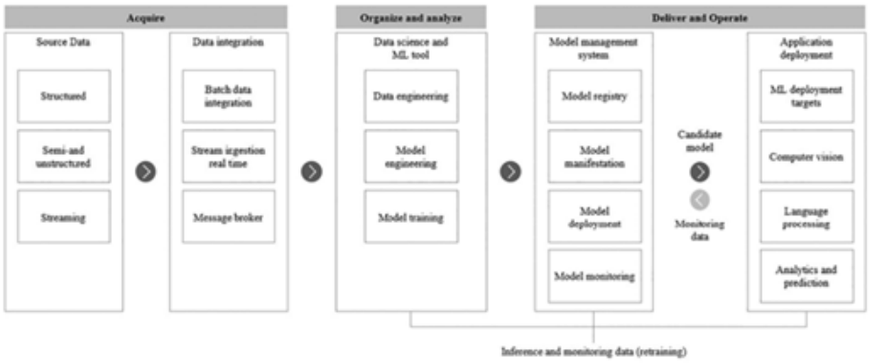


Figure 2 Respondents' company sector

Figure 3 Respondents' Country

The iterative coding process, facilitated by atlas.ti software, involved rigorous scrutiny of interview transcripts and memos, organizing respondent comments into thematic categories aligned with preliminary codes. we drew insights from relevant literature, such as works by Chen (2012), Jean, Burke et al. (2012), and ghahramani (2015), to interpret emerging concepts and strategies related to algorithmic biases.

Engaging data experts in comprehensive inquiries, our research aimed to understand their data interactions, perspectives on data strategy, awareness of evolving dynamics, enhancements in team dynamics, and identification of those benefiting from strategic choices. although unable to undertake secondary observations and field visits, our interviews provided profound insights into human impact on data, contributing significantly to the literature. the expert process model we present enriches the understanding of human influence in data technology and underscores variations in individuals' approaches to data work.

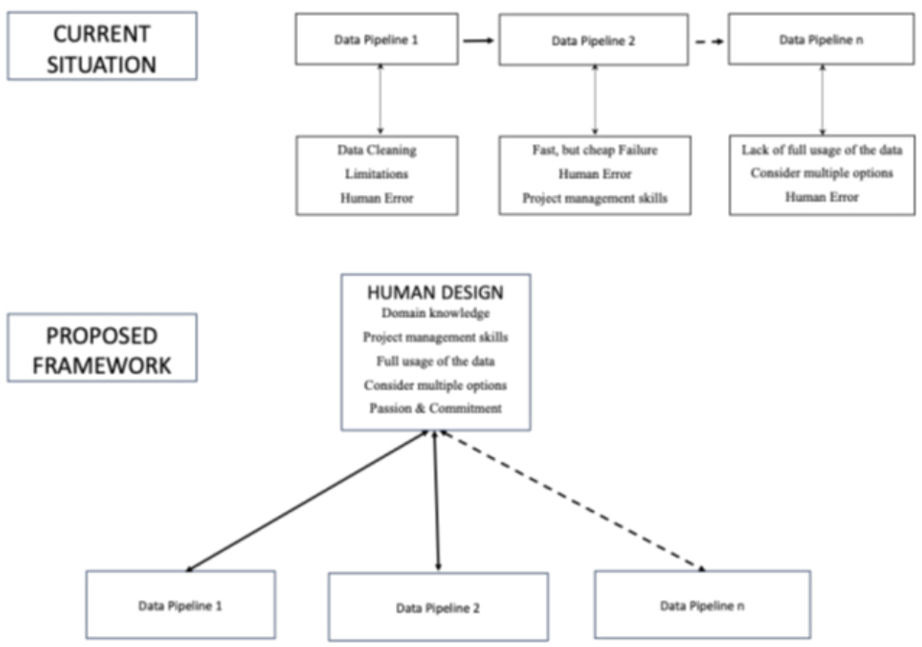
The qualitative data analysis process involved open coding, memo writing, axial coding, constant comparison, and selective coding. nine distinct categories, such as "data experts alignment," emerged from the analysis, representing discrete units of information. the foundational concept of alignment underwent iterative development during the 39th interview, involving continuous comparative analysis, additional sampling, and participant validation. the concept's robust utility and explanatory effectiveness were affirmed through evaluation among participants at later stages.

5. Findings

In the following discourse, we delineate the cognitive processes and behavioral tendencies exhibited by Data experts in relation to the various methodologies employed within the

domain of data management. Subsequently, we expound upon the foundational characteristics underpinning these approaches, categorizing them into nine distinct classes. The present categorizations, encompassing both inadvertent occurrences such as human error and essential components like project management, can be ascribed to the prevailing characteristics of the extant data pipeline within the context of the prevailing data structure. In contrast, within the framework expounded upon in this investigation, it is conceivable that a centralized human-driven design could serve as the governing entity, exercising command over all facets inherent to the data pipeline, predicated upon a set of requisite attributes, specifically: Consider multiple options, Project management skills, Passion & Commitment, Domain knowledge & Full usage of data. (Fig. 4). The findings of this study underscore the imperative for a standardized and cohesive data collection-to-response data pipeline framework. Moreover, it accentuates the indispensability of human involvement within the data processing continuum. The discernible implication is the requirement for an initial stratum of automated processing, wherein each discrete stage should be methodically designed through augmentation methodologies (Toscani, 2025B; Toscani 2025C).

Figure 4 Framework



Data experts approach to data

The approach engendered by the cognitive patterns and behaviors of data experts may give rise to challenging scenarios and offer insights into the potential optimization of data value workflows, as posited by Bamberger (2018). To illustrate, the significance of expertise as a constraining factor is exemplified in situations where professionals are recruited despite possessing limited skills within a specific

domain, based on their perceived long-term potential for delivering valuable contributions, as discussed by Snihur and Zott (2020). The introduction of a novel phenomenon into conventional software development life cycles can be discerned through behavioral inquiries, wherein, in the absence of traditional control mechanisms, collectives assume self-organizing roles, adapting their knowledge-sharing processes to enhance their capabilities in the context of deliberation and resolution, as elucidated by Kane et al. (2014). In the context of this study, nine features emerged from the comparative analysis to elucidate the characteristics of value deterioration:

1. Data Cleaning Limitations
2. Lack of full usage of the data
3. Human Error
4. Fast, but cheap Failure
5. Time Constraints
6. Consider multiple options
7. Project management skills
8. Passion & Commitment
9. Domain knowledge

In the ensuing discourse, we expound upon the intricate particulars encompassing each of the nine delineated categories, culminating with the central category of Data experts' allignment.

5.1 DATA CLEANING LIMITATIONS

The main insight from the interview is that the quality of a data model is only as good as the quality of the data used. The participant emphasizes the "garbage in, garbage out" principle, underscoring the importance of obtaining clean, noise-free data to prevent wasting time on extensive data cleaning. Consulting experts at every stage is critical to ensure the data's appropriateness for the task, which optimizes model performance and reduces inefficiencies. While data cleaning is essential for optimization, it is not the ultimate goal of big data analysis but rather a necessary tool to ensure accuracy. However, real-world datasets are often small, dirty, and even biased, which poses significant challenges in data science. These imperfections can act as limiting factors that influence or constrain model outcomes. As a result, data scientists must balance the task of refining datasets with the broader objective of deriving meaningful insights from imperfect data. This approach not only enhances workflow efficiency but also mitigates the risks associated with poor-quality data, ensuring more reliable decision-making. As per the following informants' quotes:

> #2 I also try my best to get a sense of the type of data that I can use for building this model. So sometimes it's already provided to you sometimes it's up to you to decide how to collect the data. So, the quality of your model is only as good is the quality of your data garbage in, garbage out. It's something I live by. So it's too much noise.....I could spend weeks denoising the data when I could just ask someone, can you please give me a dataset that has minimal noise from a session recorded with minimal noise where the patient for example, was not moving around a lot or are sleeping? You know, there is a lot of noise in the data. So yeah, every step along the way, I would get a sanity check from the experts in an aspect of what I'm working on.

Data highlight the challenges faced in deadline management and data organization within data science projects. Deadlines can be frustrating due to uncontrollable factors like uncertain data sources and unforeseen issues, leading to disheartenment despite a clear understanding of tasks. Additionally, there is a recognition that improvements in time management, particularly for machine learning models, could

significantly benefit the company. The need for better organization of datasets is identified as a task more suited for data engineers. Furthermore, the participant emphasizes the necessity for increased IT support for software installation and technical tasks, noting that simplifying procedures would alleviate frustration and bureaucratic hurdles.

> #53: Deadlines can be frustrating due to factors beyond my control like uncertain data sources and unexpected issues. A clear understanding of the task makes deadline management easier, though encountering obstacles can be disheartening.
>
> #58: Improvements in time use, especially for machine learning models, could benefit the company. Organizing datasets is better suited for data engineers. More IT support is needed for tasks involving software installation and technical matters, which can be frustrating and bureaucratic. Simplifying procedures would be highly beneficial.

Two critical aspects emerge here of data science workflows. First, while pre-processing and coding tasks are manageable, the duration of machine learning model training is largely dependent on computational resources. Even with optimal resources, training large models can take hours or even days, which affects the speed of feedback and iterations. Second, data experts need to continuously update their knowledge of tools and techniques to foster innovation. Engaging with various implementations can broaden their perspectives and enhance their problem-solving capabilities, ensuring they remain effective in a rapidly evolving field.

> #59: In data science, pre-processing and coding are controllable, but machine learning model duration relies on computation resources. Even with excellent resources, it can take hours or days for large models, impacting feedback and iterations.
>
> #60: Data experts must stay updated with tools and techniques for innovation. Reading different implementations expands their horizons.
>
> #62: Time constraints in big companies limit data experts' ability to innovate.

5.2 LACK OF FULL USAGE OF THE DATA

There seem to be three important aspects of data management in machine learning. First, optimization is essential for effectively handling large datasets, as it directly impacts model performance. Second, strategies such as unbiased sampling methods are implemented to manage data more efficiently, helping to overcome memory limitations and reduce loading times. Lastly, privacy concerns pose significant challenges, limiting the full utilization of data while underscoring the importance of protecting client information.

> #2 Optimization, particularly in machine learning for handling large datasets, is crucial.
>
> #12 Strategies like unbiased sampling methods are employed to manage data, overcome memory limitations, and reduce loading time
>
> #13 Privacy concerns limit full data utilization, with an emphasis on protecting client data

Respondents emphasize the iterative and experimental nature of data science and machine learning, where continuous improvement in model accuracy and performance is a key objective. The participant highlights their role in training analysts to independently manage basic statistics while encouraging a shift towards more frequent use of machine learning and predictive modeling. This transition is critical for enhancing their analytical capabilities. Additionally, the participant reflects on a past experience in banking, where the failure to fetch complete database data significantly impacted

customers, underscoring the importance of thorough and accurate data handling in preventing adverse outcomes.

> #27 Because in data science and machine learning, it's like you're starting from the scratch, it's mostly about experimentation, we start with from the scratch, we try to build up the data and we follow the whole procedure and we get the results at the end of whatever the experiment we do. And we are always working to improve the results, we are always trying to improve the accuracy or the model performance at the end of the day. So we are always aiming towards that and whatever the progress you know, I have achieved throughout the day, in particular task that is also being discussed.
>
> #58 In my role I train analysts for independent handling of basic statistics, aiming to shift their mindset toward utilizing machine learning and predictive modeling more frequently.
>
> #61 I guide them on incorporating this valuable knowledge into their work. In a past banking job (a failure to fetch all database data affected numerous banking customers, highlighting the impact of incomplete tasks. In my role I train analysts for independent handling of basic statistics, aiming to shift their mindset toward utilizing machine learning and predictive modeling more frequently. I guide them on incorporating this valuable knowledge into their work. In a past banking job (quote a failure to fetch all database data affected numerous banking customers, highlighting the impact of incomplete tasks.

5.3 Human error

Respondents exhibit meticulousness in their procedural approach, diligently posing critical inquiries, especially in the domains of data retrieval and coding. This conscientiousness enables the discernment of a fundamental error. Furthermore, the statements underscore a disparity between the initial theoretical promise of innovative ideas from another department and their practical implementation, revealing inefficacy or flaws. This accentuates the significance of practical testing and validation in the evaluation of the viability of inventive solutions.

> #5 Practical testing is crucial for assessing the viability of innovative solutions, as initial perceptions can be misleading
>
> #53Human intervention remains vital in evaluating and comprehending data, ensuring accuracy, and setting realistic expectations for algorithms
>
> #6 Adding more people to a project may lead to slower progress, with potential advantages but a risk of complexity

Furthermore, practical testing is essential for validating innovative solutions, as initial perceptions may not accurately reflect their viability. Human involvement is crucial for evaluating and understanding data, ensuring accuracy, and managing realistic expectations for algorithm performance. However, increasing the number of team members on a project can slow progress due to added complexity, despite potential advantages. Training machine learning models is time-intensive and can disrupt focus, emphasizing the need for efficient data retrieval to avoid setbacks. Additionally, human error can arise from assumptions influencing data, potentially leading to deviations from established knowledge and causing concerns for clients. Data underscore the necessity for a balanced approach that integrates human expertise, effective project management, and rigorous testing to enhance the reliability and efficiency of data-driven initiatives.

> #59 Training machine learning models is time-consuming, disrupting concentration, and efficient data retrieval is crucial to avoid setbacks
>
> #45 Human error arises when assumptions directly

> influence data, potentially diverging from established knowledge and unsettling clients

5.4 Fast, but cheap failure

The interviews highlight that fostering an experimentation culture is fundamental to success in data science. A key insight is the focus on iterative learning, where teams continuously refine their processes by applying agile methods and adopting a "learning by doing" mindset. In industrial settings, this approach is particularly important, as starting models from scratch is emphasized to ensure accurate processing and better alignment with real-world data requirements:

> #2 The experimentation culture in data science leads to key conclusions
>
> #4 In our unit we emphasizes starting models from scratch for accurate processing,
>
> #8 There is an highlight on learning by doing in industrial settings,
>
> #10 We ambraces an agile approach

Efficient time management and proactive problem-solving are repeatedly highlighted as essential traits. This means not just identifying issues but solving them in a forward-thinking manner, preventing bottlenecks. Teams also recognize the value of effective communication, as prompt feedback ensures that insights from experiments are shared early, helping to maintain momentum:

> #16 It is important to stress efficient time management. Effective communication, quality data, learning from mistakes, and proactive problem-solving are vital.
>
> #20 We advise starting with smaller data chunks to save time, communicating findings promptly, and avoiding delays.
>
> #61 uses small datasets as tests before applying to larger ones, while illustrates the importance of seeking reliable data sources.

A strategic approach to data handling is also emphasized, with smaller datasets recommended for initial testing. This method saves time and ensures that data reliability is assessed early on, minimizing the risk of errors down the line. The importance of data quality, swift feedback cycles, and an ongoing willingness to learn from mistakes are underlined as critical to refining data models and processes. By consistently focusing on these elements, data science teams can better align with industrial needs, ensuring agility and adaptability.

5.5 Consider multiple options

The qualitative interviews reveal that exploring multiple options enhances efficiency and adaptability in data collection (#61, #4). Standards for model performance are acknowledged as sufficient (#2), but bridging the gap between data science and other departments maximizes the data's potential for informed decision-making (#60). Automating tedious tasks enables a focus on more creative aspects of data science (#6), while workflow optimization ensures a continuous flow, minimizing delays (#15). The inclusion of diverse viewpoints in projects emphasizes the need for careful algorithmic application to avoid unintended consequences (#45). Together, these insights highlight the balance between automation, creativity, and cross-functional collaboration.

> #61: Exploring multiple options prevents being stuck, enhancing efficiency in tasks like data collection.
>
> #2: Model performance assessment relies on standards, acknowledging the sufficiency of existing methods.
>
> #60: Bridging the gap between data science and

other departments maximizes data potential for informed decisions.

#4: Creativity and adaptability are key in navigating challenges, emphasizing the need for dynamic planning.

#6: Utilizing machines for tedious tasks allows a focus on more exciting aspects of data science.

#15: Workflow optimization ensures a continuous flow of data, eliminating waiting periods in the job.

#45: Diverse viewpoints in data science projects highlight the importance of cautious algorithmic application.

5.6 Project management skills

the critical role of project management expertise in data science initiatives emerges: a skilled data expert underscores the importance of knowing when to initiate processes and the relevant techniques to employ. Limited participation in business-related discussions can lead to a fragmented understanding of projects, demonstrating the need for strong project management to facilitate successful outcomes and collaboration between data science and business teams. Furthermore, collaborative decision-making not only minimizes regrets but also reduces the necessity for extensive code modifications, fostering a growth-oriented mindset. Organizations that engage in diverse meetings for planning and coordination, alongside inter-team collaborations, enhance their ability to prioritize effectively and achieve their goals.

#61 A proficient data expert emphasizes the need for project management expertise, understanding when to initiate processes and the associated techniques

#12 This is evident in a case where limited involvement in business-related calls led to a fragmented understanding of the project, highlighting the necessity of competent project management for successful outcomes and collaboration between data science and business teams #37 Collaborative decision-making minimizes regrets and reduces the need for substantial code modifications, reflecting a growth-oriented mindset Organizations engaging in diverse meetings for planning and coordination, coupled with inter-team collaborations, ensure efficient prioritization and goal realization

5.7 Passion & Commitment

The insights from these interviews highlight several key factors affecting motivation and performance in data science projects. Tight deadlines can lead to missed details and dissatisfaction, underscoring the need for realistic timelines to maintain quality. Engaging in intellectually stimulating projects is vital, as individuals express disinterest in repetitive tasks; thus, maintaining focus and engagement is essential for success. When evaluating data science candidates, important attributes include passion, dedication, motivation, adaptability, a commitment to continuous learning, and strong social skills, which are crucial for collaborative environments. Additionally, positive feedback plays a significant role in motivating individuals for future projects. However, facing challenges can be demotivating, emphasizing the importance of reflecting on recent efforts and accomplishments to foster resilience and perseverance. Overall, these insights suggest that a supportive environment and meaningful work are critical to maintaining motivation and enhancing performance in data science roles.

#53 Pushing to meet tight deadlines results in missed details and dissatisfaction.

#25 Intellectually stimulating projects, expressing disinterest in repetitive tasks, and emphasizes the importance of maintaining engagement and focus"

> *#61 When assessing data science candidates, key factors include passion, dedication, motivation, adaptability, continuous learning, and strong social skills*

commitment to continuous learning, and strong social skills, which are crucial for collaborative environments. Additionally, positive feedback plays a significant role in motivating individuals for future projects. However, facing challenges can be demotivating, emphasizing the importance of reflecting on recent efforts and accomplishments to foster resilience and perseverance. Overall, a supportive environment and meaningful work are critical to maintaining motivation and enhancing performance in data science roles.

> #35 Positive feedback is identified as crucial in motivating individuals for future projects. Overcoming challenges and persevering through difficulties can be demotivating, prompting reflection on recent efforts and accomplishments

5.8 Domain knowledge

The complexities of achieving true representativity in data science, are because of its highly dependence on individual circumstances, making a universal definition difficult to establish. Engaging in business-oriented discussions allows respondents to gain valuable context and explore data iteratively, facilitating the sharing of findings with stakeholders and generating new questions that drive further inquiry.

> #12 True representativity is elusive, depending on individual circumstances, making a universal definition challenging
>
> #32 Engaging in business-oriented discussions, respondents gain context and explore data iteratively, sharing findings with stakeholders and generating new questions

Additionally, practices such as "shadowing" with business units play a crucial role in disseminating domain knowledge, effectively bridging the gap between technical and business language. This enhances communication and collaboration, ensuring that data-driven insights are relevant and actionable within the business context.

> #45 "Shadowing" with business units spreads domain knowledge, bridging the gap between technical and business language for effective communication

5.9 Time constraints

There is an inherent uncertainty in data science, where multiple initial spikes or proof of concepts (POCs) are acceptable as the investigation time remains unpredictable. Managing expectations regarding immediate results is essential, as it underscores the importance of clear project timelines and the acquisition of necessary datasets. The use of minimum viable products (MVPs) allows for continuous feedback, helping teams navigate the challenges associated with time and resource management. By setting realistic expectations and utilizing iterative approaches, data science projects can be more effectively aligned with organizational goals, enhancing their potential for success.

> #12 explains that in data science, it's acceptable to have numerous initial spikes or proof of concepts (POCs) because the time required for investigation is uncertain.
>
> #60 Our daily activity further emphasizes the challenge of getting immediate results, highlighting the need to manage expectations regarding time and resources. Setting clear project timelines, obtaining necessary datasets, and using minimum viable products (MVPs) for continuous feedback are crucial in addressing these challenges.

5.10 Data experts' allignment (core category)

Close collaboration between data experts and

other professionals is essential for delivering effective services. While data experts often prioritize insights over software quality, it's crucial to ensure alignment across the entire organization. Careful research and planning can prevent testing issues and errors, resulting in significant time savings and improved accuracy. A lack of alignment poses a major challenge, as data experts often focus on insights during tight timelines, mirroring the mindset of other roles like data engineers and QA specialists. This alignment transcends mere process coordination; it requires recognizing each individual action as part of a cohesive "cold" chain. To enhance internal processes in data science, active engagement and the suggestion of improvement ideas are necessary, alongside dedicated time for data exploration and analysis. Thorough planning is vital to avoid testing issues and minimize the need for extensive model training. By leveraging smaller, high-quality datasets and establishing robust testing protocols, organizations can save time and boost accuracy. Moreover, maintaining alignment within an external ecosystem of data experts demands effective leadership that acknowledges individual strengths and identifies collaboration needs, ensuring teams work together towards common goals and optimal results.

> #12 Improving internal processes involves active engagement, proposing ideas for enhancement, and dedicating time to data exploration and analysis

Thorough planning is vital to avoid testing issues and minimize the need for extensive model training. By leveraging smaller, high-quality datasets and establishing robust testing protocols, organizations can save time and boost accuracy. Moreover, maintaining alignment within an external ecosystem of data experts demands effective leadership that acknowledges individual strengths and identifies collaboration needs, ensuring teams work together towards common goals and optimal results.

> #16 Thorough planning can avoid testing issues and errors, reducing the need for extensive model training. Utilizing a smaller, high-quality dataset and a robust testing plan saves time and enhances accuracy

> #41 Maintaining alignment within an external ecosystem of data experts, led by a team leader, requires effective leadership, identifying strengths, and discerning collaboration needs

In an academic context, alignment among data experts is vital for the success of projects involving data collection, workflow creation, and dashboard design. When a data expert manages a project independently without adequate synchronization with overarching project objectives, it can lead to significant challenges. This alignment should ideally extend beyond academia to incorporate the Triple Helix model, which includes collaboration with industry and government. Engaging academia, industry, and government fosters a comprehensive approach to data science, ensuring that insights are effectively translated into actionable outcomes. By leveraging this collaborative framework, data experts can ensure their efforts are in sync with broader goals, enhancing the impact of their work across all sectors.

> #35 In an academic context, single-handedly managing the entire project involving data collection, workflow creation, and dashboard design may lead to issues if there's a lack of alignment or synchronization with project objectives

Value deterioration occurs due to a lack of congruence but can be addressed through a cohesive human framework and uninterrupted

technological progression.

6. Discussion

This discussion begins by exploring the question, "How do data experts' decision-making patterns and behaviors influence data value?" The study validates the socio-technical approach (Markus & Topi, 2015), which posits that successful data science projects depend not only on technical excellence but also on effective communication skills and collaboration across teams. This perspective aligns with the Triple Helix model, which stresses the collaborative interplay between government, industry, and society, including academia. The paper critiques the conventional focus on expediency and coding-centric practices, as these may hinder long-term data value, as noted by Nadkarni & Prügl (2021).

A significant focus of the research is on the interpersonal and epistemological differences among data experts, which introduces an additional layer to knowledge dissemination and collaboration. Drawing on Triple Helix literature, such tensions reflect the broader challenges in multi-stakeholder ecosystems, where clashes in epistemologies and operational frameworks often arise, as highlighted in the work of Pachidi et al. (2014). These tensions are mapped within the data pipeline dimensions in the study, underscoring the need for human-centric design and structured communication, akin to governance mechanisms found in successful Triple Helix collaborations. One of the central findings of this research is that many of the problems in data science projects stem not from technical issues, but from miscommunication between team members, underscoring the need for improved clarity and comprehension across sectors. This insight reflects broader lessons from the Triple Helix model, where cross-sectoral collaboration and clear communication channels are essential for aligning diverse stakeholder interests and maintaining focus on shared goals. The study advocates for minimizing anthropogenic intervention in data workflows by establishing a unified data "cold chain," much like the cold chain industry's use of temperature as the sole variable to ensure data consistency. This concept mirrors the structured data frameworks required in Triple Helix collaborations, where unified data metrics allow for smoother integration across governmental, industrial, and societal contexts (Toscani, 2025A). The mundane tasks that occupy data scientists' time hinder critical thinking, and this research suggests that a more integrative, human-centric approach can help safeguard data value throughout its lifecycle, fostering efficient collaboration between all stakeholders.

In terms of challenges, the research identifies that data science projects often face demotivation due to the lack of a supportive atmosphere. In the Triple Helix context, this is echoed in the struggles municipalities and organizations face when attempting to build cohesive, large-scale data ecosystems. Examples from Metro Boston and New York City illustrate the difficulties in achieving uniformity across geographical and organizational boundaries (Kitchin & Moore-Cherry, 2021; OpenData, 2024). However, efforts like cross-sectional data initiatives and public-private partnerships show promise in bridging these gaps. This study's emphasis on human engagement, workflow optimization, and clear communication is crucial for success, as seen in successful urban data ecosystems that utilize similar frameworks (Woods, Bunnell &

Kong, 2023). The integration of human elements and social interactions is vital to project success, as illustrated by Singh et al. (2018) in the cold chain sector, which demonstrates how a singular data metric can minimize value deterioration. This focus on aligning human input with technical outputs aligns with the Triple Helix literature's emphasis on structured collaboration, professional development, and ethical data practices. In particular, this study calls for corporate education initiatives that foster a data-sharing culture, essential for maintaining data value and adaptability. Finally, the Triple Helix model offers a pathway to overcoming fragmented data systems by advocating for collaboration infrastructures and public-private partnerships. Lessons from global smart city projects—like Singapore's emphasis on platformization and stakeholder engagement—show how integrating data across sectors can lead to innovation and sustainability (Woods, Bunnell & Kong, 2023; Zhu & Chen, 2022). This research proposes that, much like the cold chain's reliance on temperature as a unifying metric, the broader data science community must adopt standardized metrics and collaborative strategies to sustain data ecosystems that are both reliable and effective.

7. Theoretical contributions

The theoretical implications of this research extend the understanding of data science project management by integrating insights from the cold chain industry and the Triple Helix model, emphasizing the role of human collaboration in maintaining data value. This study reframes data science projects not only as technical endeavors but as socio-technical systems where cross-sectoral collaboration is essential. The cold chain serves as an analogy, demonstrating how a unified variable, like temperature, ensures data consistency and prevents deterioration. Similarly, the research advocates for the standardization of data metrics and processes across data ecosystems, a key principle in Triple Helix literature, which emphasizes collaboration between government, industry, and society for sustainable data practices. A key contribution of this study lies in its recognition of data experts' decision-making patterns, positioning them as central to the data creation and curation process, much like the standardization processes that optimized human expertise during the industrial revolution. Challenges such as data availability, quality, and reliability, particularly in under-digitalized contexts, are identified as critical factors that can impede the success of data projects (Schweinsberg et al., 2021). These issues reflect broader systemic challenges found in data ecosystems that lack the cohesive frameworks advocated by the Triple Helix model.

The research also highlights the importance of effective communication and collaboration within cross-functional teams, stressing that successful data science projects require not just technical excellence, but the seamless integration of outputs through well-coordinated human interactions. This echoes the Triple Helix's emphasis on creating interconnected networks of stakeholders to foster innovation and sustainability. Moreover, the study suggests that structured project planning, combined with specialized software tools that include visual representations, can simplify complex data science tasks, improve team satisfaction, and enhance overall productivity. Further, this research contributes to the understanding of how the cyclic nature of data science work—spanning business value identification, data analysis,

algorithm development, and deployment—can benefit from a unified approach, akin to the single-variable focus in the cold chain. The seamless integration of technical outputs into operational systems is critical for reducing data deterioration, ensuring that data insights remain accurate and reliable over time. By moving beyond a purely technical focus and emphasizing human-centered decision-making processes, this study supports the notion that the successful execution of data science projects depends on collaboration and alignment across various domains, in line with Triple Helix theory.

In conclusion, this research advances the theoretical discourse on data science management by emphasizing the need for standardized practices, human-centered design, and collaborative frameworks to optimize the potential of data experts. It suggests that adopting a unified, standardized approach—similar to the cold chain's temperature focus—can mitigate the risks of data deterioration and enhance the overall success of data-driven initiatives. Through the lens of the Triple Helix model, this study calls for a deeper integration of cross-sectoral partnerships and knowledge-sharing mechanisms to sustain robust data ecosystems.

8. Practical implication

The research offers valuable practical contributions to the field of data science, particularly by highlighting the importance of understanding data creation processes to prevent value deterioration. Drawing on insights from the cold chain industry, where a singular variable like temperature ensures data consistency, the study emphasizes the need for organizations and researchers to adopt standardized metrics across data workflows. This approach, akin to a "data cold chain," can help align data experts and maintain quality throughout the data lifecycle. Furthermore, integrating the Triple Helix model of collaboration between government, industry, and society proves essential for addressing data accessibility challenges. By fostering interconnectivity and partnership, diverse data sources can be integrated more effectively, yielding better insights and enhancing the decision-making process. The emphasis on sustaining motivation within data science teams underscores the importance of creating a supportive atmosphere—one that prioritizes human-centric design, open communication, and interdisciplinary collaboration. These elements are essential for overcoming the common barriers to success in data-driven projects. The research also stresses the importance of building a data-sharing culture, where data is seamlessly integrated into decision-making processes. Implementing standardized metrics and creating collaborative infrastructures, as seen in successful Triple Helix partnerships, is critical for preserving data value and preventing its deterioration. In conclusion, this study highlights that fostering collaboration, promoting a data-driven culture, and adopting human-centric approaches are key to optimizing data workflows. These strategies not only enable innovation but also help organizations overcome the complex challenges that arise in data science endeavors. Establishing a cohesive data "cold chain" within these frameworks ensures a comprehensive, end-to-end approach to managing data.

9. Limitations

The findings and conclusions of this study are contextually bound to data science projects, particularly in the realm of socio-technical collaborations and data management practices. Extrapolation to other domains or industries should be done cautiously, as unique contextual factors may influence outcomes differently. Furthermore the data collected for this study may exhibit sampling bias, potentially limiting the generalizability of the results. The study's reliance on specific industries, organizational sizes, and geographical locations may restrict the broader applicability of the findings. While the study incorporates a grounded theory approach, it does not consider diverse methodological instruments such as statistical analyses and visualization techniques, these tools may not comprehensively capture all nuances and dimensions of data value preservation. Alternative methodologies and measures could yield different insights. The grounded theory framework proposed is limited to a specific timeframe, and technological advancements or shifts in organizational practices occurring after the study's conclusion might not be adequately represented, potentially affecting the relevance of the conclusions over time.

Bibliographic references

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. Journal of the Association for Information Systems, 17(2), 3.

Ahsen, M. E., Ayvaci, M. U. S., & Raghunathan, S. (2019). When algorithmic predictions use human-generated data: A bias-aware classification algorithm for breast cancer diagnosis. Information Systems Research, 30, 97-116.

Al-Dmour, H., Saad, N., Basheer Amin, E., Al-Dmour, R., & Al-Dmour, A. (2023). The influence of the practices of big data analytics applications on bank performance: field study. VINE Journal of Information and Knowledge Management Systems, 53(1), 119-141.

Allen, R. T., & Choudhury, P. (2022). Algorithm-Augmented Performance and Domain Experience. Organization Science, 33, 149-169.

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. International Journal of Qualitative Methods, 18, 1609406919874596.

Bamberger, P. A. (2018). Clarifying what we are about and where we are going. Academy of Management Discoveries, 4, 1-10.

Belcore, E., Angeli, S., Colucci, E., Musci, M. A., & Aicardi, I. (2021). Precision agriculture workflow, from data collection to data management using FOSS tools: an application in northern Italy vineyard. ISPRS International Journal of Geo-Information, 10(4), 236.

Chen, J. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36, 1165-1188.

Cheng, Y., Li, D., Guo, Z., Jiang, B., Geng, J., Bai, W., ... & Xiong, Y. (2020). Accelerating end-to-end deep learning workflow with codesign of data preprocessing and scheduling. IEEE Transactions on Parallel and Distributed Systems, 32(7), 1802-1814.

Charmaz, K. (2014). Constructing Grounded Theory (2nd edition). Sage Publications Inc.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60, 102383.

Corbin, J., & Strauss, A. (2015). Basics of qualitative research. Techniques and procedures for developing grounded theory (4th edition). Sage Publications Inc.

Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. Harvard Business Review.

Ennen, E., & Richter, A. (2010). The whole is more than the sum of its parts—or is it? A review of the empirical literature on complementarities in organizations. Journal of Management, 207-233.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. Journal of Business Research, 69(2), 897-904.

Fan, J., Fang, H., and Han, L. (2014). Challenges of big data analysis. National Science Review, 1(2), 293-314.

Farrukh, C., & Holgado, M. (2020). Integrating sustainable value thinking into technology forecasting: A configurable toolset for early stage technology assessment. Technological Forecasting and Social Change, 158, 120171.

Fu, S., Liu, J., Chu, X., & Hu, Y. (2016). Toward a standard interface for cloud providers: the container as the narrow waist. IEEE Internet Computing, 20(2), 66-71.

Fu, J., Zhang, Y., Wang, Y., Zhang, H., Liu, J., Tang, J., ... & Zhu, F. (2022). Optimization of metabolomic data processing using NOREVA. Nature Protocols, 17(1), 129-151.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353, 790-794.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, 26(3), 191-209.

Gupta, B. B., Gaurav, A., & Panigrahi, P. K. (2023). Analysis of security and privacy issues of information management of big data in B2B based healthcare systems. Journal of Business Research, 162, 113859.

Greve, H. R. (2003). A behavioral theory of Rand D expenditures and innovations: Evidence from shipbuilding. Academy of Management Journal, 46, 685-702.

Gilli, A., & Lanata (2020). Strategies for Leading in the Age of Artificial Intelligence. NATO Defense College.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. Nature, 521, 452.

Hill, K. (2012). How target figured out a teen girl was pregnant before her father did. Forbes, Inc, 7, 4-1.

Huang, L. (2018). The role of investor gut feel in managing complexity and extreme risk. Academy of Management Journal, 61, 1821-1847.

Iansiti M., Lakhani K.R. (2020). Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World (Har- vard Business School Press, Boston).

Kahneman, D., & Tversky, A. (2000). Choices, values, and frames. Cambridge University Press.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In Handbook of the fundamentals of financial decision making: Part I (pp. 99-127).

Kane, G. C., Johnson, J., & Majchrzak, A. (2014). Emergent Life Cycle: The Tension Between Knowledge Change and Knowledge Retention in Open Online Coproduction Communities. Management Science, 60(12), 3026-3048.

Kitchin, R., & Moore-Cherry, N. (2021). Fragmented governance, the urban data ecosystem and smart city-regions: The case of Metropolitan Boston. Regional Studies, 55(12), 1913-1923.

Laureiro-Martínez, D., & Brusoni, S. (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. Strategic Management Journal, 39, 1031-1058.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To Engage or Not to Engage with DATA for Critical Judgments: How Professionals Deal with Opacity When Using DATA for Medical Diagnosis. Organization Science, 33, 126-148.

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... & Wolff, S. (2009). A brief history of the Internet. ACM SIGCOMM Computer Communication Review, 39(5), 22-31.

Lekkas, C. K., & Souitaris, V. (2023). Bureaucracy meets digital reality: The unfolding of urban platforms in European municipal governments. Organization Studies, 44(10), 1649-1678.

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, 3(1), 91-99.

Markus, M. L., & Topi, H. (2015). Big data, big decisions for science, society, and business: report on a research agenda setting workshop. ACM Technical Report.

McKinsey & Company (2017). A CEO guide for avoiding the ten traps that derail digital transformations. Digital McKinsey.

Merriam-Webster (2022). Accessed on 14th October 2022 on https://www.merriam-webster.com/dictionary/skill.

Miller, H. G., & Mork, P. (2013). From data to decisions: a value chain for big data. IT Professional, 15(1), 57-59.

Mittal, D., et al. (2023). Data management strategy for a collaborative research center. GigaScience, 12, giad049.

Nadkarni, S., & Prügl, R. (2021). Digital transformation: a review, synthesis, and opportunities for future research. Management Review Quarterly, 71, 233-341.

Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., ... & De-La-Hoz-Valdiris, E. (2022). Trends and future perspective challenges in big data. In Advances in Intelligent Data Analysis and Applications (pp. 309-325). Springer Singapore.

Namvar, M., & Cybulski, J. (2014). BI-based organizations: a sensemaking perspective. In Proceedings of the Thirty-Fifth International Conference on Information Systems, Auckland, New Zealand, December 14–17.

Obermeyer, Z., & Lee, T. H. (2017). Lost in thought–the limits of the human mind and the future of medicine. New England Journal of Medicine, 377, 1209-1211.

Pachidi, S., Berends, H., Faraj, S., Huysman, M. H., Weerd, I. (2014). What happens when analytics lands in the organization? Studying epistemologies in clash. Acad. Manage. P., Philadelphia, Pennsylvania.

Palumbo, R., Manesh, M. F., Pellegrini, M. M., Caputo, A., & Flamini, G. (2021). Organizing a sustainable smart urban ecosystem: Perspectives and insights from a bibliometric analysis and literature review. Journal of Cleaner Production, 297, 126622.

Park, J. S., Jang, J., & Lee, E. (2018). Theoretical and empirical studies on essence-based adaptive software engineering. Information technology and management, 19, 37-49.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... & Wellman, M. (2019). Machine behaviour. Nature, 568, 477-486.

Ren, S. (2022). Optimization of enterprise financial management and decision-making systems based on big data. Journal of Mathematics, 2022, 1-11.

Sayogo, D. S., Zhang, J., Luna-Reyes, L., Jarman, H., Tayi, G., Andersen, D. L., ... & Andersen, D. F. (2015). Challenges and requirements for developing data architecture supporting integration of sustainable

supply chains. Information Technology and Management, 16, 5-18.

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C., Van Assen, M. A., ... & Schulte-Mecklenbeck, M. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. Organizational Behavior and Human Decision Processes, 165, 228-249.

Sharma, G., & Bansal, P. (2020). Cocreating rigorous and relevant knowledge. Academy of Management Journal, 63, 386-410.

Singh, A. K., Subramanian, N., Pawar, K. S., & Bai, R. (2018). Cold chain configuration design: Location-allocation decision-making using coordination, value deterioration, and big data approximation. Annals of Operations Research, 270, 433-457.

Snihur, Y., & Zott, C. (2020). The genesis and metamorphosis of novelty imprints: how business model innovation emerges in young ventures. Academy of Management Journal, 63, 554–583.

Spradley, J. (1979). Asking descriptive questions. The Ethnographic Interview, 1, 44-61.

Toscani, G. (2023). The effects of the COVID-19 pandemic for artificial intelligence practitioners: the decrease in tacit knowledge sharing. Journal of Knowledge Management, 27(7), 1871-1888.

Toscani, G. (2024). Creatividad e innovación en la empresa digital: una historia muy antigua. *Harvard Deusto Business Review*, (340), 40-49.

Toscani, G. (2025A). From Code to Context: The Impact of Organizational Problem Framing on Machine Learning Success. *Journal of Decision Systems*.

Toscani, G. (2025B). AI as Augmenting Intelligence prAIority to Enhance Human judgement through Data and AI. Routeldge, London.

Toscani, G. (2025C). El arte de la aumentación: integrando IA, datos y juicio humano en la era digital. *Harvard Deusto Business Review* (356), 2025

Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. Computers in Human Behavior, 72, 577-588.

Von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. Academy of Management Discoveries.

Whang, S. E., Roh, Y., Song, H., & Lee, J. G. (2023). Data collection and quality challenges in deep learning: A data-centric AI perspective. The VLDB Journal, 32(4), 791-813.

Woods, O., Bunnell, T., & Kong, L. (2023). The state-led platformisation of financial services: Frictionless ecosystems and an expansive logic of "smartness" in Singapore. Geoforum, 146, 103849.

APPENDIX

Appendix Figure 1 Interviewee's Consent

Company (Published if permitted by Interviee) =

Date and Place (Not Published) = Barcelona/ 15th February

Type of interview = Online

Please fill the following info on type of Artificial Intelligence:

Natural Language Processing: YES/NO Machine Learning: YES/NO Deep learning: YES/NO Physical Robots: YES/NO Other Topics: Digital Transformation, Quantum Computing, HPC

Overall Work Experience (years):

Work Experience in Artificial Intelligence (years):

AI Interview

You have been selected to speak with us today because you have been identified as someone who has a great deal to share as an Artificial Intelligence expert. This research project as a whole, focuses on the process that Ai users go through in interacting with an AI tool and provides the first theoretical model of this process. This study does not aim to evaluate your techniques or experiences. Rather, we are trying to learn more about how Ai users deal with an AI-driven tool, which can be useful for the AI sector.
To facilitate our note taking, we would like to audio record our conversation today. For your information, only I will be privy to the recordings, which will be eventually destroyed after they are transcribed.
Essentially, this document states that:
(1) all information will be held confidential,
(2) your participation is voluntary and you may stop at any time if you feel uncomfortable
(3) without being asked why you are going to leave,
(4) we do not intend to inflict any harm. Thank you for your agreeing to participate. We have planned this interview to last no longer than one hour. Please sign your consent below.
I certify that all information will be held confidential
Interviewee participation is voluntary and may stop at any time if feeling uncomfortable,

There is no intention to inflict any harm during the interview