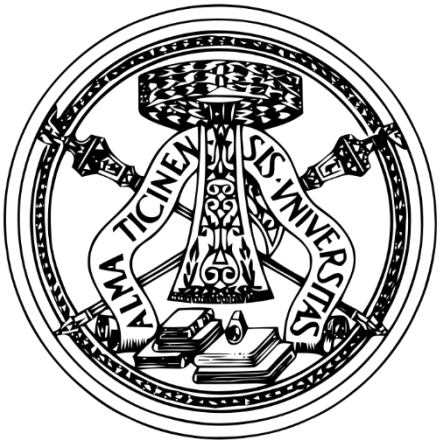


University of Pavia & University of Bergamo

PhD in Applied Economics and Management

Artificial Intelligence In Healthcare Organizations: Understanding The Multilevel Roots Of Organizational Implementation

PhD Candidate: Luigi M. Preti

Supervisor: Professor Amelia Compagni

Tutor: Professor Giovanna Magnani

ABSTRACT

Artificial Intelligence (AI) is expected to transform our societies due to its transformative, general-purpose nature. However, the implementation of AI technologies within organizational processes takes place at a path that is much slower than the rapid advancements in technology and research. Despite their large potential to improve health outcomes and operational efficiency, healthcare organizations as well suffer from a fragmented and uneven adoption and implementation of AI technologies in organizational routines. The translation of technical capabilities into routine use is hindered by the distinctive institutional, organizational, and professional features of these organizations. This dissertation contributes to the management and organization literature by investigating the multilevel roots of AI implementation in professional and knowledge-intensive organizations. It pursues a twofold objective: (i) identifying how management and organization research has conceptualized the adoption and utilization of AI and (ii) examining how organizational and professional dynamics shape the integration of AI technologies in organizational routines. First, the dissertation develops a semi-systematic review of 70 top-tier articles, highlighting the sociotechnical nature and the

interdependence of system, strategic, organizational, and individual dynamics in AI adoption and utilization. The review provides the conceptual ground to develop the empirical contribution of the dissertation. The research then explores how professional and research-oriented organizations manage AI innovation through a multiple case study of four large research and teaching hospitals in Italy. The study identifies four core tensions – organizing AI exploration, navigating the make-buy-ally continuum, managing internal demand for AI innovation, and balancing internal and external competencies – showing how organizations enact multilevel ambidextrous configurations to govern AI innovation. The second empirical contribution follows a large-scale empirical analysis of 1.5 million AI-based treatment recommendations in nephrology and provides evidence on how physicians engage with algorithmic advice. The findings reveal that professional reliance is strongly anchored in physician's norms, habits, and routines and highlight a tension between algorithmic and professional rationalities. Overall, the dissertation develops a multilevel account of AI implementation in healthcare as an industry dominated by professional, knowledge-intensive, and research-oriented organizations. It extends the literature on technology and innovation management by illustrating how knowledge arrangements, organizational capabilities, and professional logics jointly shape AI adoption and utilization. In doing so, it outlines implications for policy and management and future research directions on how organizations can navigate the opportunities and challenges of AI-driven organizational transformation.

INTRODUCTION

Over the past decade, artificial intelligence (AI) has moved from the domain of computer science into the core of social, political, organizational, and managerial debates. Two recent waves have been particularly consequential. The first unfolded in early 2010s, when the success of AlexNet – a deep convolutional neural network (CNN) that achieved record-breaking performance in ImageNet Large Scale Visual Recognition Challenge (Krizhevsky et al., 2017) – demonstrated the potential of deep learning (DL) architectures to dramatically improve performance on complex perceptual tasks. This breakthrough, enabled by advances in computing power through GPUs and the availability of large open source labeled datasets such as ImageNet (LeCun et al., 2015), catalyzed a surge of interest in neural networks, computer vision, and image recognition. A second wave emerged from late 2022 with the public release of OpenAI's generative pre-trained transformer (ChatGPT). Within two months, the application reached more than 100 million users, becoming the fastest-growing digital service in history and symbolizing the "AI boom". Generative AI differs from earlier ML and DL applications by being multimodal and multipurpose, enabling organizations and individuals to deploy a single

model across a broad set of tasks, ranging from text generation to image creation and speech synthesis. This development has also reshaped the vocabulary of the field, with a growing distinction between "traditional AI" – typically associated with narrow, task-specific ML/DL applications – and "generative AI", increasingly described as a transformative, general-purpose technology (Dwivedi et al., 2023; Finn & Downie, 2025; Gama & Magistretti, 2025). More recently, scholars and practitioner have pointed to the emergence of "agentic AI", systems that exhibit their capacity to act independently and purposefully not only by generating content but can also accomplishing a specific goal by taking autonomous actions and make decisions (Stryker, 2025). The development of agentic capabilities provides a different conceptual perspective on AI, interpreting it as a new form of *agency* rather than *intelligence*. This avoids the pitfalls of anthropocentric comparisons between AI artefacts and human capabilities (Floridi, 2025).

These advances fueled the expectations about AI's potential to boost individual and business productivity, enhance efficiency and accelerate the creation of new knowledge (Brynjolfsson & McAfee, 2014; Brynjolfsson & Mitchell, 2017). At the same time, they raise critical questions about risks and limitations. On the economic and political side, debates highlight potential disruptions to labor markets, professional work, the distribution between labor and capital, and global competition for dominance in the AI value chain (Johnson & Acemoglu, 2023; Aresu, 2024). On the social side, scholars warn about ethical risks, cognitive and behavioral transformation among individuals, and the broader consequences of decision-making, fairness, and human agency (Benbya et al., 2020; Dwivedi et al., 2021). These opportunities and challenges underscore that AI is not simply a technical innovation but implies a sociotechnical transformation with profound implications for organizations and societies alike.

Healthcare has long been among the most promising domains for AI. Advancements in DL-driven image and pattern recognition have shown substantial potential for improving clinical outcomes and operational efficiency (Topol, 2019). Regulatory authorities have begun permitting the use of AI tools in clinical practice. The FDA currently lists more than 1,200 AI-enabled medical devices cleared for routinary use in the US (FDA, 2025). Radiology remains the leading specialty, confirming how DL-driven image and pattern recognition techniques still represent the most promising domain in the healthcare domain. Recent advances in generative AI in many tasks have equaled or surpassed human clinician performance, both through generalist and specialist models (Hayat et al., 2025; McDuff et al., 2025; Singhal et al., 2025; Takita et al., 2025). Even beyond clinical diagnostic or therapeutic tasks, non-clinical areas are seeing fast growth. Domains like documentation filling, administrative burden reduction, report summarization, and patient communication are rapidly adopting GenAI tools where regulatory constraints are less strict (Bhuyan et al., 2025).

Despite rapid and promising advancements in both traditional and generative AI, the translation of technical capabilities into routine healthcare practice remains slow and uneven. A growing body of evidence points to an implementation gap, where validated applications and use cases struggle to achieve integration at scale. This gap is shaped by distinctive challenges at multiple levels (Chomutare et al., 2022; Preti et al., 2024). At organization level, healthcare providers and systems face financial and infrastructural constraints, including governance arrangements and the integration with existing IT systems and complex workflows. These barriers are further compounded by the highly regulated, professionalized, and institutionalized nature of healthcare, which makes innovation trajectories distinct from those in other industries (Barlow, 2017). At the

professional level, healthcare providers require AI outputs to be clinically meaningful and compatible with established diagnostic and therapeutic logics. Yet the opacity of many models – particularly DL systems – clashes with professional logics that emphasize interpretability, explainability, and accountability (Anthony, 2021; Lebovitz et al., 2022). Such barriers contribute to professional resistance and aversion, even when systems perform at or above human benchmarks (Faulconbridge et al., 2023). Because of their ever-growing complexity, the integration of AI systems into healthcare processes is filled with a variety of challenges that need to be addressed in a comprehensive way by various stakeholders (Petersson et al., 2022).

Building on the recent technological waves and acknowledging the distinctive multilevel features of healthcare, this dissertation addresses the overarching research question: *how can healthcare organizations deal with AI innovation and implementation of AI tools in routinary use?* The question is framed on a multilevel view of AI implementation in healthcare organizations. The idea of approaching innovation challenges through multiple levels of analysis is well established in healthcare research, where the field of implementation science and research has long examined barriers, facilitators, and strategies that operate at professional, organization, and system level (Peters et al., 2013; Nilsen, 2015). This dissertation builds on the same recognition of multilevel complexity but does so through the theoretical lens of organization and management research. By adopting this perspective, this thesis aims to contribute to not only the healthcare implementation literature but also more broadly to the field of AI innovation and its organizational implications. This general inquiry is articulated in three distinct research projects, each of which forms a distinctive dissertation chapter.

Chapter 1 presents a semi-systematic conceptual review of 70 articles published in top-tier journals between 2013 and 2024. The review adopts a problematization approach to synthetize how management and organization scholarship has theorized AI adoption and utilization. Eight thematic streams are identified, spanning from the tension between automation and augmentation, the recognition of AI's agentic role, dynamics of trust and aversion, and sociotechnical conditions that shape organizational adoption. This chapter lays the conceptual groundwork for investigating AI in healthcare organizations and draws the basis for the subsequent empirical chapters.

Chapter 2 examines how AI innovation in practice through a multiple case study of four large research-active healthcare organizations in Italy. Using the ambidexterity lens, the analysis uncovers four core tensions and the corresponding mechanisms the organizations arrange to address them: how AI exploration is legitimized and organized, how organizations navigate the make-buy-ally continuum, how internal demand for innovation is managed, and how internal and external competencies are balanced. The study contributes to ambidexterity and AI innovation management by showing how exploration and exploitation are organized in multilevel, configurational ways within professional and highly regulated contexts. It also highlights the emergence of hybrid roles, governance mechanisms, and relational capabilities that extend organizational boundaries of AI innovation.

Chapter 3 focuses on professional level and investigates physician's acceptance of AI-based recommendations in clinical decision-making. Drawing on a unique dataset of over 1.5 million treatment recommendations generated by a clinical decision-support system in nephrology care, the study analyzes how physicians across ten countries integrate or resist algorithmic advice. The findings show that alignment with clinical guidelines and

prior treatment behavior are the strongest and more consistent drivers of acceptance, while domain experience moderates physician reliance on AI. This chapter extends the literature on algorithm aversion by providing rare large-scale evidence from a real-world setting, thereby clarifying how professional norms and contextual anchors shape expert engagement with AI.

These three studies build a multilevel account of AI implementation in healthcare organizations, connecting conceptual, organizational, and professional perspectives. The concluding chapter draws on the insights accumulated across the dissertation to outline a set of implications for policy and managerial practice.

CHAPTER 1

UNRAVELLING AI IN ORGANIZATIONS: A REVIEW OF CONCEPTUAL PERSPECTIVES AND RESEARCH DIRECTIONS

1. INTRODUCTION

After several decades of intermitting interest, AI has returned to the core interest of management research and practice, fueled by the growing potential that this technology entails for transforming organizations, economic systems, and societies. Yet AI remains conceptually fluid. Scholars often describe it as the ability of a machine to perform cognitive functions commonly associated with human intelligence (Rai et al., 2019), while practitioners and regulatory agencies alternatively frame it as a comprehensive label for sequential waves of computer science innovation (Kaplan & Haenlein, 2019). Organizations now employ AI to streamline processes, develop product or services, and support – or sometimes replace – managerial decision-making. Current debate emphasizes how humans and AI can collaborate in tasks, routines, and decision processes. This collaboration promises faster and potentially more accurate decisions, yet raises several concerns over bias, fairness, transparency, explainability, as well as threats to the

identity of knowledge and professional workers involved in decision-making (Benbya et al., 2020; Shrestha et al., 2019). All this considerations contribute to make the organizational integration of AI not just a technical issue but a deeply sociotechnical and multidisciplinary challenge (Dwivedi et al., 2021)

Recent reviews illuminate specific facets of this agenda (e.g., Bankins et al., 2024; Gama & Magistretti, 2025; Matthews et al., 2025). Although valuable, these focused syntheses do not offer an integrated view across theoretical traditions and approaches. Responding to this gap, the present work asks *how management and organizational scholarship conceptualized the adoption and utilization of contemporary AI in organizations?*

To address this question, the study undertakes a semi-systematic, conceptual review based on a problematization approach (Petticrew & Roberts, 2006; Alvesson & Sandberg, 2011; Paré et al., 2015). A broad literature search was conducted to retrieve articles; after an inductive coding procedure eight thematic streams were produced.

The review reveals several cross-cutting tensions and emerging consensus points in how management and organizational literature conceptualizes AI. A central shift is the rethinking of the classic opposition between automation and augmentation. Across levels of analysis, scholars increasingly portray these not as competing logics, but as intertwined mechanisms that jointly shape the transformation of work (Coombs et al., 2020; Grønsund & Aanestad, 2020; Raisch & Krakowski, 2021). Closely linked to this is the recognition of AI's agentic role within organizations. Many studies move beyond framing AI as a neutral or purely instrumental tool and instead explore how intentionality is distributed across human and artificial actors. Concepts such as conjoined agency and hybrid problem-solving help articulate how agency, expertise, and responsibility are negotiated

in AI-mediated contexts (Maragno et al., 2023; Murray et al., 2021; Raisch & Fomina, 2024). A third line of reflection concerns the socio-cognitive conditions under which humans accept, resist, or adapt to AI systems. Rather than treating trust or aversion as static attitudes, the literature shows how they are shaped by task characteristics, system design, user expectations, and organizational context (Bankuoru Egala & Liang, 2023; Glikson & Woolley, 2020; Sturm et al., 2023). Another key insight is the embedded and multi-level nature of AI adoption. Organizational engagement with AI cannot be reduced to discrete implementation events; rather, it unfolds over time, shaped by internal capabilities, professional logics, external networks, and institutional conditions (Dahlke et al., 2024; Goto, 2023). The literature emphasizes the importance of sociotechnical alignment – between technical systems and organizational structures, between developers and users, and between emerging roles and traditional hierarchies (Faulconbridge et al., 2023; Lebovitz et al., 2021). Finally, the review surfaces two persistent areas of tension. First, ethical concerns – bias, fairness, accountability – cannot be resolved through technical optimization alone, but demand organizational responses that are normative, structural, and context-sensitive (Alon-Barkat & Busuioc, 2023; Kordzadeh & Ghasemaghaei, 2022). Second, while AI can support organizational learning and exploratory capacity, it also introduces risks of rigidity and myopia if its integration narrows experiential variety or overrides human judgment (Balasubramanian et al., 2022; Sturm et al., 2021).

While this review is not sector-specific, it is deliberately designed to inform subsequent research on AI implementation in healthcare organizations, which is the empirical and theoretical focus of the broader PhD project. The review thus serves a dual purpose: first, to map and synthesize how AI is being theorized across management and organizational

research; second, to lay the conceptual foundations for investigating how such insights apply to the specific challenges, dynamics, and institutional conditions of the healthcare sector.

The paper proceeds as follows. Section 2 describes the review methodology; Section 3 presents descriptive insights, the definitional boundaries, and the eight thematic areas; Section 4 discusses overarching findings, contributions, and limitations, and outlines avenues for future research, with a particular focus on healthcare organizations.

2. METHODOLOGY

The review seeks to identify and understand relevant research streams that have implications for the broad topic under analysis – the adoption and utilization of AI technologies in organizations – and how the topic has developed across different research traditions. Hence, to capture the diversity of theoretical approaches and contributions, I adopted a semi-systematic review approach combining systematic literature search with conceptually categorizing the different thematic streams that emerged from recent empirical and theoretical contributions to the topic (Petticrew & Roberts, 2006; Paré et al., 2015; Snyder, 2019).

I conducted a four-step process systematic literature search by collecting and reviewing studies that provide recent theoretical contributions about the macro-domain of artificial intelligence in management and organization studies. The process is outlined in Figure A1.1 as a PRISMA flowchart. First, I conducted a search on the Web of Science (WoS) database, through a list of keywords that would have reasonably captured the related topic. The keywords I used are as follows: «artificial intelligence", "machine learning", "deep learning", as well as related abbreviations ("AI", "ML", "DL"). I searched on titles,

abstracts, and keywords. Assuming a problematization approach (Alvesson & Sandberg, 2011), a broad initial search was conducted to make sure all relevant studies and framings were included in the analysis. Filters were applied to source category ("Business", "Management", "Operation Research Management Science", and "Social Science Interdisciplinary"), document type ("Article", "Review Article", and "Early Access"), and period (from 01/2013 to date of the web search, 02/2024), resulting in 11,716 articles. Second, articles were then filtered based on a selected set of journal outlets. The scope of the review was limited to 25 top-tier journals ranked 4 or 4* in the fields of general management, innovation management, organization and management science, organization studies, strategy, and public sector, according to the Chartered Association of Business Schools Academic Journal Guide (AJG) 2021. The preference for top-tier journals is due to their expectation that authors consistently make outstanding theoretical contributions in both theoretical and empirical research and is consistent with previously published review on the topic (Webster & Watson, 2002; Bankins et al., 2024; Matthews et al., 2025). I purposefully excluded journals on functional areas of management (e.g., operations and human resource management, marketing, accounting, etc.) and non-managerial fields (e.g. economics and finance, psychology), to focus on journals that engage in organization-level theorizing and broader managerial implications. I also decided to maintain journals specialized in applied settings like public management, in order to consider potential specificities of public sector organizations. Table A1.1 details the list of selected journals. Third, after filtering selected journals, 336 records were retrieved from WoS. 173 additional records were retrieved through manual search on the journals databases. The search was carried out in February 2024. A total of 509 titles and abstracts were screened. Exclusion criteria were applied to remove: (i) non-research or

non-peer-reviewed articles such as proceedings, commentaries, and editorials, (ii) articles that used AI as a method of analysis or (iii) solely cited but did not focus on AI, and articles (iv) that did not examined AI within organizational, managerial, or professional contexts. For instance, we excluded papers that focused solely on technical modeling or algorithmic performance, analyzed individual-level reactions, or solely addressed ethical or policy debates. After the screening process, 418 articles were excluded, leaving 86 articles for further screening of relevance. Fourth, each remaining record was evaluated for its theoretical contribution to the topic. Only papers that made salient and generalizable contributions for management and organization studies at micro-, meso-, and macro-level were deemed relevant. Therefore, for example, research studies that solely focus on the impact of certain AI characteristics on customer experience or those that only assess the effect of AI artifacts on workforce productivity were excluded. Additionally, non-research articles that were not identified during the first screening step were also excluded, as well as methodological articles. The entire process resulted in 65 relevant articles. Seven additional records were included after backward reference snowballing of included articles, by choosing articles published in the list of selected journals but not matching with the selected keywords, by referring for example to "algorithms", "epistemic technologies" and "agentic IT artifacts". All these contributions were deemed relevant to address the research question and hence were included in the final sample. Figure A.1 provides an overview of the search and screening process, while Table A1.2 provides the final sample of included articles.

To obtain relevant information from the articles, I adopted an inductive comparative approach on a structured data extraction form. For each article, I systematically extracted its theoretical foundation, stated theoretical contributions, key constructs used or

developed, and the definition or boundary of AI adopted (Büchter et al., 2020). The identification of themes followed a process of comparing recurring constructs and contributions across articles, grouping them into higher-level thematic clusters. After iterative rounds of refinement and reclassification, this process allowed the mergence of eight themes. In the following sections, I present a conceptual overview of literature by adopting a concept-centric approach (Fisch & Block, 2018). The findings are ordered according to the numerosity of articles included in each thematic stream and presented in accordance with the specific set of articles. However, where appropriate, I expanded the view to broader literature when key concepts and contributions were grounded in articles that were not included in the final sample.

3. RESULTS

3.1.Descriptive analysis

The distribution of the selected references indicates that more than 90% of articles have been published starting from 2021. Only 7 articles were published between 2018 and 2020 and none of them appear in issues published before 2018. The trend suggests the growing interest in the topic from top-tier journals. The 70 included articles were published in 21 journals and 29 (41%) of them were published in journals of the IS management field. The distribution of articles in IS management journals is uniform across time, with 59% of articles published in 2021-2022 and 34% published from 2023 to early 2024. In contrast, 51% (21 out of 41) of articles in other fields (general management, innovation, strategic management, organization studies, etc.) were published from 2023 to early 2024. Interest in IS journals has shown a more sustained and early engagement with the topic than other fields, even though contributions have been mainly empirical. Conversely,

conceptual and theory-oriented articles have emerged more recently and are more concentrated in non-IS management journals.

Figure 1.1. Number of included articles by year of publication

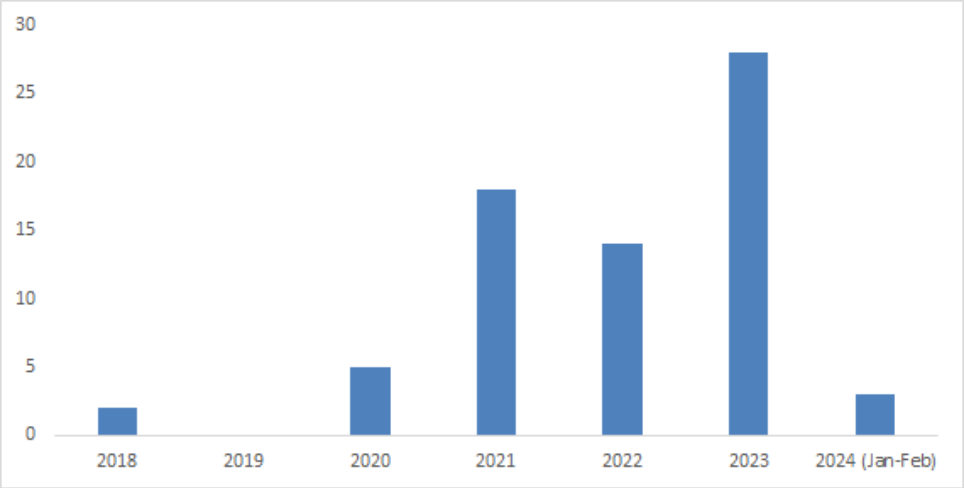


Figure 1.2. Number of included articles by period of publication and type of journal

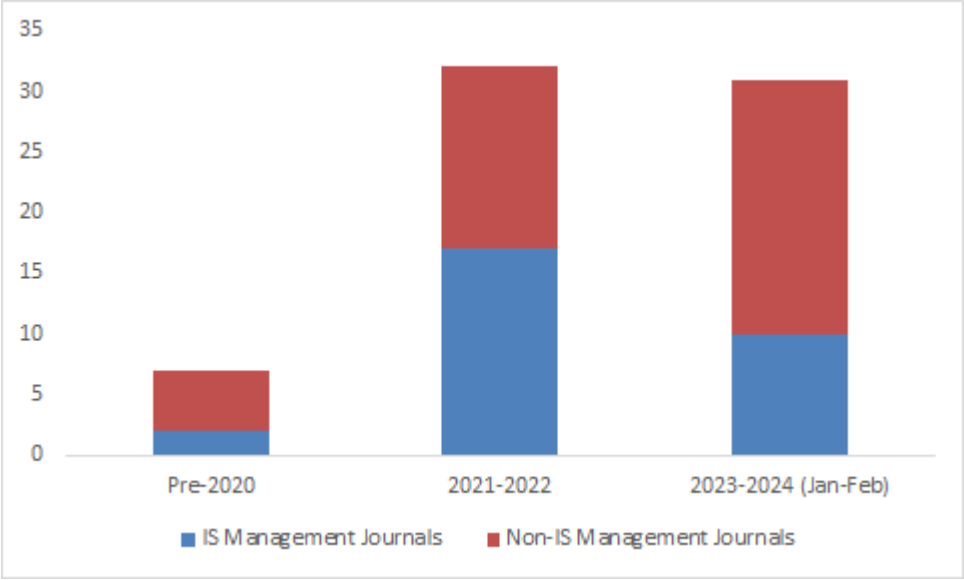
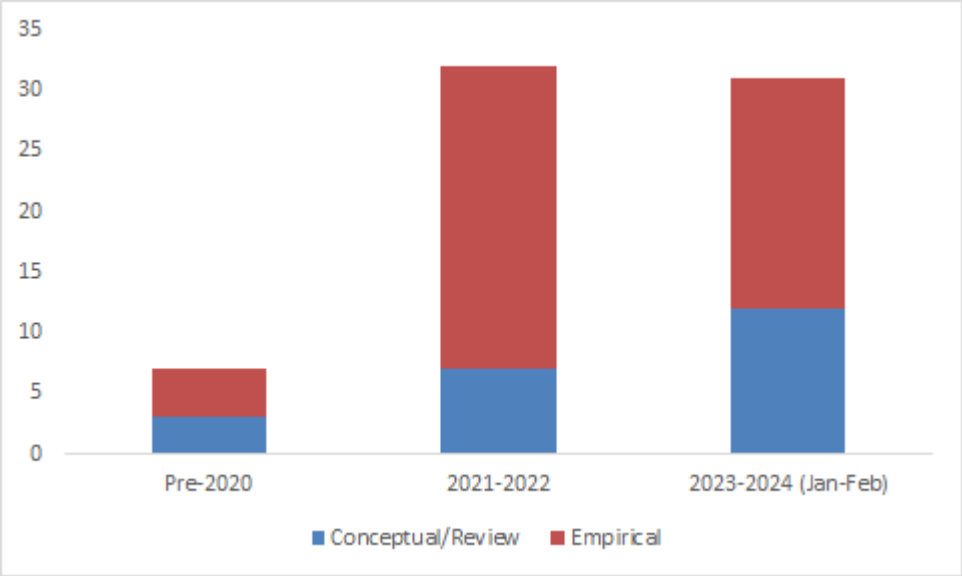


Figure 1.3. Number of included articles by period of publication and type of article



3.2.Defining the AI boundaries

The definitional boundaries of this review are shaped by the corpus of included articles and thus reflect the broad and heterogeneous use of the term Artificial Intelligence across various domains of literature. This choice is intentional and consistent with the definitional ambiguity that characterizes AI as both a technological and conceptual construct (Wang, 2019). Among the 70 articles included in the review, 39 provide some explicit definition of AI (or related concepts), while the remaining 31 refers to the term implicitly, assuming a shared understanding. Within the definitional subset, two different approaches can be identified. First, a group of studies refer to the technical capabilities of AI and its related terms (e.g., Asatiani et al., 2021; Balasubramanian et al., 2022; Choudhary et al., 2023; W. Wang et al., 2023). AI is an umbrella term encompassing a wide set of technologies and techniques, such as machine learning (ML), deep learning (DL), predictive models, generative tools, and data-driven algorithmic systems more broadly. A common reference of papers dealing with ML is to define it as a collection of statistical learning techniques that enable machines to generate models from large

volumes of data without explicit programming (e.g., Brynjolfsson & McAfee, 2014; Brynjolfsson & Mitchell, 2017). This approach reflects a technical-functional perspective on AI, foregrounding algorithmic and statistical processes over higher-level capabilities. Second, many definitions emphasize the human-like cognitive functions performed by AI systems – such as learning, reasoning, interacting, or problem solving. These accounts conceptualize AI as a technological analog to nuanced forms of human intelligence, stressing its capacity to replicate or simulate tasks traditionally associated with human cognition (e.g., Glikson & Woolley, 2020; Man Tang et al., 2022; Raisch & Krakowski, 2021). These also include the system's capacity to perform autonomous tasks and environmental adaptation, often highlighting features such as decision-making, interaction with dynamic contexts, and varying degrees of independence from human input (Baird & Maruping, 2021; Vannuccini & Prytkova, 2023; G. Wang et al., 2024). Given this landscape, the approach adopted in this review is to interpret AI in accordance with how each study defines or applies the term, whether through the lens of "AI," "algorithm," "ML," or other variants. While this flexible usage responds to the conceptual vagueness surrounding AI, it remains anchored in the academic discourse that considers such technologies part of the field. The aim is thus to embrace the evolving and plural nature of AI as constructed in scholarly work, while ensuring that only contributions situated within the boundaries of management and organizational studies are included.

3.3. Thematic summary

The conceptual diversity of the field is reflected in the eight thematic areas that emerged through inductive analysis. The most prominent stream, Human–AI Configurations, includes 16 articles (23%) and explores how human and AI agents are integrated and reconfigured in work settings. This is followed by themes such as AI Adoption and

Diffusion (11, 16%), AI & Knowledge and Professional Work (10, 14%), and AI & Decision-Making (9, 13%), which examine how AI technologies are taken up and embedded in organizations, knowledge and professional work, and decision-making processes. The stream AI & Strategic Capabilities (8, 11%) focuses on how AI contributes to building strategic capabilities and sustain competitive advantage, while Managing AI Ethical Challenges (8) engages with ethical, regulatory, and governance issues in AI deployment. Human–AI Interaction (6, 9%) concentrates on micro-level interactions between human actors and algorithmic systems. Lastly, the theme AI & Organizational Learning (2, 3%) examines how AI affects collective learning processes within organizations.

3.3.1. Human-AI Configurations

Research on Human–AI collaboration in work design distinguishes between *automation*, where AI systems substitute human work, and *augmentation*, where humans and machines jointly perform tasks (Grønsund & Aanestad, 2020; Puranam, 2021; Raisch & Krakowski, 2021).

Early work framed automation and augmentation as distinct trajectories: the former substitutes human work, the latter complements it. Recent scholarly work argues instead that the two modes are structurally intertwined and must be theorized together. Raisch & Krakowski (2021) mobilize paradox theory to argue how augmentation both drives and results from automation. They contend that *hybrid collectives* of humans and AI blur the boundary between human actors and AI artifacts, making them mutually reinforcing logics that unfold over time and across organizational spaces (Floridi & Taddeo, 2016). Anticipating this view, Coombs et al. (2020) introduce the notion of *intelligent automation*, emphasizing that AI-enabled automation occurs at the task level rather than

the work level. Because work is decomposed into granular tasks with varying degrees of automability, organizations reallocate human effort to supervision, collaboration, and oversight of AI systems. Grønsund & Aanestad (2020) provide empirical grounding to this dynamic, showing how the introduction of algorithmic systems gives rise to novel *augmentation work*. Their study highlights how human tasks such as auditing and altering the algorithm are mutually dependent and embedded in feedback loops that recursively link algorithm use and redesign. This *human-in-the-loop* configuration (Zanzotto, 2019) emerges as a new form of organizing that enables the co-evolution of human and algorithmic contributions.

Recent theorizing acknowledges the increasingly *agentic* role of AI. Murray et al. (2021) argue that contemporary technologies possess a temporally embedded capacity to constrain, complement, or substitute human intentionality in routines. When embedded in *human–nonhuman ensembles*, such technologies give rise to *conjoined agency* – defined as a shared capacity between humans and machines to shape both rules and customs development and action selection. The authors categorize four archetypes of conjoined agency, differing in how technology shapes action selection and protocol development and ranging from *assisting* technologies (fully controlled by humans), to *arresting* technologies (action selection), *augmenting* technologies (protocol development), and *automating* (both action selection and protocol development). This conceptualization resonates with Maragno et al. (2023), who extend the idea of conjoined agency into the public sector through the lens of organizational design theory and the microstructural perspective (Puranam, 2018). Their empirical study highlights how the introduction of AI systems in public sector organizations leads to the emergence of a novel organizing unit – the *AI team* – comprising three interdependent agents: the AI

solution, the AI trainer, and the public manager. This microstructure distributes intentionality across human and artificial agents, assigning to each distinct yet complementary responsibilities. Raisch & Fomina, (2024) deepen this view by mapping hybrid agency onto organizational problem-solving processes. Their *hybrid problem-solving* model articulates three collaboration modes – *autonomous*, *sequential*, and *interactive search* – that distribute agency differently across humans and AI agents. Interestingly, this is the first among the included paper to theoretically distinguish the role of predictive and generative AI. Autonomous search mirrors automating configurations, with AI independently generating solutions; sequential search parallels augmenting configurations, as AI frames problems and humans explore solutions; interactive search embodies a human-AI co-creative agency, where mutual learning shapes both problem definition and solution selection. Drawing on affordance theory (Markus & Silver, 2008), Vassilakopoulou et al. (2023) focus on how human–AI configurations unfold in practice. They identify several chatbot *affordances* – such as *filtering*, *informing*, *monitoring*, and *delegating* – that redistribute tasks between humans and AI in customer services.

As AI systems increasingly participate in decision-making, a central question emerges on whether and how tasks and authority should be sequentially distributed between humans and machines. Choudhary et al. (2023) push the argument further by proposing *human–AI ensembling* – a non-sequential division of labor without specialization in which humans and AI each generate predictions that are subsequently aggregated. Unlike delegation models that assign distinct sub-tasks to the better-suited team member, ensembling treats humans and AI as parallel decision makers whose heterogeneous decision processes can be pooled for superior performance. They also show that ensembling can arise from either augmentation (adding AI to humans) or partial

automation (replacing selected humans with AI), underscoring that hybrid agency configurations span on the automation–augmentation continuum. Other works adopt a sequential logic, using *delegation* – the transfer of decision rights and responsibilities (Baird & Maruping, 2021) – as a lens to examine specific configurations. Fügener et al. (2022) offer a foundational contribution by identifying three boundary conditions for effective delegation in human–AI configurations: the existence of complementarities, the recognition of such complementarities, and the execution of efficient delegation rules. These conditions hinge on what the authors define as *metaknowledge* – the human ability to evaluate one's own decision capability and recognize whether an AI is better-suited to perform a task. Crucially, even when AI systems outperform humans on average, humans often fail to delegate appropriately due to inaccurate self-assessment. This leads to suboptimal collaboration and reveals that effective delegation is not only a technical issue but also a cognitive and behavioral one. Extending the notion of delegation to the public sector, Dickinson & Yates, (2023) frame *automation* as a form of *technological outsourcing*, where organizations transfer control over service functions to AI systems. They caution that while responsibility can be delegated, accountability cannot. Their analysis outlines a set of principles for determining whether a task should be automated, emphasizing factors such as the centrality of the task to organizational legitimacy, the need for discretion by human judgment, and the functional quality of service provision. Revilla et al. (2023) offer a complementary conceptualization that systematizes delegation along a continuum of human involvement. In automation, decision rights are fully delegated to the AI system, with humans involved only at the design stage. *Adjustable automation* allows humans to intervene during the process, but only in select steps of the prediction. Augmentation, by contrast, entails active human involvement

throughout the predictive and decision-making stages. Their model provides a dynamic view of augmentation as an evolutionary process, whereby humans and AI learn from each other through repeated interaction. Building on the long-standing observation that IT systems progressively assume tasks once reserved to human operators, Baird & Maruping (2021) use the lens of *delegation* to recast human-IT relations. They theorize four archetypes of agentic IS artefacts – of which AI is an emblematic representation – arrayed along a continuum of autonomy, from *reflexive* artefacts reacting only to proximal stimuli to *prescriptive* artefacts that go further by prescribing or implementing courses of action. As AI artifacts move along the agency continuum, effective adoption hinges on organizing through three delegation mechanisms: *appraisal* (judging the machine's fitness for purpose), *distribution* (re-allocating rights and responsibilities), and *coordination* (establishing monitoring and assessment routines) between human and IT agents.

Regardless of its sequential or non-sequential nature, human–AI configurations reshape roles, behaviors, and performance expectations in organizational settings. Some studies move beyond structural task division and instead focus on how individual attributes, dispositions, and competencies interact with AI agents to shape emergent role configurations. Dennis et al. (2023) explore the integration of AI agents into *AI teams*, conceptualizing them as team members capable of performing support activities such as coordination, information retrieval, and reminders. Their experimental findings suggest that well-performing AI agents can reduce perceived interpersonal conflict compared to equivalent human team members, indicating that AI does not inherently disrupt interpersonal aspects of teamwork. Moving from interpersonal to individual dynamics, Jia et al. (2024) examine the creative potential unlocked through human–AI

configurations. In their field study, they find that a sequential division of labor – where AI handles routinary, repetitive tasks – can enable human employees to focus on more cognitively demanding and creative work. However, this *AI-augmented creativity* is not universal. Rather, its benefits are skill-biased: employees with higher domain expertise and job skills are better able to capitalize on AI assistance, whereas lower-skilled workers may derive fewer gains. This highlights how collaboration with AI can deepen existing inequalities in work design, requiring organizations to align skill profiles with AI capabilities to unlock complementary value. Expanding on this personal attribute perspective, Man Tang et al. (2022) investigate how individual personality traits – specifically *conscientiousness* – affect performance in human–AI pairings. Drawing on complementarity theory (Kiesler, 1983) and role theory (Sluss et al., 2011), the authors argue that conscientious employees may struggle to integrate effectively with AI agents that exhibit similar orderly and systematic behaviors. This non-complementarity produces misalignments in expectations and workflow dynamics, potentially undermining performance, suggesting AI-driven job design and task allocation may need to reflect psychological as well as functional complementarity. Fügener et al. (2021) complete this picture, showing that uniform algorithmic advice, while raising a person's accuracy, erodes its *human unique knowledge* as well as the collective knowledge of a group of individuals – the *wisdom of crowds* (Surowiecki, 2005) – leading to a loss of complementarity of humans with AI and with other humans.

3.3.2. AI & Knowledge and Professional Work

The use of AI carries distinctive implications when organizations heavily rely on knowledge or professional workers. Knowledge workers are highly qualified people using intellectual and symbolic skills in work tasks (Alvesson, 2004). In addition,

professional workers are characterized by a long and standardized formal education, a professional association governing its members, and an ideology that guides professionals to adhere to deontological rules and codes of ethics, whether explicit or implicit (Alvesson, 2004; Von Nordenflycht, 2010). Knowledge and professional workers tend to actively seek a deeper understanding of their technologies' inner working, to avoid the quality of their knowledge products be questioned or doubted and their status as experts be threatened (Anthony, 2021). Contemporary AI has raised expectations for augmenting or automating knowledge work by generating insights from data with less reliance on human expertise. Hence, the adoption of AI in professional work challenges the very identity and boundaries of professions and may cause some professional workers to resist the change (Anthony, 2021; Goto, 2021; Langley et al., 2019).

A first cornerstone in this thematic stream is Anthony (2018), a conceptual article that extends to all *epistemic technologies* – that is, artefacts explicitly designed to produce new knowledge, with AI systems as a prominent tool. Anchored in the classic literature on the development and coordination of expertise (e.g. Okhuysen & Bechky, 2009; Gittell & Douglass, 2012), Anthony argues that the quality of knowledge creation rests on the interaction among experts who interrogate the tools through which new insights are generated. The article builds from a property of epistemic technologies that come materially embedded with assumptions that are rarely transparent or intelligible even to seasoned professionals. Because these latent assumptions shape analytic outputs, the decision to probe or to trust the technology's result becomes a relational rather than a purely technical matter. Anthony theorizes two behavioral responses of knowledge workers to embedded assumptions – *questioning practices* (i.e. critical examination of hidden assumptions) and *accepting practices* (i.e. acquiescent reliance on outputs). Their

enactment is linked to contingent factors such as the salience of status differences within knowledge-work groups – conceptualized here as a mix of tenure, role, and expertise (Levina & Vaast, 2008) – perceived threats by high-status actors, and opportunities for status enhancement by low-status actors. A complementary perspective concerns individual practices through which scrutiny is enacted: these can be distinguished between *validation practices* (Anthony, 2021) and *interrogation practices* (Lebovitz et al., 2022). *Validation practices* refer to efforts by knowledge workers to ensure that analytical outputs appear trustworthy, without necessarily interrogating the assumptions or internal logics of the technology that produced them. Through an ethnography on investment-banking teams, Anthony (2021) distinguishes *validation practices* enacted in co-constructed work, where junior and senior bankers jointly walk through analytical scrutiny and interpretation, from those enacted in partitioned work, in which juniors assemble outputs while seniors merely interpret them. While the former aims to question the rationale behind the algorithmic output (i.e., opacity), those using the latter take the technology for granted without understanding how they work, even when engaging in efforts to validate analyses. Extending this argument to high-stakes medical diagnoses, Lebovitz et al. (2022) identify AI-interrogation practices as active attempts to reconcile AI claims with human knowledge. When ML technologies allow these interrogation practices to be adopted by radiologists, they observe *engaged augmentation*, whereby human and algorithmic knowledge are genuinely integrated. Otherwise, augmentation devolves into uncritical acceptance or outright dismissal (*unengaged augmentation* or *de facto* automation). The comparison between these two practices highlights a crucial distinction: whereas validation may confirm outputs without challenging the technology's

embedded assumptions, interrogation is always explicitly oriented toward exposing, understanding, and integrating these assumptions into expert judgment.

Other studies address the rationales for AI integration in knowledge work and professional practices. Two empirical studies address individual-level characteristics and scrutinize the role of knowledge workers experience. First, Allen and Choudhury (2022) propose that *domain experience* – domain-specific knowledge obtained by focused practice or learning by doing (Chari & Hopenhayn, 1991) – exerts dual and countervailing effects on human–AI teaming: it enhances a worker's *ability* to judge algorithmic advice while simultaneously fosters algorithm *aversion*. Their data from IT workers reveal a non-linear and non-monotonic (inverted-U) pattern whereby low-experience workers gain most from algorithmic augmentation, while highly experienced experts may under-perform when aversion outweighs incremental increases in ability. Second, Wang et al. (2023) sharpen this lens by disaggregating experience into *task-* or *volume-based* and *time-based* or *seniority* experience in medical coding. They find that task/volume experience complements AI, reinforcing productivity gains, whereas seniority fuels algorithmic aversion. Another perspective regards the inherently relational nature of expertise. This is the central focus of Pakarinen & Huising (2023). The relationality poses significant challenges for the augmentation potential of AI in professional contexts. Specifically, the authors identify three mechanisms through which relational expertise resists seamless integration of AI in professional practice. First, opacity arises because the tacit and situated aspects of professional knowledge are not captured by the data on which AI systems are trained, thus remaining inaccessible to algorithmic reasoning. Second, translation challenges limit the practical use of AI-generated outputs, as professionals must recontextualize AI suggestions within ongoing processes and existing

interpretive frameworks. Third, the question of accountability remains fundamentally human: despite AI's growing autonomy, society continues to hold human professionals accountable for decisions made in high-stakes contexts. Together, these mechanisms act as buffers that protect professional authority from being subordinated or displaced by AI. Instead of supplanting relational expertise, AI becomes integrated into the interactional context in which professional knowledge is generated.

This relational entrenchment also gives rise to new roles and responsibilities that serve to mediate between AI systems and expert decision-makers by translating opaque outputs into intelligible, context-sensitive insights. For example, it is the case of *knowledge* or *algorithmic brokers* discussed by Waardenburg et al. (2022). In organization theory, knowledge brokers are organizational actors enacting translation practices across knowledge boundaries (Brown & Duguid, 1998; Kellogg et al., 2020). Building on this definition, algorithmic brokers translate ML predictions to users, aiming to solve a knowledge boundary between ML communities and a user community. Through an ethnographic study, they show how these brokers enact a repertoire of translation practices – *messenger*, *interpreter*, *curator* – that not only bridge knowledge boundaries between technical and non-technical actors but also accumulate interpretive authority and produce new boundaries between themselves and the communities they serve. These findings highlight how algorithmic knowledge brokerage is not a static function but an emergent, situated process that shapes organizational structures of knowing and expertise.

Two further issues emerge from the relational view. The first concerns how AI is evaluated and integrated into knowledge work and professional settings. Lebovitz et al. (2021) highlight a fundamental tension between the dominant evaluation metrics used in

ML-based systems – such as ground truth labels and AUC[1] scores – and the criteria that professionals as a collective entity use to assess their knowledge work. While AI creators rely heavily on *know-what* – quantifiable outputs that stand as proxies for correctness – professionals rely instead on *know-how*: the embodied, tacit practices that underlie expert judgment in uncertain, situated environments. This mismatch has deep implications, as performance measures that privilege know-what over know-how risk marginalizing the profound characteristics that define professional work. While system developers often ignore the limitations of know-what measures, professionals recognize these limitations and focus their evaluative efforts on ensuring contextual adequacy. A similar tension is the focus of Van Den Broek et al. (2021). Through an ethnographic study of an ML-based hiring tool, they analyze how ML developers and domain experts navigate competing demands for *independence* versus *relevance* to situated expert knowledge. While ML developers initially aim to minimize human input to preserve algorithmic performance, they soon encounter issues – such as data selection, framing, and operational misalignment – that require reintegrating domain expertise. The process unfolds as a *mutual learning* dynamic, where developers iteratively include and exclude expert judgment to calibrate the system. This results in the emergence of a *hybrid practice* in which ML outputs are continuously shaped and validated by domain experts, reflecting not a replacement but a reconfiguration of epistemic authority in knowledge work.

The second relational issue relates to how professionals respond to perceived threats posed by AI systems. Strich et al. (2021) focus on how AI challenges the professional

[1] AUC (Area Under the Curve) score is an indicator assessing the performance of a binary classifier model, derived from ROC (Receiver Operating Characteristic) curves. It varies between 0 to 1, indicating how well a model can produce discriminate between positive or negative instances. A 0.5 AUC score indicates random guessing, while a 1.0 AUC score indicates perfect performance.

role identity in a loan consultancy firm. They show that identity threats occur at both the individual and collective levels, prompting professionals to engage in a variety of identity work mechanisms to protect or strengthen their roles. They argue how professionals actively reconstruct their identity boundaries in response to technological disruption. A similar process of collective defense and reconfiguration is theorized by Faulconbridge et al. (2023), who develop the concept of *intertwined boundary work*. They draw on the classic notion of boundary work as the effort by professional groups to influence the borders of their jurisdiction (Weber et al., 2022) and on the conceptualization of «types» and «modes» of boundary work in terms of objectives and effects that boundary work seeks to achieve (e.g., competition or collaboration with external actors) (Langley et al., 2019). They argue that in the context of AI disruptions, professionals engage in multiple, interdependent modes of boundary work, including *defensive modes* (protecting boundaries), *creative modes* (opening new boundaries), *negotiating modes* (developing collaborations), and *coalescing modes* (rework boundaries). Over time, these interconnected modes of response may reinforce one another in what they term *recursive change reinforcement*, consolidating new professional configurations.

Together, these contributions reveal a recursive dynamic: epistemic technologies such as AI are embedded with assumptions that are often opaque to users. Whether and how these assumptions are questioned depends on the status positioning and experiential background of knowledge workers. These conditions shape the practices of validation and interrogation, which in turn affect how relational expertise is enacted and how AI tools are evaluated. As these tensions surface, professionals engage in identity and boundary work to protect or redefine their roles – thereby influencing future configurations of status and expertise in AI-mediated knowledge environments.

3.3.3. AI Adoption and Diffusion

Taking a grounded, sociotechnical perspective on AI integration is essential because implementation unfolds through a set of stakeholders whose divergent assumptions shape design, deployment, and day-to-day use (Haefner et al., 2023). Rather than focusing on the moment of adoption (Battisti & Stoneman, 2003), research traces interactions and processes to understand the full cycle of organizational AI adoption and diffusion. Deployment is rarely plug-and-play, as organizations must readapt infrastructures, re-engineer processes, reskill employees, govern data, and arrange change management initiatives. Recent scholarship therefore shifts from a predominantly technical lens to one that foregrounds a social perspective on AI, i.e., how systems, people, work practices, organizational arrangements, and cultural forces interact (Sarker et al., 2019).

Two papers adopt a system-level standpoint. Anthony et al. (2023) depart from a *technology-as-a-tool* and a *technology-as-a-medium* perspective to a *technology-as-a-counterpart* one where AI is as actor within a multi-layered system. Technological artifacts are never neutral but embed designers' mental models and power relations, meaning that understanding collaboration with AI algorithms requires mapping the relational dynamics between producers, implementers, and users. Complementing this view, Vannuccini & Prytkova (2023) frame AI as a *Large Technical System* (Hughes, 1987), whose evolution is steered by *system builders* – tech giants, regulatory bodies, research institutions, and deploying organizations. These act as knowledge gatekeepers, either diffusing expertise through co-invention or strategically withholding it to consolidate advantage. The resulting ecosystem encompasses *AI creators* (producers of solutions from basic inputs and models from the giants), *AI integrators* (intermediaries that re-sell AI solutions adding some value), *AI-powered operators* (producers of AI for

internal operations), and *AI takers* (who source AI solutions to be used in-house), each occupying distinct positions in the value network and shaping which problems AI is mobilized to solve (Jacobides et al., 2021). Together, these contributions recast AI adoption as a contested, system-level process in which technological affordances and social structures co-produce organizational adoption outcomes.

Consistently with the system-ness nature underpinning AI development, the inter-organizational diffusion of AI technologies hinges on network dynamics and embeddedness within broader knowledge ecosystems. Dahlke et al. (2024) extend the notion of AI system-ness by demonstrating how AI adoption propagates through social capital and relational ties across organizations. Drawing on social capital theories and epidemic diffusion models (Karshenas & Stoneman, 1993; Tsai & Ghoshal, 1998), they identify three mechanisms by which AI knowledge disseminates. First, *indirect co-location* effects emerge when organizations operate within regional or industrial clusters associated with AI production, generating *mimetic pressures* and *indirect knowledge spillovers*. Second, *direct exposure* occurs through formal collaborations and high-intensity inter-firm ties, such as R&D partnerships, that facilitate deeper knowledge exchange. Third, *relational embeddedness* within dense AI knowledge networks amplifies the diffusion of AI know-how, though it may also result in path dependency and restricted access for peripheral actors. These findings underscore that AI diffusion is not solely a matter of technological readiness or market incentives, but also of organization's position within social and epistemic networks.

While Dahlke et al. (2024) highlight the external relational conditions for AI diffusion, a third group of studies explore the relationship between AI adoption, organization's maturity, governance structures, and capacity to re-engineer work. By shifting the focus

within the organizational boundaries, Igna & Venturini (2023) examine the role of internal innovation capabilities in shaping AI development trajectories. Their analysis reveals that organizations exhibiting AI inventive success tend to build on *dynamic returns* from prior digital innovation experiences. These complementarities suggest that AI innovation is path-dependent, favoring actors with a cumulative technological base that can be recombined and extended into the AI domain. Together, these two studies emphasize that AI diffusion is shaped by both external epidemic effects and internal absorptive capacities, reinforcing a view of adoption as a socially embedded and cumulative learning process. In public sector organizations, Neumann et al. (2024) as well argue – using an AI-adapted Technology–Organization–Environment (TOE) framework (Tornatzky et al., 1990; Pumplun et al., 2019) – that maturity and accumulated experience with AI projects in public organizations play a critical role in enabling AI systems to be applied to core functions, to leverage internal resources effectively, and to develop increasingly complex solutions in-house. Goto (2023) complements this view in professional-service firms: dedicated service-R&D units act as orchestrators that reconcile professionals' autonomy with centralized innovation. By mixing top-down steering with bottom-up idea generation, and by engaging frontline professionals while framing digital initiatives as part of wider societal change, these units legitimate AI adoption and accelerate learning cycles.

Once AI moves inside, organizations confront questions of structure and power. Pachidi et al. (2021) ethnographically investigate how the implementation of a predictive model in a telecommunications company reshaped what they term a *regime of knowing* – a process encompassing the practices through which actors generate and use knowledge, the valuation schemes by which work and expertise are assessed, and the authority

arrangements that assign control over knowledge production. Their longitudinal case traces a profound shift from a customer-oriented regime, centered on sales account managers, to a data-oriented regime, dominated by data scientists. Interestingly, the transformation was not linear nor entirely orchestrated. Rather, it unfolded through cycles of *symbolic conformity* – simulated adoption by incumbent actors intended to preserve the status quo – that paradoxically legitimized and accelerated the new order.

In the public sector domain, three contributions address how the implementation of AI-driven technologies profoundly impact the way bureaucratic organizations work. Meijer et al. (2021) coin *algorithmization* – the extensive rearrangement of routines around AI algorithms. They show that organizational culture channels this process toward two archetypes: the *algorithmic cage*, where algorithms exert control and bureaucratic power, and the *algorithmic colleague*, where algorithms act as advisers that share formal, contextual, and tacit knowledge with professionals. Vogl et al. (2020) develop a related framework of *algorithmic bureaucracy* that maps how automated processes, hierarchical roles, and traditional practice co-evolve, revealing feedback loops between street-level bureaucrats and system-level designers. Giest & Klievink (2024) compare two public-sector cases and show that AI adoption simultaneously sparks *administrative process innovation* – a restructuring of roles and formal processes around the new system – and *conceptual innovation*, a redefinition of how the underlying policy problem is addressed. Finally, Asatiani et al. (2021) propose *sociotechnical envelopment* as a design principle for balancing human and AI agency. By deliberately bounding training data, inputs, outputs, and functional goals (Floridi, 2011; Robbins, 2020), organizations construct envelopes that allow even opaque models to operate predictably while keeping humans in the loop.

These studies push the sociotechnical adoption agenda beyond initial uptake toward dynamic capability formation and organizational redesign. Maturity trajectories (Igna & Venturini, 2023; Neumann et al., 2024), coordinating structures (Goto, 2023), and culturally inflected configurations of algorithmic work (Vogl et al., 2020; Asatiani et al., 2021; Meijer et al., 2021) depict AI adoption as an iterative process in which technology and social arrangements continually reshape one another, an insight that matches with the systemic, network-based view outlined earlier (Anthony et al., 2023; Vannuccini & Prytkova, 2023). Altogether, this body of work demonstrates that AI adoption and diffusion cannot be understood as discrete events but as ongoing, contested processes in which technological artifacts and social structures coevolve.

3.3.4. AI & Decision-Making

AI tools increasingly demonstrate the capacity to outperform human judgment in a variety of decision-making tasks. By generating high-quality predictions, AI systems can provide advice that enables organizations to move from intuition-based decisions to more data-driven, evidence-informed processes. Nevertheless, despite their potential, the adoption of AI-generated advice is far from straightforward. A growing body of research has documented how human decision makers often hesitate or refuse to use algorithmic recommendations – even when these are demonstrably superior – a phenomenon typically labeled *algorithm aversion* (Dietvorst et al., 2015a, 2018; Burton et al., 2020). Recent contributions have expanded this stream of inquiry by investigating the intertwined topics of antecedents of algorithm aversion and the psychological and contextual mechanisms shaping advice-taking from AI systems. These studies reveal that algorithm aversion is neither uniform nor purely cognitive but is mediated by professional norms, contextual insights, and decision-making structures.

Focusing on clinical decision-making, Bankuoru Egala and Liang (2023) identify several sources of aversion to AI advice among physicians. Their findings suggest that algorithm aversion is not merely a function of general distrust or resistance to technological change but rather emerges when algorithmic outputs challenge clinicians' epistemic authority. Specifically, clinicians exhibit resistance when algorithmic recommendations cannot be modified in light of their professional judgment, and when the system delivers inconsistent results without allowing the user to exert interpretive control. This aversion is amplified under conditions of imperfect feedback, high cognitive load, and ethical responsibility, reinforcing the need to preserve professional discretion within AI-assisted decision processes. In parallel, Liu et al. (2023) examine decision-making among rideshare drivers and reveal that algorithm aversion is shaped by two key dynamics: *contextual experience* and *herding*. Contextual experience refers to drivers' specific knowledge of the places and times in which they operate and provides a broader range of insights compared with general domain expertise, which is often considered relevant in algorithm aversion literature (Logg et al., 2019; Allen & Choudhury, 2022). Herding, on the other hand, captures the social dimension of algorithm aversion: drivers observe and emulate their peers' behavior before deciding to accept or override algorithmic advice. The study also finds that such aversion may persist even when algorithms are designed to optimize system-level efficiency, especially when individual incentives are misaligned with global objectives. Similarly, a survey experiment with street-level bureaucrats from Wang et al. (2024) shows that while AI reduces perceived discretion in implementing policy decisions, it can also increase willingness to implement policy for *programmed decisions* – where task complexity is low and procedures are routinized. In contrast, *non-programmed decisions*, which involve higher uncertainty and require human judgment,

appear to mitigate the negative effect of AI on discretion and reduce enthusiasm for AI-based implementation. Also, De Véricourt & Gurkan (2023) challenge the assumption that algorithm aversion is driven solely by attitudes or trust deficits and propose that aversion may emerge from learning failures rooted in the structure of human–AI decision interaction. They introduce two structural features: *verification bias* and *exploration-free decisions*. The former implies that the decision maker observes the correctness of the machine's prediction only if the action is actually taken. The latter, which is common in high-stakes settings, comprises cases where experimentation is ethically or practically constrained, and decision makers are often unable to verify algorithmic accuracy unless they accept its recommendations. Over time, this learning constraint leads to either indiscriminate acceptance, indiscriminate rejection, or erratic behavior – all of which undermine the formation of calibrated beliefs on the algorithmic functioning and accuracy. Adding to this behavioral view, Boyacı et al. (2024) develop a formal decision-theoretic model to analyze how machine-generated predictions influence human judgment under cognitive constraints. Using a rational inattention framework (Sims, 2003), the authors show that while AI advice generally increases the accuracy and expected utility of human decisions, it may also produce unintended effects. Notably, machine input can exacerbate false-positive rates and raise cognitive effort, especially when decision-makers operate under time pressure. These effects emerge because the human decision-maker reallocates effort depending on how the AI output shifts prior beliefs and task complexity. The study identifies the specific conditions under which human–machine collaboration is most beneficial, i.e., the situations "confirming the rather expected", in contrast to situations "falsifying the expected" that may deteriorate the efficiency of the human by inducing more cognitive effort.

Integrating these perspectives, Sturm et al. (2023) offer a sociotechnical account of advice utilization, shifting attention from algorithm aversion to the broader conditions that influence whether algorithmic suggestions are used in organizational decision-making. Drawing on the *judge–advisor system* (JAS) paradigm (Bonaccio & Dalal, 2006), they conceptualize the AI system as the advisor and the human as the ultimate judge, responsible for the decision. Their findings show that advice utilization depends on both technical and social factors. On the technical side, the perceived quality and transparency of ML advice increase the likelihood that it is used. On the social side, the interpersonal relationship between decision makers and data scientists significantly shapes the human judge's willingness to rely on AI-based advice. Building on the same paradigm, You et al. (2022) further explore how advice-taking behavior is shaped by cognitive and relational mechanisms. Their experimental study shows that the effect of prediction performance transparency on algorithmic advice utilization is mediated by both *trust in the advisor* and *cognitive load* – one's available working memory (Kalyuga, 2011). In particular, users are more likely to rely on algorithmic advice when they trust the algorithm more than a human advisor – but only if the presentation of performance information does not induce excessive cognitive load. Moreover, the effect of transparency is moderated by the judge's *need for cognition*, that is, their intrinsic motivation to engage in complex reasoning (Cacioppo et al., 1996). Individuals with high need for cognition are more sensitive to nuanced performance and are therefore more likely to integrate more transparent AI advice into their decision-making.

A final set of contributions moves from explaining aversion to identifying mechanisms that may recalibrate it. Dietvorst et al. (2018) explore strategies to mitigate algorithm aversion. Their experiments show that people are more likely to accept algorithmic

forecasts when they are given limited discretion to adjust the prediction. Framing the decision to use an algorithm as an all-or-nothing commitment reduces uptake, whereas allowing constrained adjustments improves both perceived fairness and user satisfaction – even when the algorithm is known to be imperfect. Jussupow et al. (2021) also examine the cognitive mechanisms involved in AI-augmented decision making in medical contexts. They focus on how AI advice interacts with physicians' own diagnostic reasoning. Their findings highlight three reasons that prevent physicians from exploiting the full benefits of an AI-based decision support system. First, early exposure to algorithmic advice can hinder the development of independent reasoning, leading to biased decision-making. Second, when AI recommendations contradict a physician's opinions, the decision maker is prone to rely either on self-confidence or blind trust in the system's capabilities – both of which can produce suboptimal outcomes. Such a relationship is moderated by decision-maker's experience, with more experienced physicians more likely to ignore AI advice. Third, even if exerting *metacognitive control* - actions to monitor, evaluate, and control human decision-making (Ackerman & Thompson, 2017) – these are based on *self-monitoring* and *system-monitoring*. Used in combination, both are seen as critical tools for reconciling conflicting cues and exercising appropriate judgment. The metacognitive perspective helps to reconcile the dynamic between systematic and heuristic reasoning that is intertwined during a decision-making process (Tversky & Kahneman, 1974; Kahneman, 2013; Ferratt et al., 2018).

3.3.5. AI & Strategic Capabilities

Recent strategy scholarship positions AI as a pivotal – yet contested – source of competitive advantage. The capabilities view argue that competitive advantage arises when organizations orchestrate assets through capabilities that are idiosyncratic,

inexpensive to develop, and well aligned with both internal resources and external environmental conditions (Barney, 1991; Dosi et al., 2000; Winter, 2003). Most of the papers reviewed in this thematic stream ground on resource-based view (RBV) and dynamic capabilities view (DCV), examining how capabilities are embedded to generate competitive advantage.

The starting point focuses on the internal orchestration of capabilities. A key insight is that considering AI's general-purpose and pervasive nature, organizations face nontrivial challenges in transforming it into a source of competitive advantage (Brynjolfsson & McAfee, 2014). Kemp (2023) argues that AI is *generic* (models are not unique to single organization), *explicit* (knowledge must be encoded and hence shareable), and *myopic* (AI lacks wider context), making differentiation challenging. While Generative and Agentic AI challenge parts of this view by expanding the range of tasks and potential autonomy (Bommasani et al., 2021), this core concern remains. To overcome these limitations, the author introduces the notion of *situated AI*, emphasizing the need to circumscribe the AI's agency within the organization's unique experiential, structural, and relational systems. He proposes three situating activities – *grounding* (learn across the organization), *bounding* (learn from competitors), and *recasting* (continuous adaptation) – that enable organizations to embed generic, explicit, and myopic AI into organizational routines that integrate both human and non-human agency. When effectively orchestrated, these routines help produce *AI-driven capabilities* that are difficult to imitate and more tightly aligned with the organizations' environment. From a different angle, Rana et al. (2022) offer a negative perspective to this view by identifying the conditions under which AI fails to produce competitive advantage. Studying AI integrated business analytics (AI-BA), they conceptualize *AI-BA opacity* as stemming

from weak governance, poor data quality, and ineffective training of employees. These deficiencies compromise the explainability and reliability of AI-BA systems, leading to suboptimal decisions, operational inefficiencies, and – ultimately – competitive disadvantage. Further reasoning on potential failures, Engel et al. (2024) develop a model to assess the amenability of business processes to cognitive automation (CA), thereby helping organizations make more informed strategic decisions about where and how to apply ML technologies. Drawing on an action research study with a leading European manufacturing company, they identify four key requirement dimensions – *data*, *transparency*, *cognition*, and *relationship* – that determine the feasibility and success of CA initiatives. Complementing the previous work, they offer a diagnostic tool to anticipate barriers and reduce project failure rates, reinforcing the idea that AI capabilities must be carefully embedded within organizational and process contexts to yield competitive advantage. Lastly, drawing on upper-echelons theory (Hambrick et al., 2005; Hambrick, 2007), Li et al. (2021) examine *AI orientation* – an organization's strategic direction and goals for AI adoption – as a distinct organizational capability shaped by top management. They show that the presence of a Chief Information Officer (CIO) and board knowledge heterogeneity (educational diversity, R&D experience, AI experience) significantly strengthen AI orientation, highlighting leadership composition as a critical antecedent to effective AI capability development.

Within a product-oriented perspective, two variants of AI capability emerge. Building on RBV, Lou and Wu (2021) define *AI innovation capability* as the "ability to develop, manage, and utilize AI resources for scientific discovery and research and development" (p. 1452). This includes the orchestration of tangible assets (data, infrastructure), AI-specific skills to create, implement, and deploy AI tools, and AI-enabled intangibles such

as codified knowledge and routines. Gregory et al. (2021) take a user-centric view, introducing the concept of *platform AI capability* – the ability of a platform to learn from user data to continuously improve services. This capability gives rise to *data network effects*, whereby perceived user value increases with the scale and sophistication of the platform's AI learning processes. Thus, platform AI capability relies on internal data-governance routines (Rana et al., 2022) to sustain external data network effects. This complements traditional network externalities and extends the scope of RBV to account for market-facing AI capabilities.

Turning to value-creation mechanisms, Krakowski et al. (2023) echo Kemp's conjoined routines by arguing that AI adoption triggers a dual dynamic of substitution and complementation, underscoring the need for careful orchestration. AI can replace some human domain-specific cognitive capabilities, thus erode traditional sources of advantage, and yet enable complementation, where humans contribute creative, interpretive, or contextual capabilities that augment the machine's output. The result is a new source of sustainable competitive advantage in uniquely configured *human–machine bundles*. Finally, Shollo et al. (2022) extend this reasoning through a processual lens, identifying three *ML value creation mechanisms – knowledge creation*, *task augmentation*, and *autonomous action*. The orchestration of these mechanisms resembles a DCV in action, where data scientists and organizational actors jointly adapt value creation pathways to match changing environmental conditions. This perspective further reinforces the importance of moving beyond AI as a fixed input, toward a view of AI as embedded in evolving configurations of routines, roles, and strategic intent.

Across these strategy-oriented studies, a clear message emerges: AI becomes a sustained source of competitive advantage only when organizations actively orchestrate data

infrastructure, human expertise, and organizational routines into situated, continually re-configurable capabilities.

3.3.6. Managing AI Ethical Challenges

In contemporary organizations, the ethical concerns and social disfunctions triggered by AI adoption – most visibly bias, fairness, accountability – cannot be resolved by technical fixes alone. They require deliberate organizational actions that align technological design with human judgment (Benbya et al., 2020; Berente et al., 2021; Rai et al., 2019). The eight papers in this thematic stream therefore examine how organizations manage these ethical challenges.

From a broader perspective, Heyder et al. (2023) examine AI ethics through a sociomaterial lens, arguing that ethical outcomes emerge from the interaction between individual moral intentions (*virtue ethics*) and formal organizational rules and principles (*duty ethics*). Rather than treating AI ethics as a purely technical compliance issue, they show how responsibility can be displaced into AI systems, weakening individual accountability, or conversely reasserted through active human oversight and moral engagement. Their analysis suggests that managing AI ethics requires organizational structures that prevent the erosion of human responsibility and deliberately align technological agency with institutional norms.

All papers addressing algorithmic bias share the premise that it is a sociotechnical phenomenon whose social consequences depend on how organizations detect and mitigate it and distribute decision authority (Favaretto et al., 2019). Building on this premise, Choudhury et al. (2020) focus on *input incompleteness* as a source of bias that arises when users strategically withhold or manipulate information. Input incompleteness reflects strategic human behavior and requires human complementarity to be addressed.

The authors show experimentally that *domain expertise* and *vintage-specific skills* – familiarity with AI-executed tasks (Chari & Hopenhayn, 1991) – enable employees to recognize missing contextual cues and correct *salience bias* – when decisions are made based on what is most salient, rather than relevant. Selten et al. (2023) and Alon-Barkat & Busuioc (2023) shift the focus from algorithmic outputs to human behavioral responses in high-stakes public-sector settings – street-level bureaucrats in municipal services and police officers, respectively. Both studies interrogate *automation bias*, defined as the human propensity to defer to automated hints (Mosier et al., 1998). Their results reveal no blanket pattern of blind adherence to algorithms; yet this discretion is itself susceptible to distorted utilization. Alon-Barkat and Busuioc (2023) document *selective adherence*, whereby decision-makers adopt algorithmic advice when it aligns with pre-existing stereotypes about social groups. Selten et al. (2023) highlight an allied *confirmation bias* mechanism as users attend to AI outputs primarily when those outputs corroborate their prior judgments. All these findings carry a double-edged implication. On one hand, residual human judgment can serve as a corrective to faulty or biased AI recommendations; on the other, the discretionary judgment may re-introduce social biases the technology was expected to eliminate, limiting the promise of fairer algorithms as a solution for discriminatory decision-making.

Algorithmic bias is closely intertwined with the broader construct of *algorithmic fairness* – the extent to which an AI system does not discriminate basing on protected attributes such as race, gender, origin, etc. (Kordzadeh & Ghasemaghaei, 2022; Teodorescu et al., 2021). Kordzadeh and Ghasemaghaei (2022) articulate a causal chain in which algorithmic bias erodes perceived fairness, and diminished fairness, in turn, reduces algorithmic reliance (e.g., advice acceptance, system adoption). They propose factors that

moderate each link in the chain. Individual dispositions such as moral identity, and organizational elements such as policies, rules, and standards, map directly into the aforementioned distinction between virtue ethics (intrinsic moral intent) and duty ethics (institutionalized duties and obligations) (Heyder et al., 2023). Wang et al. (2023) adopt a game-theoretical lens to examine the organizational trade-offs of disclosing algorithmic logic to users who may act strategically. Their analysis reveals that transparency may incentivize high-performing users to invest effort into desirable causal features, thereby increasing firm performance. However, the same transparency can also enable gaming behaviors by low-performing users who exploit non-causal, correlational features. This dynamic creates an ethical dilemma: while transparency aligns with fairness and accountability principles, it may backfire by reducing overall utility or widening disparities across user types. The study further shows that firms and users often prefer opposite transparency regimes, highlighting a structural tension between organizational and ethical incentives. Extending this fairness agenda to concrete design choices, Teodorescu et al. (2021) propose that organizations should prioritize augmentation practices to achieve algorithmic fairness. By combining fairness difficulty with the locus of the final decision (AI vs human), they delineate four modes that give managers an actionable set of augmentation arrangements: *reactive oversight* (AI decision and humans ex post control), *proactive oversight* (AI decision and human ex-ante control), *informed reliance* (human decision with machine advice), and *supervised reliance* (humans decision and actively scrutinize model logic). Each mode specifies distinct allocations of verification effort, accountability, and resource intensity, thereby translating the bias-fairness chain articulated by Kordzadeh and Ghasemaghaei into design prescriptions. Finally, Qin et al. (2023) shift the lens from algorithmic design to user

perception by directly comparing perceived fairness in AI-based versus human-based performance evaluations. In an experimental setting, employees judged AI assessments to be both fairer and more accurate than those delivered by an average manager. Crucially, however, when human evaluators were viewed as fair the performance gap narrowed markedly. As AI's analytical accuracy becomes commoditized, relational and ethical skill sets grow in strategic importance for human decision makers tasked with taking ethical decisions.

3.3.7. Human-AI interaction

A large body of literature on AI has focused on how users perceive, engage with, and adapt to AI-based artifacts. These contributions build on the rich tradition of human–computer interaction (HCI) research, which has long examined how humans interact with technological artifacts (Parasuraman et al., 2000; Card, 2018). Papers in this streams work around two interrelated topics: human trust on AI and AI agency. In psychology, perception of agency is the belief that an entity can think, plan and act of its own accord (Gray & Wegner, 2012). While in literature on human trust requires the trustee to act with agency (Rousseau et al., 1998), in the human-machine domain users may have confidence in a technology's reliability without attributing any agency to it (Parasuraman & Riley, 1997). Whether an AI artifact is perceived as an intentional agent therefore becomes pivotal. Schmitt et al. (2023) focus on how users come to attribute agency. Their conceptual work highlights the role of *voice capabilities* – specifically, natural language processing performance and the naturalness of speech – in triggering agency attributions. They argue that voice affords a uniquely human cue that short-circuits the necessary embodiment usually applied to IT artifacts: the more an AI system understands and speaks like a human interlocutor, the more likely users are to treat it as an intentional agent.

Recent conceptual work argues that incorporating perceived agency into trust models yields a more fine-grained understanding of when and why people are willing to trust or distrust AI systems (Vanneste & Puranam, 2024). Trust has been identified as a crucial determinant in AI utilization, particularly because of the opaque nature of many contemporary AI systems. Glikson and Woolley (2020) provide a foundational contribution to this theme with their review of human trust in AI. They identify two key antecedents of trust: the *AI embodiment* (robotic, virtual, or embedded) and its level of machine intelligence. Their analysis distinguishes between *cognitive trust* – based on perceived competence, reliability, and transparency – and *emotional trust*, which emerges from anthropomorphic features such as human-likeness and immediacy behaviors (e.g., responsiveness or active listening). Interestingly, the impact of task characteristics differs across these dimensions. While users tend to exhibit greater cognitive trust in AI when dealing with technical or analytical problems, emotional trust is more difficult to foster in tasks requiring social or interpersonal sensitivity. This distinction is echoed in empirical studies that explore the interaction between embodiment and trust. Chandra et al. (2022), applying media naturalness theory (Kock, 2009), show that user trust in conversational AI agents is enhanced when the system demonstrates *cognitive*, *relational*, and *emotional competencies*. Their findings emphasize that *naturalness*, rather than technical performance alone, drives user engagement with AI.

Beyond trust, other affective responses to AI are beginning to receive scholarly attention. Einola et al. (2023) examine the affective dynamics of anthropomorphized AI introduction in a media consultancy firm. They find that humanizing AI can amplify negative emotions – such as frustration, confusion, and tension – particularly when there is a mismatch between management's abstract enthusiasm and employees' concrete

dissatisfaction with its present limitations. These insights complicate the more optimistic account offered by Glikson and Woolley (2020), who suggested that anthropomorphism can enhance emotional trust. A related issue concerns how perceptions of AI influence trust on algorithmic outputs. Tong et al. (2021) examine AI-based performance feedback and identify a paradox of *deployment* and *disclosure effects*. While AI-generated feedback can improve employee performance (*deployment effect*), explicitly attributing feedback to AI may trigger distrust and fears of job replacement (*disclosure effect*). The authors suggest that hybrid configurations – where AI systems generate feedback and human managers communicate it – may mitigate these tensions and support more effective implementation.

3.3.8. AI & Organizational Learning

Organizational learning research portrays firms as collective entities that encode experience into routines and, in so doing, accumulate a form of memory that is independent from individual members and sustain competitive advantage (Levitt & March, 1988; Feldman & Pentland, 2003). IT scholars have long debated whether digital tools amplify or erode this memory by shaping how knowledge is created, shared, and stored (Robey et al., 2000; Alavi & Leidner, 2001). Two recent contributions push this debate into the era of AI – and specifically ML – by treating algorithms themselves as learners whose trajectories must be coordinated with human learning processes. Sturm et al. (2021) start from the premise that effective organizational learning hinges on managing a heterogeneous population of learners. Their simulation results show that when ML systems possess strong initial learning capabilities, they can relieve humans of exploitative tasks and free managerial attention for exploration, thereby offsetting the well-known tendency toward *learning myopia*. This benefit materializes only if

organizations deliberately absorb the beliefs generated by both human actors and ML into evolving routines – an imperative that grows more acute as environmental turbulence increases. Balasubramanian et al. (2022) offers a complementary tale: substituting human decision making with ML can itself exacerbate learning myopia. Because algorithms select a narrow set of routine variants that best fit historical data (*selection effect*) and contribute little contextual knowledge (*nescience effect*), they may reduce both the diversity and richness of organizational routines. The result is a heightened risk of overlooking interdependencies, extreme events, and slow-moving change. Algorithmic substitution thus benefits learning only at the extremes – either in very stable environments, where diversity is less critical, or in highly volatile ones, where any additional predictive power is valuable despite knowledge loss. These studies recast AI not merely as an input to learning but as a second organizational learner whose influence is double-edged. Coordinating human and algorithmic learning requires balancing ML's efficiency gains against its propensity to narrow the firm's experiential base, adapting the mix of exploration and exploitation to environmental dynamism, and institutionalizing routines that integrate – rather than replace – distinctive human insight.

4. DISCUSSION AND FUTURE RESEARCH DIRECTIONS

The review allows to provide an overview of the most recent approaches to the topic of AI in organizations, as valued by the most relevant journals in the field. It offers a comprehensive, though not exhaustive, picture of the main challenges that management and organization studies are called to deepen in order to accompany a transformation of such significant dimensions as the one anticipated with AI. While the eight thematic streams provide an analytical structuring of the literature, this section discusses six cross-

cutting insights that emerge across multiple streams and reflect patterns that go through thematic boundaries.

The first insight of this review shows that the long-standing tension between automation and augmentation is no longer presented as an either/or dilemma (Raisch & Krakowski, 2021). Across perspectives, this distinction has been conceptually reframed and empirically resolved in favor of combinatory configurations. At the micro level, this shift originates from a task-based understanding of work: organizations are made of people and tasks, and tasks – rather than jobs or roles – are the true unit of AI transformation (Coombs et al., 2020; Grønsund & Aanestad, 2020). AI may automate some tasks while augmenting others, even within the same job, depending on task structure and affordances. Hence, talking about automation versus augmentation is misleading. The current literature focuses instead on how human–AI and human-in-the-loop-configurations can be designed to both leverage the distinct and complementary contributions of humans and machines and mediate risks and limitations of AI (Choudhury et al., 2020; Fügener et al., 2021; Man Tang et al., 2022; Alon-Barkat & Busuioc, 2023; Jia et al., 2024). At the organizational and strategic level, this perspective carries important implications. Studies on AI and competitive advantage converge on the idea that automation alone is not a source of sustainable advantage. Merely owning a powerful algorithm is insufficient because of AI's generic, codified and rapidly diffusing nature makes it easily imitable (Kemp, 2023). Advantage instead stems from the orchestration of AI-enabled capabilities. Grounding-bounding-recasting schema (Kemp, 2023), human-AI bundles (Lou & Wu, 2021), data-network effects (Gregory et al., 2021) all point to the same prerequisite: rich, well-governed data input and firm-specific knowledge assets that embed the AI tools in idiosyncratic contexts. including the ability

to mobilize proprietary data, integrate AI into idiosyncratic routines, and align AI tools with the strategic direction and cognitive diversity of top management (J. Li et al., 2021; Krakowski et al., 2023). Dynamic value creation unfolds not through substitution, but through the co-evolution of human–AI complementarities. The studies converge on the insight that lasting advantage derives from human–machine bundles enacted through conjoined routines and not from automation alone (Krakowski et al., 2023; Shollo et al., 2022). Together, these findings caution strategists against technology-deterministic expectations.

A second cross-cutting insight is the widespread recognition of the agentic nature of AI systems. Even before the rise of generative or autonomous models – also known as Agentic AI, to distinguish it from traditional (predictive) AI and generative AI (Finn & Downie, 2025) – several studies emphasize the ability of AI systems to perceive, reason, and act with agency (Murray et al., 2021; Maragno et al., 2023). The concept of conjoined agency and routines captures this shared intentionality between human and machine agents, distributed across task execution, protocol development, and decision-making processes (Kemp, 2023; Murray et al., 2021). Agency is not monopolized by one actor but becomes embedded in sociotechnical configurations that span levels of autonomy and responsibility.

Third, from an interactional perspective, the literature confirms that human–AI engagement is shaped by a range of cognitive, relational, and contextual factors. While early work documented widespread algorithm aversion, more recent studies show that trust – both cognitive and emotional – is a key mediator of AI use (Glikson & Woolley, 2020). Factors such as task type, transparency, embodiment, and performance feedback shape trust levels and, in turn, affect usage (Chandra et al., 2022; Einola et al., 2023).

Studies suggest that constrained discretion and performance disclosure can enhance adoption, while poor alignment between user expectations and system behavior triggers resistance.

A fourth insight concerns the multilevel nature of AI adoption and diffusion. Organizations do not operate in isolation: their ability to engage with AI depends on embeddedness in wider knowledge and relational ecosystems. At the network level, AI adoption is shaped by spillovers, partnerships, and exposure to knowledge hubs (Dahlke et al., 2024). At the institutional level, actors such as system builders, regulatory bodies, and brokers shape the value network of AI deployment (Vannuccini & Prytkova, 2023). Internally, the presence of algorithmic brokers and the ability to manage work on knowledge and professional boundaries determine how technical outputs are integrated into practice (Lebovitz et al., 2021; Anthony, 2021; Waardenburg et al., 2022). These dynamics emphasize the sociotechnical nature of adoption as an ongoing, negotiated process.

Fifth, the review reveals the co-dependence of technical and domain expertise. Multiple studies highlight tensions between AI developers and domain experts, particularly in high-stakes settings. Practices of mutual learning and translation (e.g., interrogation, validation, and reframing of algorithmic outputs) emerge as critical mechanisms for productive integration (Lebovitz et al., 2021; Van den Broek et al., 2021). The boundary between AI creators and AI users is not fixed, but constantly reconfigured through interaction, negotiation, and power shifts.

Finally, the review brings attention to two types of risks. First, ethical risks, such as bias and fairness, require not just algorithmic solutions but sociotechnical governance –

including organizational accountability structures, delegation logics, and the alignment of virtue-based and duty-based ethics (Heyder et al., 2023; Teodorescu et al., 2021). Second, organizational learning risks emerge when AI systems reduce behavioral variance, limit experimentation, or produce overfit routines. While ML can alleviate learning myopia by freeing up human exploration (Sturm et al., 2021), it can also create blind spots and reduce organizational adaptability when not properly integrated (Balasubramanian et al., 2022).

The intent of this review is to present, categorize, and reflect on the key research trajectories that management and organization theories might consider exploring in order to contribute to explaining how AI can be seamlessly integrated in organizational strategic intent, routine, tasks, and decision processes. Recent reviews on the same topic have addressed contemporary AI challenges, but often with a narrower scope. Articles published since 2024 have focused on specific fields such as innovation management (Gama & Magistretti, 2025) or organizational behavior (Bankins et al., 2024), on a specific aspect such as trust (Montealegre-López, 2025) or AI-based augmentation (Baer et al., 2025), on a specific theoretical lens (e.g., stakeholder theory: Matthews et al., 2025), or address a specific problem in a specific industry (e.g., explainable AI in healthcare: Rosenbacke et al., 2024). The review shares methodological and conceptual affinity with Ramaul et al. (2025) who adopt a problematization approach to critically re-examine how AI is theorized in organization and management theories. Like this study, their review builds on a broad literature search strategy but limited to twelve top-tier journals. However, their focus is more explicitly directed toward challenging two specific assumptions in existing literature: the rationality attributed to AI systems and the anthropomorphic framing of AI as human-like.

This review presents several limitations. First, the disciplinary scope is confined to top-tier journals in selected field of studies. While this choice ensures conceptual depth and theoretical relevance (Hiebl, 2023), it may exclude perspectives from adjacent fields such as computer science, psychology, organizational behavior, behavioral economics, which increasingly contribute to the sociotechnical framing of AI. Given the interdisciplinary nature of the topic, such contributions may have offered additional insights. Moreover, excluding functional journals does not imply that relevant contributions for specific managerial functions are absent from the review, as domain-specific studies may be published in generalist management outlets. Second, the rapid evolution of AI technologies poses inherent challenges to the stability of conceptual assumptions. For instance, widely shared claims about the generic, explicit, and myopic nature of AI (Kemp, 2023) are already being questioned in light of developments in generative and agentic AI systems. From a methodological standpoint, including non-peer reviewed records such as commentaries or editorials or conference proceedings might have allowed for closer engagement with emergent developments not yet captured in peer-reviewed scholarship. Third, the use of a concept-centric approach required adopting analytically distinct clusters and assigning each article to a single thematic stream, even when several contributions clearly intersect multiple domains. While this choice was necessary to ensure analytical clarity and thematic coherence, it may underrepresent the multidimensionality and cross-cutting relevance of some papers, as later discussed in the synthesis. Moreover, alternative clustering schemes would have been possible. Finally, the screening and thematic clustering were conducted by a single researcher. Future research could enhance methodological robustness through multiple coders and formal assessment of inter-rater agreement.

Future avenues for research in healthcare organizations

While this review is not sector-specific, the findings hold significant relevance for the healthcare domain, where the integration of AI raises unique organizational, strategic, and professional challenges. Drawing on the eight conceptual streams outlined above, several research avenues emerge for scholarly investigation of AI implementation in healthcare organizations, which can be essentially grouped into strategic-organizational and professional-clinical levels.

At the strategic level, healthcare organizations operate in environments where the logic of competitive advantage is shaped not only by market dynamics, but also by institutional constraints, regulatory pressures, and professional norms. In systems where competition is driven by patient flows, AI could support strategic positioning through improved clinical performance. In more institutionally insulated systems, however, organizational attractiveness to professionals – rather than patients – may be the key differentiator. Since technology has already been identified as a driver of professional recruitment and retention (e.g., Compagni et al., 2015), future research could explore how AI-based systems affect organizational legitimacy, professional appeal, and reputation, especially in contexts where healthcare resources are scarce.

A second avenue concerns the organizational structures and roles that foster AI adoption. The literature reviewed points to the strategic role of AI-oriented leadership and new hybrid positions (e.g., CIO, AI leads, algorithmic brokers), as well as to AI-specific organizational units such as AI R&D centers (Davenport, 2018; Goto, 2023). Future studies could investigate how such roles and units emerge, how they coordinate with clinical structures, and how they mediate tensions between exploration and exploitation – an ambidexterity challenge particularly salient in highly regulated sectors. This aligns

with recent calls for understanding how AI-driven capabilities are built and orchestrated across traditional and emergent professional logics (Benbya et al., 2020; Kemp, 2023; J. Li et al., 2021).

At the professional level, healthcare offers a rich setting to investigate individual-level responses to AI. The review confirms that AI acceptance cannot be explained by acritical trust or resistance alone, but involves complex interactions between perceived system characteristics, professional expertise, and the epistemic standards of clinical reasoning (Anthony, 2021; Lebovitz et al., 2021). In this context, one line of inquiry could examine how validation and interrogation practices evolve across clinical tasks (e.g., diagnosis, prognosis, treatment optimization), and how they affect the actual use of AI-generated outputs. A second direction concerns the source of innovation: many AI tools used in healthcare are developed externally, and their successful integration depends on the quality of interaction between developers and users (Van den Broek et al., 2021; Singer et al., 2022). Research could explore how different types of suppliers – startups, academic spin-offs, multinational vendors – shape implementation outcomes compared to traditional medical device companies.

Finally, a third avenue involves clinician experience, beliefs and experiential backgrounds, which the review identifies as a core but still underexplored factor in shaping human–AI collaboration. The emerging evidence of skill-biased augmentation suggests that experienced professionals may benefit more from AI than junior ones but may also show higher levels of algorithm aversion or disengagement. Investigating how different types of experience – clinical, procedural, chronological – interact with AI across specialties and organizational settings could yield valuable insights into how to design equitable and effective augmentation strategies in healthcare.

APPENDIX 1

Table A1.1. List of the included journals

Source title	AJG Rank (2021)	Field	Retrieved articles (n.)	Articles included in the final sample (n.)
Academy of Management Annals	4*	General Management	5	1
Academy of Management Journal	4*	General Management	6	2
Academy of Management Review	4*	General Management	14	7
Administrative Science Quarterly	4*	General Management	3	1
Information Systems Research	4*	Information Systems Management	70	2
Journal of Management	4*	General Management	4	1
Journal of the Association for Information Systems	4*	Information Systems Management	30	3
Management Science	4*	Organization & Management Science	95	6
MIS Quarterly	4*	Information Systems Management	42	8
Operations research	4*	Organization & Management Science	62	0
Organization Science	4*	Organization Studies	14	5
Public Administration Review	4*	Public Management	7	3
Research Policy	4*	Innovation Management	28	3

Strategic Management Journal	4*	Strategic Management	22	3
European Journal of Information Systems	4	Information Systems Management	15	4
Human Relations	4	Organization Studies	3	1
Information Systems Journal	4	Information Systems Management	3	0
Journal of Information Technology	4	Information Systems Management	16	2
Journal of Management Information Systems	4	Information Systems Management	31	5
Journal of Management Studies	4	General Management	5	2
Journal of Public Administration Research and Theory	4	Public Management	8	1
Journal of Strategic Information Systems	4	Information Systems Management	6	5
Organization Studies	4	Organization Studies	5	0
Public Administration	4	Public Management	6	0
Public Management Review	4	Public Management	16	5

Figure A1.1. Flow chart of search for record inclusion

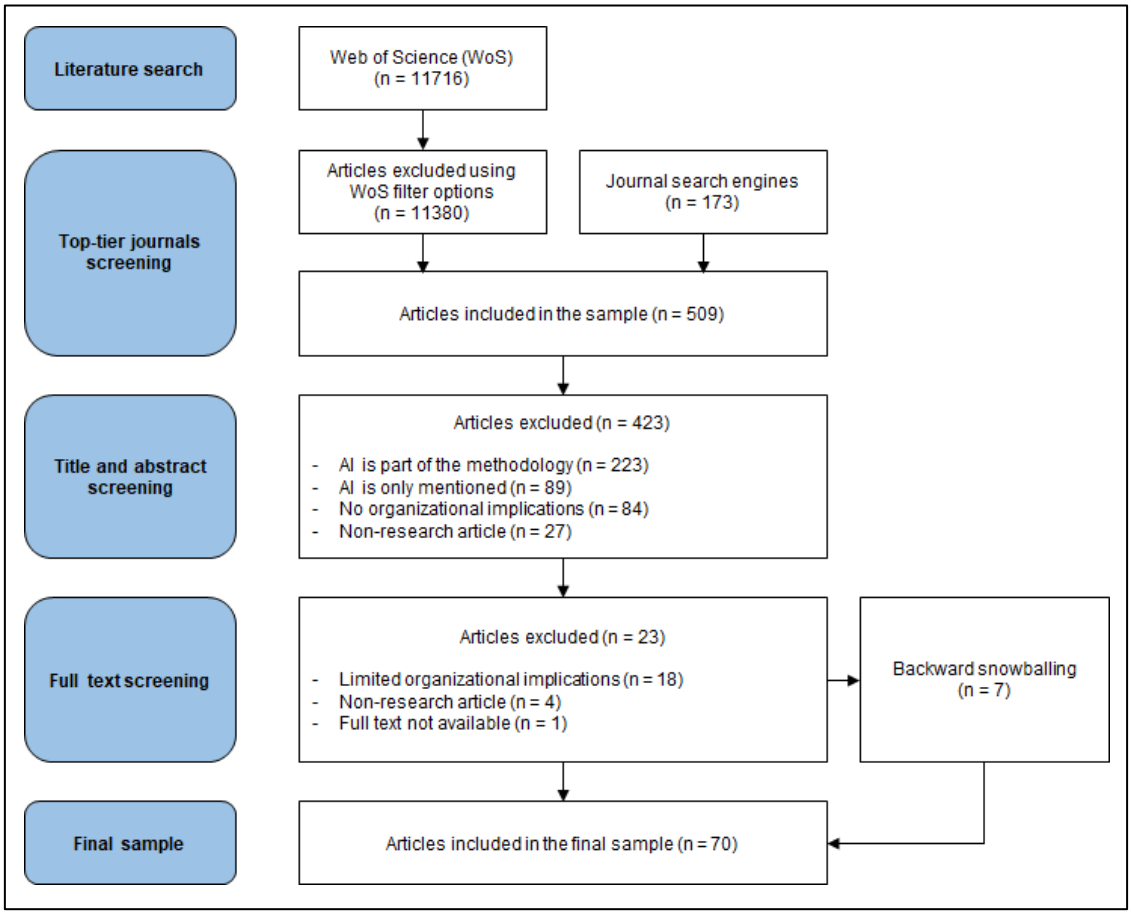


Table A1.2. Papers included

#	Authors	Title	Source	Year	Typology	Theme
1	Glikson & Woolley	Human trust in artificial intelligence: review of empirical research	Academy of Management Annals	2020	Review	Human-AI Interaction
2	Man Tang et al.	When Conscientious Employees Meet Intelligent Machines: An Integrative Approach Inspired by Complementarity Theory and Role Theory	Academy of Management Journal	2022	Empirical	Human-AI Configurations
3	Jia et al.	When and How Artificial Intelligence Augments Employee Creativity	Academy of Management Journal	2023	Empirical	Human-AI Configurations
4	Anthony (1)	To question or accept? How status differences influence responses to new epistemic technologies in knowledge work	Academy of Management Review	2018	Conceptual	AI & Knowledge and Professional Work
5	Gregory et al.	The role of artificial intelligence and data network effects for creating user value	Academy of Management Review	2021	Conceptual	AI & Strategic Capabilities
6	Murray et al.	Humans and technology: Forms of conjoined agency in organizations	Academy of Management Review	2021	Conceptual	Human-AI Configurations

7	Raisch & Krakowski	Artificial intelligence and management: the automation-augmentation paradox	Academy of Management Review	2021	Conceptual	Human-AI Configurations
8	Balasubramanian et al.	Substituting human decision-making with machine learning: implications for organizational learning	Academy of Management Review	2022	Empirical	AI & Organizational Learning
9	Kemp	Competitive Advantages through Artificial Intelligence: Toward a Theory of Situated AI	Academy of Management Review	2023	Conceptual	AI & Strategic Capabilities
10	Raisch & Fomina	Combining Human and Artificial Intelligence: Hybrid Problem-Solving in Organizations	Academy of Management Review	2024	Conceptual	Human-AI Configurations
11	Anthony (2)	When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies	Administrative Science Quarterly	2021	Empirical	AI & Knowledge and Professional Work
12	Jussupow et al.	Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence	Information Systems Research	2021	Empirical	AI & Decision-Making
13	Fügener et al. (1)	Cognitive Challenges in Human-Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation	Information Systems Research	2022	Empirical	Human-AI Configurations

14	Choudhary et al.	Human-AI Ensembles: When Can They Work?	Journal of Management	2023	Conceptual	Human-AI Configurations
15	Asatiani et al.	Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems	Journal of the Association for Information Systems	2021	Empirical	AI Adoption and Diffusion
16	Strich et al.	What Do I Do in a World of Artificial Intelligence? Investigating the Impact of Substitutive Decision-Making AI Systems on Employees' Professional Role Identity	Journal of the Association for Information Systems	2021	Empirical	AI & Knowledge and Professional Work
17	Schmitt et al.	The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution	Journal of the Association for Information Systems	2023	Conceptual	Human-AI Interaction
18	Dietvorst et al.	Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them	Management Science	2018	Empirical	AI & Decision-Making
19	Boyacı et al.	Human and Machine: The Impact of Machine Input on Decision Making Under Cognitive Limitations	Management Science	2023	Conceptual	AI & Decision-Making
20	de Véricourt & Gurkan	Is Your Machine Better Than You? You May Never Know	Management Science	2023	Conceptual	AI & Decision-Making

21	Liu et al.	Algorithm Aversion: Evidence from Ridesharing Drivers	Management Science	2023	Empirical	AI & Decision-Making
22	Wang QC et al.	Algorithmic Transparency with Strategic Users	Management Science	2023	Conceptual	Managing AI Ethical Challenges
23	Wang WG et al.	Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience	Management Science	2023	Empirical	AI & Knowledge and Professional Work
24	Baird & Maruping	The next generation of research on is use: A theoretical framework of delegation to and from agentic is artifacts	MIS Quarterly	2021	Empirical	Human-AI Configurations
25	Fügener et al. (2)	Will humans-in-the-loop become borgs? merits and pitfalls of working with ai	MIS Quarterly	2021	Empirical	Human-AI Configurations
26	Lebovitz et al. (1)	Is ai ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what	MIS Quarterly	2021	Empirical	AI & Knowledge and Professional Work

27	Li et al.	Strategic directions for AI: the role of CIOs and boards of directors	MIS Quarterly	2021	Empirical	AI & Strategic Capabilities
28	Lou & Wu	AI on drugs: can artificial intelligence accelerate drug development? evidence from a large-scale examination of bio-pharma firms	MIS Quarterly	2021	Empirical	AI & Strategic Capabilities
29	Sturm et al. (1)	Coordinating human and machine learning for effective organizational learning	MIS Quarterly	2021	Conceptual	AI & Organizational Learning
30	Teodorescu et al.	Failures of fairness in automation of human-ML augmentation	MIS Quarterly	2021	Empirical	Managing AI Ethical Challenges
31	van den Broek et al.	When the machine meets the expert: an ethnography of developing ai for hiring	MIS Quarterly	2021	Conceptual	AI & Knowledge and Professional Work
32	Pachidi et al.	Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing	Organization Science	2021	Empirical	AI Adoption and Diffusion
33	Allen & Choudhury	Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion	Organization Science	2022	Empirical	AI & Knowledge and

						Professional Work
34	Lebovitz et al. (2)	To Engage or Not to Engage with Al for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis	Organization Science	2022	Empirical	AI & Knowledge and Professional Work
35	Waardenburg et al.	In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms	Organization Science	2022	Empirical	AI & Knowledge and Professional Work
36	Anthony et al.	Collaborating with AI: Taking a System View to Explore the Future of Work	Organization Science	2023	Conceptual	AI Adoption and Diffusion
37	Vogl et al.	Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities	Public Administration Review	2020	Empirical	AI Adoption and Diffusion
38	Meijer et al.	Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems	Public Administration Review	2021	Empirical	AI Adoption and Diffusion

39	Selten et al.	'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment	Public Administration Review	2023	Empirical	Managing AI Ethical Challenges
40	Goto	Anticipatory innovation of professional services: The case of auditing and artificial intelligence	Research Policy	2023	Empirical	AI Adoption and Diffusion
41	Igna & Venturini	The determinants of AI innovation across European firms	Research Policy	2023	Empirical	AI Adoption and Diffusion
42	Dahlke et al.	Epidemic effects in the diffusion of emerging digital technologies: evidence from artificial intelligence adoption	Research Policy	2024	Empirical	AI Adoption and Diffusion
43	Choudhury et al.	Machine learning and human capital complementarities: Experimental evidence on bias mitigation	Strategic Management Journal	2020	Empirical	Managing AI Ethical Challenges
44	Tong et al.	The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance	Strategic Management Journal	2021	Empirical	Human-AI Interaction
45	Krakowski & Raisch	Artificial intelligence and the changing sources of competitive advantage	Strategic Management Journal	2023	Empirical	AI & Strategic Capabilities

46	Kordzadeh & Ghasemaghaei	Algorithmic bias: review, synthesis, and future research directions	European Journal of Information Systems	2022	Review	Managing AI Ethical Challenges
47	Rana et al.	Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness	European Journal of Information Systems	2022	Empirical	AI & Strategic Capabilities
48	Bankuoru Egala & Liang	Algorithm aversion to mobile clinical decision support among clinicians: a choice-based conjoint analysis	European Journal of Information Systems	2023	Empirical	AI & Decision-Making
49	Vassilakopoulou et al.	Developing human/AI interactions for chat-based customer services: lessons learned from the Norwegian government	European Journal of Information Systems	2023	Empirical	Human-AI Configurations
50	Einola et al.	A colleague named Max: A critical inquiry into affects when an anthropomorphised AI (ro)bot enters the workplace	Human Relations	2023	Empirical	Human-AI Interaction
51	Engel et al.	Stairway to heaven or highway to hell: A model for assessing cognitive automation use cases	Journal of Information Technology	2023	Empirical	AI & Strategic Capabilities
52	Vannuccini & Prytkova	Artificial Intelligence's new clothes? A system technology perspective	Journal of Information Technology	2023	Conceptual	AI Adoption and Diffusion

53	Chandra et al.	To Be or Not to Be Horizontal Ellipsis Human? Theorizing the Role of Human-Like Competencies in Conversational Artificial Intelligence Agents	Journal of Management Information Systems	2022	Empirical	Human-AI Interaction
54	You et al.	Algorithmic versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation?	Journal of Management Information Systems	2022	Empirical	AI & Decision-Making
55	Dennis et al.	AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With	Journal of Management Information Systems	2023	Empirical	Human-AI Configurations
56	Qin et al.	Perceived Fairness of Human Managers Compared with Artificial Intelligence in Employee Performance Evaluation	Journal of Management Information Systems	2023	Empirical	Managing AI Ethical Challenges
57	Revilla et al.	Human-Artificial Intelligence Collaboration in Prediction: A Field Experiment in the Retail Industry	Journal of Management Information Systems	2023	Empirical	Human-AI Configurations
58	Faulconbridge et al.	How Professionals Adapt to Artificial Intelligence: The Role of Intertwined Boundary Work	Journal of Management Studies	2023	Empirical	AI & Knowledge and Professional Work
59	Pakarinen & Huising	Relational Expertise: What Machines Can't Know	Journal of Management Studies	2023	Review	AI & Knowledge and

						Professional Work
60	Alon-Barkat & Busuioc	Human–AI Interactions in Public Sector Decision Making: “Automation Bias” and “Selective Adherence” to Algorithmic Advice	Journal of Public Administration Research and Theory	2023	Empirical	Managing AI Ethical Challenges
61	Coombs et al.	The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review	Journal of Strategic Information Systems	2020	Review	Human-AI Configurations
62	Grønsund & Aanestad	Augmenting the algorithm: Emerging human-in-the-loop work configurations	Journal of Strategic Information Systems	2020	Empirical	Human-AI Configurations
63	Shollo et al.	Shifting ML value creation mechanisms: A process model of ML value creation	Journal of Strategic Information Systems	2022	Conceptual	AI & Strategic Capabilities
64	Heyder et al.	Ethical management of human-AI interaction: Theory development review	Journal of Strategic Information Systems	2023	Review	Managing AI Ethical Challenges
65	Sturm et al. (2)	Machine learning advice in managerial decision-making: The overlooked role of decision makers' advice utilization	Journal of Strategic Information Systems	2023	Empirical	AI & Decision-Making

66	Giest & Klievink	More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts	Public Management Review	2022	Empirical	AI Adoption and Diffusion
67	Maragno et al.	AI as an organizational agent to nurture: effectively introducing chatbots in public entities	Public Management Review	2022	Empirical	Human-AI Configurations
68	Neumann et al.	Exploring artificial intelligence adoption in public organizations: a comparative case study	Public Management Review	2022	Empirical	AI Adoption and Diffusion
69	Dickinson & Yates	From external provision to technological outsourcing: lessons for public sector automation from the outsourcing literature	Public Management Review	2023	Review	Human-AI Configurations
70	Wang G et al.	Artificial intelligence, types of decisions, and street-level bureaucrats: Evidence from a survey experiment	Public Management Review	2024	Empirical	AI & Decision-Making

CHAPTER 2

EXPLORING AI INNOVATION DYNAMICS: A MULTIPLE CASE STUDY IN HEALTHCARE ORGANIZATIONS[2]

1. INTRODUCTION

AI is increasingly recognized as a general-purpose technology (GPT), with the potential to affect a wide range of sectors and organizational domains. Like other GPTs, AI does not simply enable isolated applications but provides a technological foundation that can generate both radical innovation of products and services and incremental – though often dramatic – improvements in existing offering (Grashof & Kopka, 2023; McAfee, 2024). The adoption of AI technologies is rarely plug-and-play. AI technologies are increasingly placing organizations before the choice to develop AI solutions internally, to co-create

[2] This chapter is part of a research conducted with Vittoria Ardito, Giulia Cappellaro, Amelia Compagni, and Francesco Petracca. The research was approved by the Bocconi University's Ethics Committee (n. RA000945).

them through alliance, or to acquire them from external providers (Hoffreumon et al., 2024). Being able to make this decision requires substantial organizational adjustments, including the reconfiguration of infrastructures, the acquisition of new competences, the redesign of routines, and the establishment of governance mechanisms capable of dealing with uncertainty and complexity.

AI innovation is still a nascent and rapidly evolving field that calls for distinguishing features compared to more established domains of technology and digital innovation (Ångström et al., 2023; Hopf et al., 2023; Lin & Maruping, 2025). Recent research has begun to identify some of these distinctive elements: AI is increasingly described in terms of its system-ness, where outcomes are shaped by the interplay of multiple actors (system builders, regulators, providers, and adopters) (Vannuccini & Prytkova, 2023); AI innovation is strongly reliant on networks and relational embeddedness, as knowledge and practice diffuse across organizations through collaborative ties and professional communities (Dahlke et al., 2024); specific AI organizational capabilities are emerging, highlighting the need for organizations to orchestrate data assets, technical skills, and organizational routines in novel ways to turn AI into a source of learning, innovation, and competitive advantage (Gregory et al., 2021; Lou & Wu, 2021; Kemp, 2023; Krakowski et al., 2023; Engel et al., 2024). In dealing with innovation, organizations have long confronted the challenge of fostering novelty while maintaining operational efficiency. The ambidexterity literature captures this enduring dilemma by distinguishing between exploration, which entails experimentation, search, and adaptation, with exploitation, which emphasizes refinement, efficiency, and incremental improvement (March, 1991; Raisch & Birkinshaw, 2008; AlKhamees & Durugbo, 2024). We draw on these perspectives as complementary lenses to investigate how organizations respond to the

distinctive challenges of AI innovation. Specifically, we want to address the following research question: *How do organizations operating in complex environments configure and govern AI innovation activities?*

Healthcare organizations provide the boundary conditions for our investigation. AI is widely acknowledged as a potentially transformative driver for healthcare organizations and systems, both in supporting clinical decision-making and improving operational efficiency (Rajpurkar et al., 2022; Topol, 2019; Wachter & Brynjolfsson, 2024). Healthcare organizations are accustomed to operate in a complex environment populated by professional dynamics, specific regulations on how clinical innovation is adopted and implemented, and strong institutional constraints and pressure (Barlow, 2017). Given their knowledge-intensive nature, healthcare organizations are under constant pressure to innovate. In everyday practice, however, innovation is largely vendor- and institution-led. Pharmaceutical, medical devices, and IT companies, as well as public agencies and regulatory authorities, introduce new technologies, while service-delivery organizations mainly assess effectiveness and economic or organizational impact before adoption. The cases in which organizations complement this habitual exploitative stance with proactive exploratory efforts are much rarer and therefore more revealing.

By relying on a multiple case study approach with healthcare organizations operating in the Italian context, we examined how such organizations orchestrate and govern AI innovation. Through our analysis, we identified four tensions that shape how AI innovation is configured across organizational levels and beyond organizational boundaries. These tensions concern (i) how AI exploration is legitimized and organized, (ii) how internal and external competences are balanced, (iii) how organizations navigate the make-buy-ally continuum, and (iv) how internal demand for AI innovation is

recognized and managed. These tensions give rise to distinct multilevel configurations through which AI innovation is structured in professional, knowledge-intensive, and highly regulated contexts.

The remainder of the paper is organized as follows. Section 2 introduces the conceptual background of the study; Section 3 details the research design; Section 4 reports empirical evidence through single case histories; Section 5 discusses the emergent framework in terms of tensions; Section 6 concludes with contribution to literature and directions for future research.

2. THEORETICAL BACKGROUND

2.1.Managing AI innovation

Managing AI innovation is a topic on a nascent stage, with growing attention that organizations and scholars devote to AI. While the literature on innovation, technology management, and IT innovation offers rich frameworks for understanding how organizations generate and implement new ideas, recent research stresses that AI innovation differs substantially from these established traditions. Treating AI simply as another form of technology or IT innovation has proved misleading, as AI artifacts embody distinct properties which challenge established way of organizing innovation and drive to develop specific conceptualization that departs from conventional innovation models and practices (Berente et al., 2021; Ångström et al., 2023; Hopf et al., 2023; Lin & Maruping, 2025).

Among these properties, two at least are worth mentioning. First, already with the rise of digital technologies, scholars noted how the locus of innovation was progressively moving beyond organizational boundaries, as digitalization has been associated with

open, networked, and collaborative forms of innovation (Nambisan et al., 2017). AI innovation amplifies this shift, exhibiting a pronounced ecosystemic character for its reliance on diverse actors, infrastructures, and data sources. In this domain, progress rarely stems from isolated organizations, but rather from the alignment of multiple system builders, AI creators, integrators, operators and simple takers, who shape the trajectory of AI development and implementation (Vannuccini & Prytkova, 2023). This implies that managing AI innovation entails navigating interdependencies across extended value chains and coordinating knowledge flows through dedicated intermediaries. Indeed, studies highlight the role of algorithmic brokers, hybrid organizational units, and inter-organizational ties in mediating AI adoption and enabling the transfer of knowledge across boundaries (Waardenburg et al., 2022; Dahlke et al., 2024).

A second distinctive feature of managing AI innovation concerns the nature of AI innovation itself. While unanimously portrayed as a disruptive and general-purpose technology (McAfee, 2024), evidence so far shows that in organizational contexts AI innovation tends to take form of incremental and process-oriented change. Rather than creating radical breakthroughs in organizations, many AI initiatives reconfigure processes, automate existing tasks, or augment established routines (Grashof & Kopka, 2023; Lin & Maruping, 2025). Moreover, AI artifacts are not neutral carriers of innovation, as their functioning and organizational implications are shaped by assumptions embedded in training data, modeling choices, and developer perspectives (Lebovitz et al., 2021; Anthony et al., 2023). This combination of process orientation, incrementalism, and embedded assumptions differentiates AI innovation from conventional models of technology and IT innovation, where artifacts were often treated as more transparent and transferable across settings.

A recurring theme in recent research portrays managing AI innovation as a field of organizational tensions. First, AI projects often oscillates between exploratory craft-like approaches embraced by data scientist versus standardized, mechanical logics favored by managers (Hopf et al., 2023). Second, organizations face the dilemma of allocating effort between internal knowledge search, leveraging proprietary data and domain expertise, and external knowledge search, often conducted through sustained collaboration with selected partners (Dahlke et al., 2024; Lin & Maruping, 2025). In the tension between domain and technical expertise, organizations are called to cope with the dilemma between separating technical expertise in separate units vis-à-vis integrating them into existing organizational structures (Selten & Klievink, 2024). Moreover, knowledge practices reveal persistent frictions between assessing know-what – validating algorithmic performance through quantitative measures – and know-how, the tacit knowledge that professionals rely on to ensure contextual adequacy (Lebovitz et al., 2021). The latter also brings forward questions of balancing centralization and formalization of decision-making in dedicated AI units with the need of local autonomy and professional discretion (Goto, 2023). Finally, at the strategic level the diffusion of AI highlights a tension between incumbent advantages and entrant agility. While incumbents benefit from data network effects and leverage AI through situated practices (Gregory et al., 2021; Kemp, 2023), new entrant AI firms leverage flexibility and speed to experiment with novel and cutting-edge applications (Jacobides et al., 2021; Vannuccini & Prytkova, 2023).

At the intra-organizational level, research identifies a set of practices and arrangements that organizations adopt to navigate these tensions. Process and knowledge management routines emerge as critical for enabling the iterative development and scaling of AI

solutions, helping balance between craft and mechanical work (Pachidi et al., 2021; Lin & Maruping, 2025). Centralized and formalized structural arrangements, including dedicated AI units or centers, provide legitimacy and resource alignment while addressing the tension between centralization and local autonomy (Goto, 2023). Algorithmic brokers and hybrid units mediate between domain and technical expertise, facilitating integration across knowledge boundaries (Waardenburg et al., 2022). Intensifying both internal knowledge search – leveraging local data, domain expertise, and contextual knowledge – and deep external search through sustained collaboration with selected partners responds to the challenge of balancing proprietary expertise with internal input (Lin & Maruping, 2025). Finally, cultural readiness, also rooted in prior experience in digital innovation, further helps to facilitate the assimilation of AI initiatives in specific organizational contexts, mitigating the risk of generic rather than situated adoption (Igna & Venturini, 2023; Neumann et al., 2024).

2.2.Ambidexterity in managing innovation

Several of the tensions identified in managing AI innovation recall a longstanding dilemma in Organization and management theory: the balance between exploration and exploitation (March, 1991; Raisch & Birkinshaw, 2008). Some tensions – such as internal versus external knowledge search and separation versus integration of technical expertise – map directly onto the classical ambidexterity dilemma described in organization theory. Others, such as the tension between generic versus situated AI, resonate with the trade-off between exploiting established routines for introducing innovation and exploring how to embed AI into idiosyncratic organizational processes and context. Against this backdrop, ambidexterity provides a particularly suitable conceptual lens to analyze how organizations navigate competing demands for efficiency and novelty.

Organizational ambidexterity has emerged as a foundational concept in the strategy and innovation literature, capturing the dual capacity of organizations to engage simultaneously in both explorative and exploitative activities (Duncan, 1976; March, 1991; Raisch & Birkinshaw, 2008; Turner et al., 2013). Although complementary in nature, exploration and exploitation compete in priorities and resources that organizations have to balance over time (Birkinshaw & Gupta, 2013; Raisch & Birkinshaw, 2008; Simsek, 2009). Achieving ambidexterity hence require the ability to manage the inherent tension between alignment with established processes and flexibility to respond to emerging opportunities (Gibson & Birkinshaw, 2004; O'Reilly & Tushman, 2013). A key advantage of ambidexterity lies in its contribution to organizational adaptability and long-term competitiveness, as organizations can both respond proactively to rapid environmental changes – such as those observed in AI-related technological and market changes – and continuously refresh their knowledge and capabilities by combining both stances (Junni et al., 2013; X. Li et al., 2025).

A substantial body of literature has explored the organizational antecedents and enabling conditions of ambidexterity (Gibson & Birkinshaw, 2004; Kassotaki, 2022; Saleh et al., 2023; Chakma et al., 2024). Given its complexity and resource intensity, ambidexterity requires organizations to make strategic choices in how they allocate effort and attention. One recurring mechanism is the externalization of one logic – typically via outsourcing or inter-organizational partnerships – while focusing internal efforts on the other. Such collaborations serve as a channel for acquiring complementary knowledge and extending capabilities (Holmqvist, 2004; Hwang et al., 2023). From a knowledge-based perspective, ambidexterity depends on the organization's ability to acquire, process, and recombine knowledge from diverse sources (Soto-Acosta et al., 2018). Similarly, organizational

learning capabilities – including mechanisms for intra-organizational learning, external knowledge exchange, and cultural openness to experimentation – are foundational to sustaining ambidexterity over time (Farzaneh et al., 2022; Jurksiene & Pundziene, 2016). Process-related enablers also play a critical role: integrating mechanisms help manage alignment between conflicting demands, while involvement mechanisms – such as supplier engagement, employee participation, and customer co-creation – enhance the exchange of information and foster shared ownership (Cao et al., 2009; Mom et al., 2019).

To structure these enabling conditions, three main approaches are often distinguished: structural, contextual, and sequential ambidexterity (Birkinshaw et al., 2016; Stelzl et al., 2020; Gibson & Birkinshaw, 2004; Maclean et al., 2021). The first suggests creating distinct organizational architectures dedicated respectively to exploration or exploitation, thus allowing them to operate under their own logic (O'Reilly & Tushman, 2013). The second emphasizes building a shared context within a single business unit, in which individuals are empowered – through processes, incentives, and leadership – to alternate between logics as needed (Gibson & Birkinshaw, 2004; Stelzl et al., 2020). The latter involves deliberately alternating between exploration and exploitation over time (Birkinshaw et al., 2016).

Finally, organizational context and culture, including values, norms, and leadership orientation, play a mediating role in shaping ambidextrous behavior. Culture that supports experimentation and promotes knowledge sharing are better positioned to reconcile the dual demands of innovation (Bodwell & Chermack, 2010; Nosella et al., 2012).

3. METHODOLOGY

The research is based on a qualitative multiple case study design, aimed at capturing the variety of approaches adopted by healthcare organizations in developing and implementing AI solutions (Yin, 2014). Case selection is based on a combination of theoretical sampling (Patton, 2002), focusing on organizations that adopt approach AI innovation through ambidexterity – combining exploration (internal development or co-development of AI solutions) with exploitation (routinary acquisition of AI solutions from external commercial providers). The cases were also chosen to reflect different organizational approaches to ambidexterity, ranging from more structural-like arrangements (e.g., dedicated units or formal separation activities) to more contextual-like arrangements (e.g., integration of AI activities within existing structures and routines).

To mitigate the risk of informant bias, participants were selected among key informants with strategic or operational roles in the development or adoption of AI solutions (Eisenhardt & Graebner, 2007). After the first contact, additional informants have been identified through snowball sampling. Informants include CIOs, Scientific Directors, Data Scientists, Directors of units involved in AI-related activities, and members of top management. All interviews were recorded with permission, transcribed and coded using ATLAS.TI.

Data collection primarily relied on semi-structured expert interviews (Flick, 2022). The interview protocol was structured around dimensions derived from our theoretical lens: (1) *antecedents* and *rationales* of exploration, including its co-existence with exploitation; (2) *resources, competences,* and *organizational implication* of exploration activities (see Appendix 2). The first section aims to explore both the rationale and the

process behind the decision to pursue exploration activities in the field of AI, as well as how co-existence with exploitation activities is managed. The second section explores which human, technological, financial, and organizational resources and units have been employed to establish exploration activities, including the potential externalization strategies adopted, and the impact that exploration activities had on the organizational routine. Interviews were conducted between April and July 2025. Secondary data included documents, websites, and news articles, used to triangulate and contextualize interviews.

We empirically explored how healthcare organizations manage ambidexterity in AI adoption in four case studies in Italy, which constitute the units of analysis. Healthcare organizations provide an appropriate setting for this investigation. While AI is widely acknowledged as a potentially transformative driver for the sector, they face strong institutional and regulatory constraints and are characterized by knowledge-intensive and professional dynamics, making the balance between exploratory and exploitative approaches to innovation particularly salient (Barlow, 2017; Topol, 2019; Wachter & Brynjolfsson, 2024). Focusing on one country enabled us to control macro-institutional, cultural, and regulatory factors, while allowing comparison of organizational approaches to ambidexterity in AI adoption.

Table 2.1 summarizes case characteristics as well as data sources. The four cases are comparable in terms of size (e.g., hospital beds and revenues) and teaching & research orientation but differ in institutional nature (public vs. private). More importantly, they display variation in how they approach ambidexterity, with two adopting structural-like and two contextual-like configurations.

Table 2.1. Case characteristics and interview sources.

	Case Alpha	**Case Beta**	**Case Gamma**	**Case Delta**
Hospital Beds	700+	1,000+	1,400+	1.300+
Revenues (mln €)	€ 600+	€ 650+	€ 800+	€ 800+
Type of organization	Private Teaching & Research Hospital	Private Teaching & Research Hospital	Private Teaching & Research Hospital	Public Teaching & Research Hospital
Approach to exploration	Structural-like	Structural-like	Contextual-like	Contextual-like
Interviews (n.)	3	1	2	1
I1	CIO (65 min)	Digital Health Manager (70 min)	COO (43 min)	Head of Analytics & OR Unit (54 min)
I2	Head of AI R&D (57 min)	-	Deputy Scientific Director (48 min)	-
I3	Head of Hematology Unit (50 min)	-	-	-

We adopted an iterative analytical strategy (Eisenhardt, 1989). First, we developed case histories for each organization. Second, we conducted within-case coding of interviews to identify emerging categories related to exploration. Third, we engaged in cross-case comparison, using replication logic to contrast structural-like and contextual-like approaches. Finally, we iterated between emerging insights and extant literature on ambidexterity and AI innovation, progressively refining our theoretical contribution. Our outcome of interest is not organizational performance per se, but the emergent configurations through which healthcare organizations organize exploration activities and combine them with exploitation.

4. CASE HISTORIES

4.1.Organization Alpha

Organization Alpha is the first and largest hospital of a private healthcare group operating primarily in northern Italy. It originates from a broader industrial conglomerate that established a research-oriented hospital that gradually evolved into a teaching hospital. This industrial heritage fostered an early emphasis on technological innovation and digital autonomy, particularly in supporting hospital operations and user access. Such commitment was exemplified by the creation of a proprietary software factory that developed internal digital solutions (e.g., software, mobile applications, and web-based services), well before the formal adoption of AI.

In 2019, the hospital management launched a dedicated strategic initiative that embedded AI in its research orientation, prior experience in digital development, and rich internal knowledge base. As the Group CIO and Center Director (A-I1) recalled, *"In 2019 my perception, and also that of the CEO and the Board at the time, was that big data and AI were topics that were coming, like when people used to talk about digital, the web, e-commerce. It was a keyword, a buzzword. So, we said: «let's try, let's understand what it is». For us it was a natural evolution."* A dedicated AI Research & Development Center was established to cover the entire cycle of innovation: from AI and clinical research to bedside implementation. The Center is substantially self-financed through national and international grants for basic, applied, and industrial research; it operates with relative autonomy within the hospital and the group. The participation to research grants and the existing clinical partnerships positioned the Center internationally by joining large research consortia.

The Center manages about fifty projects across different domains, with activities spanning diagnostic imaging, generative synthetic data, federated learning, and small language models (SLM). Clinical applications have concentrated initially on areas of high unmet clinical need, such as rare diseases, where the heterogeneity of data and the scarcity of commercial solutions provided fertile ground for exploration. These initiatives illustrate how exploration is triggered by unresolved clinical questions and the need to integrate multimodal data to support precision medicine. In the medium-term, the Center is working on the development of digital twins[3] as orchestrators of multiple models to enable individualized care.

The Center strongly supports the hospital in applying a systematic make-or-buy evaluation of both clinical and administrative AI initiatives. In the clinical domain, off-the-shelf commercial products – especially certified medical devices – are adopted when adequate. They typically require additional internal development to adapt them to organizational specificities and ensure professional acceptance. As the Head of AI R&D (A-I2) explained: *"A board has been set up within [the hospital], and it is responsible for evaluating these solutions, deciding whether to proceed with the make or buy question, and how to implement them."* In strategically distinctive areas or when no adequate commercial solutions are available, the Center pursues a make approach, leveraging in-house resources. In the administrative domain, the preference of internal development is stronger, given its strategic relevance of process knowledge, data privacy, and the possibility to use solutions as levers for organizational change. Some clinical models have been adapted for administrative use, further underlying the interdependence between the

[3] A digital twin is a virtual representation of a physical object (an organ, a physiological system, or a patient) that accurately reflects its real-world counterpart's behavior, performance, and conditions.

two domains. I1 explained: *"What many hospitals need to understand is that when you outsource something that is too close to your fundamental know-how, you lose control. That's why we keep 4–5 projects as pure make: they are too close to our core, so they must never be done outside. […] Where there is strong IP, something really relevant, you make."*

Organizationally, the Center embodies a structure-like approach to exploration, with an autonomous and dedicated unit of about fifty professionals – including engineers, developers, researchers, and project managers – organized in three areas (funding and partnership scouting, research and development, business development). Whitin R&D a sub-group is dedicated to cutting-edge research, while another focuses on engineering and software development. At the same time, exploration is not isolated, as the Center collaborates closely with hospital clinical and administrative departments (IT, Procurement, etc.) and external partners (other hospitals, research centers, industrial partners, etc.). A multidisciplinary evaluation board is being established to govern project assessment and prioritization, ensuring alignment with professional needs, internal resources, and organizational strategies. As the Head of the hematology unit (A-I3) said: *"I think that having a multidisciplinary team has enormous intrinsic value. Each professional continues to express their own expertise as physician or technician, but in an environment that facilitates interaction and a kind of osmosis of knowledge. […] It is not so much about transferring competences, but about ensuring efficient communication and entering a shared development strategy, each according to their competence and vision."* The Center trajectory reflects the hospital's longstanding innovation orientation, but also its structural dilemmas. While the creation of a separate Center highlights the

organizational choice to separate exploration, clinicians and managers emphasize the need for integration into practice and routines.

4.2.Organization Beta

Organization Beta is a large, independent, not-for-profit hospital located in central Italy, with a longstanding tradition in both research and teaching. Its research orientation was reinforced in 2016, when the hospital initiated a certification process as a research hospital[4], completed two years later. This milestone provided the trigger to reorganize the structures supporting scientific activities, particularly in the domain of data-driven research. During this period, the hospital consolidated its dispersed data science competences into a centralized center of expertise, housed within the scientific function and today employing more than thirty professionals engaged in data management, analytics, and support clinical and translational research.

At the same time, the hospital created a dedicated open innovation unit under the Scientific Directorate. Its mandate was to go beyond traditional technology transfer by engaging in co-creation and co-development initiatives with industrial partners, thereby combining the hospital's research mission with the ability to integrate capabilities not typically available within a clinical research organization. The open innovation unit expanded rapidly, supported by research and industrial grants. Over time, this growth led to its transformation into an autonomous legal entity with a clear strategic mandate from the hospital's holding. As the hospital's Digital Health Manager and spin-off's CEO (B-I1) said, *"This supply chain does not exist in a medical research institute, because it requires competences ranging from engineering to mathematics, physics, and*

[4] The Italian National Health Service recognizes research excellence by conferring the title of IRCCS (*Istituto di Ricovero e Cura a Carattere Scientifico*).

computational sciences. […] In a hospital IT department, computer science is almost never there. So, we thought that if we wanted to develop a virtuous path in digital medicine, it required cross-fertilization. We thought that open innovation was a useful paradigm in this sense." This spin-off organization operates with economic independence, sustained by external funding and commercial activities through the valorization of internally developed digital health products and services. The spin-off accompanies solutions until they reach pre-industrial maturity, including regulatory approval as medical devices, before commercialization is entrusted to external partners with industrial scaling capacity. AI has a relevant but not exclusive focus in these activities. Early projects involved ML for process optimization and federated learning across clinical centers. More recent activities focus on intelligent alert management systems and phenotyping of chronic diseases. Generative AI is currently less present due to regulatory uncertainty, although it is expected to play an influential role as frameworks evolve.

Decision-making around whether to develop or buy is guided by two central criteria. First, intellectual property (IP) ownership: all projects in which the hospital intends to retain full or partial IP are managed exclusively by the spin-off, while those sponsored by external providers or involving commercial AI products are typically overseen by internal research and data science facilities. Second, organizational adaptability: off-the-shelf solutions are preferred when they can be easily integrated into existing workflows with limited disruption, while exploratory projects are prioritized when external products are unlikely to fit the complexity of the organization's needs.

From an organizational standpoint, the spin-off mirrors a structural-like configuration, with a small top management team combining expertise in clinical regulation, digital

health, IT, and data protection, complemented by functions for legal affairs, account management, and communication, as well as a group of engineers. These act as brokers between internal stakeholders and external partners, ensuring the feasibility of initiatives in open innovation. As B-I1 recalled, *"Here we valorize all competences, including clinical ones. If company X comes and wants to develop a digital therapy for hypertension, we call our clinicians to tell us which strategies could be followed and then translate them into the digital ecosystem."* At the same time, bottom-up initiatives are increasingly frequent, as clinicians and researchers increasingly propose exploratory projects based on gaps identified in clinical practice, signaling an ongoing integration with the broader professional community.

Organization Beta illustrates an ambidextrous approach to digital and AI innovation that blends structural-like and contextual-like features. While the spin-off provides a formalized vehicle for exploration and IP retention, the integration with internal data science facilities and the active involvement of clinicians ensure that innovation is we connected with clinical and organizational needs. This dual arrangement reflects the hospital's broader attempt to balance autonomy and integration, internal control and external collaboration.

4.3.Organization Gamma

Organization Gamma is a private hospital in northern Italy, part of a larger hospital group but operating with a high degree of strategic and operational autonomy. It has a strong basic and clinical research orientation, coupled with a consolidated role as a teaching hospital. The hospital's experience with AI originated began with bottom-up initiatives during the Covid-19 pandemic, when predictive tools for diagnostic imaging were

developed to strategy the risk of disease progression. This experience served as a catalyst, reinforcing the conviction that AI development must be aimed at clinical impact.

Building on this experience, the hospital formalized its approach by creating a dedicated infrastructure for managing the full cycle of data and algorithm development. A proprietary platform was developed in collaboration with a major technological provider. This complex yet user-friendly architecture integrates access to all hospital data for clinical and research purposes and enables the generation, evaluation, and validation of predictive models through a fully traceable and automated pipeline, ensuring replicability explainability, and regulatory compliance. As the Deputy Scientific Director (G-I2) explained, *"We built a platform together with a working model, an infrastructure that was also the subject of some publications and certification. […] It was an ad hoc process to develop AI models that, if they perform as intended, are really ready for certification, because everything is done in a highly professional way – traceable, controllable, replicable, and explainable".* Over time, the platform has supported around twenty projects. Project ideas arise both from clinicians' needs and from research opportunities, such as funded grants. All proposals undergo a standardized assessment procedure evaluating data quality and validity, economic sustainability, and clinical relevance. Projects meeting these criteria proceed to regulatory definition and implementation. Alongside these platform-driven initiatives, vertical initiatives have been pursued on both the clinical and administrative domains. Governance of AI projects reflects a hybrid model. A steering committee composed of academic and operational figures guides project selection. Pragmatic criteria guide the choice between make and buy: the hospital tends to internally develop less critical modules, while for clinically sensitive solutions requiring rapid deployment it often relies on external providers. As the COO (G-I1) noted,

"We took a make approach for less sensitive projects (for example, marketing tools, where even if something goes wrong it has little impact). We went instead for buy in the more sensitive applications, both for the supplier's accountability and because at this stage the hospital still lacks a fully structured development team […]. And also for managerial reasons – speed, startup logic, showing stakeholders that it was an investment capable of bringing quick returns." The intention, however, is to progressively internalize some of the currently purchased components, as internal competences grow and the platform expands.

The core AI team consists of about fifteen professionals with backgrounds in computer science, engineering, mathematics, and physics, working in close contact with clinicians and largely embedded in the affiliated university departments and centers. The hospital provides access to data and the clinical environment, while investing in managerial expertise with backgrounds in digital transformation and AI to facilitate deployment in operational contexts.

The hospital's AI strategy today unfolds along two parallel trajectories. On one side, generative AI is being applied to both administrative and clinical functions, with active projects on clinical assistants, automated documentation, and report mining. On the other, AI is integrated into research and care pathways, with solutions remaining experimental until validation and certification are complete. Partnerships play a leading role: beyond collaborations with major technological providers, the hospital co-develops with start-ups to ensure flexibility and rapidity.

This organizational model strongly relies on bottom-up approaches, with priorities increasingly emerging from clinicians' needs and the most repetitive and labor-intensive

processes. At the same time, executives expect a return to more strategic trajectories, where processes will be radically rethought. In this scenario, technological governance – such as the choice between cloud or on-premises development – will become decisive in scaling AI sustainably and securely.

Organization Gamma thus illustrates a contextual-like approach to exploration, where exploratory activities are closely integrated into both scientific and operational units. By aligning clinical, academic, and managerial logics, the hospital balances the demands of research-driven exploration and operational efficiency, while relying on partnership to accelerate implementation when required.

4.4.Organization Delta

Organization Delta is a major public teaching and research hospital in central Italy. Deeply embedded in its regional university and healthcare environment, it combines a strong clinical mission with an increasing orientation toward research and innovation. Its strategic positioning has been reinforced by its recent transformation into a research hospital, which has reshaped its profile toward tighter integration of clinical care and scientific knowledge production. As the Head of Analytics and Operation Research (D-I1) recalled, *"Right after the pandemic, with a process that had already begun, the hospital became a research institute. The drive toward research, which before was mainly university-driven, became a positive tension – an institutional role recognized by the Ministry of Health."*

The roots of its current AI strategy can be traced back to 2017, when the hospital signed an agreement with the university to fund joint research fellowships in applied data science and operations research. This initial step was followed in 2021 by the creation of a joint

laboratory with the university's AI center, enabling the funding of PhD positions on mathematical modeling, operations research, and early applications of AI in clinical pathways. The COVID-19 pandemic then acted as an accelerator: the development of predictive models to manage hospital capacity highlighted the centrality of data as a strategic asset. This experience laid the foundation for the creation, in 2022, of a dedicated unit for clinical and research data management, later integrated with the traditional IT and data management functions. By 2024, this reorganization resulted in a single unit responsible for the full data cycle, from collection to analysis to integration with clinical and decision-making systems.

The hospital's AI initiatives today span multiple levels and are supported both by internal projects and by national programs. A new data platform is under development to enable integration of reporting, modern programming languages, and data-lake functionalities. Among ongoing projects, several include the development of clinical decision support systems, predictive models for bed management, and advanced medical imaging applications. Clinicians are frequently initiators or co-initiators of these projects, ensuring both practical relevance and smoother adoption.

In terms of make or buy choices, the hospital favors internal development or co-development with the university, rather than adopting off-the-shelf products. Projects follow an incremental logic of prototyping, testing, and validation, with particular emphasis on explainability and controlled deployment. Around fourteen professionals currently work in the dedicated data structure, including five data scientists, collaborating closely with ICT, clinical engineering, and other research departments. This configuration reinforces interdisciplinarity and leverages the hospital's academic partnerships to train

new hybrid professionals able to combine computational, clinical, and regulatory expertise.

The hospital is also active in institutional collaborations. It participates in regional AI working groups promoted by the regional government, which involve hospital managers, ICT directors, medical physicists, and universities to build a shared governance framework. Distinctively, unlike other contexts, Organization Delta has so far avoided relying on partnerships with external and commercial actors, investing instead in building internal competencies and infrastructures.

At the same time, challenges remain significant. Public sector rigidity in contracts and salaries hampers the attraction and retention of highly skilled professionals, as wage grids and seniority requirements limit competitiveness with the private sector. As D-I1 highlighted, *"In the public system, if you enter as middle manager, you have a fixed salary ceiling, equal to that of an administrative employee, and you need five years of seniority before promotion. […] Experience from PhDs or research fellowships does not count as much as consultancy in the private sector, which makes it difficult to attract and retain young professionals, especially given the higher salaries offered by private companies."* Moreover, administrative burdens and lengthy procurement procedures slow down technological adoption. These obstacles make it even more critical to maintain a long-term vision and strategic planning capacity.

Organization Delta exemplifies a contextual-like approach to exploration, in which dedicated units are formally embedded within the hospital's traditional organizational structure but operate in close synergy with academic partners. The reliance on incremental internal development and the avoidance of external partnerships reflects a preference for

a prudential approach, internal control, and gradual scaling. At the same time, the strong emphasis on regional governance and the institutionalized collaboration with the university add contextual dependency, suggesting how the hospital's mission is profoundly shaped by its nature.

5. EMERGENT TENSIONS AND MECHANISMS

Our emergent framework maps the mechanisms through which organizations configure AI innovation. These mechanisms are structured around four recurrent tensions that organizations should navigate across organizational levels and boundaries: (i) organizing AI exploration, (ii) balancing internal and external competences, (iii) the make-buy continuum, and (iv) managing internal demand for innovation.

The tensions can be understood as interconnected dimensions of a broader configurational process. For each tension, we describe the mechanisms observed and the contextual conditions under which they are activated. The contribution is configurational rather than evaluative. We do not address organizational outcomes, as in fast changing environment recurrent patterns and configurations are more informative than short-term performance.

5.1. Organizing AI exploration

The selected cases already combined exploitative with explorative approaches to AI innovation, by both relying on routinized acquisition of external solutions and assuming a more risk-taking stance by developing solutions internally (*AI exploration*). The relevant issue therefore concerns the organizational configurations that legitimize and coordinate AI exploration. In all four cases, the decision to engage in developing AI solutions was not automatic, but emerged from distinct trajectories, trigger events and pathways. Case Alpha inherited an industrial culture from its shareholders, which resulted

in a strong orientation for IT and digital innovation and a focus on R&D. This background illuminated the top-management choice to established AI development activities as a natural extension of such trajectories. In other cases, the decision to engage in AI development was associated with moments that altered the institutional and environmental position of the organization and created the opportunity and the legitimacy for in-house development. Institutional shifts such as formal recognition as a research hospital (Case Beta and Delta) embedded expectations to pursue research-driven innovation in cutting-edge domains such as AI. This status change also triggered organizational restructuring and formalized exploration alongside exploitative adoption. Environmental shocks such as COVID-19 pandemic revealed urgent demands and catalyzed rapid exploratory initiatives on predictive and adaptive tools for clinical and organizational needs (Case Gamma and Delta). Other contingencies, such as strategic opportunities in the form of a prestigious partnership (Case Beta) and the awarding of a competitive grant on cutting-edge topics (Case Alpha), fostered broader awareness of the potential of AI development and provided resources and legitimacy for more systematic exploration. Across these diverse pathways, a common antecedent was present in all cases. Each organization already possessed some form of relevant capabilities – whether embedded in prior digital initiatives, dispersed across university departments, or located in clinical research laboratories. While often fragmented or unstructured, these resources provided the substrate on which exploration was legitimized and organized.

Once exploration was legitimized, the four cases diverged in how they organized and governed it. Across cases, we observed a spectrum of organizational arrangements that blend elements of structural and contextual ambidexterity (Birkinshaw et al., 2016). The first configuration is a structure-like arrangement, through a dedicated AI R&D center

with a mandate to cover the entire exploration cycle. The center concentrates technical and managerial expertise, coordinates grant acquisitions and partnerships, and acts as a visible point of reference for the entire organization. This centralized configuration legitimizes AI development as a strategic priority and provides strong scouting and gatekeeping capacity, ensuring alignment with organizational strategic objectives. The second configuration is a hybrid model. Here, data science and analytics capabilities remain embedded within an ad-hoc research infrastructure, mainly supporting clinical research. At the same time, a separate legal entity is mandated to valorize IP and complete product development. This organizational duality reflected deliberate separation between research, development, and valorization activities, with the spin-off acting as a bridge to external partners and markets. The third configuration is a contextual-like model with dual coordination mechanisms. On the clinical-research side, exploration is structured under the supervision of the scientific directorate around a proprietary AI-based platform that enables access to hospital data, model training and validation, traceability, and compliance. This platform constitutes the main, but not exclusive, backbone of clinical AI research. In parallel, AI innovation aimed at improving clinical workflows and administrative processes sits within the operations department, which governs digital innovation and transformation initiatives at hospital level. AI exploration activities are therefore distributed between platform-mediated processes and operations-led initiatives. The last configuration resembles a pure contextual-like arrangement. AI exploration is embedded within an ad-hoc organizational unit within the IT department, which ultimately has the main responsibility for AI innovation. This arrangement reflects both the public mandate of the organization and its institutional integration with academia and the regional health system where it operates. Rather than separating AI exploration

structurally, this model pursues an embedded approach, leveraging academic collaborations and higher-level institutional governance to support innovation projects while limiting reliance on commercial providers.

Although organizational arrangements for AI exploration were the result of an explicit design, the way these same arrangements mediated the co-existence with exploitation was often the result of emergent practice. In structural-like forms, exploration is located in visible organizational nodes. By concentrating technical and managerial expertise, they become influential interlocutors for the wider organization, able to advise on exploitative initiatives on the AI domain and to channel signals from the external environment into internal initiatives, also performing market and technology scanning and foresight activities. Their role is reinforced by top managers who simultaneously occupy leadership positions in these units and in the broader organization, creating a vertical linkage between exploration and exploitation activities. Overall, they perform a stewardship role that promotes and inspires AI innovation throughout the entire organization. Contextual-like forms, by contrast, embed AI exploration within existing operational structures such as IT or Operations departments. Here AI exploration is more diffusely integrated with exploitation, even though the alignment with established innovation routine is not full.

These observations suggest that organizing AI exploration is a multilevel endeavor, as it is not only about structuring exploratory units, but also about shaping their interfaces with traditional, exploitative, innovation routines. Structural-like and contextual-like arrangements thus represent different configurational responses to the challenge of sustaining AI exploration.

5.2.Managing internal demand for innovation

A third tension concerns how organizations configurate and govern the demand for AI innovation originating from end-users. As in many professional organizations, hospitals are exposed to external pressures from commercial providers (drug and med-tech companies). Yet, the most significant driver for innovation often comes from internal professionals who interact directly with vendors and professional networks. Clinicians, in particular, act as contact points between external technological developments and organizational practice, bringing forward proposals that are framed as responses to unmet needs in both strictly clinical domains (diagnosis, prognosis) and the organization of clinical processes. In research hospitals, where clinicians are also active researchers, such initiatives are particularly frequent and legitimate.

The organizational tension under analysis lies in balancing professional autonomy and bottom-up initiatives with the need for organizational coherence, prioritization, and resource allocation. The mechanisms through which innovation demand was governed varied systematically across configurations. In contextual-like configurations (Case Gamma and Delta), platforms and related use policies were central: proposals for development projects have to comply with specific requirements in terms of potential impact, technical feasibility, resource availability, data accessibility, and overall sustainability. These mechanisms resembled evaluation funnels or stage-gate processes, which progressively screened ideas and made decisions at different phases of the development cycle. By formalizing boundaries of legitimate exploration, platforms effectively act as gatekeepers, setting clear preconditions for the evaluation of bottom-up initiatives. In structure-like configurations, the emphasis is instead on liaison and boundary-spanning roles. Hybrid professionals such as data scientists with clinical

backgrounds or “AI physicians” (Case Alpha) provided translation practices between developers and end-users, while technical staff embedded in the organizational units served as agile intermediaries with the dedicated teams in the development nodes (Case Beta). Here, multidisciplinary boards also played a role in legitimizing and prioritizing bottom-up initiatives by bringing together technical, clinical, and managerial perspectives.

While the formalization of rules and procedures seems to be useful in governing bottom-up demand, it also reveals potential limitations. As AI – especially through generative and agentic applications – becomes more pervasive, configurations relying predominantly on formalized procedures may encounter limitations, while those embedding liaison and boundary-spanning roles appear better positioned to mediate between strategic direction and professional initiatives.

5.3.The make-buy continuum

Across the four cases, once AI innovation was structured as an exploratory endeavor, organizations faced a recurrent make-buy tension regarding specific initiatives. The nature of the tension resembles a continuum and a continuous use-case-level tension rather than a one-off choice. This tension stems from a particular peculiarity of complex, knowledge-intensive organizations such as healthcare organizations, where high and low standardized activities, high and low regulated domains, and high and low sensitive processes coexist and interact. This tension is also related to the different nature of AI technologies, such as traditional and generative AI, for example, exhibiting different technical functions, integration needs, and regulatory requirements and concerns.

We observed a strong convergence across cases in how this tension is addressed, although some rationales depend strongly on the institutional nature of the organization. All organizations agreed that in clinical domains characterized by high standardization and routinization, AI innovation should be market-driven. For example, computer vision tools for diagnostic imaging were consistently regarded as preferable in their commercial form. These solutions can be readily – though not necessarily easily – integrated into existing routines and are increasingly embedded in diagnostic devices. Even though such tasks involve sensitive data in practice, their standardization makes the clinical and operational benefits of commercial AI immediately appreciable. This approach is generalizable to all domains that are highly standardized across organizations and are distant from the core capabilities and resources of the organization. Conversely, being at the core of the organizational business activities emerged as a clear rationale for pursuing make. Here the issue is twofold. On the one hand, internal development signals the willingness to retain control of the source of a competitive advantage within the organization; on the other hand, it reflects a frequent misalignment between organizational needs and market supply. Across cases, two examples illustrate this rationale consistently. First, platforms that enable and enhance research activities – such as data and analytics infrastructures that valorize organizational assets like clinical data, know-how, and research capacity – were seen as a core for a research hospital and therefore developed internally. Second, high-complexity, context-dependent, clinical and administrative activities were also strong candidates for make, because in such areas pure buy would amount to outsourcing core knowledge processes. On this point, prior work on organizations relying on knowledge-based activities has already argued that outsourcing knowledge processes rarely follows a make-buy logic, but rather a make-ally one (Mudambi & Tallman, 2010).

Alliances as a governance form are indeed central and will be discussed in a dedicated tension. A third rationale for make is closely related to knowledge processes but intersects with regulatory and compliance concerns. It is related to data and the possibility – often necessity – of developing solutions internally when sensitive or strategic data are involved in the process. Technical architectures such as federated learning were cited across cases as a mechanism that allows training models without moving raw data outside the organization. More generally, concerns emerge when commercial solutions are based on proprietary models delivered via cloud, where security guarantees and compliance with internal policies cannot be taken for granted. In these cases, the preference for make is not only strategic but also pragmatic, ensuring control over data use and regulatory adherence.

Beyond these convergences, institutional nature shapes the rationales for make. In private organizations, the economic valorization of IP emerged as a particularly strong driver. This goes beyond the retention of knowledge process: it also includes the possibility of developing entrepreneurial initiatives around the knowledge produced, which – through the organization of exploration – can be structured and leveraged. Public organizations, by contrast, face weaker institutional and organizational incentives to pursue IP valorization as a strategic goal. Yet, their status as a research hospital and their integration with universities embed logics of technology transfer that may still push in this direction, even if more indirectly and less forcefully than in private settings.

In summary, make-buy decisions never emerge as a one-off and isolated decisions, but as a recurrent configuration response reflecting organizational history, institutional mandates, and specific features and purposes of the innovation at hand.

5.4.Balancing internal and external competences

The last tension concerns the balance between internalizing and externalizing competences for AI innovation. This tension operates at the level of organizational capability architecture, shaping how AI-related skills, infrastructures, and partnerships are configured. The issue resonates with prior work on ambidexterity, which showed that organizations rarely rely exclusively on either internal or external knowledge sources (Rothaermel & Alexandre, 2009). Instead, they constantly combine and recombine capabilities under conditions of resource constraints and allocation trade-offs. In healthcare organizations, this tension is particularly salient, as developing advanced computer science skills has not traditionally pursued. AI innovation forces organizations to make challenging choices about what competencies to build internally and what to access externally.

Across cases, we observed a common pattern: research activities were consistently internalized, while development was pursued with varying intensity. Hospitals sought to retain AI exploration capacity in research and early development phases but faced two current barriers. First, extending expertise into areas such as ML engineering and MLOps went beyond their traditional skill base. Second, talent attraction and retention proved difficult in a context of high cross-sectoral demand for AI skills. This was especially true for public organizations facing binding constraints on wages and contracts.

Another recurrent dilemma concerned the choice between internalize (on-premises) and externalize (cloud) technological infrastructures. On-premises solutions were associated with stronger internalization of data governance and compliance competences but required significant investments and scarce expertise. Cloud solutions, by contrast, lowered entry barriers and offered scalability, but implied dependence on external

providers and greater exposure to regulatory uncertainty. These two options thus exemplify how infrastructural decisions become part of the broader balancing act between internal and external competences.

Scaling activities, by contrast, were consistently externalized, with different logics. In one case (Alpha), the preferred option was to spin out mature and promising projects, giving them entrepreneurial autonomy and space. In another (Case Beta), an open innovation approach led the organization to systematically scout partnerships and alliances, even in earlier stages of the development cycle, and to leverage these relationships for scaling. These differences underscore that externalization is not a residual choice, but an integral part of how organizations design their AI exploration strategies. Finally, partnerships played a crucial role in complementing internal competencies, serving not only technical but also symbolic purposes. We identified three distinct objectives for building partnerships. Some partnerships were oriented to knowledge extension, providing access to external data and know-how. Others targeted capability completion, filling gaps in technical, regulatory, or commercial expertise. Still others carried a symbolic value, as high-profile alliances signaled prestige and legitimacy, thereby reinforcing internal commitment and sometimes acting as additional triggers for the expansion of exploration. Taken together, the tension between internalizing versus externalizing competences and building partnerships emerge as a dynamic balancing that reflects both strategic priorities and institutional constraints.

6. DISCUSSION AND CONCLUSIONS

The framework that emerged from our analysis highlights three transversal insights. First, the decision to engage in exploration is never a sudden pivot but is rooted in organizational legacy and culture and catalyzed by specific trigger events. Industrial or

scientific legacies provide the substrate on which exploration can grow, while institutional shifts, external shocks, and strategic opportunities legitimate and provide resources to in-house development. This distinction clarifies that deciding to explore is a strategic-level choice, which then cascades into project-level decisions about how to source innovation. Second, governing AI innovation requires multilevel mechanisms that cut across technological, organizational, professional, and inter-organizational domains. Organizations combine architectural arrangements with market-level choices, with professional dynamics, and with partnerships that extend exploration beyond the organizational boundaries. The framework therefore shows that exploration cannot be understood in isolation, but only through its interface with exploitation components, end-user practice, and external ecosystems. Third, managing bottom-up demand and unmet end-user needs requires a balance between adaptive mechanisms and strategic direction. So far AI innovation has been largely incremental and process-oriented (Lin & Maruping, 2025), well suited to bottom-up discovery and experimentation. As generative and agentic AI technologies evolve, more radical forms of innovation may emerge that cannot be governed solely through incremental and adaptive mechanisms. This raises the prospect that bottom-up initiatives will need to be complemented by stronger top-down and strategic guidance.

Our findings need to be interpreted within specific boundary conditions. The evidence was induced from four large, research-active hospitals in the Italian healthcare system, a highly regulated, professionalized, and knowledge-intensive industry. The tensions and mechanisms we identified might be extended to large organizations with significant research capacity and strong link with academic networks. For example, smaller and non-research organizations are likely to face similar tensions in different ways. In such

organizations, the relevant sourcing dilemma is less about make-buy than about buy-ally, where the critical capability lies in commissioning and partnership management rather than in developing AI solutions internally.

Our findings extend and connect to several strands of organization and management and research. First, our study is anchored and contributes to the literature on organizational ambidexterity, which provided the lens through which we selected our cases. In line with recent reviews, we observed that ambidexterity in practice assumes multilevel configurations, cutting across organizational, group, and individual levels (Turner et al., 2013; Fernández-Pérez De La Lastra et al., 2017; Kassotaki, 2022; Chakma et al., 2024). This perspective draws attention not only to structural arrangements, but also to boundary-spanning roles, hybrid experts, and multidisciplinary boards that connect different knowledge domains. Our cases also illustrate how stewardship functions emerge from organizational configurations to guide innovation across the organization and how ambidextrous logics extend beyond the organizational boundary through partnerships, spin-outs, and collaborative arrangements (Saleh et al., 2023; AlKhamees & Durugbo, 2024). We also engaged in literature that situates ambidexterity within processes of technology sourcing (Saleh et al., 2023). Prior work emphasizes that in turbulent environments organizations constantly balance internal and external knowledge sourcing, moving along a make-buy-ally continuum (Rothaermel & Alexandre, 2009; Mudambi & Tallman, 2010; Denicolai et al., 2016). Our cases reflect this ongoing balancing act in the field of AI innovation, where in-house development is continuously weighed against external market opportunities and alliances. Our findings also resonate with calls for more attention to antecedents of ambidexterity (Saleh et al., 2023; Chakma et al., 2024), as we show that exploration can originate both from deliberate strategic choices and from

institutional or environmental pressures, including serendipitous opportunities, with different patterns emerging from peculiar legacies and trigger events.

The second contribution relates to the streams of AI innovation and organizational capabilities. Prior work defined AI innovation capabilities as the ability to develop, manage, and use AI resources for scientific discovery and R&D (Lou & Wu, 2021), while strategy scholars stress the need to situate AI in organizational routines (Kemp, 2023). Our findings are part of this view and extend it by showing that AI innovation capabilities take configurational and multilevel forms, varying across structural, contextual, and hybrid arrangements and depending on organizational objectives and institutional constraints. Hospital organizations built such capabilities not only through technical expertise but also through organizational and relational mechanisms. At the same time, our results reinforce the idea that AI innovation must be understood as a system process, as it is shaped by system builders, integrators, developers, and implementers (Vannuccini & Prytkova, 2023) and its diffusion depends on relational embeddedness within AI knowledge networks (Dahlke et al., 2024). All the observed organizations were not isolated but positioned themselves within these broader systems, building AI capabilities encompassing the ability to orchestrate inter-organizational collaboration and to extend beyond organizational boundaries.

The last contribution relates to the literature on innovation in professional organizations. Prior studies show that professional logic and knowledge practices deeply shape how new technologies are introduced and legitimized (Pachidi et al., 2021; Goto, 2023). Our findings reinforce this perspective by demonstrating that in hospital organizations, clinicians are not only adopters or resistors of innovation but often act as legitimate initiators of AI projects, especially when unmet clinical needs provide the rationale. This

bottom-up pressure places professional organizations in a unique position: they must recognize and filter initiatives that emerge from professional communities while maintaining strategic coherence. The mechanisms we observed in both structural and contextual arrangements illustrate how this translation is achieved. These mechanisms highlight that innovation in professional organizations requires both formal structures and relational interfaces to translate bottom-up demand into organizationally viable projects.

This study has limitations that also suggest avenues for future research. First, while the multiple case design enables cross-case comparison, the number and distribution of interviews across cases limit the granularity of analysis and the richness of evidence. Future research should build on more extensive field engagement to provide thicker descriptions of how these tensions unfold in everyday organizational practice. Second, it offers a snapshot of innovation at a specific moment in time. As the technological and regulatory landscape is evolving rapidly, a longitudinal research approach would be useful to trace how organizations adapt ambidextrous arrangements over time, and how mechanisms rooted in legacy and culture are reconfigured in response to ongoing technological and institutional change. Third, the rapidity of technological evolution also makes it difficult to assess organizational outcomes. Our analysis focused on recurrent tensions and mechanisms, rather than on the performance of observed configurations. As AI innovation reaches a greater degree of maturity and stability, it will become possible to evaluate more systematically how different configurations affect innovation outcomes of ambidexterity. Finally, while our findings highlight the importance of multilevel configurations, the analysis remains at a relatively broad level. Future research could probe more deeply into micro-behavioral dynamics, examining how individual roles, professional identities, and team interactions shape the functioning and effectiveness of

ambidextrous arrangements. Attention to these micro-foundations would provide more fine-grained understanding of how ambidexterity is enacted in practice and how it contributes to building innovation capabilities.

APPENDIX 2

Semi-structured interview guide

Introduction

The purpose of this interview is to explore your organization's experience with artificial intelligence (AI) solutions, in particular with the decision to develop certain applications internally. We are interested in understanding the motivation behind this choice, how it was implemented (in terms of resources and enabling conditions), and what implications it had for the organization's operations. For the purpose of this interview, internal development refers to both the in-house creation of AI solutions (e.g., ML/DL) and substantial adaptation of generative AI systems, whether developed independently or in partnership with external entities (e.g., tech companies, consulting firms, universities, other health service providers).

Section 1 – The decision to develop internally

1. When did your organization decide to start developing AI applications internally? Who was involved in that decision? What were the main motivations?
2. What are the main goals and areas of focus of your development activities? How were they selected? Please share some examples…
3. Can you provide an overview of the current development projects? A general outline of the type and application areas is sufficient.
4. Has a structured process been put in place to guide development choices? What roles are involved? Did you build on existing processes or learn from external models?

5. In what areas has the organization instead opted to acquire AI solutions externally? What are the key reasons?
6. Are there tools or mechanisms in place to coordinate or integrate internal development with external acquisitions?

Section 2 – Resources, capabilities, and organizational implications

7. What kind of investment in tangible (e.g., financial, infrastructure) and human resources have supported internal development activities? Has a specific organizational model been created?
8. Has your organization structured itself in a way that internally developed applications could also have commercial potential? What models was adopted?
9. To what extent and why have competencies been outsourced (e.g., consultants, affiliated researchers), or partnership established specifically for development? What roles does your organization play in these partnerships?
10. How would you describe the organizational impact of internal development? Was there a gap between expected and actual outcomes?
11. Regarding externally acquired solutions, how is the acquisition process structured and how do you select among variable options?
12. Based on your experience, what do you consider necessary for a healthcare organization to successfully carry out systematic internal development of AI solutions? Are there lessons learned, errors to avoid, or contextual factors you consider critical?

Table A2.1. Coding scheme

Tension	Aggregate mechanism	Characteristics	Illustrative quote
Organizing AI exploration	Triggering and legitimizing exploration	Grounding on clinical research culture	*"There was also a historical cultural push toward collecting data for research purposes. We had the cards ready, meaning we already had datasets structured at the local level that could be used to train artificial intelligence algorithms."* (A-I3)
		Grounding on the industrial culture of the group	*"This logic was adopted primarily because [we are part of] an industrial group, we were all engineers, and it made sense to move quickly in order to differentiate ourselves significantly in the market."* (A-I1)
		Covid-19 as a trigger for starting to explore	*"It came about somewhat by chance during the COVID period because one of the main problems during COVID was that patient flow management was […] not optimized."* (G-I2)
		Starting a strategic partnership	*"With [partner name], it was a direct initiative by the ownership to create this framework agreement, which then gave rise to the research platform."* (G-I1)
		Recognition as a hospital of research excellence (IRCCS)	*"The concept of evolution that led [us here] originated during the IRCCS accreditation process, which began in 2016 and was completed […] in 2018."* (B-I1)
	Configuring organizational arrangements	Establishing a dedicated unit for research	*"Those activities that were previously distributed in a granular manner, namely data collection, database creation, data analysis, and knowledge extraction for publication purposes, were assigned to one of the deputy directors of the IRCCS, specifically responsible for medical big data."* (B-I1)

		Establishing a dedicated legal entity	*"And then a project deliverable […] was precisely the creation of a vehicle, […] this gave us the opportunity to create the vehicle with a strategic mandate to promote everything I have told you about to the market."* (B-I1)
		Integrating exploration in the traditional operations	*"The person who reports directly to me [the COO], who is responsible for AI and digital transformation, is still part of the operational management team and has not yet become an operational unit."* (G-I1)
		Establishing an integrated unit for exploration	*"Since last year [2024], we have merged these experiences into a single operating unit that is responsible for both the institutional side and the more experimental and quantitative organizational research side, […], and clinical research support […]."* (D-I1)
		Establishing an AI R&D center	*"The AI Center was founded at the end of 2019, starting out as a small R&D center."* (A-I1)
	Mediating coexistence with exploitation	Establishing organizational routines for make-or-buy assessment	*"A board has been set up within [the hospital], and it is responsible for evaluating these solutions, deciding whether to proceed with the make or buy question, and how to implement them."* (A-I2)
		Sharing knowledge and experience with the procurement function	*"Yes, we continue to be involved. […] sometimes we do the scanning ourselves to evaluate different solutions, technological solutions, but also from a point of view of use and, how can I put it, deployment, installation."* (A-I2)
		Establishing a continuous relationship with the hospital	*"Then, in reality, there is constant negotiation, in some cases with a smile, in other cases with gnashing of teeth, but it is always, always like that."* (B-I1)

The make-buy continuum	Rationale for buy	Buying for standardized activities	*"Let's start with the clinical domain, which is fairly straightforward: where you mainly have standardized activities, certified with CE mark, MDR class 1,2,3, whatever that is, that's pretty much buy today."* (A-I1)
		Integrating buy-products	*"There is still a long way to go. Even when you buy, there is always an aspect of development involved, in terms of how you can use it specifically at home."* (A-I1)
		Pursuing quick returns through buy-products	*"And then also for purely managerial reasons, speed, start-ups […] and also for representation to stakeholders, that it was an investment that could lead to quick returns."* (G-I1)
	Rationale for make	Protecting intellectual property	*"We do make projects because they are too close to the company's core know-how, so they should never have done anything else. […] so to simplify the world but divide it up a bit where there is intellectual property, something strong, relevant, you make."* (A-I1)
		Favoring seamless integration into organization	*"Because at that time, […] some solutions, let's say pre-packaged solutions, would have required organizational changes that were not economically feasible."* (B-I1)
		Avoiding constraints of proprietary solutions	*"For example, a tool such as ChatGPT can be supportive with one of the applications on which ChatGPT can be used, which is, for example, RAG (Retrieval-Augmented-Generation) technology. […] So, this tool clearly has a GPT. It's already ready, but there are a few problems. What are they? If, for privacy reasons, I can't release enterprise documentation, let's always remember that ChatGPT is a closed solution, meaning that every piece of data and every question asked of ChatGPT is used for retraining and sent to OpenAI's servers, etc. When I encounter these difficulties, I have to look for alternative solutions, so in this*

			case, even though a "buy" solution is ready, I often can't use it, so here we almost go with Make". (A-I2)
		Working in complex domains	*"To solve these problems, a make approach is essential, simply because the complexity of diseases cannot be generalized; in other words, there cannot be a single tool that analyzes all diseases, so make is the application on the clinical side."* (A-I2)
Managing internal demand for innovation	Gatekeeping for bottom-up initiatives	Identifying the unmet need	*"We always start by trying to answer questions that have a high priority and an unmet need and are very well identified from a clinical point of view."* (A-I3)
		Engagement of clinicians in solutions assessment	*"The key here is to work as a team, involving the doctor from the outset and seeking a common language between doctors, engineers, technology, and the clinical side, because sometimes technology helps to better understand the clinical side, and the clinical side helps to better calibrate and develop the technology."* (A-I2)
		Embedding assessment criteria in the research platforms	*"This working group has the entire governance […] The idea comes from proposals made by doctors at the hospital or other hospitals who somehow come up with an idea. After that, a process is carried out that has been perfectly defined with a whole series of steps, which have also been the subject of ISO 2001 certification, where we carry out a technical and qualitative feasibility assessment."* (G-I2)
	Liaison roles and teams	Multi-disciplinary teams	*"We work from the outset with a multidisciplinary team, doctors and radiologists, nurses, technicians, data scientists, managers, so there is a natural evolution in the maturity of the project, nothing is set in stone, it is literally bottom-up."* (A-I1)

		Hybrid experts	*"The second is to have, as we are doing in our Cancer Center, so-called Physician-Scientists […]. The idea is to set aside part of the time for research, but focused on the hospital's strategic investment priorities, which is why we say that our research and development team with the AI Center is a truly multidisciplinary team […]. The clinical part is made up of physician-scientists, especially young people who have part of their time dedicated to participating in these projects."* (A-I3)
		Boundary-spanning roles	*"And then there is the group of designers, who create the object and draw on the expertise of their colleagues to complete it. […] They act as a link with the outside."* (B-I1)
Balancing internal and external competences	Rationale to externalize	Scaling capacity	*"The products that we have registered as co-owned are all with partners who are potentially capable of industrial scaling. So we bring the product to the end of the certification cycle, and then we bring the product to the point where it can be used on the market outside of research activities."* (B-I1)
		Computational asset	*"The point is how to systematically access, through discussions that we say we need to explore further, this [regional] high-potential computing center. […] However, it makes no sense, especially given the investment made by the [Region], not to evaluate the full use of that type of facility first."* (D-I1)
	Building partnerships	Extending knowledge	*"It could be in a subject who is in a multicenter study, […] so they are clinical experts, they are potential data providers, but then, at the implementation level, this is done centrally on anonymized data in another location."* (D-I1)

		Capability completion	*"So, in reality, we try to collaborate with those who can give us added value, both in terms of AI and the clinical side. And in recent years, as I said before, we have created a network of collaborators that has grown out of these European projects and consortia."* (A-I2)

CHAPTER 3

ALGORITHM AVERSION, DOMAIN EXPERIENCE, AND FAMILIARITY IN PROFESSIONAL DECISION-MAKING: A MULTI-LEVEL ANALYSIS OF THE DETERMINANTS OF PHYSICIAN-AI INTERACTION[5]

1. INTRODUCTION

Despite the growing availability of AI-based algorithms able to support or even improve high-stake decisions in domains such as healthcare, justice, crisis management, defense and autonomous driving, it has been repeatedly shown that, in these domains, decision-makers often discard or override algorithmic recommendations (Burton et al., 2020; Mahmud et al., 2022; Jussupow et al., 2024). In healthcare, recent data by the US Food and Drug Administration indicate that by the end of 2024 over 1,000 AI-based software were registered as medical devices in the United States (FDA, 2025), mainly to support

[5] This chapter is part of a research conducted with Giulia Cappellaro, Amelia Compagni, and Bart Vanneste. The research was approved by the Bocconi University's Ethics Committee (n. EA000831).

medical decisions in the prognosis and diagnosis of relevant diseases through image reading (Sutton et al., 2020; Esteva et al., 2021).

A large body of evidence documents that many of these AI-based solutions remain unimplemented or underutilized in clinical practice (Preti et al., 2024). This last behavior, commonly referred to as algorithm aversion, was initially documented by Dietvorst and colleagues (Dietvorst et al., 2015) as the reluctance to rely on algorithms preferring human judgement, even when algorithms ultimately outperform humans. Algorithm aversion is greatly heightened in domains dominated by professionals that hold a recognized expertise in a field and that, based on such expertise, exercise a relevant degree of subjective judgment over their decisions. Algorithmic recommendations are perceived by experts as threatening of their professional identity and status (Anthony 2018; Jussupow, et al., 2022); acceptance of their recommendations often requires the redesign of the workflow performed by these professionals and effort in terms of time and cognitive burden (Jussupow et al., 2021; Lebovitz et al., 2022), leading to an increase in the resistance to the widespread adoption and appreciation of AI-based algorithms (Anthony, 2021; Pachidi et al, 2021; Van Den Broek et al., 2021; Waardenburg et al., 2021).

Prior research has shown that algorithm aversion is a complex phenomenon influenced by a variety of factors related mainly to the algorithm itself, the decision context or the task upon which the decision is made, and the decision-maker (Burton et al., 2020; Mahmud et al., 2022). Most of these factors have been studied in isolation one from the other, often with experimental methodologies that allowed the manipulation of one factor at a time in a cross-sectional fashion (Mahmud et al., 2022). Despite returning a rich set of factors, results have not always been consistent, hinting to the fact that the conditions

under which decisions are taken might lead to the same factor contributing or detracting from algorithm aversion (Kawaguchi, 2021; Mahmud et al., 2022; Jussupow et al., 2024). Real-world studies, that examine the dynamics of algorithm aversion over time and the combination of factors as they occur, are lacking, making the understanding of this phenomenon still rather fragmented and incomplete (see Liu et al., 2023 for a notable exception). Understanding what influences algorithm aversion and its contrary, algorithm appreciation (Logg et al. 2019; Jussupow et al., 2024), is, nevertheless, of paramount importance to assure that organizations and professionals can fully reap the benefits of AI-based decision support systems. Among the factors that have been identified in the literature as influencing algorithm aversion, we are particularly interested in understanding the contribution to the acceptance of algorithmic recommendations of the *decision context* and the *experience with the AI-based algorithm* (Prahl & Van Swol, 2017; Horowitz et al., 2024; Mahmud et al., 2024). Advancing the understanding of these factors and their relation is, in our view, fundamental as they capture the essence of the main trade-off underlying algorithm aversion, i.e., relying more on the algorithmic recommendation or on human subjective judgement. So far, the literature is unclear as to direction of this relation. Experience with the AI-based algorithm has been shown to both increase and decrease algorithm aversion (Horowitz et al., 2024), while decision uncertainty is normally associated with high levels of algorithm aversion (Dietvorst & Bharti, 2020). The aim of this paper is hence to answer the following research questions: *Can the decision context and the experience with the AI-based algorithm influence the physician's algorithm aversion for treatment recommendations?*

We examine these factors and their relations in the context of healthcare and more specifically in the case of the application of AI-based algorithms to therapeutic decisions.

Treatment decisions are known for being the result of the complex process, followed by clinicians, of examining contextual factors, risks and benefits of treatments, and alternative options, often in absence of solid and certain evidence or clear-cut guidelines as to what the precise course of clinical action should be. In the sociology of professions, diagnosis, inference, and treatment are described as interrelated dimensions of the professional jurisdiction (Abbott, 1988). While deeply connected in practice, treatment decisions are particularly consequential, as they translate professional reasoning (inference) into concrete action. These decisions are intrinsically ridden by high degrees of uncertainty that require clinicians' strong reliance on their subjective judgement.

We address the research question in a large-scale, real-world setting, leveraging a unique dataset of more than 584,000 treatment recommendations generated by an AI-based CDSS used in nephrology care and built to optimize the pharmacological treatment of anemia in dialysis patients. Since its first deployment in early 2013, the algorithm has been used to treat more than 49,000 patients by roughly 1,500 physicians in 230 dialysis centers operating in more than 10 countries across the world. We adopt an empirical strategy that compares algorithmic recommendations to physician choices and identifies algorithm aversion in the form of rejection of AI advice in favor of physician's own judgement (Jussupow et al., 2024).

Our results reveal that alignment with status quo and with existing clinical guidelines are the strongest drivers for the acceptance of the AI-based recommendation. Physicians tend to favor recommendations that maintain prior dosing over recommendations to modify it, while they see high acceptance when recommendations mirror clear-cut guideline norms. Experience measures play a heterogeneous role in affecting acceptance. While the role of familiarity remains ambiguous, our findings reveal a higher propensity of more

experience physicians to rely on algorithmic recommendations and – more importantly – that experience may moderate physicians' attitude toward risk and their willingness to deviate from inertia, by inducing them to accept AI-recommendations in situations that are less exposed to risk of severe consequences.

In what follows we first summarize what the literature documents with respect to (i) algorithm aversion in the healthcare context and (ii) the relationship between clinical decision-making, decision uncertainty, and decision-support tools. We then proceed to describe in detail the empirical context of our study related to the treatment by nephrologists with ESA of dialysis patients and the characteristics of the AI-based algorithm under study. Methods, findings and discussion of contributions and implications complete the paper.

2. THEORETICAL BACKGROUND

2.1.Aversion to AI-based algorithms

Despite the growing adoption of AI-based algorithms also in high-stake decision-making domains like health care, people often fail to integrate them seamlessly into decision-making (Burton et al., 2020; Jussupow et al., 2024). This phenomenon, commonly referred to as algorithm aversion, was originally described by Dietvorst et al. (2015) as a reluctance to rely on (superior) algorithms when human decision makers know they are imperfect. Recent advancements in algorithms' learning capabilities led to a surge in the use of algorithmic technologies across organizations, with medical decision-making consistently ranked among the most exposed domains (Jussupow et al., 2020; Mahmud et al., 2022). AI-based algorithms often differ from earlier algorithmic tools not only for their accuracy, but also for their learning capability, the complexity and the scope of the

problems they can address, and the opacity of their underlying logic (Turel & Kalhan, 2023). Four broad classes of drivers of algorithm aversion emerge from prior research.

One set of factors driving algorithm aversion centers on the context and nature of the decision at hand. Aversion is pronounced for decisions marked by relevant uncertainty (Dietvorst & Bharti, 2020) and that frequently require some moral judgement (Bigman & Gray, 2018). Similarly, when the potential consequences of erroneous recommendations are severe, people become more averse to rely on algorithmic outputs (Utz et al., 2021; Filiz et al., 2023). A second set of drivers relates to the role of the decision maker's experience. In some studies this is conceptualized as an ex-ante characteristic of the decision maker in the task at hand (domain experience) (Choudhury et al., 2020; Brink et al., 2024), in others experience is the result of a process in which both time spent with (i.e. length of experience) and its frequency are relevant for the final degree of familiarity with the algorithm, i.e. individual's general knowledge and understanding of the algorithm (Mahmud et al., 2022). Across these perspectives, experience may shape how decision makers interpret algorithmic outputs, assess their reliability, and calibrate their reliance on AI-based recommendations. In professional contexts, another class of drivers is rooted in professional identity. Professionals may perceive algorithms as a threat to their authority and professional autonomy, particularly when the recommendations conflict their own assessment or presented without the possibility of negotiating or adapting the decision (Shaffer et al., 2013; Jussupow et al., 2022; Bankuoru Egala & Liang, 2023). A third set of drivers pertains to the characteristics of technology. AI-based algorithms are often perceived as opaque and recommendations poorly explained, undermining the decision maker's ability to understand, evaluate, and justify the delivered outputs. Empirical evidence suggests that explainability, perceived

justifiability, and alignment with human reasoning are useful to foster acceptance and intention to use algorithms (Jussupow et al., 2021; Burton et al., 2020; Rosenbacke et al., 2024).

However, much of this evidence isolates individual factors in experimental settings, leaving limited understanding of how contextual factors of the decision and the accumulated experience with an AI system jointly shape aversion in real-world professional environments.

2.2. Therapeutic decisions, clinical practice guidelines, and decision-support algorithms

Clinical decision-making is a form of practical reasoning that integrates experiential knowledge, prior learning, and the interpretation of complex context-specific information (Montgomery, 2006). Its complexity arises not only from objective factors – such as measurable clinical parameters or documented clinical history – but also from latent, multidimensional elements, including social, economic, and behavioral aspects that require interpretation beyond standardized rules (Nicolaus et al., 2022; Safford et al., 2007). Moreover, clinical decisions are made under conditions of uncertainty that extend beyond quantifiable risk, encompassing ambiguity due to incomplete or conflicting evidence, as well as the complexity arising from numerous interdependent clinical cues (Han et al., 2011). Among the various decisions a physician faces, treatment decisions are particularly relevant, as they involve iteratively gathering and exchanging information, deliberating about treatment options, and deciding on the treatment to achieve the best health outcome for the patient (Charles et al., 1999). These decisions are grounded in a combination of clinical theory and evidence with individual values, requiring physicians to collect and weigh information regarding benefits, costs, and

potential harms, while integrating them with their own experience, values, and patient's preferences to determine the most appropriate course of action (Eddy, 1990; Harrison et al., 2004).

Over the last decades, AI-based CDSS have been developed to enhance medical decision-making across various contexts by providing recommendations aimed at improving patient outcomes. Despite the quality of the recommendations and the effort by healthcare organizations to promote the integration of algorithmic recommendations with clinical judgment, the potential of AI-based decision-support algorithms in clinical practice remains largely unexploited. This is particularly evident in systems designed to deliver treatment recommendations. Unlike diagnostic and prognostic systems – where ground truth labels can be more readily defined – treatment recommendation algorithms attempt to infer optimal decisions by predicting the average effect of a treatment on patient outcomes (Sivaraman et al., 2023). Moreover, among the few AI-based decision-support algorithms specifically intended to generate treatment recommendations, most of them generate recommendations by leveraging existing clinical guidelines (Beauchemin et al., 2019; Duwe et al., 2024).

Clinical guidelines aim to reduce unwarranted variation in care and improve patient safety and outcomes by promoting consistent and evidence-based decisions (Woolf et al., 1999; Djulbegovic & Guyatt, 2017). Despite the well-documented benefits, clinical practice guidelines often face criticism for their limited adaptability to the nuances of individual patient cases, especially when individual characteristics, preferences, and contextual constraints complicate a straightforward application (Cabana et al., 1999; Greenhalgh et al., 2014). Suggested actions from clinical practice guidelines can vary substantially in strength depending on the quality, consistency, and quantity of supporting evidence,

creating further ambiguity in interpretation and implementation (Guyatt et al., 2011). Yet, even when a guideline offers an unambiguous recommendation, compliance is not guaranteed due to clinical inertia – the tendency to delay initiating or intensifying therapies when indicated – reflecting physicians' preferences for maintaining the status quo (Phillips et al., 2001; Giugliano & Esposito, 2011; Aujoulat et al., 2014; Camilleri & Sah, 2021). Evidence show how in certain medical domains such as oncology, decision-support tools grounded in clinical practice guidelines have proven effective in improving clinical adherence, as they offer a structured way to integrate evidence-based practice with context-specific information, enabling a more tailored by adjusting recommendations based on patient characteristics (Fox et al., 2009; Beauchemin et al., 2019; Sutton et al., 2020). On the other hand, once guidelines become embedded in clinical practice, the use of decision-support algorithms may be questioned whenever their recommendations diverge from established guidelines prescriptions, potentially undermining trust in the tool and encouraging physicians to use more familiar practices (Bouaud Jacques et al., 2015; Sivaraman et al., 2023; Duwe et al., 2024).

3. HYPOTHESES

3.1. Alignment to clinical practice guidelines

Building on prior evidence on decision justifiability and the use of guidelines-based decision-support algorithms for therapeutic decision-making, we argue that the key determinant of reliance on an AI-based treatment recommendation is its alignment with prescriptive norms. Normative behavior theories suggest that individuals are more likely to endorse decisions that conform to salient rules and standards (Kahneman & Miller, 1986; Cialdini et al., 1991). In healthcare, clinical guidelines represent institutionalized norms that define standards of accountability. Accordingly, accepting an algorithm

recommendation that aligns with such standards can be more easily, whereas accepting an algorithmic recommendation that contradicts the guidelines may increase perceived professional exposure in the advent of adverse outcomes. Accordingly, we expect physicians to display greater acceptance of recommendations that are congruent with guidelines.

Hypothesis 1 (H1). *Physicians are more likely to accept an AI-based treatment recommendation when it is consistent with the relevant clinical guidelines.*

3.2. Alignment to prior treatment

When clinical guidelines do not prescribe a unique, clearly justifiable course of action, physicians lack an external normative anchor for justifying their decisions. In such contexts, prior treatment behavior may become the salient reference point guiding subsequent decisions. Status quo bias, defined as the preference to maintain the current or previous decision state over changing it, has been extensively documented across domains, including medical decision-making (Samuelson & Zeckhauser, 1988; Camilleri & Sah, 2021). In clinical practice, this often manifests as therapeutic inertia, i.e., the tendency to continue a previous treatment approach despite changing clinical conditions (Aujoulat et al., 2014). Accepting an algorithmic recommendation that aligns with the previously adopted treatment course reduces the behavioral and professional cost of change, preserving continuity and limiting exposure to potential scrutiny. Accordingly, we expect physicians to display greater acceptance of AI-based recommendations that replicate the previously adopted treatment pattern.

Hypothesis 2 (H2). *Physicians are more likely to accept an AI-based treatment recommendation if it aligns with the prior treatment behavior.*

3.3.The role of domain experience and familiarity

While H1 and H2 capture situational mechanisms, individual-level characteristics may independently shape physicians' propensity to rely on algorithm recommendations. A significant body of literature investigates the role of experience in human-AI interaction. Domain experience frequently refers to domain-specific knowledge accumulated through focused practice or experiential learning (Simon, 1991; Salas et al., 2010; Choudhury et al., 2020; Allen & Choudhury, 2022; W. Wang et al., 2023). It is a concept related to functional expertise as the ability of individuals to perform a task with a higher level of skills, efficiency, and accuracy when they engage with the task over a long time (Heimstädt et al., 2024). Greater domain experience might strengthen reliance on established cognitive schemes and prior routines, increasing self-efficacy, and trigger professional-identity concerns, potentially diminishing physicians' propensity to rely on algorithmic recommendations (Logg et al., 2019; Jussupow et al., 2022; Mahmud et al., 2022). Although more experienced physicians might be better equipped to contextualize algorithmic outputs and detect potential errors, their stronger attachment to established practices may reduce the perceived value of the AI-based recommendation. We therefore expect domain experience to be negatively associated with acceptance of AI-based algorithm recommendations.

Hypothesis 3 (H3). *Physicians with greater domain experience are less likely to accept an AI-based treatment recommendation.*

In addition to domain experience, familiarity with the algorithm may influence the reliance on AI-based recommendations. Familiarity refers to the understanding of an entity based on prior interactions and learning experiences (Gefen et al., 2003; Komiak & Benbasat, 2006; Mahmud et al., 2022). Prior evidence suggests that its behavioral

effects are ambiguous. Limited exposure may reduce trust and uptake, while repeated use may increases both awareness of potential errors and the ability to interpret complex outputs (Prahl & Van Swol, 2017; Choudhury et al., 2020). Given these countervailing mechanisms, we do not advance a directional hypothesis and ask instead how familiarity with an AI-based algorithm influences the likelihood of accepting an AI-based treatment recommendation.

4. RESEARCH SETTING

In this section, we first provide a detailed description of the clinical context that is the object of our study, that is the treatment of anemia for patients with ESRD receiving dialysis. Then, we describe the functioning of the AI-based algorithm under analysis and how it fits in the clinical workflow. Finally, we describe the data set in detail and conclude with the description of the variables to be included in the analysis.

4.1. Clinical setting: Anemia Management in ESRD

Our research focuses on the control of anemia in patients receiving dialysis and suffering from the end stage of chronic kidney disease (CKD), that affects roughly 10% of the adult population worldwide (Kovesdy, 2022). Anemia is a common complication of CKD (Stauffer & Fan, 2014). Managing anemia in end stage renal disease (ESRD) patients receiving hemodialysis entails a delicate trade-off between reaching an optimal target in hemoglobin (Hb) levels while minimizing both potential side effects and costs associated to drug therapies (Erythropoiesis-stimulating agents – ESA – and iron). Symptoms of anemia typically include fatigue, shortness of breath and impaired quality of life. Severe anemia can increase the chance of developing heart problems and an increased risk of complications due to strokes. Clinical guidelines developed by nephrology associations inform anemia management based on the best available scientific evidence. The most

frequently referenced were developed by the independent foundation KDIGO (Kidney Disease: Improving Global Outcomes), which published its clinical practice guidelines for anemia in CKD in 2012 (McMurray et al., 2012)[6]. Guidelines set a target interval for Hb between 10 and 12 g/dL. Once all correctable causes have been addressed (e.g., iron deficiencies), guidelines suggest administering ESA when the Hb level falls below the target (10 g/dL). ESA should not be started when Hb falls above 10 g/dL or administered to move Hb above 12 g/dL. While symptoms for severe anemia (Hb < 9 g/dL) mainly concern quality of life, above the target level risks of mortality or severe consequences are much higher, while the beneficial effects of ESA are not relevant. For this reason, there is a strong recommendation not to use ESA to maintain the Hb level above 13 g/dL. However, the choice to initiate or maintain the ESA treatment and to define the optimal dosage should be individualized based on the patient's clinical history and preferences.

4.2. An AI-based decision-support system for treatment optimization

Our research is based on a CE-marked AI-based medical device (ACM) developed by a company specializing in hemodialysis products, which also operates a network of dialysis centers across various countries. ACM is designed to optimize erythropoiesis-stimulating agents (ESA) prescriptions for patients with ESRD receiving dialysis with a dual objective: (i) improving the attainment of Hb target interval and (ii) reducing the associated treatment costs. ACM operates in two stages. A feedforward artificial neural network trained on over 900,000 prescriptions simulates future Hb trajectories under multiple simulated ESA and iron doses. A subsequent policy extractor module assigns a utility score for each simulated dose via a reward function aligned with clinical practice

[6] An update of the guidelines was published in early 2026 (Tonelli et al., 2026). For the purposes of this analysis, we will refer to the 2012 guidelines as these were considered to be the standard of care for the entire period under analysis.

guidelines regarding Hb targets. The dose yielding the highest utility value ultimately represents the algorithm's recommendation.

The recommendation appears as an alert within the electronic medical record system every time the results of routine monthly blood tests are uploaded. ACM is a decision-support system; hence, physicians may accept, ignore or reject the recommendation, without altering the clinical workflow. A recommendation is ignored by the physician when a new one is available for the patient before any decision has been made on the previous one. Considering the nature of the AI-based algorithm as a CE-marked device, it did not undergo any relevant updates during the period under review. Empirical evidence over years has demonstrated that using ACM is correlated to improved patient outcomes (Bucalo et al., 2018; Barbieri et al., 2016; Garbelli et al., 2024). Although the developer has rolled out ACM across its global network of dialysis centers, actual utilization varies widely across countries, centers, and individual nephrologists, reflecting local autonomy and the professional discretion inherent in clinical practice.

4.3. Data

The dataset used in this study originates from a comprehensive database maintained by the developer. The original database includes every algorithm recommendation generated in more than 10 countries since the beginning of its deployment in 2013 through March 2024. Each record corresponds to an ACM output and is anonymized by attributing unique alphanumeric identifiers (IDs) both to the physician and to the patient. Starting from a comprehensive dataset containing roughly 1.5 million observations, we applied some exclusion criteria to ensure high standard of relevance. The exclusion criteria were defined through a collaborative review process involving the company representatives.

Table 3.1 summarizes the exclusion criteria and their impact on the sample size. First, we removed observations for which country, physician, and patient identifiers were unlikely to be uniquely assigned. Specifically, we excluded observations from countries without clearly distinguishable identifiers (1a). We also excluded observations from patients exhibiting a gap in observations within a 180-day window (1b), to ensure continuity in treatment tracking and avoid potential misidentification of patients over time. In addition, we excluded physician identifiers associated with more than 200 monthly observations (1c), as such unusually high activity levels are unlikely to correspond to a single nephrologist and may instead reflect shared or aggregated identifiers. Second, we limited the sample to observations that reflected the intended monthly routine use of the algorithm, by excluding ignored observations occurring outside the 25-35 day interval from the previous one (2). Third, we excluded outputs classified as "No suggestion", indicating that the patient was not clinically eligible to receive a recommendation and therefore did not require any decision from the nephrologist (3). Lastly, we removed observations with missing values in the relevant variables used in the empirical model (4).

Table 3.1. Sample exclusion criteria

Exclusion criteria	**Observations**	**Sample reduction**
Initial dataset	1,588,330	-
1a. Countries without clearly distinguishable IDs (Asia Pacific, Latin America, Other Europe)	1,508,235	5.04%
1b. Patient identifiers without consecutive observations within 180 days	1,410,922	6.45%
1c. Physician identifiers with more than 200 monthly observations	630,897	55.28%

2. Ignored observations occurring out of the 25-35 day interval from the previous recommendation	608,268	3.59%
3. Observations with "No suggestion" as output	584,487	3.91%
4. Observations with missing values in relevant variables	584,379	0.02%

The final dataset contains 584,379 observations, with each observation containing detailed data at the levels of observation, patient, and physician. Each entry is linked to a unique identifier for the patient, physician, and dialysis clinic. Each record displays the recommended ESA dose, the physician's response (accept/ignore/reject), blood-test results, and the prior-quarter prescription history. Patient-level variables include age, gender, and the number of comorbidities. The physician-level variable only encompasses the physician's tenure with the company.

The dataset is structured to facilitate a comprehensive analysis across multiple dimensions. It comprehends four distinct levels: patient, physician, clinic (i.e., dialysis center), and country. Each recommendation is nested on a single patient, physician, and clinic, while multiple recommendations for the same patient may be managed by different physicians and across multiple clinics. This cross-classified layout supports multilevel modelling (Goldstein, 1994, 2010).

4.4. Variables

In this section, we provide a detailed description of the dependent, independent, and contextual variables. A summary is provided in

Table 3.3.

4.4.1. Dependent Variable

The dependent variable throughout the study is the outcome of the interaction between the prescribing physician and the recommendation generated by the algorithm. A recommendation might be accepted, rejected or ignored by a physician. We measure the dependent variable as a binary variable assuming value one when the recommendation is accepted, and zero otherwise (rejected or ignored recommendations). Overall, 55.3% of the recommendations were accepted by the prescribing physician and 44.7% were not accepted, including rejected (41.4%) and ignored (3.2%) recommendations (Table 3.).

4.4.2. Independent Variables

To test H1, we constructed a six scenarios capturing the decision context by combining (i) the patient's Hb level relative to the guideline target interval – whether the observed hemoglobin level falls below ($< 10\ g/dL$), above ($> 12\ g/dL$), or within the target interval – with (ii) the treatment recommendation on whether to administer ESA ($Recommended\ dose > 0\ UI$). The three hemoglobin levels capture increasing clinical guidelines' prescriptive strength. Guidelines strongly advise against using ESA when Hb is above the target interval, while they advise using ESA when the Hb falls below the target interval. In the remaining cases, the maintenance of the Hb level within the target interval must be evaluated by physicians considering various parameters related to patient health status and require increasing, decreasing, or maintaining the ESA treatment accordingly (Table 3.2).

Table 3.2. Description of the six decision scenarios

Hemoglobin interval	**Administer ESA**	**Not Administer ESA**
Below the target	Fully aligned with guidelines calling for using ESA. Dosing decision remains on the physician	Contradicting with guidelines calling for using ESA
Within the target	Aligned with guidelines calling to define ESA dosing to maintain Hb within the target	Aligned with guidelines calling not to start ESA if already in target
Above the target	Contradicting with guidelines calling for not using ESA	Fully aligned with guidelines calling for not using ESA

To test H2, we rely on two different independent variables that capture how physicians respond to algorithmic recommendations when no strong guidance is available. The first is a continuous variable indicating the observed $Hemoglobin$ level at the time of the recommendation. The second is a categorical variable that encodes the direction of the recommendation relative to the patient's most recent prescription. This variable takes three values and distinguishes between an *increase the ESA dose*, a *decrease the ESA dose*, and a *maintain the ESA dose* recommendation. The variable captures the algorithm's stance relative to the patient's prior ESA administering, offering a proxy for how recommendations align or diverge from the established prescription pattern. To assess how the effect of recommendation direction varies across the Hb distribution, we also include an interaction term between the hemoglobin level and the recommendation direction.

At patient level, we also include categorical variables for Age and $Number\ of\ comorbidities$, both providing additional information on the complexity of the decision to take. At physician level, we include variables describing domain

experience ($Tenure\ in\ the\ clinic$) and familiarity of the physician with the algorithm ($Time\ from\ first\ decision$ and $Number\ of\ decisions$). The variable describing domain experience reflects the duration of the physician's tenure in one of the company's clinics and is among the physician-level information included in the original dataset. This measure does not represent an absolute assessment of domain experience, as it does not account for any prior professional experiences the physician may have had before joining the company. This approach reflects the concept of seniority experience and has already been used in recent scholarly works (Allen & Choudhury, 2022; Choudhury et al., 2020; W. Wang et al., 2023). Familiarity with the algorithm – or task-based experience – is another dimension of experience capturing previous exposure with the algorithm (W. Wang et al., 2023). The first variable captures the time (in months) from the day the physician addressed the first algorithmic advice. The second variable measures the cumulative volume of recommendations received by the physician throughout the entire period. When included in the models, continuous variables are always standardized to have comparable one-standard-deviation changes and facilitate comparison across measures that are expressed in different scales.

4.4.3. Contextual variables

The dataset contains contextual variables allowing us to assign each observation to a specific patient, physician, clinic, and country. Patient, physician, and clinical identifiers were used as random control variables in the model. The information on countries was used as a country fixed effect, and the categories related to eastern European countries merged.

Table 3.3. Description of the included variables

Variable	Type	Description	Time-invariant
Acceptance	Binary	Decision regarding the recommendation received (accepted vs non-accepted)	No
Decision scenarios	Categorical (6)	Decision scenarios defined from the interaction between the hemoglobin level, the target interval, and the type of recommendation (Administer vs Non administer ESA)	No
Recommendation direction	Categorical (3)	Recommendation direction defined from the difference between the recommended ESA dose and the prior patient's prescription (Increase, Decrease, Maintain the ESA dose)	No
Age	Categorical (4)	Age of the patient at the time of the recommendation, split into age groups (18-45, 45-65, 65-75, 75+)	No
Number of comorbidities	Categorical (4)	Number of comorbidities at the time of the recommendation, split into categories (None, 1, 2, 3 or more)	No
Tenure in the clinic (Std)	Continuous	Standardized number of years of tenure with one of the clinics of the company	No
Time from the first decision (Std)	Continuous	Standardized difference in months between the day of the first recommendation and the current recommendation	No
Number of decisions (Std)	Continuous	Standardized cumulative number of decisions taken by a physician, starting from the first	No
Patient ID	Index	Patient identifier	Yes
Physician ID	Index	Physician identifier	Yes
Clinic ID	Index	Dialysis center identifier	Yes
Country	Categorical (4)	Country where the clinic is located. In the models, eastern European countries were merged.	Yes

5. EMPIRICAL STRATEGY

Our empirical approach consists of three stages. First, we employ a Bayesian Multilevel Variance Decomposition to decompose the variance in acceptance across the four contextual levels of our data: patient, physician, clinic, and country. This allows us to quantify the relative importance of each level before modeling specific predictors. Second, recognizing that random intercepts may correlate with regressors, we proceed to a series of High-Dimensional Fixed Effect Linear Probability Model (LPM) to explore the association between key predictors and the probability of accepting an AI-based treatment recommendation.

5.1. Bayesian Multilevel Variance Decomposition

Multilevel data are common in social sciences and the interest in analyzing and interpreting them has its root in educational and sociological research (Hox & Roberts, 2011; Sommet & Morselli, 2017). In healthcare as well, research into the application of multilevel modelling has surged over time as many of the problems studied involve different levels and contexts (Barker et al., 2020; Leyland & Groenewegen, 2020). Context is important as many problems in healthcare are related to people's behavior and the result of the interaction between patients and providers is often the result of the interplay between different intermediate levels: the institutional system, the practice or the organization of the provider, and the interaction between the physician and the patient.

Before introducing explanatory variables, we fit a Bayesian multilevel logistic model with only an overall intercept and random intercepts for patient, physician, clinic, and country. Bayesian estimation methods such as Markov Chain Monte Carlo (MCMC) simulations have gathered a growing interest in methodological and empirical research (Fielding & Goldstein, 2006; Guo, 2017; Hox & Roberts, 2011). The advantages of Bayesian MCMC

approaches rely on their flexibility and computational feasibility with respect to more traditional maximum-likelihood approaches. Due to their flexibility, MCMC estimations are more likely to work for complex multilevel models (Berger, 2006). We estimate the following equation

$$\text{logit}(\Pr(Y_{mikcl} = 1) = \beta_0 + u_i + w_k + z_c + c_l \qquad (1)$$

where Y_{mikcl} denotes whether the recommendation m, for patient i, by physician k, in clinic c, and country l, is accepted. The terms $u_i \sim N(0, \sigma_u^2)$, $w_k \sim N(0, \sigma_k^2)$, $z_c \sim N(0, \sigma_z^2)$, and $c_l \sim N(0, \sigma_c^2)$ are random intercepts capturing heterogeneity across patients, physicians, clinics, and countries respectively. We assign all parameters weakly informative priors. Estimations via PyMC provides posterior distributions for each variance component. From these, we compute intraclass-correlation measures on the logit scale, to estimate the share of total variance attributed to each contextual level. Although this null model does not explain acceptance through covariates, it quantifies dependance structures and informs subsequent model specifications.

5.2. High-Dimensional Fixed Effect Linear Probability Models

Because random intercepts in multilevel models can correlate with observed predictors, implicating a risk of biased estimates, we adopt high-dimensional fixed effects LPM to investigate the association between acceptance and our key predictors. LPM facilitates straightforward inclusion of high-dimensional fixed effects without the incidental parameters concern that arises in nonlinear fixed-effect logit models. In this specification, acceptance Y_{mikc} is regressed on a vector of time-varying covariates X_{mikc} and different combinations of fixed effects. We estimate the following equation:

$$Y_{mikc} = \beta_0 + \beta^\top X_{mikc} + \alpha_i + p_k + \gamma_c + \tau_t + \varepsilon_{mikc} \qquad (2)$$

where X_{mikc} includes decision-scenarios indicators reflecting guideline-alignment, measures of alignment with prior treatment behavior, measures of physician's domain experience and familiarity, and patient's control variables. α_i, p_k, γ_c are patient, physician, and clinic fixed effect, absorbing all time-invariant attributes at patient-, physician-, and clinic-level. τ_t is a temporal fixed effect. Equation (2) represents the baseline modeling structure. We estimate a sequence of models with progressively richer fixed-effects structure to assess robustness and capture potential interaction among levels. We also include physician-patient interaction fixed effects to control for pair-specific tendencies, physician-year interaction fixed effects to capture physician-specific trends in reliance on the algorithm, and patient-quarter interaction fixed effects to absorb seasonal variation in patient's health status and conditions.

6. RESULTS

6.1. Descriptive Statistics and Variance Decomposition

Table 3.4 summarizes the descriptive statistics of the full sample (N = 584,431). The average Hb level across observations is 11.22 g/dL (SD = 1.26), placing most patients within the target guideline-recommended target interval (10-12 g/dL). Figure 3.1 also shows the distribution of Hb levels across all observations, reporting a concentration (65%) of observation within the target interval and a slightly right-skewed distribution. Physicians in the dataset have on average 12.89 years (SD = 3.39) in their clinic at the time of each observation, have been using the algorithm for approximately 44.4 months (SD = 32.46), and made around 1,697 decisions using the algorithm (SD = 1,719), indicating substantial heterogeneity in experience and familiarity at the point of decision-making. Regarding treatment decisions, the average recommended ESA dose is 15,089 UI (SD = 17,760), closely matching the prescribed dose (15,433 UI; SD = 20,260),

suggesting an alignment between algorithm recommendations and physician prescriptions. In terms of patient patients' demographics, 60.4% of observations refer to male patients, approximately 28.7% of observations are for patients with one comorbidity, while about 43.5% have two or more comorbidities. Age distribution is skewed toward older adults, with 26.9% of observations involving patients aged 65–75 years and 39.9% patients aged over 75 years. Across all records, physicians accepted the AI-based recommendation in 55.3% of cases, rejected it in 41.4%, and ignored it in 3.2%. Data span eight countries, with Portugal (57.0%) and Spain (22.9%) providing the majority of observations. The sample includes unique identifiers from 26,819 patients, 1,076 physicians, and 170 clinics.

Table 3.4. Summary statistics of the study sample

Variables	**Mean**	**Standard deviation**
Hemoglobin level (g/dL)	11.221	1.256
Physician's tenure in clinic (years)	12.892	3.385
Physician's time from first decision (months)	44.372	32.458
Physician's number of decisions	1,696.697	1,718.804
Recommended dose (UI)	15,089.29	17,760.15
Prescribed dose (UI)	15,433.12	20,259.67
Variables	**Count**	**Percentage**
Patient gender: male	353,011	60.41
Number of comorbidities		
None	162,402	27.79
1	167,756	28.71
2	125,883	21.54
3+	128,338	21.96
Patient age (years)		
18-45	41,707	7.14
45-64	152,130	26.03
65-75	157,138	26.89
75+	233,404	39.94

Physician's decision		
Accepted	323,355	55.33
Rejected	242,156	41.44
Ignored	18,868	3.23
Country		
Czechia	5,441	0.93
France	23,454	4.01
Hungary	2,047	0.35
Poland	19,936	3.41
Portugal	333,384	57.05
Romania	37,996	6.50
Slovakia	28,313	4.84
Spain	133,809	22.90
Patients (count)	26,814	-
Physicians (count)	1,066	-
Clinics (count)	170	-
N.	584,379	-

Figure 3.2 shows the average ESA dose recommended by the ACM across the distribution of observed Hb levels. As expected, the suggested dose decreases steadily as Hb

increases, with a sharp decline as levels approach the upper threshold. Above 12 g/dL, recommended doses approach zero.

Figure 3.3 plots the proportion of algorithm outputs recommending not to administer ESA, across Hb levels. The proportion remains near zero for low Hb values, begins to rise around 11 g/dL, and converges to 100% above 13 g/dL. These patterns reinforce the idea that the algorithm is calibrated to align with clinical guidelines, which discourage ESA treatment above the target range, and inform our classification of decision scenarios based on hemoglobin intervals and ESA recommendation. Finally,

Figure 3.4 plots our dependent variable – the average acceptance rate of the algorithm's recommendation – across the observed Hb levels. Acceptance remains below 50% throughout the discretionary zone (< 12 g/dL), then rises steeply and exceeds 90% once Hb is clearly above 12 g/dL. This pattern already mirrors the norm-strength pattern that is documented in the following empirical models.

Figure 3.1. Distribution of hemoglobin levels

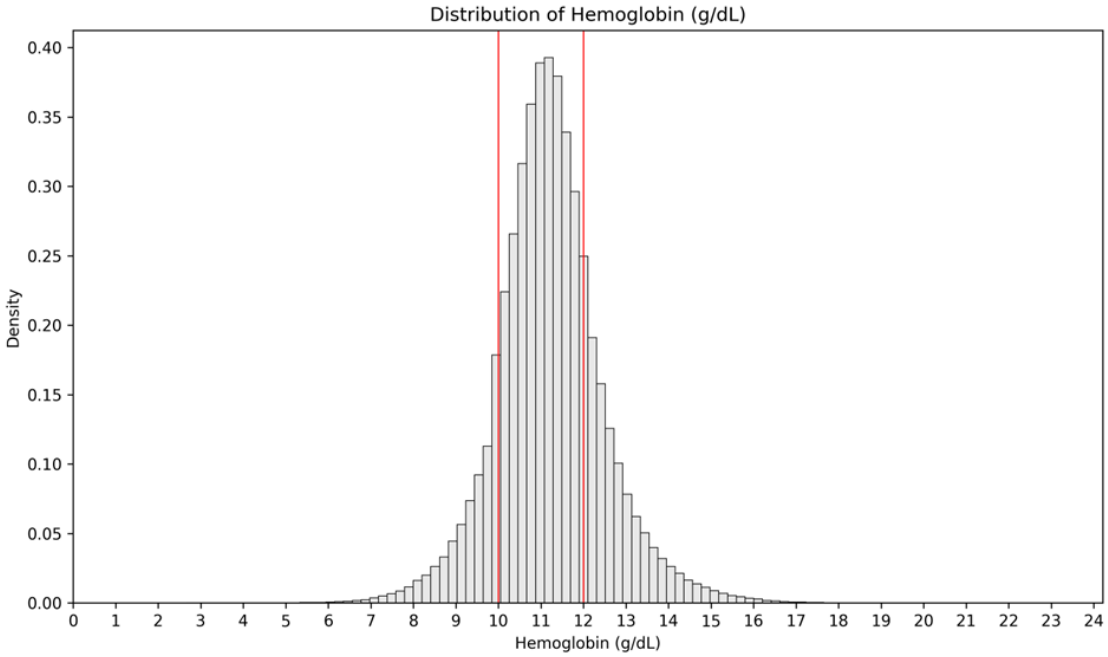


Figure 3.2. Suggested dose and hemoglobin level (p1, p99)

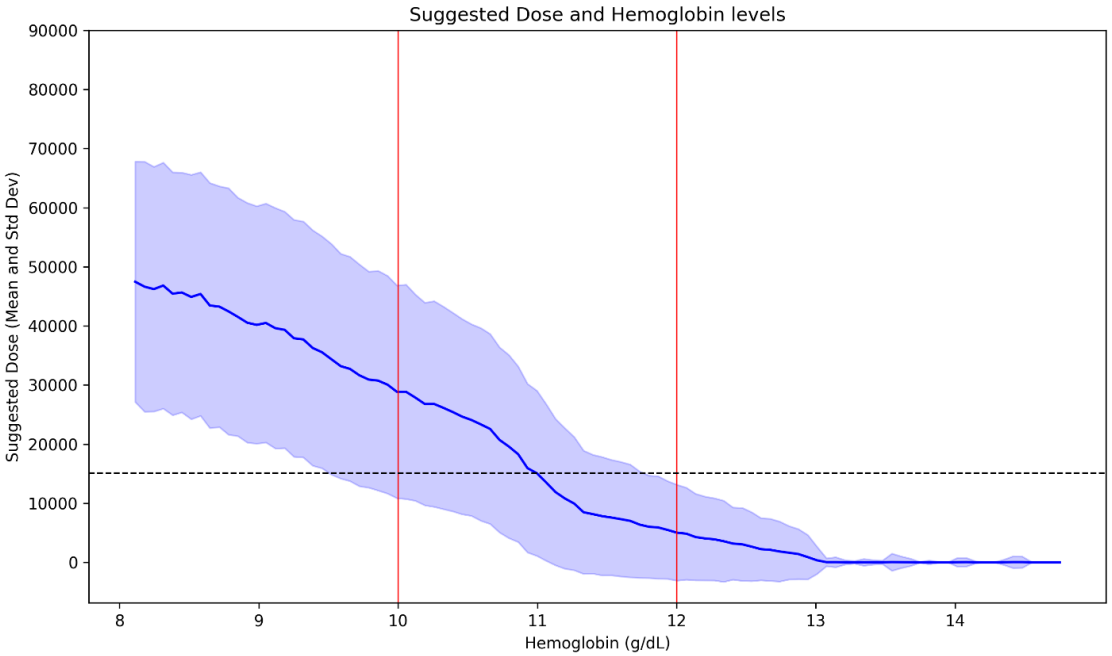


Figure 3.3. Recommendation not to administer ESA and hemoglobin level (p1, p99)

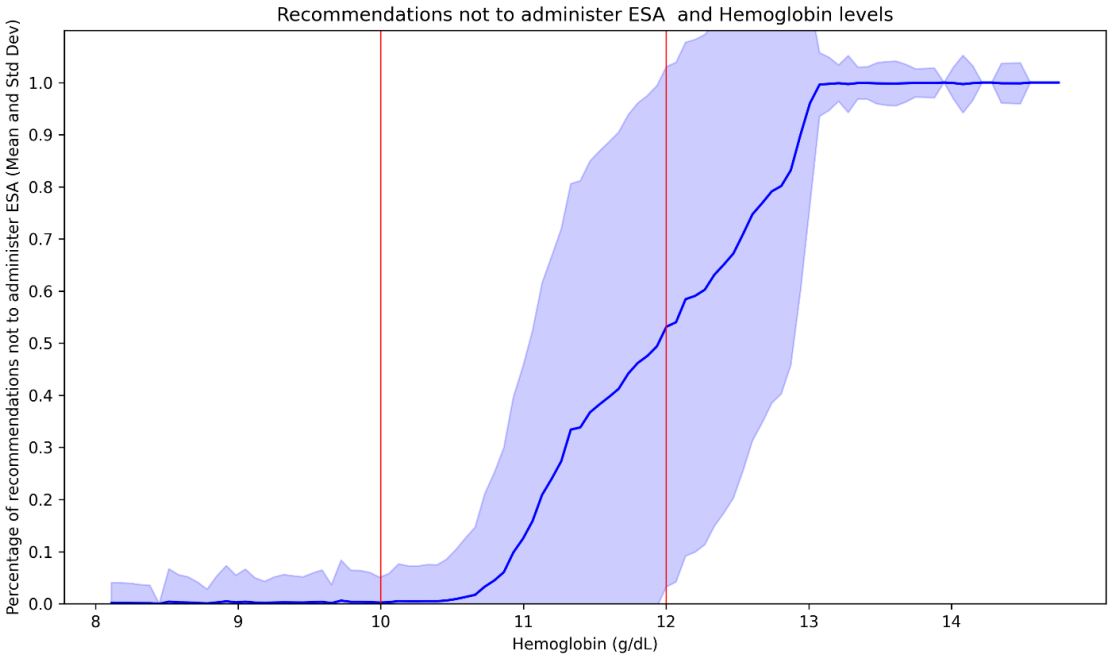
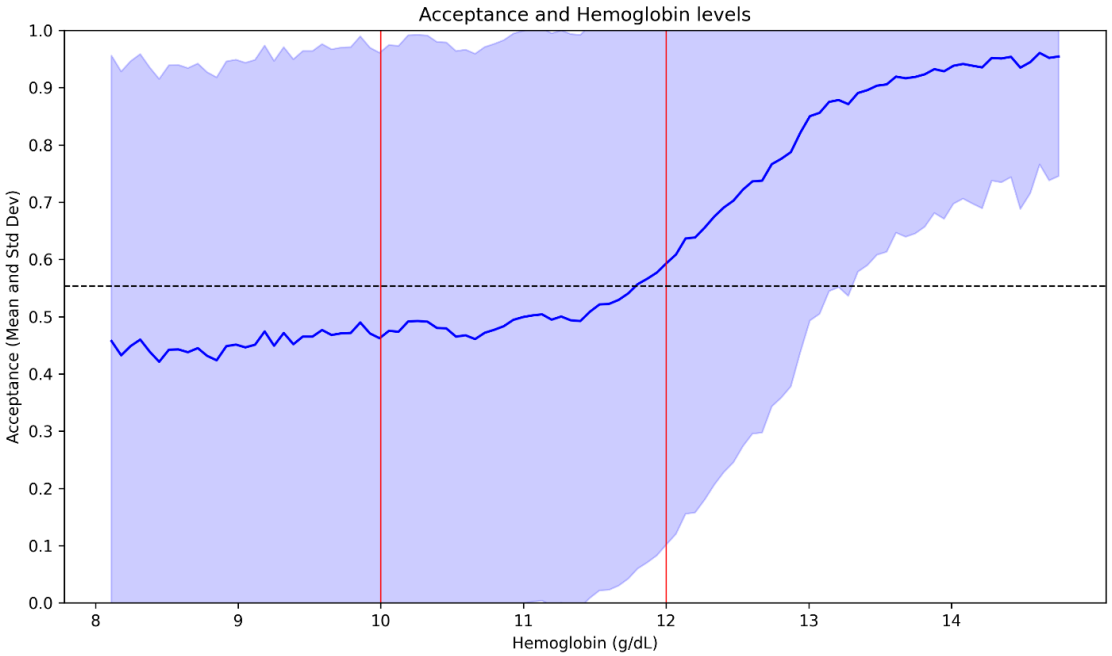


Figure 3.4. Acceptance rate and hemoglobin level (p1, p99)



The first empirical result provides an overview of the contextual factors shaping physician's acceptance of algorithmic recommendations (Table 3.5). By partitioning the total variance, we can see that the majority of variance is systematic and can be attributed to the specific context of the clinic, the patient, the physician, and the country. The largest share of variance is attributable to the dialysis centers (ICC = 20.3%), suggesting how factors related to the organizational environment, decision-making culture, and local protocols shape how physicians interact with the system. The patient is the second most important factor (ICC = 16.4%), implying that decisions are anchored to patient's specific case. This variance likely captures time-invariant patient's characteristics like the overall health status and the relationship with the prescribing physician, as well as time-varying characteristics related to patient's condition at a specific point in time. Physician's habits and characteristics are less influential than the previous contextual variables (ICC = 9.1%), suggesting how physicians are more constrained by patient- or clinic-related factors when exercising their judgement. Country-level effects are just as strong as physician-level effect (ICC = 9.6%), indicating that system-level factors are as important

as the human decision-maker making the final decision. However, this last result should be interpreted in light of the clusters at each level. While the point estimate for the country-level suggests an influence comparable to that of individual physician, the estimate is associated with considerable uncertainty (the 95% credibility interval for the ICC is between 2.1% and 15.2%). This is an expected consequence of the relatively small number of clusters (8 countries) at the highest level of our model. In contrast, the variance components of the clinic (N = 170), physician (N = 1,066), and patient (N = 26,814) are estimated with much higher precision.

Table 3.5. Results from the Bayesian Variance Decomposition analysis

Contextual Level	**N (clusters)**	**Variance Component (σ^2)**	**Intra-Class Correlation (ICC)**
Patient	26,814	1.206 [1.171 – 1.240]	16.4% [14.8% - 18.5%]
Physician	1,066	0.669 [0.582 – 0.756]	9.1% [9.0% - 9.2%]
Clinic (dialysis center)	170	1.491 [1.143 – 1.839]	20.3% [18.1% - 21.9%]
Country	8	0.706 [0.130 – 1.282]	9.6% [2.1% - 15.2%]
Residual	-	-	44.7% [39.1% - 52.1%]

Notes: The residual variance is a fixed property of the logistic distribution and is not estimated. 95% Credibility Intervals are shown in parentheses for the estimated parameters.

6.2. Determinants of algorithm acceptance: decision- and physician-level predictors

Table 3.6 presents the results of our High-Dimensional Fixed Effect Linear Probability Models, which test our hypotheses on the drivers of algorithm acceptance. We specify a sequence of six models to show the robustness of our findings to an increasingly specific set of fixed effects. The most comprehensive specification (Column 6) includes a set of

clinic, patient, physician, and interaction fixed effects to absorb a wide range of unobserved heterogeneity.

Confirming our H1, we find that alignment with clinical guidelines is a powerful predictor of acceptance. This is evident from the coefficients of the decision scenarios variables, which are stable and significant across all models. In the most comprehensive specification, a recommendation that is fully aligned with a strong guideline (Scenario 1: not administer ESA when Hb falls above the target range) is 23 p.p. more likely to be accepted than the baseline category (administer ESA when Hb falls within the target range). Conversely, recommendations that directly contradict clear guidelines are strongly rejected: Scenario 5 is associated with a 37 p.p. decrease in probability of acceptance. This large negative effect is particularly noteworthy given that such clear contradictions by the algorithm are infrequent in our data (N = 222), suggesting that while the algorithm seldom deviates from unambiguous guideline norms, physicians' rejection is decisive when it does.

The results show a large and consistent effect for a preference for status quo (H2). The coefficient for the recommendation that is aligned with the prior treatment is positive, highly significant, and substantial in magnitude across all the relevant specifications. In the most comprehensive model, a recommendation to maintain the prior dose is approximately 30 p.p. more likely to be accepted than a recommendation that suggests a change, holding all other factors constant. This indicates that therapeutic inertia is one of the most dominant behavioral forces in our setting.

We test H3 and H4 regarding the role of physician-level attributes. Contrary to our expectations (H3), our models reveal a positive effect for domain experience (a one

standard deviation increase in tenure increases acceptance probability by 6 p.p. in the most comprehensive model). Specifications that do not control for time trends find a negative association, as more experienced physicians are naturally observed in later years, which exhibits a general negative trend in overall acceptance. The inclusion of year fixed effects absorbs these confounding trends. The remaining positive coefficient therefore represents the within-physician effect of experience.

*Table 3.*Table 3.7 presents a separate set of models that replace the measure of domain experience with the two distinct measures of familiarity: the time elapsed since the first use and the cumulative volume of recommendations each physician has handled. When testing time from the first use, the coefficient becomes positive and slightly significant once the full suite of fixed effects is included. This provides some evidence that physicians with longer familiarity in using the system might be more likely to accept the recommendation. As for the alternative measure (cumulative use), across both the simpler specification and fully specified model the coefficients are close to zero and not statistically significant.

We do not include the measure for domain experience and time-based familiarity in the same model simultaneously. As both are time-based measures, they are highly collinear, and our models already include a full set of year fixed effects to absorb time trends, their joint inclusion would measure the same within-physician phenomenon and lead to unstable and unreliable coefficient estimates. We therefore chose to present their effects in separate models to provide a clearer and more stable assessment of each construct's potential influence.

Table 3.6. Results from HDFE Linear Probability Model (1)

Linear Probability Model Full sample	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Scenario 1: Above target & Not Administer ESA	0.34***	0.19***	0.34***	0.20***	0.28***	0.23***
Scenario 2: In target & Not Administer ESA	0.17***	0.05***	0.17***	0.05***	0.11***	0.05***
Scenario 3: Below target & Administer ESA	0.01	0.00	0.01***	0.01***	0.03***	0.04***
Scenario 4: Above target & Administer ESA	-0.06***	-0.03***	-0.05***	-0.03***	-0.02***	-0.002
Scenario 5: Below target & Not Administer ESA	-0.36***	-0.43***	-0.35***	-0.42***	-0.33***	-0.37***
Prior treatment: Recommendation = prior treatment	-	0.38***	-	0.37***	-	0.30***
Tenure in the clinic (std)	0.01	0.01	0.01*	0.01**	0.09**	0.06*
Patient control (age, gender, comorb)	X	X	X	X	X	X
Physician FE	X	X	X	X	X	X
Clinic FE	X	X	X	X	X	X
Patient FE	X	X	X	X	X	X
Physician x Patient FE	-	-	X	X	X	X
Patient x Quarter FE	-	-	-	-	X	X
Physician x Year FE	-	-	X	X	X	X
Year FE	X	X	-	-	-	-
N.	581,203	581,203	562,886	562,886	523,332	523,332

Adj R^2	0.3057	0.3912	0.3264	0.4081	0.4037	0.4516
Within R^2	0.0549	0.1712	0.0514	0.1664	0.0242	0.1025

*Notes: Robust standard errors are used. Significance levels are denoted as follows: * p < 0.10, ** p < 0.05, *** p < 0.01.*

Table 3.7. Results from HDFE Linear Probability Model (2)

Linear Probability Model ***Full sample***	**(1)**	**(2)**	**(3)**	**(4)**
Scenario 1: Above target & Not Administer ESA	0.19***	0.23***	0.19***	0.23***
Scenario 2: In target & Not Administer ESA	0.05***	0.05***	0.05***	0.05***
Scenario 3: Below target & Administer ESA	0.001	0.04***	0.001	0.04***
Scenario 4: Above target & Administer ESA	-0.03***	-0.002	-0.03***	-0.002
Scenario 5: Below target & Not Administer ESA	-0.43***	-0.37***	-0.43***	-0.37***
Prior treatment: Recommendation = prior treatment	0.38***	0.30***	0.38***	0.30***
Time from the first use (months, std)	0.01	0.05*	-	-
Cumulative use (std)	-	-	0.00	0.02
Patient control (age, gender, comorb)	X	X	X	X
Physician FE	X	X	X	X
Clinic FE	X	X	X	X
Patient FE	X	X	X	X
Physician x Patient FE	-	X	-	X
Patient x Quarter FE	-	X	-	X
Physician x Year FE	-	X	-	X
Year FE	X	-	X	-
N.	581,255	523,340	581,255	523,340
Adj R^2	0.3912	0.4517	0.3912	0.4517
Within R^2	0.1712	0.1026	0.1712	0.1025

*Notes: Robust standard errors are used. Significance levels are denoted as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Having established the primary, direct effects of decision- and physician-level factors, we also test how physician characteristics moderate the influence of these powerful recommendation-based effects. Table 3.8 presents the results of two interaction models, where we interact both domain experience and algorithm familiarity measures with the key drivers of acceptance: guideline alignment and prior-treatment alignment. Here, guideline alignment is measured by merging homogeneous decision scenarios. Scenario 1 remains the one aligned with the strongest guideline recommendations, while Scenarios 4 and 5 diverge from the suggestions in the guidelines. The remaining scenarios are the baseline categories, representing situations where the prescriptive strength of the guidelines is weaker. The results for guideline alignment (increase of acceptance probability by 17 to 19 p.p. when the recommendation is aligned with guidelines and decrease by 1 to 4 p.p. when the recommendation is diverging from the guideline) and prior treatment alignment (increase by 31 to 38 p.p. in acceptance probability) are consistent with those presented in the previous model specifications. The first two models explore the moderating effect of a physician's domain experience. First, the main effect of tenure is still positive and marginally significant, suggesting that in baseline scenarios, more experienced physicians are slightly more likely (5 p.p.) to accept AI's recommendation once all interaction effects are accounted for. We also find positive and significant interaction effects. The positive effect is amplified when the recommendation is supported by a strong guideline and more experienced physicians become even more likely to accept the recommendation when it is consistent with norms. Even if statistically significant, the same coefficient for the interaction with guideline-diverging recommendation is close to zero, as well as the coefficient for the interaction with the

prior treatment alignment. The results for familiarity show both similarities and differences. Similar to domain experience, time-based familiarity has a positive effect on acceptance in the baseline decision scenarios (5 p.p.) and in presence guideline-aligned recommendations (3 p.p.). However, the key difference lies in the other interaction effects, as familiarity does not significantly moderate the response to guideline-diverging and prior treatment-aligned recommendations. However, these findings are highly sensitive to model specification, as the effects only become statistically significant once a full set of interaction fixed effects are included. Given this sensitivity, the finding cannot be considered robust across different specifications.

Table 3.8. Results from HDFE Linear Probability Model (3)

Linear Probability Model ***Full sample***	**(1)**	**(2)**	**(3)**	**(4)**
Guideline-aligned: Above target & Not Administer ESA	0.17***	0.19***	0.17***	0.19***
Guideline-diverging: Above target & Administer ESA Below target & Not Administer ESA	-0.04***	-0.01***	-0.04***	-0.01***
Prior treatment: Recommendation = prior treatment	0.38***	0.31***	0.38***	0.31***
Tenure in the clinic (years, std)	0.004	0.05*	-	
Time from the first use (months, std)	-	-	0.002	0.05*
Tenure in the clinic x Guideline-aligned	0.02***	0.03***	-	-
Tenure in the clinic x Guideline-diverging	0.01*	0.01**	-	-
Tenure in the clinic x Prior treatment	0.01***	-0.005*	-	-
Time from start x Guideline-aligned	-	-	0.02***	0.03***
Time from start x Guideline-diverging	-	-	-0.01	0.003
Time from start x Prior treatment	-	-	0.01***	-0.01***
Patient control (age, gender, comorb)	X	X	X	X

Physician FE	X	X	X	X
Clinic FE	X	X	X	X
Patient FE	X	X	X	X
Physician x Patient FE	-	X	-	X
Patient x Quarter FE	-	X	-	X
Physician x Year FE	-	X	-	X
Year FE	X	-	X	-
N.	581,203	523,332	581,255	523,240
Adj R^2	0.3905	0.4508	0.3906	0.4509
Within R^2	0.1702	0.1011	0.1705	0.1012

*Notes: Robust standard errors are used. Significance levels are denoted as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

6.3. Drivers of AI acceptance in ambiguous decision scenarios

To further disentangle the influence of different behavioral drivers, we estimate our models on a subsample of observations that exclude scenarios with strong and unambiguous guideline recommendations. This subset only includes Scenario 2 ("In target & Not Administer ESA") and Scenario 3 ("Below target & Administer ESA") with "In target & Administer ESA" as a baseline scenario. We also differentiate the reference to the prior treatment by distinguishing non-aligned recommendation between recommendation to increase the dose and recommendations to decrease the dose. Table 3.9 presents the results from two specifications of our HDFE models estimated on the restricted sample. First, in absence of strong prescriptive guidance, physicians exhibit stronger therapeutic inertia: recommendations to increase (-0.40 to -0.41 p.p. on the acceptance probability) or decrease (-0.36 to -0.44 p.p.) the prior ESA dose are significantly less likely to be accepted compared to recommendations that replicate prior prescription. These effects are robust across specifications and highlight the strong preference for status quo when physicians are not normatively guided. Second, the

interaction terms between recommendation direction and the two decision scenarios reveal a more nuanced behavioral pattern. In Scenario 3, a recommendation to increase the dose significantly offsets the status quo penalty – associated with a positive marginal effect of 27–33 p.p. – suggesting that physicians are particularly receptive to upward corrections in the dose when the algorithm signals that the patient's condition might require it. These effects reveal that discretion does not lead to randomness but is still sensitive to implicit clinical expectations and the direction of dose adjustment. Finally, we explore how physician tenure moderates these patterns of decision-making under discretion. Interaction terms show that more experienced physicians are more likely to accept recommendations in Scenario 2 and more reluctant to accept recommendations in Scenario 3 compared to the baseline category. The negative and significant interaction between tenure and the decrease recommendations suggests that more experienced physicians are less likely to accept downward adjustments in ESA dose. These findings suggest that experience may shape physicians' perception of risk and their willingness to deviate from the status quo – particularly when the algorithm proposes clinically consequential changes. Together, these results provide further evidence that algorithmic acceptance is not only sensitive to guidelines but also shaped by behavioral tendencies like inertia, which become particularly salient in ambiguous clinical contexts.

Table 3.9. Results from HDFE Linear Probability Model (4)

Linear Probability Model ***Subsample: Ambiguous Decision Scenarios***	**(1)**	**(2)**
Scenario 2: In target & Not Administer ESA	0.08***	0.01
Scenario 3: Below target & Administer ESA	-0.19***	-0.21***
Increase wrt prior treatment:	-0.40***	-0.41***

Recommendation > prior treatment		
Decrease wrt prior treatment: Recommendation < prior treatment	-0.44***	-0.36***
Decrease x Scenario 2	-0.07***	0.04***
Increase x Scenario 2	(no obs.)	(no obs.)
Decrease x Scenario 3	0.01	0.10***
Increase x Scenario 3	0.27***	0.33***
Tenure in the clinic (years, std)	0.01	-0.03
Tenure in the clinic x Scenario 2	0.01***	0.02***
Tenure in the clinic x Scenario 3	-0.01***	-0.01***
Tenure in the clinic x Decrease	-0.02***	-0.01**
Tenure in the clinic x Increase	0.002	0.01***
Patient control (age, gender, comorb)	X	X
Physician FE	X	X
Clinic FE	X	X
Patient FE	X	X
Physician x Patient FE	-	X
Patient x Quarter FE	-	X
Physician x Year FE	-	X
Year FE	X	-
N.	449,628	382,910
Adj R^2	0.3551	0.4137
Within R^2	0.1680	0.1190

*Notes: Robust standard errors are used. Significance levels are denoted as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

7. DISCUSSION AND CONCLUSIONS

This study investigates how physicians respond to AI-based treatment recommendations in a clinical setting where algorithmic recommendations are calibrated to data and clinical practice guidelines. Our results shed light on the multilevel determinants of physicians'

reliance on AI-based algorithms and highlight the interplay between normative anchors, habitual behavior, and individual-level characteristics. We find that alignment with clinical guidelines and prior treatment behavior are the strongest and most consistent driver of algorithm acceptance. In the absence of unambiguous guidance, physicians stand firm on the prior behavior, exhibiting a pronounced tendency toward inertia. Moreover, domain experience operates as an individual-level moderator rather than a simple main effect. In ambiguous decision scenarios, where prescriptive guidance is weaker, acceptance is strongly shaped by whether the recommendation implies a change relative to prior dosing and by the direction of that change. Experience further differentiates these patterns: more experienced physicians respond differently to upward versus downward adjustments across scenarios, suggesting that the same recommendation may be interpreted differently depending on accumulated experience.

Our findings support the argument that physicians' acceptance of algorithmic recommendations is governed by the presence of strong prescriptive norms defined by their professional community. When the recommendation is fully aligned with guidelines that clearly indicate a unique course of action, physicians are much more likely to accept it. This result is consistent with norm theory (Kahneman & Miller, 1986), which posits that people are more comfortable taking actions that can be justified with reference to socially or institutionally grounded norms. In medical decision-making, this translates into a preference for decisions that are justifiable in terms of established clinical practice. Alignment with strong guidelines appears to serve as a normative anchor that legitimizes experts' reliance on algorithmic recommendations, thereby reducing the perceived need for individual accountability in case of poor outcomes. This tendency becomes especially

salient in high-risk contexts such as higher-than-normal hemoglobin, where decision errors can have significant patient consequences.

In decision contexts characterized by ambiguity or weaker guidance, we observe that physicians exhibit a strong preference for maintaining the status quo. Our results confirm that recommendations aligned with prior treatment behavior are much more likely to be accepted than those suggesting a change. This pattern aligns with prior research on status quo bias (Samuelson & Zeckhauser, 1988) and clinical inertia (Camilleri & Sah, 2021; Giugliano & Esposito, 2011), which document a tendency among physicians to maintain continuity with prior actions and treatment decisions. Such continuity may reflect both behavioral anchoring and clinically reasonable caution in the face of uncertainty. Especially when guidelines do not dictate a clear course of action, the prior behavior becomes the salient benchmark, and deviating from it introduces reduces the availability of external justification. In this sense, therapeutic inertia serves as an internalized norm that substitutes for external prescriptive guidance and hampers the physician's willingness to rely on external algorithmic support.

Our findings also suggest that physicians exhibit a form of asymmetrical risk perception when evaluating whether to accept a recommendation that implies a change. Specifically, changes that involve an increase or decrease in dosage are more likely to be rejected when they deviate from prior behavior, especially in the absence of strong guideline backing. This suggests that physicians are more sensitive to potential negative outcomes (e.g., adverse reactions, over- or under-treatment) than to potential gains in efficacy or safety. Such behavior is consistent with the predictions of prospect theory (Kahneman & Tversky, 1979), where losses seem larger than gains, and decision-makers exhibit regret aversion when anticipating potential errors.

Taken together, our findings point to a misalignment between the logic of AI-based decision systems and the behavioral tendencies of human physicians. AI systems are built on probabilistic action-oriented, utility-maximizing logic: they use probabilistic models optimize recommendations based on current patient data, seeking to maximize expected treatment effectiveness. Physicians, by contrast, operate with rationality shaped by institutional norms, prior behavior, and personal accountability. This mismatch becomes particularly salient in ambiguous clinical scenarios, where AI proposes a change while the physician lacks a strong justification for deviating from the current practice. In these cases, algorithmic rationality – focused on what should be done now - clashes with the physician's backward-looking reasoning about what has worked before. This highlights a broader challenge for clinical AI adoption: beyond accuracy and performance, algorithmic systems must either be consistent to the cognitive structure of medical decision-making or be interpretable for physicians to understand and endorse the logic guiding their recommendations.

The main findings reported in this study are robust to a range of additional checks designed to assess the stability of our estimates. Specifically, we verified that results remain consistent when restricting the sample to the first observation per physician, when distinguishing between prior treatments prescribed by the same physician and those prescribed by a different physician, and when defining the status quo based on treatment decisions older than the immediately preceding one. Furthermore, exclusion tests that systematically drop one country or one year at a time confirm the persistence of the main effects across subsamples. These checks lend further credibility to our interpretation of the behavioral patterns underlying algorithm acceptance.

Nonetheless, this work has several limitations. The main one is the lack of a clear quasi-random source of variation in algorithm recommendations. While the algorithm is designed to reflect clinical norms, we cannot fully disentangle whether changes in physician behavior are due to the content of the recommendation or to unobserved patient or physician factors. Moreover, our data does not include observations prior to the introduction of the AI system, preventing us from conducting within-physician before-after comparisons. Identifying a natural experiment or other quasi-experimental variation in future research would strengthen the causal interpretation of the observed behavioral responses and enable more definitive conclusions about the effect of algorithmic advice on clinical decision-making.

CONCLUSIONS

This dissertation develops a multilevel account of how AI is implemented in large, research-oriented, professional organizations. The findings show that contemporary scholarship and practice increasingly move beyond the dichotomy between automation and augmentation, recognizing adoption as a sociotechnical process that unfolds across technical, organizational, and professional domains. AI implementation is not a discrete event but an ongoing negotiation, in which the interplay of infrastructures, governance mechanisms, and professional logics continuously shape outcomes. At the organizational level, decisions surrounding AI can be understood as arrangements for producing, sharing, and valorizing knowledge. These include choices on exploration versus exploitation, whether to leverage internal research and development for new solutions, or acquire them externally, or ally with partners, and how to balance internal competencies with external sources of knowledge. Such arrangements are not uniform, but path-dependent and conditioned by organizational legacies, institutional constraints, and the specificities of healthcare as a professionalized and highly regulated sector. The professional dimension emerges as equally central. Professionals act both as gatekeepers

and enablers of innovation. As gatekeepers, they shape internal demand from interactions with the market and the professional networks they are embedded in. As end-users, their responses are strongly influenced by established norms and behaviors, and professional status considerations. This reflects a deeper tension between AI rationality and professional rationality. The insights together suggest that successful implementation of AI requires recognizing of this multilevel complexity. Rather than focusing on technical performance and short-term benefits, organizations must orchestrate strategies that align knowledge arrangements with professional logics, while acknowledging the institutional and regulatory environment in which they operate.

The findings of this dissertation suggest several implications for policy and practice in the governance of the adoption of AI in healthcare. From a regulatory perspective, certification processes and frameworks should consolidate approaches that recognize the learning nature of AI technologies. Unlike other digital and traditional technologies, AI systems evolve through continuous improvement, and policies must ensure that evaluation and oversight mechanisms are able to accompany this dynamic dimension while safeguarding quality and patient safety. Economic policies should foster investment in data, infrastructure, and human capabilities at sufficient scale to sustain breakthrough innovation. Fragmented or undersized initiatives are unlikely to deliver transformative outcomes. At the same time, the appropriate scale and mode of investment should be assessed case by case, as some cases require large, coordinated effort, while others may benefit from smaller-scale, local experimentation. Healthcare organizations need to strengthen internal capabilities to govern AI adoption. This includes data governance, technical expertise, and hybrid roles that can translate between technological and professional logics. In many settings, central units or shared hubs may provide these

capabilities, making them accessible to organizations that cannot develop them independently. Professional engagement remains essential for effective implementation. Involving clinicians in the design and adoption of AI tools enhances relevance and trust on adopted technologies. Yet, as innovation becomes more radical, there will be circumstances where top-down decisions are needed, shifting the focus from co-construction to the rationalization of why adoption is necessary. Balancing these dynamics across use cases will be critical. Finally, decisions regarding whether to build, buy, or ally for AI solutions should be supported by transparent and shared evaluation metrics that consistently take into account know-what and know-how. In public healthcare systems, where many organizations lack the scale to pursue independent strategies, system-level coordination becomes indispensable. Without such coordination, smaller actors risk being left behind or becoming dependent on external interests not fully aligned with organization or public goals.

REFERENCES

[1] Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. University of Chicago press.

[2] Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617.

[3] Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 107–136.

[4] AlKhamees, S. B., & Durugbo, C. M. (2024). Organisational ambidexterity and innovation: A systematic review and unified model of 'CODEC' management priorities. *Management Review Quarterly*. https://doi.org/10.1007/s11301-024-00474-5

[5] Allen, R., & Choudhury, P. (Raj). (2022). Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion. *Organization Science*, *33*(1), 149–169. https://doi.org/10.1287/orsc.2021.1554

[6] Alon-Barkat, S., & Busuioc, M. (2023). Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice. *Journal of Public Administration Research and Theory*, *33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

[7] Alvesson, M. (2004). *Knowledge work and knowledge-intensive firms*. Oxford University Press.

[8] Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, *36*(2), 247–271.

[9] Ångström, R. C., Björn, M., Dahlander, L., Mähring, M., & Wallin, M. W. (2023). Getting AI Implementation Right: Insights from a Global Survey. *California Management Review*, *66*(1), 5–22. https://doi.org/10.1177/00081256231190430

[10] Anthony, C. (2018). To Question or Accept? How Status Differences Influence Responses to New Epistemic Technologies in Knowledge Work. *Academy of Management Review*, *43*(4), 661–679. https://doi.org/10.5465/amr.2016.0334

[11] Anthony, C. (2021). When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies. *Administrative Science Quarterly*, *66*(4), 1173–1212. https://doi.org/10.1177/00018392211016755

[12] Anthony, C., Bechky, B. A., & Fayard, A.-L. (2023). "Collaborating" with AI: Taking a System View to Explore the Future of Work. *Organization Science*, *34*(5), 1672–1694. https://doi.org/10.1287/orsc.2022.1651

[13] Aresu, A. (2024). *Geopolitica dell'intelligenza artificiale* (1. ed. in Scintille). Feltrinelli.

[14] Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems. *Journal of the Association for Information Systems*, *22*(2), 325–352. https://doi.org/10.17705/1jais.00664

[15] Aujoulat, I., Jacquemin, P., Darras, E., Rietzschel, E., Wens, J., Hermans, M., Scheen, A., & Trefois, P. (2014). Factors associated with clinical inertia: An integrative review. *Advances in Medical Education and Practice*, 141. https://doi.org/10.2147/AMEP.S59022

[16] Baer, I., Waardenburg, L., & Huysman, M. (2025). What Is Augmented? A Metanarrative Review of AI-Based Augmentation. *Journal of the Association for Information Systems*, *26*(3), 760–798. https://doi.org/10.17705/1jais.00921

[17] Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*, *45*(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

[18] Balasubramanian, N., Ye, Y., & Xu, M. (2022). Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning. *Academy of Management Review*, *47*(3), 448–465. https://doi.org/10.5465/amr.2019.0470

[19] Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, *45*(2), 159–182. https://doi.org/10.1002/job.2735

[20] Bankuoru Egala, S., & Liang, D. (2023). Algorithm aversion to mobile clinical decision support among clinicians: A choice-based conjoint analysis. *European Journal of Information Systems*, 1–17. https://doi.org/10.1080/0960085X.2023.2251927

[21] Barbieri, C., Molina, M., Ponce, P., Tothova, M., Cattinelli, I., Ion Titapiccolo, J., Mari, F., Amato, C., Leipold, F., Wehmeyer, W., Stuard, S., Stopper, A., & Canaud, B. (2016). An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. *Kidney International*, *90*(2), 422–429. https://doi.org/10.1016/j.kint.2016.03.036

[22] Barker, K. M., Dunn, E. C., Richmond, T. K., Ahmed, S., Hawrilenko, M., & Evans, C. R. (2020). Cross-classified multilevel models (CCMM) in health research: A systematic review of published empirical studies and recommendations for best practices. *SSM - Population Health*, *12*, 100661. https://doi.org/10.1016/j.ssmph.2020.100661

[23] Barlow, J. (2017). *Managing innovation in healthcare*. World Scientific.

[24] Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1177/014920639101700108

[25] Battisti, G., & Stoneman, P. (2003). Inter- and intra-firm effects in the diffusion of new process technology. *Research Policy*, *32*(9), 1641–1655. https://doi.org/10.1016/s0048-7333(03)00055-6

[26] Beauchemin, M., Murray, M. T., Sung, L., Hershman, D. L., Weng, C., & Schnall, R. (2019). Clinical decision support for therapeutic decision-making in cancer: A systematic review. *International Journal of Medical Informatics*, *130*, 103940. https://doi.org/10.1016/j.ijmedinf.2019.07.019

[27] Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, *19*(4).

[28] Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). *Managing Artificial Intelligence*. 1433–1450.

[29] Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*(3). https://doi.org/10.1214/06-BA115

[30] Bhuyan, S. S., Sateesh, V., Mukul, N., Galvankar, A., Mahmood, A., Nauman, M., Rai, A., Bordoloi, K., Basu, U., & Samuel, J. (2025). Generative Artificial Intelligence Use in Healthcare: Opportunities for Clinical Excellence and Administrative Efficiency. *Journal of Medical Systems*, *49*(1), 10. https://doi.org/10.1007/s10916-024-02136-1

[31] Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

[32] Birkinshaw, J., & Gupta, K. (2013). Clarifying the Distinctive Contribution of Ambidexterity to the Field of Organization Studies. *Academy of Management Perspectives*, *27*(4), 287–298. https://doi.org/10.5465/amp.2012.0167

[33] Birkinshaw, J., Zimmermann, A., & Raisch, S. (2016). How Do Firms Adapt to Discontinuous Change? Bridging the Dynamic Capabilities and Ambidexterity Perspectives. *California Management Review*, *58*(4), 36–58. https://doi.org/10.1525/cmr.2016.58.4.36

[34] Bodwell, W., & Chermack, T. J. (2010). Organizational ambidexterity: Integrating deliberate and emergent strategy with scenario planning. *Technological Forecasting and Social Change*, *77*(2), 193–202. https://doi.org/10.1016/j.techfore.2009.07.004

[35] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). *On the Opportunities and Risks of Foundation Models* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2108.07258

[36] Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151.

[37] Bouaud Jacques, Spano Jean-Philippe, Lefranc Jean-Pierre, Cojean-Zelek Isabelle, Blaszka-Jaulerry Brigitte, Zelek Laurent, Durieux Axel, Tournigand Christophe, Rousseau Alexandra, Vandenbussche Pierre-Yves, & Séroussi Brigitte. (2015). Physicians' Attitudes Towards the Advice of a Guideline-Based Decision Support System: A Case Study With OncoDoc2 in the Management of Breast Cancer Patients. In *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/978-1-61499-564-7-264

[38] Boyacı, T., Canyakmaz, C., & De Véricourt, F. (2024). Human and Machine: The Impact of Machine Input on Decision Making Under Cognitive

Limitations. *Management Science*, *70*(2), 1258–1275. https://doi.org/10.1287/mnsc.2023.4744

[39] Brown, J. S., & Duguid, P. (1998). Organizing knowledge. *California Management Review*, *40*(3), 90–111.

[40] Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & company.

[41] Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

[42] Bucalo, M. L., Barbieri, C., Roca, S., Ion Titapiccolo, J., Ros Romero, M. S., Ramos, R., Albaladejo, M., Manzano, D., Mari, F., & Molina, M. (2018). The anaemia control model: Does it help nephrologists in therapeutic decision-making in the management of anaemia? *Nefrología (English Edition)*, *38*(5), 491–502. https://doi.org/10.1016/j.nefroe.2018.10.001

[43] Büchter, R. B., Weise, A., & Pieper, D. (2020). Development, testing and use of data extraction forms in systematic reviews: A review of methodological guidance. *BMC Medical Research Methodology*, *20*(1), 259. https://doi.org/10.1186/s12874-020-01143-3

[44] Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

[45] Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P.-A. C., & Rubin, H. R. (1999). Why Don't Physicians Follow Clinical Practice Guidelines?: A Framework for Improvement. *JAMA*, *282*(15), 1458. https://doi.org/10.1001/jama.282.15.1458

[46] Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197.

[47] Camilleri, A. R., & Sah, S. (2021). Amplification of the status quo bias among physicians making medical decisions. *Applied Cognitive Psychology*, *35*(6), 1374–1386. https://doi.org/10.1002/acp.3868

[48] Cao, Q., Gedajlovic, E., & Zhang, H. (2009). Unpacking Organizational Ambidexterity: Dimensions, Contingencies, and Synergistic Effects. *Organization Science*, *20*(4), 781–796. https://doi.org/10.1287/orsc.1090.0426

[49] Card, S. K. (2018). *The psychology of human-computer interaction*. Crc Press.

[50] Chakma, R., Paul, J., & Dhir, S. (2024). Organizational Ambidexterity: A Review and Research Agenda. *IEEE Transactions on Engineering Management*, *71*, 121–137. https://doi.org/10.1109/TEM.2021.3114609

[51] Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To Be or Not to Be …Human? Theorizing the Role of Human-Like Competencies in Conversational Artificial Intelligence Agents. *Journal of Management Information Systems*, *39*(4), 969–1005. https://doi.org/10.1080/07421222.2022.2127441

[52] Chari, V. V., & Hopenhayn, H. (1991). Vintage Human Capital, Growth, and the Diffusion of New Technology. *Journal of Political Economy*, *99*(6), 1142–1165. https://doi.org/10.1086/261795

[53] Charles, C., Gafni, A., & Whelan, T. (1999). Decision-making in the physician–patient encounter: Revisiting the shared treatment decision-making model. *Social Science & Medicine*, *49*(5), 651–661. https://doi.org/10.1016/S0277-9536(99)00145-8

[54] Chomutare, T., Tejedor, M., Svenning, T. O., Marco-Ruiz, L., Tayefi, M., Lind, K., Godtliebsen, F., Moen, A., Ismail, L., Makhlysheva, A., & Ngo, P. D. (2022). Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators. *International Journal of Environmental Research and Public Health*, *19*(23), 16359. https://doi.org/10.3390/ijerph192316359

[55] Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. (2023). Human-AI Ensembles: When Can They Work? *Journal of Management*, 01492063231194968. https://doi.org/10.1177/01492063231194968

[56] Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381–1411. https://doi.org/10.1002/smj.3152

[57] Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. In *Advances in Experimental Social Psychology*

(Vol. 24, pp. 201–234). Elsevier. https://doi.org/10.1016/S0065-2601(08)60330-5

[58] Compagni, A., Mele, V., & Ravasi, D. (2015). How Early Implementations Influence Later Adoptions of Innovation: Social Positioning and Skill Reproduction in the Diffusion of Robotic Surgery. *Academy of Management Journal*, *58*(1), 242–278. https://doi.org/10.5465/amj.2011.1184

[59] Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems*, *29*(4), 101600. https://doi.org/10.1016/j.jsis.2020.101600

[60] Dahlke, J., Beck, M., Kinne, J., Lenz, D., Dehghan, R., Wörter, M., & Ebersberger, B. (2024). Epidemic effects in the diffusion of emerging digital technologies: Evidence from artificial intelligence adoption. *Research Policy*, *53*(2), 104917. https://doi.org/10.1016/j.respol.2023.104917

[61] Davenport, T. H. (2018). *The AI advantage: How to put the artificial intelligence revolution to work*. mit Press.

[62] De Véricourt, F., & Gurkan, H. (2023). Is Your Machine Better Than You? You May Never Know. *Management Science*, mnsc.2023.4791. https://doi.org/10.1287/mnsc.2023.4791

[63] Denicolai, S., Ramirez, M., & Tidd, J. (2016). Overcoming the false dichotomy between internal R&D and external knowledge acquisition: Absorptive capacity dynamics over time. *Technological Forecasting and Social Change*, *104*, 57–65. https://doi.org/10.1016/j.techfore.2015.11.025

[64] Dennis, A. R., Lakhiwal, A., & Sachdeva, A. (2023). AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With. *Journal of Management Information Systems*, *40*(2), 307–337. https://doi.org/10.1080/07421222.2023.2196773

[65] Dickinson, H., & Yates, S. (2023). From external provision to technological outsourcing: Lessons for public sector automation from the outsourcing literature. *Public Management Review*, *25*(2), 243–261. https://doi.org/10.1080/14719037.2021.1972681

[66] Dietvorst, B. J., & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

[67] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015a). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

[68] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015b). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

[69] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly)

Modify Them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

[70] Djulbegovic, B., & Guyatt, G. H. (2017). Progress in evidence-based medicine: A quarter century on. *The Lancet*, *390*(10092), 415–423. https://doi.org/10.1016/S0140-6736(16)31592-6

[71] Dosi, G., Nelson, R. R., & Winter, S. G. (Eds.). (2000). *The nature and dynamics of organizational capabilities*. Oxford University Press.

[72] Duncan, R. B. (1976). The ambidextrous organization: Designing dual structures for innovation. *The Management of Organization*, *1*(1), 167–188.

[73] Duwe, G., Mercier, D., Wiesmann, C., Kauth, V., Moench, K., Junker, M., Neumann, C. C. M., Haferkamp, A., Dengel, A., & Höfner, T. (2024). Challenges and perspectives in use of artificial intelligence to support treatment recommendations in clinical oncology. *Cancer Medicine*, *13*(12), e7398. https://doi.org/10.1002/cam4.7398

[74] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

[75] Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H.,

Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[76] Eddy, D. M. (1990). Anatomy of a Decision. *JAMA: The Journal of the American Medical Association*, *263*(3), 441. https://doi.org/10.1001/jama.1990.03440030128037

[77] Einola, K., Khoreva, V., & Tienari, J. (2023). A colleague named Max: A critical inquiry into affects when an anthropomorphised AI (ro)bot enters the workplace. *Human Relations*, 00187267231206328. https://doi.org/10.1177/00187267231206328

[78] Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

[79] Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

[80] Engel, C., Elshan, E., Ebel, P., & Leimeister, J. M. (2024). Stairway to heaven or highway to hell: A model for assessing cognitive automation use cases. *Journal of Information Technology*, *39*(1), 94–122. https://doi.org/10.1177/02683962231185599

[81] Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *Npj Digital Medicine*, *4*(1), 5. https://doi.org/10.1038/s41746-020-00376-2

[82] Farzaneh, M., Wilden, R., Afshari, L., & Mehralian, G. (2022). Dynamic capabilities and innovation ambidexterity: The roles of intellectual capital and innovation orientation. *Journal of Business Research*, *148*, 47–59. https://doi.org/10.1016/j.jbusres.2022.04.030

[83] Faulconbridge, J., Sarwar, A., & Spring, M. (2023). How Professionals Adapt to Artificial Intelligence: The Role of Intertwined Boundary Work. *Journal of Management Studies*, joms.12936. https://doi.org/10.1111/joms.12936

[84] Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: Perils, promises and solutions. A systematic review. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0177-4

[85] FDA. (2025, July 10). *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. Food & Drug Administration. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices

[86] Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*(1), 94–118.

[87] Fernández-Pérez De La Lastra, S., García-Carbonell, N., Martín-Alcázar, F., & Sánchez-Gardey, G. (2017). Intellectual capital role in ambidexterity

emergence: A proposal of a multilevel model and research agenda. *Journal of Intellectual Capital*, *18*(4), 733–744. https://doi.org/10.1108/JIC-10-2016-0100

[88] Ferratt, T. W., Prasad, J., & Dunne, E. J. (2018). Fast and slow processes underlying theories of information technology use. *Journal of the Association for Information Systems*, *19*(1), 3.

[89] Fielding, A., & Goldstein, H. (2006). *Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review* (No. Research Report RR791). Department for Education and Skills.

[90] Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2023). The extent of algorithm aversion in decision-making situations with varying gravity. *PLOS ONE*, *18*(2), e0278751. https://doi.org/10.1371/journal.pone.0278751

[91] Finn, T., & Downie, A. (2025, February 11). *Agentic AI vs. Generative AI*. IBM. https://www.ibm.com/think/topics/agentic-ai-vs-generative-ai

[92] Fisch, C., & Block, J. (2018). Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly*, *68*(2), 103–106. https://doi.org/10.1007/s11301-018-0142-x

[93] Flick, U. (2022). *An introduction to qualitative research* (Seventh edition. Core new edition). SAGE Publications Ltd.

[94] Floridi, L. (2011). Children of the fourth revolution. *Philosophy & Technology*, *24*(3), 227–232.

[95] Floridi, L., & Taddeo, M. (2016). What is data ethics? In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Vol. 374, Issue 2083, p. 20160360). The Royal Society.

[96] Floridi, L. (2025). AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. *Philosophy & Technology*, *38*(1), 30, s13347-025-00858–00859. https://doi.org/10.1007/s13347-025-00858-9

[97] Fox, J., Patkar, V., Chronakis, I., & Begent, R. (2009). From practice guidelines to clinical decision support: Closing the loop. *Journal of the Royal Society of Medicine*, *102*(11), 464–473. https://doi.org/10.1258/jrsm.2009.090010

[98] Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. *MIS Quarterly*, *45*(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

[99] Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, *33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

[100] Gama, F., & Magistretti, S. (2025). Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. *Journal of Product Innovation Management*, *42*(1), 76–111. https://doi.org/10.1111/jpim.12698

[101] Garbelli, M., Baro Salvador, M. E., Rincon Bello, A., Samaniego Toro, D., Bellocchio, F., Fumagalli, L., Chermisi, M., Apel, C., Petrovic, J., Kendzia, D., Ion Titapiccolo, J., Yeung, J., Barbieri, C., Mari, F., Usvyat, L., Larkin, J., Stuard, S., & Neri, L. (2024). Usage of the Anemia Control Model Is Associated with Reduced Hospitalization Risk in Hemodialysis. *Biomedicines*, *12*(10), 2219. https://doi.org/10.3390/biomedicines12102219

[102] Gefen, Karahanna, & Straub. (2003). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, *27*(1), 51. https://doi.org/10.2307/30036519

[103] Gibson, C. B., & Birkinshaw, J. (2004). The antecedents, consequences, and mediating role of organizational ambidexterity. *Academy of Management Journal*, *47*(2), 209–226. https://doi.org/10.2307/20159573

[104] Giest, S. N., & Klievink, B. (2024). More than a digital system: How AI is changing the role of bureaucrats in different organizational contexts. *Public Management Review*, *26*(2), 379–398. https://doi.org/10.1080/14719037.2022.2095001

[105] Gittell, J. H., & Douglass, A. (2012). Relational Bureaucracy: Structuring Reciprocal Relationships into Roles. *Academy of Management Review*, *37*(4), 709–733. https://doi.org/10.5465/amr.2010.0438

[106] Giugliano, D., & Esposito, K. (2011). Clinical Inertia as a Clinical Safeguard. *JAMA*, *305*(15), 1591. https://doi.org/10.1001/jama.2011.490

[107] Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

[108] Goldstein, H. (1994). Multilevel Cross-Classified Models. *Sociological Methods & Research*, *22*(3), 364–375. https://doi.org/10.1177/0049124194022003005

[109] Goldstein, H. (2010). *Multilevel Statistical Models* (1st ed.). Wiley. https://doi.org/10.1002/9780470973394

[110] Goto, M. (2021). Collective professional role identity in the age of artificial intelligence. *Journal of Professions and Organization*, *8*(1), 86–107. https://doi.org/10.1093/jpo/joab003

[111] Goto, M. (2023). Anticipatory innovation of professional services: The case of auditing and artificial intelligence. *Research Policy*, *52*(8), 104828. https://doi.org/10.1016/j.respol.2023.104828

[112] Grashof, N., & Kopka, A. (2023). Artificial intelligence and radical innovation: An opportunity for all companies? *Small Business Economics*, *61*(2), 771–797. https://doi.org/10.1007/s11187-022-00698-3

[113] Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130.

[114] Greenhalgh, T., Howick, J., Maskrey, N., & for the Evidence Based Medicine Renaissance Group. (2014). Evidence based medicine: A movement in crisis? *BMJ*, *348*(jun13 4), g3725–g3725. https://doi.org/10.1136/bmj.g3725

[115] Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The Role of Artificial Intelligence and Data Network Effects for Creating User

Value. *Academy of Management Review*, *46*(3), 534–551. https://doi.org/10.5465/amr.2019.0178

[116] Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, *29*(2), 101614. https://doi.org/10.1016/j.jsis.2020.101614

[117] Guo, G. (2017). Demystifying variance in performance: A longitudinal multilevel perspective. *Strategic Management Journal*, *38*(6), 1327–1342. https://doi.org/10.1002/smj.2555

[118] Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., & deBeer, H. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026

[119] Haefner, N., Parida, V., Gassmann, O., & Wincent, J. (2023). Implementing and scaling artificial intelligence: A review, framework, and research agenda. *Technological Forecasting and Social Change*, *197*, 122878. https://doi.org/10.1016/j.techfore.2023.122878

[120] Hambrick, D. C. (2007). Upper Echelons Theory: An Update. *Academy of Management Review*, *32*(2), 334–343. https://doi.org/10.5465/amr.2007.24345254

[121] Hambrick, D. C., Finkelstein, S., & Mooney, A. C. (2005). Executive Job Demands: New Insights for Explaining Strategic Decisions and Leader

Behaviors. *Academy of Management Review*, *30*(3), 472–491. https://doi.org/10.5465/amr.2005.17293355

[122] Han, P. K. J., Klein, W. M. P., & Arora, N. K. (2011). Varieties of Uncertainty in Health Care: A Conceptual Taxonomy. *Medical Decision Making*, *31*(6), 828–838. https://doi.org/10.1177/0272989X10393976

[123] Harrison, C., Canadian Paediatric Society (CPS), & Bioethics Committee. (2004). Treatment decisions regarding infants, children and adolescents. *Paediatrics & Child Health*, *9*(2), 99–103. https://doi.org/10.1093/pch/9.2.99

[124] Hayat, H., Kudrautsau, M., Makarov, E., Melnichenko, V., Tsykunou, T., Varaksin, P., Pavelle, M., & Oskowitz, A. Z. (2025). *Toward the Autonomous AI Doctor: Quantitative Benchmarking of an Autonomous Agentic AI Versus Board-Certified Clinicians in a Real World Setting* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2507.22902

[125] Heimstädt, M., Koljonen, T., & Elmholdt, K. T. (2024). Expertise in Management Research: A Review and Agenda for Future Research. *Academy of Management Annals*, *18*(1), 121–156. https://doi.org/10.5465/annals.2022.0078

[126] Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems*, *32*(3), 101772. https://doi.org/10.1016/j.jsis.2023.101772

[127] Hiebl, M. R. W. (2023). Sample Selection in Systematic Literature Reviews of Management Research. *Organizational Research Methods*, *26*(2), 229–261. https://doi.org/10.1177/1094428120986851

[128] Hoffreumon, C., Forman, C., & Van Zeebroeck, N. (2024). Make or buy your artificial intelligence? Complementarities in technology sourcing. *Journal of Economics & Management Strategy*, *33*(2), 452–479. https://doi.org/10.1111/jems.12586

[129] Holmqvist, M. (2004). Experiential Learning Processes of Exploitation and Exploration Within and Between Organizations: An Empirical Study of Product Development. *Organization Science*, *15*(1), 70–81. https://doi.org/10.1287/orsc.1030.0056

[130] Hopf, K., Müller, O., Shollo, A., & Thiess, T. (2023). Organizational Implementation of AI: Craft and Mechanical Work. *California Management Review*, *66*(1), 23–47. https://doi.org/10.1177/00081256231197445

[131] Hox, J. J., & Roberts, J. K. (2011). *Handbook of advanced multilevel analysis*. Routledge.

[132] Horowitz, M. C., Kahn, L., Macdonald, J., & Schneider, J. (2024). Adopting AI: How familiarity breeds both trust and contempt. *AI & SOCIETY*, *39*(4), 1721–1735. https://doi.org/10.1007/s00146-023-01666-5

[133] Hughes, T. P. (1987). The evolution of large technological systems. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, *82*, 51–82.

[134] Hwang, B.-N., Lai, Y.-P., & Wang, C. (2023). Open innovation and organizational ambidexterity. *European Journal of Innovation Management*, *26*(3), 862–884. https://doi.org/10.1108/EJIM-06-2021-0303

[135] IBM. (2025). *Agentic AI vs. Generative AI*. https://www.ibm.com/think/topics/agentic-ai-vs-generative-ai

[136] Igna, I., & Venturini, F. (2023). The determinants of AI innovation across European firms. *Research Policy*, *52*(2), 104661. https://doi.org/10.1016/j.respol.2022.104661

[137] Jacobides, M. G., Brusoni, S., & Candelon, F. (2021). The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science*, *6*(4), 412–435.

[138] Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and How Artificial Intelligence Augments Employee Creativity. *Academy of Management Journal*, *67*(1), 5–32. https://doi.org/10.5465/amj.2022.0426

[139] Johnson, S., & Acemoglu, D. (2023). *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. Hachette UK.

[140] Junni, P., Sarala, R. M., Taras, V., & Tarba, S. Y. (2013). Organizational Ambidexterity and Performance: A Meta-Analysis. *Academy of Management Perspectives*, *27*(4), 299–312. https://doi.org/10.5465/amp.2012.0015

[141] Jurksiene, L., & Pundziene, A. (2016). The relationship between dynamic capabilities and firm competitive advantage: The mediating role of

organizational ambidexterity. *European Business Review*, *28*(4), 431–448. https://doi.org/10.1108/EBR-09-2015-0088

[142] Jussupow, E., Benbasat, I., & Heinzl, A. (2020). *Why are we averse towards Algorithms? A comprehensive literature review on algorithm aversion*. 28th European Conference on Information Systems (ECIS).

[143] Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An Integrative Perspective on Algorithm Aversion and Appreciation in Decision-Making. *MIS Quarterly*. https://doi.org/10.25300/MISQ/2024/18512

[144] Jussupow, E., Spohrer, K., & Heinzl, A. (2022). Identity Threats as a Reason for Resistance to Artificial Intelligence: Survey Study With Medical Students and Professionals. *JMIR Formative Research*, *6*(3), e28750. https://doi.org/10.2196/28750

[145] Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

[146] Kahneman, D. (2013). *Thinking, fast and slow* (1st pbk. ed). Farrar, Straus and Giroux.

[147] Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136

[148] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

[149] Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19.

[150] Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

[151] Karshenas, M., & Stoneman, P. L. (1993). Rank, stock, order, and epidemic effects in the diffusion of new process technologies: An empirical model. *The RAND Journal of Economics*, 503–528.

[152] Kassotaki, O. (2022). Review of Organizational Ambidexterity Research. *Sage Open*, *12*(1), 21582440221082127. https://doi.org/10.1177/21582440221082127

[153] Kawaguchi, K. (2021). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business. *Management Science*, *67*(3), 1670–1695. https://doi.org/10.1287/mnsc.2020.3599

[154] Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

[155] Kemp, A. (2023). Competitive Advantage Through Artificial Intelligence: Toward a Theory of Situated AI. *Academy of Management Review*, amr.2020.0205. https://doi.org/10.5465/amr.2020.0205

[156] Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological Review*, *90*(3), 185.

[157] Kock, N. (2009). Information systems theorizing based on evolutionary psychology: An interdisciplinary review and theory integration framework. *Mis Quarterly*, 395–418.

[158] Komiak & Benbasat. (2006). The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly*, *30*(4), 941. https://doi.org/10.2307/25148760

[159] Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

[160] Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: An update 2022. *Kidney International Supplements*, *12*(1), 7–11. https://doi.org/10.1016/j.kisu.2021.11.003

[161] Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, *44*(6), 1425–1452. https://doi.org/10.1002/smj.3387

[162] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

[163] Langley, A., Lindberg, K., Mørk, B. E., Nicolini, D., Raviola, E., & Walter, L. (2019). Boundary Work among Groups, Occupations, and Organizations: From Cartography to Process. *Academy of Management Annals*, *13*(2), 704–736. https://doi.org/10.5465/annals.2017.0089

[164] Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

[165] Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022a). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. *Organization Science*, *33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

[166] Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022b). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. *Organization Science*, *33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

[167] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

[168] Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How are They Related? Do We Need Them? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2129878

[169] Levina & Vaast. (2008). Innovating or doing as Told? Status Differences and Overlapping Boundaries in Offshore Collaboration. *MIS Quarterly*, *32*(2), 307. https://doi.org/10.2307/25148842

[170] Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology*, *14*(1), 319–338.

[171] Leyland, A. H., & Groenewegen, P. P. (2020). *Multilevel Modelling for Public Health and Health Services Research: Health in Context*. Springer International Publishing. https://doi.org/10.1007/978-3-030-34801-4

[172] Li, J., Li, M., Wang, X., & Bennett Thatcher, J. (2021). Strategic Directions for AI: The Role of CIOs and Boards of Directors. *MIS Quarterly*, *45*(3), 1603–1644. https://doi.org/10.25300/MISQ/2021/16523

[173] Li, X., Li, X., & Ding, S. (2025). Digital transformation and innovation ambidexterity: Perspectives on accumulation and resilience effects. *European Journal of Innovation Management*, *28*(4), 1601–1624. https://doi.org/10.1108/EJIM-07-2023-0526

[174] Lin, Y.-K., & Maruping, L. M. (2025). Organizing for AI Innovation: Insights From an Empirical Exploration of U.S. Patents. *MIS Quarterly*, *49*(3), 1095–1122. https://doi.org/10.25300/MISQ/2025/18765

[175] Liu, M., Tang, X., Xia, S., Zhang, S., Zhu, Y., & Meng, Q. (2023). Algorithm Aversion: Evidence from Ridesharing Drivers. *Management Science*, mnsc.2022.02475. https://doi.org/10.1287/mnsc.2022.02475

[176] Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

[177] Lou, B., & Wu, L. (2021). AI on Drugs: Can Artificial Intelligence Accelerate Drug Development? Evidence from a Large-Scale Examination of Bio-Pharma Firms. *MIS Quarterly*, *45*(3), 1451–1482. https://doi.org/10.25300/MISQ/2021/16565

[178] Maclean, M., Harvey, C., Golant, B. D., & Sillince, J. A. (2021). The role of innovation narratives in accomplishing organizational ambidexterity. *Strategic Organization*, *19*(4), 693–721. https://doi.org/10.1177/1476127019897234

[179] Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

[180] Mahmud, H., Islam, A. K. M. N., Luo, X. (Robert), & Mikalef, P. (2024). Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance. *Decision Support Systems*, *179*, 114168. https://doi.org/10.1016/j.dss.2024.114168

[181] Man Tang, P., Koopman, J., McClean, S. T., Zhang, J. H., Li, C. H., De Cremer, D., Lu, Y., & Ng, C. T. S. (2022). When Conscientious Employees Meet Intelligent Machines: An Integrative Approach Inspired by Complementarity Theory and Role Theory. *Academy of Management Journal*, *65*(3), 1019–1054. https://doi.org/10.5465/amj.2020.1516

[182] Maragno, G., Tangi, L., Gastaldi, L., & Benedetti, M. (2023). AI as an organizational agent to nurture: Effectively introducing chatbots in public entities. *Public Management Review*, *25*(11), 2135–2165. https://doi.org/10.1080/14719037.2022.2063935

[183] March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

[184] Markus, M. L., & Silver, M. (2008). A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit. *Journal of the Association for Information Systems*, *9*(10), 609–632. https://doi.org/10.17705/1jais.00176

[185] Matthews, M. J., Su, R., Yonish, L., McClean, S., Koopman, J., & Yam, K. C. (2025). A Review of Artificial Intelligence, Algorithms, and Robots Through the Lens of Stakeholder Theory. *Journal of Management*, *51*(6), 2627–2676. https://doi.org/10.1177/01492063241311855

[186] McAfee, A. (2024). *The Economic Impact of Generative AI*.

[187] McDuff, D., Schaekermann, M., Tu, T., Palepu, A., Wang, A., Garrison, J., Singhal, K., Sharma, Y., Azizi, S., & Kulkarni, K. (2025). Towards accurate differential diagnosis with large language models. *Nature*, 1–7.

[188] McMurray, J., Parfrey, P., Adamson, J. W., Aljama, P., Berns, J. S., Bohlius, J., Drüeke, T. B., Finkelstein, F. O., Fishbane, S., & Ganz, T. (2012). Kidney disease: Improving global outcomes (KDIGO) anemia work group. KDIGO clinical practice guideline for anemia in chronic kidney disease. *Kidney International Supplements*, 279–335.

[189] Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. *Public Administration Review*, *81*(5), 837–846. https://doi.org/10.1111/puar.13391

[190] Mom, T. J. M., Chang, Y.-Y., Cholakova, M., & Jansen, J. J. P. (2019). A Multilevel Integrated Framework of Firm HR Practices, Individual Ambidexterity, and Organizational Ambidexterity. *Journal of Management*, *45*(7), 3009–3034. https://doi.org/10.1177/0149206318776775

[191] Montealegre-López, N. (2025). Exploring the role of trust in AI-driven decision-making: A systematic literature review. *Management Review Quarterly*. https://doi.org/10.1007/s11301-025-00526-4

[192] Montgomery, K. (2006). *How doctors think* (Vol. 8). Oxford University Press New York.

[193] Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation Bias: Decision Making and Performance in High-Tech Cockpits. *The International Journal of Aviation Psychology*, *8*(1), 47–63. https://doi.org/10.1207/s15327108ijap0801_3

[194] Mudambi, S. M., & Tallman, S. (2010). Make, Buy or Ally? Theoretical Perspectives on Knowledge Process Outsourcing through Alliances. *Journal of Management Studies*, *47*(8), 1434–1456. https://doi.org/10.1111/j.1467-6486.2010.00944.x

[195] Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and Technology: Forms of Conjoined Agency in Organizations. *Academy of Management Review*, *46*(3), 552–571. https://doi.org/10.5465/amr.2019.0186

[196] Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management. *MIS Quarterly*, *41*(1), 223–238. JSTOR.

[197] Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review*, *26*(1), 114–141. https://doi.org/10.1080/14719037.2022.2048685

[198] Nicolaus, S., Crelier, B., Donzé, J. D., & Aubert, C. E. (2022). Definition of patient complexity in adults: A narrative review. *Journal of Multimorbidity and Comorbidity*, *12*, 263355652210812. https://doi.org/10.1177/26335565221081288

[199] Nicolini, D., Mørk, B. E., Masovic, J., & Hanseth, O. (2017). *Expertise as Trans-Situated* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198806639.003.0002

[200] Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*(1), 53. https://doi.org/10.1186/s13012-015-0242-0

[201] Nosella, A., Cantarello, S., & Filippini, R. (2012). The intellectual structure of organizational ambidexterity: A bibliographic investigation into the state of the art. *Strategic Organization*, *10*(4), 450–465. https://doi.org/10.1177/1476127012457979

[202] Okhuysen, G. A., & Bechky, B. A. (2009). Coordination in Organizations: An Integrative Perspective. *Academy of Management Annals*, *3*(1), 463–502. https://doi.org/10.5465/19416520903047533

[203] O'Reilly, C. A., & Tushman, M. L. (2013). Organizational Ambidexterity: Past, Present, and Future. *Academy of Management Perspectives*, *27*(4), 324–338. https://doi.org/10.5465/amp.2013.0025

[204] Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing. *Organization Science*, *32*(1), 18–41. https://doi.org/10.1287/orsc.2020.1377

[205] Pakarinen, P., & Huising, R. (2023). Relational Expertise: What Machines Can't Know. *Journal of Management Studies*, joms.12915. https://doi.org/10.1111/joms.12915

[206] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253.

[207] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*(3), 286–297.

[208] Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

[209] Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3 ed). Sage Publications.

[210] Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (2013). Implementation research: What it is and how to do it. *British Medical Journal*, *347*. https://doi.org/10.1136/bmj.f6753

[211] Petersson, L., Larsson, I., Nygren, J. M., Nilsen, P., Neher, M., Reed, J. E., Tyskbo, D., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research*, *22*(1), 850. https://doi.org/10.1186/s12913-022-08215-8

[212] Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell publ.

[213] Phillips, L. S., Branch, W. T., Cook, C. B., Doyle, J. P., El-Kebbi, I. M., Gallina, D. L., Miller, C. D., Ziemer, D. C., & Barnes, C. S. (2001). Clinical Inertia. *Annals of Internal Medicine*, *135*(9), 825–834. https://doi.org/10.7326/0003-4819-135-9-200111060-00012

[214] Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464

[215] Preti, L. M., Ardito, V., Compagni, A., Petracca, F., & Cappellaro, G. (2024). Implementation of Machine Learning Applications in Health Care Organizations: Systematic Review of Empirical Studies. *Journal of Medical Internet Research*, *26*, e55897. https://doi.org/10.2196/55897

[216] Pumplun, L., Tauchert, C., & Heidt, M. (2019). *A new organizational chassis for artificial intelligence-exploring organizational readiness factors*.

[217] Puranam, P. (2018). *The microstructure of organizations*. Oxford University Press.

[218] Puranam, P. (2021). Human–AI collaborative decision-making as an organization design problem. *Journal of Organization Design*, *10*(2), 75–80. https://doi.org/10.1007/s41469-021-00095-2

[219] Qin, S. (Marco), Jia, N., Luo, X., Liao, C., & Huang, Z. (2023). Perceived Fairness of Human Managers Compared with Artificial Intelligence in Employee Performance Evaluation. *Journal of Management Information Systems*, *40*(4), 1039–1070. https://doi.org/10.1080/07421222.2023.2267316

[220] Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *Mis Quarterly*, *43*(1), iii–ix.

[221] Raisch, S., & Birkinshaw, J. (2008). Organizational Ambidexterity: Antecedents, Outcomes, and Moderators. *Journal of Management*, *34*(3), 375–409. https://doi.org/10.1177/0149206308316058

[222] Raisch, S., & Fomina, K. (2024). Combining Human and Artificial Intelligence: Hybrid Problem-Solving in Organizations. *Academy of Management Review*, amr.2021.0421. https://doi.org/10.5465/amr.2021.0421

[223] Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

[224] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

[225] Ramaul, L., Ritala, P., Kostis, A., & Aaltonen, P. (2025). Rethinking How We Theorize AI in Organization and Management: A Problematizing Review of Rationality and Anthropomorphism. *Journal of Management Studies*, joms.13246. https://doi.org/10.1111/joms.13246

[226] Rana, N. P., Chatterjee, S., Dwivedi, Y. K., & Akter, S. (2022). Understanding dark side of artificial intelligence (AI) integrated business analytics: Assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems*, *31*(3), 364–387. https://doi.org/10.1080/0960085X.2021.1955628

[227] Revilla, E., Saenz, M. J., Seifert, M., & Ma, Y. (2023). Human–Artificial Intelligence Collaboration in Prediction: A Field Experiment in the Retail Industry. *Journal of Management Information Systems*, *40*(4), 1071–1098. https://doi.org/10.1080/07421222.2023.2267317

[228] Robbins, S. (2020). AI and the path to envelopment: Knowledge as a first step towards the responsible regulation and use of AI-powered machines. *Ai & Society*, *35*(2), 391–400.

[229] Robey, D., Boudreau, M.-C., & Rose, G. M. (2000). Information technology and organizational learning: A review and assessment of research. *Accounting, Management and Information Technologies*, *10*(2), 125–155.

[230] Rosenbacke, R., Melhus, Å., McKee, M., & Stuckler, D. (2024). How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. *JMIR AI*, *3*, e53207. https://doi.org/10.2196/53207

[231] Rothaermel, F. T., & Alexandre, M. T. (2009). Ambidexterity in Technology Sourcing: The Moderating Role of Absorptive Capacity. *Organization Science*, *20*(4), 759–780. https://doi.org/10.1287/orsc.1080.0404

[232] Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

[233] Safford, M. M., Allison, J. J., & Kiefe, C. I. (2007). Patient Complexity: More Than Comorbidity. The Vector Model of Complexity. *Journal of General Internal Medicine*, *22*(S3), 382–390. https://doi.org/10.1007/s11606-007-0307-0

[234] Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-Based Intuition and Decision Making in Organizations. *Journal of Management*, *36*(4), 941–973. https://doi.org/10.1177/0149206309350084

[235] Saleh, R. H., Durugbo, C. M., & Almahamid, S. M. (2023). What makes innovation ambidexterity manageable: A systematic review, multi-level model and future challenges. *Review of Managerial Science*, *17*(8), 3013–3056. https://doi.org/10.1007/s11846-023-00659-4

[236] Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

[237] Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, *43*(3), 695–719. https://doi.org/10.25300/misq/2019/13747

[238] Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2023). The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution. *Journal of the Association for Information Systems*, *24*(4), 980–1004. https://doi.org/10.17705/1jais.00827

[239] Selten, F., & Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*, *41*(1), 101885. https://doi.org/10.1016/j.giq.2023.101885

[240] Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review*, *83*(2), 263–278. https://doi.org/10.1111/puar.13602

[241] Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R., & Medow, M. A. (2013). Why Do Patients Derogate Physicians Who Use a Computer-Based Diagnostic Support System? *Medical Decision Making*, *33*(1), 108–118. https://doi.org/10.1177/0272989X12453501

[242] Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. *The Journal of Strategic Information Systems*, *31*(3), 101734. https://doi.org/10.1016/j.jsis.2022.101734

[243] Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, *61*(4), 66–83. https://doi.org/10.1177/0008125619862257

[244] Simon, H. A. (1991). Bounded Rationality and Organizational Learning. *Organization Science*, *2*(1), 125–134. https://doi.org/10.1287/orsc.2.1.125

[245] Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

[246] Simsek, Z. (2009). Organizational Ambidexterity: Towards a Multilevel Understanding. *Journal of Management Studies*, *46*(4), 597–624. https://doi.org/10.1111/j.1467-6486.2009.00828.x

[247] Singer, S. J., Kellogg, K. C., Galper, A. B., & Viola, D. (2022). Enhancing the value to users of machine learning-based clinical decision support tools: A framework for iterative, collaborative development and implementation.

Health Care Management Review, *47*(2), E21–E31. https://doi.org/10.1097/HMR.0000000000000324

[248] Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., Hou, L., Clark, K., Pfohl, S. R., & Cole-Lewis, H. (2025). Toward expert-level medical question answering with large language models. *Nature Medicine*, *31*(3), 943–950.

[249] Sivaraman, V., Bukowski, L. A., Levin, J., Kahn, J. M., & Perer, A. (2023). Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3544548.3581075

[250] Sluss, D. M., Van Dick, R., & Thompson, B. S. (2011). *Role theory in organizations: A relational perspective.*

[251] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

[252] Sommet, N., & Morselli, D. (2017). Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-Step Procedure Using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*(1), 203–218. https://doi.org/10.5334/irsp.90

[253] Soto-Acosta, P., Popa, S., & Martinez-Conesa, I. (2018). Information technology, knowledge management and environmental dynamism as drivers

of innovation ambidexterity: A study in SMEs. *Journal of Knowledge Management*, *22*(4), 824–849. https://doi.org/10.1108/JKM-10-2017-0448

[254] Stauffer, M. E., & Fan, T. (2014). Prevalence of anemia in chronic kidney disease in the United States. *PloS One*, *9*(1), e84943.

[255] Stelzl, K., Röglinger, M., & Wyrtki, K. (2020). Building an ambidextrous organization: A maturity model for organizational ambidexterity. *Business Research*, *13*(3), 1203–1230. https://doi.org/10.1007/s40685-020-00117-x

[256] Strich, F., University of Bayreuth, Germany, Mayer, A.-S., University of Passau, Germany, Fiedler, M., & University of Passau, Germany. (2021). What Do I Do in a World of Artificial Intelligence? Investigating the Impact of Substitutive Decision-Making AI Systems on Employees' Professional Role Identity. *Journal of the Association for Information Systems*, *22*(2), 304–324. https://doi.org/10.17705/1jais.00663

[257] Stryker, C. (2025, February 11). *Agentic AI*. IBM. https://www.ibm.com/think/topics/agentic-ai#7281537

[258] Sturm, T., Gerlacha, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organization Learning. *MIS Quarterly*, *45*(3), 1581–1602. https://doi.org/10.25300/MISQ/2021/16543

[259] Sturm, T., Pumplun, L., Gerlach, J. P., Kowalczyk, M., & Buxmann, P. (2023). Machine learning advice in managerial decision-making: The overlooked role of decision makers' advice utilization. *The Journal of Strategic Information Systems*, *32*(4), 101790. https://doi.org/10.1016/j.jsis.2023.101790

[260] Surowiecki, J. (2005). *The Wisdom of Crowds*. Knopf Doubleday Publishing Group.

[261] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *Npj Digital Medicine*, *3*(1), 17. https://doi.org/10.1038/s41746-020-0221-y

[262] Takita, H., Kabata, D., Walston, S. L., Tatekawa, H., Saito, K., Tsujimoto, Y., Miki, Y., & Ueda, D. (2025). A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *Npj Digital Medicine*, *8*(1), 175. https://doi.org/10.1038/s41746-025-01543-z

[263] Teodorescu, M., Morse, L., Awwad, Y., Center for Complex Systems, King Abdulaziz City for Science & Technology, & Kane, G. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation. *MIS Quarterly*, *45*(3), 1483–1500. https://doi.org/10.25300/MISQ/2021/16535

[264] Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, *42*(9), 1600–1631. https://doi.org/10.1002/smj.3322

[265] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[266] Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). The processes of technological innovation. *(No Title)*.

[267] Tsai, W., & Ghoshal, S. (1998). Social capital and value creation: The role of intrafirm networks. *Academy of Management Journal*, *41*(4), 464–476.

[268] Turel, O., & Kalhan, S. (2023). Prejudiced against the Machine? Implicit Associations and the Transience of Algorithm Aversion. *MIS Quarterly*, *47*(4), 1369–1394. https://doi.org/10.25300/MISQ/2022/17961

[269] Turner, N., Swart, J., & Maylor, H. (2013). Mechanisms for Managing Ambidexterity: A Review and Research Agenda. *International Journal of Management Reviews*, *15*(3), 317–332. https://doi.org/10.1111/j.1468-2370.2012.00343.x

[270] Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. JSTOR.

[271] Utz, S., Wolfers, L., & Göritz, A. (2021). The Effects of Situational and Individual Factors on Algorithm Acceptance in COVID-19-Related Decision-Making: A Preregistered Online Experiment. *Human-Machine Communication*, *3*, 27–46. https://doi.org/10.30658/hmc.3.3

[272] Van Den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

[273] Vanneste, B. S., & Puranam, P. (2024). Artificial Intelligence, Trust, and Perceptions of Agency. *Academy of Management Review*, amr.2022.0041. https://doi.org/10.5465/amr.2022.0041

[274] Vannuccini, S., & Prytkova, E. (2023). Artificial Intelligence's new clothes? A system technology perspective. *Journal of Information Technology*, 02683962231197824. https://doi.org/10.1177/02683962231197824

[275] Vassilakopoulou, P., Haug, A., Salvesen, L. M., & Pappas, I. O. (2023). Developing human/AI interactions for chat-based customer services: Lessons learned from the Norwegian government. *European Journal of Information Systems*, *32*(1), 10–22. https://doi.org/10.1080/0960085X.2022.2096490

[276] Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities. *Public Administration Review*, *80*(6), 946–961. https://doi.org/10.1111/puar.13286

[277] Von Nordenflycht, A. (2010). What Is a Professional Service Firm? Toward a Theory and Taxonomy of Knowledge-Intensive Firms. *Academy of Management Review*, *35*(1), 155–174. https://doi.org/10.5465/amr.35.1.zok155

[278] Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms. *Organization Science*, *33*(1), 59–82. https://doi.org/10.1287/orsc.2021.1544

[279] Wachter, R. M., & Brynjolfsson, E. (2024). Will Generative Artificial Intelligence Deliver on Its Promise in Health Care? *JAMA*, *331*(1), 65. https://doi.org/10.1001/jama.2023.25054

[280] Wang, G., Xie, S., & Li, X. (2024). Artificial intelligence, types of decisions, and street-level bureaucrats: Evidence from a survey experiment. *Public Management Review*, *26*(1), 162–184. https://doi.org/10.1080/14719037.2022.2070243

[281] Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

[282] Wang, Q., Huang, Y., Jasin, S., & Singh, P. V. (2023). Algorithmic Transparency with Strategic Users. *Management Science*, *69*(4), 2297–2317. https://doi.org/10.1287/mnsc.2022.4475

[283] Wang, W., Gao, G. (Gordon), & Agarwal, R. (2023). Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience. *Management Science*, mnsc.2021.00588. https://doi.org/10.1287/mnsc.2021.00588

[284] Weber, C. E., Kortkamp, C., Maurer, I., & Hummers, E. (2022). Boundary Work in Response to Professionals' Contextual Constraints: Micro-strategies in Interprofessional Collaboration. *Organization Studies*, *43*(9), 1453–1477. https://doi.org/10.1177/01708406221074135

[285] Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, xiii–xxiii.

[286] Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, *24*(10), 991–995. https://doi.org/10.1002/smj.318

[287] Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: Potential benefits, limitations, and harms of clinical guidelines. *BMJ*, *318*(7182), 527–530. https://doi.org/10.1136/bmj.318.7182.527

[288] Yin, R. K. (2014). *Case study research: Design and methods* (5. edition). SAGE.

[289] You, S., Yang, C. L., & Li, X. (2022). Algorithmic versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation? *Journal of Management Information Systems*, *39*(2), 336–365. https://doi.org/10.1080/07421222.2022.2063553

[290] Zanzotto, F. M. (2019). Viewpoint: Human-in-the-loop Artificial Intelligence. *Journal of Artificial Intelligence Research*, *64*, 243–252. https://doi.org/10.1613/jair.1.11345

FUNDING

The empirical part of this thesis was conducted as part of the MUSA (Multilayered Urban Sustainability Action) project, which is funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D "innovation ecosystems," set up of "territorial leaders in R&D.".

RINGRAZIAMENTI

Alla fine di questo lungo, intenso, strano percorso, sento la necessità di dedicare una piccola ma significativa parte di questa tesi alle persone che hanno contribuito con la loro presenza a rendere il sentiero meno impervio.

Il primo e più importante ringraziamento va a Cristina, prima e principale "vittima" delle mie preoccupazioni e dei miei affanni, che nonostante le ore, le serate e i fine settimana sacrificati, ha deciso comunque, nel frattempo, di sposarmi. Sono pronto a ricambiare il favore, con gli interessi.

Un ringraziamento perenne alla mia famiglia, i miei genitori, Daniela e Giuseppe, e mio fratello Damiano, punti fermi sui quali, ne sono certo, potrò sempre confidare. Un grazie sincero anche ai miei suoceri, Maria e Gaetano, e ai miei cognati, Marta e Salvatore, per l'affetto costante e discreto, e alla piccola Elide, che ha contribuito a rendere più lieti gli ultimi mesi di scrittura di queste pagine.

La mia gratitudine va ad Amelia, che ha fatto da supervisor a questa tesi, per la guida, il supporto e i lunghi momenti di confronto e riflessione. Buona parte di questo lavoro è il risultato di un impegno corale e per questo desidero rivolgere un ringraziamento anche a chi ha fatto parte di questa squadra: Amelia, Giulia, Vittoria, Francesco e Bart. Un grazie, in particolare, a Vittoria e Francesco per aver condiviso buona parte degli sforzi e dei piccoli traguardi di questi anni.

Grazie anche a Giovanni Fattore e Amit Nigam, per il tempo e l'attenzione che hanno dedicato a rendere migliore questo lavoro.

Infine, il mio pensiero costante a chi, ne sono certo, continua a scattare istantanee da lassù.