

Influence of AI on strategic decision-making

Delegation versus deliberation: the impact of access architecture on managerial teams

Degroote Robbe, De Ryck Lukas, Devoldere Thomas

0841482, 0808271, 1103223

Thesis submitted to obtain the degree of
MASTER IN HET MANAGEMENT

Promoter: Prof. Dr. Iryna Nyenno

Academic year 2025-2026

Copyright Information:

*student paper as part of an academic education
and examination.*

*No correction was made to the paper
after examination.*

Influence of AI on strategic decision-making

Delegation versus deliberation: the impact of access architecture on managerial teams

Abstract

As artificial intelligence becomes increasingly embedded in organisational practice, a critical question emerges. Not whether AI should be used in managerial decision-making, but under what conditions its use supports rather than substitutes human judgment. This thesis investigates how the architecture of AI access in managerial teams shapes ethical reflection, strategic decision quality, and group deliberation in managerial contexts.

A quasi-experimental business game design was employed in which 27 master's-level participants, divided into nine groups of three, solved a constrained strategic business crisis under one of three conditions: No-AI access, Semi-AI access via a single terminal, and Full-AI access. Evidence was gathered across four complementary instruments: a rubric-based evaluation of final strategic presentations, an Interaction Process Analysis (IPA) of audio transcripts, a mapping of AI prompts and search histories, and a post-experiment perception survey.

The results reveal that AI access architecture, rather than mere presence, determines decision quality and process. Although overall differences in strategic coherence and analytical rigor did not reach statistical significance, unrestricted individual access (Full-AI) produced the highest performance ceiling but the lowest floor in final deliverables, driven by variance in participants' prompting capacity. Furthermore, process data revealed that Full-AI access condensed deliberation and reduced human-to-human interaction. While Full-AI users expressed higher AI scepticism, actual factchecking decreased as groups increasingly relied on direct task delegation and treated AI as an ultimate authority. Conversely, while shared Semi-AI access improved systemic thinking, it might simultaneously increase premature strategy commitment. While survey results highlighted an inverse relationship between efficiency and ownership: Full-AI participants reported a significant drop in strategic confidence, despite faster perceived decision speeds and a maximized perceived strategic influence of the tools on the decision-making process. Shared access (Semi-AI), by contrast, produced consistent outcomes, higher systemic organisational thinking, and sustained cross-functional deliberation. The most robust finding concerns ethical reflection: shared access (Semi-AI) scored significantly higher on ethical consideration than both Full-AI and No-AI groups. While Full-AI did not score significantly lower than the No-AI baseline, the data suggests individual AI mediation bypasses the collective ethical reasoning fostered by a shared bottleneck.

These findings contribute to the literature on human-AI collaboration by demonstrating that AI access architecture functions as a governance variable with measurable consequences for both decision quality and ethical accountability. The results carry practical implications for organisations designing AI deployment frameworks, suggesting that constrained, shared access may better preserve the deliberative conditions under which high-quality and ethically grounded managerial decisions are made, although it could simultaneously increase the risk of premature strategy commitment (tunnel vision).

Acknowledgments

We would like to express our gratitude to everyone who contributed to the completion of this thesis.

First and foremost, we extend our sincere thanks to our promotor, Prof. Dr. Iryna Nyenno, for her guidance, expertise, and valuable feedback throughout the entire research process.

We would also like to thank our partners Lou'Ann Jacquemin, Jana Decroubele, and Alexe Heleu for their friendship and support during this journey. Their support has made the past years considerably more endurable.

We would furthermore like to thank each other for the excellent collaboration, mutual trust, and collegiality that made working together such a rewarding experience.

Finally, a special word of thanks goes to our parents and parents-in-law. Their unconditional support and the knowledge that we could always turn to them has meant more than words can express. We would like to express our gratitude to everyone who contributed to the completion of this thesis.

Table of contents

1	Introduction	1
2	Literature review	2
3	Methodology	7
3.1	<i>Hypothesis</i>	7
3.2	<i>Business case</i>	7
3.3	<i>Measuring</i>	12
3.3.1	Slide deck	12
3.3.2	Audio recordings and transcriptions	14
3.3.3	AI usage and search history	17
3.3.4	Survey perception participants	18
4	Results	20
4.1	<i>Slide deck</i>	20
4.1.1	Overview of scores	20
4.1.2	Dimension 1: Strategic coherence	22
4.1.3	Dimension 2: Analytical rigor	22
4.1.4	Dimension 3: Feasibility and constraint alignment	23
4.1.5	Dimension 4: Trade-off and decision transparency	24
4.1.6	Dimension 5: Ethical and human consideration	24
4.1.7	The Full-AI polarization effect	26
4.1.8	Summary of hypothesis evaluation	26
4.2	<i>Audio recordings and transcriptions</i>	26
4.2.1	Quantitative analysis	26
4.2.2	IPA analysis	28
4.2.3	Conclusion	30
4.3	<i>AI usage and search history</i>	31
4.3.1	Discussion of the ethics and legality of uploading files	33
4.4	<i>Survey perception participants</i>	36
5	Discussion	39
5.1	<i>AI as amplifier, not equaliser</i>	39
5.2	<i>The paradox of access and the erosion of cognitive ownership</i>	39

5.3	<i>Cross-functional integration & the structural value of constraint</i>	40
5.4	<i>The ethical gap between deliberation and output</i>	40
5.5	<i>What the No-AI condition reveals</i>	41
5.6	<i>Practical implications</i>	41
5.7	<i>Limitations</i>	42
5.8	<i>Further research</i>	43
5.8.1	<i>Architectural variations in AI-access</i>	43
6	<i>Conclusion</i>	44
6.1	<i>H1: Cognitive delegation (partially supported)</i>	44
6.2	<i>H2: Cross-functional integration (directionally supported)</i>	45
6.3	<i>H3: Ethical reflection (supported)</i>	46
6.4	<i>The Architecture of access as the central variable</i>	46
7	<i>Bibliography</i>	48
7.1	<i>Sources</i>	48
7.2	<i>References</i>	49
8	<i>Attached files</i>	54
8.1	<i>List of tables</i>	54
8.2	<i>List of Figures</i>	54
8.3	<i>Appendix A, E-Trax the business case</i>	56
8.3.1	<i>Appendix A1, Business case – Online documents – CFO CTO CHR – spreadsheets</i>	56
8.3.2	<i>Appendix A2, Business case – Offline documents – CFO CTO CHR</i>	56
8.4	<i>Appendices B through E; Excel files with the results and the necessary Calculations</i>	56
8.4.1	<i>Appendix B, Prompt and search history</i>	56
8.4.2	<i>Appendix C, Slide deck analysis</i>	56
8.4.3	<i>Appendix D Survey</i>	56
8.4.4	<i>Appendix E Transcriptions</i>	56
8.5	<i>Appendix F Turn-around pitches participants</i>	56

1 Introduction

The broad adoption of artificial intelligence (AI) in organisations, governments, business and society seems inevitable. However, specific methods for implementation and their eventual impacts remain speculative. As Bill Gates once famously put it: *“The first rule in any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency.”*

This phenomenon requires rigorous study, since data from McKinsey & Company suggests that 80 % of organizations worldwide utilize generative AI (McKinsey & Company, 2025). Other data suggests that when organisations don't implement it themselves, their employees would. A study by the University of Melbourne and KPMG found that 56% of employees utilized AI without knowing if it was allowed (Gillespie et al., 2025, p. 75).

This thesis attempts to add to the ongoing academic debate by proposing and conducting an experiment using a business game format. This is a simulation which requires participants to analyse and solve a business case under controlled conditions. In this study several groups were given the same business case, while having different levels of access to AI. These circumstances made it possible to compare the way in which people make managerial decisions when confronted with differing access to AI. Focusing on ethics and reliance, this research addresses the primary question: “How do human-AI interactions influence ethical reflection, strategic decision quality, and group deliberation in managerial contexts?”

2 Literature review

To evaluate how AI alters managerial work throughout the last hundred or so years, several authors created definitions and wrote down their ideas on this topic (Fayol et al., 1971, pp. 1-53). One influential approach to defining managerial work is Fayol's framework, which conceptualises management through five core functions: planning, organising, commanding, coordinating and controlling. The planning function consists of defining goals, establishing strategy and developing the necessary plan of action. This action plan relies on the firm's resources, the nature of ongoing work, and future trends. Furthermore, Fayol describes the organising function as follows: *"To organise a business is to provide it with everything useful to its functioning: raw materials, tools, capital, personnel"* (Fayol et al., 1971, p. 53). He defines the command function as: *"For each manager the object of command is to get the optimum return from all employees of his unit in the interest of the whole concern"* (Fayol et al., 1971, p. 97). Coordination, on the other hand, means ensuring that all parts of an organisation, such as its resources, activities and functions, are properly balanced and aligned to achieve overall success. It requires aligning expenditures with financial capacity and synchronizing production with demand. In essence, it means to ensure that all things and actions are to the necessary proportions. Lastly, controlling entails the need to verify that everything occurs in adherence to the established plan. The goal is to identify faults and errors in the execution so that these can be resolved (Fayol et al., 1971, pp. 103-109).

Building further upon this, Mintzberg programmed the unstructured nature of managerial work into ten roles spanning three domains: interpersonal, informational, and decisional. Interpersonal roles include acting as a figurehead or liaison, while informational roles involve monitoring and disseminating information. Crucially, decisional roles position the manager as an entrepreneur, disturbance handler, resource allocator, and negotiator (Mintzberg, 1973). Another method to define managerial work is the skills-approach by Katz. In his article "Skills of an effective administrator" he proposes that an effective administrator or manager needs three skill sets. Being technical, human and conceptual skills. His thesis goes further in stating that all three of these skill sets can be learned by an individual, which makes it especially relevant for training future managers. A second important caveat is that these skills have varying levels of importance depending on the hierarchical level of an individual. Technical skills illustrate this dynamic, while essential at lower levels, their importance diminishes relative to conceptual skills at higher organizational tiers (Katz, 1974).

In the context of decision-making, Herbert Simon's trichotomy further divides the administrative process into three stages: intelligence (identifying problems), design (generating alternatives), and choice (evaluating and choosing) (Simon, 1977). Establishing these classical frameworks allows for a structured comparison of which specific managerial tasks are augmented, substituted, transformed, or newly created by AI systems.

Within these established frameworks, AI is increasingly playing an essential part in supporting dynamic decision making in many fields. AI systems are capable of doing this by leveraging the vast amount of data they have access to, in order to learn and make predictions. The main advantage and goal of AI-Human cooperation is to combine human intuition and contextual judgment with the sheer computational power to make decisions

better than either could make on their own. AI is already being utilized in this way in certain domains, with healthcare being a prime example (Gonzalez & Heidari, 2025).

However, the integration of AI is not without friction. A comprehensive review of these sectors reveals persistent technical, regulatory, and organisational challenges. Including the "black box" nature of complex algorithms and the potential erosion of human judgment. Addressing these risks requires technical solutions, such as Explainable AI (XAI), and robust governance models to ensure technology serves the common good (BaniHani et al., 2024).

Consequently, when evaluating this integration of AI, a primary concern is the impact it has on organisational productivity. The Article "*Artificial Intelligence and firm-level productivity*" by Dirk Czarnitzki, Gaston Fernandez, and Christian Rammer investigated this relation. They did this by measuring firm output in a Cobb-Douglas production function of several inputs such as capital, labour and AI as an additional intangible input. The data they used came from the community innovation survey, conducted for the European Commission, more specifically the German contribution. The German contribution was chosen due to the inclusion of questions about the use of AI that were not found in the surveys from other countries. After a regression analysis they found a positive relation between AI usage/investment and productivity (Czarnitzki et al., 2023).

These findings from the European market are corroborated by Gao and Feng (2023), who analysed micro-level manufacturing data in China. They found that a 1% increase in AI penetration results in a 14.2% increase in total factor productivity, primarily driven by technology upgrading and skill-biased enhancements (Gao & Feng, 2023). An article written by Dritjon Gruda and Brad Aeon has a slightly different opinion. They state in "*Seven Myths about AI and Productivity: what the evidence really says*" that although AI can increase productivity, this isn't necessarily always the case. They argue that AI usage has a significant positive impact on less experienced workers for certain tasks, but a smaller and sometimes even negative impact for more experienced workers working on the same task (Gruda & Aeon, 2025a).

This variance across experience levels is further supported by experimental evidence synthesized by Calvino, Reijerink, and Samek (2025). Their review confirms that productivity gains are predominantly concentrated among lower-skilled workers executing well-defined tasks, suggesting that generative AI may help narrow skill-based productivity gaps. On innovation, the review finds that while AI can enhance idea generation and novelty, AI-assisted outputs tend toward homogeneity, potentially reducing collective creative diversity. On entrepreneurship, early experimental evidence suggests that generative AI lowers entry barriers and accelerates early-stage growth, though outcomes vary considerably by entrepreneur characteristics and usage patterns. Across all three domains, the authors emphasize that effectiveness depends critically on human-AI collaboration, user expertise, task complexity, and trust in AI outputs, with AI functioning best as a complement to rather than a substitute for human capabilities (Calvino et al., 2025). Furthermore to the point of idea diversity, Gruda and Aeon also warn that AI can cause a decline in idea diversity, which can have a harmful impact on innovation (Gruda & Aeon, 2025b, p. 5).

Dell'Acqua et al. introduce the concept of a jagged technological frontier in the paper "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality". This concept entails that what is difficult for AI might be different than what a human perceives as difficult, and that tasks equally difficult for a human might be accomplished to different levels of correctness by an AI. Which can create problems as the utilizer might not realize this and become over reliant on AI for tasks it is not skilled at. The authors also note that productivity gains by AI are concentrated mostly amongst lower skilled workers, instead of higher skilled workers (Dell'Acqua et al., 2026).

In contrast to the focus on vertical upskilling, Vendraminelli et al. (2025) shift the focus to horizontal upskilling. Their paper titled "*The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders*" examines the horizontal upskilling effect that artificial intelligence can have. In this context, 'horizontal' refers to comparisons across different occupation, while 'vertical' refers to comparisons within the same occupation between individuals with varying skill levels. The researchers assigned the same article-writing task to three groups: insiders, adjacent outsiders, and far outsiders. Both insiders and adjacent outsiders operated in similar occupational roles. For the experiment, the insiders decomposed the primary task into two distinct subtasks. The first subtask was conceptualizing the article, the second writing the article itself.

Results indicated that generative AI could bridge the gap between insiders and all outsiders in the conceptualization stage, whilst only being able to do the same for adjacent outsiders compared against insiders. This suggests that there exists a "wall" of needed horizontal knowledge, which an individual needs to have before generative AI is able to bridge the skill gap (Vendraminelli et al., 2025).

However, the assumption that AI universally increases output overlooks significant operational frictions. Simkute et al. (2024) challenge the standard productivity narrative by identifying four "ironies of automation" that induce productivity loss during human-AI interaction. These include a cognitive shift from actual production to demanding evaluation tasks, unhelpful workflow restructuring, frequent task interruptions, and task-complexity polarization. Which points to a phenomenon where AI makes easy tasks easier, but complex tasks harder to execute (Simkute et al., 2024).

The shift towards AI-augmented decision architectures necessitates an evolution in managerial competencies. Research on digital transformation, such as studies exploring CEO competencies in family firms (e.g., *Navigating Uncertainty: The Role of CEO Competencies in Strategic Decision-Making for Digitalisation in Family Firms*), provides empirical evidence that digital, analytical, and financial competencies among top decision-makers significantly influence strategic choices under uncertainty. The integration of AI demands a hybrid strategic-computational mindset. As AI systems assume heavy data-processing tasks and the identification of intricate patterns, the human manager's role shifts heavily towards the "soft elements" that AI currently cannot replicate. These include empathy, ethical judgment, complex negotiation, and the assessment of morale (Anwar et al., 2026).

When looking at firm management it is also important to evaluate the implications that AI adoption has for corporate social responsibility. Schmidpeter and Altenburger state in "*Responsible Artificial Intelligence: Challenges for sustainable management*." that the organizational and social value of artificial intelligence will mainly depend on how it is

governed and integrated. Improper AI implementation may intensify systemic biases and increase operational fragility (Schmidpeter, 2023).

Adjacent to this is the ethical question that comes with using artificial intelligence. As artificial intelligence uses historical training data there is a risk that it will amplify historical biases and wrongs. This subject is explored in "*Why you need an AI ethics committee*" by Reed Blackman. An example Blackman highlighted was the risk of a bank using artificial intelligence to determine how it will grant loans. Training the AI-supported model solely on historical data carried a significant risk of displaying racial biases. The reason for this is that according to 2019 study, banks were 80% more likely to deny loans to black people than to white people. Based on this example and others, Blackman suggests that organisations need AI ethics committees which can identify and mitigate the ethical and legal risks of AI usage (Blackman, 2022).

Humphrey's et al. (2024) examines the cybersecurity and more general risks associated with the rapid adoption of generative AI in business environments. The authors argue that the current widespread trend surrounding generative AI has encouraged organizations to adopt these tools faster than the required governance structures can adequately develop. Which as a result may expose these organizations to significant operational, ethical and data-security risks. The article identifies several cybersecurity risks associated with generative AI, including data poisoning, training-data extraction, model backdoors, and adversarial prompting. These risks are significant because generative AI systems may process proprietary, personal, or commercially sensitive information, while also being vulnerable to manipulation or misuse. The authors connect these risks to two key ethical concerns: overreliance, where users place excessive confidence in AI-generated outputs, and over-trust, where users assume AI systems are safe places to input sensitive information. Both behaviours can increase organizational exposure to cybersecurity breaches and poor decision-making. The authors further frame responsible generative AI implementation through the ethical principles of beneficence, non-maleficence, autonomy, justice, and explicability. They argue that businesses have an ethical obligation to protect company, employee, and customer data when adopting generative AI. To reduce risk, they recommend secure and ethical AI model design, fair and trusted data collection, secure data storage, ethical model maintenance, and employee training to mitigate overreliance and over-trust. Overall, the article contributes to the literature by linking generative AI adoption, cybersecurity risk, and business ethics, showing that AI implementation should be treated as a governance and moral responsibility issue rather than simply a productivity or innovation strategy (Humphreys et al., 2024).

The necessity of such ethics committees stems from a broader, fundamental issue regarding accountability. When AI systems augment or execute managerial decisions, responsibility can become more difficult to assign clearly. Nyenno, Truba, and Tokarchuk (2023) argue that AI participation creates additional managerial responsibility and legal risks, particularly because AI behaviour may be unpredictable and where AI behaviour remains unpredictable and its capacity to autonomously perform core managerial roles remains limited and contested. This makes human oversight, structured expert consultation, and responsible managerial governance necessary. (Nyenno et al., 2023). This principle of human-centric control is also emphasized in the *European Declaration on Digital Rights and Principles* (2022), which asserts that technological systems must not undermine human autonomy or accountability (Metsola et al., 2022). Consequently, organizations are forced to adopt 'responsible AI' frameworks. As previously noted by Schmidpeter and Altenburger (2023), responsible AI requires management to ensure that

AI deployment aligns with core organizational values and societal expectations, preventing situations where no specific individual can be held accountable for system failures or biased outcomes (Schmidpeter, 2023, pp. 2-8).

Another important concern is the legal aspect of AI integration. Although this is most definitely interesting and has an important part to play concerning the integration of artificial intelligence, it will not be a consequential part of this literature review. The reason for this is that none of the authors possesses a substantial legal background, making a meaningful contribution impossible. Nevertheless, this subject does merit some discussion. The book *EU digital Law in the AI era* discusses the legal implications of widespread AI adoption. It examines these through the lens of existing EU law while highlighting necessary regulatory improvements (Synodinu et al., 2025). The most relevant chapters from the book are "*The AI liability Puzzle: Rethinking Defective Product liability for AI*" written by Rodriguez-de-las-Heras and "*Regulating Algorithmic Management of work at EU Level: the adequacy of the platform work directive and the AI act*" by Yannakourou.

Rodriguez examined in her article in what way the current regulatory framework is ready for the unique challenges that accompany widespread AI adoption. According to her, the introduction of AI presents a unique challenge because of several characteristics differentiate it from other products. Characteristics being, the opacity of AI systems, their data-dependency, vulnerability, complexity, autonomy and openness. These challenges are currently in the process of being addressed by European Union, but the author questions if they are being adequately addressed. What this highlights for managers is the high degree of uncertainty that currently exists in the regulatory climate concerning AI, which comes with unique challenges and questions concerning the rate and way in which AI is adopted (Rodriguez-de-las-Heras, 2025).

The second chapter, written by Yannakourou explores how, the rise of Algorithmic management (and the potential role AI plays in this) and its transformation of managerial functions, lead to changes in EU law, using the AI act and the Platform Work Directive (PWD) as examples. Algorithmic management, although often more efficient, comes with risks due to the opaqueness, complexity and possible biases in the way AI makes such decisions. According to Yannakourou, the PWD attempts to do this by granting or confirming several rights to workers. Next to this, it also promotes human oversight in the implementation of decisions made by AI. The AI act takes a risk-based method, which entails that AI systems are regulated depending on the perceived or potential harm they can cause. AI systems for hiring, evaluation or workforce management are classified as high-risk systems according to Yannakourou. A problem with the AI act is that instead of viewing AI as a governing issue, it sees it as a technical compliance issue, thus weakening employee protection. The primary takeaway for managers is that while the regulatory landscape remains unsettled, a significant push exists to implement AI personnel management systems that respect employee rights and to include governance structures which mitigates legal exposure (Yannakourou, 2025).

Thus, a conclusion that can be made from these articles is that AI introduces entirely novel responsibilities for leaders. Managers must now possess the digital literacy to evaluate algorithmic outputs, monitor the ethical conduct of AI systems, and actively motivate staff to adopt and integrate these technologies into their daily routines. This evolution in required capabilities forms the theoretical basis for the variables measured in the subsequent business game experiment.

3 Methodology

3.1 Hypothesis

Three distinct experimental conditions (No-AI, Semi-AI and Full-AI) will be evaluated to investigate the main research question: "How do human-AI interactions influence ethical considerations, strategic decision-making, and planning processes in managerial group tasks?"

While the specific contents of these conditions are detailed thoroughly later in this chapter, the core distinction lies in the architecture of Large Language Model (LLM) integration. The No-AI group serves as a traditional baseline with zero AI access, while the Semi-AI group operates with shared access (one shared workstation) and the Full-AI group utilizes unrestricted, individual AI assistance. Building on classical theories of managerial decision-making (Fayol, Mintzberg and Simon) and contemporary literature on human-AI collaboration, this study expects that the degree of AI access will systematically shape group dynamics.

Rather than assuming a uniformly positive effect of technology, the study explicitly recognizes that AI may alter *how* decisions are made. Anticipating shifts in analytical rigor, cross-functional reasoning, and ethical reflection, the following directional and falsifiable hypotheses are proposed:

- H1. Cognitive delegation: Full-AI groups will score higher on analytical rigor but exhibit a lower frequency of critical questioning and explicit trade-off articulation compared to No-AI groups.
- H2. Cross-functional integration: Semi-AI groups will demonstrate more balanced cross-functional integration than both No-AI and Full-AI groups, as the necessity to coordinate a shared AI terminal reinforces verbal information exchange across role-based silos.
- H3. Ethical reflection: Ethical considerations will be articulated more explicitly in No-AI and Semi-AI conditions than in the Full-AI condition, where individual algorithmic delegation may bypass collective ethical reasoning.

Given the exploratory nature of the study and the limited sample size ($N = 27$, divided over 9 groups), these hypotheses are tested with an emphasis on pattern detection and the identification of emerging trends rather than definitive causal inference.

3.2 Business case

The central objective of this research is to evaluate how varying degrees of access to artificial intelligence, in this case Large Language Models (LLM's), influence managerial group decision-making, specifically regarding ethical considerations, strategic decision-making, planning processes. To achieve this, a quasi-experimental business-game design was implemented centred on a multi-role business case simulation. The experiment is structured as a "hidden profile task", a paradigm in which information is distributed

unequally among group members. This design necessitates high-level verbal collaboration and information synthesis, as no individual participant possesses the complete dataset required for an optimal solution.

The experiment manipulated the primary independent variable, the level and architecture of AI integration, across three distinct conditions. The control group (No-AI) operated without AI assistance, establishing a baseline for human-only deliberation. The second condition, called "Semi-AI", was provided with a shared workstation utilizing Gemini Pro as a centralized "external consultant". The third condition, "Full-AI," granted each participant an individual AI agent. This triadic structure allows for a comparative analysis of how AI access architecture shifts the reliance from conventional research methods to algorithmic outputs.

The primary instrument for this experiment is a proprietary business case involving E-Trax BV, a fictitious Belgian manufacturer of light electric vehicles (LEVs). To ensure the simulation met the rigorous standards of postgraduate management education, the case was modelled after established pedagogical frameworks and industry-standard consulting resources. The case used for the experiment was mainly based on insights and examples retrieved from several professional works, most notably the Darden Casebooks, specifically the 2018-2019, 2023-2024, and 2024-2025 editions (Constantine et al., 2018; Darden Casebook 2023-2024, 2023; Darden Casebook 2024-2025, 2024), and the 2019 HEC case book (Pasotti, 2019). These casebooks were produced by the Darden Consulting Club and the HEC Consulting Club (French business school), respectively. The Darden Consulting Club is a student organization dedicated to informing Darden alumni and students about opportunities within the consulting industry and providing the resources necessary to secure those opportunities (MBA - Culture and Community - Clubs and Organizations - Consulting Club, n.d.). Similarly, the HEC Consulting Club focuses on these aims while also offering direct consulting services to companies (HEC Junior Conseil | Junior Company of HEC Paris, n.d.). Because these casebooks are designed to provide tips, tricks, and modelled business cases for aspiring consultants, they served as an ideal foundation for building a realistic, high-stakes experimental environment. The final book that is worth mentioning is called "Top team bedrijfssimulatie" by Rob Frederix, Dave Hendriks and Marie Van Biesen. It relates to the use of business simulations as a didactical tool to teach entrepreneurship and strategic management and contains several simulations (Frederix et al., z.d.). Adapting these professional materials, the research ensured that the strategic trade-offs and financial complexities faced by participants were grounded in real-world consulting logic.

The simulation is set in March 2026, placing E-Trax BV in an existential crisis. A critical design flaw in the steering column of the flagship Apex 350 e-scooter has led to accidents and a subsequent mandatory European recall of 12,500 units. The crisis is multifaceted with the company facing a liquidity collapse, impending legal claims, and a brain drain resulting from organizational fatigue.

Participants, acting as a management team, were mandated by the Board of Directors and KBC Corporate Banking to execute a data-driven turnaround plan. The task was designed to be mathematically and operationally "constrained," meaning it was impossible to fund all initiatives simultaneously. This high-pressure environment with a pre-set time limit of 120 minutes forced definitive capital allocation choices, thereby exposing the participants' propensity to either engage in rigorous critical thinking or delegate complex trade-offs to the available AI tools.

To ensure a comprehensive analysis across management functions, teams were divided into three roles: Chief Financial Officer (CFO), Chief Technology Officer (CTO), and Chief of Human Resources (CHR). Each role was provided with a unique dossier, enforcing a strict information asymmetry. The CFO was tasked with financial survival, utilizing an interactive Excel model with pre-programmed formulas to track liquidity and medical claim provisions. By using this pre-programmed Excel, it was possible to produce substantiated forecast. Which would cancel out unnecessarily long calculations that would reduce time for strategic thinking. The CTO focused on technical feasibility and factory floor bottlenecks. While the CHR managed human capital and morale using qualitative and quantitative personnel metrics.

Because of this setup it was possible to ensure that all management functions as identified by Fayol were present in the case. The planning function was present in all three roles, the most obvious being the CFO, who had to define the global goals for the organisation and establish a strategy to accomplish the financial target end states. The CHR had to develop a plan of action to deal with the crisis in the human resources department and the CTO had to assist the CFO in creating the global strategy with technical expertise and production data. The organizing function was also present, as all participants had to ensure they could provide the organisation with the necessary tools for it to function. Examples of this being the fact that the CHR had to maintain the necessary personnel and the CFO had to find the necessary funding to maintain the company during the crisis period. The command function was present, but mainly at a high level, since there were no actual employees present during the case. The CHR again had to ensure that the plan of action they came up with maintained the employees that created the most value. Something similar was true for the CTO, who had to make choices regarding the prioritization of certain projects to ensure this value creation. The coordination function was mainly present in the case through the fact that the information needed to create a solution, was divided between the participants. The CFO had to ensure that all decisions led to a balanced whole that achieved success. The last function that Fayol described was the controlling function. This function entails the fact that managers must verify that everything occurs in adherence to the established plan. The controlling function was present in the case in the early stages, where participants had to identify the problems that were occurring in the period prior to them taking over the reins (Degroote et al., 2026a; Fayol et al., 1971).

The three domains that Mintzberg described were also present in the case. These three domains are known as the interpersonal, informational and decisional. The interpersonal role was mainly present between the participants. Since there were no employees present, the figurehead or liaison role could not be included. The informational role was however thoroughly present; the participants had to monitor the information they were provided in the beginning of the case and disseminate the required information throughout their team. The decisional role was the most present role, the participant had to make decisions that would decide the future of the company, by allocating the resources available to them and handling the disturbances (Degroote et al., 2026a; Mintzberg, 1973).

The documentation was divided into digital (Appendix A1) and physical formats (Appendix A2). Physical documents, such as internal correspondence, represented internalized background knowledge. Digital files, primarily data-heavy spreadsheets and production schedules, were intended for active processing and potential AI input. Although participants were not restricted from digitizing physical documents via OCR or photography, this

distinction was implemented to observe whether participants would treat AI as a tool for "big data" processing or as a general-purpose cognitive assistant for all available information.

The main Large Language Model (LLM) employed was Gemini Pro, although certain participants did use ChatGPT pro. It was always ensured that the participants had the ability to submit an unlimited number of prompts, to ensure that they would not be limited, which could interfere with the study. Critically, no system prompts or specialized instructions were used. This raw implementation reflects how contemporary managers interact with general-purpose AI in a low-friction environment. In the "Semi-AI" condition, the shared workstation acted as a bottleneck, so that the LLM would only be used for high-value, consensus-driven queries. All digital documents were available on the shared desktop to replicate an external consultancy workflow.

In the "Full-AI" condition, individual agents had exclusively access to the digital files of their respective users. This mirrored an agentic workflow where AI supports individual productivity but maintains information silos. This setup sought to identify whether unlimited access leads to algorithmic dependency, where the depth of cross-departmental verbal reasoning is potentially sacrificed for the speed of AI-generated summaries.

The study utilized a purposive sampling strategy, targeting Master-level students and recent graduates, with maximum of one year work-experience, from KU Leuven. The decision to target this specific demographic served a dual methodological purpose. Primarily, it functioned as a critical control for digital literacy and age-related variance, mitigating potential confounding variables that could otherwise skew the analysis of human-AI collaboration. By ensuring a narrow age bracket, the study isolated the impact of the experimental conditions from intergenerational differences in technological proficiency. Furthermore, this setup provided an ideal environment to observe strategic decision-making in its most fundamental state. By recruiting participants who lacked extensive industry experience, the research established a *tabula rasa* regarding organizational norms. This allowed for an unfiltered observation of how individuals navigate complex, realistic scenarios without the interference of professional entrenchment, entrenched hierarchical biases, or preconceived notions of corporate workflows. Consequently, the results reflect the direct influence of AI integration on the cognitive and collaborative processes themselves more accurately, rather than the adherence to established professional habits.

Participants were recruited using a "Business Case Challenge" framing, designed to attract high-performance individuals and elicit competitive, rigorous behaviour. The recruitment survey was distributed through a multi-channel approach, including classroom presentations, social media networks, and targeted word-of-mouth outreach, to ensure a broad and diverse reach within the target student population. To preserve the internal validity of the experiment and mitigate the Hawthorne effect or pre-emptive bias, the specific focus on human-AI interaction was intentionally withheld during the recruitment phase. Instead, the study was presented as a general inquiry into management research and collaborative problem-solving. This approach ensured that participants did not adjust their behaviour in anticipation of AI-specific performance metrics. To incentivize participation and recognize the time commitment, all participants received a formal certificate of participation, acknowledging their contribution to academic research. The survey categorized participants based on their academic specializations, aligning Engineering students with the CTO role, Economics/ Management students with the CFO

role and Sociology/ Psychology students with the CHR role. This alignment increased the ecological validity of the simulation, as participants operated within their areas of expertise. Group assignment followed a chronological order based on availability, ensuring a randomized distribution across the three experimental conditions.

The experimental procedure was defined by a high-stakes, time-constrained task requiring the synthesis of fragmented data into a cohesive strategic "turnaround pitch". Teams were allotted a strictly enforced 120-minute window to navigate their dossiers and finalize a five-slide presentation. To maintain the integrity of the time-pressure variable, no additional time was permitted for formatting or editing after the simulation concluded. This standardized deliverable served as the primary instrument for objective measurement, allowing managerial performance to be quantified across the different AI-integration conditions.

The content of the pitch was engineered to demand the deep integration of the information silos held by the CFO, CTO, and CHR. The initial requirement focused on a definitive executive strategy regarding capital allocation and product portfolio management, which necessitated a concise strategic justification for which products were prioritized and which were divested. Beyond this high-level strategy, the team had to provide empirical proof of financial viability over a 12-month horizon. This included the quantification of Capital Expenditure for line conversions and validation, ensuring the liquidity ratio remained above the mandatory 1.6 threshold while covering a minimum €1.5 million provision for medical claims.

This financial discipline was required to align with operational and engineering feasibility, specifically regarding the physical and temporal allocation of the factory floor and the distribution of scarce R&D capacity. Participants were tasked with formulating clear timelines and evaluating the necessity of outsourcing to meet production targets. Parallel to these technical requirements, the management of human capital necessitated a quantitative substantiation of workload to ensure that proposed R&D and operator hours were consistent with available Full-Time Equivalents. The plan also had to detail a leadership strategy to mitigate talent drain and crisis fatigue without provoking labour disruptions.

The strategic roadmap concluded with a mandatory risk assessment, identifying the two most critical strategic threats, such as regulatory delays or supply chain failure, and formulating immediate financial and operational fallback options. By mandating these specific, data-driven outputs, the methodology ensured that the final decision-making quality was not merely an abstract observation but a measurable result. This structure highlights whether the presence of AI led to more rigorous data alignment or if it inadvertently caused teams to overlook the complex interdependencies between the financial, technical, and human layers of the case.

3.3 Measuring

To answer the research question, a multi-method measurement strategy was used. Given the socio-technical nature of human–AI interaction, a combination of process data (capturing how decisions are made) and outcome data (the quality of the final decisions) has been done to ensure that both the influence of AI during the decision-making process and the resulting managerial outputs are systematically evaluated.

The combination of outcome-based evaluation (“Slide deck analysis”) with the process-based methods (“Audio recordings and transcription analysis”, “Prompt strategy mapping”, “Post-experiment survey analysis”), enables a comprehensive understanding of both how decisions are made and how they manifest in final strategic outputs, thereby strengthening the internal validity of the findings.

3.3.1 Slide deck

One of the primary outcome measures consisted of a qualitative assessment of the final strategic deliverables produced by each group. These deliverables take the form of a standardized five-slide “turnaround pitch”, submitted (Appendix F) at the end of the 120-minute-long simulation. This output served as a structured representation of the group’s final decision-making quality, integrating financial, operational, and human resource considerations under time pressure.

The slide decks were used to evaluate strategic coherence, analytical rigor, feasibility and constraint alignment, trade-off and decision transparency, and ethical and human consideration. The “turnaround pitch was structured in 5 slides.

1. Executive strategy and capital allocation: Final strategic decision and justification.
2. Financial survival and compliance (CFO): Liquidity constraints, capital expenditure, and financial feasibility.
3. Operational and engineering roadmap (CTO): Allocation of R&D capacity and production planning.
4. Human capital and risk management (CHR): Workforce allocation, retention, and leadership approach.
5. Contingency plan: identification of key risks and fallback strategies

Each dimension was scored by 6 independent/blind raters (3 field-experts and the 3 researchers) that scored every pitch on an ordinal scale from 0-4, with anchored descriptors at each level to maximize scoring consistency. The full score output is provided in Appendix C. To ensure systematic and comparable evaluation across experimental conditions (No-AI, Semi-AI, and Full-AI), each presentation was assessed using this standardized rubric (table 1), allowing for consistent comparison between groups while preserving the richness of qualitative judgment. The scoring focuses primarily on the underlying reasoning, the explicit articulation of trade-offs, and the alignment between proposed actions and available constraints.

3.3.1.1 Scoring Framework

Dimension	0	1	2	3	4
Strategic coherence	No clear strategy	Fragmented and inconsistent	Basic strategy, weakly justified	Logical and consistent	Highly coherent, clear priorities and focus
Analytical rigor	No use of data	Incorrect use of data	Limited analysis	Correct and relevant analysis	Comprehensive and logically consistent
Feasibility and constraint alignment	Unrealistic	Major inconsistencies	Partially feasible	Realistic	Fully aligned with time, budget, and capacity constraints
Trade-off & decision transparency	No trade-offs identified	Implicit trade-offs	Limited articulation	Clearly identified	Explicitly articulated and well justified
Ethical & human consideration	Ignored	Minimal consideration	Basic acknowledgment	Clearly present	Strongly integrated into decision-making

Table 1: Slide deck analysis – Scoring framework

3.3.1.2 Statistical analysis

Because the outcome scores are measured on an ordinal scale and the sample consists of only 18 observations per condition (N = 54 total), parametric tests relying on normality assumptions were deemed inappropriate. Accordingly, group differences across each scoring dimension were evaluated using the Kruskal-Wallis H test, a non-parametric equivalent of the one-way ANOVA that tests whether the rank-order distributions of scores differ significantly across three or more independent groups (No-AI, Semi-AI, and Full-AI). If the Kruskal-Wallis test reveals significant variance, a post-hoc analysis utilizing the Mann-Whitney U test with a Bonferroni correction will be conducted to identify specific pairwise differences between conditions.

While this multi-layered analysis provides a structured view of the data, several methodological limitations regarding statistical power must be acknowledged. A sample size of nine participants per test group inherently increases the risk of a Type II error, where the analysis fails to detect an effect that actually exists due to individual variance.

Given the more exploratory nature of this study, prioritizing the identification of emerging trends over strict hypothesis testing is deemed appropriate. Therefore, to balance the risk of Type I errors against the substantial risk of Type II errors, the threshold for statistical significance in the omnibus Kruskal-Wallis test is set at $\alpha = 0.10$. Consequently, the

adjusted threshold for the three pairwise post-hoc comparisons is set at $p < 0.033$ ($\alpha/3$). This transparent adjustment ensures that the analysis remains sensitive to patterns in human-AI collaboration that might otherwise be obscured by more rigid parameters, providing a more nuanced foundation for future, larger-scale research.

To provide an intuitive overview of performance profiles across all five dimensions and all three conditions, mean dimension scores are additionally visualized in a radar diagram. This visualization illustrates the relative strengths and weaknesses of each condition across the full scoring framework, making it possible to identify patterns of balanced versus polarized performance that may not be immediately apparent from individual test results alone.

Finally, boxplots are used to visualize the distribution of scores per dimension per condition, displaying the median, interquartile range, and potential outliers. Given the ordinal scale and small sample size, boxplots are preferred over bar charts with means, as they more faithfully represent the underlying data structure without implying interval-level precision.

3.3.2 Audio recordings and transcriptions

A rigorous analysis of the actual group deliberation processes was conducted. Audio and video recordings of the experimental sessions were fully transcribed, capturing the real-time interactions, debates, and decision-making workflows of the nine participating groups (appendix E, transcriptions, IPA codes qualitative page). This was done with the use of Notta.ai, which was able to transcribe the audio files that were retained, in a cost effective and accurate way.

Analysing the transcripts allowed for the triangulation of perceived behaviours with observed, objective group dynamics, mitigating the risk of subjective bias. The transcription analysis was executed in two complementary phases: a quantitative assessment of conversational structure and an LLM-assisted deductive thematic analysis based on an Interaction Process Analysis (IPA) framework.

3.3.2.1 Quantitative conversational analysis

Before analysing the semantic content of the deliberations, the structural and volumetric characteristics of the conversations were measured. This step was designed to objectively test hypotheses related to deliberation paths and cross-functional integration (H2). Variables extracted from the transcripts included three primary metrics to quantify group dynamics:

1. Total conversational volume: The absolute word count generated by each group.
2. Overall interaction frequency: The total sum of speaking turns per group, regardless of the speaker.
3. Role-based participation: The number of individual speaking turns taken by each specific organizational role (CEO, CFO, HR).

The statistical analysis of these metrics required distinguishing between group-level and individual-level data. Group-level metrics (total word count and total group speaking turns) are inherently restricted to a smaller sample size ($n = 3$ groups per condition, $N = 9$ groups total). Conversely, the analysis of role-based participation operates at the individual level, yielding a larger sample size ($n = 9$ participants per condition, $N = 27$ participants total).

Given the small sample sizes and the assumption of non-normal distribution inherent in both data tiers, boxplots were utilized to visualize median values and interquartile ranges (IQR) across conditions for all three metrics.

To identify statistically significant structural shifts induced by AI access, the omnibus Kruskal-Wallis H test was employed for both the group-level and individual-level analyses. This aligns with the slide deck's established statistical methodology. The threshold for statistical significance was maintained at $\alpha = 0.10$. Where the Kruskal-Wallis test indicated significant variance, post-hoc pairwise comparisons were conducted using the Mann-Whitney U test against a Bonferroni-adjusted threshold of $p < 0.033$.

3.3.2.2 Deductive thematic analysis (IPA framework)

To analyse how the groups interacted with the technology and each other, a deductive thematic analysis was conducted using a customized Interaction Process Analysis (IPA) codebook. The codebook was specifically developed and mapped to the study's three primary hypotheses, namely cognitive delegation (H1), cross-functional integration (H2) and ethical reflection (H3).

Category	Variable	Code	Definition	H
AI-RELIANCE & DELEGATION CODES	Direct task delegation	AI 1	Group explicitly transfers substantive analysis or decision-making to the AI	H1
	AI as Authority Appeal	AI 2	AI output is used as a standalone argument or ultimate proof to shut down debate	H1
	Blind Acceptance	AI 3	AI-generated output is accepted and integrated into the plan without critical discussion or verification.	H1
	AI Scepticism	AI 4	Expression of doubt, mistrust, or critical questioning regarding the validity of the AI's output	H1
	Human Override	AI 5	A human participant explicitly corrects, alters, or rejects the AI's output based on their own reasoning	H1
VERIFICATION & ACCOUNTABILITY	Fact Checking	RA 1	Actively verifying facts, data, or AI claims against the provided case materials	H1
	Source Demand	RA 2	A participant explicitly asks another for the source or origin of a proposed claim or data point	H1
	Traceability Statement	RA 3	Explicitly stating the source of information to back up an argument	H1
	Logical Reconstruction	RA 4	The group actively rebuilds or traces the step-by-step logic of an (AI-generated) solution to understand why it works	H1
CAUSAL REASONING	Structured Causal Reasoning	CR 1	Building a clear, rational cause-and-effect argument supported by case data	H1

	Unsubstantiated Causal Link	CR 2	Accepting or stating a plausible but unverified or hallucinated reasoning chain (often AI-driven)	H1
	Statistical Prudence	CR 3	Expressing caution about assuming causation from correlation or questioning data validity	H1
ETHICAL VIGILANCE	Explicit Ethical Concern	EV 1	Actively introducing ethical dimensions.	H3
	Ethical Principle Invoked	EV 2	Reasoning based on moral duties, transparency, or corporate responsibility rather than purely financial metrics.	H3
	Ethical Mitigation Step	EV 3	Translating an ethical concern into a concrete action plan or policy adjustment within the strategy.	H3
	Ethical Neglect	EV 4	Explicitly dismissing or minimizing ethical concerns in favor of pure efficiency, speed, or profit.	H3
STRATEGIC DEPTH	Long-term Consideration	SD 1	Explicitly factoring in long-term consequences over short-term gains during the decision-making process.	H3
	Scenario Thinking	SD 2	Discussing multiple future scenarios or "what-if" situations to ensure the robustness and fairness of the plan.	H3
	Explicit Trade-off Articulation	SD 3	Explicitly weighing conflicting goals against each other	H1
STRUCTURED PLANNING	Role Assignment	SP 1	Explicitly discussing how different roles must collaborate or share information to solve the case.	H2
TEAM DYNAMICS	Constructive Disagreement	TD 1	Voicing conflicting role-based perspectives.	H2
PEOPLE/ STRUCTURE ATTENTION	Human Impact Mention	PS 1	Acknowledging the effect of decisions on stakeholders, particularly employees	H3
	Systemic Organizational Thinking	PS 2	Discussing how a decision impacts the broader organizational structure beyond one's own specific department or silo.	H2
TUNNEL VISION	Single Solution Lock-in	TV 1	Prematurely committing to one strategy without exploring or discussing alternatives.	H1
	Exploration Suppression	TV 2	Actively discouraging the generation of alternative ideas or differing viewpoints.	H1

Table 2: Deductive Thematic Analysis - IPA Framework

3.3.2.3 LLM-Assisted coding procedure and validation

Given the substantial volume of qualitative data generated by the nine transcripts, an LLM was utilized as a first-pass annotator, a method increasingly recognized in computational qualitative research for enhancing coding efficiency. To ensure strict methodological rigor and prevent algorithmic hallucination, the LLM was controlled via a highly structured "master prompt" and only 5 pages at a time were fed into the LLM to mitigate possible laziness of the model (Appendix E, transcripts, IPA Codebook page). The LLM was instructed to perform strict deductive coding based solely on the predefined definitions within the IPA codebook and it was explicitly forbidden from generating inductive codes. The model was obliged to return the following elements:

1. Participant: e.g. CTO
2. Code: e.g., AI 1
3. Evidence: "Exact Dutch quote from the text."
4. Reasoning: "1 English sentence explaining why the code applies to this quote."

To guarantee the validity and reliability of the LLM-assisted coding, a robust human verification process was implemented. Twenty percent of the highlighted Dutch sentences were manually reviewed and coded by the research team in a blinded fashion. The human-assigned codes were cross-referenced with the LLM-generated output to assess inter-rater reliability and confirm that the model accurately interpreted the nuanced group dynamics and accurately applied the IPA definitions. When the correlation of the coding produced a score of 99.8%, the full dataset of LLM-extracted codes was accepted for final analysis (appendix E, transcripts).

3.3.2.4 Quantitative transformation of qualitative codes

To move beyond purely descriptive qualitative insights, the validated IPA codes were aggregated into quantitative frequencies per group. By calculating how often specific behaviours occurred (e.g., the ratio of 'fact checking' instances versus 'blind acceptance' instances), the qualitative data was transformed into measurable variables.

These aggregated code frequencies were subsequently subjected to the same rigorous non-parametric statistical pipeline utilized for the conversational metrics. Kruskal-Wallis H tests ($\alpha = 0.10$) and Bonferroni-adjusted Mann-Whitney U post-hoc tests ($p < 0.033$) were executed to determine whether the frequency of specific cognitive, integrative, or ethical behaviors differed significantly depending on the architecture of AI access (No-AI, Semi-AI, or Full-AI). This mixed-methods approach ensures that the qualitative observations regarding human-AI collaboration are grounded in statistically resilient evidence.

3.3.3 AI usage and search history

For groups with access to AI or external search tools, AI chat logs and search histories were collected after the experiment. Participants were instructed to share the relevant parts of their interaction data, which were subsequently catalogued.

The web history and prompts were analysed to assess multiple things, the first being the intensity of AI usage. This first element is interesting because it gave us the ability to measure the likelihood that the participants utilize AI, especially when comparing the

groups who had limited and unlimited access to AI. A second element that was analysed is the type of the prompts or searches, namely if they were meant to generate a solution, verify a solution the team / AI came up with, structure the slide deck or if it was purely to search for a term. For the AI prompts an additional category is for prompts which served to add information into the AI model, this category is only applicable for prompts which solely add information. If in the prompt there was also a query for a solution or something else, this was seen as the dominant classification. Another additional category is case data extraction, which occurs when the participant uses AI to extract data from the case without interpreting the data itself. For example, asking AI what the requirements from the bank were would be an example of data extraction. Asking AI which personnel do not comply with safety standards would be classified as solution generation. The final category that requires some additional discussion is the sparring partner type of prompt, this entails that the prompter includes both a solution the prompter themselves came up with but also asks the AI to generate potential different solutions. This is distinct enough from either the solution verification or generation as to require a distinct label (Appendix B, "Respon" page).

Another way in which, especially the prompts for the artificial intelligence were catalogued was their relation to the management functions as set out by Fayol. The goal of this was to analyse which functions, if any, the participants were more likely to outsource to the AI. An important caveat is that not all prompts can be categorized as pertaining to a management function, this is especially true for informative prompts. If a participant's queries "*What does the abbreviation FTE stand for?*", this was not categorized as pertaining to a specific management function. However, a prompt such as "*Which employee is the most likely to depart?*" was categorized. A second level on which the relation between the prompts and the management functions were analysed was again the difference between the groups who had their access to AI throttled and those who did not. The results were subjected to a chi-square test to identify the statistical significance of the results.

The reliance the participants had on AI when making their decisions was also measured, this together with the way the participants challenged the recommendations AI made. To analyse this both the way the participants prompted and the way they discussed the results was considered. To ensure that any connection was statistically relevant a Chi Square test was run, with ($p = 0.10$) as significance benchmark.

The dataset enables a direct measurement of AI dependency and interaction patterns, making it possible to distinguish between AI used as a supportive tool and AI functioning as a substitute for human reasoning. It also allows tracing how AI-generated outputs are translated into final managerial decisions.

3.3.4 Survey perception participants

Following the business case, participants completed an 8-item survey to capture their internal perceptions of the decision-making process, collaboration quality, and tool utility across the three experimental conditions (*No-AI*, *Semi-AI*, and *Full-AI*):

1. To what extent did you feel that you fully understood the problem before the group made a decision?
2. What role did the AI-tool (*search engine*) play in structuring the problem and generating solutions?

3. To what extent did your group critically question the proposed solutions before making a final decision?
4. How confident are you that the chosen strategy was the best decision for the company?
5. To what extent do you think the use of AI influenced (*access to AI would influence*) your group's final decision?
6. To what extent did the AI-tool (*Google*) influenced the speed at which your group reached a decision?
7. To what extent do you think it would have been more difficult (*easier*) to solve the case without (*with*) the help of AI?
8. Do you have something you would like to share about your experience?

Questions 1–7 utilized a 5-point Likert scale to triangulate the objective case results with subjective team experiences. Question 8 provided qualitative depth, allowing participants to highlight unforeseen aspects of the collaboration.

The Likert data is treated as ordinal because equidistant intervals between scale points cannot be assumed. Given the small sample size ($n = 9$ per condition, $N = 27$) and the likelihood of a non-normal distribution, the analysis prioritizes non-parametric measures. Following the methodology established in the slide deck- and audio transcription analysis, this analysis utilizes the Kruskal-Wallis H test and subsequent Mann-Whitney U post-hoc comparisons. To mitigate the substantial risk of Type II errors, where the analysis might otherwise fail to detect existing effects, the significance threshold is maintained at $\alpha = 0.10$. Consistent with previous justifications, the Bonferroni-corrected alpha for the three pairwise post-hoc comparisons is set at $p < 0.033$ ($\alpha/3$). This method prioritizes identifying emerging human-AI collaboration trends over more rigid parameters that could obscure nuanced patterns in decision-making.

Boxplots serve as the primary visualization tool to represent median values and the interquartile range (IQR), providing a robust overview of participant consensus and shifts in the centre of gravity regarding decision confidence or difficulty.

To further explore the relationships between participants' tool reliance, cognitive load, and decision confidence, a correlation analysis was conducted. Given the ordinal nature of the 5-point Likert scales and the non-normal distribution of the responses, Spearman's rank correlation coefficient (ρ) was utilized. Unlike Pearson's correlation, Spearman's ρ assesses monotonic relationships by ranking the data points, making it more robust for small-sample, non-parametric analysis. This method allows for the identification of underlying patterns, such as the potential trade-off between AI-assisted speed and individual comprehension depth. To determine the statistical significance of these observed relationships, the correlation coefficients were evaluated using a t-test to compute their corresponding two-tailed p-values. Consistent with the exploratory framework established for this study, the threshold for statistical significance was maintained at $\alpha = 0.10$.

4 Results

4.1 Slide deck

This section presents the findings of the outcome-based measurement of the slide deck presented after every simulation. Results are organized by scoring dimension, followed by an integrated discussion linking the observed patterns to the three a priori hypotheses.

4.1.1 Overview of scores

Prior to discussing individual dimensions, it is informative to consider the overall scoring landscape. Table 3 presents the median (Mdn) scores per presentation per dimension, aggregated across all six raters.

Presentation	Strategic Coherence	Analytical Rigor	Feasibility	Trade-off Transparency	Ethical Consideration
No AI 1	2.0	2.5	2.5	2.0	1.0
No AI 2	2.0	2.0	2.0	2.0	1.5
No AI 3	2.0	1.0	2.0	2.0	1.5
Semi AI 1	3.0	3.0	3.0	2.5	2.5
Semi AI 2	3.0	2.0	3.0	2.0	2.0
Semi AI 3	2.5	1.5	2.0	2.0	2.5
Full AI 1	<u>4.0</u>	<u>4.0</u>	<u>4.0</u>	<u>3.5</u>	2.0
Full AI 2	2.0	1.5	2.0	1.5	0.5
Full AI 3	2.0	2.0	2.0	2.0	<u>0.0</u>

Table 3: Median scores per presentation per dimension (scale 0-4)

Presentation	Strategic Coherence	Analytical Rigor	Feasibility	Trade-off Transparency	Ethical Consideration
No AI 1	2.2	2.3	2.7	2.0	1.3
No AI 2	2.2	1.7	2.3	2.0	1.7
No AI 3	2.3	1.3	1.8	2.2	1.5
Semi AI 1	2.8	3.0	2.7	2.7	2.3
Semi AI 2	2.5	2.3	2.7	2.3	2.3
Semi AI 3	2.5	1.5	1.8	2.3	2.7
Full AI 1	3.8	3.8	3.8	3.5	1.8
Full AI 2	2.2	1.5	1.8	1.7	0.8
Full AI 3	2.0	2.2	2.0	2.0	0.7

Table 4: Mean scores per presentation per dimension (scale 0-4)

A striking pattern immediately emerges from Table 3: the Full-AI condition exhibits the highest degree of internal variance of the three conditions. Full-AI 1 consistently produced the highest scores across almost all dimensions, while Full-AI 2 and Full-AI 3 performed at or below the No-AI baseline. This polarization within the Full-AI condition is a key finding that will be discussed further below.

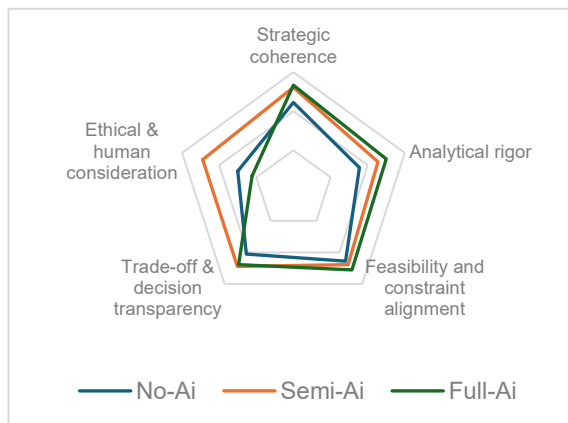


Figure 1: Radar diagram - Average scores

The radar diagram (Figure) provides a first synthetic overview of mean scores (M) per condition across all five dimensions simultaneously. The Semi-AI profile encompasses the largest and most balanced area, with relative strength on strategic coherence (M = 2.61), trade-off transparency (M = 2.44) and ethical consideration (M = 2.44). The No-AI profile forms a smaller but geometrically consistent pentagon, without pronounced peaks or troughs. The Full-AI profile is the most asymmetric. It achieves the highest mean score on strategic coherence (M = 2.67) and performs comparably to Semi-AI on feasibility (M = 2.56) and analytical rigor (M = 2.50), but displays a sharp decline in ethical consideration (M = 1.11), where it falls substantially below both other conditions (No-AI M = 1.5 and Semi-Ai M = 2.44). This visual

asymmetry already anticipates the central statistical finding reported in the dimensional analyses below.

4.1.2 Dimension 1: Strategic coherence

The Kruskal-Wallis test for strategic coherence did not yield a statistically significant result ($H = 3.16$, $p = 0.205$), indicating that differences across conditions cannot be distinguished from random variations at this sample size. At the condition level, median scores were identical for No-Ai and Full-Ai (Mdn = 2.0), while the Semi-Ai condition scored marginally higher (Mdn = 3.0). In terms of means, the pattern was No-Ai ($M = 2.22$), Semi AI ($M = 2.61$), and Full AI ($M = 2.67$).

The boxplot for strategic coherence (Figure 2) makes the distributional differences visually apparent. The No-Ai condition produced a moderately narrow distribution. The Semi-AI condition shows a similar interquartile range (IQR = 2.0 - 3.0) with a median of 3.0, indicating consistent strategic quality across its three presentations.

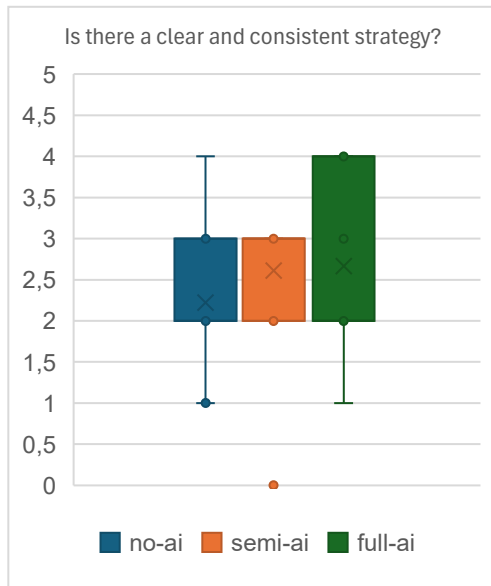


Figure 2: Boxplot – Strategic Coherence

The Full-AI condition displays the widest spread (IQR = 2.0 – 4.0, visually confirming a strong polarization pattern. Although non-significant, the descriptive pattern is noteworthy. One group achieved the maximum score of 4.0 while the other scored 2.0 pulling the condition median down despite having the highest mean of all three conditions ($M = 2.67$, Mdn = 2.0).

Although non-significant, the descriptive pattern is noteworthy. The Semi-AI condition produced the most consistently coherent strategies, while Full-AI exhibited the highest within-condition variance. This polarisation suggests that unrestricted AI access does not uniformly improve strategic coherence but rather amplifies existing differences in how groups engage with and direct AI-tools.

4.1.3 Dimension 2: Analytical rigor

Analytical rigor also did not reach statistical significance ($H = 4.43$, $p = 0.109$). Though the result fell close to the borderline threshold and is classified as a directionally supported finding. Median scores at the condition level were identical across all three conditions (Mdn = 2.0), while means showed a clearer gradient: No-AI ($M = 1.78$), Semi-AI ($M = 2.28$), and Full-AI ($M = 2.50$). Two borderline pairwise differences were observed: No-AI vs. Semi-AI ($p = 0.091$) and No-AI vs. Full-AI ($p = 0.060$), neither of which survives the Bonferroni-corrected threshold ($\alpha < 0.033$). Given that the omnibus Kruskal-Wallis test itself is non-significant, these post-hoc comparisons are reported for descriptive purposes only and should not be interpreted as confirmed effects.

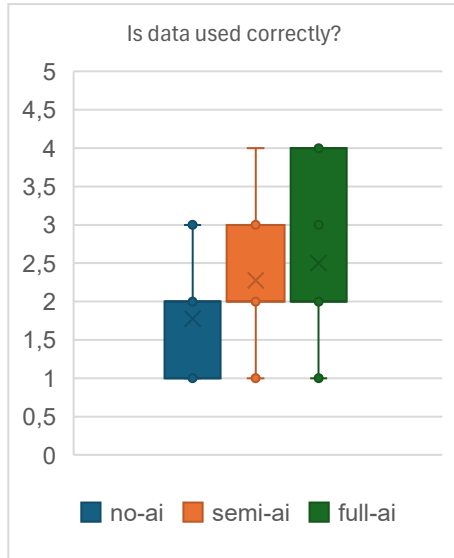


Figure 3: Boxplot – Analytical Rigor

As illustrated in Figure 3, all three conditions overlap considerably on this dimension, with identical condition-level medians of 2.0. The IQR tells a more differentiated story however: No-AI shows a narrow spread (IQR = 1.0 – 2.0), indicating consistently modest analytical performance. Semi-AI shows a wider range (IQR = 2.0 – 3.0) with presentation medians of 3.0, 2.0, and 1.5, reflecting more variability across groups. The Full-AI condition again displays the widest spread (IQR = 2.0 – 4), driven by an extreme divergence between presentations: Full-AI 1 achieved the highest possible score (Mdn = 4.0), consistent with a group that effectively leveraged individual AI access to perform rigorous data analysis, while Full-AI 2 and Full-AI 3 scored 1.5 and 2.0 respectively.

The findings in this dimension were consistent with H1, which predicted that Full-AI groups would score higher on analytical rigor, but the effect is driven entirely by a single high-performing group rather than a systematic condition effect and cannot be statistically confirmed at this sample size.

4.1.4 Dimension 3: Feasibility and constraint alignment

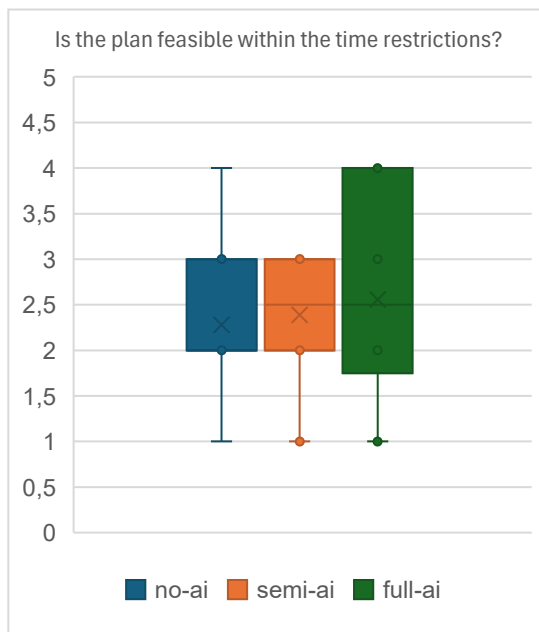


Figure 4: Boxplot – Feasibility and constraint alignment

Feasibility showed the least variation across conditions of all five dimensions ($H = 0.78$, $p = 0.678$), with near-identical condition-level medians: No-AI (Mdn = 2.0), Semi-AI (Mdn = 3.0), and Full-AI (Mdn = 2.0). The means were closely clustered together: No-AI ($M = 2.28$), Semi-AI ($M = 2.39$), and Full-AI ($M = 2.56$). No pairwise comparisons approached significance (No-AI vs. Semi-AI $p = 0.467$, Semi-AI vs. Full-AI $p = 0.669$, No-AI vs. Full-AI $p = 0.477$).

Figure 4 confirms this visually. All three boxplots share similar medians and overlapping interquartile ranges (No-AI: IQR = 2.0 – 3.0; Semi-AI: IQR = 2.0 – 3.0; Full-AI: IQR = 1.75 – 4.0). Full-AI again displays the greatest internal spread, driven by the outlying high performance of Full-AI 1 (Mdn = 4.0) against the two lower-scoring groups (Mdn = 2.0 each).

This finding is classified as a null result. No directional signal was detected across

conditions, and feasibility does not contribute to the evaluation of any of the three hypotheses. The level of AI access had in our exploratory research no meaningful influence

on the ability of groups to produce plans that were realistic and consistent with the financial, temporal, and operational constraints of the case.

4.1.5 Dimension 4: Trade-off and decision transparency

Trade-off transparency did not reach statistical significance ($H = 1.78$, $p = 0.41$). Condition-level medians were identical across all three conditions ($Mdn = 2.0$), with means of No-AI ($M = 2.06$), Semi-AI ($M = 2.44$), and Full-AI ($M = 2.39$).

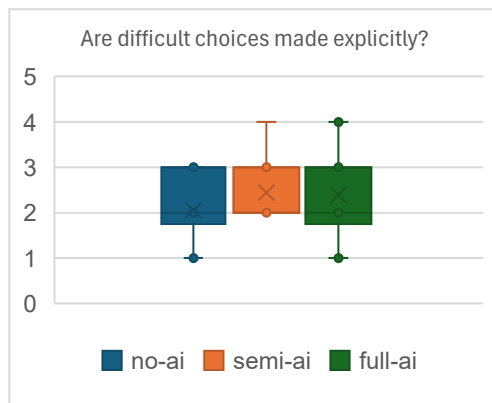


Figure 5: Boxplot –Trade-off Transparency

As shown in Figure 5, all three conditions share the same median but differ in spread. No-AI shows a compact distribution ($IQR = 1.75 - 3.0$) with all three presentations scoring identically at $Mdn = 2.0$, indicating a consistently modest but uniform level of trade-off articulation. Semi-AI displays a similarly narrow range ($IQR = 2.0 - 3.0$) with presentation medians of 2.5, 2.0, and 2.0. The Full-AI condition again shows the widest internal divergence ($IQR = 1.75 - 3.0$), driven by Full-AI 1 ($Mdn = 3.5$) against Full-AI 2 and Full-AI 3 scoring 1.5 and 2.0 respectively, indicating that trade-offs were either implicit or only partially articulated in the lower-performing groups.

The Semi-AI condition demonstrated the most consistent performance across its three presentations, which is directionally consistent with H2, which predicted that Semi-AI groups would demonstrate more balanced cross-functional integration. The pattern is coherent across presentations and aligned with the theoretical prediction, but the effect is too small to support a firm conclusion at this sample size, and no pairwise comparison approaches significance.

4.1.6 Dimension 5: Ethical and human consideration

Ethical and Human Consideration is the only dimension that produced a statistically significant result in the Kruskal-Wallis test ($H = 13.00$, $p = 0.002$), and its boxplot (Figure 6) is visually the most distinctive of all five dimensions.

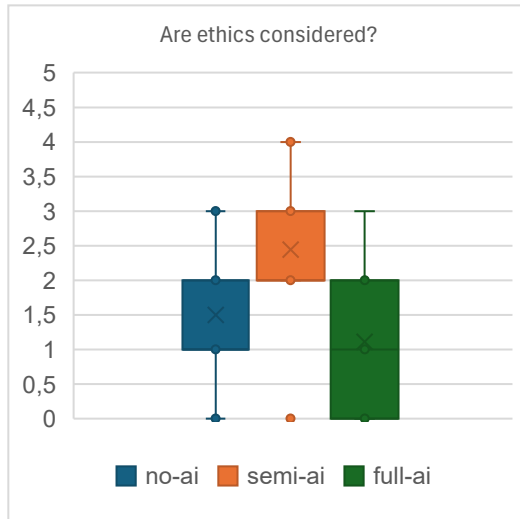


Figure 6: Boxplot –Ethical & Human Consideration

The Semi-AI condition occupies the highest position, with a median of 2.5 and an IQR of 2.0 – 3.0, with a whisker extending to 4.0, indicating that ethical considerations were consistently and sometimes strongly integrated in Semi-AI presentations. The No-AI condition sits in a lower range with a median of 1.5 and an IQR of 1.0 – 2.0. Most strikingly, the Full-AI condition is clearly separated downward, with a median of 0.5 but an IQR of 0.0 – 2.0, indicating that ethical dimensions were largely absent from the majority of Full-AI presentations.

At the condition level, median scores were No-AI (Mdn = 1.5), Semi-AI (Mdn = 2.5), and Full-AI (Mdn = 0.5), with means of No-AI (M = 1.50), Semi-AI (M = 2.44), and Full-AI (M = 1.11).

The post-hoc Mann-Whitney U tests with Bonferroni correction ($\alpha < 0.033$) confirmed two statistically significant pairwise differences: Semi-AI scored significantly higher than both No-AI ($p = 0.006$) and Full-AI ($p = 0.001$). The difference between No-AI and Full-AI did not reach significance ($p = 0.296$).

The individual presentation scores illustrate this pattern in stark terms. Within the Full-AI condition, Full-AI 3 scored 0.0 and Full-AI 2 scored 0.5, meaning that the majority of raters judged ethical dimensions to be either entirely absent or minimally acknowledged in those presentations. Even the strongest Full-AI presentation (Full-AI 1) only reached a median of 2.0, corresponding to basic acknowledgement. By contrast, the Semi-AI presentations scored 2.5, 2.0, and 2.5, and the No-AI presentations ranged between 1.0 and 1.5.

This finding lends empirical support to the argument that accountability risks emerge when managerial decisions are increasingly mediated by algorithmic outputs. identifies the delegation of decisions to algorithmic systems as a key accountability risk, noting that such delegation reduces human discretion and diffuses responsibility across multiple actors (Wieringa, 2020, pp. 9-10). This creates conditions in which it becomes unclear who should account for what. The near-complete absence of ethical consideration in two of the three Full-AI presentations is consistent with this dynamic. When individual participants had unrestricted access to their own AI agent, the deliberation process appears to have shifted towards speed and efficiency rather than reflective deliberation. The delegation of strategic reasoning to AI agents without this reflective deliberation caused ethical dimensions to fail in the final output, illustrating Wieringa's findings.

In the Semi-AI condition, the shared bottleneck of a single AI terminal appears to have forced conditions under which ethical considerations were more likely to be raised and subsequently included in the final deliverable. It should be noted, however, that this mechanism remains a theoretically grounded interpretation. The quasi-experimental design does not allow for a direct causal claim.

This finding strongly supports H3, which predicted that ethical considerations would be articulated more explicitly in No-AI and Semi-AI conditions than in the Full-AI.

4.1.7 The Full-AI polarization effect

A cross-dimensional pattern that deserves particular attention is the extreme internal variance within the Full-AI condition, visible across all five boxplots. Across all dimensions, Full-AI 1 consistently outperformed not only its condition peers but often the entire sample, while Full-AI 2 and Full-AI 3 performed at or below the No-AI baseline. This suggests that unrestricted individual AI access does not produce a uniform uplift in decision quality but rather creates a high-variance environment in which outcomes depend heavily on how individual group members choose to engage with their AI-tools.

4.1.8 Summary of hypothesis evaluation

Hypothesis	Prediction	Finding	Assessment
H1: Cognitive delegation	Full-AI scores higher on analytical rigor but lower on critical questioning	Full-AI 1 scored highest on analytical rigor; Full-AI 2 and 3 did not. No significant condition effect overall.	Directionally supported. Effect present but not uniform
H2: Cross-functional integration	Semi-AI shows most balanced integration	Semi-AI showed the most consistent scores across presentations and performed best on trade-off transparency	Directionally supported
H3: Ethical reflection	No-AI Semi-AI show more ethical consideration than Full-AI	Only Semi-AI > Full-AI ($p = 0.001$), No-AI > Full-AI ($p = 0.296$)	Supported

Table 5: Summary of hypothesis evaluation based on slide deck analysis

4.2 Audio recordings and transcriptions

4.2.1 Quantitative analysis

By examining the structural conversational metrics within an exploratory framework ($\alpha = 0.10$), clear patterns emerge regarding how AI architecture shapes team interaction and cross-functional participation. The analysis progresses from high-level group metrics to granular, role-specific behaviours (Appendix E, transcriptions).

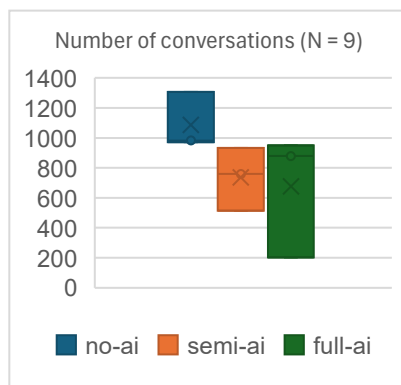


Figure 7: Boxplot – Number of conversations (1)

The analysis initially examined the structural characteristics of the deliberations at the macro group-level ($N = 9$). The Kruskal-Wallis test revealed no statistically significant difference in the total word count across the three conditions ($H = 1.16$, $p = 0.56$). However, when analysing the total number of speaking turns at this same group level, a significant difference emerged ($H = 5.42$, $p = 0.066$). Despite this, post-hoc pairwise comparisons (Bonferroni-corrected: $\alpha < 0.033$) revealed no statistically significant differences in the number of speaking turns between the No-AI condition and either the Semi-AI

($p = 0.049$) or Full-AI ($p = 0.049$) conditions. The lack of significance in the total word count, contrasted with the significant difference in speaking turns, is likely due to the small sample size at the group level. A few very lengthy monologues in a single group can easily skew the total word count, making the frequency of speaking turns a much more robust indicator of actual interaction.

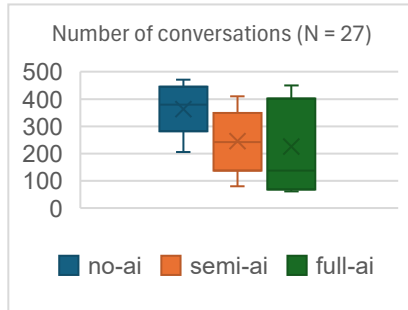


Figure 8: Boxplot – Number of conversations (2)

the number of conversations between the No-AI and Full-AI participants ($p = 0.085$) and Semi-AI and Full-AI participants ($p = 0.72$), although the median score of Full-AI (Med = 138) was the lowest of the three groups. These findings suggest that access to AI (especially in Semi-AI), might limit the conversational volume of managerial teams.

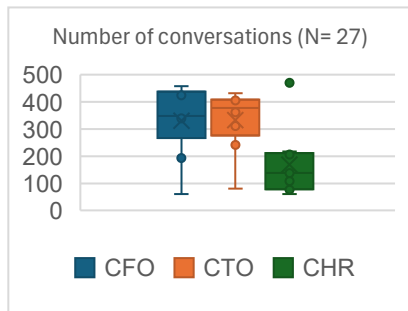


Figure 9: Boxplot – Number of conversations (3)

indicates a baseline structural imbalance within the groups, where the human resources perspective is marginalized compared to the others.

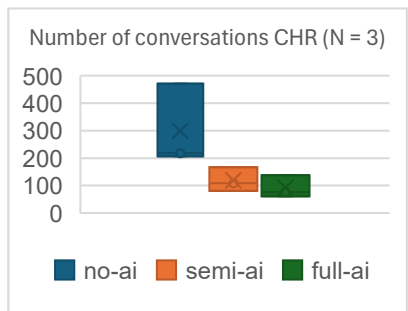


Figure 10: Boxplot – Number of conversations (4)

of conversations was higher (206-407) than the Semi- (80-167) or Full-AI (61-138) groups. There was no significant difference between the two AI-assisted conditions ($p = 0.28$). This

To achieve higher statistical power and validate the trends observed at the group level, the analysis shifted to the individual level, measuring the total conversations per participant ($N = 27$). The omnibus Kruskal-Wallis test confirmed a significant variance in total number of conversations based on AI access ($H = 5.13, p = 0.077$). Post-hoc comparisons using the Mann-Whitney U test showed that individuals in the No-AI condition engaged in significantly more conversations than those in Semi-AI conditions ($p = 0.031$). There was no significant difference in the

Beyond the impact of AI, the transcript data was analysed to determine if participation was balanced across the specific organizational roles ($N = 27$). The Kruskal-Wallis test revealed a significant disparity in how frequently different roles participated in the overall discussion ($H = 6.14, p = 0.046$). Post-hoc analysis showed no significant difference in participation between the CFO versus CTO roles ($p = 0.89$) and the CFO versus CHR roles ($p = 0.052$). However, the HR role exhibited significantly fewer speaking turns than the CTO ($p = 0.021$). This

To investigate whether AI integration exacerbates or mitigates this role-based marginalization, a targeted analysis was conducted solely on the CHR participants across the three conditions ($N = 9$). The Kruskal-Wallis test indicated a significant difference in HR participation depending on the AI environment ($H = 5.96, p = 0.051$). Post-hoc comparisons revealed that HR participants in the No-AI condition did not have significantly more conversations than their counterparts in both the Semi-AI ($p = 0.049$) and Full-AI ($p = 0.049$) conditions, though the range in number

could indicate that the introduction of AI does not help integrate the already marginalized HR role, but must be interpreted with caution, due to the limited data pool (N=9).

4.2.2 IPA analysis

To quantitatively substantiate the qualitative observations regarding group dynamics, the aggregated frequencies of the validated Interaction Process Analysis (IPA) codes were subjected to rigorous non-parametric testing.

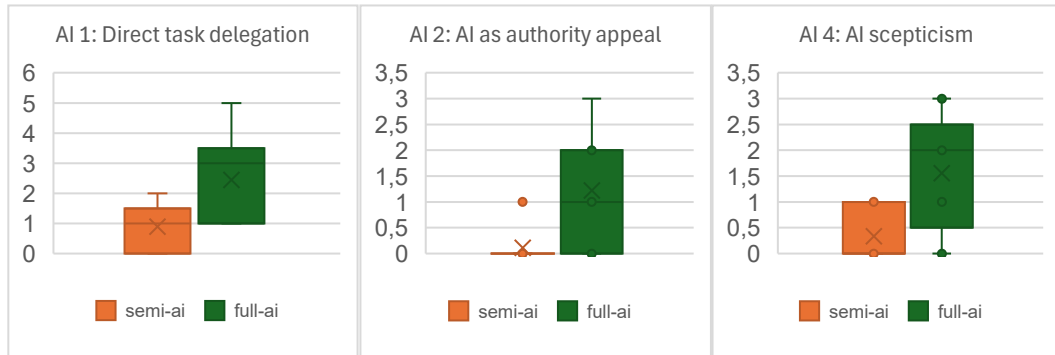


Figure 11: Boxplot – AI 1: direct task delegation

Figure 12: Boxplot – AI 2: AI as authority appeal

Figure 13: Boxplot – AI 4: AI scepticism

Evaluating cognitive delegation (H1), the analysis focused exclusively on the AI-assisted conditions (Semi-AI and Full-AI), as AI-specific behaviours are structurally absent in the No-AI baseline. Mann-Whitney U tests revealed that direct task delegation to the AI (AI 1) was significantly higher in the Full-AI condition compared to the Semi-AI group ($p = 0.031$). This trend of algorithmic reliance was mirrored in the use of AI as an ultimate authority appeal (AI 2), which also occurred significantly more often in Full-AI groups ($p = 0.031$). Interestingly, this heightened reliance in the Full-AI condition was accompanied by a significant increase in explicit AI scepticism (AI 4), which was expressed more frequently than in the Semi-AI condition ($p = 0.024$).

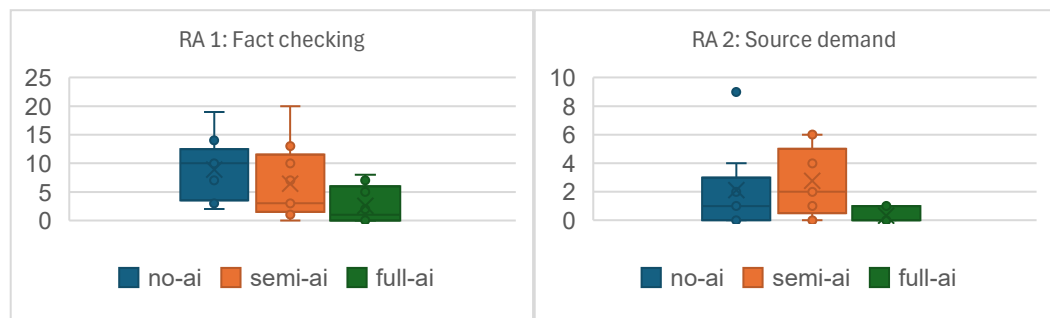


Figure 14: Boxplot – RA 1: Fact checking

Figure 15: Boxplot – RA 2 Source demand

Despite this scepticism, fact checking (RA 1: $H = 7.19$, $p = 0.028$) dropped significantly in the Full-AI environment compared to the No-AI setup ($p = 0.009$). Concurrently, demand for (the origin of) sources (RA 2: $H = 6.34$, $p = 0.042$) decreased, with the Full-AI groups

being significantly less critical than Semi-AI groups ($p = 0.017$). This shows a clear gap between expressing doubt and checking facts. It suggests that unlimited AI access leads to reduced cognitive effort. The convenience and confident tone of the LLM supersede the team's initial doubts, stopping further investigation of the information.

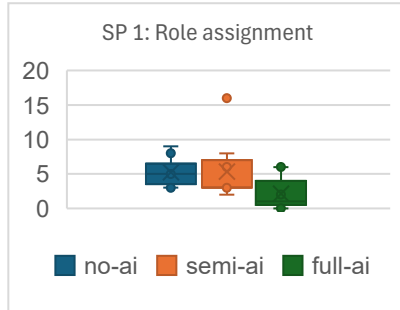


Figure 16: Boxplot – SP 1: Role assignment

The type of AI access did affect how groups organized their teamwork. The analysis of explicit role assignment (SP 1) showed a significant difference across the three conditions ($H = 7.82, p = 0.020$). The post-hoc test revealed that Full-AI teams discussed less frequently how their different roles should work together to solve the case than No-AI ($p = 0.019$) and Semi-AI ($p = 0.019$) groups. This indicates that providing unrestricted, individual AI access reduces the team's effort to openly coordinate their specific functions, resulting in a more isolated workflow.

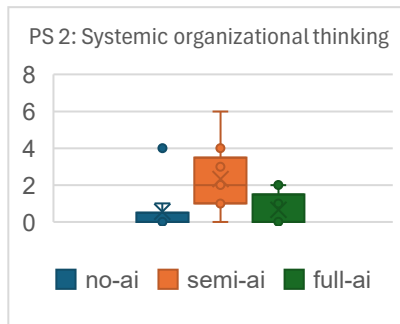


Figure 17: Boxplot – PS 2: Systemic organizational thinking

The architecture of AI-access also impacted cross-functional integration and systemic organizational thinking (H2). The Kruskal-Wallis test identified a significant shift in the groups' ability to maintain a macro-level perspective of the organization (PS 2: $H = 7.2, p = 0.027$). Post-hoc analysis confirmed that systemic thinking was significantly better in the Semi-AI condition compared to the No-AI control group ($p = 0.017$). This indicates that having access to an AI model, functioning as a shared consultant with knowledge of the whole organization, improves the ability to maintain an overview. Furthermore, this shared configuration appears superior to providing unlimited individual access, as the data reveals a marginally significant decline in systemic thinking between the Semi-AI and Full-AI conditions ($p = 0.034$), since the Bonferroni corrected alpha is equal to 0.033.

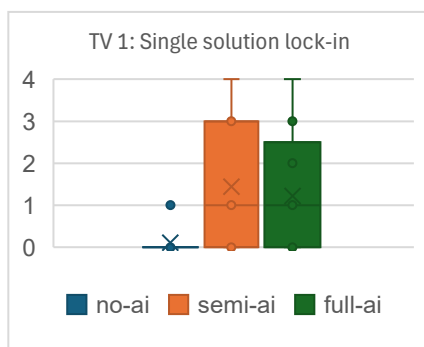


Figure 18: Boxplot – TV 1: Single solution lock-in

The architecture of AI-access significantly influenced the groups' tendency to prematurely commit to a single strategy without adequately exploring alternatives (TV 1: $H = 5.14, p = 0.076$). However, post-hoc comparisons evaluated against a strict Bonferroni-corrected alpha of 0.033 revealed that the pairwise differences narrowly missed strict significance. Specifically, the higher frequency of premature commitment in the Semi-AI groups compared to the unassisted No-AI baseline was only marginally significant ($p = 0.034$). The difference between the Semi-AI and Full-AI groups was not statistically significant ($p = 0.72$) and the difference

between the No-AI and Full-AI conditions also failed to reach the adjusted post-hoc significance threshold ($p = 0.085$). Despite failing strict post-hoc thresholds, the overall trend suggests that the presence of a shared AI-model might sufficiently narrow a group's

exploratory focus, driving them to anchor on a single, AI-supported solution path rather than debating multiple scenarios.

When examining integration at the specific role level, a significant disparity emerged regarding the explicit articulation of strategic trade-offs ($H = 8.78$, $p = 0.012$). The CHR role engaged in this critical integrative behaviour significantly less frequently than both the CTO ($p = 0.008$) and the CFO ($p = 0.019$), indicating a persistent marginalization of the human resources perspective during the deliberation process.

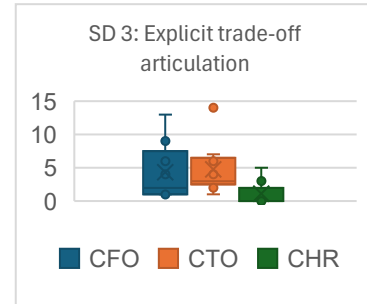


Figure 19: Boxplot – SD 3: Explicit trade-off articulation

Finally, the analysis investigated whether the integration of large language models altered the teams' ethical vigilance and strategic depth (H3). Crucially, the Kruskal-Wallis tests revealed that the presence of AI, regardless of whether access was shared or individual, produced no statistically significant shifts in core ethical reflection. Key variables such as the introduction of explicit ethical concerns, reliance on moral principles, the formulation of ethical mitigation steps, and engagement in long-term scenario thinking yielded no significant variance across the three experimental conditions ($p > 0.10$).

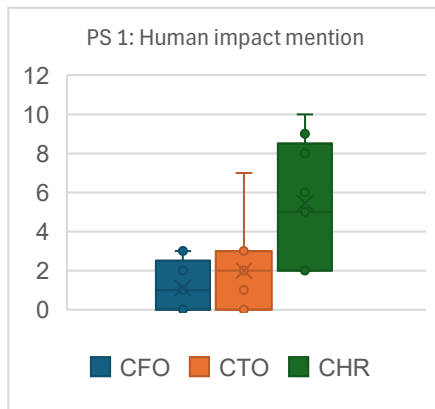


Figure 20: Boxplot – Human impact mention

An analysis of role-based differences in mentioning human impact (PS 1) revealed a significant variance ($H = 9.58$, $p = 0.008$). The CHR role raised this issue significantly more often than both the CTO ($p = 0.024$) and CFO ($p = 0.004$) roles, demonstrating the critical function of the HR perspective. However, the robust absence of differences across the primary ethical codes suggests that AI access neither substantially elevated nor diminished the groups' intrinsic moral reasoning. Consequently, the overall depth of ethical reflection remained static, indicating that algorithmic assistance does not automatically translate into a more responsible or forward-looking corporate strategy.

4.2.3 Conclusion

In conclusion, integrating AI fundamentally changes team decision-making by reducing human interaction significantly and increasing reliance on algorithmic outputs. While shared AI-access improves systemic thinking, unrestricted individual access leads to reduced cognitive effort, reduced fact checking, and a more isolated workflow. Furthermore, AI most likely does not correct existing team imbalances, the human resources perspective remains marginalized regardless of the technology used. Finally, AI-assistance does not improve a group's moral reasoning or ethical focus. Overall, while AI condenses the deliberation path and drives teams to commit to solutions faster (as reported in the survey by participants) it reduces specific verification behaviours, namely factchecking and source demand, without making the final strategy more responsible.

4.3 AI usage and search history

The statistical analysis of the different types of groups also yielded interesting results. The analysis itself was run on two levels, the first level compared both the google search and ai results from all three types of groups. Whilst the second level only compared the AI prompts and was thus limited to the semi and full ai groups.

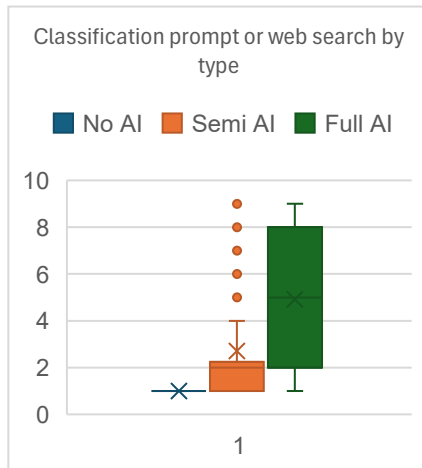


Figure 21: boxplot-Classification prompt or web search by type

A first interesting discovery is that when comparing the types or goals of the prompts and web searches on the first level there is a significant difference between all three groups ($p < 0.001$). All web searches were given an informative query classification (code 1), this since without fail all of these were the participants looking up certain terms. Thus explaining the observed spread for the No-AI groups. When comparing these groups there is also a noticeable difference between the types of prompts and searches utilized by the Semi and Full-AI groups. Where the core distribution for the Semi-AI group was limited to informational queries, solution generation (code 2) and information addition (code 3) with outliers in the other classifications. The Full-AI group however did not limit itself and saw a much more diverse usage of

the AI tools. An interesting note is that the Full-AI groups employed the tool considerably less in an informational query way. A possible explanation for this being the fact that AI gave this information as part of its answers to broader prompts. An interesting addition to this is that there were no web searches from the groups that had full access to AI (Appendix A, Boxplots & Chi Square sheets)

When looking at the types through the lens of the second level the difference between the Semi and Full-AI groups is still observable, however it is less pronounced ($p < 0.001$). A first observation is that the core distribution for the Semi-AI groups shifted upwards, which is a result of the informative query web searched being taken out of account. This function was “replaced” by the structuring (code 4) type of prompt (Appendix A, Boxplots & Chi Square sheets).

When examining which management functions, the AI prompts were meant to address several observations were made. Similar to the type of prompt/search there is a significant statistical difference, both when taking the non ai groups into account ($p < 0.001$) and excluding those ($p < 0.001$). For the remainder of the analysis, the No-AI results will not be further discussed. The reason for this is that, with a single exception, none were assigned a function due to the nature of the search¹ (Appendix A, Boxplots & Chi Square sheets).

¹ The web-search “stopping talent drain in a company” was assigned the management function commanding due to it going further than simply searching for the definition of a term. It was assigned the informative query type designation since the search engine did not have any information regarding the situation of the company

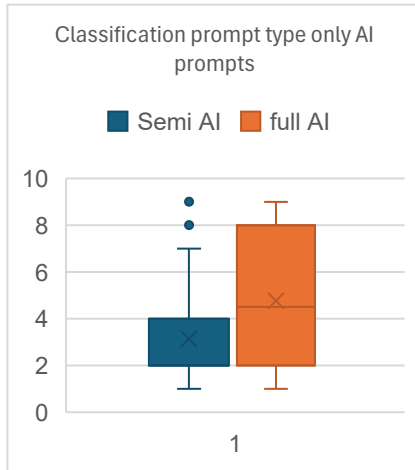


Figure 22: boxplot-Classification prompt type, only AI prompts

In the Semi-AI condition, interactions are mainly concentrated in planning (accounting for 37% of the prompts) and controlling (which accounted for 33% of the prompts). This indicates a strong focus on structured goal-oriented, and monitoring activities. Commanding (15%) and organizing (13%) also play a noticeable role, while coordinating is almost absent (2%). This suggests that Semi-AI prompts tend to emphasize directive and internally structured management functions (Appendix A, Boxplots & Chi Square sheets).

In contrast, the Full-AI condition shows a differently diversified distribution. While planning remains the most frequent category (46%), its dominance is less pronounced compared to Semi-AI. Notably, coordinating increases substantially to 22%, compared to only 2% in the Semi-AI condition. In this context, coordinating does not refer to interaction between employees, but rather to prompts that involve integrating, aligning, or managing multiple elements, processes, or information streams within the AI-generated task. At the same time, controlling (19%), commanding (9%), and organizing (3%) are less prominent than in the Semi-AI condition (Appendix A, Boxplots & Chi Square sheets).

Management functions Fayol				
	Semi AI		Full AI	
not applicable	0	0%	0	0%
planning	17	37%	31	46%
commanding	7	15%	6	9%
controlling	15	33%	13	19%
organizing	6	13%	2	3%
coordinating	1	2%	15	22%
sum	46	100%	67	100%

Table 6: Tabulation of management functions for semi ai and full AI conditions

(Appendix A, boxplots sheet)

Overall, the statistical analysis reveals significant differences in how the three groups interacted with their available tools, both in terms of query types and management functions addressed. The Semi-AI group used AI primarily for informational queries and solution generation, with a strong focus on planning and controlling functions. The Full-AI group demonstrated the broadest and most diverse usage, shifting away from purely informational prompts toward more integrative, coordinating interactions, with planning remaining dominant but complemented by a notably higher share of coordinating prompts. This suggests that greater AI access not only changes how participants use the tool but also influences the nature of the managerial thinking it supports.

4.3.1 Discussion of the ethics and legality of uploading files

When examining the way that the participants treated the files they were given in relation to the AI they had access to there is a small yet difference. Where the participants in the Full-AI condition, always upload nearly all files, almost always in one of the first prompts. The uploads of the Semi-AI condition were more spread out, usually being uploaded in relation to a specific prompt.

When examining the ethics and legality of the files uploaded to the AI it would be pointless to examine each file individually, therefore a selection was made to examine solely the most sensitive. The most sensitive were an email concerning a bank notice about the potential for additional funding, the executive briefing itself and the personnel files. The bank notice is confidential due to fact that it's becoming public could trigger additional investor panic. The briefing gives a complete overview of the crisis and would thus be an immense asset to any competitor who could obtain it. The personnel files would very likely constitute a GDPR breach if uploaded into an LLM without appropriate safeguards or anonymisation, as they contain sensitive personal data such as addresses, dates of birth, and health-related information (*Data Protection under GDPR*, 2025). In the briefing itself it was explicitly mentioned that participants should: "*Be discreet with potential sensitive information towards competitors, employees, ...*" (Appendix A1, the case, briefing)

The briefing and the bank notice can be treated jointly, as participant interaction with both documents reveals the same underlying behavioural pattern. A clear trend emerges in the Full-AI condition: participants consistently began the experiment by uploading all available files to their AI system, including both the briefing and the bank notice. This typically as part of their very first prompt and often prior to reading even the most basic introductory material. This suggests a strong tendency to outsource the initial stage of information processing entirely to AI when managerial personnel are given new information and unrestricted access to AI is provided. Compared to the Semi-AI condition, where none of the participants uploaded either of the files.

This pattern points to a key managerial risk: when employees are given unrestricted access to AI, they may default to immediately externalizing newly received information without first engaging with it themselves. In practice, this can limit individual judgment, critical reading, and increases the likelihood that sensitive or strategically important data is shared with AI systems without proper consideration.

The main risk from the uploading of these files comes from a potential data breach and the impact this can have on the company rather than legal or ethical jeopardy. This risk is heightened depending on the way the company has set up the LLM and its training data

that the management staff is using. The article AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business, discusses the potential risk of training data being extracted by external users (Humphreys et al., 2024, p. 798). Combine this with the risk of a more general data breach at an AI centre and uploads in this vein could be extremely risky for any company. This risk is amplified by the potential over-trust employees place in the way AI companies protect their data (Humphreys et al., 2024, p. 800). Data from a recent study by the University of Melbourne and KPMG highlight this. Their findings showed that 48-49% of employees uploaded sensitive information into public AI tools (Gillespie et al., 2025, p. 75). Our results show a similar, if higher figure.

The risk profile of these files is slightly different from the last, the personnel files. The difference stems from two main concerns. The first being the already mentioned legality of uploading sensitive files, especially personnel files, into artificial intelligence. Depending on which information is included the upload would not be GDPR compliant (*Data Protection under GDPR*, 2025). A second risk is the way the information itself is used. AI tools for the management of workers are designated as high risk under the EU AI act. This designation comes with strict obligations. Such as the traceability of results and appropriate human oversight measures (*AI Act | Shaping Europe's Digital Future*, 2026).

The personnel files were uploaded by two CHR participants for both the Semi-AI condition and the Full AI condition (Full AI one and two, Semi AI two and three). Of these the most interesting are the full AI one (FAI1) and Semi-AI two (SAI2) groups, these were the only groups who made a, explicit decision to demote or fire an employee. The person that FAI 1 demoted was mentioned as a risk in a response to the CHR for that group starting in the first prompt, even though the participant did not specifically ask for this. Throughout the chat this person (Olivier Claeys) was brought up as the LLM's go to example of an employee who should be "avoided" without the participant directly asking for this (CHR Participant Full AI 1, 2026). There was never a discussion during the case between this participant and the other members of their group surrounding this decision (Degroote et al., 2026b). Due to this a definitive appreciation of the impact that AI had on the deliberations cannot be made, yet speculation is possible. It is highly likely, however, that the AI made a significant contribution to the decision to demote Olivier. The participant never challenged the LLM on this recommendation and copied its output when they asked it to generate the CHR's part of the slide deck (anonymous participants Full AI 1, 2026; CHR Participant Full AI 1, 2026). It is possible that the participant exercised the appropriate oversight, but due to absence of evidence regarding this it might put the company in legal jeopardy. The decision itself also rests in a clear ethical grey area.

The other CHR for SAI2 made their decision regarding the firing of several employees in an even more dubious way. This group decided to fire instead of demoting an employee. The impact of AI on this decision and the ethical implications of this was also more clearly discussed. Both of the individuals that were eventually fired "Lisa Martens" and "Arne Willems" were identified as not matching the compliance culture by the AI (CHR Participant Semi AI 2, 2026). The decision regarding what to do with Lisa was not discussed in the group, Arne however was discussed. The CHR said that they had achieved a breakthrough in a "maybe not so ethical way" (Degroote et al., 2026c, 01:00:40). After this the participant proceeded to read a summarized version of the response to their prompt to the rest of the group (CHR Participant Semi AI 2, 2026; Degroote et al., 2026c, 01:01:00). Although this shows that the participant relied to some degree on AI there is also evidence of the participants did not rely solely on the AI to make their decisions. Since, in an earlier prompt

AI recommended the participants that they should aim their retention efforts at the employees they ended up firing. Although AI did raise some potential issues concerning Arne. It also identified the employees based on a direct query from the CHR. The participants also filtered the recommendations AI provided, this can be gleaned from the fact that they did not promote the top recommendation of the AI. Instead they chose others, further down the line (anonymous participants Semi AI 2, 2026; CHR Participant Semi AI 2, 2026). The evidence provided by the chat logs, recordings and presentation thus suggest that the participants exercised oversight over the recommendations made by AI but were influenced by it.

What these two case studies suggest is that an AI model can impact the decision-making process in multiple ways. The first is the system being able to plant an idea that can be based on incomplete information, AI in the FAI1 condition for example only had access to the digital personnel files, whilst the hall files were never uploaded to it. In the second case the AI was directly involved in the decision-making process. Both eventualities risk being an ethical breach, apart from the fact that the AI can stealthily influence such decisions, due to similar reasons. The article *“AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business”* by Humphries et al. highlights some of the risks of utilising AI in business, most of which are applicable to this example. A first major ethical risk when making, especially this kind of decisions with AI is the risk of training data being poisoned. A model can only be as good as the data it was trained on. If this data is based on, for example, a data set with extremist undertones, it could result in biased outcomes (Humphreys et al., 2024, p. 797). A second risk would be overreliance on AI, which the article defines as people; *“Instead of combining critical thinking and their own insights into a problem along with an AI model, people frequently over-rely on the AI even if they would have made a better choice on their own”* (Humphreys et al., 2024, p. 799)

The article by Blackman highlights similar risks as the work by Humphreys, risks that the business game experiment suggest would become reality if AI were to be implemented without proper oversight. It highlights the risks of AI being trained on racist training data, which would result in inequitable outcomes (Blackman, 2022, p. 122). If this is combined with the reliance and trust the participants placed in the AI. They had access to, this would confirm the fears highlighted by Blackman. The data from the experiment does add credit to the idea of an ethics committee that might curtail some of the worst excesses that unregulated AI use might entail (Blackman, 2022, p. 123). However the risk remains that employees can use personal LLM's for work without properly notifying or bringing the company in the loop (Gillespie et al., 2025, p. 75). This complicates any governance structures that organizations try to implement. This is an area which requires more study, since the hybrid ethical threat stemming from both misuse by organizations themselves or their employees cannot be ignored.

4.4 Survey perception participants

The analysis of the post-simulation survey provides (see appendix E, Survey) critical contextual depth to the objective performance metrics of the E-Trax BV business case. By examining the self-reported data within an exploratory framework ($\alpha = 0.10$), several distinct patterns emerge regarding how the architecture of AI access influences managerial cognition and group deliberative processes. The averages of all answers are visible in the radar plot to the right.

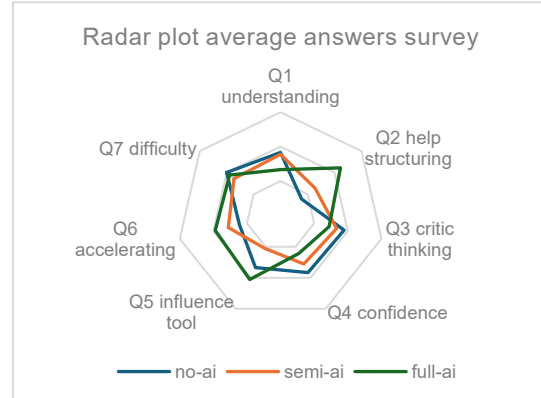


Figure 23: Radar diagram – Average answers survey

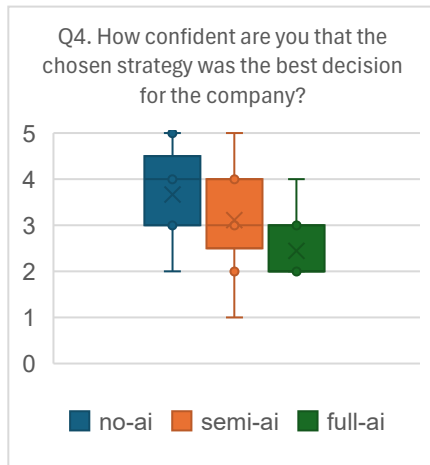


Figure 24 - Boxplot – Q4 Confidence

A primary finding of this study is the observed trend between the degree of AI-access and the participants' confidence in their chosen strategy (Q4: $H = 5.94, p = 0.051$). Post-hoc pairwise comparisons demonstrate that the No-AI group reported significantly higher confidence levels than the Full-AI group ($p = 0.017$). No statistically significant difference was found between the No-AI and Semi-AI conditions. These findings indicate an inverse relationship between unrestricted AI support and strategic confidence. Participants operating without AI assistance reported higher certainty in their final decisions than those with full AI access, suggesting that heavy reliance on external algorithms may diminish a decision-maker's sense of ownership over the final strategy.

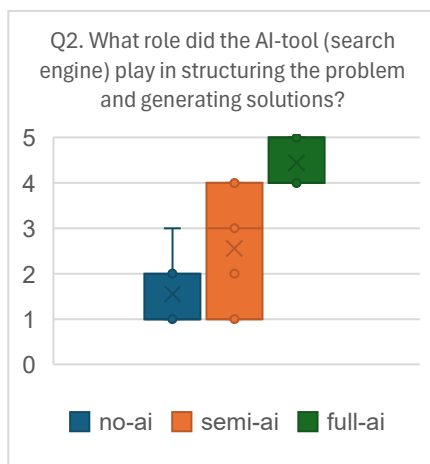


Figure 25 - Boxplot – Q2 Role of AI/search engine

The architecture of AI access fundamentally altered how teams utilized technology to structure the complex E-Trax crisis. Significant variance was observed regarding the role of AI in generating solutions and organizing the case data ($H = 15.88, p < 0.001$). The post-hoc analysis confirmed that the Full-AI group relied significantly more on AI for problem structuring than the No-AI group relied on the standard search engine ($p < 0.001$). It also confirmed that the Full-AI group perceived to have relied more on the AI-tool than the Semi-AI group ($p = 0.004$). The lack of a statistically significant difference between the No-AI and Semi-AI conditions ($p = 0.12$) indicates that restricted, shared access to AI does not increase teams' reliance on the tool for problem structuring.

compared to the baseline. Participants only exhibited a high reliance on algorithmic output for organizing case data and generating solutions when provided with unrestricted, individual access (Full-AI).

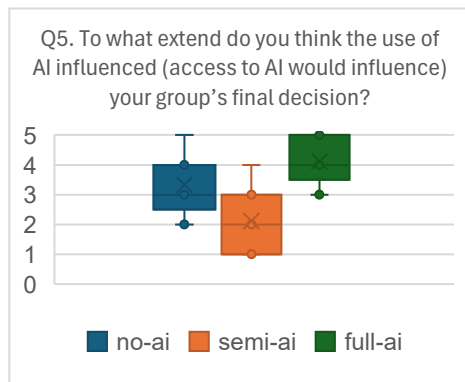


Figure 26 - Boxplot – Q5 Influence on decision

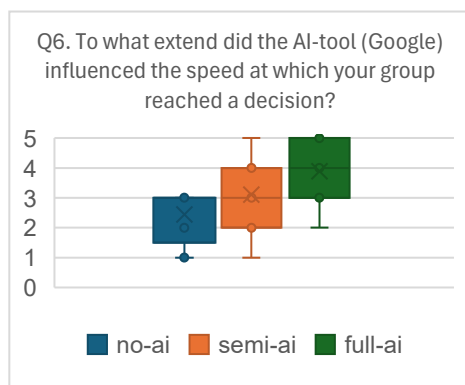


Figure 27 - Boxplot – Q6 Influence on speed

Regardless of the specific access architecture, the presence of AI tools altered the participants' perception of the decision-making process. The omnibus test identified significant differences in both the perceived influence of AI on the final decision ($H = 10.96, p = 0.004$) and the speed at which that decision was reached ($H = 6.22, p = 0.044$). It is important to note a methodological distinction between these metrics: for decision influence (Q5), the No-AI group was asked to estimate the hypothetical impact AI would have had, allowing for a comparison between expected and realized AI impact. For decision speed (Q6), the No-AI group evaluated their actual use of Google, enabling a direct comparison between the efficiency of a standard search engine versus AI tools. Post-hoc pairwise comparisons using a Bonferroni-corrected alpha ($\alpha = 0.033$) demonstrated two matters. First, the actual technological impact reported by the Full-AI group was significantly higher than the impact experienced by the Semi-AI group ($p = 0.002$). There was no statistically significant difference between the hypothetical expectations of the No-AI baseline and the realized experiences of either the Full-AI group ($p = 0.11$) or the Semi-AI group ($p = 0.034$,

borderline). Secondly, The Full-AI group reported that unrestricted AI access provided significantly higher perceived efficiency gains than the No-AI group experienced using Google ($p = 0.013$). The perceived speed gains reported by the Semi-AI group did not differ significantly from the No-AI group's experience with Google ($p = 0.233$), nor did the two AI-assisted groups differ significantly from each other ($p = 0.2$).

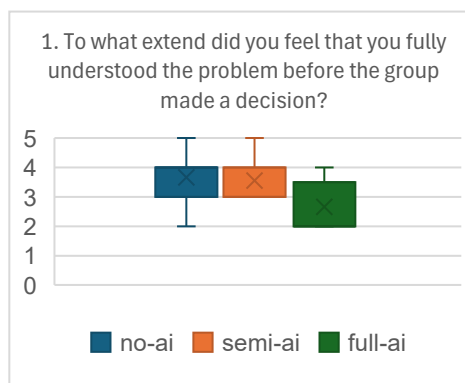


Figure 28 - Boxplot – Understanding

The Kruskal-Wallis test indicated variance in problem comprehension among the three groups ($H = 5.59, p = 0.061$). However, post-hoc pairwise comparisons evaluated against the Bonferroni-corrected threshold ($\alpha = 0.033$) revealed no statistically significant differences between the Full-AI group and either the Semi-AI ($p = 0.047$) or No-AI ($p = 0.042$). Furthermore, analysis of questions three and seven yielded no significant differences across the conditions.

To explore the underlying drivers of the observed group differences, a Spearman rank correlation analysis was conducted. This analysis identified a significant tension between technological assistance and cognitive mastery. A moderate negative relationship exists between the role of AI in structuring the problem and perceived understanding ($\rho = -0.57, p = 0.002$). There is a strong negative relationship between the role of AI in structuring the problem and strategic confidence ($\rho = -0.64, p < 0.001$). These results suggest that as the AI takes a more dominant role in organizing the case, individual managers feel less "ownership" of the problem, leading to a measurable decline in their subjective certainty. Furthermore, the data reveals a critical impact on group deliberation. Tool reliance for structuring was negatively associated with the degree of critical questioning within the group ($\rho = -0.54, p = 0.004$). This implies a potential "automation bias" where AI-generated structures are accepted with less scrutiny than human-generated ones. Conversely, a strong positive correlation was found between tool reliance and perceived decision speed ($\rho = 0.67, p < 0.001$), confirming that while AI integration may erode the depth of deliberation, it successfully facilitates a faster sense of progress.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Q1	1,00	-0,57	0,30	0,47	-0,42	-0,40	-0,11
Q2	-0,57	1,00	-0,54	-0,64	0,60	0,67	0,23
Q3	0,30	-0,54	1,00	0,39	-0,27	-0,36	-0,15
Q4	0,47	-0,64	0,39	1,00	-0,46	-0,21	-0,05
Q5	-0,42	0,60	-0,27	-0,46	1,00	0,37	0,36
Q6	-0,40	0,67	-0,36	-0,21	0,37	1,00	0,29
Q7	-0,11	0,23	-0,15	-0,05	0,36	0,29	1,00

Figure 29 – Spearman rank correlation

In summary, the subjective data indicates that while unrestricted AI integration outpaces standard search tools in perceived decision speed and maximizes the tool's strategic influence, it simultaneously erodes the individual manager's strategic confidence. Furthermore, as demonstrated by the variance in problem structuring, unrestricted access fosters a heavy reliance on algorithmic output. The negative correlations between this tool reliance and critical questioning confirm that high levels of AI integration create a measurable trade-off between perceived efficiency and analytical rigor.

Consequently, the Semi-AI condition occupies an effective middle ground. By limiting access to a shared terminal, this architecture prevents the structural over-reliance observed in the Full-AI group and maintains the strategic confidence levels characteristic of the No-AI baseline, while still capturing the perceived efficiency gains of the technology. These findings provide a rigorous foundation for evaluating how AI architecture can be optimized to support managerial group tasks, indicating that shared access is superior to individual access for balancing operational efficiency with high-stakes decision quality.

5 Discussion

The findings of this study invite for reflection at several levels: How they relate to the existing literature, what they add to it, where they challenge prevailing assumptions, and what their practical implications are for organisations navigating AI integration.

5.1 AI as amplifier, not equaliser

The most pervasive pattern across all four instruments is that AI access, particularly unrestricted individual access, does not produce uniform effects. It amplifies. Where groups possess the conceptual scaffolding to direct AI effectively, unrestricted access produces the best outcomes observed in this study. Where that scaffolding is absent, performance is indistinguishable from or worse than the No-AI baseline. This finding stands in productive tension with much of the productivity literature reviewed. The articles by Czarnitzki et al. and Gao & Feng document substantial average productivity gains from AI adoption at the firm level, but averages conceal distributional effects (Czarnitzki et al., 2023; Gao & Feng, 2023). Dell'Acqua et al. comes closest to anticipating the pattern observed here, noting that gains are concentrated among lower-skilled workers on well-defined tasks and characterising AI's frontier as "jagged" (Dell'Acqua et al., 2026). The present study extends this insight to the managerial domain: even within a homogeneous, high-skill student population working on the same constrained task, AI access introduced substantial performance variance rather than convergence. The implication is that the average productivity gain reported in macro-level studies may mask a polarisation effect at the individual and team level that is consequential for how organisations design and govern AI deployment.

Vendraminelli et al. provides the most direct conceptual parallel, demonstrating that AI-enabled horizontal expertise transfer has limits that are determined by the depth of domain knowledge the user brings to the interaction (Vendraminelli et al., 2025). The Full-AI polarisation effect observed in this study is consistent with this finding. Participants who could decompose the E-Trax crisis into precise, well-bounded queries and critically evaluate the AI's outputs produced analytically rigorous presentations. As illustrated most clearly by Full-AI 1. Those who uploaded their files wholesale and accepted AI-generated frameworks without interrogation did not. What this study adds is the process-level evidence for this mechanism. The IPA transcript and prompt analysis document the behavioural difference between these groups directly, showing that the divergence in outcomes is preceded by a divergence in how the technology is engaged with.

5.2 The paradox of access and the erosion of cognitive ownership

A second theme that runs across the instruments concerns the relationship between AI access and perceived cognitive ownership. The survey identified an inverse relationship between unrestricted support and strategic ownership: Full-AI participants reported the highest degree of AI structuring and the lowest strategic confidence. The paradox is driven by the significant erosion of subjective certainty despite access to the most powerful analytical support. This is counterintuitive from a purely instrumental perspective, but

consistent with the ironies of automation described by Simkute et al, who argue that AI integration can produce a cognitive shift away from active production toward evaluation tasks, reducing the sense of ownership over the final output (Simkute et al., 2024).

Simon's trichotomy of intelligence, design, and choice is instructive here (Simon, 1977). In the No AI condition, participants were forced to perform all three stages themselves. They had to identify what the problem was, generate possible solutions from the available data, and then choose among them. In the Full-AI condition, the intelligence and design stages were substantially outsourced to the AI, leaving participants primarily in the choice role, choosing between AI-generated options rather than constructing their own. This compression of the decision-making process may explain both the speed gains reported in the survey and the confidence deficit. When the reasoning process is externalised, participants may struggle to fully internalise or defend the outputs, producing a situation in which the decision is made but not owned. This interpretation is consistent with the observed patterns across the survey and IPA instruments but remains a theoretically grounded inference rather than a directly tested claim.

5.3 Cross-functional integration & the structural value of constraint

The Semi-AI findings introduce a nuance that is absent from most of the AI-productivity literature. The possibility that constrained access outperforms unconstrained access on dimensions that require collective reasoning. The IPA analysis showed that Semi-AI groups demonstrated significantly higher systemic organisational thinking than No-AI groups, and the slide deck analysis showed the most consistent and balanced performance profiles for this condition. The mechanism proposed here that the shared bottleneck of a single AI terminal forces verbal synthesis across role-based information silos, is consistent with Fayol's coordination function and with Mintzberg's emphasis on the interpersonal dimension of managerial work. In the Semi-AI condition, the AI functioned as an external consultant that the team had to brief collectively, requiring participants to articulate their role-specific knowledge to each other before it could be submitted. This enforced a degree of information sharing that the Full-AI architecture appears to have reduced. This claim was not statistically proved.

5.4 The ethical gap between deliberation and output

The ethical findings of this study deserve particular attention because they are both the most statistically robust and the most practically consequential. The gap between the IPA result, no significant difference in ethical concern codes across conditions, and the slide deck result, a highly significant difference in ethical consideration scores ($H = 13.00$, $p = 0.002$), points to a mechanism that is subtle but important. Ethical considerations were not absent from the Full-AI deliberations. They were raised at comparable rates to other conditions. What differed was whether they were carried through to the final output.

Two explanations for this gap are plausible and not mutually exclusive. The first is attentional. In the Full-AI condition, individual participants were simultaneously managing their own AI interaction and contributing to group deliberation. This dual demand may have created an attentional bottleneck in which ethical dimensions, which tend to be softer and

less immediately quantifiable than financial or operational considerations, were crowded out when the group synthesised its final presentation. The second is structural. Because each participant had their own AI-agent structuring their individual dossier, the integration of role perspectives into the final presentation may have been less dependent on verbal deliberation and more dependent on everyone's own AI-output.

The file upload analysis makes this concrete. In documented cases, Full-AI participants uploaded personnel files to AI-systems that had access only to partial information, received recommendations based on that incomplete data, and incorporated those recommendations into consequential HR decisions, including demotions and dismissals, without apparent critical review. This is a direct instantiation of the accountability gap described by Wieringa), in which delegation to algorithmic systems reduces human discretion, diffuses responsibility across actors (Wieringa, 2020, pp. 3-4). The regulatory implications flagged by Rodriguez and Yannakourou are directly relevant here. The EU AI Act's emphasis on human oversight in high-risk AI applications is not an abstract governance concern but a practical necessity for exactly the kind of individually mediated, consequential decision-making observed in this study (Rodriguez-de-las-Heras, 2025, pp. 7-55).

5.5 What the No-AI condition reveals

It is worth pausing on what the baseline condition shows, because it is easy to overlook in a study framed around AI-effects. No-AI participants reported the highest strategic confidence. Their slide deck scores formed the most geometrically consistent profile across dimensions without the pronounced peaks or troughs that characterised both AI conditions. The structural transcript analysis suggested descriptively higher conversational volume in the No-AI condition, though this was not statistically tested at condition level. What they could not do was process the data-intensive case materials as effectively as the best Full-AI group, and their analytical rigor scores reflected this limitation, though this difference did not reach statistical significance.

This profile, cognitively engaged, ethically present, analytically constrained, is a reasonable description of the pre-AI manager. It is not a description of incompetence. The risk, when framing AI-adoption as an unambiguous upgrade, is that the capabilities preserved by No-AI deliberation, critical engagement, ethical vigilance, cognitive ownership, are treated as inefficiencies to be eliminated rather than as properties to be preserved as AI is introduced.

5.6 Practical implications

Three practical implications follow from these findings. First, organisations should not treat AI-access as a binary variable. The architecture through which AI is accessed, individual or shared, constrained or unconstrained, shapes deliberation quality, ethical output, and outcome variance in ways that the simple presence or absence of AI does not predict. The design of AI deployment within managerial teams is a governance question, not merely a technology question.

Second, AI literacy training should be understood not as training in how to use AI-tools, but as training in how to delegate to them, specifically, in how to identify the stages of the decision-making process that require human judgment and to protect them from premature externalisation. The findings suggest that the participants who performed best under Full-AI access were not necessarily those who used AI most, but those who used it most selectively. Illustrated by the contrast between Full-AI 1 and the other Full-AI groups in the prompt analysis, through this interpretation cannot be formally tested at this sample size.

Third, the ethical risks associated with unrestricted AI-access in personnel and strategic decision-making contexts are not theoretical. This study documents two cases in which AI recommendations based on incomplete information directly influenced consequential HR decisions. While these represent only two documented cases and cannot be generalised, they illustrate a category of risk that needs systemic attention. As AI-access in organisations continues to expand, and as the 56% of employees using AI without employer knowledge cited in the introduction suggests it already has, the gap between actual AI use and governance frameworks designed for that use represents a live organisational risk (Gillespie et al., 2025, p. 75).

5.7 Limitations

Several limitations bound the conclusions of this study. The sample of nine groups provides insufficient statistical power for definitive causal inference, and the homogeneity of the participant population, master's students with limited professional experience, limits generalisability to practising managers, for whom domain knowledge, professional norms, and organisational accountability structures may substantially alter the observed dynamics.

Another recurring limitation in the quantitative analysis is the statistical discrepancy between the omnibus test results and the subsequent pairwise comparisons. Specifically, the Kruskal-Wallis tests frequently identified significant or borderline variance across the three conditions, but the post-hoc Mann-Whitney U tests frequently failed to identify statistically significant differences between specific pairs. This discrepancy stems from the interaction between the small sample size ($N = 27$ for all tests, except the slide deck analysis which contained $N = 54$) and the strict application of the Bonferroni correction.

The Kruskal-Wallis test evaluates all data points simultaneously, utilizing the maximum available statistical power of the dataset. Pairwise comparisons restrict the analysis to only 18 or 36 data points ($n = 9$ per group for all tests, except the slide deck which contained $n = 18$), which inherently reduces the test's statistical power and its capacity to detect an effect. The Bonferroni correction is applied to control the Type I error rate (the probability of detecting a false positive) when conducting multiple comparisons. By dividing the exploratory alpha ($\alpha = 0.10$) by the three necessary pairwise comparisons, the threshold for significance is strictly lowered to $\alpha = 0.033$. In small-sample research, enforcing this strict threshold severely inflates the risk of a Type II error (a false negative, or failing to detect a true effect). Consequently, the absence of statistical significance in the post-hoc tests cannot be definitively interpreted as the absence of a true behavioural effect. It is highly probable that the study simply lacked the statistical power required to satisfy the stringent Bonferroni threshold at the subgroup level. As a result, valid differences in human-AI collaboration may have been downgraded to non-significant trends due to the strictness of the statistical parameters rather than the reality of the observed data.

The use of Gemini Pro as the primary LLM, without system prompts, means the findings are specific to a particular tool configuration and may not generalise to systems with built-in ethical guardrails or structured prompting frameworks. Finally, the business game format, while providing experimental control, compresses dynamics that in real organisations unfold over days or weeks; time pressure and novelty effects may have influenced behaviour in ways that would not persist in routine managerial practice.

5.8 Further research

The conclusions of this paper are necessarily exploratory. A sample of nine groups does not permit causal inference, and the homogeneity of the participant population, KU Leuven master's students with limited professional experience, restricts generalisability to experienced managerial populations.

Future research should replicate this design with larger and more professionally experienced samples and examine whether explicit AI literacy training attenuates the polarisation effect observed in the Full-AI condition. The ethical risk associated with unrestricted file uploading and AI-mediated personnel decisions warrants particular attention, as it points to a category of organisational risk that current governance frameworks do not yet adequately address. More broadly, the question of how AI access should be architecturally designed for managerial teams, not whether AI should be used, but under what structural conditions its use supports rather than substitutes human judgment, remains open and, given the pace of adoption, urgent.

5.8.1 Architectural variations in AI-access

This study tested three conditions that differ in both the degree and the architecture of AI access. The Semi-AI condition combined throttled access with a shared terminal, making it impossible to disentangle the effect of access volume from the effect of the shared deliberation structure. Future research should introduce conditions that isolate these variables, for example, comparing individual access at low frequency against shared access at high frequency, to determine which architectural feature drives the systemic thinking and ethical output advantages observed in the Semi-AI condition. The introduction of structured prompting frameworks, system prompts with ethical guardrails, or mandatory reflection checkpoints within Full-AI conditions would also allow for a test of whether the ethical gap can be closed by design without reducing the analytical potential of unrestricted access.

6 Conclusion

This thesis set out to investigate how the architecture of human-AI access in managerial teams shapes ethical reflection, strategic decision quality and group deliberation in managerial contexts. With the central research question, “How do human-AI interactions influence ethical considerations, strategic decision-making, and planning processes in managerial group tasks” was examined through a quasi-experimental business game design in which nine groups of master's level participants solved a constrained strategic crisis under three distinct conditions: No-AI access, Semi-AI access, and Full-AI access. Evidence was gathered across four complementary instruments:

1. Rubric-based evaluation of the final strategic presentations
2. IPA-coded analysis of audio transcripts
3. Mapping of AI prompt and search history
4. Post-experiment perception survey.

This conclusion synthesises the findings by returning to the three hypotheses before drawing broader implications.

- H1. Cognitive delegation: Full-AI groups will score higher on analytical rigor but exhibit a lower frequency of critical questioning and explicit trade-off articulation compared to No-AI groups.
- H2. Cross-functional integration: Semi-AI groups will demonstrate more balanced cross-functional integration than both No-AI and Full-AI groups, as the necessity to coordinate a shared AI terminal reinforces verbal information exchange across role-based silos.
- H3. Ethical reflection: Ethical considerations will be articulated more explicitly in No-AI and Semi-AI conditions than in the Full-AI condition, where individual algorithmic delegation may bypass collective ethical reasoning.

6.1 H1: Cognitive delegation (partially supported)

H1 predicted that Full-AI groups would score higher on analytical rigor but exhibit lower frequencies of critical questioning and explicit trade-off articulation. The evidence is mixed but directionally consistent with this prediction.

On analytical rigor, the slide deck analysis yielded no statistically significant difference across conditions ($H = 4.43$, $p = 0.109$). The condition-level medians were identical across all three groups ($Mdn = 2.0$). However, the Full-AI condition produced the widest internal spread of any condition: one group achieved the maximum score of 4.0, while the other two scored 1.5 and 2.0, values comparable to the weakest No-AI outputs. This polarisation suggests that the capacity for analytical uplift exists under Full-AI access but is not systematically realised.

The process-level evidence more clearly supports the critical questioning component of H1. The IPA analysis confirmed a significant surge in direct task delegation in the Full-AI

condition, while fact checking dropped significantly compared to No-AI groups, and source scrutiny was significantly lower than in the Semi-AI condition. The prompt analysis reinforced this: Full-AI participants uploaded their files immediately and used AI across a broader functional range, including coordinating a function almost entirely absent in Semi-AI prompts. The survey confirmed this subjective dimension regarding certainty and reliance: Full-AI participants reported the highest degree of AI structuring and the lowest strategic confidence of all three conditions. However, it is important to reiterate a methodological limitation: the No-AI group's responses regarding AI influence were hypothetical estimations, whereas the AI-assisted groups reported on actual experiences. Consequently, these direct comparative survey findings should be interpreted with caution. H1 assessed as partially supported: the delegation and de-skilling effects are clear and consistent across instruments, but the analytical rigor uplift is neither uniform nor statistically confirmable at this sample size.

6.2 H2: Cross-functional integration (directionally supported)

H2 predicted that Semi-AI groups would demonstrate more balanced cross-functional integration than both other conditions, driven by the coordination demands of a Semi-AI terminal. The evidence provides directional support across several instruments, though not without complication.

The IPA transcript analysis found a significant improvement in systemic organisational thinking in the Semi-AI condition compared to No-AI groups, consistent with the hypothesis. Full-AI teams, by contrast, discussed role coordination significantly less frequently than both other conditions, indicating that unrestricted individual access reduced the need, or the habit, of explicit cross-functional deliberation. The slide deck analysis showed the most consistent performance profiles for Semi-AI across all five dimensions, without the within-condition variance that characterised Full-AI. The trade-off transparency scores for Semi-AI were the most stable of any condition (2.5, 2.0, 2.0), suggesting that the shared bottleneck encouraged more deliberate verbal reasoning.

However, the evidence also introduces complications that qualify this conclusion. The IPA analysis found that Semi-AI groups showed a significantly higher frequency of tunnel vision behaviour, premature commitment to a single strategy without exploring alternatives, compared to No-AI groups. Semi-AI access, while improving macro-level systemic thinking, may simultaneously reduce the diversity of options considered. Additionally, the CHR role remained structurally marginalised in trade-off articulation across all conditions, indicating that the imbalances inherent in cross-functional team composition were not resolved by any AI architecture. The survey did not distinguish Semi-AI from No-AI on most perception variables, and the prompt analysis showed that Semi-AI prompts were more directive and control-oriented than Full-AI prompts. H2 is assessed as directionally supported: shared access appears to sustain more coherent and balanced deliberation, but the bottleneck effect does not uniformly produce cross-functional integration across all dimensions of measurement.

Finally, a critical limitation of cross-functional integration emerged regarding role-based marginalization. Across all conditions, the human resources perspective (CHR) was structurally marginalized in terms of speaking turns and trade-off articulation. Importantly, AI access, whether shared or individual, failed to mitigate this baseline disparity,

demonstrating that technological architectures alone cannot resolve entrenched interpersonal team imbalances.

6.3 H3: Ethical reflection (supported)

H3 predicted that ethical considerations would be articulated more explicitly in No-AI and Semi-AI conditions than in the Full-AI condition. This is the most clearly and consistently supported hypothesis across all instruments.

The slide deck analysis produced the only statistically significant outcome finding: Ethical and Human Consideration differed significantly across conditions ($H = 13.00$, $p = 0.002$). Post-hoc Mann-Whitney U tests confirmed two statistically significant pairwise differences. Semi-AI scored significantly higher than both No-AI ($U = 75.5$, $p = 0.006$) and Full-AI ($U = 58.0$, $p = 0.001$). The difference between No-AI and Full-AI did not reach statistical significance ($p = 0.296$). Within the Full-AI condition, two of three groups scored at or near zero, meaning external raters judged ethical considerations to be either absent or minimally acknowledged in their final presentations.

The process instruments add critical nuance to this finding. The IPA transcript analysis found no statistically significant difference in explicit ethical concern codes across conditions, suggesting that ethical topics were raised verbally at broadly comparable rates during deliberation regardless of AI access. The gap therefore lies not in whether ethical considerations entered the conversation, but in whether they were carried through to the final output. The prompt and file analysis makes this dynamic concrete: Full-AI participants consistently uploaded sensitive personnel files at the outset without evident reflection on confidentiality. In addition to this, during two documented cases participants acted on AI-generated workforce recommendations based on incomplete information, directly influencing decisions to demote or dismiss named employees. In one case the AI recommended demoting an individual, without being prompted to, and snuck this decision into the final presentation. This gap between deliberation and output, visible across the transcript, prompt, and slide deck instruments simultaneously, represents the most robust cross-instrument finding of the study. This finding is consistent with the work of Wieringa et al. (Wieringa, 2020) H3 is assessed as supported.

6.4 The Architecture of access as the central variable

Beyond the three individual hypotheses, the results collectively point to a conclusion that transcends any single finding: it is not the presence of AI that determines decision quality, but the architecture through which it is accessed.

Unrestricted individual access, the Full-AI condition, produced the highest ceiling of performance observed in the study, but also the lowest floor. One group achieved maximum scores across nearly all dimensions while two others performed at or below the No-AI baseline. This polarisation is not incidental. It reflects a structural property of an architecture that places the full burden of AI direction on the individual: whether that individual can formulate effective prompts, critically evaluate outputs, and synthesise AI-generated content with role-specific and cross-functional knowledge determines the result

in part. This is consistent with the jagged frontier described by the Dell'Acqua (Dell'Acqua et al., 2026). The practical implication is consequential: organisations that deploy unrestricted individual AI access without investing in the underlying human capabilities to direct it cannot reliably expect an improvement in decision quality and may introduce greater variance and ethical risk than they resolve.

The Semi-AI condition produced neither the highs nor the lows of Full-AI. Its defining characteristic is consistency: coherent strategies, stable scores across presentations, and a level of systemic thinking that exceeded the No-AI baseline. The bottleneck of a shared terminal appears to function as a structuring constraint that forces verbal synthesis across role-based information silos, sustaining cross-functional deliberation rather than replacing it. This does not make shared access uniformly superior, the tunnel vision effect and the absence of analytical extremes suggest its own limitations, but it does make it more predictable and more protective of the deliberative process. It is recommended that organizations employing shared AI access, must implement active facilitation to prevent teams from prematurely committing to the first plausible, AI-supported strategy.

The No-AI baseline confirms that unaided human deliberation, while limited in analytical reach on data-intensive tasks, preserves a level of cognitive ownership and ethical awareness that neither AI condition fully replicates. Participants in this condition reported the highest problem understanding and the highest strategic confidence, suggesting that the absence of external structuring forces individuals to engage more deeply with the material themselves. That this depth does not translate into systematically better outcomes underscores the genuine trade-off at the centre of this research question.

7 Bibliography

7.1 Sources

Anonymous participants Full AI 1., Degroote, R., De Ryck, L., & Devoldere, T. (2026b, maart 16). *Transcription Full AI 1* [Audio recording].

Anonymous participants Semi AI 2., Degroote, R., De Ryck, L., & Devoldere, T. (2026c, april 1). *Transcription Semi AI 2* [Audio recording].

Anonymous participants Full AI 1. (2026, maart 16). *Appendix F, Presentation Full AI 1* (T. Devoldere, L. De Ryck, & R. Degroote, Red.).

Anonymous participants Semi AI 2. (2026, april 1). *Appendix F, Presentation Semi AI 2* (T. Devoldere, L. De Ryck, & R. Degroote, Red.).

CHR Participant Full AI 1. (2026, maart 16). *Appendix B, AI chat history CHR participant Full AI 1* (T. Devoldere, L. De Ryck, & R. Degroote, Red.).

CHR Participant Semi AI 2. (2026, april 1). *Appendix B, AI chat history CHR participant Semi AI 2* (T. Devoldere, L. De Ryck, & R. Degroote, Red.).

Degroote, R., De Ryck, L., & Devoldere, T. (2026a). *Appendix A, E-Trax Business case files* [Business case].

7.2 References

- AI Act | Shaping Europe's digital future*. (2026, mei 8). <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- Anwar, M., Hani, I. B., & Hujran, O. (2026). Navigating uncertainty: The role of CEO competencies in strategic decision-making for digitalisation in family firms. *Sustainable Technology and Entrepreneurship*, 5(1), 100119.
<https://doi.org/10.1016/j.stae.2025.100119>
- BaniHani, I., Alawadi, S., & Elmrayyan, N. (2024). AI and the decision-making process: A literature review in healthcare, financial, and technology sectors. *Journal of Decision Systems*, 33(sup1), 389-399. <https://doi.org/10.1080/12460125.2024.2349425>
- Blackman, R. (2022). Why you need an AI ethics committee. *Harvard Business Review*, 07, 118-125.
- Calvino, F., Reijerink, J., & Samek, L. (2025). *The effects of generative AI on productivity, innovation and entrepreneurship* (39ste dr., OECD Artificial Intelligence Papers) [OECD Artificial Intelligence Papers]. OECD. <https://doi.org/10.1787/b21df222-en>
- Czarnitzki, D., Fernández, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, 211, 188-205.
<https://doi.org/10.1016/j.jebo.2023.05.008>

Data protection under GDPR. (2025, March 3). Your Europe.

https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm

Dell’Acqua, F., McFowland, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Kraye, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. *Organization Science*, 37(2), 403-423. <https://doi.org/10.1287/orsc.2025.21838>

Fayol, H., Urwick, L. F., & Fayol, H. (1971). *General and industrial Management* (Repr). Pitman.

Frederix, R., Hendriks, D., & Van Biesen, M. (z.d.). *Top team bedrijfssimulatie*.

Gao, X., & Feng, H. (2023). AI-Driven Productivity Gains: Artificial Intelligence and Firm Productivity. *Sustainability*, 15(11), 8934. <https://doi.org/10.3390/su15118934>

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hasted, G. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025* (p. 4974511 Bytes). The University of Melbourne. <https://doi.org/10.26188/28822919>

Gonzalez, C., & Heidari, H. (2025). A Cognitive Approach to Human–AI Complementarity in Dynamic Decision-Making. *Nature Reviews Psychology*, 4(12), 808-822. <https://doi.org/10.1038/s44159-025-00499-x>

- Gruda, D., & Aeon, B. (2025a). Seven Myths about AI and Productivity: What the Evidence Really Says. *California Management Review Insights*. <https://cmr.berkeley.edu/2025/10/seven-myths-about-ai-and-productivity-what-the-evidence-really-says/>
- Gruda, D., & Aeon, B. (2025b). Seven Myths about AI and Productivity: What the Evidence Really Says. *California Management Review Insights*. <https://cmr.berkeley.edu/2025/10/seven-myths-about-ai-and-productivity-what-the-evidence-really-says/>
- Humphreys, D., Koay, A., Desmond, D., & Mealy, E. (2024). AI hype as a cyber security risk: The moral responsibility of implementing generative AI in business. *AI and Ethics*, 4(3), 791-804. <https://doi.org/10.1007/s43681-024-00443-4>
- Katz, R. (1974). Skills of an effective administrator. *Harvard Business Review*, 52, 90-102.
- Mckinsey & Company. (2025, december 5). *Adoption of artificial intelligence among organizations worldwide from 2017 to 2025, by type [Graph]*. Mckinsey & Company. Statista. <https://www.statista.com/statistics/1545783/ai-adoption-among-organizations-worldwide/>
- Metsola, R., Fiala, P., & Von Der Leyden, U. (2022, december 15). *European Declaration on Digital Rights and Principles for the Digital Decade*. <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles>
- Mintzberg, H. (1973). *The nature of managerial work*. Harper & Row.

- Nyenno, I., Truba, V., & Tokarchuk, L. (2023). Managerial Future of the Artificial Intelligence. *Virtual Economics*, 6(2), 72-88.
[https://doi.org/10.34021/ve.2023.06.02\(5\)](https://doi.org/10.34021/ve.2023.06.02(5))
- Rodriguez-de-las-Heras, T. (2025). The AI liability Puzzle: Rethinking Defective Product liability for AI. In T.-E. Synodinu, P. Jougoux, C. Markou, & T. Prastitou-Merdi (Red.), *EU digital law in the AI era* (pp. 7-33). Springer. <https://doi.org/10.1007/978-3-031-96743-6>
- Schmidpeter, R. (with Altenburger, R.). (2023). *Responsible Artificial Intelligence: Challenges for Sustainable Management*. Springer.
- Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A. E., Sellen, A., & Rintel, S. (2024). *Ironies of Generative AI: Understanding and mitigating productivity loss in human-AI interactions* (Versie 1). arXiv. <https://doi.org/10.48550/ARXIV.2402.11364>
- Simon, H. A. (1977). *The new science of management decision* (rev. ed). Prentice-Hall.
- Synodinu, T.-E., Jougoux, P., Markou, C., & Prastitou-Merdi, T. (Red.). (2025). *EU digital law in the AI era*. Springer.
<https://doi.org/10.1007/978-3-031-96743-6>
- Vendraminelli, L., Dossantos Disorbo, M., Hildebrandt, A., Mc Fowland III, E., Karunakaran, A., & Bojinov, I. (2025, september 8). *The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders* [Working Paper]. Harvard Business School.

https://www.hbs.edu/ris/Publication%20Files/26-011_04dcb593-c32b-4e4e-80fc-b51030cf8a12.pdf

Wieringa, M. (2020). What to account for when accounting for algorithms:

A systematic literature review on algorithmic accountability.

Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 1-18. <https://doi.org/10.1145/3351095.3372833>

Yannakourou, S. (2025). Regulating Algorithmic Management of work at EU Level: The adequacy of the platform work directive and the AI act. In T.-E. Synodinu, P. Jougoux, C. Markou, & T. Prastitou-Merdi (Red.), *EU digital law in the AI era* (pp. 33-55). Springer.

<https://doi.org/10.1007/978-3-031-96743-6>

8 Attached files

8.1 List of tables

Table 1: Slide deck analysis – Scoring Framework . 1	Fout! Bladwijzer niet gedefinieerd.
Table 2: Deductive Thematic Analysis - IPA Framework.....	16
Table 3: Median scores per presentation per dimension (scale 0-4).....	20
Table 4: Mean scores per presentation per dimension (scale 0-4).....	21
Table 5: Summary of hypothesis evaluation based on slide deck analysis	26
Table 6: Tabulation of management functions for semi ai and full AI conditions.....	32

8.2 List of Figures

Figure 1: Radar diagram - Average scores.....	21
Figure 2: Boxplot – Strategic Coherence	22
Figure 3: Boxplot – Analytical Rigor	23
Figure 4: Boxplot – Feasibility and constraint alignment	23
Figure 5: Boxplot –Trade-off Transparency	24
Figure 6: Boxplot –Ethical & Human Consideration.....	25
Figure 7: Boxplot- Number of conversations (1)	26
Figure 8: Boxplot – Number of conversations (2)	27
Figure 9: : Boxplot – Number of conversations (3)	27
Figure 10: Boxplot- Number of conversations (4)	27
Figure 11: Boxplot – AI 1: direct task delegation	28
Figure 12: Boxplot- AI2: AI as authority appeal	28
Figure 13: Boxplot – AI 4: Ai scepticism	28

Figure 14: Boxplot – RA 1: Fact checking.....	28
Figure 15: Boxplot – RA 2 Source demand	28
Figure 16: Boxplot – SP1: Role assignment	29
Figure 17: Boxplot – PS 2: Systemic organizational thinking	29
Figure 18: Boxplot – TV 1: Single solution lock-in	29
Figure 19: Boxplot – SD 3: Explicit trade-off articulation	30
Figure 20: Boxplot – Human impact mention	30
Figure 21: boxplot-Classification prompt or web search by type	31
Figure 22: boxplot-Classification prompt type, only AI prompts.....	32
Figure 23: Radar diagram – Average answers survey	36
Figure 24: Boxplot – Q4 Confidence	36
Figure 25: Boxplot – Q2 Role of AI/search engine	36
Figure 26: Boxplot – Q2 Role of AI/search engine	37
Figure 27: Boxplot – Q6 Influence on speed.....	37
Figure 28: Boxplot – Understanding	37
Figure 29: Spearman rank correlation	38

8.3 Appendix A, E-Trax the business case

Attached in the upload zone there are two files containing all documents the participants had access to whilst solving the business case. One document is an excel file, the other is a word document containing an amalgamation of all text documents.

8.3.1 Appendix A1, Business case – Online documents – CFO CTO CHR – spreadsheets

8.3.2 Appendix A2, Business case – Offline documents – CFO CTO CHR

8.4 Appendices B through E; Excel files with the results and the necessary Calculations

8.4.1 Appendix B, Prompt and search history

8.4.2 Appendix C, Slide deck analysis

8.4.3 Appendix D Survey

8.4.4 Appendix E Transcriptions

8.5 Appendix F Turn-around pitches participants

FACULTY OF BUSINESS AND ECONOMICS

Naamsestraat 69 bus 3500

3000 LEUVEN, België

feb.leuven@kuleuven.be

www.feb.kuleuven.be



LID VAN **ASSOCIATIE
KU LEUVEN**