The Role of Transparency in AI Systems on Human Idea Evaluation Processes

Nicolai Hessing

Dissertation written under the supervision of Univ.-Prof. Dr. Christopher Lettl (WU)

Dissertation submitted in partial fulfilment of requirements for the MSc in International Management with Specialization in Entrepreneurship and Innovation, at Universidade Católica Portuguesa and for the MSc in Strategy, Innovation, and Management Control (SIMC) at WU University of Economics and Business, 31.08.2023.

Abstract

Title: The Role of Transparency in AI-Systems on Human Idea Evaluation Processes

Author: Nicolai Alexander Heßing

The evaluation of innovative ideas is a key process for successful new product development (NPD), which traditionally relies on human input. The rise of big data has introduced *artificial intelligence* (AI) as a new promising source of idea evaluation. Characterized by enhanced analytical power, heightened predictive accuracy, and advanced learning capabilities, AI emerges as an efficient alternative to human intelligence and has the potential to revolutionize the entire innovation process. Nonetheless, the intricate nature of AI systems also raises concerns about their trustworthiness and accountability, thereby underscoring the need for enhanced *transparency*. However, there is ambiguity concerning the potential implications of enhanced transparency in AI systems for human decision-making processes (e.g., idea evaluations). This thesis investigates the impact of enhanced transparency in AI systems on entrepreneurs' inclination to utilize AI advice within the context of idea evaluation processes. Furthermore, potential moderating effects of *confidence* and *familiarity with AI* are examined. Data derived from an experimental survey encompassing 366 entrepreneurs is examined employing the *Judge-Advisor System* (JAS). The findings reveal a significant positive relationship between the degree of transparency in AI systems and AI advice utilization. The relationship is moderated moderately by the level of familiarity with AI and surprisingly not moderated by the level of confidence in the entrepreneurs' own evaluation.

Keywords: Artificial Intelligence, Transparency, Idea Evaluation, Judge-Advisor System

Sumário

Título: O Papel da Transparência em Sistemas de IA nos Processos de Avaliação Humana de Ideias.

Autor: Nicolai Alexander Heßing

A avaliação de ideias inovadoras é um processo-chave para o sucesso do desenvolvimento de novos produtos (DNP), que tradicionalmente dependia do contributo humano. A ascensão de grandes volumes de dados introduziu a *inteligência artificial* (IA) como uma nova fonte promissora de avaliação de ideias. Caracterizada por uma potência analítica aprimorada, maior precisão preditiva e capacidades avançadas de aprendizagem, a IA surge como uma alternativa eficiente à inteligência humana e tem o potencial de revolucionar todo o processo de inovação. No entanto, a natureza intrincada dos sistemas de IA também suscita preocupações quanto à confiabilidade e responsabilidade, ressaltando assim a necessidade de uma maior *transparência*. No entanto, há ambiguidade em relação às implicações potenciais de uma transparência aprimorada nos sistemas de IA para os processos de tomada de decisão humana (por exemplo, avaliações de ideias). Esta tese investiga o impacto de uma transparência aprimorada em sistemas de IA na inclinação dos empreendedores em utilizar os conselhos da IA no contexto dos processos de avaliação de ideias. Além disso, são examinados os possíveis efeitos moderadores da *confiança* e *familiaridade* com a IA. Os dados derivados de um questionário experimental abrangendo 366 empreendedores são analisados empregando o *Sistema Juiz-Consultor*. Os resultados revelam uma relação significativamente positiva entre o grau de transparência nos sistemas de IA e a utilização dos conselhos da IA. A relação é moderadamente moderada pelo nível de familiaridade com a IA e, surpreendentemente, não é moderada pelo nível de confiança na própria avaliação dos empreendedores.

Palavras-chave: Inteligência Artificial, Transparência, Avaliação de Ideias, Paradigma JAS

Acknowledgement

First and foremost, I would like to express my profound gratitude to my family for their constant support and encouragement throughout my entire academic career. This master's thesis signifies the culmination of my academic journey, and I attribute a significant part of my accomplishments to their love and guidance. I am especially thankful to my parents, Dr. Christine Hessing and Michael Hessing. Their belief in me and boundless support have been a guiding light throughout my academic journey, gifting me far more than just the opportunity to study.

Furthermore, I would like to express my sincere gratitude to my supervisor, Univ.-Prof. Dr. Christopher Lettl, whose guidance and mentorship have been invaluable. Professor Lettl's expertise, and commitment to academic excellence have inspired me to push the boundaries of my knowledge and capabilities. I am immensely grateful for the opportunity to work under his supervision, which has undoubtedly enriched my academic growth. I would also like to thank Erik Kommol and Tahnee Mahr, for their valuable insights, encouragement and feedback on my work.

Table of Contents

List of Appendices

List of Figures

List of Tables

List of Abbreviations & Symbols

AI	Artificial Intelligence
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
FFE	Fuzzy Front End
GPT	Generative Pre-trained Transformer
IoT	Internet of Things
JAS	Judge-Advisor System
M	Mean
ML	Machine Learning
NPD	New Product Development
R&D	Research and Development
SD	Standard Deviation
VC	Venture Capital
WOA	Weight of Advice
XAI	Explainable Artificial Intelligence
α	Significance Level
&	And

1 Introduction

"The rise of powerful AI will be either the best or the worst thing ever to happen to humanity. We do not yet know which."

Stephen Hawking, 2016

1.1 Motivation

Generating and evaluating innovative ideas are key processes for successful new product development (NPD) (Eling et al., 2015; Kornish & Ulrich, 2014). Companies usually rely on human input to evaluate new ideas, particularly through expert panels, focus groups or advisory boards (Criscuolo et al., 2017; Krueger, 2014; Schweitzer et al., 2012). However, the emergence of big data has increased the availability of a new source of advice for evaluating ideas: *Artificial intelligence* (AI) (Logg et al., 2019). The clear superiority in analytical capacity, predictive accuracy, and learning capabilities establish AI as a fast, cost-effective, and scalable alternative (or supplement) to human intelligence (Jarrahi, 2018; Piller et al., 2023). Due to rapid advancements and substantial investments in AI technology over the last decades, AI has garnered a progressively central position in an expanding array of domains, including corporate decision-making processes (Makridakis, 2017). In the context of NPD, AI holds the potential to fundamentally change the idea evaluation process in its entirety by harnessing its inherent capabilities (Cockburn et al., 2018).

However, with the increasing application and establishment of AI systems, concerns about their trustworthiness and accountability have also increased (Glikson & Woolley, 2020). This is particularly attributed to the fact that AI systems often operate as complex *black boxes*, posing a challenge for individuals to grasp the underlying process of output generation or verify the credibility of evaluations (Arrieta et al., 2020; Ribera & Lapedriza, 2019). This inherent opacity of AI algorithms raises demands for greater *transparency* in AI systems (Editorials, 2023; Glikson & Woolley, 2020). However, there is still ambiguity and disagreement in research about what transparency in AI systems actually constitutes and, in particular, what increased transparency might imply for human decision-making (Andrada et al., 2022; Zhao et al., 2019). It is essential to acknowledge that, at least at this point, AI systems are not autonomous decision-making systems. Instead, they function more as decision-making *support* systems and serve as advisors in the context of the idea evaluation process. In light of the rapid proliferation

of AI, the role of humans as actual decision-makers (e.g., evaluators) who are potentially influenced by AI advice is a factor that has often been overlooked in prior research (Ribeiro et al., 2016).

1.2 Research Objective

Some research approaches have attempted to explore how people respond to AI advice and have yielded contradictory results. The research of Logg et al. (2019) has demonstrated consistent evidence that people prefer and utilize AI advice, despite its opacity. They coined the notion of *algorithm appreciation*, which underlines the potential and capabilities of AI. In contrast, prior research has simultaneously shown that individuals exhibit a certain degree of distrust and reluctance in utilizing AI advice, leading to an opposing notion regarding AI advice: *algorithm aversion* (Dietvorst et al., 2015; Dzindolet et al., 2002). Given the contrasting concepts of algorithm appreciation and algorithm aversion, the question arises: What specific factors influence or determine people's willingness to utilize AI advice? This question is of central importance in understanding the reasons behind the appreciation of algorithms by some individuals and the aversion of algorithms by others. This thesis addresses this research gap by examining the role of transparency in AI systems for the utilization of AI advice.

Furthermore, this thesis focuses specifically on the context of *idea evaluation processes*. While research on NPD has made remarkable progress in idea generation, relatively little is known about idea evaluation (Basadur, 1995; Lonergan et al., 2004). Previous research often adopted a one-size-fits-all approach without distinguishing between the domain of judgment and the type of judges. This research gap is also addressed by specifically focusing on the process of idea evaluation (domain of judgement) and entrepreneurs (type of judges). The following research questions are addressed:

How does transparency of AI-based idea evaluation systems affect the utilization of human evaluation in an entrepreneurial context?

What factors moderate this relationship?

The answer to the research question is by no means apparent. On the one hand, in theory, greater transparency of AI systems could increase human utilization of AI advice as the perceived black box nature of AI systems is migrated (Chao et al., 2016). Through the transparent disclosure of

verifiable and traceable information, the decision-making process becomes comprehensible, thereby fostering increased trust in the evaluation outcome (Glikson & Woolley, 2020). In addition, human biases in the idea evaluation process could be identified and migrated. Thus, increased transparency would enhance algorithm appreciation. On the other hand, there are reasonable arguments that greater transparency of AI systems could actually decrease human reliance on AI advice. Several AI models, such as deep learning or machine learning models, are complex black box models with processes that are inscrutable and, particularly for lay users, often not understandable. Disclosing information could paradoxically raise more concerns and doubts instead of alleviating them. This can lead to information overload (Bertrand et al., 2022) and reduced operational efficiency (Samek et al., 2019). Idea evaluators might feel intimidated and uncertain about relying on an idea evaluation process they do not fully comprehend, thus causing algorithm aversion.

In this thesis, a *positive* linear relationship is hypothesized between transparency in AI systems and advice utilization for the evaluation of ideas. Additionally, potential moderator effects of the variables *confidence in own evaluation* and *familiarity with AI* are tested.

For this purpose, a quantitative between-subjects research design was developed. Drawing from the *Judge-Advisor System* (JAS) paradigm, an experimental survey involving 366 entrepreneurs was designed to explore how transparency in AI systems impacts advice utilization. In the course of the survey, participants (*judges*) evaluated a fictional idea based on specific criteria. After being presented with AI advice (*advisor*) with different manipulated degrees of transparency (experimental conditions), participants reevaluated the idea. This allows statistical examination of the extent to which participants adjusted their initial evaluations in response to varying degrees of transparency in AI advice.

The results provide compelling evidence, indicating a significant positive relationship between the degree of transparency in AI-based idea evaluation systems and human decision-makers utilization of AI advice. However, this is only a general tendency. Certain individuals completely disregard AI advice, even in conditions with high transparency, while others rely entirely on AI advice in conditions with no transparency. This is consistent with the research of Parasuraman and Riley (1997), who demonstrated that human judges tend to either underutilize or overly rely on automated advice. Even though the results indicate that enhanced transparency causally explains a general tendency of algorithm appreciation to some degree, they also

highlight the complex nature of algorithm appreciation and algorithm aversion as the concepts coexist. This calls for future research to explore additional factors (besides transparency) that contribute to explaining algorithm appreciation and algorithm aversion. Furthermore, the findings indicate that the positive effect of transparency on advice utilization is *moderately* moderated by the level of familiarity with AI (the effect is more pronounced among individuals with lower familiarity with AI) and surprisingly *not* influenced by the level of confidence in own evaluation.

1.3 Structure of this Thesis

This thesis consists of seven Chapters, following the subsequent outline: This first Chapter introduces the motivation and research objectives, including specific research questions. Chapter 2 presents the theoretical background through a comprehensive literature review, covering the role of idea evaluation for innovation (Section 2.1), AI (Section 2.2), and transparency in AI systems (Section 2.3). Chapter 3 deduces hypotheses from the theoretical background and introduces the conceptual framework. Subsequently, Chapter 4 describes the methodology used to empirically test the derived theoretical hypotheses. Chapter 5 is dedicated to presenting the results of the thesis. Section 6 discusses the results, implications, and limitations, along with incentives for further research. Finally, Chapter 7 summarizes the core takeaways and concluding remarks.

2 Theoretical Background

This Chapter provides a literature review of relevant prior research. First, Section 2.1 discusses the role of idea evaluation for innovation, relevant constructs, and human biases in idea evaluation. Subsequently, Section 2.2 provides an overview of the theoretical background of AI and its role in the idea evaluation process. Moreover, state-of-the-art research on transparency in AI systems is demonstrated (Section 2.3).

2.1 The Role of Idea Evaluation for Innovation

Innovation is recognized in research and practice as a critical component for the performance and growth of companies and the creation of competitive advantage (Drucker, 2014; Van der Panne et al., 2003). In contrast to an earlier conception that innovations cannot be planned and require free thinking, recent research state that successful innovations result from rigorous processes that require the assessment of critical factors (Du Preez & Louw, 2008). The *innovation funnel* emerged as one of the most prominent processes for managing innovations (Cooper, 1990; Franken et al., 2020).

2.1.1 The Innovation Funnel

The innovation management process is often presented as a funnel with multiple stages (O'Sullivan & Dooley, 2008). For example, Franken et al. (2020) depicted the innovation funnel (adapted from the stage-gate system of Cooper, 1990) as a five-stage process (Figure 1) from *idea generation* (stage 0), *idea scoping/evaluation* (stage 1), *detailed investigation/ business case* (stage 2), *development* (stage 3), *testing & validation* (stage 4) to *launch & commercialization* (stage 5).

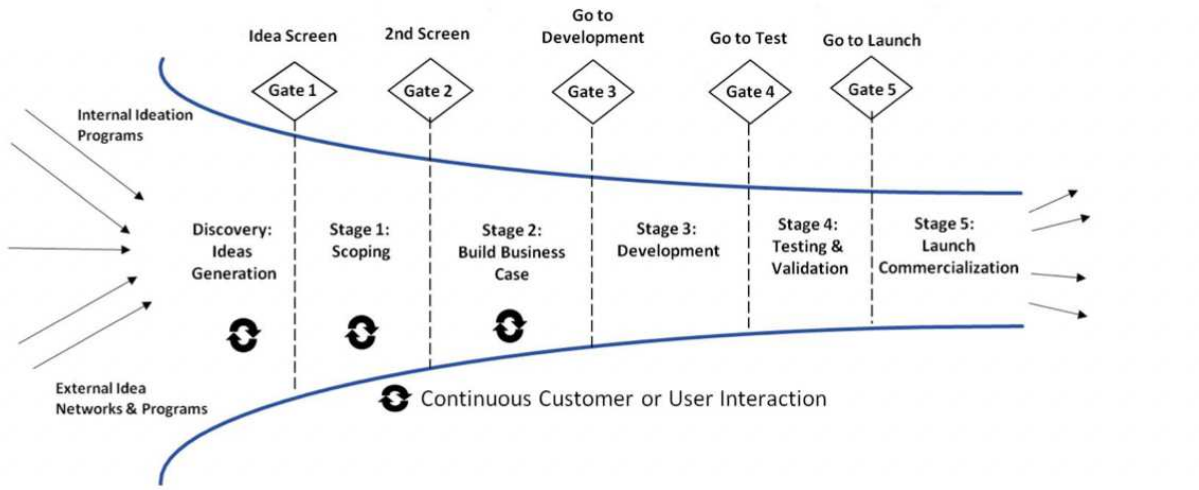

Figure 1: Innovation funnel (Franken et al., 2020, adapted from Cooper, 1990)

The innovation funnel represents a selective stage-by-stage approval process based on a set of criteria that are central to the company (Bink & Marsh, 2000). After each stage of the innovation funnel, evaluations are conducted at the "gates" to decide whether the idea has the potential to be further developed and whether it is reasonable to invest further (financial) resources for the next stage (Franken et al., 2020). The percentage of ideas that successfully passed through all gates of the innovation funnel is remarkably low. According to Barczak et al. (2009), only 14% of generated ideas were successfully commercialized, while 86% were discarded along the innovation funnel[1]. Moreover, over 60% of the initially generated ideas are already discarded in the first two stages of the innovation funnel (idea scoping/evaluation & investigation/ business case) (Barczak et al., 2009).

These early stages of the innovation funnel are referred to as the *"fuzzy front end"* (FFE) of innovation, which is "intrinsically non-routine, dynamic, and uncertain" (Kim & Wilemon, 2002, p. 270; Soukhoroukova et al., 2012). The uncertainty (*fuzziness*) in these early stages results from a lack of information about "technology, markets, required resources, company-fit and capabilities, and company limits" (Kim & Wilemon, 2002, p. 270). This uncertainty makes

[1] Based on self-reported data of the 2004 version of the comparative performance assessment study (Barczak et al., 2009)

it difficult to evaluate an idea and decide whether to pursue or discard it at the gates (Jarrahi, 2018). While the FFE may seem chaotic and unstructured, it provides the necessary freedom and flexibility to interact with customers or users, challenge assumptions, and refine ideas before committing to a specific concept (Franken et al., 2020). The FFE is a crucial stage to foster creativity, encourage interdisciplinary collaboration, and experimentation to drive innovation. Effective management of the FEE is essential for the successful development of a product or service, as it sets the foundation for subsequent stages, guiding the allocation of resources and shaping the overall direction of the project (Dwyer & Mellor, 1991).

The gate (or time) at which a go/ no-go decision is made has important implications for companies. On the one hand, as mentioned above, the early stages of the innovation funnel (FFE) are associated with high *uncertainty* and *ambiguity* about the quality of an idea (Kim & Wilemon, 2002). In a context of high uncertainty, decision-makers have vague knowledge derived from incomplete information, making it difficult to accurately evaluate the full potential and impact of a new idea (Metzger & Spengler, 2019). On the other hand, later stages of the innovation funnel are associated with *higher costs of evaluation*. Valuable resources have already been invested in advancing to the later stages of the innovation funnel, which generally are more resource-intensive (e.g. R&D) (Robertson, 1971). The required investments aim to minimize the uncertainty of the early stages but simultaneously create a stronger commitment to the idea due to higher costs (lock-in). Thus, companies face a trade-off between *minimizing uncertainty* and the *cost of evaluation* in regard to the respective stage of the innovation process. In research, this trade-off is called the *Collingridge Dilemma* (Collingridge, 1982).

Effectively managing the FFE can contribute directly to the success of a new product or service (Dwyer & Mellor, 1991). Accurate idea evaluation in the FFE plays a pivotal role in shaping the trajectory for developing a product or service, as decisions made during this stage have significant implications for the subsequent stages of the innovation funnel (Kim & Wilemon, 2002). Generally, two basic errors can occur when evaluating ideas in the FFE: Type I errors of rejecting a superior idea and type II errors of pursuing an inferior idea (Knudsen & Levinthal, 2007). For example, a company that (falsely) rejects an idea at the beginning of the funnel (type I error) could never reap the benefits of potentially successful commercialization of the idea. And if a company fails to recognize and create a steady stream of innovations, it can quickly become irrelevant in a dynamic market due to competitors who adapt better to the environment

(Geroski et al., 1993). Additionally, inaccurate idea evaluation can also lead companies to continue investing in ultimately unsuccessful ideas (type II error), which is a significant financial risk for the company, especially in industries with relatively high R&D costs. For example, the process of developing a drug can take over a decade and costs US$2.8 billion on average (Paul et al., 2021). Hence, developing a drug without carefully evaluating potential problems can have significant financial consequences for the company when it proves ineffective. The research of Thomke and Fujimoto (2000) demonstrated the costly impact of (engineering) changes later in the in the development process. Therefore, it is important for companies to minimize both type I and type II errors through more accurate evaluations early in the FFE.

2.1.2 The Process of Idea Evaluation

While research on the FFE has made remarkable progress on *idea generation* (stage 0), relatively little is known about *idea evaluation* (stage 1) (Basadur, 1995; Lonergan et al., 2004). However, given the severe consequences of type I and II errors (Section 2.1.1) and the high discarding rate of ideas in the FFE (Barczak et al., 2009), the process of idea evaluation is equally important as idea generation in terms of innovation and creative achievement (Lonergan et al., 2004). Mumford et al. (2002) state that idea evaluation is a distinctive and systematic process based on a contextual appraisal of the consequences of pursuing an idea. The predicted consequences of pursuing an idea within a particular context are then evaluated based on a set of various critical factors such as strategic alignment, market potential, (technical) feasibility, resource requirements, and risk assessment (Du Preez & Louw, 2008; Kuipers et al., 1988). The objective of the idea evaluation process is to minimize uncertainty about the quality of an idea based on the evaluated factors to subsequently decide whether an idea has the potential to be developed (stage 3) (Blair & Mumford, 2007). Moenaert et al. (1995) characterize successful teams by their ability to minimize uncertainty (fuzziness) through accurate evaluations early in the FFE.

The paper of Kim and Wilemon (2002) illustrates the pattern of the fuzziness level that decreases with each stage of the innovation funnel (Figure 2). Depending on risk aversion, strategy, technology and availability of resources, companies have an “approval level” for fuzziness (a). Once the fuzziness level falls below this threshold due to idea evaluation, the management commits to developing a product or service (b) (stage 3). In light of the

Collingridge dilemma, accurate idea evaluation in the FFE is essential to avoid potentially costly consequences of type II errors at later stages of the innovation funnel.

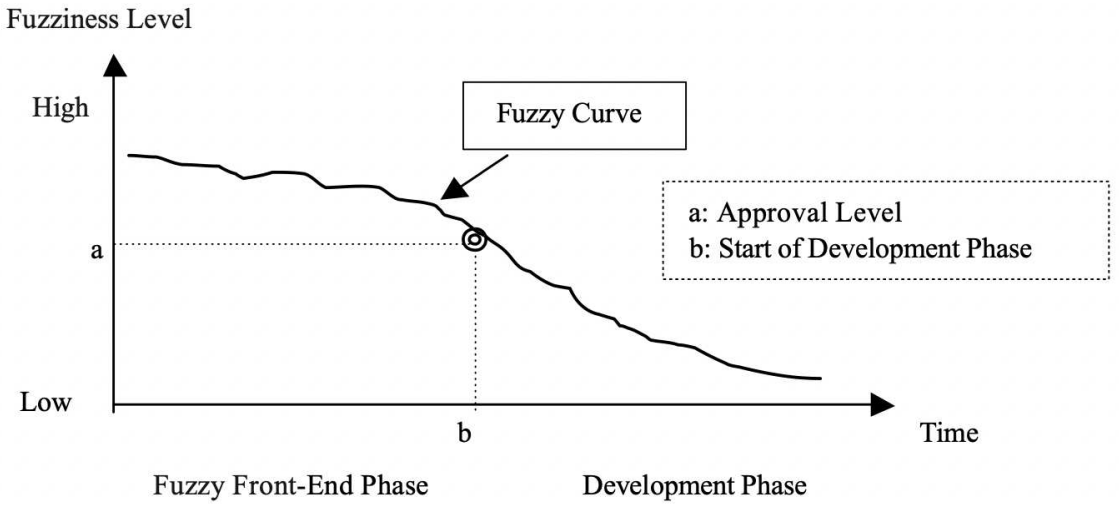

Figure 2: The Fuzziness Level along the Innovation Funnel (Kim & Wilemon, 2002)

Most companies rely on panels of internal company members and external experts to evaluate ideas (Criscuolo et al., 2017). However, there are various forms of idea evaluation, including crowdsourcing, focus groups, or advisory boards (Krueger, 2014; Schweitzer et al., 2012; Velamuri et al., 2017). The choice of evaluation form depends on factors such as the nature of the idea, available resources, time constraints, and the desired level of expertise and objectivity required for evaluation. For all forms of idea evaluation, it is important that the *idea quality* is sufficiently evaluated to make an informed decision about the allocation of resources for the development of the product or the service.

2.1.3 Constructs for Operationalizing Idea Quality

The ability to create processes and methods that generate, evaluate and implement ideas is critical to fostering innovation (Dean et al., 2006). Researchers and practitioners have been exploring techniques and instruments for years to measure and improve the quantity and quality of ideas (Cooper & Kleinschmidt, 1996; Driva et al., 2000; Griffin & Page, 1993). Today, there is a variety of methods, techniques and criteria that people use to evaluate the quality of ideas (Du Preez & Louw, 2008; Rebernik et al., 2008). This led to great inconsistency in

operationalizing the multidimensional construct *idea quality* and consequently in the comparability of previous studies (Dean et al., 2006). To address this limitation, Dean et al. (2006) examined 90 scientific papers on idea generation and idea evaluation and derived four dimensions for idea quality: *Novelty*, *workability* (*feasibility*), *relevance*, and *specificity*. The dimensions are also constructs that are operationalized based on sub-dimensions (Appendix 1). The authors conducted a confirmatory factor analysis demonstrating high loadings between the sub-dimensions and high validity between the dimensions. The constructs that determine the *idea quality* are briefly described below:

2.1.3.1 Novelty

Novelty is a key construct for creativity (Dean et al., 2006). In research, novelty is recognized as a characteristic that can influence the potential and success of an idea. A novel idea is rare, unusual and uncommon (Connolly et al., 1993). In the paper of Dean et al. (2006), novelty is a construct operationalized by two sub-dimensions: *Originality* and *paradigm relatedness*. That is, novel ideas are not only rare but also ingenious and imaginative (originality) as well as transformational (paradigm relatedness) (Besemer & Treffinger, 1981).

2.1.3.2 Feasibility

Feasibility refers to the (potential) execution or *implementation* of the idea within technological, social, legal, political, and economic *constraints* (Briggs et al., 1997; Dean et al., 2006). An accurate evaluation of the construct feasibility in the early stages of the innovation process is critical, as it represents a basis for deciding whether to pursue an idea.

2.1.3.3 Relevance

Relevance is determined by the ability of an idea to specifically *refer* to a problem and simultaneously provide a solution to *address* the stated problem (Dean et al., 2006). A relevant idea, therefore, meets the needs and expectations of its target audience and gains traction. Thus, the relevance of an idea impacts the level of support, funding, and interest an idea receives from stakeholders, including investors, customers, and partners. This is an essential step in reducing uncertainty and securing funding for the further stages of the innovation funnel.

2.1.3.4 Specificity

Specificity is a construct that measures how *clearly* and *complete* an idea is elaborated (Dean et al., 2006). Ideas with a high degree of specificity are more useful and allow for more accurate evaluations than vague and incomplete ideas. Specific ideas can more effectively reduce the high level of uncertainty at the beginning of the innovation process and enable more precise and faster decision-making processes.

2.1.3.5 Market Potential

The paper of Dean et al. (2006) derived four key constructs for idea evaluation from an extensive corpus of prior research. Evaluation these constructs is intended to reduce uncertainty and help innovators to make an informed decision about whether to pursue an idea. In an attempt to minimize type I errors (Section 2.1.1), research has also addressed the evaluation of potential *outcomes* of innovative ideas (e.g. market performance). Even though this approach is somewhat limited, since success is unknown ex-ante, the criteria of *market potential* is an important determinant for innovation potential (Frederiksen & Knudsen, 2017). In the course of this work, therefore, the dimension of market potential is taken into account in addition to the criteria of Dean et al. (2006).

2.1.4 Biases in Human Idea Evaluation

Idea evaluation processes are highly influenced by *heuristics* and *cognitive biases* (Klayman, 1995; Li, 2017; Montibeller & Von Winterfeldt, 2015; Mueller et al., 2012; Reitzig & Sorenson, 2013). Heuristics are "rules or strategies that simplify the decision-making process" (Cossette, 2014, p. 473). People use heuristics (mostly unconsciously) to make decisions faster without wasting cognitive resources, especially when the decision context involves high complexity and uncertainty (Cossette, 2014). These mental shortcuts can lead to systematic "errors in prediction or estimation" of cognitive processes, defined in research as *biases* (Tversky & Kahneman, 1974, p. 1130). Whereas heuristics reflect the cognitive processes of human decision-making, biases are the results and systematic errors that arise from limitations of those underlying cognitive processes (Cossette, 2014). In the context of idea evaluation, there is a great corpus of research demonstrating how cognitive biases affect the perception and judgement of individuals, resulting in inaccurate idea evaluations and missed opportunities for organizations (Blair & Mumford, 2007; Boone et al., 2019; Klayman, 1995; Li, 2017). Ultimately, cognitive

biases can lead to type I and type II errors in idea evaluation (see Section 2.1.1). The following paragraph introduces biases that can influence the evaluation of ideas and their impact on innovation outcomes.

The formation of judgments for accurate evaluation involves various complex cognitive processes (Klayman, 1995). However, people have *limited cognitive resources* to process large amounts of information and tend to rely on heuristics, especially in the face of uncertainty in the FFE (Metzger & Spengler, 2019). For example, the research of Klayman (1995) has shown that decision-makers selectively consider evidence that confirms their pre-existing knowledge and beliefs. This phenomenon is referred to as *confirmation bias*. It reflects a huge threat for accurate idea evaluation, as people may be more likely to select ideas that align with their existing beliefs and reject those that challenge them (Klayman, 1995; Montibeller & Von Winterfeldt, 2015). However, relying on existing beliefs may not lead to embracing innovative ideas.

Prior research has also identified several biases, particularly concerning the evaluation criteria of novelty, which is a key construct for the evaluation of ideas (Dean et al., 2006). For example, Blair and Mumford (2007) have demonstrated that people tend to reject original and risky ideas and prefer ideas that provide short-term benefits and are easy to implement. The findings of Licuanan et al. (2007) also support the notion of a *novelty bias* by identifying that people tend to underestimate and discount the value of original ideas, especially in more complex settings. Eidelman et al. (2009) and Zajonc (2001) show that people tend to reject new and risky ideas regardless of the context due to *personal preferences* for familiar ideas and maintaining the status quo. According to Li (2017), the influence of personal preferences is particularly noticeable among experts. People with expertise have an informational advantage in evaluating ideas, but also have greater personal preferences that bias their objectivity (Li, 2017). Prior research has also identified a negative *bias against creativity* when people want to reduce uncertainty, even though creativity is paradoxically desirable (Mueller et al., 2012). Innovators seeking to reduce uncertainty in FFE are ironically less able to recognize creative ideas.

Since the screening process of idea evaluation (see Section 2.1.2) is traditionally conducted by a panel of internal and external experts, the process is also subject to *in-group biases* that may distort accurate evaluations (Blair & Mumford, 2007). Research in social sciences has shown that people tend to identify with groups and favor members of those groups (Brewer, 1979;

Mullen et al., 1992). Building on that research, Reitzig and Sorenson (2013) demonstrated that managers tend to (consciously or unconsciously) favor ideas from people they know or someone who resembles them. Therefore, idea evaluations might also be biased by group dynamics, as people prefer ideas of their own social group. Furthermore, the impact of social presence from panel members on idea evaluation can vary based on individual characteristics, including the level of anxiety experienced during group communication (Gudykunst, 1993). High levels of anxiety can have a substantial detrimental impact on individuals' cognitive abilities, as their worries and fears about others overpower their capacity to fully concentrate on the task (Hwang & Won, 2021). These are just a few examples of cognitive biases that can influence the decision processes. For a detailed overview of other biases, see Montibeller and Von Winterfeldt (2015).

Cognitive biases can have a significant impact on how people perceive and evaluate innovative ideas, potentially leading to type I and II errors in the uncertain FFE. Thus, companies seek to create structured processes to minimize the impact of biases on idea evaluation (George et al., 2000; Montibeller & Von Winterfeldt, 2015). Emerging technologies also offer high potential to minimize biases for optimizing idea evaluation. In the next Chapter *artificial intelligence* (AI) is introduced (Section 2.2.1) and defined (Section 2.2.2). Additionally, this Chapter addresses the criticism surrounding AI in Section 2.2.3 and explores both the potential (Section 2.2.4) and limitations (Section 2.3.5) of AI in the idea evaluation process. Furthermore, Section 2.2.6 delves into the concept of hybrid intelligence.

2.2 Theoretical Background of Artificial Intelligence

2.2.1 The Rise of Artificial Intelligence

The total volume of data created, captured, copied, and consumed globally is projected to increase rapidly, reaching 181 zettabytes by 2025[2] (see Figure 3) (Statista, 2021). One of the primary drivers of this exponential data growth is the proliferation of the *Internet of Things* (IoT). The proportion of data captured by machines and sensors is increasing rapidly compared

[2] 1 zettabyte is equivalent to 1 trillion gigabytes

to data generated by humans due to the diffusion of IoT devices such as smart home appliances or industrial sensors (Sides et al., 2019).

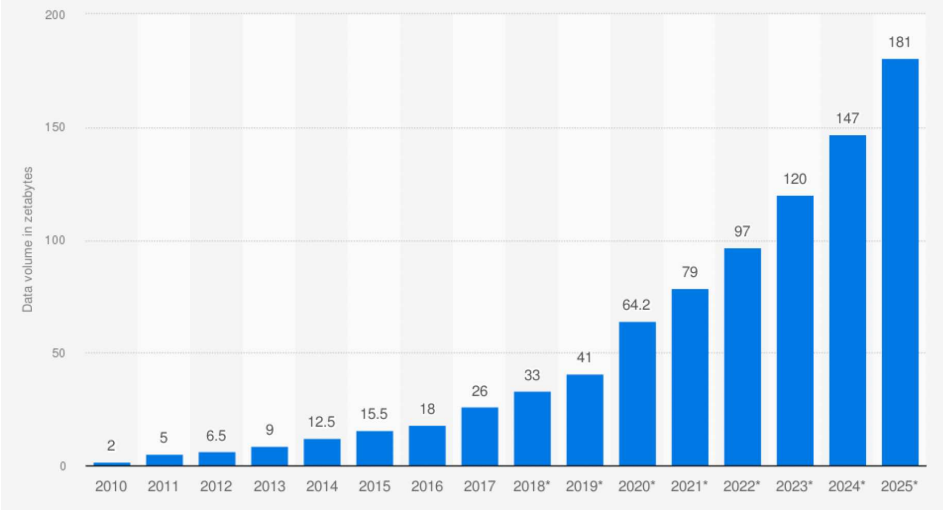

Figure 3: Volume of data (in zettabytes) created, captured, copied and consumed worldwide from 2010-2020, with forecasts to 2025 (Statista, 2021)

This rise of *Big Data* represents both great potential and challenges for people. On the one hand, Big Data analytics could reveal new insights and knowledge about society, markets, and the environment, enabling humans to anticipate opportunities, drive change, and support decision-making processes. On the other hand, the explosion of (machine-generated) data poses a major challenge for humans to create a computational infrastructure to process, analyze and manage Big Data (Chen et al., 2013; Idrees et al., 2019). According to Regalado (2013), only 0.5% of all generated data is analyzed. To overcome the challenge and exploit the great potential of Big Data, *advanced algorithms* and computing infrastructures have been developed and implemented in the last decades, driving the fourth industrial revolution (Cheng & Hackett, 2021). Algorithms are a set of instructions or rules that computer programs follow to solve a problem or accomplish a specific task. For example, clustering algorithms have emerged as a highly effective tool to categorize and analyze large amounts of data (Fahad et al., 2014). As companies grappled with the challenge of extracting insights from vast amounts of data, the proliferation of advanced algorithms for analyzing Big Data and increasing computational power laid the foundation for a new era: The rise of *artificial intelligence* (AI) (Ferràs-Hernández, 2018).

2.2.2 Definition and Capabilities of Artificial Intelligence

In management research, Ferràs-Hernández (2018, p. 260) defines AI as a "new generation of machines capable of (a) interacting with the environment, gathering information from outside (including from natural language) or from other computer systems; (b) interpreting this information, recognizing patterns, inducing rules, or predicting events; (c) generating results, answering questions; or giving instructions to other systems; and (d) evaluating the results of their actions and improving their decision systems to achieve specific objectives". While advanced algorithms form the core of AI systems by providing computational instructions, the interactive capabilities of AI allow it to learn and intelligently change its behavior in response to environmental stimuli (Frantz, 2003).

AI has witnessed rapid advancements in recent years, offering immense potential to reshape various industries including finance, education and healthcare (Cockburn et al., 2018). AI clearly outperforms humans in terms of data processing, analytical capacity, and learning capabilities (Jarrahi, 2018). For example, Dawes et al. (1989) demonstrated a general superiority of algorithmic judgment in diagnosing and predicting human behavior compared to human judgement. The meta-analysis of Grove et al. (2000) supports the notion of superior mechanical predictive power compared to human experts. State-of-the-art research characterizes AI as "fast, efficient, cheap, scalable, and consistent, enabled its core abilities of *pattern recognition*, *predictions*, and *generative* abilities" (Piller et al., 2023, p. 410). The core abilities of AI are outlined in Figure 4.

Pattern recognition abilities	An algorithm's ability to learn relevant relationships in data and spot emerging patterns, e.g. to recognize certain objects in a picture after being trained on large amounts of labeled data, to find emerging topics in large repositories of texts, to analyze relationships in data about social networks or technologies (like patent data).
Predictive abilities	An algorithm's ability to dynamically predict trends or future occurrences by techniques like support vector machines, random forest algorithms, or regression trees. Prediction tasks are perhaps the most fundamental ability of machine intelligence and have been subject of some of the most significant recent advances in ML.
Generative ("creative") abilities	An algorithm's ability to produce not just mere numerical predictions, but output considered novel and understandable by humans. Generative adversarial networks (GANs) or (transformer) language models allow machines to generate outcomes indistinguishable from human-generated output, for example artificial pictures instead of photographs, virtual objects instead of craft or design work, and computer-generated text instead of authored articles.

Figure 4: Core Abilities of AI (Piller et al., 2023)

Overall, the abilities of AI represent a potential paradigm shift how people approach problem-solving, enabling machines to not only follow predetermined rules but also learn from data and make autonomous decisions. According to Ferràs-Hernández (2018), AI is a *disruptive technology* that is increasingly able to take decisions in complex environments, driving a revolution similar to the one witnessed during the rise of the internet.

2.2.3 Criticism of Artificial Intelligence

AI has the potential to revolutionize various industries and improve human lives. However, as the AI revolution advances, criticism and concerns about the technology are increasingly being raised (Arrieta et al., 2020; Kuner et al., 2018).

One commonly raised concern is the potential impact of AI on *employment* and the *economy*. The intelligent capabilities of AI have the potential to disrupt labor markets, leading to job displacement and economic inequality. Since AI is capable of learning and intelligently changing behavior in response to environmental cues, not only automatable jobs but also creative jobs that require thinking could be disrupted or displaced (Ferràs-Hernández, 2018; Frantz, 2003). Even though AI may create new job opportunities, there is a risk that the benefits of automation primarily accrue to those with the skills and resources to adapt, exacerbating the gap between the technologically proficient and the marginalized. This division has already been observed in previous (technological) revolutions (Makridakis, 2017). In the course of the Industrial Revolution, for example, the manual labor of people was supplemented but also substituted by the power of machines (Makridakis, 2017).

As AI systems become increasingly autonomous, concerns regarding the *accountability* and *transparency* of their decision-making processes are growing (Diakopoulos, 2016; Glikson & Woolley, 2020; Pasquale, 2015). The opacity of AI systems can make it difficult for people to comprehend the reasoning behind their outputs, hindering the ability to hold them accountable for potential biases (Arrieta et al., 2020; Ribera & Lapedriza, 2019). This lack of transparency also raises concerns about the potential for AI systems to perpetuate or amplify existing societal biases, leading to discriminatory outcomes in areas such as hiring, lending, and criminal justice (Bertrand et al., 2022; G. Harris, 2020).

Privacy and (data) security represent another critical aspect associated with AI. As AI systems heavily rely on vast amounts of data, there is a heightened risk of privacy breaches and unauthorized access to sensitive information (Kuner et al., 2018). The collection, storage, and analysis of personal data raise concerns about consent, data ownership, and copyright. Inadequate data protection measures or the mishandling of data can have severe consequences, including identity theft ("deepfake"), profiling, and breaches of confidentiality.

In conclusion, the AI revolution is associated with enormous potential for a variety of industries and society, while at the same time, criticism surrounding AI is growing and multifaceted. But what is the specific role of AI in the innovation process, particularly in idea evaluation? The following Sections address the potential (Section 2.2.4) and limitations (Section 2.2.5) of AI regarding idea evaluation processes.

2.2.4 Potential of AI in Idea Evaluation Processes

AI is becoming a crucial component of many corporate decision-making processes (Makridakis, 2017). However, Cockburn et al. (2018) state that the biggest economic impact of AI is yet to come: AI as a *new method of invention*. AI has the potential to transform and accelerate the innovation process in its entirety (Cockburn et al., 2018). Griliches (1957, p. 502) designates such enabling technologies as an "invention of a method of invention" that have a greater impact on society and the economy compared to single product or service innovations. That is, AI not only has the potential to improve the automation and efficiency of end-use technologies but also to transform the innovation process itself (Cockburn et al., 2018). The following paragraphs describe how the core abilities of AI, i.e., *pattern recognition abilities*, *prediction abilities*, and *creative abilities* (Section 2.2.2), can transform the idea evaluation process.

First, the computational advantages and *pattern recognition abilities* of AI have the potential to *reduce cognitive biases* in the idea evaluation process. As described in Section 2.1.4, the idea evaluation process is strongly influenced by cognitive biases. For decades, researchers have been trying to find ways to reduce the effects of biases in decision-making, e.g. through decision support systems (George et al., 2000; Montibeller & Von Winterfeldt, 2015). Since biases in human decision-making result from limited cognitive abilities and the use of heuristics, AI can significantly reduce these biases, given its superior computational capabilities in interpreting

data, recognizing patterns, and preventing errors (Ferràs-Hernández, 2018). For example, emerging trends and customer behavior can be identified by leveraging AI to recognize patterns in data such as social media networks or patents (Piller et al., 2023). This information helps determine the key features of a new product to be developed and to evaluate its novelty or market potential. AI systems can help humans access and analyze a broader range of information and perspectives, migrating the inherent *novelty bias* that may arise from limited or incomplete information. That is, by analyzing and evaluating vast amounts of data, AI can help people recognize the value of new ideas in complex settings, overcoming the bias against novelty outlined by Licuanan et al. (2007). Through algorithmic processing, AI can assess ideas independently, focusing objectively on their merits rather than personal beliefs or preconceptions, reducing the impact of *confirmation biases* and *in-group biases*.

Second, the *predictive abilities* of AI have the potential to efficiently *reduce uncertainty* in the innovation process. As described in Section 2.1.1, the FFE is associated with high uncertainty (Kim & Wilemon, 2002). The objective of idea evaluation processes is to reduce uncertainty about the quality of the idea in order to subsequently decide whether to invest resources for developing the product or service (Blair & Mumford, 2007). AI algorithms can help reduce uncertainty in FFE and assist innovators in making informed decisions by creating accurate predictions and scenarios about the potential success or failure of new ideas (Davenport et al., 2020; Piller et al., 2023). AI can predict broad potential risks and challenges associated with an idea, such as financial, legal, or market risks. By creating models and simulations, AI can explore different scenarios and their potential outcomes. This allows innovators to evaluate the impact of different factors based on those scenarios. Consequently, AI helps innovators to prioritize ideas based on various criteria and make more informed decisions. The potential of the predictive abilities of AI is exemplified by the drug development and approval process in the pharmaceutical industry. The complex NPD of a drug can take over a decade, costs US$ 2.8 billion on average, and even then, 9 out of 10 drugs fail to make it through the stage of clinical trials due to safety or efficacy concerns (Paul et al., 2021). Accurate prediction of the success or failure of a drug is therefore of great importance to avoid costly type I or type II errors. Li et al. (2020) developed an AI model that predicts drug safety concerning drug-induced liver injuries (DILI), which accounts for 11% of liver failure cases. The AI-based model outperformed existing prediction methods, demonstrating an accuracy of 68.7% in predicting DILI occurrence for new drugs (Connor et al., 2022). This example demonstrates how the

predictive abilities of AI can help reduce the uncertainty of NDP early in the FFE to prevent costly type I and II errors.

Third, the *generative (creative) abilities* of AI represent an *additional method for invention* by providing a novel approach to generating and evaluating new ideas (Cockburn et al., 2018). AI models can simulate creative thinking processes and explore uncharted territories, thereby expanding the scope of innovation (Ferràs-Hernández, 2018). For example, AI-based (transformer) language models are able to produce and recognize output that is considered novel by humans (Piller et al., 2023). By leveraging AI algorithms to generate different concepts, innovators can break out conventional thinking patterns, stimulate creativity, and reignite the spirit of invention. Combining these creative abilities with pattern recognition abilities also enables autonomous evaluation of AI-generated ideas, e.g., concerning novelty by analyzing patent data. The concept of creative abilities in machines challenges the traditional notion that creative thinking solely originates from human abilities.

2.2.5 Limitations of AI in Idea Evaluation Processes

AI has the potential to revolutionize the process of idea evaluation, offering innovative approaches and unprecedented efficiency. However, despite the remarkable progress, it is essential to critically examine the limitations inherent in AI-based idea evaluation. Understanding these limitations is crucial for striking a balance between leveraging the potential of AI and recognizing the role of human judgment in the complex realm of idea evaluation.

First, AI systems lack *human intuition* and *emotional intelligence* (Ferràs-Hernández, 2018). In the ambiguous FFE, the decision-making process frequently necessitates utilizing limited and imperfect information (Kim & Wilemon, 2002). If there is no data or incomplete data, AI cannot make accurate predictions and assist in decision-making (Vincent, 2021). In those uncertain situations, decision-makers have often relied on *intuition*, “the ability to understand and act without the need for conscious reasoning” (Ferràs-Hernández, 2018, p. 261; Vincent, 2021). Dane and Pratt (2007) state that intuition is effective when the decision-maker has domain expertise and the decision is unstructured under uncertainty. Human intuition is a cognitive area in which humans are superior to machines as intuition is honed through experience, contextual understanding, and tacit knowledge, which are challenging to formalize and encode into AI

algorithms (Dane & Pratt, 2007; Ferràs-Hernández, 2018). Furthermore, human intuition encompasses a holistic understanding of social, cultural, and ethical dimensions, which play a crucial role in decision-making (Sinclair & Ashkanasy, 2005). AI lacks the ability to grasp the full spectrum of human values, ethical considerations, and social dynamics that influence complex decision-making processes (Ferràs-Hernández, 2018).

While the primary objective of AI is to make data-driven decisions devoid of (emotional) biases, it remains crucial to comprehend the nuanced aspects of human emotions during the evaluation process. For example, when evaluating a brand with aspects such as emotional brand attachment. However, *emotional intelligence* presents a limitation in AI due to its inherent complexity and difficulty replicating human emotional understanding and empathy. Similar to human intuition, emotional intelligence relies on contextual understanding, nonverbal cues, and the ability to navigate complex interpersonal relationships, which are challenging for AI systems to replicate accurately. Amabile (2020) and van der Aalst (2021) advocate for the concept of superior human intelligence in areas such as empathy, contextual knowledge, and common sense. Moreover, emotional intelligence is closely tied to social and cultural contexts, which vary across individuals and communities. AI systems trained on specific datasets may struggle to generalize emotional understanding across diverse populations or adapt to rapidly changing social norms. Nevertheless, considering the swift progress in the field of AI, it is reasonable to anticipate that this limitation will likely diminish in the near future. The demonstration of the creative abilities of AI is an example of how an attribute previously associated exclusively with human cognition has become a facet of AI as well, even if not yet fully developed.

Second, *AI output* critically depends on the *quality* and *quantity* of the *input data*. The quality of the training data can impact the applicability and generalizability of AI algorithms (Daneshjou et al., 2021). For example, standard AI-based systems (e.g., Chat GPT) are trained on texts on the internet. It is impossible to determine precisely which texts were used and how credible they are. Insufficient data may result in limited patterns and representations, reducing the system's ability to make accurate predictions or understand complex contexts (Halevy et al., 2009). Incomplete, old or not representative data, for example, can prevent an accurate evaluation of an idea. Adequate data quality and quantity are vital to capture the variations and complexities inherent in the problem domain, enabling the AI system to make well-founded

evaluations. In the technology lifecycle, AI is at the beginning of the growth phase and still relatively limited in capacity compared to its anticipated advances in the future. However, with increasing computing capacity and the growing amount of data available, this may not be a limitation for much longer. The development of newer versions of AI models, such as ChatGPT-4, aims to overcome the limitations of input data.

Third, *AI systems* introduce *new biases* that result from the data they were trained on or the biases embedded by their creators (Guidotti et al., 2018). In the pursuit of mitigating human biases in decision-making, AI ironically gives rise to various alternative forms of biases. For example, inherent biases in the input data used to train AI models can result in distorted evaluation of ideas as the AI system learns from this biased data (Guidotti et al., 2018). Algorithms that learn from biased human judgments consequently acquire those biases (G. Harris, 2020). When AI is considered as a tool for decision support rather than an autonomous decision-making system (see Section 2.2.6), a variety of biases can be detected that are created or reinforced by the interaction between humans and AI. For example, research in social psychology has shown that people automatically defer to automated systems and disregard contradictory information from other sources or do not search for additional information (Alon-Barkat & Busuioc, 2023). This undue overreliance on automated systems is referred to as *automation bias* (Alon-Barkat & Busuioc, 2023; Cummings, 2006; Lyell & Coiera, 2017). Schemmer et al. (2022) support the notion of human overreliance on AI advice, as humans are unable to disregard faulty AI advice.

Fourth, AI systems their underlying algorithms are *not transparent*. AI systems often operate as complex *black boxes*, posing a challenge for individuals, especially lay users, to grasp the underlying process of output generation (Arrieta et al., 2020; Ribera & Lapedriza, 2019). This opacity of AI algorithms raises concerns about accountability, interpretability, and potential biases (Glikson & Woolley, 2020). In the process of idea evaluation, decision-makers cannot verify the credibility and legitimacy of evaluations generated by AI systems. The discussion about the transparency of AI systems is a central part of this thesis and will be addressed in detail in Section 2.3.

2.2.6 The Concept of Hybrid Intelligence

As the rise of AI unfolds, the boundaries of what algorithms can achieve are expanding, empowering people to reshape their relationship with technology and redefine what it means to be creative and intelligent. Until now, human idea evaluation and AI idea evaluation have been predominantly discussed separately. However, recent research postulates a shift towards a more integrated approach known as *hybrid intelligence* (Jarrahi, 2018; Maher & Fisher, 2012; Piller et al., 2023; Siemon, 2022; van der Aalst, 2021; Vincent, 2021). By combining the collective power of human intuition and contextual understanding alongside the analytical capabilities of AI, hybrid intelligence holds the potential to enhance the accuracy and effectiveness of idea evaluation (Maher & Fisher, 2012; Piller et al., 2023). Through this collaborative relationship, humans and machines can mutually mitigate their respective biases and leverage their individual strengths, creating synergies of intelligence (Piller et al., 2023). For instance, van der Aalst et al. (2021) demonstrated that AI systems are unable to effectively manage disruptions, such as the impact of the COVID-19 pandemic on the supply chain. In the presence of sudden and substantial changes, predictive models may fail, regardless of the available data (van der Aalst et al., 2021). The unique capabilities of human intelligence in terms of adaptability, contextual knowledge, and empathy can complement machine intelligence to create synergies of hybrid intelligence (van der Aalst, 2021). Furthermore, as outlined in Section 2.2.4, the computational advantages of machine intelligence can simultaneously complement human intelligence (Ferràs-Hernández, 2018). Thus, „human and machine intelligence adapt to and collaborate with each other, forming a two-way information exchange and control" (Piller et al., 2023, p. 411). Figure 5 illustrates the concept of hybrid intelligence.

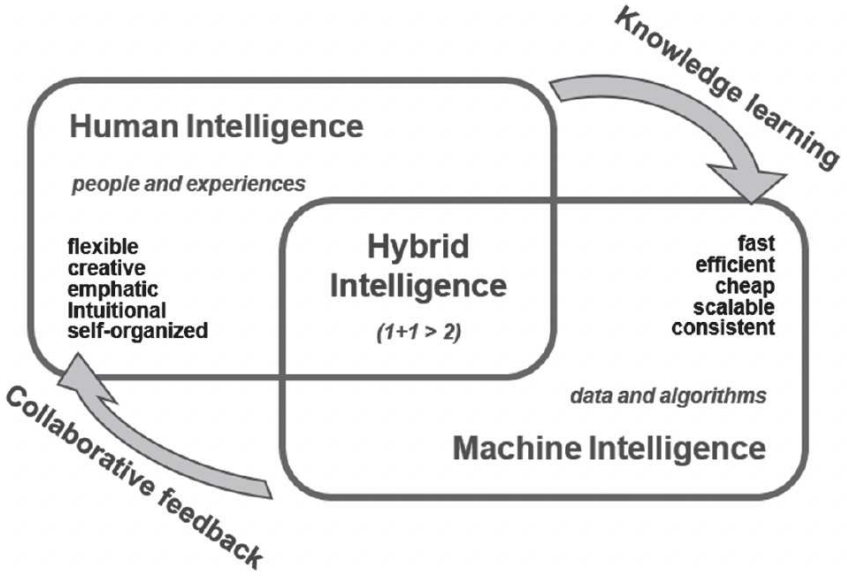

Figure 5: A Model of Hybrid Intelligence (Piller et al., 2023)

The emerging paradigm of hybrid intelligence indicates that AI is more a decision-support system rather than an automated decision-making system (Alon-Barkat & Busuioc, 2023). Facilitating effective and meaningful interaction between humans and machines constitutes a crucial aspect of hybrid intelligence. The challenge for managers is to enable employees to critically assess the quality of AI advice and prevent automation bias (Lee & See, 2004). Simultaneously, humans must enable AI to exploit its potential by providing high quality and quantity of data and provide feedback to the system in cases where low-quality outputs are generated (Piller et al., 2023). Effective interaction between humans and machines is of central importance for creating synergies. Lee and See (2004) state that poor interactions between humans and machines can be very costly and limit the potential of hybrid intelligence. Hence, understanding this interaction and human utilization of AI-based decision-support systems opens a vast demand for research. However, understanding these complex interactions requires a profound comprehension of how AI systems operate. The following Section discusses the role of transparency in AI systems.

2.3 Transparency in AI Systems

AI systems are often perceived as opaque black boxes (Arrieta et al., 2020; Guidotti et al., 2018; Pasquale, 2015). Therefore, as described in Section 2.2.5, users of AI systems may be unable to comprehend the internal workings and processes of the algorithms (Arrieta et al., 2020; Ribera & Lapedriza, 2019). The integration and advancement of algorithmic decision-making systems have provoked a tense debate about the trustworthiness of AI systems and increased demands for *transparency* (Burrell, 2016; Diakopoulos, 2016; Pasquale, 2015). For example, in the 2019 guidelines presented by the High-Level Expert Group on Artificial Intelligence of the European Commission, transparency is identified as a critical requirement for (trustworthy) AI technology (Larsson & Heintz, 2020). However, as transparency is a "multifaceted concept used by various disciplines" (Larsson & Heintz, 2020, p. 2), there is ambiguity about the definition of transparency in AI systems and, thus, what greater transparency might entail (Andrada et al., 2022). According to the comprehensive review of Glikson and Woolley (2020, p. 13), transparency "reflects the level to which the underlying operating rules and inner logics of the technology are apparent to the users". However, this perspective only encompasses a partial understanding of transparency in AI systems. Transparency in AI systems includes different types, and several neighboring concepts of transparency (e.g. explainability or interpretability) play an essential role in clarifying and explaining different facets of human interactions with AI (Andrada et al., 2022). The following Sections present research on the types of transparency in AI systems (Section 2.3.1), the degrees of transparency (Section 2.3.2), as well as the positive (Section 2.3.3) and negative effects of transparency (Section 2.3.4) on idea evaluation.

2.3.1 Types of Transparency in AI Systems

The research of Andrada et al. (2022) distinguished multiple types of transparency in AI systems (e.g., information transparency, material transparency, or transformational transparency). *Information transparency* has attracted the most attention in shaping cognitions such as judgments, decisions, or evaluations (Andrada et al., 2022; Diakopoulos, 2016). Consequently, this thesis will mainly focus on information transparency. More information and definitions regarding the other types (e.g. material transparency) can be obtained from the paper of Andrada et al. (2022).

Information transparency refers to the "disclosure of information about an AI system, typically to support judgements concerning the system's fairness, trustworthiness, safety, efficacy, accountability, and compliance with regulatory and legislative frameworks" (Andrada et al., 2022, p. 4). The notion of information transparency can be further subdivided based on the sort of information that is disclosed. Specifically, information pertaining to *input data* (Section 2.3.1.1), *explanations* (Section 2.3.1.2), and *algorithms* (Section 2.3.1.3) can be disclosed (see Figure 6).

2.3.1.1 Data Transparency

The prevalence of data breaches and misuse (see Section 2.2.3) highlights the relevance of data in AI systems and emphasizes the growing importance of *data transparency* in various contexts (Bertino, 2020). Bertino et al. (2019, p. 20) define data transparency as the "ability of subjects to effectively gain access to all information related to data used in processes and decisions that affect them". In the context of AI, data transparency refers to the disclosure and accessibility of information about the (training) data used by an AI system (Andrada et al., 2022). For example, data transparency in the evaluation of ideas entails providing information about the specific data and sources used for an evaluation outcome.

2.3.1.2 Explainable AI (XAI)

Another crucial aspect of transparency in black boxes includes explanations why a particular decision was made (Gilpin et al., 2018; Glikson & Woolley, 2020). *Explainable artificial intelligence* (XAI) has recently experienced a resurgence (Miller, 2019), driven by evidence that explanations of AI decision-making processes make opaque AI models more transparent and interpretable (Arnold et al., 2006; Lundberg & Lee, 2017; Ribeiro et al., 2016). Despite rapid growth in research on, for example, XAI techniques and user needs for XAI, there is still no consensus on the precise definition and essential factors that constitute a transparent explanation (Carvalho et al., 2019; Ehsan et al., 2021; Ribera & Lapedriza, 2019). Notwithstanding the poorly defined concept and factors of explainability, previous literature commonly agrees on the *objective* of XAI: making the operation or decisions of an AI system easy for humans to understand (Ehsan et al., 2021; Glikson & Woolley, 2020). In the context of idea evaluation, XAI provides insights into the decision-making process, revealing the significance of features in the decision and their respective weights, and incorporates clear and

comprehensible explanations and arguments, elucidating why an AI system produced a specific evaluation.

2.3.1.3 Algorithm Transparency

Algorithm transparency represents another type of information transparency, encompassing the disclosure of information concerning the underlying "nature of the algorithms" employed within AI systems (Andrada et al., 2022, p. 4). This type of transparency emphasizes the technical aspects of AI systems, such as the models employed (e.g., neural networks, k-nearest neighbor, regression, deep learning, classification, et cetera.), as well as information about the purpose of the algorithm (Zouave & Marquenie, 2017). Despite current research efforts to develop methods that enhance the transparency of AI algorithms, achieving *full* algorithm transparency remains a considerable challenge due to the complexity of advanced AI algorithms (Ananny & Crawford, 2018). Models like deep learning consist of multiple layers and millions of parameters, making it difficult for experts to trace the reasoning behind their predictions or outputs. The notion of algorithm transparency in the context of idea evaluation refers to the ability of people to understand and assess the complex algorithms and methods used by an AI system to evaluate and prioritize ideas.

Figure 6 illustrates the hierarchy of transparency within AI systems, based on Andrada et al. (2022).

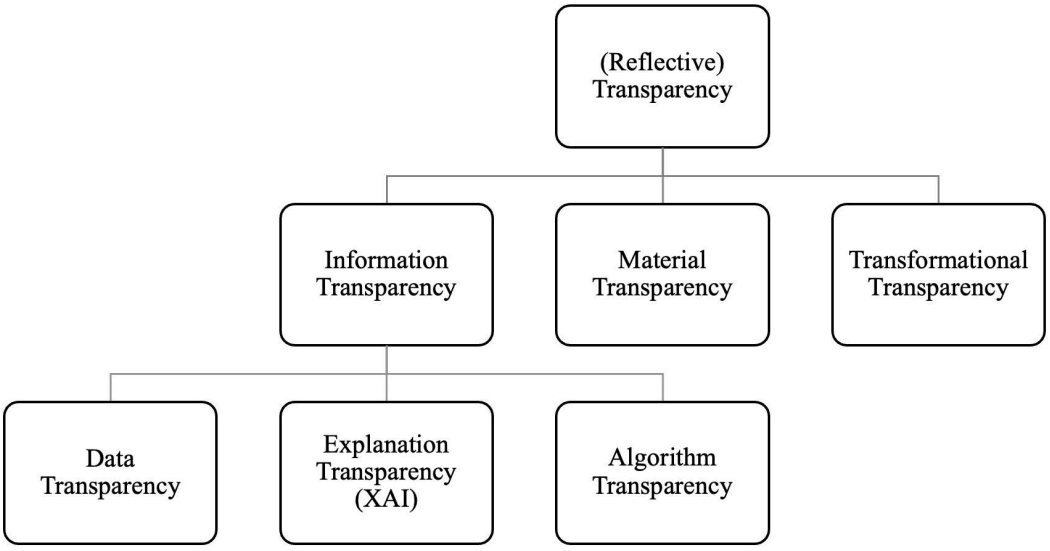

Figure 6: The Hierachy of Transparency (Own Illustration based on Andrada et al., 2022)

2.3.2 Degrees of Transparency

The construct transparency consists of multiple types (e.g., *information transparency*) and sub-types (e.g., *data transparency*) (Figure 6), with each (sub-) dimension capturing a distinct aspect of transparency (Andrada et al., 2022). This *multidimensionality* of the construct implies that transparency can have different *degrees,* contingent on the distinct manifestations of its (sub-) dimensions. The following Section distinguishes between objective and subjective degrees of transparency.

2.3.2.1 Objective Degree of Transparency: Quantity and Scope

Objective transparency can be defined as the extent to which the *quantity* and *scope* of the (sub-) types of transparency are disclosed (Zhao et al., 2019). The *quantity* of transparency types describes the degree of transparency depending on the number of *different* disclosed (sub-) types. For instance, an AI output that discloses information about multiple (sub-) types, such as data (2.3.1.1), explanations (2.3.1.2) and algorithms (2.3.1.3), has a higher degree of transparency compared to an AI output that only provides information about one (sub-) type, such as data used. The *scope* of transparency types describes the degree of transparency depending on the extent of disclosed information *within* each (sub-) type that is disclosed. For example, an AI output that discloses more comprehensive information regarding data transparency (e.g., a greater number of data sources) has a higher degree of transparency compared to an AI output that discloses less information regarding data transparency.

2.3.2.2 Subjective Degree of Transparency: Perception

Although the extent to which the quantity and scope of the (sub-) types of transparency in AI systems are disclosed contributes to a higher degree of (objective) transparency, it is crucial to acknowledge that the degree of transparency is also subjectively construed, influenced by factors such as familiarity with AI, depth of knowledge, and individual level of expertise in the field (Zhao et al., 2019). For instance, Miller (2019) and Ehsan et al. (2021) argue that, given a human-machine interaction (see Section 2.2.6), the transparency sub-type XAI is not only a product (e.g. an explanation about the system's reasoning) but also a "knowledge transfer process" that requires mutual exchange (Ehsan et al., 2021, p. 3). This implies that the degree of transparency is subject to individual variations in the ability to perceive, comprehend and interact with the disclosed information (Andrada et al., 2022). This relational notion of the

extent to which humans perceive transparency is defined as *subjective transparency* (Zhao et al., 2019). Subjective transparency encompasses the subjective perception and interpretation of objective transparency, which can vary among individuals or stakeholder groups based on their knowledge and experiences. For instance, AI developers may require less detailed algorithm transparency, whereas inexperienced users may require more information about algorithms to perceive the AI system as equally transparent. Therefore, even though AI systems may theoretically be fully (objectively) transparent, there will be no complete (subjective) transparency in practice throughout society (Andrada et al., 2022).

The increasing demands for transparency in AI systems highlight the prevailing importance attributed to this facet (Editorials, 2023; Glikson & Woolley, 2020). Yet, a deeper examination of the motivations and anticipated outcomes arising from transparency becomes imperative. The subsequent Section expounds upon the positive and advantageous impacts of integrating transparent AI systems.

2.3.3 Positive Effects of Transparency in AI Systems

2.3.3.1 Accountability and Compliance

The demand for transparency in AI systems is often rooted in the philosophical assumption that "truth is correspondence to, or with, a fact" (David, 2002, p. 1). Transparency reveals facts about the inner workings of AI systems and thus accumulates the *"truth"* required to govern and hold systems accountable (Ananny & Crawford, 2018). For example, XAI (see Section 2.3.1.2) enhances human comprehension of the decision-making processes employed by AI algorithms and thus holds the potential to enable responsible and ethical AI deployments across various domains (Arrieta et al., 2020; Gilpin et al., 2018). This is essential to ensure compliance with regulatory and ethical standards that protect the rights and interests of individuals and society (Ehsan et al., 2021).

2.3.3.2 Bias Detection and Migration

The impact of transparency on (cognitive) biases is a subject of ongoing and contradictory discussions within research. In principle, transparency serves as a fundamental pillar for the detection and mitigation of inherent biases by providing an improved understanding of opaque processes. For example, if an AI system is trained on historical data that reflects societal prejudices or discriminatory practices, transparency may expose those biases by revealing how

the AI system arrived at its decisions. This is of great importance for the implementation and improvement of fair AI systems, which could reduce discrimination and promote social justice. However, it is important to acknowledge that a high degree of transparency may also present certain negative implications with regard to cognitive biases. These implications will be explored in detail in Section 2.3.4.2.

2.3.3.3 Trust

The comprehensive review of recent empirical research on human trust in AI by Glikson and Woolley (2020) reveals that transparency plays a pivotal role in developing trust. The prevailing hypothesis posits that transparency in AI systems, especially XAI, fosters a greater understanding of the technology and consequently leads to increased trust (Arnold et al., 2006; Mercado et al., 2016; Miller, 2019). Explanations of how AI systems work shape human perception regarding the competence and capabilities of AI (Pieters, 2011). Chao et al. (2016) demonstrated that the perception of the technological capabilities of AI can contribute to a greater reliance on the technology, thereby facilitating the utilization of the potential of AI (see Section 2.2.4). Hence, trust in AI represents a fundamental prerequisite to effectively utilize and fully realize the potential of AI systems (Venkatesh et al., 2016).

2.3.3.4 Improvement of the System and Learning from the System

Moreover, the objective of transparency in AI systems is to enhance the system's performance and functionality while facilitating more efficient learning from the system (Ribera & Lapedriza, 2019). By providing visibility into the inner workings of AI algorithms, transparent systems allow developers and researchers to gain insights into the system's strengths and weaknesses and to provide AI with the right quantity and quality of data. That can lead to iterative improvements and reduced failures, and improve the safety of the AI system in terms of data breaches and data misuse (Gilpin et al., 2018). Additionally, transparency enables users and stakeholders to better understand the system's behavior and outcomes, facilitating effective knowledge acquisition and utilization. This could, for example, accelerate the innovation process. Simultaneous improvement of the system and learning from the system underlines the collaborative synergy between humans and machines, exemplifying the concept of hybrid intelligence (see Section 2.2.6).

2.3.4 Negative Effects of Transparency in AI Systems

Despite the strong demand for transparency in AI systems and the benefits mentioned, research has also examined the potential negative effects of transparency.

2.3.4.1 Information Overload and Decreasing Efficiency

The extensive quantity and scope of information required for achieving comprehensive (objective) transparency have the potential to result in information overload (Bertrand et al., 2022; Hosseini et al., 2018; Zhao et al., 2019), consequently leading to reduced operational efficiency (Samek et al., 2019; Tintarev & Masthoff, 2007). The phenomenon of information overload arises due to the inherent limitations in human cognitive capacity, described in Section 2.1.4. With the promise of AI to overcome human cognitive limitations through superior computational power, the extensive disclosure of information about the AI system may also paradoxically overwhelm and confuse users. The detrimental impact of information overload manifests in reduced efficiency, as it increases processing time (Tintarev & Masthoff, 2007) and distracts users from essential information (Ananny & Crawford, 2018). Arrieta et al. (2020, p. 83) report a "trade-off between the performance of a model and its transparency". Especially people without technical expertise or familiarity with AI systems often have difficulties understanding and interpreting the information provided by the system, which also may lead to reduced trust and confidence in AI systems (Hosseini et al., 2018).

2.3.4.2 Creation and Reinforcement of Biases

As described in Section 2.3.3.2, increased transparency in AI-assisted decision-making has the potential to detect and thus migrate inherent cognitive biases that might exist in training data or are learned by algorithms. However, a high degree of transparency in AI systems ironically creates and/or reinforces alternative forms of biases (Bertrand et al., 2022). This, in turn, influences human decision-making processes through the interaction between humans and AI. For example, Bertrand et al. (2022) state that XAI, as a sub-type of information transparency (Section 2.3.1.2), can cause or reinforce confirmation bias and other reasoning errors. The more detailed explanations and information users receive about the AI system and its reasoning, the greater the potential for them to selectively choose evidence or information supporting their pre-existing beliefs. This highlights the complexity and ambivalence of the effect of transparency in AI systems on biases. On the one hand, transparency increases the likelihood

of biases being detected and mitigated, while on the other hand, it can reinforce or create other biases. Undoubtedly, deploying XAI without ensuring that explanations are aligned with the cognitive processes/ constraints of the user poses a significant risk for decision-making processes (Bertrand et al., 2022).

2.3.4.3 Overreliance

Similar to the discussion about biases, it becomes imperative to critically assess the relationship between transparency and trust. While trust in AI facilitates the realization of its inherent potential (Section 2.2.4), it is also essential to be aware of the potential consequences of blind and unjustified trust, which can lead to dependencies and overreliance. Bertrand et al. (2022), for example, demonstrated that the confirmation bias triggered by XAI could lead to overreliance, emphasizing the need for caution and critical thinking in interpreting the provided explanations. Overreliance on AI can pose significant risks, particularly because algorithms learning from biased data tend to internalize those biases (G. Harris, 2020) or commonly utilized generative AI applications such as Chat GPT occasionally "invent" sources (Mirowski, 2023). Overreliance on AI systems without critical evaluation or independent verification can lead to potential biases going unnoticed and human judgment and expertise being neglected. The tendency to heuristically rely on AI, can also be referred to as *automation bias* (see Section 2.2.5).

3 Hypothesis Development and Conceptual Framework

The previous Chapter presented a literature review of relevant research on idea evaluation (2.1), the role of AI in idea evaluation (2.2), and transparency in AI systems (2.3). This Chapter addresses the specific research question outlined in Section 1.2. Hypotheses are derived from prior research, and a conceptual model is presented to provide a structured framework for the thesis.

3.1 Utilization of AI Advice in Idea Evaluation Processes

Many decisions in our daily or professional lives are made with the help of advisors (Schemmer et al., 2022). A great corpus of research demonstrated how people receive and utilize advice from other people like experts, friends or family (Yaniv & Kleinberger, 2000). The *Judge-Advisor System* (JAS) paradigm has emerged as a prominent and widely adopted framework in cognitive sciences for investigating the multifaceted dynamics of advice-giving and reception (Sniezek & Buckley, 1995; Tauchert & Mesbah, 2019). The JAS typically involves a scenario where one participant takes on the role of a judge who faces a decision-making task, and another participant acts as an advisor who provides guidance or suggestions to the judge (Sniezek & Buckley, 1995). The JAS paradigm allows researchers to examine various aspects and interactions of the advice-giving and receiving process, such as factors influencing the advice utilization of the judge (e.g., confidence or trust in the advisor). Previous research has primarily focused on examining these factors within the context of interactions between human judges and human advisors (Bonaccio & Dalal, 2006). The rise of AI has increased both the "availability and utility of a new source of advice: algorithms" (Logg et al., 2019, p. 90). The role of AI as an advisor in the JAS paradigm, and its consequential effect on advice utilization of human judges, remains a relatively underexplored domain within the realm of research. Utilization of AI advice refers to the extent to which individuals adapt their initial judgement towards advice provided by AI. Considering the complexity of the decision-making research stream, it is not surprising that various studies have yielded different findings. Human judges tend to either underutilize or overly rely on automated advice (Parasuraman & Riley, 1997). The generalized question of whether AI advice leads to a higher advice utilization compared to human advice remains a subject of ongoing debate and results in two conflicting notions: *algorithm aversion* and *algorithm appreciation*.

3.1.1 Algorithm Aversion

Even though advanced algorithms often outperform human judgement, empirical studies have presented compelling evidence contributing to *algorithm aversion*, whereby individuals exhibit a certain degree of distrust and reluctance in utilizing AI advice (Dietvorst et al., 2015; Dzindolet et al., 2002). The research of Dietvorst et al. (2015) demonstrates a consistent tendency for individuals to prefer human advice over AI advice and a higher level of tolerance towards errors originating from human advisors compared to algorithmic errors. According to the authors, this phenomenon can be attributed to a loss of confidence in algorithms when individuals witness repeated instances of the same mistake being made. The research of Yeomans et al. (2019) indicates that people seem averse to AI systems because they do not understand the recommendation process of black box algorithms and therefore prefer and trust human recommenders. Concerns regarding the credibility of AI are also compounded by the fact that certain AI systems, such as Chat GPT, have been observed to invent sources (Mirowski, 2023).

3.1.2 Algorithm Appreciation

Logg et al. (2019) documented a contrasting phenomenon in utilizing AI advice: *algorithm appreciation*. The notion of algorithm appreciation refers to the consistent adherence and preference of individuals to utilize AI advice (over human advice). In a series of decision-making experiments, Logg et al. (2019) have presented compelling evidence indicating that, given equal advice, individuals exhibit a greater tendency to adhere to advice when they believe it originates from AI systems[3]. This is because humans acknowledge the potential of AI (see Section 2.2.4) and attribute more rationality and objectivity to AI systems compared to humans (Beerbaum & Puaschunder, 2019). Therefore, it is not surprising that algorithm appreciation dominates, especially in domains that require a certain degree of logic and accuracy (Logg et

[3] Additionally, Logg et al. (2019) examined factors influencing algorithm appreciation such as the expertise/familiarity of the judges (Experiment 4) and the role of one's own judgment (Experiment 3). These aspects are further explored in Section 3.3 and 3.4, respectively.

al., 2019). Due to the complexity and multifaceted nature of the role of AI within the JAS paradigm, it is inevitable to differentiate the underlying decision-making context. On the one hand, individuals tend to rely more on advice from other humans (e.g. friends) in subjective domains, such as decisions involving personal preferences (e.g. book recommendations) (Sinha & Swearingen, 2001). On the other hand, individuals tend to prefer AI advice in objective domains that require a certain degree of logic and accuracy (Dijkstra et al., 1998).

As algorithm appreciation or algorithm aversion depend on the decision-making context, further research is required to delineate the exact role of AI within the JAS across various domains, particularly in entrepreneurship. More specifically, there is a notable gap in understanding the utilization of AI advice in the process of idea evaluation (Section 2.1). Even if it could be argued that the process of idea evaluation involves personal preferences to a certain extent, the evaluation of an idea is fundamentally based on objective and measurable indicators. Therefore, this thesis builds on the notion of *algorithm appreciation* by Logg et al. (2019).

3.2 The Effect of Transparency in AI-Systems on Advice Utilization

The findings of Logg et al. (2019) imply that humans utilize the advice of AI systems, even though they may not fully understand the underlying processes. The inherent trust people place in black boxes is evident in various aspects of everyday life, such as the reliance on planes or autonomous vehicles. Based on the notion of algorithm appreciation, the question arises of how an increasing degree of transparency in AI systems affects the utilization of AI advice. Does a higher degree of transparency lead to greater advice utilization through accountability, bias detection, trust and system optimization (Section 2.3.3) or do negative effects of transparency such as information overload, new biases and overreliance (Section 2.3.4) predominate, leading to lower advice utilization? This paper delves into investigating the question above within the context of idea evaluation.

I hypothesize that the positive effects associated with transparency in AI systems (as discussed in Section 2.3.3) and the recognition of the potential of AI in facilitating idea evaluation processes (as explored in Section 2.2.4) dominate the negative effects and limitations that transparency in AI systems may present. A higher degree of transparency neutralizes some of the arguments of algorithm aversion advocates. For instance, as mentioned in Section 3.1.1, judges seem averse to AI systems because they do not understand the recommendation process

and have no direct access to the underlying explanations of the advisor. As a subtype of transparency, XAI aims to provide precise explanations for AI systems, countering algorithm aversion.

Additionally, as described in Section 3.1.1, people are especially averse to AI systems when they observe system errors. Such errors significantly diminish people's trust in AI systems, thereby reducing technology adoption and acceptance (Tauchert & Mesbah, 2019; Venkatesh et al., 2016). However, the research of Dzindolet et al. (2002) demonstrated that providing *explanations* for potential errors actually increased trust and technology adoption. Chao et al. (2016) support the positive influence of transparency on trust in AI systems. They found that understanding the capabilities of AI can contribute to a greater reliance on technology. In the context of idea evaluation, I hypothesize that:

H1: *There is a positive relationship between the degree of (information) transparency in AI-based idea evaluation systems and the utilization of AI-advice by human decision makers.*

3.3 The Role of Familiarity with AI on Advice Utilization

The importance of familiarity with AI has been recognized as an important yet confounded factor within the JAS paradigm, impacting the utilization of advice (Belanche et al., 2019; Logg et al., 2019; Sinha & Swearingen, 2002).

On the one hand, it is reasonable to argue that increasing familiarity with AI leads to greater utilization of AI advice. Individuals with greater familiarity with AI have better capabilities to process and understand AI advice (Zhao et al., 2019), resulting in higher levels of *subjective transparency* (see Section 2.3.2.2) and trust. Analogous to the first hypothesis and the notion of algorithm appreciation, a higher degree of transparency would, in turn, lead to increased utilization of AI advice. Furthermore, a higher degree of familiarity with AI fosters seamless collaboration with AI systems, resulting in enhanced effectiveness in improving the system and gaining valuable insights from its operations (see Section 2.2.6). This iterative process can lead to greater trust and utilization of AI advice.

On the other hand, contrary to expectations, Logg et al. (2019) have paradoxically shown that users with high familiarity with AI (e.g. experts in forecasting) tend to utilize AI advice less. In contrast to lay users, individuals with a high level of familiarity with AI demonstrate a deeper understanding of the operational boundaries and biases of AI systems. This can be exemplified by considering the fact that AI might cite non-existent sources (Mirowski, 2023). The presence of sources can increase the credibility and trustworthiness of AI outputs for people who are not familiar with AI. However, users who are familiar with AI may be more sceptical because they know that AI systems occasionally make errors and „invent" sources. While lay users are often impressed by the capabilities and enthusiastic about the potential of AI, users with greater familiarity may exhibit more caution and hesitancy when it comes to relying solely on AI advice.

Whether people with higher familiarity with AI are more averse to AI advice or appreciate it is controversial and might depend on the context. The second hypothesis addresses the question of how familiarity with AI influences the relationship between the degree of transparency and the utilization of AI advice in the context of idea evaluation.

I argue that users with greater familiarity tend to perceive AI systems as more (subjectively) transparent, thereby requiring less information to consider a system as transparent compared to lay users. As a result, a high degree of transparency has a more significant effect on the utilization of AI advice for users with lower familiarity[4]. Additionally, the substantial volume of information in systems with high transparency increases the likelihood that potential biases and false information will be detected. Individuals with a higher degree of familiarity with AI are more likely to identify these issues compared to users with lower familiarity with AI. Based on the findings of Logg et al. (2019), the following hypothesis is stated:

[4] AI-advice utilization refers to the extent to which individuals *relatively* adapt their initial judgement towards advice provided by AI

H2: *Familiarity with AI moderates the relationship between the degree of (information) transparency in AI-based idea evaluation systems and the utilization of AI advice by human decision-makers, such that human decision-makers with high familiarity with AI decreases the positive effect.*

3.4 The Role of Confidence in own Evaluation on Advice Utilization

Another critical factor influencing advice utilization within the JAS paradigm is the confidence judges have in their own judgement (Grimaldi et al., 2015; Lewandowsky et al., 2000). Confidence can be defined as a metacognition that relies on a subjective feeling about own thoughts and judgements (Luttrell et al., 2013).

Research in cognitive science has yielded robust findings indicating a common tendency among individuals to disregard the advice of others and assign greater weight to their own judgment (Harvey, 1997; Yaniv & Kleinberger, 2000). This phenomenon can be primarily attributed to concepts such as (over)confidence (Harvey, 1997), egocentrism (Soll & Mannes, 2011), or differential information (Glikson & Woolley, 2020). The research of Woolley and Risen (2018) indicates that individuals have a strong attachment to their intuition and tend to prioritize intuitive judgements, even when they acknowledge that an alternative judgment is more likely to be correct. In light of the rise of AI, researchers have increasingly shifted their focus towards examining AI as a source of advice. Consistent with prior research on confidence, Logg et al. (2019, p. 97) demonstrate a "reduction in algorithm appreciation when individuals are forced to choose between their own judgement and that of an algorithm". Additionally, Lewandowsky et al. (2000) state that people with great confidence in their capabilities are less trustful and tend to rely less on AI.

I argue that those empirical findings on confidence negatively moderate the relationship between the degree of transparency in AI-based idea evaluation systems and the utilization of AI-advice. The following hypothesis is stated:

H3: *Perceived confidence in own evaluation moderates the relationship between the degree of (information) transparency in AI-based idea evaluation systems and the utilization of AI advice by human decision-makers, such that higher perceived confidence in own evaluation decreases the positive effect.*

The hypotheses are visualized in Figure 7. The conceptual diagram illustrates the main hypothesis of transparency in AI systems on utilization of AI advice, as well as two moderator effects.

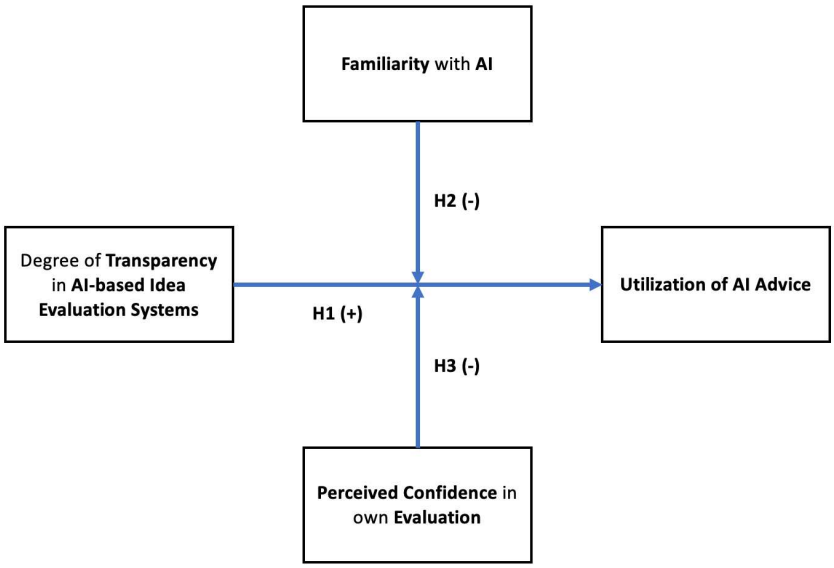

Figure 7: Conceptual Diagram of the Multiple Moderation Model

4 Methodology

This thesis examines the role of transparency in AI systems on human idea evaluation processes. In Chapter 3, the relationships between the degree of transparency, advice utilization, familiarity with AI and confidence in own evaluation are hypothesized. This Chapter describes the methodology used to test the derived theoretical hypotheses on an empirical level. Section 4.1 provides an introduction to the empirical setting, followed by a comprehensive description of the sample and data collection procedure in Section 4.2. Furthermore, Section 4.4 outlines the variables used for the subsequent analysis.

4.1 Empirical Setting

A quantitative research design was developed to explore the role of transparency in AI systems on the utilization of AI advice. Building on previous research on the JAS paradigm, an *experimental survey* was designed to test causal relationships (Malhotra et al., 2017). The experimental survey follows a randomized *between-subjects design* where participants are assigned to different hypothetical experimental conditions, and each participant experiences only one condition. The survey questionnaires were based on descriptions of hypothetical scenarios. Those hypothetical scenarios were systematically manipulated to causally explain which factors had a significant impact on the participants' judgments. Applying experimental designs in survey research enables researchers to determine causal relationships while integrating the experimental intervention into representative surveys. This approach enables researchers to maintain experimental control without sacrificing generalizability (Krupnikov & Findley, 2015). However, as participants engage with hypothetical scenarios (unlike real-life situations in conventional experiments), there is a risk that participants may not be able to immerse themselves in the described scenario and thus generate biased responses. Therefore, it is important to present scenarios realistically and plausibly and conduct manipulation checks.

4.2 Sample & Survey Distribution

The experimental survey conducted in this study has a clear focus on the entrepreneurial context, with an emphasis on *idea evaluation processes*. Hence, it is crucial to ensure that the participants possess the requisite experience or expertise in the entrepreneurial context. Therefore, the survey targeted only entrepreneurs who met specific criteria. These criteria were assessed through a series of selective questions at the beginning of the experimental survey,

ensuring a representative sample of entrepreneurs with increased response quality. Participation was open to individuals who a) have (co-)founded a start-up, b) have the intention to found a start-up, c) have worked in a start-up within the last two years, or d) have attended courses at university on entrepreneurship, innovation or business planning. Only participants who met at least one criterion were eligible for participation (the experimental survey can be found in Appendix 2). The required sample size was determined a priori by running a G Power analysis. Analogous to Logg et al. (2019), 0.80 power and medium effect sizes (0.25) were considered, indicating a required sample size of 179 participants.

The distribution strategy for the survey includes various channels with a particular emphasis on social media platforms. In addition to direct outreach to entrepreneurs within my personal network, anonymous survey links were posted in different LinkedIn groups and university communities (WU Vienna, LMU Munich and Católica Lisbon). Cold acquisition methods, such as QR codes placed in libraries and offline events, were also employed to reach potential participants. Furthermore, the survey was shared within two startup communities where I was employed and was distributed through the WU E-Mail distributor. This multifaceted approach aimed to maximize survey participation and capture diverse perspectives of entrepreneurs.

4.3 Systematic Data Collection Procedure

The following paragraph describes the data collection procedure (for the entire survey, see Appendix 2). Subsequent to the participants' consent to the *survey conditions* and their response to the aforementioned selective questions regarding *entrepreneurial experience*, they first had to indicate their *familiarity with AI* based on a 7-point Likert scale.

Afterwards, all participants were presented with a fictional scenario. They were instructed to adopt the perspective of a business development manager of a large (fictional) clothing manufacturer named Mosaic, which has recently encountered a decline in its operational performance. Their responsibility was to evaluate a potential innovative business model to counteract the negative trend. This business model, referred to as Mosaic+, was described as a rental subscription service where customers pay a monthly fee to access a collection of clothing

items that they can wear and return via delivery or local stores[5]. Participants were initially provided with a description of the dimensions of idea quality (*novelty*, *feasibility*, *relevance*, and *market potential*) based on the definitions of Dean et al. (2006) (see Appendix 1). Subsequently, all participants were instructed to evaluate each quality dimension of Mosaic+ (*initial estimate*)[6] and indicate their level of confidence in their evaluations. For this purpose, a numerical scale ranging from 0 to 100 was employed, with 0 representing the lowest evaluation (respectively, the lowest level of confidence) and 100 indicating the highest evaluation (respectively, the highest level of confidence).

After all participants evaluated the idea quality of Mosaic+ (average of *novelty*, *feasibility*, *relevance* and *market potential*), they were randomly assigned to treatment groups. Across all treatment groups, the scenario explained that participants would receive assistance from an AI-based idea evaluation system, which evaluated Mosaic+ on the same scale. (*AI advice*). However, the degree of *transparency* of the AI advice varied depending on the treatment group. The precise elucidation of the distinct degrees of transparency is outlined in Section 4.4.1.

Following the disclosure of systematically manipulated AI advice, participants were asked to re-evaluate Mosaic+ (*final estimate*). Hence, participants were confronted with the decision of either adhering to their initial estimate or determining the appropriate extent of revision/adjustment in response to the (manipulated) advice generated by the AI system (*advice utilization*). Figure 8 illustrates the randomized assignment of participants to one of four treatment groups, each representing a different degree of transparency of the AI advice.

After re-evaluation of Mosaic+, a questionnaire was employed to measure the perceived *subjective transparency* (sub_trans) of the AI output in the respective treatment group. This assessment involved the utilization of a 7-point Likert scale across four items (see Appendix 2). The underlying objective of this questionnaire is to conduct a manipulation check to ascertain whether participants in treatment groups with higher degrees of transparency also

[5] Mosaic+ is a suitable exemplification of a business model innovation due to its inherent comprehensibility, thereby obviating the necessity for an exhaustive elaboration that might overwhelm the participants' attention

[6] It should be noted that the dimension of *specificity* (Dean et al., 2006), was omitted due to the fictional nature and lack of detailed elaboration of the idea. Additionally, in line with previous research by Frederiksen & Knudsen (2017), the dimension of *market potential* was additionally incorporated for the evaluation of idea quality.

perceived the AI advice as significantly more transparent compared to treatment groups with lower transparency.

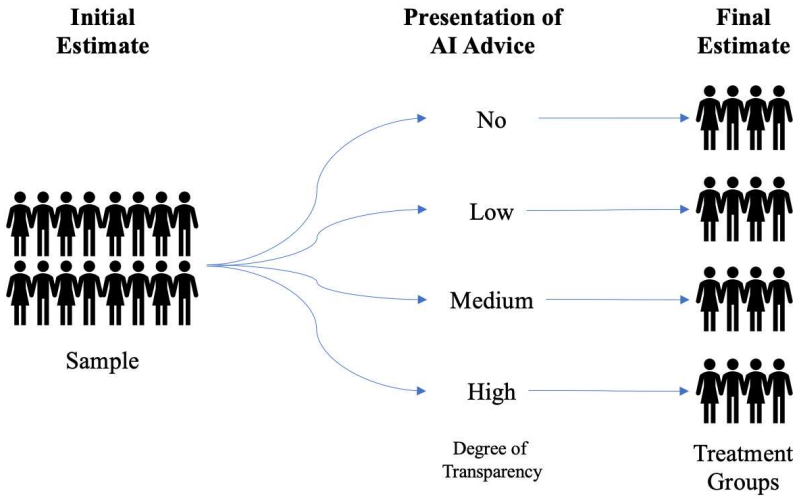

Figure 8: Randomized Allocation of four Treatment Groups in a Between-subjects Design

4.4 Variables

In the previous Sections, the constructs and the hypotheses have been discussed theoretically. This Section is dedicated to the empirical level of the study, focusing on the operationalization of the independent, dependent, moderator, and control variables within the framework of the experimental survey.

4.4.1 Independent Variable: Degree of (objective) Transparency

In this experimental survey, the *degree of transparency* in AI-based idea evaluation systems was systematically manipulated (independent variable) to examine the impact on *advice utilization* (dependent variable). As described in Section 2.3.2, the degree of transparency depends on the extent to which the quantity and scope of the (sub-)types of transparency are disclosed. In this thesis, I focus exclusively on the quantity of (sub)types of transparency (see Section 2.3.1). That is, the degree of transparency is operationalized by the number of distinct (sub)types of transparency that are concurrently presented. The overview of the treatment groups is presented in Table 1.

Treatment Group	Presented Types of Transparency	Reference
1) No Transparency	AI Rating (only the score (0-100) of the four dimensions is presented, without any further information)	*Appendix 3*
2) Low Transparency	AI Rating + Data Transparency	*Appendix 3* *Section 2.3.1.1*
3) Medium Transparency	AI Rating + Data Transparency + XAI	*Appendix 3* *Section 2.3.1.1* *Section 2.3.1.2*
4) High Transparency	AI Rating + Data Transparency + XAI + Algorithm Transparency	*Appendix 3* *Section 2.3.1.1* *Section 2.3.1.2* *Section 2.3.1.3*

Table 1: Degrees of Transparency

All participants complete the same survey; however, the presented AI advice only discloses information about the type of transparency corresponding to their assigned treatment group. The information about each type of transparency was generated utilizing Chat GPT-3.5. Appendix 3 shows the input prompt employed and the output (evaluation) generated by Chat GPT. However, in the survey, the AI evaluation was not linked to Chat GPT to avoid biases associated with Chat GPT´s brand reputation (MacKenzie et al., 1986). Hence, the study distinguishes four distinct treatment groups, each representing different degrees of transparency based on the inclusion of various (sub)types of transparency.

4.4.2 Dependent Variable: Weight of Advice

Consistent with numerous studies on the JAS paradigm, the variable *Weight of Advice* (WOA) is employed as the dependent variable to operationalize advice utilization (Bailey et al., 2022; Gino & Moore, 2007; Logg et al., 2019; Tauchert & Mesbah, 2019). WOA is calculated by

dividing the difference between the final and initial estimate generated by each participant by the difference between the AI advice and the initial estimate (Logg et al., 2019)[7]:

$$WOA = \frac{Final\ Estimate - Initial\ Estimate}{AI\ Advice - Initial\ Estimate}$$

WOA captures the magnitude of adjustment of the participant´s initial estimate as a function of exposure to the AI advice. For rational decision-makers, the continuous range of WOA is usually between 0 (0%) and 1 (100%) (Tauchert & Mesbah, 2019). WOA of 0% occurs when participants completely ignore AI advice (final estimate = initial estimate). WOA of 100% occurs when participants completely adopt AI advice (final estimate = AI advice). Values of WOA between 0% and 100% represent a partial adoption of AI advice. For example, WOA of 50% indicates that a participant assigns equal weight to their own initial estimate and the AI advice. This suggests that the participant considers both sources of information to be equally valuable in forming the final estimate. Irrational and, therefore, not very common are values below 0% or above 100%. Values below 0% indicate that individuals adjust their initial estimate in a direction opposite to the AI advice and values above 100% indicate that participants overly rely on AI advice and adjust their initial estimate even further than the AI advice (Tauchert & Mesbah, 2019). A common approach in prior research is to "*winsorize*" those values (Logg et al., 2019). Winsorization involves replacing values below 0% by 0%, and values above 100% by 100% to ensure that the values are within a meaningful range and maintain the interpretability and validity of the analysis. The calculation of WOA is not possible when the initial estimate is equal to the AI advice, as it leads to a mathematical division by zero. This scenario indicates that participants completely disregard and simultaneously fully adopt AI advice. These cases cannot be used for analysis, but they are also unlikely to be encountered frequently since the idea evaluation scale ranges from 0 to 100.

[7] For each dimension of idea quality, participants initially provide their own evaluation (initial estimate). Subsequently, participants are presented with an AI evaluation that varies in the degree of transparency, depending on their assigned treatment group (AI advice). Following this, participants are instructed to re-evaluate the idea (final estimate).

By operationalizing AI advice utilization through WOA, values can be computed for each participant and dimension of the construct idea quality. The *WOA of idea quality* can be calculated by averaging the WOA values for each dimension of all participants. Thus, the hypotheses can be tested in relation to idea quality in general. Additionally, exploratory analyses can be conducted to investigate which specific dimensions are statistically significant and whether variations among the dimensions collectively constitute the construct idea quality.

4.4.3 Moderator Variables: Familiarity with AI & Confidence in own Evaluation

Familiarity with AI is an important factor within the JAS paradigm, impacting the utilization of AI advice (Belanche et al., 2019; Sinha & Swearingen, 2002). The measure of Belanche et al. (2019) was applied to operationalize the participant´s familiarity with AI. However, the scale of familiarity was adapted to the context of AI instead of robots (see Appendix 2). Therefore, familiarity with AI is operationalized by the following three items: 1) *I have worked with AI or taken classes on AI*, 2) *I have had experience interacting with AI throughout my life*, 3) *I am familiar with AI contents (produced texts, pictures, et cetera.)*. All items are measured through self-reports based on a seven-point Likert scale, ranging from 1 (strongly agree) to 7 (strongly disagree). The average of these three items measures familiarity with AI. In accordance with the recommendations of Hinshaw (2007), the moderator variable was assessed prior to the manipulation. This ensures the independence of the moderator from the treatment assignment and eliminates any systematic biases that may affect the validity.

The third hypothesis investigates the potential moderating effect of the metacognition *confidence* (Section 3.4). In prior research on human decision-making, confidence is commonly assessed through retrospective judgments (Grimaldi et al., 2015). Participants typically provide a self-reported confidence rating subsequent to a primary cognition (e.g., judgment). The confidence level can range from a state of complete certainty to a state of complete uncertainty regarding the validity of a primary cognition (Luttrell et al., 2013). In previous studies on decision-making, various scales were employed to quantitatively assess the level of confidence (e.g., binary choices or open-ended ratings). The most commonly used scale is a continuous confidence scale, where participants self-report their level of confidence on a spectrum ranging from 0% (representing complete uncertainty) to 100% (representing complete certainty) (Grimaldi et al., 2015). In this experimental survey, the continuous confidence scale is

employed to quantify the metacognitive judgment of confidence after each evaluation of a dimension (primary cognition).

4.4.4 Covariates

For the robustness of the analysis, several covariates are also included to account for potential confounding factors. The objective of incorporating these additional variables is to control for their effects and enhance the accuracy of the findings.

Drawing from the contributions of Minson et al. (2011) and Logg et al. (2019), the control variable *distance of advice* has also been incorporated. The distance of advice is calculated as follows:

$$Distance\ of\ Advice = Initial\ Estimate - AI\ Advice$$

By incorporating the variable *distance of advice*, the potential influence of the alignment between the AI advice and the initial estimate is controlled for evaluating idea quality and respective dimensions.

The participants' *demographics* are collected in the final step of the experimental survey. Specifically, the demographic variables *age*, *gender*, *education*, *employment*, and *nationality* were measured. This information allows for a comprehensive understanding of the sample, facilitating the exploration of potential relationships between demographic factors and the dependent variable. Demographic variables aid in the generalizability and interpretation of the study findings within specific population segments. The specific questions can be found in Appendix 2.

5 Results

This Chapter presents the results of the thesis. Section 5.1 provides a comprehensive overview of the sample composition, and Section 5.2 elucidates the data pre-processing procedures employed. Subsequently, Section 5.3 offers a detailed presentation of the descriptive statistics. Finally, Section 5.4 is devoted to the hypotheses testing.

5.1 Sample Description

After conducting a pilot test to assess the clarity, comprehensibility, consistency, and logical coherence of the survey design, the data for the experimental survey was collected via Qualtrics from April 27th 2023 to June 19th 2023. Throughout this period, a total number of 380 responses have been recorded[8]. However, 9 participants did not fulfil any criteria outlined in Section 4.2, which were established to ascertain the requisite level of expertise within the entrepreneurial context, and 5 participants failed the attention check. Consequently, these responses were excluded from the analysis, resulting in a final sample size of 366, surpassing the minimum sample size of 179 determined a priori in Section 4.2. The participants were randomly assigned to four distinct treatment groups (Figure 8), yielding a total of 95 responses in group 1 (no transparency), 91 responses in group 2 (low transparency), 88 responses in group 3 (medium transparency), and 92 responses in group 4 (high transparency). The observed differences in the number of responses per group are minimal and negligible.

The observed range of the variable *Age* (in years) goes from *min* = 16 to *max* = 59 with a mean (M) of 26.6 and a standard deviation (SD) of 5.07. The distribution is slightly right-skewed, with a relatively young average age for entrepreneurs (Appendix 4). In the sample, 52.18% (191) of participants identified as female, 46.45% (170) identified as male, and a small proportion of 1.37% (5) did not specify their *Gender* (Appendix 5). An examination of the composition of the variable *Education* indicates that the majority of participants have achieved a Bachelor's degree (55.46%) or a Master's degree (37.16%). A small proportion of 5.74% possess a high school diploma, while 1.64% have obtained a doctoral degree (Appendix 6).

[8] The recorded responses were from participants who willingly provided their consent and completed the survey in its entirety

This distribution highlights the prominence of individuals with undergraduate and postgraduate qualifications within the sample. Regarding the participants' *Employment* status, a substantial portion are students (44.81%). Furthermore, 38.25% of the participants are employed full-time, while 12.57% work part-time. For a detailed overview of the remaining employment structure, please refer to Appendix 7. Due to limited resources for this thesis and the emphasis on leveraging my personal network, the majority of participants are predominantly from Germany (57.34%), Austria (13.56%), and Portugal (5.08%). The complete list of *Nationalities* can be found in Appendix 8.

5.2 Data Pre-Processing

5.2.1 Scale Reliability

Testing for scale reliability is crucial when using multi-item scales to operationalize constructs or variables. Reliability refers to the internal consistency and stability of the measurements obtained from the scale. In other words, reliability is the extent to which the items in the scale measure the same underlying construct consistently across different participants. Especially in the case of self-reported measures, it is important to conduct a test of scale reliability to mitigate potential measurement biases. Consistent with previous studies, the scale reliability of the variables *familiarity with AI* and perceived *subjective transparency* of each treatment group was evaluated using *Cronbach's alpha* (Tauchert & Mesbah, 2019). The scale adapted from Belanche et al. (2019) to measure *familiarity with AI* has a Cronbach's alpha of 0.81 (Appendix 9). According to Streiner (2003), this indicates a *very good* level of reliability among the items in the scale[9]. Cronbach's alpha was also computed for the variable of subjectively perceived transparency in each of the four treatment groups. The first treatment group (no transparency) exhibited a Cronbach's alpha of 0.90, indicating a very good level of internal consistency. The second (low transparency) and third group (medium transparency) had acceptable Cronbach's alphas of 0.61 and 0.63, respectively. The fourth group (high transparency) demonstrated a Cronbach's alpha of 0.77, indicating a good level of internal consistency. Overall, these findings

[9] A value that is > 0.6 is acceptable, > 0.7 is good, > 0.8 is very good and a value of > 0.9 can mean that there may be redundant items (Streiner, 2003)

indicate that the scales measuring familiarity with AI and subjective transparency in all treatment groups possess acceptable to very good levels of reliability (Appendix 9).

5.2.2 Manipulation Check

The experimental survey distinguishes four distinct treatment groups, each representing different (*objective*) degrees of transparency based on the inclusion of various (sub)types of transparency. The validity of the study crucially depends on whether these objectively distinct degrees of transparency are also subjectively perceived as different (see Section 2.3.2.2), thus ensuring a successful manipulation of transparency. The variables *sub_trans* (of treatment groups 1-4) will be described descriptively, followed by testing for significant differences. This entails determining whether *sub_trans* in treatment group 1 is perceived as significantly less transparent compared to *sub_trans* in treatment group 2, et cetera.

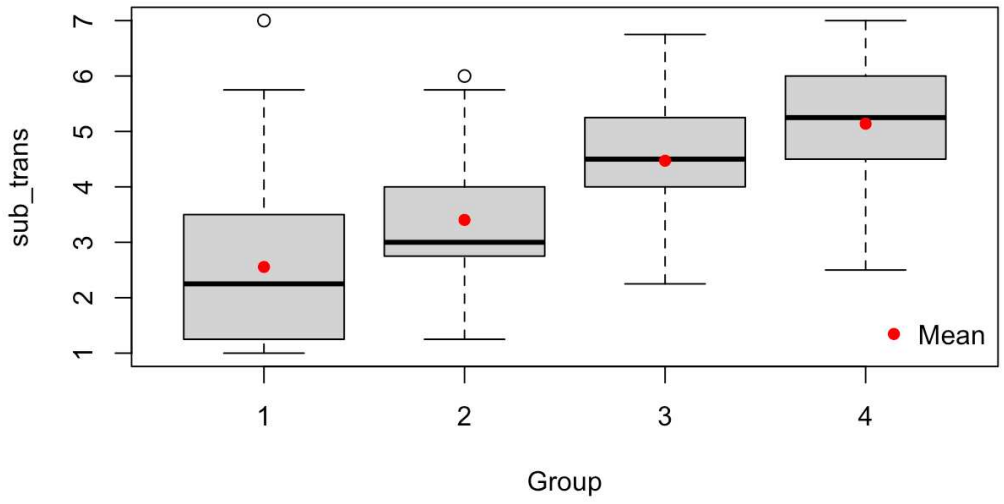

Figure 9: Box-plots of subjective transparency per Group

Based on the visual representation of the boxplots (Figure 9), it can be observed that there is a positive association between the grouping variable (degrees of transparency) and the means (and medians) of *sub_trans*. Treatment groups that include progressive enhancements of (objective) transparency of the AI evaluation conducted on Mosaic+ tend to have higher mean values of perceived (subjective) transparency. Notably, a few isolated outliers were detected within groups 1 and 2. Furthermore, the SD in Group 1 (1.42) was comparatively higher than in the other groups. For a comprehensive overview of the precise descriptive statistics, please refer to Appendix 12.

An ANOVA and pairwise t-tests were conducted to compare the means of the *sub_trans* variable across the treatment groups to determine whether the differences were significant. The results of the ANOVA yielded a large and significant F-value of 268.5 and a small p-value ($p < 0.001$). The results suggest strong evidence against the null hypothesis (no group differences) and indicate that the observed differences in means are highly unlikely to occur by chance alone. There is a significant difference in the (increasing) mean values of *sub_trans* across the four treatment groups. Therefore, the manipulation of transparency was successful. For all values of the ANOVA analysis and the pairwise t-tests, refer to Appendix 13.

5.3 Descriptive Statistics

5.3.1 Idea Evaluation

5.3.1.1 Idea Quality

Idea quality is determined by calculating the average of the evaluations across the individual dimensions, including *novelty*, *feasibility*, *relevance*, and *market potential*. Before the treatment (*initial estimate*), the participants evaluated the idea quality of Mosaic+, with $M = 58.92$ and $SD = 16.97$. The lowest evaluation was $min = 13.75$, and the highest evaluation was $max = 97.25$. See Appendix 14 for the histogram and descriptive statistics.

Subsequent to the randomized allocation of participants to the four treatment groups (*AI evaluation 1-4*), another evaluation of idea quality was conducted. This evaluation was conducted after the participants were exposed to AI advice, with different degrees of transparency (Section 4.3). In the first treatment group (no transparency), the mean value of the evaluation of idea quality is $M = 66.22$ ($SD = 15.88$); in the second treatment group (low transparency) $M = 66.44$ ($SD = 14.37$); in the third treatment group (medium transparency) $M = 68.96$ ($SD = 14.43$); and in the fourth treatment group (high transparency) $M = 70.34$ ($SD = 13.17$) (Appendix 15). All mean values of the treatment groups are significantly higher compared to the mean value of idea quality before the treatment ($p < 0.0001$) (see Appendix 16). However, there are no significant differences in the mean values of idea quality between the treatment groups (Appendix 15).

5.3.1.2 Dimensions of Idea Quality

In addition to considering idea quality as an average measure, it is valuable to differentiate the underlying dimensions that constitute idea quality (i.e., *novelty*, *feasibility*, *relevance* and *market potential*). Examining these individual dimensions can provide more nuanced insights into the construct idea quality.

The AI evaluation of Mosaic+ generated by Chat-GPT is significantly higher than the evaluation of individual dimensions by humans. Chat-GPT has rated the dimensions of *novelty* and *market potential* with a score of 80 and *feasibility* and *relevance* with 90. Figure 10 demonstrates that the initial *human evaluation* (pre-treatment) was significantly lower for all dimensions. It can be observed for all dimensions that the ratings improve with increasing transparency towards the AI advice. Detailed descriptive values can be found in Appendices 17-21.

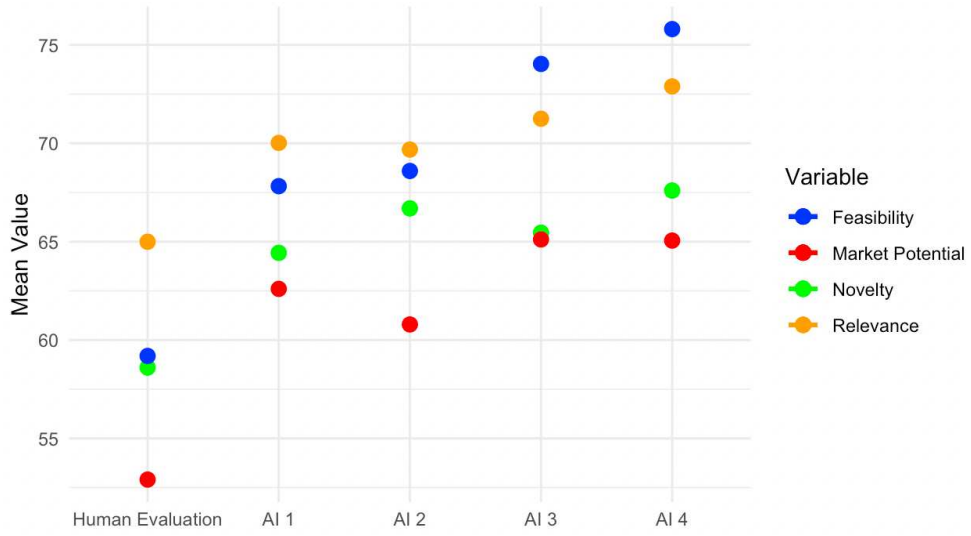

Figure 10: Evaluation of Idea Quality Dimensions across Treatment Groups

5.3.2 Weight of Advice (WOA)

So far, only the *absolute ratings* have been compared. The following Section focuses on the actual adjustment of the initial evaluation in response to AI advice through the variable WOA.

5.3.2.1 Idea Quality

Firstly, the variable WOA (Section 4.4.2) is computed for the overall *idea quality*, taking into account the average evaluation of the four dimensions and the average AI advice ((80 + 80 + 90 + 90) / 4 = 85). Following the methodology employed by Logg et al. (2019), values of less than 0 (0%) and greater than 1 (100%) are *winsorized*. Additionally, cases where the initial estimate is equal to the AI advice are excluded in accordance with the guidelines outlined in Section 4.4.2[10].

On average, participants revised their evaluation of idea quality by 38.43% (*SD* = 0.24) towards AI advice in treatment group 1 (no transparency), 43.78% (*SD* = 0.26) in treatment group 2 (low transparency), 46.65% (*SD* = 0.27) in treatment group 3 (medium transparency) and 50.60% (*SD* = 0.30) in treatment group 4 (high transparency) (Appendix 22). That is, in treatment groups with higher transparency, the mean values of the variable WOA are also higher for idea quality (Figure 11). The histograms of WOA (idea quality) in Appendix 22 show that for an *average* WOA in treatment group 1, a relatively large number of participants completely disregarded the AI advice. However, this proportion decreases in groups with higher transparency. The investigation of the statistical significance of these differences is tested in Section 5.4.1.1.

[10] Since the scale for the idea evaluation ranges from 0-100, an exact match of the initial estimate and the AI advice occurred on average only 3 times per group. These values were not considered for the analysis.

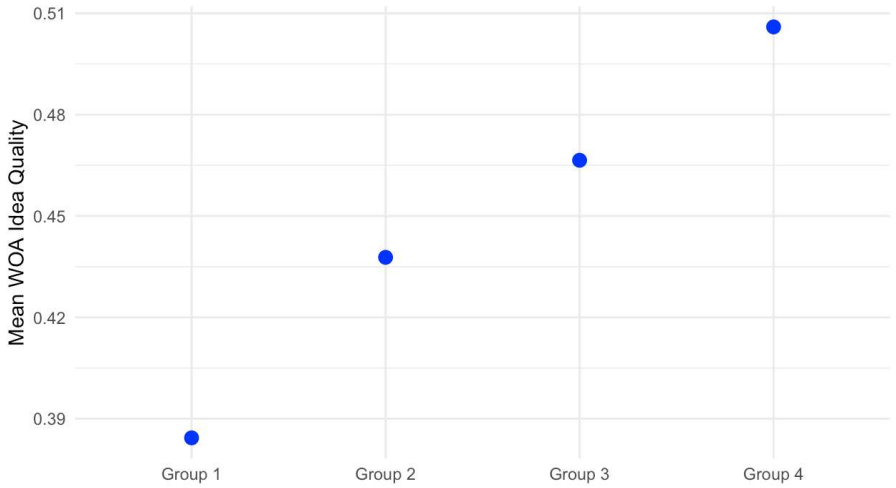

Figure 11: Mean Values of WOA across Groups (Idea Quality)

5.3.2.2 Dimensions of Idea Quality

For a more comprehensive analysis, this paragraph distinguishes the underlying dimensions of idea quality. Figure 12 demonstrates the mean WOA values per dimension across the four experimental conditions. The data reveals a notable increase in the mean values of WOA as the degree of transparency increases with each treatment group. However, not all dimensions experience equal increases. Specifically, the dimension of *relevance* exhibits a slight decrease from treatment group 3 to group 4. The mean value of WOA for the dimension *feasibility* increased the most from the lowest WOA value in treatment group 1 (35.70%) to the highest in group 4 (55.79%). This leads to the question of whether transparency has a different effect on AI advice utilization depending on the dimension. Section 5.4.1.2 is dedicated to test hypotheses.

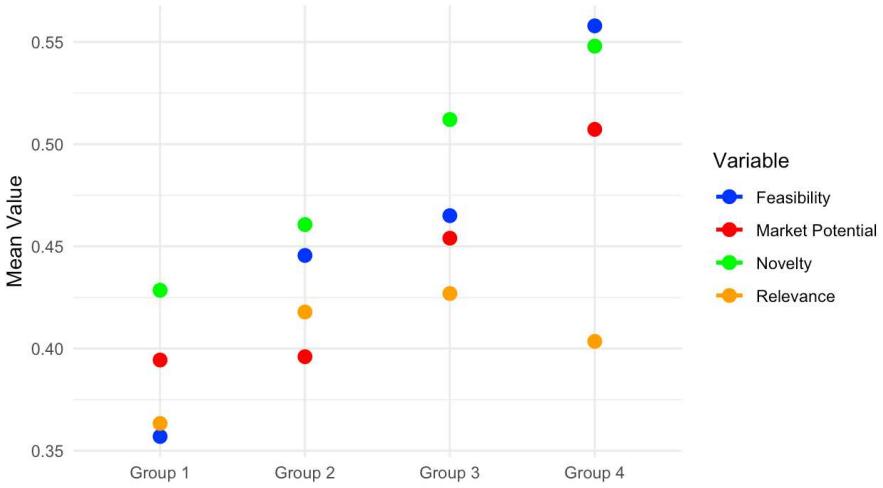

Figure 12: Mean Values of WOA across Groups (per Dimension)

5.3.3 Moderator Variables

Since all items are reliable for measuring the underlying construct, familiarity with AI (*FAM*) and subjective transparency (*sub_trans*) were determined by averaging the items. The variable *Familiarity with AI* exhibits a slightly left-skewed distribution (Figure 13), with $M = 5.21$ and $SD = 1.32$ (Appendix 10).

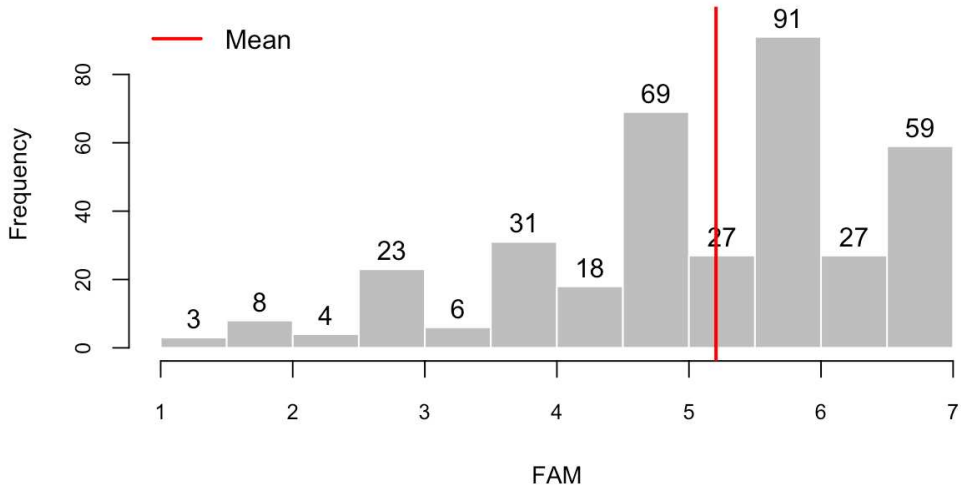

Figure 13: Histogram Familiarity with AI

It is not necessary to determine the scale reliability of the variable *confidence in own evaluation* since the confidence scale is a single-item scale. However, the confidence rating was measured for each evaluation dimension (i.e., *novelty*, *feasibility*, *relevance*, *market potential*). Figure 14 demonstrates the *average confidence rating* derived from these four dimensions. The distribution is approximately normally distributed with M = 68.91 and SD = 15.53 (Appendix 10).

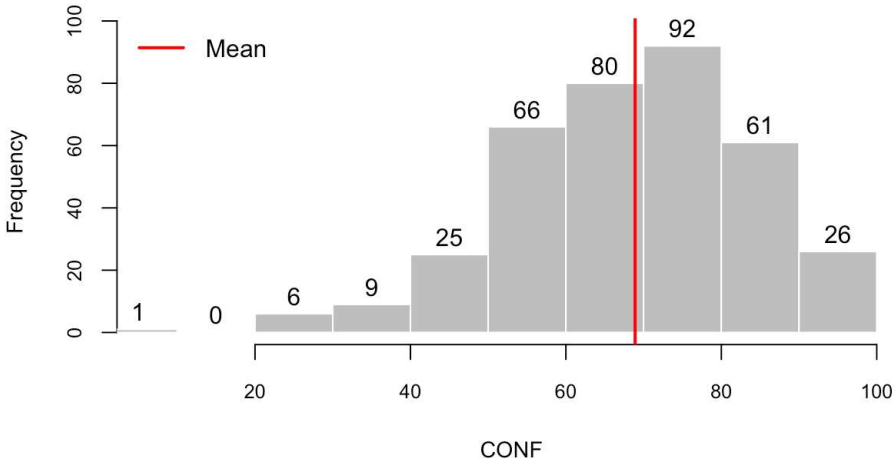

Figure 14: Histogram Confidence in Own Evaluation

Figure 15 presents a distinction in the confidence rating across the various evaluation dimensions. The confidence ratings exhibit only minimal variations among the dimensions (Appendix 10). However, an *Analysis of Variance* (ANOVA) revealed that those differences are significant (F = 6.73), as indicated in Appendix 11.

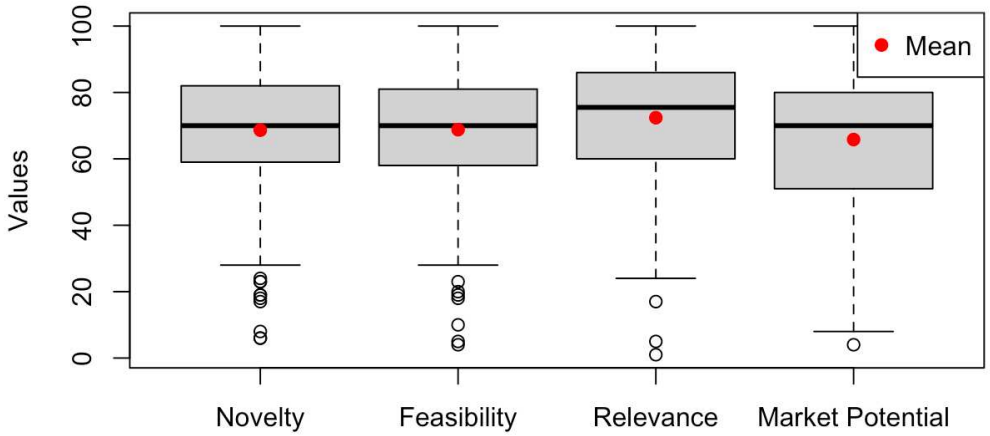

Figure 15: Level of Confidence per Evaluation Dimension

5.4 Hypotheses Testing

The hypotheses (Chapter 3) are tested in various models using the statistical software R. Null hypotheses are rejected at a significance level of $\alpha = 5\%$.

5.4.1 Hypothesis 1

The first hypothesis posits a positive relationship between the degree of transparency in AI-based idea evaluation systems and the utilization of AI advice (WOA) by human decision-makers. Descriptive statistics have shown that increased transparency is associated with higher mean values of the variable WOA regarding *idea quality* (Section 5.3.2.1) and the underlying *dimensions of idea quality* (Section 5.3.2.2). The following paragraphs will test these differences for their statistical significance and examine causal relationships.

5.4.1.1 Idea Quality

Model 1 utilizes ANOVA as a statistical methodology to determine significant differences in the mean WOA values among the various treatment groups. Before examining the validity of the results, it is crucial to assess the assumptions underlying ANOVA. First, the assumption of independence of observations is satisfied due to the random sampling employed in the between-subjects design. Consequently, the measurement outcome of one treatment group does not influence the measurement outcome of another treatment group. Second, the dependent variable WOA adheres to a ratio scale measurement, enabling meaningful comparisons of the ratios

between its values. Third, the variability of the variances (homoscedasticity) of the dependent variable WOA is approximately equal across treatment groups. The result of *Levene's Test for Homogeneity of Variance* (Appendix 27) indicates no significant evidence to reject the null hypothesis of homogeneity of variance among the groups ($p = 0.48$). However, the fourth assumption of ANOVA pertaining to the normality of the dependent variable WOA is violated. The results of the conducted Shapiro-Wilk tests (see Appendix 28) present robust evidence to reject the null hypothesis that the distribution of the variable WOA adheres to a normal distribution within all treatment groups. However, the deviation of normality does not necessarily invalidate the use of ANOVA. Recent simulation studies have demonstrated the robustness of ANOVA to violations of the normality assumption (Blanca Mena et al., 2017). Nonetheless, in addition to conducting ANOVA (model 1), the non-parametric Kruskal-Wallis test (model 2) is also employed to compare the medians of the treatment groups.[11] This non-parametric test is an alternative to ANOVA when the assumptions of ANOVA, such as normality, are violated or not met. This increases the robustness of the analysis.

Both model 1 and model 2 indicate that the differences among the treatment groups are significant. The ANOVA analysis reveals a statistically significant effect of the treatment groups on the dependent variable WOA, with a p-value of $p = 0.002$ (Appendix 29). The observed variation among the groups (higher *mean values* of the variable WOA in treatment groups with higher transparency) is unlikely to be due to chance. The robust Kruskal-Wallis rank sum test also reveals a significant difference in the medians of the dependent variable WOA among the groups with an associated p-value of $p = 0.002$ (Appendix 30).

Model 3 is based on the statistical method ANCOVA (Analysis of Covariance). This method extends the concept of ANOVA through regression analysis by incorporating categorical independent variables *and* continuous covariates to explain the variation of the dependent variable WOA. The main objective of ANCOVA is to determine if there are significant differences in the means of the dependent variable across different levels of the categorical independent variable (treatment groups) while controlling for the effects of covariates.

[11] It should be noted that ANOVA compares means, while the Kruskal-Wallis test compares medians.

Incorporating covariates in the statistical model enables the consideration and adjustment of their potential influence on the dependent variable. This adjustment helps mitigate the risk of confounding factors and enhances the accuracy and reliability of the statistical analysis. In Model 3, dummy variables for all categorical covariates were created. The reference categories chosen for gender, education, employment, and nationality were as follows: *female*, *bachelor*, *student*, and *Germany*.

Model 3 also reveals a statistically significant effect of the treatment groups on the dependent variable WOA (Appendix 31). The F-value of $F = 6.14$ indicates a robust relationship between the predictor variable and the response variable. The extremely low probability of observing this F-value under the null hypothesis ($p < 0.001$) provides compelling evidence to reject the null hypothesis. Furthermore, the group variable, with a sum of squares of 1.82, explains a notable amount of variability in the response variable (WOA) across different groups. Considering the significant outcomes of models 1, 2, and 3, there is substantial evidence to reject the null hypothesis associated with hypothesis 1 confidently.[12]

5.4.1.2 Dimensions of Idea Quality

For a more nuanced analysis, in this Section, ANOVA is applied to the underlying dimensions of idea quality. Models 4 - 7 (Appendix 32) respectively focus on *novelty*, *feasibility*, *relevance*, and *market potential*. Thus, it is possible to determine significant differences of the mean WOA-values per dimension among the treatment groups.

The data suggests that there are statistically significant differences in the mean values of WOA among the groups only in regard to the dimension *feasibility* ($F = 5.02$; $p < 0.01$). There are no significant differences of the variable WOA among the dimension *novelty* ($F = 1.69$; $p = 0.17$) *relevance* ($F = 0.44$; $p = 0.72$) and *market potential* ($F = 1.83$; $p = 0.14$) (Appendix 32). Overall, it can be concluded that the null hypothesis associated with *idea quality* can be rejected, although not all dimensions show statistical significance.

[12] It is worth noting that *familiarity with AI* is also significant with a p-value of $p < 0.05$. Familiarity with AI is separately analyzed in Section 5.4.2.

5.4.2 Hypothesis 2

The second hypothesis posits that human decision-makers with high familiarity with AI *decrease* the positive effect of hypothesis 1. Model 8 is based on a *linear regression* that captures the moderation effect through an interaction term between the treatment groups and the variable *familiarity with AI*. Additionally, covariates were taken into account in this model (Appendix 33).

The interaction term has a coefficient of -0.019, with a standard error of 0.01. The t-value for the interaction term is t = -1.73, with a corresponding p-value of p = 0.085, which indicates that the interaction, and consequently, hypothesis 2, is significant *only* for α = 10%. Figure 16 illustrates that the relationship between the treatment groups and WOA is contingent on the level of familiarity with AI. In alignment with hypothesis 2, individuals with greater familiarity with AI (+ 1 SD) tend to respond less to the treatment groups concerning the dependent variable WOA compared to those with lower familiarity (- 1 SD). In other words, the influence of enhanced transparency on the variable WOA is more pronounced among individuals with lower familiarity with AI compared to individuals with higher familiarity with AI.

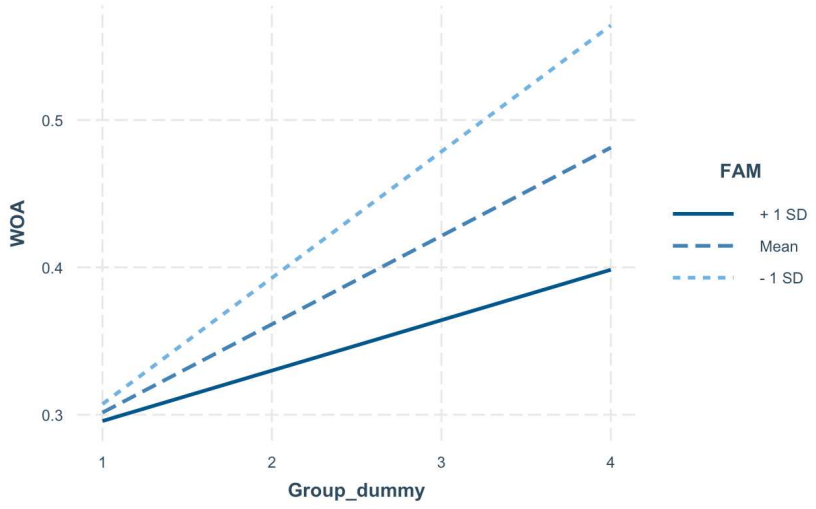

Figure 16: Moderation Effect of Familiarity with AI

In model 8, the treatment group variable, representing different degrees of transparency is highly significant, with p = 0.008. On average, the variable WOA increases by 0.1611 (16.11%) when comparing one treatment group to another, ceteris paribus. All other covariates, except

the dummy variable *male* with t = -1.88 and $p < 0.1$, are not significant. The overall model fit indicates a multiple R-squared value of 0.1028, suggesting that 10.28% of the variance of the variable WOA can be explained by the combined effects of the predictors in the model. The model as a whole is statistically significant ($F = 2.66$; $p < 0.001$).

Given that the significance level was set a priori at $\alpha = 5\%$, the null hypothesis (no moderation effect) cannot be rejected. However, the results indicate that the moderation effect is significant for $\alpha = 10\%$. Thus, it can be concluded that the effect of the treatment groups on WOA is only *moderately* influenced by the level of familiarity with AI, with an increased risk of type I errors.

5.4.3 Hypothesis 3

The third hypothesis posits that human decision-makers with high *confidence* in their own evaluation *decrease* the positive effect of hypothesis 1. Analogous to hypothesis 2, model 9 is based on a *linear regression* that captures the moderation effect through an interaction term between the treatment groups and the variable *confidence in own evaluation* (Appendix 34).

Model 9 has no significant interaction term (t = -0.355; p = 0.72). The relationship between the treatment groups and WOA is not contingent on the level of confidence in own evaluation (Figure 17). Neither the individual group variable (t = 1.27; p = 0.21) nor the variable confidence in own evaluation (t = 0.87; p = 0.38) show significance in this model. Only the variable familiarity with AI is significant with t = -2.49 and p = 0.01. Consequently, the model itself is significant; however, the null hypothesis (no moderation effect) cannot be rejected. Thus, it can be concluded that the effect of the treatment groups on WOA is not influenced by the level of confidence in own evaluation.

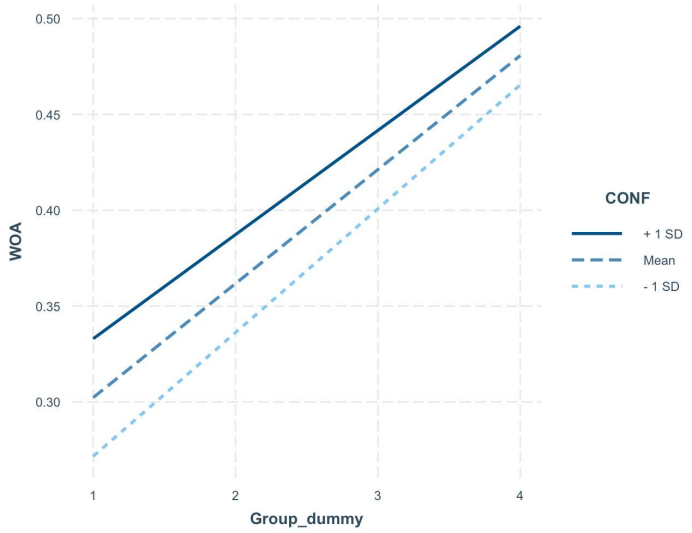

Figure 17: Moderation Effect of Confidence in Own Evaluation

In the context of an exploratory analysis, the variations in confidence ratings within and between the two conditions *before* and *after* the treatment were examined. Figure 18 illustrates the mean values of the confidence ratings before and after the disclosure of AI advice across all treatment groups. The t-tests revealed that for each treatment group, the confidence rating *after* the disclosure of AI advice was significantly higher compared to the confidence rating *before* the disclosure of AI advice (all $p < 0.05$) (Appendix 35). Initially, both conditions showed an increase in the confidence rating up to treatment group 3, followed by a decline in group 4. However, the ANOVA results for both conditions (Model 10 for pre-treatment and Model 11 for post-treatment) did not provide sufficient evidence to reject the null hypothesis (no significant difference (increase) in the mean values of confidence across the four groups) (Appendix 36). Therefore, based on the p-values ($p > 0.05$), we cannot conclude that the treatment groups have a significant effect on the confidence rating.[13]

[13] Since no manipulation occurred during the condition before treatment, the focus here lies on the condition *after* treatment.

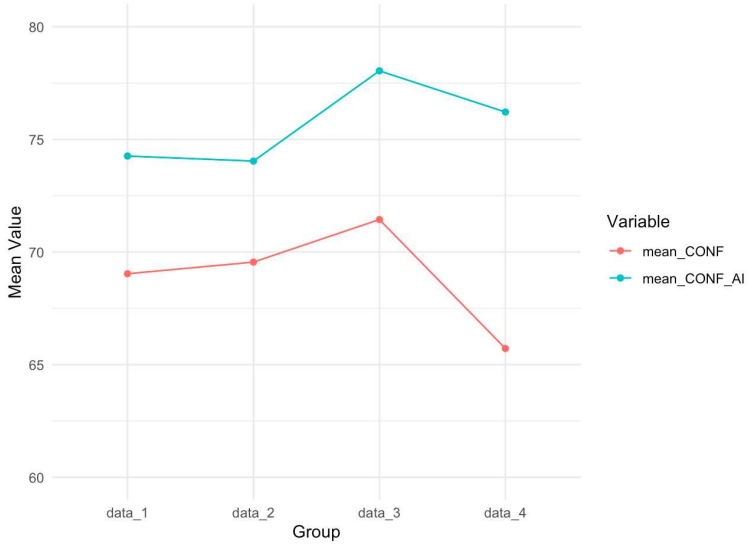

Figure 18: Comparison of Confidence Ratings (Mean Values from 0-100) Before (red) and After (green) Disclosure of AI Advice across Treatment Groups

Although the variable *confidence in own evaluation* does not have a significant impact on the dependent variable WOA, the exploratory analysis revealed that the level of confidence for each treatment group is significantly higher after the disclosure of AI advice compared to before the AI advice.

6 Discussion

In this Section, the obtained results will first be elaborated and interpreted in light of the existing literature (Section 6.1). Furthermore, the theoretical and managerial implications will be outlined (Section 6.2) before reflecting on the limitations of this thesis (Section 6.3). Lastly, potential future research avenues will be explored in Section 6.4, elucidating promising directions that can further advance knowledge in this domain.

6.1 Research Findings

6.1.1 Hypothesis 1

This thesis examines the role of transparency in AI systems on human idea evaluation processes. The findings provide compelling evidence, indicating a significant positive relationship between the degree of transparency in AI-based idea evaluation systems and the utilization of AI advice by human decision-makers. This aligns with the findings of Chao et al. (2016) and Yeomans et al. (2019). Even in the condition of no algorithm transparency (treatment group 1), participants demonstrated algorithm appreciation, which is consistent with the results of Logg et al. (2019), who posit that humans utilize advice from AI systems despite lacking a comprehensive understanding of the underlying processes.

However, this pertains solely to the generalized construct of idea quality. While all mean values of the variable WOA for the underlying dimensions of idea quality exhibit continuous increments with rising transparency, only feasibility emerged as statistically significant. Consequently, transparency manifests distinct effects on advice utilization depending on the dimensions of idea quality. A potential explanation for the insignificance of the other dimensions could be that the WOA values for the dimensions are not normally distributed.

A considerable number of participants completely ignored AI advice (WOA = 0%) or completely relied on it (WOA = 100%). This is consistent with the research of Parasuraman and Riley (1997), who assert that human judges tend to either underutilize or excessively depend on automated advice. The histograms (Appendix 23-26) illustrate this tendency. The

substantial variability within each group, relative to the differences between the means, could lead to the insignificant outcomes for the underlying dimensions of idea quality [14].

Although the number of participants completely ignoring AI advice decreases with increasing transparency, it remains relatively high, indicating that the support for hypothesis 1 may not apply to all participants uniformly. This implies that even with high transparency, some participants remain completely averse to AI advice. Thus, it becomes imperative for future research to identify additional moderators that elucidate the characteristics or conditions favoring algorithm aversion. This highlights that algorithm appreciation and algorithm aversion do not represent binary oppositions but instead form a complex continuum, necessitating further exploration of their predictors.

A substantial body of advice-taking research has consistently demonstrated that individuals who aim to maximize the accuracy of their judgments should average their own judgment with the advice received from a randomly selected individual (*WOA* = 50%) (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975). However, research showed that, on average, individuals only adjust 30-35% towards the advice they receive (Liberman et al., 2012; Logg et al., 2019; Soll & Larrick, 2009). The findings of this study revealed that individuals adjusted their judgment of idea quality by 38.43% in treatment group 1, slightly exceeding the 30-35% range observed in advice-taking research. As transparency increases, there is a gradual rise in the value of WOA, reaching 50.60% in treatment group 4 (Appendix 22), which corresponds to the level associated with maximum accuracy in judgments. Drawing on the research of Dawes and Corrigan (1974) and Einhorn and Hogarth (1975), the observed positive relationship between transparency and the degree to which individuals incorporate advice into their judgments consequently results in more informed and accurate decision-making processes. Providing transparent advice from AI systems emerges as an effective approach to enhance AI advice utilization.

[14] By averaging and taking into account all dimensions when calculating the *idea quality*, the variability within each group is reduced compared to the differences between the means.

6.1.2 Hypothesis 2

The results of the second hypothesis revealed that the positive effect of transparency on advice utilization is only moderately influenced by the level of familiarity with AI (significant for α = 10%). This is consistent with experiment 4 by Logg et al. (2019, p. 98), who demonstrated that "participants who claimed greater familiarity with algorithms took less advice". Thus, participants with higher self-reported familiarity with AI exhibited a stronger inclination to adhere to their initial evaluations and displayed a greater tendency to disregard the disclosed information. This could be explained by the fact that individuals with greater familiarity with AI possess better capabilities to process and comprehend AI advice, thus requiring less information to perceive a system as transparent. Consequently, each additional information has a more substantial positive impact on individuals with lower familiarity with AI. Simultaneously, individuals with greater familiarity with AI can leverage their expertise to detect operational boundaries, limitations, and biases of AI systems more effectively. This is consistent with Li (2017), who states that people with expertise have an informational advantage in evaluating ideas. This heightened awareness and discernment may lead to a more cautious approach in embracing AI advice, potentially resulting in a decreased utilization of AI advice. The exact reasons for this need to be clarified through further research.

6.1.3 Hypothesis 3

The results of hypothesis 3 surprisingly demonstrated that the effect of transparency in AI systems on advice utilization is *not* moderated by the level of confidence in their own evaluation. Neither the moderation effect nor the variable confidence per se are significant in the linear regression model. This is in contrast to the robust findings of Lewandowsky et al. (2000), Grimaldi et al. (2015), and Logg et al. (2019), who acknowledged that confidence judges have in their judgement is an essential factor that (negatively) influences advice utilization within the JAS paradigm. The reasons for this discrepancy could be attributed to various factors, such as the specific characteristics of the participants in the study, the methodology employed or the specific entrepreneurial context of the study. Further investigation and analysis are required to determine the exact reasons for the differing results from previous research.

In addition to the insignificant moderation effect, the exploratory analysis did not reveal any significant differences in the confidence rating between the treatment groups. This indicates

that participants in groups with higher transparency (more information) did not rate their confidence significantly higher than participants who received limited information about the AI system. However, within each treatment group, significant differences in confidence ratings were observed before and after the disclosure of AI advice. Participants rated their confidence significantly higher after the AI advice was disclosed, regardless of the degree of transparency. Participants thus became significantly more confident in their evaluations after the AI advice was disclosed. However, as tested with hypothesis 3, the higher confidence ratings after the disclosure of AI advice do not significantly influence advice utilization. This means that participants rated their confidence significantly higher after the AI advice was disclosed, regardless of whether and to what extent they adjusted their initial evaluation in favor of the AI advice.

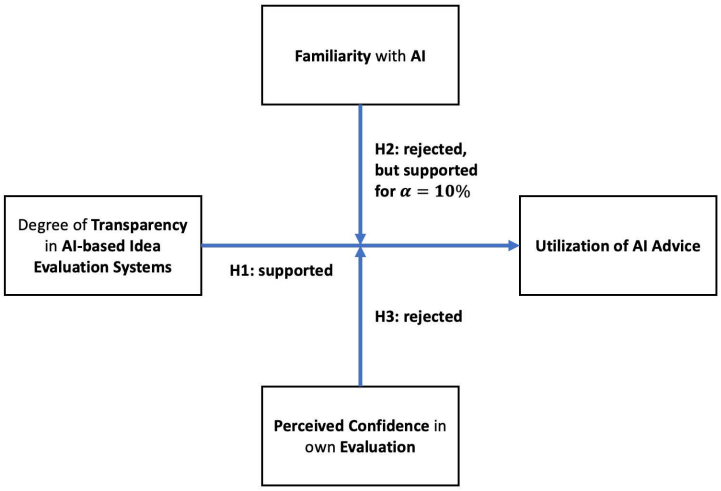

Figure 19: Summarized Results of the Hypotheses

6.2 Implications

6.2.1 Theoretical Implications

The results suggest that one way to enhance AI advice utilization in the context of idea evaluation is to increase the transparency of AI systems. In consonance with the argument of Yeomans et al. (2019), the absence of transparent explanations elucidating black-box algorithms thus accounts for why individuals exhibit apparent aversion to AI systems. Nevertheless, it is important to acknowledge that the positive influence of transparency on AI

advice utilization only represents a general tendency. Certain individuals completely disregard AI advice, even in conditions with high transparency, while others completely rely on AI advice in conditions with no transparency. This underlines the complex nature of algorithm aversion and algorithm appreciation and implies that these notions are not binary opposites and coexist in society.

The results are consistent with prior research with respect to the role of familiarity with AI (Logg et al., 2019), yet they challenge the commonly accepted notion regarding the moderating role of confidence (Lewandowsky et al., 2000; Logg et al., 2019). These results open avenues for further investigation into the complex interaction between familiarity, confidence, and other potential mediating or moderating factors. Future research could explore the mechanisms underlying these relationships, exploring how they vary across different contexts, user groups, and types of AI advice.

6.2.2 Managerial Implications

For decades, researchers have been trying to find ways to increase advice utilization and reduce the effects of biases in decision-making. Drawing on the research of Dawes and Corrigan (1974) and Einhorn and Hogarth (1975), the positive effect of transparency on AI-advice utilization results in more informed and accurate decision-making processes (maximized accuracy of judgments with *WOA* = 50%). The findings underline the crucial role and value of transparency in AI systems for idea evaluation, as human decision-makers are more inclined to adopt AI advice. This suggests that transparency can act as a catalyst in promoting the acceptance and integration of AI technology in decision-making processes. Accordingly, the potential of AI, such as mitigating cognitive biases, can be leveraged through superior computational capabilities (see Section 2.2.4). For managers, this implies that utilizing information from AI evaluations, such as insights about markets, technology, or required resources, can reduce the uncertainty in the FFE faster and effectively while concurrently decreasing the cost of evaluation. Consequently, the results indicate that the *Collingridge Dilemma* in the FFE (Section 2.1.2) is mitigated, enabling companies to expedite decisions regarding the allocation of (financial) resources for the initiation of the development phase. This more efficient management of the FFE in innovation can directly contribute to the success of a new product or service (Dwyer & Mellor, 1991).

Leveraging these insights to develop evaluation support systems that harness the potential of AI while addressing human preferences for understandable and transparent processes represents a major challenge for managers. This is particularly notable because individuals have differing levels of familiarity with AI, which leads to variations in how they perceive AI output in terms of transparency. Managers must be aware that individuals may react differently to AI advice based on their familiarity with AI. Therefore, strategies should be developed to address this variation, tailoring communication and guidance to effectively accommodate both those familiar and less familiar with AI.

A major challenge for managers is to enable and educate all employees to critically assess the quality of AI advice to prevent automation bias (Lee & See, 2004) while concurrently integrating AI into the operational NPD processes. Failing to critically assess or independently verify AI advice through integrated processes could lead to the oversight of potential errors or biases due to an excessive overreliance on AI. Currently, AI evaluation systems are not autonomous decision-making systems and must, therefore, be optimally aligned with employees to harness the full potential of *hybrid intelligence* and foster the synergies between human and machine intelligence (Piller et al., 2023). Managers need to acknowledge and comprehend the inherent weaknesses and strengths of both human and machine intelligence and subsequently complement their strategies and decisions. This encompasses providing a substantial volume of high-quality data to the AI system and educating employees to effectively engage with machines. Considering the availability and low cost of AI advice, effectively managing hybrid intelligence could be a potential competitive advantage for companies (Piller et al., 2023).

In essence, the results imply that managers should create integrated NPD processes that empower employees to exploit the potential of *hybrid intelligence* while recognizing the complexity and diversity of human responses to AI. A balanced, informed approach to incorporating AI advice leads to increasingly effective and accurate idea evaluation processes.

6.3 Limitations

Like all empirical research, this thesis should be considered in light of its limitations.

First, the assumption of ANOVA pertaining to the normality of the dependent variable WOA is violated. Even though ANOVA exhibits a certain degree of robustness against violations of

the normality assumption (Blanca Mena et al., 2017), the violations nonetheless have implications for the validity and reliability of the results. Specifically, this could distort p-values and influence type I and type II errors, potentially leading to incorrect determinations about the significance of effects. In order to enhance the robustness of the analysis, the non-parametric Kruskal-Wallis test was additionally employed as an alternative approach. Nevertheless, it is important to carefully assess the impact of the violation on the results and interpret the findings with caution.

Second, the operationalizations of constructs might be susceptible to measurement errors, affecting the accuracy of interpretations. The problem of appropriate operationalization is a central topic of empirical research. In this thesis, the degree of *transparency* is operationalized by the number of distinct (sub)types of information transparency that are concurrently presented. However, the degree of transparency also depends on the scope of information within each (sub)type and on other types of transparency (e.g., material or transformational transparency introduced by Andrada et al. (2022)). For instance, a condition presenting four (sub)types of transparency simultaneously but with limited information per dimension could be perceived as equally transparent as a condition involving only two dimensions but with greater depth and completeness of information. Thus, the operationalization of transparency in this thesis might not fully capture all nuances of transparency. Furthermore, it was not examined whether all (sub)types of transparency exert an equal effect on AI advice utilization, implying that the sequence of disclosed(sub)types of transparency might play a role (*data transparency* & *XAI* vs. *XAI* & *algorithm transparency* for the second degree of transparency).

The operationalization of the construct of *idea quality* is based on the findings of Dean et al. (2006). However, for the operationalization of the construct of idea quality, only the dimensions were employed (e.g., *novelty*) without considering the underlying sub-dimensions (e.g., *originality* and *paradigm relatedness*) (Appendix 1). Although the sub-dimensions were referenced in the description of the dimensions within the experimental survey (Appendix 2), their separate measurement was not conducted, considering the limited scope of this thesis. Consequently, the findings may not provide a comprehensive understanding of the complete spectrum of idea quality as defined by the underlying sub-dimensions and limit the ability to capture the full complexity of the construct.

Although the constructs *confidence in own evaluation* and *familiarity with AI* demonstrate internal consistency through single-item measurements or tested scale reliability, respectively, it is crucial to acknowledge that they rely on *self-reported measures*. Since these measures lack external validation, potential response biases may have influenced these constructs, including social desirability bias or strategic responding. As a result, this may have introduced inaccuracies in measuring those variables, potentially affecting the overall reliability of these measures.

Third, due to the quantitative nature of the results, they might not capture the nuances of participants' thought processes and justifications behind their evaluations, potentially missing valuable insights. Consequently, the findings can only affirm that increased transparency in AI systems leads to higher advice utilization. However, they do not elucidate the underlying reasons or aspects of transparency that contribute to this effect.

Fourth, due to specific context and sample characteristics, the findings might not be *generalizable* to broader populations or real-world scenarios (limited external validity). While the focus of the thesis on entrepreneurs is valuable for understanding idea evaluation processes within an entrepreneurial context, it also introduces a certain degree of restriction regarding the diversity and representativeness of the sample. Entrepreneurs often possess distinctive characteristics, motivations, and risk-taking tendencies that distinguish them from individuals in other professional domains. As a result, their reactions to AI-generated advice and their behaviors in decision-making scenarios may not align with those of individuals in different contexts. Therefore, any conclusions drawn from this study's findings should be applied carefully when attempting to understand how transparency in AI systems influences advice utilization in domains outside of entrepreneurship.

Fifth, this thesis relies on participants' responses to *hypothetical scenarios*, which might not accurately reflect real-world idea evaluation processes. As a result, the idea evaluation process relies solely on descriptive characteristics, lacking the inclusion of direct experiences and situational factors. Thus, participants may struggle to fully immerse themselves in the scenarios, potentially leading to biased responses or artificial reactions. For instance, participants may be influenced by the social norms related to the use of AI systems, the user interface or the perceived expertise of the AI system.

6.4 Future Research Avenues

Given the complex and multifaceted nature of AI advice utilization, future research could mitigate the limitations mentioned above by reproducing this study, taking into account a range of (contextual) variables. For instance, the degree of *trust* individuals place in technology (AI) could be introduced as a mediating variable to test the hypothesis that increased AI advice utilization is not driven by transparency but rather by enhanced trust in the system instilled by transparency. Furthermore, it would be reasonable to control for the *complexity* of AI systems, the *strength* of the advice, and distinctions between *positive AI evaluations* (when the rating of AI advice is higher than one's own evaluation) and *negative AI evaluations* (when the rating of AI advice is lower than one's own evaluation). The construct transparency should be differentiated not only by the number of distinct (sub)types of transparency but also by the depth of information within each (sub)type (scope). Furthermore, an exploration could be undertaken to determine which combination of (sub)types of transparency led to the observed outcomes and whether all are significant. In order to migrate the limitation of generalizability and enhance the robustness of the conclusions, future research could consider replicating the study across various professional domains and demographic groups. For this purpose, real-world evaluations should be used, and self-reported measures should be externally validated. Additionally, the role of confidence in own evaluation should be critically reexamined, as the findings of this thesis contrast with the robust results of previous research. Finally, the effectiveness of different transparency approaches and long-term effects of transparency on evaluation outcomes could also be explored.

The reflection of this thesis underlines the need for more in-depth and critical research on the subjective perception of transparency in AI systems. Future work should focus on qualitative research to gain nuanced insights into the underlying thought processes, justifications and mechanisms behind AI advice utilization. For example, future research could conduct interviews with focus groups to understand their subjective experiences and perceptions of transparency in AI advice utilization. Qualitative research can uncover the underlying emotions and concerns of AI-supported idea evaluation processes. Special emphasis should be placed on elucidating the factors that make transparency trustworthy and meaningful from the user's perspective in order to design effective user experiences with AI systems. This is important

because recent research indicates that transparency in AI systems can improve the output but at the expense of enjoyable usability (Schaffer et al., 2015).

Another avenue for future research involves conducting a direct comparative analysis of AI advice utilization across different sources, including AI advice vs. human advice vs. hybrid advice. This research could investigate whether advice with identical content yields varied levels of utilization depending on the source of delivery. The comparative approach could shed light on the nuanced interactions between different sources of advice and their impact on idea evaluation processes.

Furthermore, future research could investigate whether exceptionally high degrees of transparency migrate the positive effect on advice utilization and potentially even become negative. In other words, the disclosure of excessively detailed information could potentially amplify the negative effects associated with transparency in AI systems, such as information overload (Hosseini et al., 2018; Zhao et al., 2019), decreasing efficiency due to a longer time to comprehend the information (Tintarev & Masthoff, 2007), inherent biases within AI systems (Bertrand et al., 2022), or diminished trust resulting from an inability to fully grasp the information provided (Hosseini et al., 2018). This may result in individuals returning to a greater reliance on their own evaluations. The research could assess the existence of a nonlinear relationship, specifically an inverted U-shaped curve, between transparency and AI advice utilization. This investigation may reveal an optimal degree of transparency, representing the peak of advice utilization, where the positive effects are maximized before the negative consequences begin to outweigh the benefits.

7 Conclusion

The rapid technological advancements and investments in AI are reshaping the innovation processes of companies, offering a promising alternative (or complement) to human intelligence. AI can help entrepreneurs identify market trends and customer needs, assess potential risks and benefits of new ideas and consequently evaluate and prioritize new ideas fast and cost efficient. Since AI systems in the context of entrepreneurship are not (yet) autonomous decision-making systems, the focus in the future will be primarily on creating collaborative relationships between humans and machines. Effective and meaningful interactions between humans and machines aim to mitigate respective biases and harness their strengths to synergistically combine intelligence (Piller et al., 2023). Understanding this interaction, particularly how entrepreneurs utilize information from AI systems, is of central importance to identify and exploit the opportunities presented by technological advances (Logg et al., 2019). However, prior research has predominantly focused on the output of AI systems while paying comparatively less attention to how individuals respond to AI advice.

The results of this thesis demonstrate that increasing transparency in AI systems enhances the extent to which entrepreneurs integrate information into their evaluation processes. Moreover, the condition characterized by the highest degree of transparency has yielded an optimal level of accuracy in judgments (*WOA* = 50%), as stated by Dawes and Corrigan (1974). For the innovation process, accurate evaluations signify the ability to minimize uncertainty in the FEE, thereby reducing the likelihood of costly type I or type II errors in subsequent stages of the innovation funnel. However, as some participants, even in the high transparency condition, completely ignored the AI advice, future research should explore additional factors that determine AI advice utilization. Managers should empower employees to critically evaluate the quality of AI advice and mitigate the risks of automation bias and overreliance. Transparency is important for understanding and evaluating AI output but it should not kill usability (Arrieta et al., 2020).

The potential of AI in the innovation process is enormous and, at the same time, poses major challenges for companies. It is becoming increasingly important for companies to develop competencies to use AI properly and to recognize possible biases of the systems. The competence to use AI reasonably could become a critical factor in the competitiveness of companies. Thereby, innovation processes can be adapted to the technology in such a way that

the full potential and benefits of transparent AI can be effectively exploited while recognizing its limitations.

Appendices

Dropbox Link

The precise regulations outlined by Universidade Católica Portuguesa, prescribe a maximum limit of 30 pages for the appendices within the final document. Considering the broader scope required by WU and the detailed quantitative nature of my thesis, my appendices exceed the specified limit of 30 pages. Therefore, I have included the appendices in a designated Dropbox folder, in accordance with the guidance provided by Católica.

You may access the *List of Appendices* and the *Appendices* through the following link:

https://www.dropbox.com/scl/fo/qoq3on0j08rvxn4zifnaf/h?rlkey=smdszr2uholdu0pwxnksxawko&dl=0

For questions, please contact *nicolai.hessing@gmx.de*

Bibliography

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.

Amabile, T. M. (2020). Creativity, artificial intelligence, and a world of surprises. *Academy of Management Discoveries*, *6*(3), 351-354.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society*, *20*(3), 973-989.

Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. *AI & society*, 1-11.

Arnold, V., Clark, N., Collier, P. A., Leech, S. A., & Sutton, S. G. (2006). The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *Mis Quarterly*, 79-97.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, 82-115.

Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2022). A meta-analysis of the weight of advice in decision-making. *Current Psychology*, 1-26.

Barczak, G., Griffin, A., & Kahn, K. B. (2009). Perspective: trends and drivers of success in NPD practices: results of the 2003 PDMA best practices study. *Journal of product innovation management*, *26*(1), 3-23.

Basadur, M. (1995). Optimal ideation-evaluation ratios. *Creativity research journal*, *8*(1), 63-75.

Beerbaum, D., & Puaschunder, J. M. (2019). A Behavioral Approach to Irrational Exuberances–An Artificial Intelligence Roboethics Taxonomy. *Scientia Moralitas-International Journal of Multidisciplinary Research*, *4*(1), 1-30.

Belanche, D., Casaló, L. V., & Flavián, C. (2019). Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers. *Industrial Management & Data Systems*.

Bertino, E. (2020). The quest for data transparency. *IEEE Security & Privacy*, *18*(3), 67-68.

Bertino, E., Merrill, S., Nesen, A., & Utz, C. (2019). Redefining data transparency: A multidimensional approach. *Computer*, *52*(1), 16-26.

Bertrand, A., Belloum, R., Eagan, J. R., & Maxwell, W. (2022). How cognitive biases affect XAI-assisted decision-making: A systematic review. *In Proceedings of the 2022 AAAI/ACM conference on AI, ethics, and society*, 78-91.

Besemer, S. P., & Treffinger, D. J. (1981). Analysis of creative products: review and synthesis. *The Journal of Creative Behavior*.

Bink, M. L., & Marsh, R. L. (2000). Cognitive regularities in creative activity. *Review of General Psychology*, *4*(1), 59-78.

Blair, C. S., & Mumford, M. D. (2007). Errors in idea evaluation: Preference for the unoriginal? *The Journal of Creative Behavior*, *41*(3), 197-222.

Blanca Mena, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 2017, vol. 29, num. 4, p. 552-557*.

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.

Boone, L. E., Kurtz, D. L., & Berston, S. (2019). *Contemporary business*. John Wiley & Sons.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological bulletin*, *86*(2), 307.

Briggs, R. O., Reinig, B. A., Shepherd, M. M., Yen, J., & Nunameker, J. (1997). Quality as a function of quantity in electronic brainstorming. *Proceedings of the thirtieth Hawaii international conference on System sciences*, *2*, 94-103.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, *3*(1), 2053951715622512.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), 832.

Chao, C.-Y., Chang, T.-C., Wu, H.-C., Lin, Y.-S., & Chen, P.-C. (2016). The interrelationship between intelligent agents' characteristics and users' intention in a search engine by making beliefs and perceived risks mediators. *Computers in Human Behavior*, *64*, 117-125.

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of computer Science*, *7*, 157-164.

Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, *31*(1), 100698.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda* (pp. 115-146). University of Chicago Press.

Collingridge, D. (1982). The social control of technology.

Connolly, T., Routhieaux, R. L., & Schneider, S. K. (1993). On the effectiveness of group brainstorming: Test of one underlying cognitive mechanism. *Small Group Research*, *24*(4), 490-503.

Connor, S., Li, T., Roberts, R., Thakkar, S., Liu, Z., & Tong, W. (2022). Adaptability of AI for safety evaluation in regulatory science: A case study of drug-induced liver injury [Review]. *Frontiers in Artificial Intelligence*, *5*.

Cooper, R. G. (1990). Stage-gate systems: a new tool for managing new products. *Business horizons*, *33*(3), 44-54.

Cooper, R. G., & Kleinschmidt, E. J. (1996). Winning businesses in product development: The critical success factors. *Research-technology management*, *39*(4), 18-29.

Cossette, P. (2014). Heuristics and cognitive biases in entrepreneurs: a review of the research. *Journal of Small Business & Entrepreneurship*, *27*(5), 471-496.

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal*, *60*(2), 433-460.

Cummings, M. L. (2006). Automation and accountability in decision support system interface design.

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of management review*, *32*(1), 33-54.

Daneshjou, R., Smith, M. P., Sun, M. D., Rotemberg, V., & Zou, J. (2021). Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. *JAMA dermatology*, *157*(11), 1362-1369.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*, 24-42.

David, M. (2002). The correspondence theory of truth.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological bulletin*, *81*(2), 95.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668-1674.

Dean, D. L., Hender, J., Rodgers, T., & Santanen, E. (2006). Identifying good ideas: constructs and scales for idea evaluation. *Journal of Association for Information Systems*, *7*(10), 646-699.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56-62.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of experimental psychology: General*, *144*(1), 114.

Dijkstra, J. J., Liebrand, W. B., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*, *17*(3), 155-163.

Driva, H., Pawar, K., & Menon, U. (2000). Measuring product development performance in manufacturing organisations. *International Journal of Production Economics*, *63*(2), 147-159.

Drucker, P. (2014). *Innovation and entrepreneurship*. Routledge.

Du Preez, N. D., & Louw, L. (2008). *A framework for managing the innovation process* PICMET '08 - 2008 Portland International Conference on Management of Engineering & Technology, 546-558.

Dwyer, L., & Mellor, R. (1991). Organizational environment, new product process activities, and project outcomes. *Journal of product innovation management*, *8*(1), 39-48.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human factors*, *44*(1), 79-94.

Editorials, N. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, *613*(612), 10.1038.

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in ai systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-19.

Eidelman, S., Crandall, C. S., & Pattershall, J. (2009). The existence bias. *Journal of Personality and Social Psychology*, *97*(5), 765.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational behavior and human performance*, *13*(2), 171-192.

Eling, K., Langerak, F., & Griffin, A. (2015). The performance effects of combining rationality and intuition in making early new product idea evaluation decisions. *Creativity and Innovation Management*, *24*(3), 464-477.

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Foufou, S., & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, *2*(3), 267-279.

Ferràs-Hernández, X. (2018). The future of management in a world of electronic brains. *Journal of Management Inquiry*, *27*(2), 260-263.

Franken, R., Heringa, M. B., Oosterwijk, T., Dal Maso, M., Fransman, W., Kanerva, T., Liguori, B., Poikkimäki, M., Rodriguez-Llopis, I., & Säämänen, A. (2020). Ranking of human risk assessment models for manufactured nanomaterials along the Cooper stage-gate innovation funnel using stakeholder criteria. *NanoImpact*, *17*, 100191.

Frantz, R. (2003). Herbert Simon. Artificial intelligence as a framework for understanding intuition. *Journal of Economic Psychology*, *24*(2), 265-277.

Frederiksen, M. H., & Knudsen, M. P. (2017). From creative ideas to innovation performance: The role of assessment criteria. *Creativity and Innovation Management*, *26*(1), 60-74.

G. Harris, C. (2020). Mitigating cognitive biases in machine learning algorithms for decision making. *Companion Proceedings of the Web Conference 2020*, 775-781.

George, J. F., Duffy, K., & Ahuja, M. (2000). Countering the anchoring and adjustment bias with decision support systems. *Decision Support Systems*, *29*(2), 195-206.

Geroski, P., Machin, S., & Van Reenen, J. (1993). The profitability of innovating firms. *The RAND Journal of Economics*, 198-211.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, 80-89.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21-35.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660.

Griffin, A., & Page, A. L. (1993). An interim report on measuring product development success and failure. *Journal of product innovation management*, *10*(4), 291-308.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, 501-522.

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews*, *55*, 88-97.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment*, *12*(1), 19.

Gudykunst, W. B. (1993). Toward a theory of effective interpersonal and intergroup communication: an anxiety/uncertainty management (AUM) perspective.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, *51*(5), 1-42.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE intelligent systems*, *24*(2), 8-12.

Harvey, N. (1997). Confidence in judgment. *Trends in cognitive sciences*, *1*(2), 78-82.

Hinshaw, S. P. (2007). Moderators and mediators of treatment outcome for youth with ADHD: Understanding for whom and how interventions work. *Journal of pediatric psychology*, *32*(6), 664-675.

Hosseini, M., Shahri, A., Phalp, K., & Ali, R. (2018). Four reference models for transparency requirements in information systems. *Requirements Engineering*, *23*, 251-275.

Hwang, A. H.-C., & Won, A. S. (2021). IdeaBot: investigating social facilitation in human-machine team creativity. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-16.

Idrees, S. M., Alam, M. A., & Agarwal, P. (2019). A study of big data and its challenges. *International Journal of Information Technology*, *11*, 841-846.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business horizons*, *61*(4), 577-586.

Kim, J., & Wilemon, D. (2002). Focusing the fuzzy front–end in new product development. *R&d Management*, *32*(4), 269-279.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, *32*, 385-418.

Knudsen, T., & Levinthal, D. A. (2007). Two faces of search: Alternative generation and alternative evaluation. *Organization Science*, *18*(1), 39-54.

Kornish, L. J., & Ulrich, K. T. (2014). The importance of the raw idea in innovation: Testing the sow's ear hypothesis. *Journal of marketing research*, *51*(1), 14-26.

Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.

Krupnikov, Y., & Findley, B. (2015). Survey experiments: Managing the methodological costs and benefits.

Kuipers, B., Moskowitz, A. J., & Kassirer, J. P. (1988). Critical decisions under uncertainty: Representation and structure. *Cognitive Science*, *12*(2), 177-210.

Kuner, C., Cate, F. H., Lynskey, O., Millard, C., Ni Loideain, N., & Svantesson, D. J. B. (2018). Expanding the artificial intelligence-data protection debate. In (Vol. 8, pp. 289-292): Oxford University Press.

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2).

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80.

Lewandowsky, S., Mundy, M., & Tan, G. (2000). The dynamics of trust: comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104.

Li, D. (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics*, *9*(2), 60-92.

Li, T., Tong, W., Roberts, R., Liu, Z., & Thakkar, S. (2020). DeepDILI: deep learning-powered drug-induced liver injury prediction using model-level representation. *Chemical research in toxicology*, *34*(2), 550-565.

Liberman, V., Minson, J. A., Bryan, C. J., & Ross, L. (2012). Naïve realism and capturing the "wisdom of dyads". *Journal of Experimental Social Psychology*, *48*(2), 507-512.

Licuanan, B. F., Dailey, L. R., & Mumford, M. D. (2007). Idea evaluation: Error in evaluating highly original ideas. *The Journal of Creative Behavior*, *41*(1), 1-27.

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.

Lonergan, D. C., Scott, G. M., & Mumford, M. D. (2004). Evaluative aspects of creative thought: Effects of appraisal and revision standards. *Creativity research journal*, *16*(2-3), 231-246.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Luttrell, A., Briñol, P., Petty, R. E., Cunningham, W., & Díaz, D. (2013). Metacognitive confidence: A neuroscience approach. *Revista de Psicología Social*, *28*(3), 317-332.

Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association*, *24*(2), 423-431.

MacKenzie, S. B., Lutz, R. J., & Belch, G. E. (1986). The role of attitude toward the ad as a mediator of advertising effectiveness: A test of competing explanations. *Journal of marketing research*, *23*(2), 130-143.

Maher, M. L., & Fisher, D. H. (2012). Using AI to evaluate creative designs. *DS 73-1 Proceedings of the 2nd International Conference on Design Creativity Volume 1*.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, *90*, 46-60.

Malhotra, N., Nunan, D., & Birks, D. (2017). *Marketing research: An applied approach*. Pearson.

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors*, *58*(3), 401-415.

Metzger, O., & Spengler, T. (2019). Modeling rational decisions in ambiguous situations: a multi-valued logic approach. *Business Research*, *12*(1), 271-290.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325-1338.

Mirowski, P. (2023). The glassy essence of transparency. In: Springer.

Moenaert, R. K., De Meyer, A., Souder, W. E., & Deschoolmeester, D. (1995). R&D/marketing communication during the fuzzy front-end. *IEEE transactions on Engineering Management*, *42*(3), 243-258.

Montibeller, G., & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk analysis*, *35*(7), 1230-1251.

Mueller, J. S., Melwani, S., & Goncalo, J. A. (2012). The bias against creativity: Why people desire but reject creative ideas. *Psychological science*, *23*(1), 13-17.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European journal of social psychology*, *22*(2), 103-122.

Mumford, M. D., Lonergan, D. C., & Scott, G. (2002). Evaluating creative ideas: Processes, standards, and context. *Inquiry: Critical thinking across the disciplines*, *22*(1), 21-30.

O'Sullivan, D., & Dooley, L. (2008). *Applying innovation*. Sage publications.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230-253.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug discovery today*, *26*(1), 80.

Pieters, W. (2011). Explanation and trust: what to tell the user in security and AI? *Ethics and information technology*, *13*, 53-64.

Piller, F. T., Bouschery, S. G., & Blazevic, V. (2023). Hybrid Intelligence for Innovation: Augmenting NPD Teams with Artificial Intelligence and Machine Learning. *The PDMA Handbook of Innovation and New Product Development*, 407.

Rebernik, M., Bradač, B., Rebernik, M., & Bradač, B. (2008). Idea evaluation methods and techniques. *Institute for Entrepreneurship and Small Business Management, University of Maribor, Slovenia*, *27*.

Regalado, A. (2013). Big data gets personal. In (Vol. 116, pp. 63-64): MIT Technology Review.

Reitzig, M., & Sorenson, O. (2013). Biases in the selection stage of bottom-up strategy formulation. *Strategic Management Journal*, *34*(7), 782-799.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). " Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135-1144.

Ribera, M., & Lapedriza, A. (2019). Can we do better explanations? A proposal of user-centered explainable AI. *IUI workshops*, *2327*, 38.

Robertson, T. S. (1971). *Innovative behavior and communication*. Holt McDougal.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (2019). *Explainable AI: interpreting, explaining and visualizing deep learning* (Vol. 11700). Springer Nature.

Schaffer, J., Giridhar, P., Jones, D., Höllerer, T., Abdelzaher, T., & O'donovan, J. (2015). Getting the message? A study of explanation interfaces for microblog data analysis. *Proceedings of the 20th international conference on intelligent user interfaces*, 345-356.

Schemmer, M., Hemmer, P., Kühl, N., Benz, C., & Satzger, G. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. *arXiv preprint arXiv:2204.06916*.

Schweitzer, F. M., Buchinger, W., Gassmann, O., & Obrist, M. (2012). Crowdsourcing: Leveraging innovation through online idea competitions. *Research-technology management*, *55*(3), 32-38.

Sides, R., Marsh, M., Goldberg, R., & Mangold, M. (2019). Consumer Privacy in Retail: The Next Regulatory and Competitive Frontier. *Deloitte Development,, LLC*.

Siemon, D. (2022). Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration. *Behaviour & Information Technology*, 1-19.

Sinclair, M., & Ashkanasy, N. M. (2005). Intuition: Myth or a decision-making tool? *Management learning*, *36*(3), 353-370.

Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. *CHI'02 extended abstracts on Human factors in computing systems*, 830-831.

Sinha, R. R., & Swearingen, K. (2001). Comparing recommendations made by online systems and friends. *DELOS*, *106*.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*(2), 159-174.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of experimental psychology: Learning, memory, and cognition*, *35*(3), 780.

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, *27*(1), 81-102.

Soukhoroukova, A., Spann, M., & Skiera, B. (2012). Sourcing, filtering, and evaluating new product ideas: An empirical exploration of the performance of idea markets. *Journal of product innovation management*, *29*(1), 100-112.

Statista. (2021). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes)*. Statista.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103.

Tauchert, C., & Mesbah, N. (2019). Following the Robot? Investigating Users' Utilization of Advice from Robo-Advisors. ICIS,

Thomke, S., & Fujimoto, T. (2000). The effect of "front-loading" problem-solving on product development performance. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, *17*(2), 128-142.

Tintarev, N., & Masthoff, J. (2007). A survey of explanations in recommender systems. *2007 IEEE 23rd international conference on data engineering workshop*, 801-810.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124-1131.

van der Aalst, W. M. (2021). Hybrid Intelligence: to automate or not to automate, that is the question. *International Journal of Information Systems and Project Management*, *9*(2), 5-20.

van der Aalst, W. M., Hinz, O., & Weinhardt, C. (2021). Resilient digital twins: organizations need to prepare for the unexpected. In (pp. 1-5): Springer.

Van der Panne, G., Van Beers, C., & Kleinknecht, A. (2003). Success and failure of innovation: a literature review. *International journal of innovation Management*, *7*(03), 309-338.

Velamuri, V. K., Schneckenberg, D., Haller, J. B., & Moeslein, K. M. (2017). Open evaluation of new product concepts at the front end of innovation: objectives and contingency factors. *R&d Management*, *47*(4), 501-521.

Venkatesh, V., Thong, J. Y., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the association for Information Systems*, *17*(5), 328-376.

Vincent, V. U. (2021). Integrating intuition and artificial intelligence in organizational decision-making. *Business horizons*, *64*(4), 425-438.

Woolley, K., & Risen, J. L. (2018). Closing your eyes to follow your heart: Avoiding information to protect a strong intuitive preference. *Journal of Personality and Social Psychology*, *114*(2), 230.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260-281.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32*(4), 403-414.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science*, *10*(6), 224-228.

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019). Do users always want to know more? Investigating the relationship between system transparency and user´s trust in advice-giving systems. *In Proceedings of the 27th European Conference on Information Systems (ECIS)* Research-in-Progress Papers.

Zouave, E. T., & Marquenie, T. (2017). An inconvenient truth: algorithmic transparency & accountability in criminal intelligence profiling. *2017 European Intelligence and Security Informatics Conference (EISIC)*, 17-23.