universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Empirically understanding high-risk AI: A case study of an AI system that performs matchmaking to connect freelancers with companies to develop projects

verfasst von | submitted by

Ingrid Paola Marin Cabezas

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 906
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Science-Technology-Society
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Maximilian Fochler

Acknowledgements

I want to express my gratitude to the startup for its willingness to grant access to documents and, most importantly to the opportunity to interview team members. I learned so much about AI.

Thank you to my supervisor, Max Fochler, who provides essential advice that help me to continue my research during periods of uncertainty.

Thank you most of all to my family and friends for their support and enthusiasm in this research that constitutes not only a master´s thesis, but also the culmination of a dream and a significant phase of my life.

Special thanks to Marby Duarte, who facilitated the contact with the startup that I finally did my research with. Without your collaboration, the topic of this research most probably be completely different.

Abstract (English)

This thesis embarks on an empirical exploration of high-risk Artificial Intelligence (AI), specifically examining its development within the context of Small and Medium-sized Enterprises (SMEs). The motivation for this research arose in 2022 amidst discussions and the proposed regulation of AI in the European Union, which introduced risk levels of AI: unacceptable, high-risk, limited risk, and minimal risk. Of particular interest is the "high-risk" category, subject to stringent obligations. This thesis positions itself not to analyse the categorisation, but rather to empirically understand high-risk AI through a real-world case study. The overarching question guiding this research is *How do startups develop what would be most likely high-risk AI?*

The selected case focuses on a startup that has developed an AI system for matchmaking freelancers with companies for specific projects. Given the final version of the EU Artificial Intelligence Act (EU AI Act), published on June 13, 2024, specifically Annex III High-risk AI systems, which addresses “Employment, workers’ management and access to self-employment” this startup's AI system is likely to be classified as high-risk.

The empirical findings highlight several aspects of high-risk AI development. For instance, AI products not inherently grounded in widespread societal needs can nonetheless transition into high-risk AI. Moreover, the research reveals that the startup actively shapes its AI and, consequently, public perceptions surrounding it. The vision of the startup about reinventing work creates a powerful narrative that influences how the technology is conceptualized by investors, customers, and employees. This demonstrates the relevance of narratives embedded within AI systems, and their capacity to influence the realities of the freelance job market.

Furthermore, a significant impact identified relates to data, particularly the simplification of complex personal preferences and motivations into static information within the AI system. This simplification raises questions about whether AI-driven decisions truly reflect real-life complexities. The thesis also addresses the challenge of understanding and contesting AI decisions, especially for "black box" models, emphasizing the necessity of human involvement from diverse disciplines at various stages of high-risk AI production. Moreover, the research raises concerns about potential overreliance on AI recommendations and the need for future research on its impact on user decision-making and dependency.

The core finding of the research is concerned with the startup's valuation process of AI. Employing a valuation studies framework, the research reconstructs how the startup conceives high-risk AI through

the "triad of positions" -valuee, valuator, audience (Waibel et al., 2021). It demonstrates that the performance of the algorithm is the primary valuee, and surprisingly, the AI product itself, not the end-users, is the true audience for this valuation. The research reveals that the value attributed to high-risk AI, despite its risks, drives and shapes the startup's business goals, processes, and vision, significantly also shaping the freelance sector. This thesis attempts to contribute to reduce the "black box" nature of high-risk AI by providing an empirical account of its creation, function, and multifaceted impacts, advocating for a broader understanding of AI as a social construction shaped by, and shaping, society.

Abstract (German)

Diese Arbeit unternimmt eine empirische Untersuchung von risikoreicher künstlicher Intelligenz (KI) und untersucht insbesondere deren Entwicklung im Kontext kleiner und mittlerer Unternehmen (KMU). Die Motivation für diese Forschung entstand im Jahr 2022 im Zuge von Diskussionen und der vorgeschlagenen Regulierung von KI in der Europäischen Union, die Risikostufen für KI einführte: inakzeptabel, risikoreich, begrenztes Risiko und minimales Risiko. Von besonderem Interesse ist die Kategorie „hohes Risiko", die strengen Auflagen unterliegt. Diese Arbeit hat nicht zum Ziel, die Kategorisierung zu analysieren, sondern hochriskante KI anhand einer Fallstudie aus der Praxis empirisch zu verstehen. Die übergeordnete Frage, die diese Untersuchung leitet, lautet: *Wie entwickeln Start-ups KI, die mit hoher Wahrscheinlichkeit als hochriskant einzustufen ist?*

Der ausgewählte Fall konzentriert sich auf ein Start-up, das ein KI-System entwickelt hat, um Freiberufler für bestimmte Projekte an Unternehmen zu vermitteln. Angesichts der endgültigen Fassung des EU-Künstlicher-Intelligenz-Gesetzes (EU-KI-Gesetz), das am 13. Juni 2024 veröffentlicht wurde, insbesondere Anhang III „Hochriskante KI-Systeme", der sich mit „Beschäftigung, Arbeitnehmermanagement und Zugang zur Selbstständigkeit" befasst, wird das KI-System dieses Start-ups wahrscheinlich als hochriskant eingestuft werden.

Die empirischen Ergebnisse heben mehrere Aspekte der Entwicklung risikoreicher KI hervor. So können beispielsweise KI-Produkte, die nicht von vornherein auf weit verbreiteten gesellschaftlichen Bedürfnissen basieren, dennoch zu risikoreicher KI werden. Darüber hinaus zeigt die Untersuchung, dass das Start-up seine KI aktiv gestaltet und damit auch die öffentliche Wahrnehmung beeinflusst. Die Vision des Start-ups, die Arbeit neu zu erfinden, schafft eine starke Erzählung, die die Wahrnehmung der Technologie durch Investoren, Kunden und Mitarbeiter beeinflusst. Dies zeigt die Relevanz von Narrativen, die in KI-Systemen eingebettet sind, und ihre Fähigkeit, die Realitäten des Freiberufler-Arbeitsmarktes zu beeinflussen.

Darüber hinaus wurde ein signifikanter Einfluss im Zusammenhang mit Daten festgestellt, insbesondere die Vereinfachung komplexer persönlicher Präferenzen und Motivationen zu statischen Informationen innerhalb des KI-Systems. Diese Vereinfachung wirft Fragen auf, ob KI-gesteuerte Entscheidungen wirklich die Komplexität des realen Lebens widerspiegeln. Die Arbeit befasst sich auch mit der Herausforderung, KI-Entscheidungen zu verstehen und anzufechten, insbesondere bei „Black-Box"-Modellen, und betont die Notwendigkeit der Einbeziehung von Menschen aus verschiedenen Disziplinen in verschiedenen Phasen der Produktion von risikoreicher KI. Darüber hinaus wirft die

Untersuchung Bedenken hinsichtlich einer möglichen übermäßigen Abhängigkeit von KI-Empfehlungen und der Notwendigkeit weiterer Forschung zu deren Auswirkungen auf die Entscheidungsfindung und Abhängigkeit der Nutzer auf.

Die zentrale Erkenntnis der Untersuchung betrifft den Bewertungsprozess des Start-ups im Hinblick auf KI. Unter Verwendung eines Bewertungsstudien-Rahmenwerks rekonstruiert die Studie, wie das Start-up risikoreiche KI anhand der „Triade der Positionen“ – Wertgeber, Bewerter, Zielgruppe (Waibel et al., 2021) – konzipiert. Sie zeigt, dass die Leistung des Algorithmus der primäre Wertgeber ist und überraschenderweise nicht die Endnutzer, sondern das KI-Produkt selbst die eigentliche Zielgruppe dieser Bewertung ist. Die Studie zeigt, dass der Wert, der risikoreicher KI trotz ihrer Risiken beigemessen wird, die Geschäftsziele, Prozesse und Visionen des Start-ups vorantreibt und prägt und damit auch den Freelancer-Sektor maßgeblich beeinflusst. Diese Arbeit versucht, einen Beitrag zur Verringerung der „Black-Box“-Natur risikoreicher KI zu leisten, indem sie einen empirischen Überblick über ihre Entstehung, Funktion und vielfältigen Auswirkungen liefert und für ein breiteres Verständnis von KI als soziales Konstrukt eintritt, das von der Gesellschaft geprägt ist und diese wiederum prägt.

Table of contents

1. Introduction

Artificial Intelligence (AI) is considered as one of the defining technological advancements of the 21st century, permeating various aspects of daily life and reshaping the present and the future of the world. The AI areas encompass machine translation, computer vision, speech recognition, natural language processing, information retrieval and reasoning, and cognitive modelling and affective computing. Its potential to automate tasks, enhance decision-making, and create efficiencies is widely acknowledged. However, like all technologies AI is two-sided, its transformative power comes with a growing recognition of ethical, social, and economic implications, particularly concerning systems that operate in critical domains of the society. The discourse around "high-risk AI" has intensified, driven by concerns about fairness, transparency, accountability, and fundamental rights, and the efforts towards harness AI´s full potential and benefit as much as possible but taking into consideration its downsides and risks. In response to that, the European Union introduced the first regulatory framework of AI around the world EU AI Act.

This thesis delves into the practical realities of developing high-risk AI, moving beyond theoretical discussions to provide an empirical understanding of what is done by those at the forefront of AI creation: Small and Medium-sized Enterprises (SMEs). These organisations often drive innovation, yet their processes for developing AI systems, especially those with significant societal impact, remain underexplored. Understanding how they navigate the complexities of identifying societal needs, conceiving AI solutions, and their impact is crucial for raising awareness about high-risk AI.

The motivation to research about high-risk AI came from 2022, when there were discussions around the categorization deployed in the proposal of regulation in the European Union. Those categories encompass different levels of risk: unacceptable, high-risk, limited risk and minimal risk. Among those, high-risk level will be subject to strict obligations before the systems classified in this level can be put on the market. From that context, I decided to take as the entry point of the research considerations of high-risk AI provided in the official proposal of AI Act (version of 21.04.2021), specifically, the Explanatory Memorandum section 5.2.3. HIGH-RISK AI SYSTEMS (TITLE III) that states "**Title III** contains specific rules for AI systems that create a *high risk to the health and safety or fundamental rights of natural persons* [emphasis added]" (p.13).

Since this research takes as a case study one startup whose business scope is software development to connect companies and freelancers, I will consider the following area addressed in the final version of

the regulation published on 13th. of June 2024 - the EU Artificial Intelligence Act (EU AI Act)- Annex III High-risk AI systems referred to in Article 6(2)).

> 4. Employment, workers' management and access to self-employment:
>
> (a) AI systems intended to be used for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates;
>
> (b) AI systems intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and evaluate the performance and behaviour of persons in such relationships. (The European Parliament and The Council of The European Union, 2024)

The research is guided by a central overarching question which is, *How do startups develop what would be most likely high-risk AI?* To comprehensively address this, the following three sub-questions are considered, *How are problems, concerns, and needs of society identified to develop high-risk AI? How do startups conceive high-risk AI?* and *How what the startup pursues when developing high-risk AI impact the society?*

To explore these questions, this thesis adopts a qualitative case study approach, focusing on a Spanish startup that has developed an AI system for matchmaking freelancers with companies for specific projects. The relevance of this particular system is evidenced by the AI Specialist of the startup confirmed, who confirmed that it is most probably to be classified as high-risk under the EU AI Act, given its application in the employment and worker management context. Needed to mention that the research is not focused on analyse the categorisation of AI risks or the conditions stated in the regulation to consider what is high-risk AI, but to present a case study to understand or attempt to empirically understanding high-risk AI.

The thesis is structured into five chapters, each building upon the previous one to provide a comprehensive analysis. In Chapter 1, I present academy literature related to the topic, including dominant frames of AI, concepts and visions, as well as an overview of the EU AI Act. Chapter 2 is dedicated to the research questions, methods of the research and ethical considerations and limitations. Afterwards, I discuss the theoretical lens through which I analysed the empirical findings (Chapter 4). In Chapter 5, I dive deeper into the interviews and documents analysis, which the objective of providing and

analysing the empirical findings. I conclude my thesis synthetizing and discussing the findings and connecting those with the literature review and analytical framework, and summarizing the key contributions of the thesis.

2. Literature Review

Since artificial intelligence (AI) is researched from different perspectives, I will attempt to explore contributions made in this domain by Science and Technology Studies (STS) as well as by other disciplines. The literature review focuses on research related to key topics of the present Master´s thesis, such as AI discourses, use of AI by companies, risks, and societal implications. This section is an attempt to comprehensively describe the most important contributions that matter within those themes. Salient topics from the literature review will also be included. This considers providing, if possible, definitions to become acquainted with the terminology around AI. I will also take a look at topics in which STS particularly seems to be lacking.

2.1 Dominant framing of AI

In order to understand the dynamics around AI, studies of Ulnicane et al. focused on AI governance pointed out that it is important to clarify that AI is identified as an emerging technology because of its characteristics of new, fast growth, big impact, uncertainty and ambiguity. Besides, it is relevant to acknowledge the association of emerging technologies with hypes and expectations and the performative functions of those as "collectively pursued explorations of the future that affect activities in the present" (Ulnicane et al., 2021, p. 162) and shape emerging technologies.

In that regard, Science and Technology Studies (STS) have reflected on how the future is mobilised around technologies including AI. In Williams´ (2006) contributions to this topic, the author criticized the tendency of advocates and critics to project specific visions of high-tech futures encompassing prospects and implications. By mapping anticipations of technical and social impacts of technologies and attempting to foresight, the future is portrayed as imminent and nearly determined (Williams, 2006). Therefore, the distinction between imagined and actual futures is blurred. In this regard, S. Lindgren & Holmström (2020) make reference to Lagerkvist (2020), who examines how the self-presentation of AI constructs its future as inevitable. Likewise, the author argues that we are in a "digital limit situation" because imagined AI futures are presented as the unique solution to the issues of societies through forecasting and prediction. This effectively limit other possible futures to be considered.

Another central concern in STS is the way AI is framed as an autonomous entity rather than as a set of interwoven practices and power relations that scholars should analyse. For instance, Suchman (2023) criticizes what she calls the uncontroversial thingness of AI, arguing that the status and agency of this technology requires to be critically examined in light of the work of AI in specific context. This critique

aligns with broader STS perspectives that call to slow down and use ethnography, history, and description to examine current cases of AI, providing detailed empirical and critical insights (Jaton & Sormani, 2023). In the same line, S. Lindgren & Holmström (2020) proposed what they name 'four building blocks' to approach AI not only from the dominant framing of this technology as a shaper of society. Starting from analysing the interaction between humans and machines in a societal context. Then, identify and examining AI agents as social actors. Subsequently, addressing AI as a social construction framed and understood as a phenomenon with hopes and fears. Then, reflect on how the AI and datafication could affect social science research topics and methods.

In the same direction, other aspects addressed in literature encompass the vision of AI, the vision of society and the effects of this technology. S. Lindgren & Holmström (2020) underline that AI inherently carries social intentionality and must be recognized as a site of power. Sartori & Theodorou (2022) pointed out that this technology is organized around narratives of progress, automation, potential uses and projecting the future, but it is also entangled with traditional structures of social, economic, and political inequalities. Regarding the development and future effects of AI, Peeters et al. (2021) present three perspectives that mobilize debates. The technology-centric that points out that AI's potential will be greater than human capabilities; the human-centric perspective that focuses on ethical risks and social autonomy of human over AI; and the collective intelligence perspective that claims intelligence as a result of a collective group of humans and AI. Williams (2006) emphasizes that technology futures are portrayed through the promotion of expectations and revolutionary breakthroughs, and the economic and social benefits being promised to justify large-scale investments in technologies. Those constitute powerful visions that entangle anxieties and controversies over the risks of new technologies. In a complementary manner, the same author highlighted that society is in a paradoxical situation because advancement is imperative yet vulnerable. Concerns regarding the risks and ethical and social implications of technologies have the potential to obstruct acceptance and result in the failure of those advancements (Williams, 2006). In another contribution, scholars mentioned that "the scale at which AI is now distributed, multiplied, adapted, and vastly interconnected allows this technology to generate massive impact on society, at a rate that no longer allows for careful consideration of future consequences" (Peeters et al., 2021, p. 231).

In regard to the recent debates around artificial general intelligence (AGI) Gerdes (2021) discusses that overconfident predictions of AI, overpromising language, and assumptions of intelligence as individualistic reward maximization, are the arguments that mobilise the idea of AI reaches human – level intelligence. The author dismisses the idea of a foreseeable future approach of AGI, arguing, among other factors, intelligence requires tacit knowledge, that is acquired through socialization to form social intelligence, which is a hurdle to AGI.

2.2 Corporate vision of AI

Regarding the use of AI by companies, scholars have examined the topic through the lens of corporate imaginaries and economic motivations. Hsiao & Shorey (2025) explore the corporate vision of technologies, arguing that how companies conceptualized and explained vision of the technologies they produce shape how it is used and ultimately its social consequences. Those authors analysed vision of technologies in the context of work, finding that corporate discourses reconfigure labor practices. They argued that narratives of future potential and future uncertainty of AI lie not only in the value of profits of sorting products, but mostly in the data created by AI that is collected through monitoring in real-time then transforming labor and material waste into data. In other sectors, such as medicine, Jaton & Sormani (2023) show how the construction of benchmarked datasets is central to legitimize 'AI-enabled cancer immunotherapy' despite the limitations, costs and contingencies of these datasets. Other authors such as Sartori & Theodorou (2022) also point out that narratives configure a technological imaginary, hopes and fears of programmers, entrepreneurs or citizens may impact AI´s technical development, and how they depict and account for AI.

Concerning also to companies, scholars from management and design fields questioned the awareness of the social impact of AI by companies. Baker-Brunnbauer (2021) argues that social impact is addressed as a competitive element, not as a social problem because management is focused on cost savings, process optimization, and profitability. Gerdes (2021) emphasizes that only when societal issues gain public attention, the tech industry tries to address them. The author argued that commercial interests that drive industry tend to overlook societal problems such as inequality and racial discrimination. From the business perspective, Leszkiewicz et al. (2022) develop a concept of the social value of AI synthesized as a trade-off between the benefits for stakeholders and the costs and concerns of AI. According to the authors the focus is on the benefits because these will exceed the concerns over time due to laws, self-regulated companies and advancements in algorithms and technical capabilities, which consequently will increase the social value of AI.

2.3 Roles and moral status of AI

Another salient topic in the research of AI is the roles played by AI. Academics from different fields have examined those. Ma et al. (2024) identify three roles named as recommender, analyzer, or devil's advocate, and effects of the interaction of those roles and AI performance in decision-making. The research demonstrates that AI roles are influenced by how those are integrated into decision-making processes and by AI performance, highlighting issues such as rarely analyses by people when they

receive suggestions from the AI role of recommender. Other scholars such as H. Lindgren (2024) has researched roles of AI in relation to how AI is perceived and used in different contexts. For instance, the role of a robot as leader, follower, tutor, even victim and bully, or the role of AI as an oracle for decision support and prediction. The author pointed out that roles and human AI- relationships should be considered when designing AI. Likewise, Kim et. al (2023) study AI roles based on human involvement and AI autonomy identifying four different AI role clusters: AI tools, AI servants, AI assistants and AI mediators. The outcomes indicated that users value autonomous AI but prefer AI that allows human control. In alignment with (H. Lindgren, 2024), Kim et al. (2023) highlighted also the design of AI, emphasizing the necessity for a balanced consideration between AI autonomy and human involvement. Besides, concerning to the design and development of AI systems, Peeters et al. (2021) mentioned that the perspective used to develop AI systems normally is not planned neither encompassing a wide-angle view of problems and solutions, rather it depends on "a coincidental matter of who happens to be in the development team" (2021, p. 232). Something that partially counteracts STS scholars, such as Redaelli (2023) who mentions that in many instances impacts cannot be solely ascribed to the intentions of designers and users.

Building on the roles of AI, STS has interrogated the moral status of AI, a question that remains unresolved. From a theoretical perspective, Redaelli (2023) analyses Socio-technical Systems Theory (STST), Mediation Theory (MT) and postphenomenological MT. The author finds that, on one hand, theories state that non-humans and humans influence each other, and it is within that collective where the moral framework is formed. On the other hand, technologies are understood as either morally neutral or as moral agents similar to humans. The author highlighted that in the case of AI and humans would be considered to share the same moral agency, companies would transfer the responsibility to technical systems. To address that concern, the author focuses on the postphenomenological MT that emphasizes the intentionality relations between human and technology where AI plays a mediating role, but it does not possess any moral responsibility.

2.4 Risks of AI

As mentioned in the introduction, risk is also a topic to address with the literature review. Studies have drawn attention to the risks of AI, concepts, and other stances. Ulnicane et al. (2021) explain that AI development and use encompass concerns about potential concentration of power and increase of inequality. Crawford and Whittaker (2016, as cited in (Ulnicane et al., 2021)) highlighted that 'the risk that AI systems will not only aggregate power by reducing the ability of the weakest to contest their own treatment, but also that those who design the systems will be able to define what counts as ethical

behaviour". The research of Gerdes (2021) indicates that the increasing use of AI into high-risk domains requires the research agenda focus on model explainability and interpretability to comprehend how AI´s decisions are made and allow humans to scrutinize. The author raised concerns about the potential unintended consequences of AI because of the current influence of companies on AI research, and the overconfident predictions that inform the decisions to introduce AI in high-risk sectors. Other authors provide definitions of the risk factors of AI that should be taken into account by business:

> *Bias*…is embedded into organizational or industrial cultures, personal, unconscious, and human bias and data representation bias…*Fairness* is about treating people equally through developing models that encapsulate moral standards in the decision-making process. *Explainability* …all stakeholders, including people impacted by the decision of automated systems, can understand how a decision is made and the user knows why a system has made a decision. *Societal impacts* (potential benefits and harms) must be considered by a business, not only just to mitigate reputational damage …but also to meet or exceed the minimum standards of business ethics Business must question where *responsibility* (task and obligations) lies within their AI governance framework and define *accountability* (oversight and liability) to roles across the design/development/deployment life cycle. (Crockett et al., 2023, p. 780)

Other scholars from different fields have been focused on the ethical issues that entangle risk and should be taken into account by companies. Cunhna & Estima (2023) define six groups of AI ethics issues: privacy and regulation violations, intellectual property rights violations, environmental and socioeconomic harms, enabling malicious actors and harmful actions, broken systems, and hallucinations. The two former terms which are perhaps not totally clear in their name signify "Broken systems: ...situations where the algorithm or the training data lead to unreliable outputs" (Cunha & Estima, 2023, p. 98), and hallucinations as erroneous information in the outputs from AI, which is a potential risk visible for instance in chatbots, because "these conversational AIs present true and false information with the same apparent "confidence" instead of declining to answer when they cannot ensure correctness. With less knowledgeable people, this can lead to the heightening of misinformation and potentially dangerous situations" (p.99). Companies need to prevent unethical AI outcomes from biased learning or input data, prevent discrimination and mitigate risks for liability (Jöhnk et al., 2021).

Furthermore, the debate of the risk of general AI (AGI) because of a technological singularity that would become uncontrollable has been pointed out by H. Lindgren (2024), who also addresses interests in public debates about developing regulations and practices that ensure an explainable, responsible and

trustworthy AI. Those concepts that are part of the Ethical, Legal and Social Impacts (ELSI) have been criticized by STS. For instance, Williams (2006) argues that the increase of ELSI requirements encompasses hazards of simplification, framing and influence, reducing complex sociotechnical dilemmas to checklist audits, and replicating narrative bias by assuming static and not dynamically relationships between design intentions and outcomes.

Scholars have also analysed risk from another perspectives. S. Lindgren & Holmström (2020) emphasize that the process of datafication represents a risk for social scientists. The enormous volume of data and the culture shaped around numbers valuing only what is quantified, it could diminish societal topics and analytical methods for researching about beliefs, social change, expression, and other critical perspectives. In addition, Gerdes (2021) points out that the research agenda should be reconsider because of the huge carbon footprint involved in training models. Analyses of Suchman (2023) questioned the proposal of algorithmic intensification as a solution. The author argues that it could increase global problems such as climate crisis, inequality, forced migration, and ignore those concerns. Albeit STS scholars acknowledge the risk associated with technologies, analyses by Williams (2006) underscore the necessity to pay attention to the presumptions that certain technologies inherently involve significant risks, and that these risks are likely to be different from those posed by previous technological developments.

2.5 AI Act

In April 2021, the European Commission presented a proposal of regulation of artificial intelligence, the first attempt at global level to regulate AI. The regulation is part of EU strategy that includes resources by the public sector and investment by private sector that will be mobilised for a decade in order to reach the goal of "allowing Europe to amplify its ambitions and become a global leader in developing cutting-edge, trustworthy AI" (European Commission, (n.d.-b). The proposal of the regulation contemplates different levels of risk: unacceptable, high, limited and minimal risk, explained by Tomada (2022) as follows:

> First, it prohibits the implementation and use of AI systems that present unacceptable risks. Further on, it permits the uses of both AI systems presenting high-risks and the ones with limited risks. The high-risk systems are subject to compliance with specific requirements and obligations, while the limited risk systems must comply with transparency obligations. Lastly, the proposal mentions AI systems which present only minimal risks, and which are not directly targeted in the AI Act.

The final version of the regulation was published on 13th. of June 2024 - the EU Artificial Intelligence Act (EU AI Act). The purpose of the regulation as states in the EU AI Act (Whereas 1) is

> to improve the functioning of the internal market by laying down a uniform legal framework … of artificial intelligence systems …, in accordance with Union values, to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights… (The European Parliament and The Council of The European Union, 2024)

The regulation presents what is called a solid risk methodology to classify AI systems. The criteria to classify those as high-risk comprises function, purpose, and uses of AI applications. Based upon that, the regulation states in Chapter III, SECTION 1 Classification of AI systems as high-risk, Article 6 Classification rules for high-risk AI systems in the Artificial Intelligence Act:

> 1. Irrespective of whether an AI system is placed on the market or put into service independently of the products referred to in points (a) and (b), that AI system shall be considered to be high-risk where both of the following conditions are fulfilled:
>
> (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;
> (b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I. (The European Parliament and The Council of The European Union, 2024).

In addition to the criteria and categories of high-risk AI systems mentioned above, the regulation in Annex III lists the areas that will be a matter of high risk, making clear that the list may expand in the future. Those areas include infrastructure, education, employment, public services, law enforcement, migration, and others. Similarly, in Article 3 Chapter I General Provisions, the EU AI Act provides definitions of AI systems and risks, as well as of the concepts and terms used in the document to define the scope of the regulation. The EU AI Act will gradually come into full force by 2026, with some exceptions by 2027 (European Commission, n.d.-a).

2.6 Conclusion

In sum, STS and other research fields have made contributions to understanding the visions of AI futures and limitation of other possible futures, and narratives of AI. Researchers also provide valuable

perspectives on future effects of AI, corporate imaginaries, social value from business discourse, and interests that mobilize AI industries. Furthermore, studies have focused on examining roles of AI, roles in relation to human involvement, and the moral status of AI. Concepts related to risk of AI have also been explained such as bias, societal impact, fairness, explainability. Likewise, critiques of ELSI and further analyses of risks for social studies and presumptions have been analysed. Despite the ample literature that has researched AI from different perspectives to my knowledge there are no studies to date focusing on the Small and Medium-Sized enterprises (SMEs) ecosystem, specifically on SMEs that produce high-risk AI. Based on the literature review, I think that empirical studies of current cases of high-risk AI produced by those enterprises will provide valuable insight to learn more about this categorization of AI that will be subject to specific legal requirements in the European Union, with a particular focus on SMEs and start-ups.

3. Research Design, Methods, and Ethics

The following section will outline initially how the research was designed, the accounting of the field access, as well as the research questions. Subsequently, I will lay out the methods used to analyse the case- study, then the ethical considerations and limitations.

3.1 Research Design and Field Access

The research I conducted is qualitative research. Often this kind of research is inductive and has the aim to establish understanding of themes that have not been so much researched, also develop theories or discover processes in human interactions (Jensen & Lauren, 2016). This speaks to the topic of this thesis, high-risk AI, a concept posed by the regulation of AI whose research of this topic is in its infancy. It also speaks to the approach of this thesis, which is a case- study that helps me to attend to practices of high-risk AI in the European Union context amidst a regulation.

Due to the focus of the regulation on Small and Medium-Sized Enterprises ('SMEs') and start-ups, I decided to search a case study of one enterprise. Acknowledging that the proposal of the regulation of AI encompasses contentious debates; to design a research based on a case-study was a challenge. In the process of attempting to recruit participants, I made notes of the successes and failures (Silverman, 2006) that I will inform as follows. I started my searching in Austria, I attended one event where I contacted a person of Austria Wirtschaftsservice Gesellschaft mbH (AWS) who was by that time in charge of what they called Deep Technologies. The person expressed a willingness to help me to find a suitable startup. Some days after the person indicated that had initiated contact with AI startups and was waiting for responses. The final email received was a response with a contact of a not high-risk AI startup but with the explanation that might be an interesting company to exchange information with. I proceed to contact them, but I did not receive response. I cannot ensure that the person of AWS contacted startups, although there is indication that at least some approach was made, if so, it remains uncertain the reasons for not having satisfactory responses. Subsequently, seeking for more options I proceed to contact an acquaintance lawyer who had analysed the AI Act, who directly contacted a colleague to ask about it. Coincidentally, the answer was a contact from AWS. I decided not to contact that institution again but to approach a university assistant (prae doc) who researches about the co-production of artificial intelligence and society, and had explored the AI Act proposal. She gave me valuable insight into the proposal of the regulation, and to find a suitable startup she referred to contact the same institution (AWS). Following in the course of thesis sessions with the supervisor and fellow students were identified potential reasons of the difficulties encountered to contact startups in the field of AI, along with options for redefining the thesis

topic. Some reasons encompass companies do not want to be labelled as high-risk, I might probably avoid using the word “high-risk” when explaining my research interests at first, and I might be careful with that categorisation because the regulation was not in force yet. The potential topics for researching include startups whose products or services are not necessarily in the category of high-risk AI and how startups are trying to categorise their AI products or services in light of the proposal of the regulation. I thought about both, the potential causes of difficulties to reach a case study and the possible research topics, but I decided to persist extending the search for a case-study in another country of the European Union. I sent a general description of my thesis topic to some contacts in Spain. One friend introduced me to one enterprise.

The approach to the startup began with an email about the scope of the research. Subsequently, an exploratory meeting was arranged, whose organisation encompassed searching general information in internet about the startup, writing a guideline for personal use about the analysis of the AI Act and the factors that personally seemed to make the startup in mention to most probably be classified as high-risk, and about the scope of the research and the methodology intended to use. Subsequently, the online exploratory meeting was carried out with one of the co-founders of the startup and the AI specialist. The conversation focused on the proposal of the AI Act, mainly the areas and classification of high-risk AI systems, to which the AI specialist mentioned that the startup will most probably be into the high-risk category. That comment was the main reason for choosing that enterprise as a case-study because the acceptance as a likely to be a high-risk AI signalled a perfect case to analyse in light of the research interest on high-risk AI solutions. Besides, the willingness of the startup to participate in the research could facilitate the aim of providing a detailed exploration of a concrete case of high-risk AI. Furthermore, the selection of the case-study was informed by the considerations stated in the AI Act proposal (version 2021) concerning high-risk AI, specifically the list of areas that would be deemed high-risk. After the exploratory meeting a writing document was requested by the startup with the scope of the research, type of information to analyse in institutional documents, type of questions in interviews, profiles to interview, and timeline. Months after the research began with this startup whose business scope is freelance platforms and software development to connect or, as the participant in the exploratory meeting mentioned, to match freelancers with companies to develop projects in different fields. The startup develops an AI system that, among other tasks that will be addressed in the present document, it provides companies with a recommendation of three freelancers from a list of candidates who are deemed to be suitable for the development of a project. The startup´s mission is to reinvent the way of working.

Returning to the topic of the challenges associated with finding a case-study in Austria, the following statement, which was encountered right after selecting the Spanish startup, provide a potential

explanation of that issue. According to the European Union project AI Social Design Thinking Lab (Interreg ATCZ 271) “data show that in Southern Bohemia, Upper Austria and Vienna the potential of AI is used much less than in other European and non-European countries - mainly in the area of research and development and only rarely … in companies” (Institut für Höhere Studien Wien, 2022).

3.2 Research questions

[MQ] How do startups develop what would be most likely high-risk AI?

Before explaining the main question of my research, I would like to share that initially the question sought to analyse how startups envision the society when develop high-risk AI. However, I realize that the analytical process of the empirical material invites me first of all to attempt to understand high-risk AI meaning and how high-risk AI is made. Therefore, I decided to lead my research endeavour to empirically understand high-risk AI.

[SQ1] How are problems, concerns and needs of society identified to develop high-risk AI?

This question examines whether the society is considered by startups when they decide to develop AI, also have insight of startups´ AI developing process in relation to what the society needs.

[SQ2] How do startups conceive high-risk AI?

Address this question contributes to explore how startups understand, create, use, and form a mental idea of AI, and the product they are developing which most likely be categorised as high-risk.

[SQ3] How what the startup pursues when developing high-risk AI impact the society?

With this question I attempt, first, to identify what startups are trying to achieve with high-risk AI and based on that, to investigate the impact of that goal or motivation for the society.

3.3 Methods

Given my interest in exploring a high-risk AI, a topic that has been widely discussed recently due to the regulation of this field, but not easy to understand in practice, the methodology I developed comprises document analysis and semi-structured qualitative interviews. Acknowledging that the former method resulted in detailed audio data that required me major effort to transcribe and analyse (Jensen & Laurie, 2016), but it allowed me to follow up interesting points in the interviews that otherwise would have been difficult to understand. In what follows, I will lay out the methods and ethical considerations for the research.

3.3.1 Document Analysis

Once the startup agreed to participate in the research they committed to provide internal documents. Those were sent hours before the first interview; hence, those were not taken into account immediately but analysed afterwards. The documents in question correspond to two drafts of internal documents. The first document addresses the application of AI within the company, and the second the infrastructure of AI. During the course of the analysis of documents, a public knowledge base emerged as a third source for analysis. Besides, I came across with the website and information in various formats about the AI product and vision of the startup more generally. I consider those to be supplementary/background material in the context of this research and use it accordingly.

For the analysis I adopted an explanatory approach to the documents to enhance the research. First, I paid special attention to the explanation of AI, the aims of the AI product, its functionality, as well as terminology used. Then, after coding the interviews I went back to documents and other sources of information mentioned above with focus on some codes such as, generating value for freelancers with AI, using technology to impact the future of work, identifying needs, automate processes, improving the algorithm and impacting further public. In documents I encountered some information of interest to me in technical language, that requires me further research through online sources but mostly to delve into other collected data - interviews and public knowledge base. For instance, the overall functionality of the matching algorithm, the protocols and explainability metrics.
I attempted to have a comprehensive understanding of it since one of my aims is, to the best of my ability, to explain as clear as possible the topic of high-risk AI.

My thesis process includes also the proposal of the regulation of artificial intelligence -AI Act (version 2021), a key document that constitutes the basis of the research topic. Even though I decided not include the outcomes of analysis of that document in the thesis, certainly, those analyses provided me knowledge and also influenced how I shaped my final research topic, therefore, I consider relevant to inform about the process of that document analysis. As Prior (2008) mentioned “that in matters of social research, documents do much more than serve as informants and can, more properly, be considered as actors in their own right.” (p.822) and that is the proposal of the AI Act, a performative document produced to shape the future of AI in Europe. My approach to that document had different phases and purposes. Initially, I laid down categories to analyse the document such as, the vision of the future of society depicted in the regulation, definition of risk levels, concepts and actors of high-risk, intertextuality, tensions, and guidelines for AI systems. First, I took the Explanatory Memorandum, then I took the proposal itself, specifically those titles that agree into the categories previously defined. In that analysis, I extracted

general concepts, understood the structure of the AI Act (titles, chapters and annexes), had sights of the vision of the society addressed in the regulation, and brought up the context of the AI and valuable information to approach startups in the early stages of the thesis process. Subsequently, I researched the valuing process of AI in the proposed regulation. That involved the identification of the valuation constellation, which comprises actors, positionality, and roles played in the dynamic of valuation in the proposal of the AI Act. In addition, I made a general identification of valuation devices to ascribe value to AI, as well as the temporalities of the valuation process. Perhaps, the reader might be curious as to why those outcomes are not included in the thesis. It is important to clarify that the preliminary research of the proposal of the regulation highlighted the relevance for society, from a personal perspective, of understanding the practical implications of high-risk AI. Therefore, the interests shifted towards focusing on a case-study rather than on the AI Act.

3.3.2 Semi-structured qualitative interviews

As I already mentioned previously, analysing the proposal of the regulation of AI provided me valuable insight about the research topic. Collecting literature on the topic of interest is a way of working with interview data that gives clues of the potential interviewees and questions to ask (Silverman, 2006). Further, the exploratory meeting with the startup mentioned in the Research Design section accounted for outlining the first questionnaire. The questions attempted to obtain insight into how the identification of problems, concerns and needs of society is considered to produce high-risk AI [SQ1], the conception of high-risk AI [SQ2], and the impact on society when developing high-risk AI [SQ3].

Overall, the questionnaires for semi-structured qualitative interviews were design with the intention of following recommendations by Jensen & Laurie (2016) of writing a list of questions to cover the research coherently while maintaining the questionnaires flexible to adapt questions during the interview. The first questionnaire includes open-ended questions to get an overview of the operation of departments involved in the creation of AI-based solutions (general process, professionals involved, technology used), to get insight of what is the AI solution that is likely to be high-risks, how it works and the vision towards that product. Also, to comprehend social, technical and other aspects that are taken into account for the development of AI solutions, and perspectives of the impact of the proposal of the AI regulation on the AI solution created. Subsequently, as one of the practical examples provided by Silverman (2006) to work with interview data, I wrote up notes on my reactions and observations of the interview, then I analysed the direction of the research and identified ideas I was interested to discuss with the next interviewee. The second questionnaire aimed to cover questions about actors considered in the identification of needs and problems to solve, history of the creation of the AI product, process of

ideation and implementation of the matching, and vision of the product. During the interviews I attempted to do analysis in the making of the interview and tell the participants about my thoughts or pose questions to somehow 'test my analyses' (Silverman, 2006). For instance, talking about the probabilities of the AI system to be considered high-risk AI because of the use of IA to recruit candidates, evaluate them, also to announce vacancies, all that part of the selection process that is done by the algorithm in a certain way. Both questionnaires comprise nine questions, different for each participant, with relevant questions marked in bold, whose answers will constitute valuable insight for the [SQ]. Only one question about the vision of the AI product appeared in both questionnaires. The flexibility of the semi-structured interviews allowed to ask extra questions that came out during the interviews and obtain detailed information for a more fine-grained analysis.

The two semi-structured interviews were carried out via Zoom of each 55 minutes and 1 hour in length. I included the questions of the interviews as an appendix (Annex I: Interview guidelines), since the interviews were in Spanish I translated those into English. Important to mention that interview guidelines do not include questions that I developed spontaneously during the interview. The transcriptions were made with the available feature of Microsoft Word, the review and correction of the transcriptions were done manually, and it was decided to translate into English only the quotations included in the analysis.

Having describe how I structured the interviews, I now briefly introduce the participants. Consider the participants encompasses thinking about potential interviewees, special insights and kind of information they could provide (Jensen & Laurie, 2016). The participant in the first interview was the AI specialist of the startup. He attended the exploratory meeting mentioned further above, where I had the opportunity to know about his role as the leader of Data and AI department, also noted his knowledge of the product and the regulation, as well as his background which caught my attention. The second participant I had in mind would be Product/Services Engineer or Product/Services, however, after conducting the first interview, I told the AI specialist my objectives for the second interview, and he told me to carry out it with the Product Manager (PM). This participant has extensive knowledge about the history of the startup and the product due to the different roles he has played since the beginning of the enterprise as a UX Designer, project consultant and the current role.

Regarding the coding process, as it is said that gathering data is learning by doing, the coding task was a learning process that I decided to do without a software, meaning that I labelled content, I found patterns, connect data to (SQs) and notes, and made a list of codes. The first attempt of coding resulted in medium- sentences "codes", no more than two per question/answer. Those looked like a small resume of what questions/answers were about, but I still considered that as the early coding stage. Then, I went through notes written down during and after interviews, translated those into key phrases and add

them to the list of “codes”. I highlighted them to make a differentiation from the “codes” of the interview itself. After, I read a chapter about coding from the book *Constructing grounded theory* (2nd edition.) (Charmaz, 2014), which gave me valuable insight into this process and motivated me to attempt to follow the recommendations. Certainly, I did not follow the order exactly as the book described because I had already done a first coding attempt, but I moved into a second phase based on that. The second phase involves selecting the most telling codes to synthesize, integrate, and organize the information (Charmaz, 2014). I coded in-depth mostly each line of each question/answer of the two interviews. Then, I identified the most common codes from both, early coding and second coding of the two interviews. Subsequently, I moved to the focused coding (Charmaz, 2014) determining the most salient codes from coding the initial codes, comparing those with the data and identifying the most conceptual ones. That process gave me confidence in my analysis of the data.

Then I went back to the interview´s notes, the ones that I highlighted initially. Those were topics that drew my attention during the midst of the interviews and after which I decided to take into consideration because as Silverman (2006) said “The world never speaks directly to us but is always encoded via recording instruments like fieldnotes and transcripts” (p.113). Reading the list of the most salient codes and the interview´s notes I attempted to think carefully what path to follow according to the findings and my interests. Those topics that were part of the salient codes and of the interview´s notes were the aspects that I considered as possible paths to go - in depth. From that point on, I attempted to take a pragmatic approach making a list of how much insight (information) I had about those identified aspects (topics), and the lack of needed data, then I stopped for a couple of days to reflect about what I had done so far and the following steps. Such reflexivity process brought me to inquiry about my research questions because I felt the necessity to have a ground pole to be able to continue towards the next stage of the research. Did the topics identified through interviews answered /resonate with the research questions? How did I feel about my research questions at that moment? After many deliberations with myself I went back and forward considering my interests, and the insights that I got from the coding I made some changes to the research questions. Those changes accounted, in my case, as an evolution process considering basis stated in the initial research questions, definitions identified in the second update handled in the last thesis seminar, and as Charmaz (2014) mentioned, unforeseen aspects that emerged from the coding.

3.4 Ethical Considerations and Limitations

The treatment given in the present thesis to the system of the case - study as most probably high-risk AI come from my understanding of examples and the list of areas provided in the AI Act, and the initial

analysis of the regulation made by the startup to consider that its product could be in that category. However, whether that AI system meets the criteria for high-risk classification will be determined first by the startup and finally by the regulatory bodies in 2026. Therefore, this thesis is for researching purposes only and no classification can be derived from it.

Based upon that, I decided to protect identities of participants and as much as possible the enterprise. Personal names are not mentioned, only their roles at the organisation. The name of the enterprise is not released neither names of internal documents and public online resources. Likewise, links of the public online resources and website are not included in the references. However, analysis of the documents and the interviews raise ethical questions about de-identification, specifically of the startup itself. My interest in explaining as clear as possible the topic of high-risk AI brought me to provide information about the AI system whose data might still be linked to an specific enterprise, which could allow the identity of the startup to be revealed. Therefore, it is pertinent to mention that such information is analysed in the context of the thesis as well as used only for scientific and educational purposes. as stated in the written informed consent created for the interviews. Following Jensen & Laurie (2016) it was asked to participants to review the informed consent document beforehand and sign it. Furthermore, information analysed is also drawn from the public knowledge base of the startup, which as its name suggests, it is a publicly available source.

The main limitation of the research is the analysis of only one AI product of a specific industry/sector, that might be a small participant sample in the endeavour of understanding high-risk -AI. Besides, the research was carried out from theoretical and practical perspective, taking into consideration interviews, documents and public information of the startup, but it does not analyse perceptions of the end-users (freelancer and companies). Furthermore, the research attempted to identify a constellation of impacts, however, those do not capture all potential impacts in the corresponding sector or beyond.

4. Analytical Framework

This section presents studies, concepts and notions that sensitize me and guide me throughout the journey of addressing the sub research questions [SQs] in my attempt of understanding empirically high-risk AI.

4.1 Valuation studies

In my approach of the [SQ2] I started questioning directly the startup regarding aspects taken into account to use AI rather than any other technology, the answers bring me to consider valuation studies as the lens for the analysis. For the startup, whatever task they want to do that somehow motivates the deployment of AI the most important is to identify the value added by that technology, but they also point out that value is the freelance community they have. In that sense, valuation studies enable me to look at how the valuation of AI has been ascribed by the enterprise, at what is the real meaning the startup expresses regarding value. Specifically, the concept of valuation constellations of Waibel et al. (2021) sensitizes me to attempt to understand the elements of that process of attributing worth to AI.

Waibel et al. (2021) points out that a valuation constellation entails three components: positions and relations, rules, and infrastructures. Those are key components in moments of valuation, therefore, useful to analyse valuation phenomena. The use of the concept with its three components most probably allow to understand the whole connections and aspects that intervene in a situation of valuation. That would be useful for a study focused on valuation processes, but that is not the aim of the present research. However, digging more into the concept, I found useful to analyse the [SQ2] in light of the ´triad´, the first component of the concept of valuation constellation. It integrates a triad of valuator, valuee, and audience. Waibel et al. (2021) highlight that it is a triad of positions and relations essential for identifying how valuations are interconnected, that triad "is the crucial starting point for every investigation into the worlds of valuation" (p.37). Certainly, it constitutes only the first component, but perhaps, its use in the present research might also contribute to seed the beginning of studies of valuation of high-risk AI.

That notion of ´triad´ enables me to examine the valuator which I understand as the positionality in relation to valuee and audience; the valuee, that entails the dimensions that are exactly valuated, and the audience, which refers to whom the valuation is addressed. Adapting to the research that notion, I will analyse the triad in a different sequence: valuee, audience, valuator. That order allows me identifying positions and the relations among the three. That means, analysing first the valuee provides a perspective of what is truly considered in the valuation ascribed by the startup, and it enables me to question the considerations that perhaps are assumed, as well as make visible others that are not evident at first

glance. With that insight, the real audience of that valuation will be identified clearly. Once the valuee and audience are identified, a deeper level of understanding is gained with regard of the positionality of the startup in relation to valuee and audience. As can be observed, the three: valuee, audience and valuator are interconnected. The specific relations of that ´triad´ will be elucidated upon completion of the empirical analysis.

4.2 Additional concepts and frameworks

In addressing other subordinate questions, additional concepts and frameworks are useful in the process of analysing the research material. In the [SQ1] is taken in consideration the explanation of different AI roles in AI-assisted decision-making address by Ma et al. (2024). The authors present three roles: recommender, which provides recommendation and explanations; analyzer, which analyses options providing evidence for and against each one but not recommending, and the role of devil´s advocate, which presents counterarguments to initial choices, even to AI's own choices (Ma et al. 2024). Specifically, for the research the notion of ´role of recommender´ is considered because that is most probably the main role played by the AI systema developed by the startup of the case study.

In the [SQ3] the 'Socio-Technology Analysis and Prescriptions (STAP) framework of Baek & Lee (2021) provides me with tools to analyse the impact of technology. The whole framework is not used in the research, I intend to use concepts and aspects that resonate with the purpose of identifying and analysing how what the startup pursues when developing high-risk AI impacts the society. Baek & Lee (2021) based on a review of main STS theory, determine six basic attributes for the STAP framework: target technology, society scopes (micro/meso/macro), result sentiment (positive/negative), impact, causes, solutions. The authors define "attribute as anything related to influencing society or technology in the relationship between technology and society described by the three theories: TD, SCOT, and ANT" (p.3). STAP attributes become a list of conceptual questions posed by Baek & Lee (2021). I select a number of these that allow me to navigate my SQ3, I adapt these to the research accordingly: what is pursued by the startup, what is the technology influenced or influencing, what are the groups influenced with the AI product, how are influenced, and what is the impact. Furthermore, to analyse the social group influenced, the STAP framework encompasses the ´Society scopes´ that are categorised as Micro, Meso or Macro, which I include in my analyses because those might serve to explain impacts across the evolution of the AI product.

The research overall is also informed by Suchman (2023) and her reflexive process of research practices regarding AI, that entails questioning what and according to whom the problems must be solve with AI, the often-uncontested autonomous nature of AI, and the translation of social practices into

statistics. Besides, for the analysis of [SQ2 and SQ3] result useful considerations of Jaton & Sormani (2023) regarding human intervention to make AI with common factors with society and real world. Also, examinations of Hsiao & Shorey (2025) concerning explanations of emerging technology with the ideologies of AI producers and vision of companies.

5. Empirical Analysis

Considering that the focus of the research is high-risk AI, the approach of the empirical chapter is descriptive and analytical in order to attempt to understand this ´recent´ category introduced in the regulation of AI in the European Union. That approach is inspired by what Jaton & Sormani (2023) questioned "Doesn't the social analyst have to stay close to concrete actions—if only for him/her to become able to document and describe the activities that participate in the local production of meaning and order?" (p. 627).

To account for that curiosity of understanding empirically high-risk AI, I will start with the first sub-question*: how do startups identify problems, concerns and needs of society to develop this technology?* Next section will try to address the second sub-question: *how do startups conceive high-risk AI?* and third section will focus on: *how what startups pursue when developing high-risk AI impact the society?* The three analytical sections will draw upon insights from theoretical approaches and concepts from the chapters Literature Review and Analytical Framework.

5.1 Consider problems, needs, concerns whose?

Attempting to understand how the startup identifies problems, needs and concerns of society allows me to interestingly get insight into who was considered initially for the development of an AI system, and the variety of roles have been given to technology. Furthermore, it contextualises the use of AI by the enterprise.

5.1.1 Problems and necessities of someone

Entrepreneurs usually disrupt the market with new ideas to solve something, to meet or fulfil necessities or to create better products or services, but not to build solutions for a desire, problem or a necessity no one has. However, it does not mean that they always consider the society as a whole, or think about the most relevant problems to solve, but at the beginning of their entrepreneurship they always identify people that will solve issues with those ideas. In the present case study the founders of the startup initially had a project agency, but at a certain time they faced a challenge. The number of projects increased, and to some extent, the motivation decreased to work on specific projects that were not so much in their interest. To deal with that, they started to hire freelancers specialized in particular topics. However, the requests for projects boosted, hence the looking for and hiring of specialized freelancers as well. Such situation motivated the creation of a product to make that process scalable to solve their issues,

according to the PM, who has been since the beginning of the startup with different roles such as project consultant and UX Designer:

> We said, we want to do that, how is the way to do it? ... there is nothing more than a similarity algorithm to do it, we couldn't think of any other option, ..., we didn't do so much to identify all the users' problems because we didn't have users.

The product started as a similarity algorithm to match freelancers with customers that have projects to develop. Background, motivations, and a variety of other topics were contemplated in the algorithm

> At the beginning we asked [to the freelancers] what do you do? …filling out a profile, what characteristics do you have? we gave four examples of personalities following the model of the Disc type of psychology... who did you identify with? …and then we asked, what do you like to do…?
>
> And we asked clients, what do you like to do? … what characteristics do you have?... and…What do you need for your project? And we made a direct match assuming that people who like the same things will get along better and then they will work more comfortably... That was version number one, what was happening? …it was not very effective because…it is not easy to identify with one of those models. (Product Manager)

That first outcome seemed to be no good, but as one of the characteristics of entrepreneurs is persistence, they moved their idea forward with technology. They had a first version based on a similarity algorithm, an initial idea of what they wanted to achieve, and data from freelancers, companies and projects, thus they decided to move ahead. The next step was to hire an AI Specialist, who came from the academia, which is particularly interesting because of the impact that seems to have on the AI product amid an upcoming AI regulation. I will come back later on this topic.

With an AI Specialist the startup continued working on the idea of matching freelancers with customers, but now with a specific technology -AI- to make that possible. According to the AI Specialist, the subsequent versions of the product include artificial intelligence in the first and second steps

> The first step is to put participants into our market- place, ...that means, that a company… gives us all the information possible about what company is, what it wants, what is looking for, what it needs, what project wants to carry out. The same for talents [freelancers], that people apply to, pass the interviews, approve quality filters, fill out the profile, that we have the information about those talents...The second step is for people to decide to work with each other…we algorithmically make a match, so to speak…we recommend this

> project to you [freelancer], or for a client …these are the talents [freelancers] that we think are the best for your project.

At this point the analysis indicates that initially it was an issue of a few (founders with a project agency overloaded) that was solved with one public (freelancers), then it turned to be an idea to integrate freelancers and companies that have projects to be develop. The latter public did not have a visible role in the issue before. That integration has been done with technology with those two publics at play (freelancers and companies) and taking into consideration specific information from them. The first target (freelancers) and the subsequent idea to integrate two actors to develop projects, have become a vision of matching necessities, skills, backgrounds, and motivations as accurate as possible with technology, specifically with AI, which means that there is much more at stake than there was initially.

5.1.2 Roles of AI

Currently, the startup identifies necessities and problems of different publics: freelancers, companies and internal team. To that end the startup uses different methods according to the public. In this research the analysis will be in freelancers, which are the target of the startup. In its terminology, they use the word ´talent´ or ´supply´ to refer to freelancers, and ´customers´ or ´demand´ for companies. The methods to know what they called ´pains of users´ and needs vary between conversations one to one with talent and customers, carry out surveys, receiving feedback, and observing how users interact with the platform. Some problems and necessities are mentioned by the users, while other that even users do not know or do not tell they have come out through analysis. Then the startup creates solutions using technology as the Product Manager (PM) explains

> I am analysing the flow of how a customer publishes a project in [name of the startup] …, then, ... I realise ... that we are asking him to tell us the budget, but the problem he has is that …he doesn't know it, ... So how can we solve that? Well, maybe ... by using an algorithm to get the budget for your [customer] project.

In that stage, it can be notice that technology plays a role of ´solution´, and also that machine learning and artificial intelligence start to show up in that process, considering that algorithms are the building blocks that make up those. It is precisely because of the use of machine learning and artificial intelligence that other roles emerge, from technology to solve problems and meet necessities to technology to identify those. For instance, when users joint or show interest to join the platform of the startup for the first time, they are asked about desires, goals, and kind of projects intendent to develop.

Some of those questions apply for both (freelancers and companies), with that information the startup develops some algorithms to detect necessities. Considering how the AI Specialist describes

> ...we automatically estimate how many people you [customers] might need in your team and what kind of profiles. ..., there are many people who know this, ... but when a company comes to us and says I want to set up my website or I want to set up my platform ...or I want this and that... people say ..., I have no idea, ...I don't know how many profiles, what profiles ...Well ...obviously people are experts in their business, ..., but of course, you [customer] don't have to be technical and we have to detect that need and automate that process as best as possible.

Constantly, technology has been used to tackle problems and needs, which explains the role of ´solution´ that technology has, but also due to technological advancements some other roles are played for instance for exploration/investigation, identification, anticipation, prediction, forecasting and recommending. In the present case study, specifically in the process of identifying problems and need of customers, technology plays the role of solution and the role of detecting necessities of costumers as those were identified above. But, to play those roles AI takes into consideration different aspects and suggests what is good or suitable for a particular customer to solve a problem or meet necessities, hence, AI starts playing another role, which is recommender. Recommendations vary, for example, introductory messages for freelancers to contact a company, proposal of the budget or number of members of the workforce and the kind of profiles required for a project. However, in that role, AI provides recommendations in areas that might be a matter of risk.

> “The main objective of the matching system is to offer a personalized and individual recommendation of available positions for talent or vice versa”. (Internal document of the startup)

The role of recommender, name borrowed from Ma et al. (2024), in the case study encompasses a variety of tasks, such as collect information, filter, weight variables, evaluate, prioritize and decide, then the AI system that will most probably be considered high-risk gives an outcome, it recommends three people to develop a project. What companies do when they receive the outcome of the AI system, do they analyse the outcome? Do they look at the fourth, the fifth …profiles that were consider initially by the AI system? Do they ask for explanation how the AI system choose the profiles? Do they rely on the recommendations of the AI system? Those questions are posed to highlight that the role of AI as a recommender encompasses the handing over of decision-making, which among other factors, might influence the reliance and dependency on AI systems.

5.1.3 Discussion

The first part started by elucidating that for producing AI at the beginning, the problems, necessities or concerns of society were not consider, but only problems and necessities of the co-founders of the startup. Another factor that emerged is that the initial idea and the target evolved with the use of AI and have become a vision that would probably impact or is impacting a broader society. The next part, first enlightened about the roles assigned to technology to identify and solve problems and necessities of customers. Second, it presents the role of recommender assigned to AI and what that role comprises specifically for the AI product of the startup. Also, it provides an indication of the risk involved in the AI product and points out potential consequences of the role of recommender. Overall, a contextualisation of the use of AI by the enterprise was also portrayed.

In this way, the first section of the Empirical Chapter attended to answer inquiries of identification of problems, concerns and needs, which in turn, open the door to the world of high-risk AI, bringing about questions of what the meaning of that technology for that startup is, and how they truly understand and use it. With that in mind, the following section will start.

5.2 AI: core, value, risk

The following statement of the draft version of one internal document [1]explains what the startup of the case study is

> … [name of the firm] has implemented and refined an AI system aimed at performing automatic matching or matchmaking tasks to connect talent [freelancers] with business needs in projects taking into account professional and individual skills. Thus, [name of the firm] represents a bilateral marketplace, where companies and freelancers... meet to develop projects in a dynamic and joint way.
>
> The AI system is integrated within the … platform, which includes all the necessary steps for the development of projects… (p.1)

Based upon that it becomes clear the relevance of artificial intelligence for the core business. That keeps a direct correlation with elements that came to light in the previous section that delineated how AI started to become relevant for them, and for what the startup uses AI, which in a broad sense is for placing freelancers and companies inside of what they call their marketplace, and for the match between those publics to work together. The latter represents the reason why the AI developed by the

[1] Reference derived from the draft version of internal document of the startup (not publicly accessible)

startup will most likely be considered in the high-risk AI categorization. The relevance and use of AI raise interest regarding how they conceive high-risk AI.

When discussing the aspects taken into consideration to use AI rather than machine learning or any other kind of technology, the startup manifests that they build from the bottom to the top. Identifying problems or necessities come first, followed by identification of potential solutions, and then the technology to use for, always cautious to apply complex technological solutions to problems that can be solved simple. In addition, in that discussion a key element arises, the value.

> ... I come from statistics, so machine learning, artificial intelligence, whatever it is, at the end of the day it is matrices with matrices, so when doing any kind of task, for me the *key is where the value is generated* [emphasis added]. Value I mean that it can contribute to the business. That is to say, we never approach a problem with a perspective of doing machine learning for the sake of doing it or doing artificial intelligence because everyone does artificial intelligence. (AI specialist)

Naturally, the ´value´ is understood as the contribution to business because the startup in question is a business in essence. Nonetheless, it does not involve only business but also people as the same respondent clarified "We understand that when we talk about value, that is, our most important value within [name of the enterprise] is the freelance community that we have". Therefore, technology must generate value also for freelancers. That might mean, solve problems, anticipate needs, provide alternatives for performing task efficiently, optimize time, address necessities, so on and so forth. This conception of value of AI makes this cutting-edge technology highly relevant for startups because the value that enterprises expect to bring for their business and their target audience (in this case freelancers) is becoming apparently achievable mainly with artificial intelligence. One example is the basic functionality that gives the freelancer an introductory message to reach the company whose project the freelancer is interested to work on. The suggested introductory message includes information about freelancer's background and interests related to characteristics or requirements of the project, as well as correct grammar and compelling content. Once that functionality was implemented, the number of accepted job proposals were tripled, which was considered beneficial for freelancers because of their interest to work, also for companies because they seek people to carry out their projects, and of course it is beneficial for the startup.

Values in technology such as usefulness, profitability, and growth, are stable values since long time ago, but how the value is ascribed to each technology is what makes a difference in shaping actors, sectors, and practices around. The previous section described how for solving problems of the founders of the startup and achieve their goals, the startup initially created a similarity algorithm, and subsequently

transitioned to use artificial intelligence. In present section, to some extent is now evident that the deployment of AI is not because of the tendency among companies to adopt this technology, but it is driven by the value that according to the startup this technology offers to the business and the target audience. Hence, the transition between similarity algorithm, AI, to what will probably be considered as high-risk AI have been characterised by the valuation of that technology, which is not straightforward. Some of the multiple factors that take place in valuation are the dimensions that are exactly valuated, the positionality of the startup in relation to those dimensions, and the audience to whom are the valuation addressed. In valuation studies those factors respond to valuee, valuator, and audience respectively, named the ´triad of positions´ by Waibel et al. (2021). As was explained in the chapter Analytical Framework, the analysis of those factors in the present research will start with valuee, followed by audience and then valuator. First, to understand the valuee, it will be analysed aspects considered by the enterprise in its business beyond the mere economic factors.

5.2.1 Aspects considered beyond economic factors

Throughout analysis of interviews, website, a public knowledge base and accessible institutional documents, two aspects emerged: *talent first* and *adequate work*, as the aspects taken into account by the startup beyond the economic factors. *Talent first* means to take into consideration what is beneficial and desirable for freelancers (talent). As a case in point, in the platform of the startup freelancers are able to apply to as many projects as they want, while companies cannot invite as many freelancers as they want to join the selection process to develop projects. As another case, the negotiation of salary and other economic conditions starts with economic proposal made by freelancers not by companies. In the first case, the platform is parameterized in accordance with the conditions to prioritize freelancers, and for the second case, the negotiation between both sides (freelancers and companies) takes place in the platform. The identification of that variable – *talent first* – and examples of what it does mean, resonate with what was mentioned some paragraphs above that freelancers are the most important value for the startup, and it also points out talent first as one of the priorities of the startup. Furthermore, although the two cases in point do not require the use of AI, one question arises of whether this aspect *talent first* is taken into account in any extent in the use of AI.

> "… our most important value within [name of the enterprise] is the freelance community that we have. We started what we call Talent First. That is, for example, we only allow a client to invite three freelancers to a project, but the freelancers can apply for as many projects as they want and when a financial agreement has to be reached, it is the talent that has to set the conditions for the client, not the client for the talent (AI Specialist)

The second aspect that emerged was *adequate work*, which constitutes one of the fundamental principles that underlie the philosophy of the startup. It encompasses two elements, one is what the startup called valuable projects, that means medium-long term projects where freelancers can work, short-term projects are not contemplated. The second element is what the enterprise mentions as positive labour economy for both freelancer and company in charge of the project. That element of labour economy entails companies that will have an expert (talent/freelancer) not only as a consultant who is paid to give advice or training on the particular project but an expert who is doing the job, sharing knowledge and experience, and contributing for the know-how of the company for specific time. On the other hand, the same element encompasses what the startup names ´freeworkers´, which are freelancers that earn usually a higher average wage, want to be part of a team in the company, share his/her knowledge, develop the project they are interested in, and leave the company afterwards usually within a time frame of months. The startup supports a working model of no long-term attachment neither to one company nor one employee, focus that will be developed later in this research.

> …the philosophy is to create a work economy that is positive for both parties. Some gain flexibility, earn money … above what their market reference is, but the company also has the luxury because you bring in experts… who do the work… It is a new model of integration between people who do not want to be tied to any company and the company that also does not want to be tied to any external worker and can collaborate for a period of time for a specific project. (AI Specialist)

Overall, it seems that to ensure the aspect *adequate work* into the platform is not required the use of AI, nevertheless, acknowledging the relevance of that aspect for the startup, it leads to question of whether is somehow considered in the use of AI, which is basically the same question that was raised by the other aspect *talent first*.

Both *adequate work* and *talent first* were identified as not economic aspects the startup takes into account and are important in its business. Apparently, to ensure those into the platform artificial intelligence is not required, which prompts a further question of what is considered in the use of AI. It makes sense that the research continues to seek to address that question.

5.2.2 Aspects considered in the use of AI

In the analysis two aspects emerged: *product optimisation* and *customer experience*. *Product optimisation* indicates “All AI applications are focused on optimizing processes within our product” (AI Specialist). Processes that comprise detecting necessities of freelancers and companies, identifying what is required for those necessities, identifying which of those needs can generate business for the enterprise and how to create profits, translating necessities into professional skills according to the project field,

providing list of requirements that might be considered by companies, providing information as the input for future functionalities of the product, and recommending freelancers to project´s companies.

The second aspect identified is *customer experience*, which using AI seeks to facilitate activities of freelancers and companies, "… when client registers at [platform] … we ask them to write us … what are your objectives? What is the title of your project? What things do you want to do? And we automatically … detect needs and transform those into our skills so the person only has to select, … So, what we do is make the client's job easier" (AI Specialist). Practices of *customer experience* consist of deliver the best services of processes mentioned above. Among the most relevant are estimating resources for a project regardless the field (human resources, team member profiles, skills, budget), select and recommend to companies the freelancers that might be suitable for developing projects, suggest to freelancers projects to apply for, and provide personalised message to freelancers for the first contact with the company.

Both aspects *product optimisatio*n and *customer experience* are impacted to a large extent one to another. What is done with AI to optimise processes within the product, in most cases means features that impact positively on the customer experience. The same happens in the search of "making life easier for everybody", as the AI leader said, because identifying with AI potential improvements or necessities for/of customers is translated into the optimisation of the product. Hence, the use of AI and the aspects in question are interdependent. It is needed to acknowledge constraining elements of that interdependence, such as tasks made by the startup to determine whether profits can be made after weighing both aspects and making decisions.

At this point, on one side, it is clear that the use of artificial intelligence is oriented by specific aspects (*product optimisation* and *customer experience*). On the other side, the startup considers in its business other aspects beyond the economic ones (*talent first* and *adequate work*). It is pertinent to return to the question posed in previous section regarding whether those - *talent first* and *adequate work* are somehow taken into account in the use of AI. Based on what has been examined, it suggests that those are not considered because the aspects that guide the use of AI are different and do not include elements related to *talent first* and *adequate work*. As it can be noticed, the startup of this case study uses AI for its main business, appealing to scepticisms, do they really not take into account in the use of that technology aspects that emerged from the analysis that are not economic but relevant for its business (*talent first* and *adequate work)*?

From that sceptical position, a more fine-grained analysis was carried out of *product optimisation* and *customer experience*- aspects considered in the use of AI. The analysis reveals that within those there are processes that at least are related to *adequate work*. For instance, as part of *product*

optimisation, necessities of freelancers and companies are identified, that is the first step to find benefits for freelancers and companies in charge of projects, thereby, it impacts the positive labour economy-an element of *adequate work*- that seeks to benefit both actors. Another key process of *product optimisation* is the recommendation of freelancers to projects. That process impacts *adequate work* because the recommendation entails components that ultimately aim to benefit freelancers and company, such as choosing freelancers that have expertise in the field of the project, affinity with the project and cultural fit with the company, motivation and willingness to work in that project. Therefore, *adequate work* is taken into account, if only indirectly, in the use of AI throughout some processes of *product optimisation*.

It seems reasonable to posit that at least one of the non-economic aspects considered by the startup in its business (*adequate work*) is considered in the deployment of AI, but it also questions what about the other aspect *talent first*. In the course of this research, returning to the first part of this section, the AI specialist stated that the most important value for the enterprise is the freelance community. Indeed, talent first was established because of the priority that freelancers represent for the startup, so the aspect *talent first* should also be considered in the use of AI somehow, should not be? That uncertainty brings up to extend the analysis of aspects considered by the startup to develop high-risk AI.

5.2.3 Aspects considered for high-risk AI

Before moving on, it is important to briefly recall that high-risk is one of the levels of risk set in the regulation of artificial intelligence in the European Union (EU AI Act). The criteria for classifying systems, services and products within this category comprises function, purpose, and uses of AI applications. In addition, it is relevant to remind that the startup of this research has identified itself as being most probably classified within the high-risk category due to the nature of its matchmaking task. After examining documents, public information and interviews three aspects emerged as taken into account by the enterprise to develop high-risk AI: *performance of the matchmaking algorithm, data,* and *regulation*.

About *performance of the matching*, its relevance is a salient point from one statement of the Product Manager, who has been working at the startup since the beginning: “[the match] is very critical and very important within [name of the startup], …the algorithm … can never stop improving”. In the analysis was identified that *performance of the matchmaking* is characterised by five elements: protocols, explainability metrics, similarity, measuring point and validation by humans (expert knowledge). Protocols are used to determine performance and problems of the algorithm; it includes golden cases accepted or failed historically, i.e. cases where the three people recommended by the algorithm to carry out a project were not the adequate profiles for that project. Those cases are used to analyse the performance of new versions, e.g. one case whose recommendation was wrong in the past, once it is tested with the new

version of the algorithm, it performs better so it provides a more accurate recommendation of people to develop projects. Protocols also include statistic metrics such as improvements in selecting the top three freelancers compared to previous months, and recall metrics, for instance, the ability of the algorithm to recover the five most recommended freelancers in the history. The element of explainability metrics is used to explain why the algorithm recommends one person over another, also to understand how the algorithm learns and how variables influence predictions. Explainability metrics consist of analysing different features, cases, do exploration, and permutations, those useful for identifying features that are less or more important, as well as for examining other set of factors. Those are focus on understanding how the matching algorithm operates. It is pertinent to mention that explainability is a fundamental factor of transparency obligations set out in the regulation of AI.

Within the aspect *performance of the matching* exists another element pointed out as similarity. It constitutes the base match percentage for the algorithm, e.g. “a user who has the "Wordpress" or "Web Design" skills validated in their profile will have a high "base" match percentage with projects that require such knowledge” Public knowledge base (n.d.). It identifies whether knowledge, skills and abilities acquired and developed by freelancers in previous jobs are similar to the project in question. It does not take into account if the person has worked in the specific role requested. According to one internal document, that is the same recommendation and matching model applied in the field of Human Resources Technology (HR-tech) by different firms. However, as it has been mentioned throughout the analysis, key characteristics of the matching functionality developed by the startup are not only to match skills and technical knowledge, but also motivations and cultural fit, which as far as it is known by the time of this research, any other firm has done. The algorithm adjusts the match percentage based on the engagement of freelancers in the platform, and the assessment of personal profile that derives from two tests created by the startup, one motivational and one of personality. Those tests are filled out by freelancers and companies responsible for the projects. Despite the limitation of this research about having access to those tests, in the public knowledge base (n.d.) the startup mentioned “…we have designed a test to identify the interests of a freelancer, and to give a higher percentage of matches to projects that fit the talent's motivations”. That demonstrates that the aspect *talent first*, which was not found before in any other aspect considered in the use of AI, it is taken into account in *performance of the matching*. At this point, it makes relevant to have extended the analysis to high-risk AI since the question posed in previous section is answered. And, what else entails the development of high-risk AI?

Continuing with the aspect *performance of the matching,* it comprises also measuring points, evaluation, and validation. One of the measuring points is asking companies about three factors regarding freelancers after the first conversation between them. Those factors are capability to develop the project,

motivation for the project, and perception of the cultural fit with the company. For the evaluation, the matching accuracy is determined among other factors by freelancers, i.e., “For each recommendation we have the opportunity for the talent [freelancers] to tell whether the recommendation is appropriate or not. There are up and down buttons” (AI Specialist). That confirms once again the importance of *talent first* in the *performance of the matching*. Besides, a salient point in evaluation and validation is the considerable degree of human involvement in different stages. For instance, new versions or new features of the algorithm are validated first by the product department, followed by people of sales, talent or customer departments who have experience dealing with different cases, and identifying not accurate ranking of freelancer and recommendations made by the algorithm. Another example is the oversight of the matching which is made by team members of the product department who detect problems of the recommendation system and modify what is needed. Furthermore, the algorithm is continuously analysed by the AI specialist in light of metrics of statistic, usability, user/customer experience, and other factors. Activities that require expert knowledge and plays an essential role in the control of the algorithm. According to the AI specialist, the algorithm per se does not have limitations because it must facilitate processes, there are internal rules within the system, but not limitations. Therefore, the *performance of the matching* requires expert knowledge in order to ensure identification of complex failures and solve those consequently.

The second aspect considered in high-risk AI is *data.* According to the analysis two elements are the most relevant: information collected and data curation. As it was explained in previous sections, to put participants into the market-place (platform) requires companies and freelancers provide different type of information. The aim is to collect as much information as possible in all stages. For companies the information comprises name, industry, size, all the information about the project including requirements, and mention if the role of the freelancer would be individual or if it would be performed within a team. The latter resonates with the previous analysis of the non- economic aspect *adequate work*, specifically about the element of positive labour economy which encompasses the desire of freelancers to be part of a team to share knowledge and experience as experts in the firms. In addition, companies must explain how they will support the freelancer in terms of communication, tools, information for the tasks, and also, they must provide information about the duration of the project and whether it is a project for specific hours or a complete project to be developed. That is also related to both non-economic aspects *adequate work* and *talent first*, particularly about the elements of valuable projects that consider the time of duration as a key element, and the pursuing working model of no long-term attachment neither to one company nor one employee. Here it is noticeable again that the non-economic aspects identified in the research as the ones considered by the startup in its business (*talent first* and *adequate work)* are taken into account to develop high-risk AI.

In line with the relevance of collecting as much data as possible, a key element of improving the matching functionality is data curation considering this quote by the AI Specialist "we realize that the most important thing for us is to have very curated data, that is, data that we trust very well, that this data reflects the real processes". Startups are newly established business that are validating their business model and evolving continuously, hence the data changes constantly, which might increase probabilities of many data used by the algorithm worsen the performance of the matching. Therefore, for the enterprise, improving data quality, organising, and maintaining data sets is a priority. Factors such as temporality of data, when matches were made, active or inactive freelancers in the platform are some of the factors that might affect the matching. One of the tasks to keep curated data is data distillation, which involves processes of extracting insights and patterns from the data to have data in a usable form. Overall, tasks involved in data curation reiterate a previous finding regarding of the involvement of human that is essential for this aspect of *data.*

The third and last aspect that emerged regarding the use of high-risk AI is *regulation,* specifically the EU AI Act.

> …the ideation is the same, you have a prediction problem, a classification problem, or you want to apply a GPT or whatever and you are going to do it… the ideation is not lost. But if the execution used to take you a week, now the execution (obtaining the metrics and obtaining the adaptation to the regulation), maybe the day you are going to present it will take you another three [weeks] (AI Specialist)

The aspect of *regulation* has three elements: monitoring, expert knowledge, and delay execution. The first entails design transparent system, review protocols, set metrics and monitoring systems, as well as other processes and practices to develop responsible systems, as AI Specialist mentioned. The second element refers to expert knowledge of the AI regulatory framework that will be necessary for the years ahead as the same participant comment "if a startup… wants to use artificial intelligence… suddenly has to hire an expert in [the regulation] … that expert also must know how to adapt to the regulation". This is another example of the relevance of human involvement in high-risk AI. The last element, delay execution indicates that startups will take time to adapt the high-risk AI system to the regulation because the requirements and obligations that entail the AI Act, and the lack of resources regarding knowledge of the regulation and money to hire consultants. Whether this aspect of *regulation* will impact the product itself might probably be determined sometime after August 2026 when the regulation will come into force for high-risk AI systems.

> …what I think is going to happen is that people will continue to innovate, the problem is that there will be times when they will have to adapt everything to the regulation, and then months of work and/or money will be lost in consulting to do it. (AI Specialist)

This aspect of *regulation* in general does not take into account non-economic aspects considered by the startup in its business - *talent first* and *adequate work*, either the aspects considered in the use of AI - *product optimisation* and *customer experience*.

Overall, this section has required a longer analysis due to the complexity involved in high-risk AI. In addition to the aspects identified for high-risk AI, it has been elucidated that other aspects that emerge in the first section are also taken into account, specifically the non-economic aspects – *talent first and adequate work.* Furthermore, descriptions and analyses have shed light on how ´the most important value for the enterprise´, which is freelancer community- *talent first*, is enacted in the product through the *performance of the matching algorithm* and *data.* This section also reveals the key role of humans in measurements, evaluations, validation, data curation and regulation of high-risk AI. There is a last but not least topic this section signals: risk.

As has been demonstrated throughout the section, from the three aspects considered for high-risk AI, in two of them is where the risk emerges, those are *data* and *performance of the matching algorithm*. As was explained and analysed *performance of the matching algorithm* encompasses particularly more elements than any other aspect identified and involve non-economic aspects, it requires not only automation for metrics, measurements and other elements analysed above, but with the *data* aspect, those two currently also requires a high-degree of human participation, many departments of the startup involved, the public/target of the solution, and expert knowledge.

> …we make sure to try…to find where it fails, because in extreme cases … we have to go back and investigate what happened. …and it often happens, especially with black box models, where you don't explicitly control the weights or what is happening or could happen inside, so you have to go to the specific case…, and often it forces you to do almost detective work because since you often don´t have idea why it happens, go back, go to the training data, … Then you see the pattern that the model learns…
>
> We can still do that because our volume of data is very large, but since the cases we have… are so broad, we can still use time doing this type of lucubration and still not be very sure because in the end that is the problem with this type of more complex model, especially when you put in certain statistical models that are difficult to explain, where you still don't know what is really happening. But in those extreme cases, we review the training data. (AI specialist)

That indicates the scope of the risks associated with AI. Likewise, the statement pointed out that the quantity of data is large, but the cases are still broad, which means that the data sets are large and diverse in examples that allow the model to learn general patterns that apply to different situations, which

facilitates the identification of failures. That might happen because they are an enterprise that started in 2021, but they are growing.

> Until last year (2013), …, we were able to compensate the shortcomings of the match with people. That means, if the match gave a recommendation that was not adequate, …in the end …, we do not have such a large community, … so we have been able to manoeuvre if the match failed. … What happens now? We have more and more volume, so everything must be more and more automatic, so if the match does not work, nothing works. (Product Manager)

5.2.4 Triad of positions: entry point in the research of valuation

Analyses in this section have the purpose of understanding how startups conceive high-risk AI, that brought me to a journey through valuation process of AI with focus on valuee, valuator, and audience - the ´triad of positions´ by Waibel et al. (2021). The analysis of the first position - the valuee has taken the whole section until now in order to understand what is mainly considered in the valuation.

Examination elucidated aspects considered by the startup in its business beyond the economic factors. One is *talent first*, that represents what is beneficial and desirable for freelancers. The second is *adequate work*, whose elements encompass valuable projects understood as medium-long term projects where freelancers can work in, and positive labour economy that comprises the concept freeworkers – freelancers, experts working, sharing knowledge and expertise within the companies. Those two non-economic aspects were identified relevant for the startup, but at first glance, the analysis showed that the use of AI is not needed to achieve those. Precisely, that prompted questions of what aspects are considered in the use of AI, hence the subsequent step was to analyse those.

Two aspects were identified as the ones considered by the startup in the use of AI, one is *product optimisation* that entails optimising processes within the product/platform including detecting needs of freelancers and companies, and recommendation of freelancers to projects and projects to freelancers. The other aspect, *customer experience* comprises creation of features to facilitate tasks for freelancers and companies. The analysis showed that *product optimisation* and *customer experience* are interdependent with the constraint factor that the startup decides whether profits can be made by weighing up the two aspects. Further analysis revealed that within *product optimisation* there are elements related to one of non- economic aspects -*adequate work*. That suggests that *adequate work* is considered in the use of AI, but not the other aspect *talent first*, which questioned if that aspect claimed by the startup as one of the most relevant for them is truly taken into account or not. Subsequently, the analysis of high-risk AI was done with the emergence of three aspects: *performance of the matchmaking algorithm*, *data*

that includes data curation and information collected in all stages, and *regulation* that entails monitoring, expert knowledge and delay in execution. In addition, ´risk´ emerged as what might be a cross-cutting topic in the analysis of high-risk AI.

The examination showed that non-economic aspect - *talent first* that was initially missing, it is certainly taken into account specifically in high-risk AI for the *performance of the matching* and for *data* aspects, and the other non-economic aspect *adequate work* is also considered. However, I was still questioning whether those aspects that the startup claims are the most relevant to their practices are in fact the most relevant. The compendium of findings and analyses in the section brought me to finally elucidate the ´triad of positions´ as follows.

5.2.4.1 Valuee

The research sheds light on what is mainly considered in the valuation of AI is the performance of the recommendation algorithm. It encompasses protocols, explainability metrics, similarity elements, measuring point and validation elements that in turn comprises data curation, interests of freelancers, expert knowledge, optimisation of the product, tasks for freelancers and companies, customer experience. It constitutes what is really considered by the startup.

> I think that [the matching] is the most critical and most important thing in [the startup]. In the end we bring together people from one side with people from the other, and there is something in the middle that must bring them together, if that tree of the middle doesn´t work, doesn't matter if you have something here and maybe there, if you're not able to connect the dots, you're not going to get it. (Product Manager, who has played different roles since the beginning of the startup)

5.2.4.2 Audience

The research revealed that the product that will most likely be considered high-risk AI is to whom the valuation is truly addressed. Through the analysis there was a continuous question about the aspect *talent first* that emerged as a non-economic aspect. Hesitation of why it was not apparently considered in the use of AI motivated further analysis because *talent first* refers to the freelancers, who are the most important value for the enterprise according to interviews, documents and website. Therefore, it seemed unlikely that *talent first* and its entangled elements for the benefit and desire of freelancers, were not taken into consideration in the use of AI. At the end, the research enlightens that *talent first* was intrinsically part of the high-risk AI within different elements of *performance of the matching* and *data*. Based on that, it seemed that *talent first* -freelancers- is the audience of the valuation. Although, endeavours are led to

impact the community of freelancers and the future of work, the underlying audience that came out from the research is the high-risk AI product.

The valuation of AI is truly addressed to that product, the use of AI by the startup is focused on the high-risk AI product. The non-economic aspect *talent first* focuses on tasks and features on the platform-product that are beneficial for freelancers. The other non-economic aspect *adequate work* emphasizes valuable projects, expert freelancers, freeworkers, and working model of no long-term attachment, those for the operation of that platform/product. Furthermore, the aspects considered for the use of AI that were identified as interdependent- *product optimisation* and *customer experience* - focus on processes within the platform/product, and automation of tasks in the platform/product to make easier activities of freelancers and companies. The aspects taken into account for high-risk AI constitute what the product is.

> "All the uses we make of artificial intelligence are very oriented to our … product" (AI specialist)

Moreover, considering that the core of the startup is the platform/product, whose aim is performing automatic matching to connect freelancers with companies, that is consistent with the value of the valuation - the *performance of the matching*- and the audience to whom the valuation is addressed - the high-risk AI product.

5.2.4.3 Valuator

All previous analyses have illuminated the factors of the valuator in the valuation process of AI. Valuator encompasses positionality of the startup regards valuee and audience, that comprises the understanding and position about AI, benefits, downsides and risks, implications, considerations, and management of that technology. Insights in the course of the research has indicated that AI represents an enabler of value the startup seeks for the business and for the publicly manifested target audience (freelancers). That lead the startup to transition from other technologies to AI. Regardless the risks entangled in AI, the value expected in terms of optimisation, benefits, market differentiation, so on and so forth, is promising based upon what has been reached so far. That encourages the enterprise to continue the ride of startups lifecycle, but the aim they have is still not achieved

> …, even if we wouldn´t grow in demand, it [the matching] always must be better, for me it seems to be the most difficult. For me, it [the matching] is something that no one has

> still been able to solve through technology without human intervention. If we are able to solve it, then we stop the game... (Product Manager)

5.2.5 Conclusion

Having presented an analysis of the entry point in the research of valuation process- the ´triad of positions´ - I can conclude this section making explicit what the analysis enlightened in regard to how the startup conceive high-risk AI.

The assigned valuation of AI signals how the startup conceives high-risk AI. The value ascribed to AI has shaped the business goals, target audience (freelancers), organisation itself, and practices around the sector where they operate. The valuee of the valuation identified in the research– *performance of the matching algorithm*-, and the audience towards the valuation is addressed – *the high-risk AI product*-, those entangle meanings, purposes and practices that have been addressed in the research. The valuator entails positions regarding that product such as, the main aim has still not been achieved, which means not able to do the matching without human intervention, as well as the risk associated with that product.

> [Product name] is a platform, so to speak, unitary. Although it has steps from where you enter until you leave, it is a single product. We will probably be considered a high-risk application due to the context in which it is produced, which is the work context. The matchmaking … where … the regulation I think is going to apply, … is when we recommend talents to projects or projects to talents. (AI specialist)

The factor of risk, according to the EU AI Act, "'(2) 'risk' means the combination of the probability of an occurrence of harm and the severity of that harm". (Article 3: Definitions -Artificial Intelligence Act, Official Journal version of 13 June 2024). Despite of the factor of risk seen in different forms, such as the design and quality of data sets in the system and outputs of the algorithm, and the implications that it represents for the startup in terms of measurements to manage risk for instance, protocols and human oversight, and now the regulation, they conceive high-risk AI as their core. The probability of harm exists as the uncertainty is created. The startup is developing an AI system that entangles risk; hence, uncertainties are produced in doing so; nevertheless, internal processes, departments and business are being shaped by the value the startup has assigned to AI. The risk exists, but the driving force to continue is how they have conceived that technology because of the value they assigned to it. They conceive high-risk AI as the core of what they are, what they do, and as an enabler of impact they want to have in the society. This implies a need to interrogate what they pursue and how that might impact the society. That will occupy the following section.

5.3 In the pursuit of impact

The first version of the product was created based upon desires and necessities of the founders of the startup, however, the product and the target evolved into a vision that has been built hand in hand with technology to constitute what the startup is today for: a product that will most probably be in the category of high-risk AI, freelancers, companies, projects and the proposal of reinventing work. What has been identified as the pursuit in the development of high-risk AI is to match freelancers with companies to develop projects, taking into account technical knowledge, skills, motivations, and cultural fit. That encompasses different roles assigned to that technology, those addressed in the first section of this document, and the valuation of high-risk AI that constitutes the driving force of the startup, analysed in the previous section. The question now is how what the startup pursues with high-risk AI impact the society.

Impact of technologies on society has been analysed in different sectors with expected and unexpected results. The aim of this section is building on what has been emerged from the previous sections to perhaps uncover impacts of high-risk AI that have not been analysed. Attempt to understand impacts encompasses asking first questions about the technology, the impact and the society, and if the aim is also to suggest actions, then interrogate about solutions. Those questions refer to the STAP attributes of the framework of Baek & Lee (2021). As was indicated in the chapter Analytical Framework, for this section some basis notions of the 'Socio-Technology Analysis and Prescriptions (STAP) framework by Baek & Lee, will be considered and adapted to the research.

5.3.1 Impact framing

In the first section of the present analysis was identified how the startup began to embark themselves in the freelancing industry and how AI enters at play. The co-founders started hiring freelancers specialized in specific topics to develop those projects that they [founders] could not develop because of the increasing number of requests, and to some extent, the lack of motivation to work on projects that were not so much in their interest. They decided to create a technological product to make scalable the process of looking for freelancers and hiring them to develop projects for different customers. To that end they developed the match between freelancers and customers with artificial intelligence. Currently, even though freelancers have been matched with companies, projects have been developed, and the startup is advancing in the entrepreneurial journey, according to the startup, what they pursue has not yet been realized.

> …other platforms have their algorithms that are surely much more evolved than ours, but none of them take all variables into account, … If we manage to make that within 200 profiles capable of doing your project, we highlight the 3 that do it best, then that is where I think we will stop the game. (Product Manager)

The aim of developing high-risk AI to connect freelancers with companies remains, but there is more pursuing: consider all variables -technical knowledge, skills, motivation and cultural fit, and recommend the best three out of a candidate pool. That has been shaped by the context and the technology used for. First, concerning the context, any other company or startup have been able to automate processes of connecting freelancers and companies taking into account all variables together and making decisions. The startup of this research is apparently the first that has advanced in that matter, but not fully automate it. The requirements for automation, optimisation, customisation, and decision-making that characterized technological developments do not change what the startup pursues but make it more challenging to reach in an increasingly competitive environment. Second, regarding the technology, nowadays, to reach what this startup pursues with the aforementioned requirements means to embrace AI. The startup has adopted AI and built its purpose around that technology, but AI entails risks in different forms mentioned in the previous session, and like any other technology, AI is shaped by the context and impacts the society.

The aim of the match encompasses directly freelancers and customers that have projects to develop, those are the specific group the startup has had in mind to influence with the AI product, emphasizing on freelancers. They have reached both successfully "… We have a repeat rate, half of the people [companies and freelancers] who start working with [us] decide to continue" (AI Specialist). They mentioned the impact is limited to the users of the platform not further because it is a closed platform, "Our platform is closed and only the people we allow have access to it... only the people we validate can enter" (Product Manager). Beyond questions, it is a closed platform or a closed community, but what if the business model change and the reached public is broader? Startups are keen on exploring business models so that could happen but let only consider the present. According to the official website, at the end of 2024, the startup has around 3.500 validated freelancers in the platform, and more than 400 companies work with the startup. Hence, the startup has reached with the product the groups they desired, emphasizing on freelancers. Significant or not those figures represent a group of society that is reached with the AI product.

The outcome is going further. Throughout the research it has been highlighted the term ´freeworkers´ and the focus on supporting a working model of no long-term attachment neither to one company nor one employee. According to the public knowledge base (n.d.), usually ´freelancers´ and

´freelancing´ are terms comprising connotations such as workers only for specific tasks, and cheap, unmotivating and temporal projects. That is a problem from the perspective of the startup because of the entailed limitations for freelancers and the potential of their businesses. Based upon that, the startup coined its own term ´freeworkers´.

> When we define the worker of the "future", who enjoys full flexibility, is delocalized, super-specialized, works by milestones, etc., we are constructing a true portrait of the freelancer. **Profiles that**, for the most part, **at the top of their career, decide to bet on their talent and develop it to its maximum potential by working on projects that motivate them**. (Public knowledge base, n.d., emphasis included in the original quote)

Likewise, the startup creates the term ´freeworking´ explained in the same online resource, as hiring fast, project work, efficiency and agility in projects, build problem-solving teams and hybrid teams, flexible work, offshoring, maximising motivation of freeworkers, and access to the best professionals -freeworkers.

> "It's simple, whoever wants to have the best talent in their teams will have to jump on the freeworking bandwagon…". (Public knowledge base, n.d.)

That assertion and the characteristics of freeworking and freeworkers mentioned above entail considerations, such as best talent is equivalent to freeworkers, and freeworking is a must-have for companies. Similarly, the statement of freelancing as a "mindset of the past" and freeworking a "mindset of the future", publish on the public knowledge base (n.d.), entails considerations of freelancing as something that happened or existed before now, and freeworking as something that will come, happen or exist. Those considerations are not ideas or statements that just came out to add value to the market, those have been conceived, developed, and evidently mobilized through public spaces. Why the startup decided to conceptualize and name freelancers and freelancing differently is not the interest of this research, rather analyse whether, in light of the impact on society, AI has a role in framing those concepts and considerations and the public reached.

The startup has had a specific purpose since the beginning "…we were born, …to generate an ecosystem where talent can have decent work, adequate work and can have this experience of being freelancer for as long as they want" (AI Specialist), once they started to adopt AI, that specific aim became broader and the impact pursued when developing high-risk AI has been shaped. Connecting freelancers with companies whose projects are interesting for freelancers has been the aim and still remains, but now is part of a broad vision.

> "[name of the startup] was born with the vision of reinventing work. We imagine a world where companies and digital workers can connect in a much faster and more efficient way to develop high-impact projects" (Public statement of Co-founder)

For the startup, the motivations and specialized profiles of freelancers have been contemplated since the beginning as part of the aim of developing the AI product, but the conceptualization of freeworkers and freeworking and the vision have been developed at pace of the evolution of the startup and the AI product. Analysing the characteristics attributed to freeworkers as super-specialized profiles that works on projects that motivate them, it can be considered that high-risk AI plays a role to enhance those because super-specialized profiles is a key element of the matching functionality. The same for the concept "freeworking", specifically, the consideration about maximising motivation of freeworker because, as it was analysed in the previous chapter, the motivation of freelancers is one of the most relevant aspects of the matching. The startup manifests that freeworkers are super-specialized workers that work on projects that motivate them, and freeworking is hiring fast, maximising motivation of freeworkers, and having access to the best professionals. How could the startup conceive, develop and mobilize these concepts if it did not have a product that does that? It is not pretentious to assert that somehow the AI product enables those concepts because it suggests the best workers for developing projects based on skills, motivations and cultural fit with companies, making it possible to hire faster. With the matching functionality developed with AI the startup proactively and deliberately aims to bring about impact in the freelance work sector, but there is impact beyond the scope of the startup.

5.3.2 Impact beyond the startup scope

As it has been mentioned throughout the analysis, the key characteristic of the matching is not only to match skills and technical knowledge, but also motivations and cultural fit. That might impact the recruitment and selection processes not only for freelancing but for other kind of jobs. Currently, in the market there are tools that read curriculum vitae identifying knowledge and skills, filter candidates based on the criteria introduced on those tools and score candidates. These generative AI recruitment platforms are used in the recruitment and selection process providing shortlisted candidates. These perform similarly to the matching developed by the startup that filters the most appropriate candidates and recommends the best three, but there is a difference that impact the usual hiring process. The task of identifying affinity of candidates with the culture of companies and motivations of candidates to work for a company or develop a project and analysing and selecting people in light of those factors are tasks usually done by human resources professionals after the outcomes of the AI recruitment platforms. However, the matching developed by the startup already filters based also on cultural fit and motivations before providing the three best candidates. Certainly, from the people suggested by the matching, the customer carries out interviews and select one. Nevertheless, it does mean that now in the hiring process of freelancers (in the future could be any other employee not only freelancers), human resources

professionals have the option of selecting but only at the very end and they rely on a previous selection of an AI system.

> …obviously the customer can invite a talent [freelancer] that has a 30% match, but the customer is going to invite … the three that have the most match. By far, if the customer is curious and sees the profiles will invite the fifth, the sixth (AI Specialist)

Besides, there is an impact on the accuracy and reliability of the selection process. The more customers and data, the most difficult is to be able to identify failures and errors of the algorithm. Hence, there are possibilities that the performance of the matching may not be fully accurate also with biases in the recommendation of three candidates over other people. The possibility to identify that may be complex for a human. If the system provides that recommendation whereby the company selects one person, then the final decision will be impacted by that and if not accurate and/or biased may not be easily identified, corrigible or reversible. There are mechanisms to prevent or substantially minimise risks or ensure bias detection and correction, but it does not guarantee the accuracy and reliability of the performance of the matching, therefore, exist an impact on the selection process.

Another impact is the role of humans described in the previous sessions of the present document, specifically for the performance of the matching where there is a considerable degree of human involvement in different stages. Beyond the team of Data and AI, purposely the startup involves people from different departments and expertise to support testing and validation of changes on the matching to oversight outputs and to identify failures. How companies handle and manage AI internally varies, but what is remarkable from this case-study is the involvement of humans and the role they play regarding high-risk AI. The startup involves people to validate, evaluate and oversight based on the expert knowledge, that refers to the knowledge that employees and even customers have of different tasks that are intended to be done by high-risk AI, in this case by the matching. Those practices might influence processes and structures of SMEs, startups or companies that develop this kind of technology. Clearly, high-risk AI management requests more than only 1 person and 1 department, even external people such as customers might be involved. Besides, the tasks assigned to humans regarding high-risk AI includes checking the accuracy, assessing the outcomes and the impact, and supervising and monitoring.

In that regard, knowledge of people plays a central role in the context of high-risk AI. Apart from the knowledge translated into data and introduced to the AI, knowledge is used to determine the limits or boundaries of the high-risk product. It encompasses knowledge of people about the aims of the developed AI solution, of processes and tasks addressed by it, and knowledge of the technology itself.

> …validating expert knowledge of what people do as an easy task [e.g. identifying failures in the recommended profiles], but which is very difficult for machines. And when we

> release into production [changes of the algorithm], for each recommendation we have the opportunity for the talent [freelancers] to tell whether the recommendation is appropriate or not … And those signals are introduced to retrain the model and see exactly what could have been happening… let's say, the way to delimit it [the algorithm] and somehow control it, it is knowledge. (AI Specialist)

What the startup pursues with the development of high-risk AI entails also a particular profile to lead such task. It is relevant to recall that at the beginning of the entrepreneurial different attempts did not work as expected to accomplish the idea of connecting freelancers and companies to develop projects taking into account skills and motivation. Then, with the arrival to the team of an AI specialist with studies in psychology has contributed not only to develop a system based on artificial intelligence, but also to include aspects of personality, preferences, and motivations of people.

> …we kept adding things and changes to improve the matching at a level that we call cultural fit and motivational. So, well, … we did two tests, one of personality and one of motivation… And since [name of AI specialist] also has a PhD in Psychology, well, he added his whole layer of psychology on top of it… at the beginning none of us were experts in anything…now, that test is done and validated, that is, it has a lot of science behind that we didn't have before. (Product Manager)

For the specific aim this startup pursues, it appears ´normal´ to have an AI specialist with studies in Psychology, so if that is the case of one startup, which is a small business that has just been started recently, then arises the question if that would become ´normalize´ for other startups, SMEs and companies developing AI platforms whose data is about motives and thoughts of humans. Besides, there is another factor which is the experience of the AI Specialist in academia that seems to impact how the regulation might or not affect the way of addressing AI by specialists or leaders of AI departments in companies.

> It's hard to say, in my case, I think the impact is going to be relative… I come from a university background, so, I am used to a series of quality and rigor processes that are not relatively similar, but they follow a bit of the path that is required when designing transparent systems and designing responsible systems. (AI Specialist)

This case - study brings up the scenario of professionals of natural and social sciences along with experience in academia as the ones that might be considered when developing high-risk AI systems. Needed to mention that the presence of an AI leader with the mentioned profile does not guarantee that AI will be free from bias, privacy concerns, lack of algorithmic transparency, and ethical responsibility. Analysing the profile of the AI Specialist contributes to inform what profiles startups are taking into account to develop high-risk AI.

On the other hand, there is an impact on the information that people must provide to systems. For the matching it is requested to incorporate motivational aspects and interests of people to be taken into account to recommend three candidates for a job and not only be based on the skills. That means that to be included in the job data base and maybe to be considered for a selection process, people provide information to a system about what motivate them and what interest they have. That is not sensitive data, but still, it is individual data that normally people could provide or not, however, for the matching that information must be provided. Likewise, it creates a necessity of informing to system individual changes of likes or dislikes, a sort of updating. For instance, if a person feels motivations, likes or interests different from the ones that she/he introduced initially to incorporate in the algorithm, it seems that those changes of feelings should be informed, otherwise, the matching will perform with “old” data and that will impact the consideration or not of that person for a job.

5.3.3 Conclusion

This section is an attempt to comprehensively identify and analyse how the startup´s pursuit of high-risk AI impacts the society. As is evident, the impact extends beyond the scope of the startup when developing high-risk AI. What startups pursue when developing any kind of technology might vary from one enterprise to another, and subsequently the specific impact on society. Once they move forward in the entrepreneurial journey validating their business models there is an overarching aim that startups pursue, which is to bring about changes in a broader society. Analysis of this research case study has revealed that by developing high-risk AI that overarching aim is feasible, but the impact is not limited to what the startup pursues.

In the course of developing what will most likely be high-risk AI the startup has been impacted itself (Micro- impact scope) in its internal processes, business goals and also team member profiles, such as the profile of the AI Specialist to lead the development of the product. Since the beginning the startup proactively and deliberately has pursued to bring about a specific impact in the freelance work sector (Meso- impact scope) with the matching functionality developed with AI, specifically on how freelancers and companies connect each other to develop projects. Then the startup moved to pursue a broader impact, reinventing work (Macro -impact scope), which entails freeworkers defined as the best talent, freeworking a way of work that will be a must-have for companies that need and desire the best talent to develop projects, it also comprises fast hiring and other characteristics mentioned before. The startup has framed that vision with the adoption of AI and the development of high-risk AI, that technology became for the startup an enabler to pursue that broader impact in society. In addition, there are impacts on the recruitment and selection processes, the role of humans to handle high-risk AI, the particular profiles to

lead high- risk AI adoption and development, and impact on the information that people must provide to systems.

6. Discussion and Conclusion

The interest of my research is to attempt to understand empirically how high-risk AI is made, what is its meaning and elucidate its impact. From a personal impression, when the proposal of the regulation of AI in EU was discussed and even now that the regulation is on placed, I noticed that I simply assumed a lack of understanding of what seemingly other people understand clearly regarding high-risk AI. That makes me think of Suchman´s observations

> While interpretive flexibility is a feature of any technology, the thingness of AI works through a strategic vagueness that serves the interests of its promoters, as those who are uncertain about its referents (popular media commentators, policy makers and publics) are left to assume that others know what it is. (Suchman, 2023, p. 3)

Perhaps, policy makers, popular speakers and companies or SMEs that produce that technology know or have insight about high-risk AI, but I do not, and I am not sure if general public understand it.

For the empirical research I addressed three research sub questions encompassing concerns and needs taken into account to develop high-risk AI, understandings and uses of this technology by the startup of the case-study, as well as what that enterprise pursues and how that impacts the society. In this section I discuss the most relevant findings of the research in light of contributions of the Literature Review and Analytical Framework presented in the thesis.

6.1 AI not grounded on societal concerns might become high-risk to society

One of the first seemingly obvious findings of the case study was that at the beginning of the entrepreneurial endeavour, the identification of problems, concerns and necessities of society were not factors for the development of the AI product. Certainly, as Gerdes (2021) pointed out, the industry is guided by commercial interests paying few attention to societal issues. Nevertheless, it does not imply that no one was considered, it is relevant to recognise that entrepreneurs create products or service in response to desires, problems or necessities that someone has. The point here is that society as a whole, communities or broader societal issues were not taken into consideration in the initial stages. Instead, there were considered only the necessities and ´problems´ of three people (the co-founders) to produce the first version of the AI system. This situation is not new but sometimes uncontested. As Suchman (2023) questions, because of the apparent agency and inevitability of AI most critical debates of AI are not addressing basic questions such as “What is the problem for which these technologies are a solution? According to whom?” (2023, p. 4)

For whom and for what are produced AI services/solutions do not appear to represent a barrier in moving towards commercial interests and adapting the product to meet society´s needs, request or expectations. As Gerdes wrote "problems increasingly enter the surface and catch public attention, the tech industry tries to mitigate them. But not before" (2021, p. 4). While this is a usual situation, what is important is to acknowledge that although AI products/services that are not grounded in societal concerns or necessities might also become high-risk AI. In this regard, the empirical analysis reveals that multiple factors influence such as roles assigned to AI, roles of humans in relation to AI, valuation ascribed to AI as a pivotal element in how the startup conceives high-risk AI, the aims pursued and the impact.

6.2 Startup shapes AI and public perceptions around it

Based on the empirical outcome is evident that the startup shapes AI and perceptions around it. In their approach of machine vision from a sociotechnical perspective, Hsiao & Shorey (2025) discuss issues related to the significant role played by promoters and designers of AI in shaping use and social consequences of technology. With the matching functionality, most probably high-risk AI, the startup assigns uses of AI that deliberately seek to bring about impact in the freelance work sector, and incidentally, purposely or not, impact on hiring process, information required for AI, and role of humans in relation to AI. Similarly, regarding visions of companies, Hsiao & Shorey emphasize that "their visions act as primary agents that organize public conceptions of emerging technology" (2025, p. 405). While it cannot be said that a startup in its early stages impacts public perceptions, its vision is certainly an initial factor to build upon. The public vision articulated by the startup in question is based on the notion of reinventing work. This entails a new concept named freeworking where fast and effective connections between companies and freeworkers to develop projects are possible with AI. Investors, customers (freelancers and companies) and employees are involved in that construct, which serves to frame their conceptualization about that technology.

Terminology created by the startup and conceptualization that people have framed around the system that would be most probably high-risk AI, entail narratives that Sartori & Theodorou highlighted might reflect visions for society and dominant social, economic and political inequalities, "they reveal important glimpses on the existing structure of inequalities that, through an increasing reliance on formalized algorithmic and AI techniques, could be automated" (2022, p. 7). In the present case - study, we are observing the relevance of narratives in preserving main dominant interests while shaping reality of the freelance job with new narratives. On one hand, efficiency, project work, flexibility and agility in projects are structures that reflect traditional lines of narratives of the freelance job present on the widespread message of the startup. On the other hand, motivations, hiring fast, super-specialized

professionals, and access to best talent, are factors that are building an own narrative embedded within the AI system through its processes, data, validations, automation, and so on, and within the social context of employees, investors and companies looking for workers, and the freelancers. Indeed, as it was found in the empirical chapter, somehow the enablement of those narratives come from the development of the AI systems, and those have been materialized with the same AI product. The startup has conceived, developed and mobilized those narratives because the AI system they produce is made to take into consideration skills, motivations, cultural fit with companies, delocalization of the job, project work, among other factors that are the core to match freelancers with companies, hiring faster and develop projects. The perception of those narratives by customers (freelancers and companies) should be subject of future research.

6.3 Decisions based on data that does not necessarily reflect real life

Technology entails impact that is not necessarily straightforward but instead diverse and complex (Baek & Lee, 2021). In this sense, one of the possible impacts identified in the research is related to data. Specifically, the information that seems is going to be required to update regarding changes in personal preferences, motivations, likes or interests of freelancers. That kind of data is necessary for the matching purposes, the algorithm will work with the information introduced initially, unless it is updated regularly. That might impact who is recommended by the AI system to carry out a project since considerations of freelancers for vacancies are determined by that data in the AI system. Furthermore, what would happen is a simplification of those personal feelings into information that cannot vary so much. It means, freelancers will not be allowed to enter different motivations, likes or interests from one vacancy to another simultaneously, only options to change or update those from time to time. That founds an argument in what Suchman (2023) highlighted as the process of data reduction or cleaning information that does not fit with the only purpose to translate social practices into statistics. In real life, one person can have different motivations and interests from one job to another because of multiple factors such as passion for the topic, salary, status, experience, necessity of having a job, so on and so forth. In the real life, when applying for a job, in many cases people adapt their motivations, likes and interests to the vacancy.

> 'Data is made by people going around and counting things or made by sensors that are made by people. In every seemingly orderly column of numbers, there is noise. There is mess. There is incompleteness. This is life'. Yet dirty data confounds reliable computation; anomalies must be cleaned up to make functions run smoothly, and in that process, the irremediable contingency of signification disappears. As is now widely recognized among science and technology scholars, categorization is performative in that it works to write itself in and through the worlds that it orders. (Suchman, 2023, p. 3)

Apart from the impact emerging from high-risk AI functionalities that comprises process of data sorting, placing information into groups or categories that shape sectors, like the job search sector of this case-study, there are other related impacts on understanding decisions made by AI and the possibility of question those. Transparency and explainability of AI certainly contribute to understand why an AI system decides what it decides. However, as Sartori & Theodorou emphasized, to question a decision "requires us to understand and review a decision made by a system within the socio-legal context of where and when it was taken" (2022, p. 5). In the empirical chapter was identified that understand and contest decisions made by high-risk AI is a challenge for the producers, hence, may be even more difficult for customers since they do not know about the multiple variables taken into account.

As Sartori & Theodorou (2022) point out, complex AI is linked with different stakeholders, organisations, norms and ethical, social and cultural values. Therefore, to understand and have the possibility to contest AI decisions, as the authors argue, it is needed to consider perspectives of all stakeholders because their interpretation of values might differ. Also, focus the efforts on attempt that any definition is explicit and transparent, and not to try to a ´one-size-fits all´ definition for values. Most importantly, the authors mentioned that computer science can offer tools to formally represent value interpretations, and the participation of different people from different disciplines is required to understand the context in which an interpretation is valid and reliable. In light of the case study, the startup, specifically in the performance of the matching, has integrated at a certain degree metrics, measurements, employees from different departments of the startup, rules within the system, customer experience and expert knowledge. Nevertheless, understand and question decisions made by high-risk AI constitutes a complex process "we have to go back and investigate what happened. …and it often happens, especially with black box models, where you don't explicitly control the weights or what is happening or could happen inside, so you have to go to the specific case…" (AI Specialist, participant of the research). Decisions of AI systems as decisions made by humans are context dependent, hence, human involvement as inputs and analyses from different perspectives should be a must – have for producing AI.

6.4 Human involvement and knowledge

When analysing impacts is also relevant to identify the involvement of human beings in the production of high-risk AI and what it entangles. The research of Gerdes (2021) about the pathway to artificial general intelligence (AGI) provides an example of how the required and significant human role in AlphaGo to defeat the world champion of Go in 2016 was not the highlighted point in the hype around that, but the ´self-learning´ of the AI system. Similarly, Jaton & Sormani (2023) address what they name ´commensuration practices in the emergence of 'AI' and the invisibility of the work necessary for that

commensuration - the state of having common factors with society and the ´real´ world. The authors provide examples of different sectors to show that AI is enacted through shadow activities that require human intervention such as training algorithms, automation, environments for trials. The present research identified a considerable degree of human involvement in different stages of the production of high-risk AI beyond the design. For instance, before releasing new version or new features, those are validated by the departments of sales, talent, product and customers. After launching feature improvements or new functionalities, people from product department supervise the performance of the AI system. The outcomes of the AI system regarding the recommendation of three profiles are evaluated by freelancers and the matching accuracy by companies after interviewing the three people. Likewise, continuous analyses are done by the AI Specialist, and the data curation by the data specialists. Certainly, no one denies that producing AI requires human action, but the amount or level of that involvement and which forms it takes might be explained and acknowledged, among other things, to understand what high-risk AI means.

Considering that involvement of humans regarding high-risk AI includes checking the accuracy, assessing the outcomes and the impact, and supervising and monitoring, it is clear that knowledge is essential to do that, that is knowledge that people have. Knowledge about the technology, the tasks, the processes, the aims…"expert knowledge of what people do as an easy task, but it is very difficult for machines" (AI Specialist, participant of the research). I argue that knowledge that people should acquire is most importantly about AI societal implications. Jöhnk et al. clarify that companies need to implement protocols to avoid unethical AI outcomes. Otherwise, organisations will be exposed to rely on biased learning and input data and face liability for discrimination. The authors mention as an example that "gender bias in data sets for AI hiring tools can result in biased candidate selection" (2021, p. 13). In that regard, the startup of the case study explains "Gender is something we didn't ask about …. However, we're now ask about it to monitor how many men and women we recommend for positions and to take necessary action where we can observe gender gaps" (AI Specialist, participant of the research). In addition, as it was addressed in the empirical chapter, the startup has protocols, set metrics, monitor the system, and establish other processes and practices to develop, "design transparent systems, design responsible system" (AI Specialist, participant of the research). All of those practices undoubtedly contribute to AI ethically ready companies (Jöhnk et al., 2021). Nonetheless, it is relevant to remember that biases are not inherent in AI, those exist in society (Sartori & Theodorou, 2022). Therefore, people as social actors might identify inequalities, discrimination, understand the social impact of producing AI and attempt to address those. "Not only should there be proper training for developers, but also for all

other stakeholders; including users and the general public that is indirectly affected by the technology" (Sartori & Theodorou, 2022, p. 6).

6.5 AI performing the role of recommender

The roles identified in the empirical chapter show that the startup has assigned roles to AI for identifying problems and necessities, solving problems, predict aspects of projects such as the number of people needed at a specific time, and forecasting for example revenues of a project. Furthermore, and most importantly, it was identified an assigned role for recommending, which is named by the startup as an ´intelligent assistant´ "The AI system is designed as an intelligent assistant that allows: a) organizations to assess and rank talents based on their project fit; b) talents to assess and rank projects based on their project fit" (Internal document of the startup).

According to Ma et al. (2024) despite the other roles that AI could perform in decision-making such as analyzer and devil´s advocate, the focus is on the role of AI as recommender mostly. In the case-study one of the outcomes of that role, and perhaps the most relevant and controversial is the recommendation of three people, preferably to hire one of them, for developing a project. That outcome comes from different factors considered by the matching functionality that will most probably be high-risk AI. In an internal document the startup mentioned that "The AI system facilitates and enriches the decision-making process but does not respond to a closed system where the user gives up control over the final decision". However, as it was analysed in the empirical chapter, provide a recommendation of three people over many other candidates even though is not the startup in charge of hire any of them, it certainly influences the decision that a company will make.

It is also relevant to point out what the research of Ma et al., showed when AI performs mainly the role of recommender, "one notable issue is that individuals, when passively receiving AI suggestions, seldom engage in analytical thinking. Furthermore, people frequently inappropriately rely on the AI's recommendations (such as over reliance and under-reliance) and the mere provision of AI explanations can, paradoxically, exacerbate overreliance". (2024, p. 1). In the present study the effects of the recommender role of AI on user experience are not analysed neither the accuracy of the AI system or the confidence of companies in the recommendations, but based on findings of the empirical chapter seems that users are engaging with the AI suggestions "… We have a repeat rate, half of the people [companies and freelancers] who start working with [us] decide to continue" (AI Specialist).

Furthermore, according to Ma et al. (2024) advantages and limitations of AI roles are influenced by how are integrated into decision-making processes and by the performance of AI. While the

performance of the AI system is not subject of this research, in the empirical section when analysing how the startup conceives high-risk AI, the performance of the recommendation algorithm was identified as the valuee of the valuation of that technology, which means that performance is the main aspect considered in the valuation of AI. Moreover, the product which is the AI system was identified as the audience to whom the valuation of that technology is addressed. Therefore, all the efforts are towards maximizing the role of AI as recommender with its advantages and limitations. This implies a need for subsequent research to be conducted on the influence of this role on decision-making of companies selecting people for a job and the dependency on AI systems.

6.6 Profiles to lead AI development

Another outcome of the empirical chapter was the kind of profile the startup of the case-study takes into account to lead the development of high-risk AI. An AI Specialist with studies in psychology and statistics, and experience in industry and academia. Such profile has enabled the startup to include, apparently successfully, key factors on the AI system, which are aspects of personality, preferences, and motivations of people. Also, as the own AI Specialist mentioned, his profile has contributed, at least at first glance of the regulation, to meet what would be required in designing a transparent and responsible system. As Baker-Brunnbauer (2021) highlighted "Diversity is one factor to reduce bias" (p.178). To have professionals of natural and social sciences along with experience in academia leading the development of AI does not guarantee bias-free, ethical responsibility and transparency. However, considering that part of the ethical responsibility of AI system lies on the producers, to have people with different studies background is required to deal with social and technical complexities of AI (Baker-Brunnbauer, 2021).

6.7 Valuation process of AI

A core theme that lies at the heart of the empirical chapter is the valuation process of AI. My interest was to understand how the startup conceives high-risk AI, that brought me to analyse the valuation of that technology by the startup, hence I attempted to reconstruct the valuation process around it. Based upon valuation studies in STS, the analysis focus on the valuee, the valuator, and the audience. Those three constitute part of the first component of the concept "valuation constellation" proposed by Waibel et al. (2021) - triad of positions and relations-, which is a key entry point in the research of valuation but normally assumed rather than analysed.

The research elucidates that valuee, valuator and audience are in one way or another, intrinsically tied to the AI product. First, the main dimension considered by the startup in the valuation of AI is the performance of the recommendation algorithm, that represents the valuee. Second, surprisingly, the

audience to whom the valuation is truly addressed is the AI product, not the freelancers which is the target audience manifested by the startup publicly. The third element of the 'triad of positions' -valuator encompasses who evaluates, which role and positionality regards valuee and audience, which in this case- study means the positionality of the startup regarding the AI product they are developing. The analysis shows that the startup considers downsides and risks of AI, the probability to be categorised as high-risk AI in European Union and the implications of that, nevertheless the startup conceives AI an enabler of value.

The assigned valuation to high-risk AI constitutes the core of what the startup is, including business goals, target audience (freelancers), processes, and also encompass impact on practices within the sector in which they operate. In the research of valuation was identified and analysed how the startup transitioned from similarity algorithm, AI to high-risk AI, also the aspects considered by the enterprise in its business beyond the economic interests, aspects for the use of AI, and specifically for the deployment of high-risk AI, as well as the scope of risks associated with AI and how the startup works to address those. The worth attributed to high-risk AI is the driving force of the startup even though the AI product they develop might represents risk for the society. Furthermore, that ascribed value to high-risk AI constitutes the fuel for the vision the startup has now. Connecting freelancers with companies whose projects are interesting for freelancers has been the initial aim and still remains, but now that is only one part of a broad vision of reinventing work. That vision has been shaped and being enacted by high-risk AI. Thereby, this research of the ´triad of positions´ of the valuation process of high-risk AI sheds light on "how valuing enacts and provokes something (new)" (Asdal et al., 2024, p. 80). This might be a motivation to consider future research to analyse the whole ´valuation constellation´ of high-risk AI.

6.8 Concluding Thoughts

Concerns about the black-box ´nature´ of AI has increased research and policies to open and understand this technology. Insufficient information regarding how AI systems work can lead to make limited and uninformative interactions of people with the AI, as well as, over trust or little trust on this technology, and disuse or misuse of it (Sartori & Theodorou, 2022). With this research I attempt to contribute to reduce the black box character of AI, specifically about the recent introduced category of high-risk AI in the regulation of European Union. The aim was to analyse how it functions, who participates in its development, roles, considerations taken into account and other aspects entangled with high-risk AI. As Ehsan et al. (2021) point out, to understand how people make sense of a system "is not just about opening the closed box of AI, but also about who is around the box, and the sociotechnical factors that govern the use of the AI system and the decision" (p.2). Likewise, to open and examine black box AI is not only about making transparent and explainable AI to understand why an AI system produces specific

outputs or how it makes decisions, but opening black boxes is also about how AI is built and then communicate that to the broader public to raise awareness.

Based on the empirical findings this research demonstrates that the startup of the case-study shapes AI, and the startup is shaped by AI as well as the society. Perhaps, one of the challenge of humans regarding AI is to understand that ´AI is a social construction´. Borrowing that theoretical insight from S. Lindgren & Holmström (2020), despite AI is a shaper of society, AI is also shaped by society, and its social significance is based upon hopes, fears, and how it is conceptualised and understood. The presence of societal biases tied to AI are also a factor. In my effort to understand empirically high-risk AI, I address the topic from the experience of the producers, which have technical and social science background. My purpose was to let them explain to me – hopefully also to the readers, how they do artificial intelligence. As a counterbalance I approach the topic from a position of social scientist with a set of research questions, analytical framework and curiosity to interrogate, analyse, enhance discussions, and if possible, raise awareness among actors of a sociotechnical environment in which all of us are involved.

7. References

Asdal, K., Doganova, L., & Fochler, M. (2024). Valuation studies as a frame in STS. In U. Felt, & A. Irwin (Eds.), *Elgar Encyclopedia of Science and Technology Studies* (pp. 79-86). Edward Elgar.

Asdal, K., Doganova, L., & Fochler, M. (2024). Valuation studies as a frame in STS. In U. Felt & A. Irwin (Eds.), *Elgar Encyclopedia of Science and Technology Studies* (pp. 79–86). Edward Elgar Publishing. https://doi.org/10.4337/9781800377998.ch08

Baek, H., & Lee, H. (2021). Framework of Socio-Technology Analysis and Prescriptions for a sustainable society: Focusing on the mobile technology case. *Technology in Society*, *65*, 101523. https://doi.org/10.1016/j.techsoc.2020.101523

Baker-Brunnbauer, J. (2021). Management perspective of ethics in artificial intelligence. *AI and Ethics*, *1*(2), 173–181. https://doi.org/10.1007/s43681-020-00022-3

Crockett, K., Colyer, E., Gerber, L., & Latham, A. (2023). Building Trustworthy AI Solutions: A Case for Practical Solutions for Small Businesses. *IEEE Transactions on Artificial Intelligence*, *4*(4), 778–791. https://doi.org/10.1109/TAI.2021.3137091

Cunha, P. R., & Estima, J. (2023). Navigating the Landscape of AI Ethics and Responsibility. In N. Moniz, Z. Vale, J. Cascalho, C. Silva, & R. Sebastião (Eds.), *Progress in Artificial Intelligence* (Vol. 14115, pp. 92–105). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49008-8_8

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. https://doi.org/10.1145/3411764.3445188

European Commission. (n.d.-a). AI Act. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission. (n.d.-b). A European approach to artificial intelligence. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Gerdes, A. (2021). AI can turn the clock back before we know it. 2021 IEEE International Symposium on Technology and Society (ISTAS), 1–6. https://doi.org/10.1109/ISTAS52410.2021.9629161

Hsiao, W.-J., & Shorey, S. (2025). Machine visions: A corporate imaginary of artificial sight. New Media & Society, 27(1), 404–423. https://doi.org/10.1177/14614448231176765

Jaton, F., & Sormani, P. (2023). Enabling 'AI'? The situated production of commensurabilities. Social Studies of Science, 53(5), 625–634. https://doi.org/10.1177/03063127231194591

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or Not, AI Comes—An Interview Study of Organizational AI Readiness Factors. Business & Information Systems Engineering, 63(1), 5–20. https://doi.org/10.1007/s12599-020-00676-7

Kim, T., Molina, M. D., Rheu, M. (Mj), Zhan, E. S., & Peng, W. (2023). One AI Does Not Fit All: A Cluster Analysis of the Laypeople's Perception of AI Roles. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–20. https://doi.org/10.1145/3544548.3581340

Leszkiewicz, A., Hormann, T., & Krafft, M. (2022). Smart Business and the Social Value of AI. In T. Bondarouk & M. R. Olivas-Luján (Eds.), Advanced Series in Management (pp. 19–34). Emerald Publishing Limited. https://doi.org/10.1108/S1877-636120220000028004

Lindgren, H. (2024). Emerging Roles and Relationships Among Humans and Interactive AI Systems. International Journal of Human–Computer Interaction, 1–23. https://doi.org/10.1080/10447318.2024.2435693

Lindgren, S., & Holmström, J. (2020). A Social Science Perspective on Artificial Intelligence: Building Blocks for a Research Agenda. Journal of Digital Social Research, 2(3), 1–15. https://doi.org/10.33621/jdsr.v2i3.65

Ma, S., Zhang, C., Wang, X., Ma, X., & Yin, M. (2024). Beyond Recommender: An Exploratory Study of the Effects of Different AI Roles in AI-Assisted Decision Making (No. arXiv:2403.01791). arXiv. https://doi.org/10.48550/arXiv.2403.01791

Peeters, M. M. M., Van Diggelen, J., Van Den Bosch, K., Bronkhorst, A., Neerincx, M. A., Schraagen, J. M., & Raaijmakers, S. (2021). Hybrid collective intelligence in a human–AI society. AI & SOCIETY, 36(1), 217–238. https://doi.org/10.1007/s00146-020-01005-y

Prior, L. (2008). Repositioning Documents in Social Research. Sociology, 42(5), 821–836. https://doi.org/10.1177/0038038508094564

Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

Redaelli, R. (2023). Different approaches to the moral status of AI: A comparative analysis of paradigmatic trends in Science and Technology Studies. Discover Artificial Intelligence, 3(1), 25. https://doi.org/10.1007/s44163-023-00076-2

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and

Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)Text with EEA relevance. (n.d.).

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. Ethics and Information Technology, 24(1), 4. https://doi.org/10.1007/s10676-022-09624-3

Suchman, L. (2023). The uncontroversial 'thingness' of AI. Big Data & Society, 10(2), 20539517231206794. https://doi.org/10.1177/20539517231206794

Tomada, L. (2022). Start-ups and the Proposed EU AI Act: Bridges or Barriers in the Path from Invention to Innovation? Journal of Intellectual Property, Information Technology and Electronic Commerce Law, 13(1), 53–66.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2021). Framing governance for a contested emerging technology:insights from AI policy. Policy and Society, 40(2), 158–177. https://doi.org/10.1080/14494035.2020.1855800

Waibel, D., Peetz, T., & Meier, F. (2021). Valuation Constellations. Valuation Studies, 8(1), 33–66. https://doi.org/10.3384/VS.2001-5992.2021.8.1.33-66

Williams, R. (2006). Compressed Foresight and Narrative Bias: Pitfalls in Assessing High Technology Futures. Science as Culture, 15(4), 327–348. https://doi.org/10.1080/09505430601022668

8. Annex I: Interview guidelines

This annex contains the interview guidelines I developed to carry out interviews. Those guidelines do not include follow-up questions.

8.1 Interview I

1. As far as I know there is a department of Data and AI in [name of the startup] and your role there is AI Specialist of Algorithms and Product. Could you describe your role, please?

2. [Name of the startup] facilitates collaboration between companies and freelancers and has several products for this purpose. Which of those products use artificial intelligence?

3. There are technical and economic aspects taken into account when deciding to use artificial intelligence instead of other types of technology such as machine learning, the Internet of Things (IoT), or others, but are there other aspects that [name of the startup] consider?

4. When you [the startup] uses artificial intelligence, do you [the startup] use it to solve problems, improve a product or service, to fill a gap you identify that doesn't exist and could be useful, or for other reasons?

5. In our exploratory meeting, we discussed that in the current context of AI regulation, [name of the startup] identified that one of its products could possibly be part of the high-risk category. Which product is that?

    a. How is it work?
    b. What makes that product possibly fall into that category?

6. When [name of the startup] created that product based on AI, did you think it might be high risk?

7. From my vague understanding of AI, I understand that it's fuelled by data, and the more data it has, and the more implementation or practice or training it has, the better it will work because it identifies more patterns and offers more options, alternatives, solutions, and predictions. If this is true, what is the vision of [name of the startup] for this product?

    a. Do you [the startup] in any way define your expectations for that product?
    b. Can you [the startup] control the scope of that product?

8. Understanding that [name of the startup], like most startups, is in the process of understanding and adapting to the regulation. Based on what you know so far about this regulatory proposal, do you think it will impact the way [name of the startup] approaches or uses artificial intelligence?

    a. Will it in any way limit the ideation stage and then execution and production stages?

9. Has [name of the startup] considered any possible misuses of this/those products?

8.2 Interview II

1. I understand you have been with [name of the startup] for 4 years, which means you have been since the very beginning of the startup. What has your career at [name of the startup] been like?

2. What is your current role? Overall, what do you in that role?

3. One of the main issues that I understand is part of your role, is making the projects happen and designing the product. In that process you have some challenges such as, identifying what the client and the freelancer need, what is the problem to be solved, what are the expectations (customers, freelancer and [name of the startup]). How do you identify those needs, problems, expectations?

    a. At some point in the identification "phase", do you take into account other actors beyond customers, freelancers and [name of the startup]?
    b. What type of problems, needs, and expectations have you identified?
    c. Do you ever consider society in general when carrying out a project or designing a product, or do you respond more to what the customer requires, or what is better for freelancers?

4. I understand that another challenge of your role is to bring what you have identified (need, problem, etc.) and what based on your knowledge, might be a potential solution to the Data and AI team, and they analyze what type of technology is best suited. Is that correct? How is that process with the Data and AI team?

    a. When bringing out these requirements to the Data and AI team, do you consider the potential risks that may arise with this potential technological solution, or does the Data and AI team do that when defining the technology to be used? Or is it done at a later time? Or not at all?

5. If I understand correctly, there are two phases for developing or improving the product´s features: ideation and execution. Is that correct? What happens in each phase?

6. Speaking more specifically about the platform's matching process for connecting freelancers with companies, were you involved in the development and implementation of this matchmaking feature? Could you explain how it was like, please?

    a. Do you remember what motivated you [name of the startup] to create this type of feature? Is there any documentation?
    b. Once the matching feature was ready, were any concerns identified that might arise among your audience (freelancers and companies) or among society in general?
    Expectations?
    Needs?
    Risks?
    Is there any documentation?
    c. What was your role in the development and implementation of that feature?

7. From your perspective, what impact does this matchmaking feature have on the audience of the startup?

    a. Do you think there is any impact beyond the audience [name of the startup] wants to reach?

8. In your role, do you have a vision for where matching functionality should be headed? Anything that could guide the development of long-term improvements to this feature?

9. Are you aware of the regulation of AI that will come into effect in the EU?

    a. Have you considered the implications for the product?

b. Have you considered the implications for the product implementation stage?

9. Annex II: Original quotes

This annex contains the original version of each quote included in the present document.

1.AI Specialist “Nosotros probablemente seamos considerados como una aplicación de alto riesgo por el contexto en el que se produce, que es el contexto laboral. Todo lo que es la parte del matchmaking, todo lo que es la parte realmente peliaguda donde nosotros podemos tener, o que la regulación va a actuar yo creo, sí o sí, es en la parte de cuando nosotros recomendamos talentos a proyectos o proyectos a talentos”.

2.Product Manager “Bueno, básicamente ahí dijimos, queremos hacer eso. ¿Cómo es la forma de hacerlo? …no hay nada más que un algoritmo de similaridad para hacerlo. No se nos ocurría ninguna otra,…, ahí no hicimos tanto de ver todos los problemas de los usuarios, ni nada, porque no teníamos usuarios”

3.Product Manager “Al principio, cómo las preguntábamos, ¿qué sabes hacer?... rellenando un perfil, ¿cómo eres? poníamos cuatro ejemplos de personalidades siguiendo el modelo del Disc type de psicología, que como cuatro modelos de personalidad, y te decíamos que ¿con quién te identificabas? …y luego te preguntábamos ¿qué te gusta hacer…?

Y a los clientes les preguntábamos, oye, ¿qué te gusta hacer?... ¿cómo eres? … ¿Qué necesitas para tu proyecto? Y hacíamos un match directo, asumiendo que la gente que le gusta lo mismo se va a llevar mejor entonces van a trabajar más a gusto…Esto era la versión número uno, ¿qué pasaba? … no era muy efectivo porque … no es fácil identificarte con uno de esos modelos…”

4.AI Specialist “ El primero de los pasos es meter a participantes dentro de nuestro market place, .. es decir, que una empresa … nos dé toda la información posible de que empresa es, qué quiere, qué busca, qué necesita, qué proyecto quiere realizar. Lo mismo para los talentos, que la gente pues se aplique, que pasen las entrevistas, que pasen los filtros de calidad, que rellenen el perfil, que tengamos la información de esos talentos...

El segundo paso es que la gente decida trabajar entre ellos…nosotros algorítmicamente hacemos un match, puedes decirlo, hacemos oye,…te recomendamos este proyecto, o a un cliente, … estos son los talentos que mejor pensamos que vienen para tu proyecto”

5.Product Manager “Yo estoy analizando el flujo de cómo un cliente publica un proyecto en [name of the startup], … entonces … yo me doy cuenta que nosotros le estamos pidiendo que nos diga el presupuesto, pero el problema que tiene él es que …él no lo sabe, ... ¿Entonces, cómo podemos solucionar eso? Pues quizás … mediante un algoritmo que saque el presupuesto de tu proyecto”

6.AI Specialist “nosotros automáticamente estimamos cuántas personas podrías necesitar dentro de tu equipo y qué tipo de perfiles… hay mucha gente que lo sabe, …pero cuando nos viene una empresa que dice yo es que quiero montar mi página web o yo quiero montar mi plataforma … yo quiero tal …la gente dice …no tengo ni idea … no sé ni que cuántos perfiles, qué perfiles… Pues …obviamente la gente es experta en su negocio,…pero no claro, no tienes porque ser técnico y nosotros tenemos que detectar esa necesidad y automatizar ese proceso lo mejor posible.”

7.Internal document “El objetivo principal del sistema de emparejamiento es ofrecer una recomendación personalizada e individual de las posiciones disponibles para talento o viceversa”.

8. Internal document "…[name of the startup] ha implementado y perfeccionado un sistema de IA destinado a realizar tareas de emparejamiento automático o matchmaking para conectar talento con necesidades empresariales en proyectos teniendo en cuenta habilidades profesionales e individuales. Así, [name of the startup] representa un marketplace bilateral, donde empresas y freelances … se encuentran para desarrollar proyectos de forma dinámica y conjunta.

El sistema de IA está integrado dentro de… plataforma, que comprende todos los pasos necesarios para el desarrollo de proyectos…".

9.AI Specialist "…yo vengo de estadística, entonces machine learning, inteligencia artificial, lo que sea, al final es matrices con matrices a la hora de hacer cualquier tipo de tarea para mí, la clave es donde se genera valor. Y con valor me refiero a que pueda aportar el negocio. Es decir, nosotros nunca tomamos un problema con una perspectiva de queremos hacer machine learning por hacer machine learning, o queremos hacer inteligencia artificial porque todo el mundo hace inteligencia artificial…"

10.AI Specialist "entendemos que cuando hablamos de valor, osea nuestro valor más importante dentro de [name of the startup] es la comunidad de freelance que tenemos".

11.AI Specialist "…nuestro valor más importante dentro de [name of the startup] es la comunidad de freelance que tenemos. Nosotros empezamos lo que llamamos Talent first. Es decir, por ejemplo, nosotros a un cliente sólo dejamos que invite a un proyecto a tres freelances, pero los freelances pueden aplicar a todos los proyectos que quieran y cuando tiene que llegar un acuerdo económico, es el talento el que tiene que poner las condiciones al cliente, no el cliente al talento".

12.AI Specialist "…la filosofía de sí que es generar una economía del trabajo que sea positiva para ambas partes. Unos ganan flexibilidad, ganan dinero…por encima de lo que es su referencia de mercado, pero para la compañía también le viene el lujo porque traes expertos…que hacen el trabajo….Esta es un nuevo modelo de cómo de integración entre gente que no quiera atarse a ninguna compañía y la compañía es que tampoco quieren atarse a ningún trabajador de manera externa y poder colaborar durante un tiempo para un proyecto determinado".

13.AI Specialist "todas las aplicaciones de AI están centradas en optimizar procesos dentro de nuestro dentro de nuestro producto".

14.AI Specialist "cuando un cliente quiere se registra en [name of the startup] … Nosotros le decimos que nos escriba … ¿qué objetivos tienes? ¿cuál es el título de proyecto? ¿qué cosas quieres hacer? Y automáticamente …detectan esas necesidades y lo transforman en nuestras skills o habilidades que la persona solamente tiene que seleccionar…Entonces nosotros lo que hacemos es facilitar el trabajo al cliente".

15.Product Manager "es muy crítico y muy importante dentro de [name of the startup],… el algoritmo es que no puede dejar de mejorar nunca".

16.Public knowledge base "un usuario que tenga las habilidades "Wordpress" o "Diseño Web" validadas en su perfil, tendrá un porcentaje de match "base" alto con los proyectos que demanden dichos conocimientos".

17. Public knowledge base "hemos diseñado un test para identificar los intereses de un freelance, y otorgarle un mayor porcentaje de match a los proyectos que encajen con las motivaciones del talento".

18.AI Specialist “para cada una de las recomendaciones tenemos la oportunidad de que el talento me diga esta recomendación es adecuada o no es adecuada. Tiene los botones unm up down”.

19. AI Specialist “nos damos cuenta de que lo más importante para nosotros es tener datos muy curados, es decir, datos que confiemos muy bien, que esos datos reflejan los procesos reales”.

20. AI Specialist “…la ideación es la misma. Tú tienes un problema de predicción o de clasificación, o tú quieres aplicar un GPT a hacer lo que sea y tú lo vas a hacer… la ideación no se pierde. Pero si la ejecución antes te va una semana, ahora la ejecución (obtener las métricas y obtener la adaptación a la regulación), a lo mejor el día que lo voy a presentar va a otras tres”.

21. AI Specialist “ si una startup…quiere utilizar inteligente artificial … y de repente dicen tienes que contratar un experto en este tema…ese experto también debe saber adaptar a la regulación”.

22. AI Specialist “…lo que creo que va a pasar es que la gente va a trabajar y va a seguir innovando, el problema es que luego pasarán épocas en las que tendrán que adaptar todo a la regulación, y ahí se perderán, pues meses de trabajo y/o dinero en consultoría para hacerlo”.

23. AI Specialist: “…nos aseguramos es intentar … encontrar dónde falla, porque es en esos casos extremos …tenemos que volver e investigar qué ha pasado… y muchas veces ocurre, y sobre todo, con modelos de caja negra que como tú no controlas explícitamente los pesos o qué es lo que está ocurriendo o puede ocurrir dentro, pues te toca irte al caso concreto …,y muchas veces lo que te obliga a eso es hacer un trabajo casi de detective de que, como no tienes muchas veces ni idea de por qué ocurre irte atrás y ir a los datos de entrenamiento …Entonces tú ves el patrón que aprende el modelo…

Obvio que eso lo podemos hacer todavía porque nuestro volumen de datos es muy grande, pero como las casuísticas que tenemos son tan amplias todavía hoy podemos perder el tiempo en hacer este tipo de lucubraciones, y aún así no tienes muy seguro porque al final es el problema de este tipo de modelos más complejos, sobre todo cuando metes ciertos modelos estadísticos difícilmente explicables en el que aún así no sabes qué realmente está pasando, que ha captado esos patrones o que está pasando que dentro de ello. Pero en esos casos extremos, nosotros revisamos los datos de entrenamiento”.

24.Product Manager “hasta el último año (2023), …, no teníamos un volumen suficiente como para no poder compensar las carencias del match con personas. Es decir, que si el match daba una recomendación que no era adecuada,…al final…, no tenemos tanta comunidad tan grande, … entonces hemos podido tener margen de maniobra si el match fallaba…¿qué pasa? que cada vez tenemos más volumen, entonces cada vez sí tiene que ser todo más automático, entonces si el match no funciona, no funciona nada”.

25. Product Manager “yo creo que es la cosa más crítica y más importante que hay en todo [name of the startup], al final nosotros juntamos una gente de un lado con otra gente de otro, y hay algo en el medio que les tiene que juntar, si ese árbol del medio no funciona, da igual que tengas a lo mejor aquí y a lo mejor aquí, que si no eres capaz de unir los puntos, no lo vas a conseguir”.

26. AI Specialist “Todos los usos que hacemos de inteligencia artificial están muy orientados a nuestro producto”.

27. Product Manager "… aunque no creciéramos en demanda, siempre va a tener que ser mejor, o sea, me parece lo más difícil. Me parece que es algo que nadie ha sabido resolver mediante tecnología, todavía sin intervención humana. Y que, si somos capaces de resolverlo, pues nos paramos el juego…".

28.AI Specialist "[name of the product] es una plataforma, por así decirlo unitaria. Aunque tengamos pasos desde que entras hasta que sales, es un solo producto. Probablemente seamos considerados como una aplicación de alto riesgo por el contexto en el que se produce, que es el contexto laboral. Todo lo que es la parte del matchmaking…la regulación va a actuar, yo creo…es en la parte de cuando nosotros recomendamos talentos a proyectos o proyectos a talentos".

29. Project Manager "…otras plataformas tienen sus algoritmos que están seguramente hasta mucho más evolucionados que el nuestro, pero no tienen en cuenta ninguno estas variables, …, si nosotros conseguimos hacer que dentro de 200 perfiles capaces de hacer tu proyecto, destacarte los 3 que mejor lo hacen, pues ahí es donde creo que pararemos el juego".

30. AI Specialist "…tenemos una tasa de repetición de la mitad de la gente que empieza a trabajar con [name of the startup] decide continuar".

32. Product Manager "nosotros tenemos cerrada la plataforma y solo entran las personas que nosotros validamos".

33. Public knowledge base "Cuando definimos a ese trabajador del "futuro", que disfruta de plena flexibilidad, deslocalizado, superespecializado, que trabaja por hitos etc, estamos construyendo el fiel retrato del freelance. **Perfiles que**, en su mayoría, **en el punto más top de su carrera, deciden apostar por su talento y desarrollarlo en su máximo potencial trabajando en proyectos que les motivan**".

34. Public knowledge base "Es sencillo, quien quiera contar con el mejor talento en sus equipos tendrá que subirse al carro del freeworking…".

35. AI Specialist "…obviamente el cliente puede invitar a un talento que tenga un 30% de match, pero el cliente va a invitar… a los tres que tienen más match. Por mucho si tiene curiosidad y se ve los perfiles, invitará al quinto, al sexto".

36. AI Specialist "…validando ese conocimiento experto que las personas hacemos una tarea muy fácil, pero que a las máquinas les cuesta mucho. Y cuando nosotros lo sacamos a producción, nosotros para cada una de las recomendaciones tenemos la oportunidad de que el talento me diga esta recomendación es adecuada o no es adecuada… Y esas señales son introducidas en el modelo para reentrenar el modelo y ver exactamente qué ha podido estar pasando…digamos la forma de delimitarlo controlarlo si alguna manera es conocimiento".

37.Product Manager "…seguimos metiéndole cosas y cambios a mejorar el matching a nivel, lo llamamos, cultural, fit cultural y a nivel motivacional. Entonces, …pues hicimos dos test, uno de personalidad y uno de motivación …Y como [name of AI specialist], además es doctor en psicología también, pues ya le metió encima toda su capa de psicología …al principio ninguno éramos expertos en nada, y ahora, pues ese test está hecho y validado, o sea, tiene como mucha ciencia detrás que antes no teníamos".

38.AI Specialist “Es difícil de decir, en mi caso, creo que el impacto va a ser relativo… yo vengo de la universidad, entonces yo estoy acostumbrado a una serie de procesos de calidad y rigor que son relativamente no, similares, pero bueno que siguen un poco la estela de lo que se piden a la hora de diseñar sistemas transparentes, diseñar sistemas responsables”.

39. AI Specialist “conocimiento experto que las personas hacemos una tarea muy fácil, pero que a las máquinas les cuesta mucho”.

40. AI Specialist “Género es algo que hasta hace muy poquito, nosotros no preguntábamos…Sin embargo, ahora lo vamos empezado a preguntar para precisamente tener un control de cuántos hombres mujeres recomendamos dentro de posiciones y poner las acciones necesarias en aquellos sitios donde podamos observar gaps de género”.

41. AI Specialist “diseñar sistemas transparentes, diseñar sistemas responsables”.

42. Internal document of the startup “El sistema de IA de Shakers está diseñado como un asistente inteligente que permita: a) que las organizaciones tengan una valoración y ordenación de los talentos relativos a su afinidad a el proyecto; b) que los talentos tengan una valoración y ordenación de los proyectos en función de su afinidad con los mismos”.

43. Internal document of the startup “El sistema de IA facilita y enriquece el proceso de toma de decisión, pero no responde a un sistema cerrado donde el usuario cede el control sobre la decisión final”.

44. AI Specialist ““…tenemos una tasa de repetición de la mitad de la gente que empieza a trabajar con [name of the startup] decide continuar”.