Are We Nearly There Yet? A Desires & Realities Framework for Europe's AI Strategy

Dr. Ariana Polyviou, Department of Management, School of Business, University of Nicosia, 2417, Nicosia, Cyprus, polyviou.a@unic.ac.cy

Dr. Efpraxia D. Zamani, Information School, The University of Sheffield, S1 4DP, Sheffield, U.K., e.zamani@sheffield.ac.uk

Abstract

Of all emerging technologies, Artificial Intelligence (AI) is perhaps the most debated topic in contemporary society because it promises to redefine and disrupt several sectors. At the same time, AI poses challenges for policymakers and decision-makers, particularly regarding formulating strategies and regulations to address their stakeholders' needs and perceptions. This paper explores stakeholder perceptions as expressed through their participation in the formulation of Europe's AI strategy and sheds light on the challenges of AI in Europe and the expectations for the future. Our analysis reveals six dimensions towards an AI strategy; ecosystems, education, liability, data availability sufficiency & protection, governance and autonomy. It draws on these dimensions to construct a desires-realities framework for AI strategy in Europe and provide a research agenda for addressing existing realities. Our findings contribute to understanding stakeholder desires on AI and hold important implications for research, practice and policymaking.

Keywords: Artificial Intelligence, AI Strategy, technology policy, technology regulation, AI ethics.

1. Introduction

Artificial intelligence (AI) brings a plethora of benefits to citizens, enterprises and the economy, and it is expected to transform our lives. For example, it is expected to change healthcare by assisting the diagnosis of diseases (Yu, Beam, & Kohane, 2018), to contribute to reaching greater efficiency in energy systems (Kastner, Kofler, Reinisch, 2010), to increase the security of critical systems (Karagiannis et al., 2020), and to bring forward a digital revolution across the industry (Sigov et al., 2022).

Presently, there is a variety of available definitions for AI. In this paper, we define AI as "a collection of technologies that combines data, algorithms and computing power" (European Commission, 2020, p.3). Given that not all digital tools favour humanity (e.g., dark web, fake news) (Stahl, 2021), it is essential to carefully design strategies that will prevent AI's fail-safes and promote its wider deployment for the benefit of the economy and society. Specifically, AI's black-boxed algorithms must be scrutinised carefully and regulated so we can shift beyond the risks of uncontrolled algorithms and towards a trustworthy AI (Meske et al., 2022).

Along these lines, AI is one of the top priorities on the European Commission’s digital agenda. Indeed, drawing on its technological and industrial strengths, high-quality digital infrastructure and regulatory framework, Europe holds the potential to become a leader in innovation in the data economy and its applications. To foster this, the EC has derived a list of actions encapsulated in the 'European approach to excellence in AI'. Furthermore, aiming to identify Europe's vision and priorities on AI (European Commission, 2021a), between February and June 2020, the EC launched a consultation on Artificial Intelligence, inviting relevant stakeholders to share their views on the recently published White Paper on 'Artificial Intelligence: A European approach to excellence and trust' (European Commission, 2020).

Existing literature emphasises reviewing extant research and policy documents on AI to conclude how to better facilitate AI for excellence and trust and identify the challenges in AI implementation (Enholm et al., 2021; Ulnicane et al., 2021). To enhance the relevance (Benbasat & Zmud, 1999) of this body of research, our work complements existing literature through empirical data supplied by key stakeholders and enriches it by identifying pathways for future research and development. We do this by analysing stakeholders' perceptions towards formulating an integrated European strategy on AI. Stakeholders in our case are individuals with a good understanding of AI who are actively engaged in expressing their views on how to reach excellence and trust in AI (i.e., academics, researchers, policymakers, key industry players). Our study sheds light on the realities of potential issues of using AI and the desires on how these issues could be best addressed in the EU's forthcoming AI strategy. We consider this work to be timely and vital in light of the increased attention paid to AI and AI-enabled information systems, whereby governments, national and international agencies and the industry as a whole need to consider the needs and expectations of their stakeholders and those impacted by the technology when designing policies, products and services.

Drawing on a qualitative approach, our work addresses the following research question: “What are the needs and expectations of stakeholders with regards to the EU’s AI strategy?” and it analyses stakeholders’ perceptions towards excellence, organised into six dimensions: ecosystems, education, liability, data availability, sufficiency & protection, and governance. We then reflect on our findings regarding the realities and expectations that emerge in the backdrop of the AI Strategy and derive a list of research questions to guide future research in this area and support the application of a forthcoming AI Strategy.

The paper is structured as follows. In what follows, we provide an overview of the existing literature on AI and particularly regarding stakeholder perceptions. Next, we describe the theoretical background for structuring and analysing desires and expectations, forming our study's analytical device. We then describe the methodological approach employed and present our findings from analysing the stakeholders' letters. This is followed by a discussion of our findings against the current literature, and we develop a framework mapping the desires regarding Europe's AI strategy against the realities. Finally, we conclude our work by identifying limitations and future directions.

2. Literature Review

2.1. Definitions of AI

Ulnicane, Knight, Leach, Stahl & Wanjiku (2021) note that 'Artificial Intelligence' tends to be used as an umbrella term for several technologies and domains, ranging from cognitive computing and augmented intelligence to machine learning algorithms. Therefore, a single accepted definition of AI does not exist (Rizzi & Perra, 2020, p. 7). Indeed, there exists a variety of definitions of AI. For example, an early definition described AI as "computer programmes that model aspects of intelligent behaviour" (Fleck, 1982, p. 169). Similarly, Afiouni (2019) describe AI as systems that execute tasks that generally require natural human intelligence and may or may not be rule-based. Other definitions, however, are detailed in how AI achieves this type of behaviour. For example, Mikalef and Gupta (2021) define AI as a system that identifies, interprets and makes inferences, and learns from the data to accomplish goals. Across these and other definitions, it is noted that AI is meant to instil systems with "human-like capabilities (...) such as understanding, reasoning, and problem-solving" (Enholm et al., 2021).

Such definitions of AI that approach the technology as one that can resemble human behaviour have been quite influential and have shaped policy regarding how the technology is perceived. For example, Vinnova, the Swedish Innovation Agency, describes AI as "the ability of a machine to imitate intelligent human behaviour", which is closely related to Fleck's very early definition. The German Federal Government differentiates between strong and weak AI, where strong AI denotes systems with the same or more intellectual capabilities as humans, whereas weak AI denotes systems that leverage computer science techniques to support their self-optimisation (German Federal Government, 2018, p. 4). An early definition by the European Commission describes AI as "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals (European Commission, 2018). In this study, as we are interested in the perceptions of the EU's upcoming AI strategy, we adopt the definition that is included in the European Commission (EC)'s "White Paper on Artificial Intelligence: A European approach to excellence and trust" (European Commission, 2020), according to which "AI is a collection of technologies that combines data, algorithms and computing power" (p.3).

2.2. Perceptions towards AI

AI is argued to be at the forefront of digital transformation and disruptions, whereby the technology can lead to considerable benefits for societies and businesses, such as improving welfare, supporting sustainable development, creating efficiencies and automating processes (Enholm et al. 2021). As a result, in recent years, AI has attracted much attention from researchers and practitioners alike, who have explored several aspects of it, such as reliability, trustworthiness, responsibility and ethics (e.g. (Kerr, Barry, & Kelleher, 2020; Stahl, 2021).

However, most research focuses on deriving empirical data and exploring the public's perceptions of AI. For example, in (Kelley, 2021), the author explored the perceptions of over ten thousand citizens in eight countries and six continents, aiming to understand the views of individuals on AI. In the analysis, the researcher identifies different sentiments associated with AI responses (exciting, useful, worrying, and futuristic). Similarly, in (Güngör, 2020), the author focused on examining professionals' perceptions in Europe. The research highlighted a positive perception towards using AI for financial benefit and value creation; however, stakeholders perceive that the wide adoption of AI will negatively impact employees and society. In another study, the authors draw on a survey to explore the AI-based service trustworthiness in organisational stakeholders (Araujo, Helberger, Kruikemeier, De Vreese, 2020). Their results highlight that the attitudes are associated with concerns about risk and mixed views on AI's fairness and usefulness. However, on their evaluation of potential fairness, usefulness, and risk of specific decisions taken automatically by AI in comparison to human experts, respondents consider it as equal or even better for making high-impact decisions. Other studies also explored the perceptions of stakeholders on the use of AI in specific domains (e.g., financial security (Melnychenko, 2020), public sector (Qian Sun, Medaglia 2019), health care sector (Blomqvist, Van Der Werff, 2020) and human resource management (Bankins et al., 2022).

The common denominator across such studies is that while AI undoubtedly promises to provide solutions to significant business and societal problems, at the same time, it raises concerns about potential problematic future scenarios. For example, Winfield and Jirotka (2018) note that, despite that the public holds overall positive opinions regarding AI, this seems to be declining, with the majority indicating the need for careful management and transparency. Indeed, it is essential to carefully design strategies that will prevent fail-safes of AI and promote its wider deployment to

benefit the economy and society. AI's black-boxed algorithms need to be scrutinised carefully and regulated so that we can shift beyond the risks of uncontrolled algorithms and towards a trustworthy AI (Trocin, Mikalef, Papamitsiou, Conboy, 2021). This need has resulted in an increasing number of convergent and divergent ethics guidelines targeting the design and development of AI systems (Ryan and Stahl, 2020) and the turn to the Responsible Research and Innovation agenda (Stahl, 2021).

In the next section, we draw our attention to the literature that focuses on ethical and responsible AI to understand the relevant principles and guidelines better.

2.3. Reliable, ethical and trustworthy AI

Several studies discuss the need for incorporating ethical principles within the AI discourse and practice (Stahl et al., 2021). Specifically, current discourse focuses on 'what ethical AI is' (Firth-Butterfield, 2021)

Moreover, it summarises existing or develops new suggestions regarding principles and guidelines for ethical and responsible AI. For example, Jobin, Ienca, & Vayena (2019) identify the converging views on the ethical principles of AI. In particular, the paper identifies five distinct ethical principles, including transparency, justice and fairness, non-maleficence, responsibility and privacy. However, the paper also highlights that, despite the emergence of the five principles, there is notable divergence on how they are interpreted and how they should be implemented. Similarly, Hagendorff (2020) reviews principles and opinion papers on the ethics of new AI technologies. More specifically, it reviews 22 papers on ethical AI, highlighting that most ethical considerations are relevant to aspects of accountability, privacy or fairness. Others emphasise specific aspects of ethical AI, such as education and teaching AI to data scientists (e.g., Garzcarek & Steuer, 2020,

Goldsmith & Burton, 2017). In contrast, discussions have also been initiated on ethics washing to avoid or escape governmental regulations (Rességuier & Rodrigues, 2020).

Literature also focuses on designing systems that can facilitate reliability and trustworthiness by design. For example, in Shneidermam (2020), the author emphasises human-centred AI and suggests a framework that can assist the balancing between human control and computer automation, identify situations in which full human or full machine control is required and minimise the risks arising from excessive human or machine control. In Güngör (2020), researchers introduce a framework that facilitates the interplay of human values, interpersonal dynamics, and AI systems' socially situated nature. Focusing on the digital health domain, in particular, Trocin et al. (2021) further develop a research agenda for the design and use of AI in digital health. While the authors do not provide specific design principles, they highlight issues regarding inconclusive, inscrutable and misguided evidence, unfair outcomes, transformative effects and traceability, which they then use as a springboard for querying how knowledge is constructed through AI; how results are interpreted; for arguing in favour of sensitivity with regards to disproportionate impacts on underrepresented groups; and being reflexive regarding privacy and autonomy. Finally, Ryan and Stahl (2020) have developed a framework for identifying ethical principles, such as justice and fairness, non-maleficence, dignity and solidarity) for responding to constituent ethical issues (e.g., equality and equity, plurality, non-discrimination, cohesion), and provide tangible examples for achieving these. For example, they note a great need to increase diversity in AI teams and datasets for addressing gender-based biases in AI systems.

Our work complements this stream of work, highlighting the considerations of European stakeholders on the use of AI in Europe and their implications for better accommodating these considerations in a forthcoming AI Strategy in Europe. We believe that such an engagement with

stakeholders' perceptions will enable "moving from 'is' to 'ought'" and support the design and development of AI systems within the EU region at least, whose purpose is human flourishing (Stahl, 2021). In the next section, we describe the actions of the European Commission toward the formulation of a European Strategy on AI.

2.4. The EU approach to AI

The European Commission is committed to supporting scientific breakthroughs and promoting the EU's technological leadership while ensuring that technological innovations are made available to all Europeans to improve their lives and respect their rights. While some of the EU Member States had already defined and announced national initiatives on AI, in April 2018, the EU Member States signed up a cooperation agreement on Artificial Intelligence (European Commission, 2018a). Through this agreement, they declared an intention to collaborate and contribute toward a European AI approach which will enable them to ensure the opportunities of AI for Europe fully and address upcoming challenges collectively. Furthermore, the EU aims to define its own way of addressing the opportunities and challenges arising from AI, such that it can sketch its unique way to promote the development and deployment of AI-based on EU values. Following up on the agreement, the EU has organised many actions aiming at the development of a coordinated EU strategy on AI, such as a Coordinated Plan on AI (European Commission, 2018c), leading to the and the publication of a white paper on AI for a European approach to excellence and trust (European Commission, 2020). In light of the publication of the White Paper, the EU Coordinated Plan aimed at the engagement of citizens stakeholders to contribute with their views on the upcoming European policy and regulatory steps on AI.

Europe has the potential to become a global leader in AI as it can combine technological and industrial strengths with high-quality digital infrastructure and a good regulatory framework. The

White Paper (European Commission, 2020) aimed at identifying the policy options for addressing the human ethical implications arising from AI and improving the use of big data for innovation purposes. This will enable it to reach sufficient scale while facilitating the emergence of a dynamic European policy and can bring benefits to citizens, businesses and the general public. The White Paper consists of two building blocks: i) a policy framework towards an ecosystem of excellence that will encapsulate research and innovation community, SMEs, partnerships with the private sector, adoption of AI by the private sector, and collaboration with like-minded global players, ii) main components of a future regulatory framework for AI which will facilitate an ecosystem of trust in alignment with the EU values on human and consumer rights such that citizens will have the confidence to use AI applications and provide businesses with the legal certainty to innovate via AI.

After the publication of the White Paper, the European Commission invited the Member States, other European institutions, citizens and relevant stakeholders (e.g., industry, social partners, researchers, general public) to provide their views and contribute to the EU's future decision making on AI. This was achieved through a public consultation open from February 19th to June 14th 2020. AI stakeholders could provide their consultation online through an online survey and/or letters submitted to the European Commission. Data arising from this public consultation and other activities have led to the publication of the European Commission's Proposal for a Regulation on Artificial Intelligence (AI) systems (European Commission, 2021).

3. An Analytical Tool for our study: ‘Desires’ and ‘Realities’

Emerging and disruptive technologies are often understood and assessed based on the potential stakeholders see in them and relative to their expectations, rather than tangible outcomes stemming

from their implementation (Alkeemade and Suurs, 2012). In their study, Venters and Whitley (2012) have argued that often, such expectations are further compounded by the hype that surrounds emerging and disruptive technologies, whereby whichever technology happens to be rising in popularity is expected "to solve all organisational needs often on the basis of limited or non-existent empirical evidence" (p. 182). This holds for AI, too. Businesses, governments and national and international research programmes approach AI as a core technology for delivering benefits across a range of domains, shaping expectations around AI, with implications for funding and policy (Kerr et al., 2020). Such expectations strongly influence research activities and funding and create legitimacy (Borup, Brown, Konrad & Van Lente, 2006). However, despite such expectations, the reality may "fall short of the hype" for several reasons, including due to the complexity of the technology, lack of expertise and issues concerning data (Holford et al., 2019, p. vi). At the same time, hype in itself is often perceived as deceptive, where the capabilities of a technology are being exaggerated (Van Lente, Spitters, & Peine, 2013).

Inspired by the sociology of expectations (Van Lente, 2012; Kerr et al., 2020), we draw from Venters and Whitley (2012) to develop an analytic device for identifying and structuring the repertoire of EU's and stakeholders' expectations (or desires) with regards to AI. This approach is further discussed next.

3.1. Repertoires of Desires and Realities regarding AI

We frame our study in terms of 'desires' and 'realities'. In our study, desires reflect the expectations of the EU's AI stakeholders regarding what they hope the EU AI strategy will look like and what AI itself can potentially support in the future. Such expectations, although forward-looking, shape the development, adoption and regulation of emerging technologies, and in this case, AI. In a way,

acknowledging today's capabilities of AI influences expectations with regards to what future technological advances will allow AI to achieve and the impacts on our everyday lives (Stahl et al., 2021). Along these lines, Ulnicane et al. (2021) have used a rhetorical governance frame as their analytic device to structure the discourse around AI policy and identify existing prescriptions. The authors discuss that the state is seen as a regulator and a mitigator of potential risks and explain that, at the same time, high expectations seem to exist for public engagement as a way of ensuring equality and diversity in AI development and AI-enabled responsible products, systems and services.

3.2. Identifying challenges through the Desires-Realities approach

Structuring our study around desires is beneficial, theoretically and practically, for several reasons. First, AI stakeholders often desire many things from each other and the technology, much as it happens with other technologies (Venters and Whitley, 2012). However, each of these stakeholder groups has different agendas, resulting in conflicting interests and tensions within the AI ecosystem. In addition, in line with the sociology of expectations, expectations are not always realistic nor achievable (Borup et al., 2006). Whether they can be fulfilled depends on a range of factors, including the content of the expectations and who are the actors expressing said expectations (Alkemade and Suurs, 2012). Therefore, identifying and elaborating on today's expectations regarding AI can help us consider the realities of AI, identify the challenges that need to be addressed to achieve stakeholder expectations, and explore a pathway towards reconciling expectations and realities.

4. Methodology

In this study, we focus on the priorities and needs of stakeholders relative to the EU AI strategy, with the underlying aim to explore how an AI strategy could be designed based on responsibility and EU values. For this purpose, we adopt a qualitative research approach because it allows us to develop in-depth insights and access rich meanings as we engage with our empirical material. For our empirical material, we use secondary data, namely letters submitted by stakeholders during the open consultations around the European Commission's "White Paper on Artificial Intelligence: A European approach to excellence and trust". In more detail, the European Commission launched a Consultation on Artificial Intelligence on February 19th, 2020, inviting commentary from a range of stakeholders, including scholars, businesses, organisations, and citizens. The consultation was open to feedback and comments until June 14th 2020, inviting stakeholders to put forward their views on the White Paper and, therefore, contribute to shaping the European Commission's forthcoming AI strategy. Contributors included EU and non-EU citizens, Member States and all relevant stakeholders (including civil society, industry and academics, such as PICPU, CLAIRE, Booking.com, Nokia and Stanford University labs), who provided their opinions, concerns and recommendations through publicly available letters.

There were 422 letters, of which 71 were not written in English (e.g., German and Italian). In addition, several of the submissions were not stakeholder letters but general contributions, such as academic journal and conference papers, weblinks to multimedia (such as YouTube), interview transcripts, among other things (21 submissions). As such, 92 submissions were excluded from subsequent analysis, resulting in a final dataset of 330 letters. Each letter was assigned a letter plus a six-digit number (e.g., [X111111]) for identification purposes. The data analysis method we followed was that of thematic analysis. Thematic analysis offers a flexible yet rigorous approach

for identifying and organising patterns within the empirical material (Braun & Clarke, 2006). We followed a six-step approach (Griva, Kotsopoulos, Karagiannaki & Zamani, 2021):

Step 1 - Familiarisation with the data: We downloaded, stored and assigned unique identifiers to all letters in our final dataset, and we then read them carefully, making memos around the preliminary ideas emerging from them;

Step 2 - Initial coding: We developed a preliminary coding scheme based on the emerging ideas and codes, examining and comparing them against each other;

Step 3 - Theme Identification: We began comparing the emerging coding scheme to the existing literature, which resulted in reducing the initial set of codes by merging them into broader concepts;

Step 4 - Review: We compared the evolving coding scheme against our dataset to examine whether the empirical material confirms our scheme. We also reviewed the coding scheme itself to ensure that the resulting codes are distinct from one another;

Step 5 - Definition of themes: We named the resulting themes and subthemes;

Step 6 - Write up, at which point we analysed our results and extracted quotations to develop the chain of evidence of our study.

Steps 1 and 2 were performed by the first author, while Steps 3 and 4 were conducted in consultation with the second author to confirm the analysis and the coding scheme. This consultation process allowed both authors to overcome potential personal biases in the interpretation of the findings and supported the validity of the study. Validity concerns were further considered by juxtaposing our own interpretations against the current discourse surrounding AI while accounting for the potential interpretations of our findings from multiple perspectives (e.g.,

industry representatives, academics, policymakers). During this process, no major discrepancies in understanding were noted, and minor ones were addressed via discussion and consulting the relevant literature. Based on this analysis, we identified the stakeholders' desires for a human-centred AI strategy, which we organised into dimensions. This led to a list of high-level thematic codes that underpins our analysis and correspond to the dimensions mentioned above. As we were specifically interested in exploring stakeholder perceptions through the lens of expectations, we did not consider alternative theoretical frameworks other than that put forward by the Sociology of Expectations because we aimed to draw from prior research on emerging technologies within the disciplines of Innovation, Information Systems and others (e.g., Van Lente, 2012; Kerr et al., 2020; Venters and Whitley, 2012).

We note that we did not code all letters in the final pool: after reading and coding approximately the first half of the pool (150), we observed that no new themes were emerging, but rather that the same themes were recurring and that the content of these was essentially the same, suggesting that we had achieved saturation in our analysis (Saunders et al., 2017). In what follows, we present our findings.

5. Findings and Analysis

Our analysis focused on the priorities and needs of stakeholders regarding an EU-driven AI strategy with the view to support it. Our findings show that stakeholders today operate within an AI ecosystem, whereby they experience many obstacles and challenges due to the realities of AI. Therefore, their contributions to the consultation reveal, in effect, stakeholders' recommendations with regards to how, on the one hand, the EU's AI strategy can be implemented while, on the other

hand, their desires regarding how the EU can support them in overcoming the obstacles stemming from the realities surrounding the AI ecosystem today. These also constitute their desires for the EU strategy on AI to be formulated.

In what follows, we present our findings based on six major thematic dimensions that emerged from the thematic analysis of the stakeholders' letters, namely: ecosystems, education, liability, data availability sufficiency & protection, governance and autonomy.

5.1. Ecosystems

Enhanced collaboration and cooperation across the EU member states are desired. EU aims to develop an AI regulatory framework across the EU in the frame of the Digital Single Market, as this will eliminate the possibility that "the Member States would start to implement their own regulatory frameworks" [F528880]. Failing to achieve this will not be helpful for any of the member states in the long run as it would inhibit technology scaling and prevent Europe from speeding up the development of AI solutions and allow more advanced solutions developed worldwide to take over the EU market [F528880]. Overall, enhancing collaboration among member states on AI would assist in creating a forum for the exchange of information, best practices and provisioning of guidance when needed [F519146].

Beyond member state collaboration, the EU also needs to facilitate the cooperation and involvement of different public and private sector actors, including government, businesses and society. Europe currently lacks stakeholder and civil society engagement, and this is weakening the use of AI [F514698] and the potential for facilitating open dialogue. This prompts the desire for a more active government role in establishing and maintaining multi-stakeholder collaboration, which will also be politically backed [F514732]. This urges the EU to place "societal interests and

values at the centre of its approach towards AI, it requires robust engagement and relationships between governments and many diverse actors from civil society" [F514698]. Governments could also invest in building the necessary oversight tools and regulatory toolboxes to best address citizens' expectations and AI development and deployment [F514735]. Additionally, research collaborations on AI in Europe need to be enhanced. This could be served through the development of "dedicated lighthouse centres of research, as well as excellence and testing centres" [F516955], which will further boost the research, development and industrialisation of AI and help Europe act as a "single economic area for data and AI […] and can make a substantial contribution to the Sustainable Development Goals" [F516955].

To address global competition, stakeholders suggest that Europe should encourage the involvement of global actors to participate in European AI ecosystems. First, the lack of collaboration with leading global actors holds drawbacks for Europe's potential to participate and compete in the global market. As noted in [F514725], Europe currently lacks large tech giants compared to its counterparts as the American and Chinese technological companies. Without their involvement, Europe will miss the opportunity to build AI ecosystems that involve an integrated spectrum of actors, including essential market players. Second, Europe must collaborate with leading world organisations and associations that can serve as valuable sources of information and guidance in specific sectors. For example, in the healthcare sector, the Commission may consider "early future partnership with the World Health Organization (WHO), the International Council for Harmonisation (ICH), the International Medical Device Regulators Forum (IMDRF), and professional organisations like the Institute of Electrical and Electronics Engineers (IEEE), which is dedicated to advancing technology for the benefit of humanity". [F519146].

5.2. Education & Upskilling

AI brings many opportunities across various roles and industries. When coupled with other technological advances, e.g., Blockchain and the Internet of Things, it can support digital transformation and innovation, and lead to benefits, such as more accurate decision-making, improved forecasting and security in supply chains (Ali et al., 2021; Roe et al., 2022). However, the use of AI in different industries poses challenges to the upskilling of employees in certain industries and the ability of the general working population to respond to the imminent job market requirements. Workers in specific industries will need upskilling in order to be able to utilise AI tools. For example, in the healthcare sector, AI can guide physicians and assist them to make better informed, effective and confident decisions. However, further training of healthcare practitioners on how to use the technology in their decision making is needed [F514709]. While elevating the skills of AI researchers and developers in Europe is important, Europe's upskilling policy should also focus on providing citizens and the workforce with the qualifications necessary to allow them to apply AI and deal with the associated changes [F516955]. In particular, while AI is expected to disrupt the job market, digital literacy is low concerning AI systems, and the skills gap will impact citizens looking to secure employment [F514729]. Efforts to support the human element across the lifecycle of AI systems can help mitigate potential discrimination against the digitally excluded or the low-skilled and support the inclusion of, e.g., women, who traditionally are underrepresented in STEM jobs [F514729].

To address the skills gap, stakeholders highlight that regulation, funding, and trust should be considered with the role of the human element in mind, and specifically what skills are needed, today and for the future, for designing and developing AI-powered systems, but also using such systems with confidence. On the one hand, upskilling is needed to ensure that capacity and the

right expertise exist within the EU, as these are necessary to support the further development of AI-powered systems [F514589]. In addition, they further warn that upskilling is required within the broader ecosystem of AI systems to support the EU to catch up with "the other major powers" [F530483]. On the other hand, however, upskilling will support greater uptake of AI and wider adoption across sectors [F528943]. While "provid[ing] truthful and easy to understand representations regarding intended use and risks" can support adoption, AI systems and their outcomes must be "reasonably understood by those intended" [F530501] and build a sense of trust [F530476].

Along these lines, several recommendations are provided. For example, it is noted that “AI systems have to be part of curricula in basic medical education, specialist training and continuing medical education to broaden knowledge” [F514709], and it is highlighted that across industries, “learn[ing] new skills and competencies relevant to the new technical developments" will be required, and should thus be supported [F514589]. In fact, any occupation that involves the use of AI to develop, operate and/or make decisions […] must have the necessary expertise and appropriate-to-scale understanding of how the technology functions and its potential effects [F514777]. Such training and upskill should enable individuals to perform checks and identify whether an objective can be achieved without a significant loss in quality by using a less complex algorithmic system that involves an easier-to-understand mode of operation. Alongside upskilling AI from a technical perspective, individuals also need to be educated on promoting interdisciplinary exchange across tasks and remain open to anybody who might be interested or affected. This also involves strategies for improving the EU's "communication and visibility to retain and attract talents" in this field [F514725]. Furthermore, regarding talent acquisition, Europe will need to implement and maintain a risk-based framework as the market for such talent is

already highly competitive as companies from other countries are eager to recruit professionals in Europe. Finally, trade unions could also contribute to establishing upskilling strategies, clarifying the competencies and training needed for the workplace of the future [F514753]. As part of such upskilling and training, stakeholders propose the use of a certification system, which will allow and require that "AI engineers […] to obtain additional skills for developing complex AI systems (such as IT ethics, socio-economic impact of new technologies, legal responsibility of AI etc.)." [F514713].

5.3. Liability & Accountability

The broader use of AI requires that the framework clarify liability and accountability for this technology. In their letters, stakeholders draw on European values and fundamental rights to build a trustworthy AI. In particular, they highlight that their view of trustworthy AI is not only relevant to testable and tested technology but that this is "coupled with a proper liability regime that ensures that those who take the hazardous risk with other people's interests will pay the price and therefore think twice" [F515025]. Thus, it is necessary to develop rules on product safety and liability which will offer adequate legal certainty to businesses for the development, market launch, operation and application/use of products and services [F515025]. To address this issue, other stakeholders also propose adding AI algorithms and software as products under the Product Liability Directive to address liability [F514713]. Although the majority of the stakeholders agree with the need to address AI liability, they also highlight that overregulation may alter innovation potential and reduce interest in the development of AI in Europe [F515025].

The stakeholders also stress the need for AI systems to facilitate accountability. Accountability requires identifying an entity or person responsible for the tasks conducted using an AI algorithm

or piece of software. In the context of AI, this is complex, as multiple entities, individuals or organisations may share the responsibility, the allocation of responsibility needs to be documented and made available to all parties involved [F514777]. Alternatively, accountability by design could be another option. Such accountable systems need to "include provisions explaining their conduct and decision-making, namely choices, assumptions and trade-offs made by the people who designed this system" [F514732]. Impact assessments are also relevant to AI and algorithmic systems, and the information underlying algorithmic systems should be documented. This should eliminate the possibility of discrimination and other consequences for individuals and communities [F514777].

5.4. Data availability, sufficiency and protection

The use of AI assumes the availability and validity of data and requires practices that will protect the sharing of sensitive data while ensuring data sufficiency for AI algorithms.

First, there is currently a lack of valid data for AI tools. Stakeholders report that the data which is currently available "is often incorrect (e.g. outdated), incomplete, biased, or irrelevant" [F515025]. Despite the efforts in sharing data to address this problem, they report that access is limited and data quality is low [F519146]. Lack of data is also linked to unpredictability and lack of incentives for data sharing, such as opportunities for proof of concepts and profitable application [F528880]. Indeed, without a strict policy on data sharing, the Single European Data Space is likely to create far more problems than it solves [F515025]. To address these issues, stakeholders desire the support of the European Commission such that more complete and valid datasets can be generated. As highlighted in [F514735], "Funding the creation of these datasets and making them accessible will not only give guidance on what training data should look like but will also provide a form of

soft governance by incentivising researchers and developers to create more representative AI applications via the use of these datasets". This could assist several actors in the AI ecosystem, including developers and users, who could benefit because only a small number of actors can make valuable training data available. Along the same lines, large datasets are produced and collected in cities, and thus local authorities could also contribute to making data available [F514729]. Also, it facilitates a culture that will enable entities to utilise and understand the value of data and generate incentives that will push public and private organisations to share data [F514880]. In some sectors (e.g., healthcare), this is very valuable for the public good as it "could enable improved disease understanding, real-world monitoring for safety of pharmaceutical products, and help to realise the vision for pragmatic clinical trials that are more representative of the patients who will eventually receive a medicine in clinical practice" [F519146]. While regulation could also ensure that European values and rights are protected, Europe should carefully regulate data sharing for AI to facilitate a culture of understanding the value of data availability for innovation [F528880]. For example, AI-based decision making can be employed in the form of benefit-risk evaluation, which can provide internal decision support on safety through an AI-based tool [F514146]. In such cases, employing a set of standards and rules on how much data should be retained to document a decision at a point in time can ensure data sufficiency in AI decision making. However, given the complexity of AI tools, this might not be possible in all cases. As highlighted by [F528867] for an AI-based tool that enables an internal decision point on safety, it is difficult to determine how much data should be retained to document a decision at a point in time. "For example, if a company used AI to determine seriousness of a post-marketing safety case and an inspector enquired as to the decision process for that case, it may be difficult, if not impossible, to fully document the step-wise process, as AI continually adjusts to new data

[F528867]. Thus, setting up general guidance and an appropriate window for regulations to take effect will make compliance more feasible to maintain [F528867]. However, other stakeholders also note that the focus must be on facilitating "greater productivity, product safety, performance and objectivity" rather than implementing a strict regulation [F516955]. Overall, further actions to boost data availability would enable Europe to serve as a single data market for the benefit of all ecosystem players [F514729].

Second, the reliability and robustness of AI also rely on ensuring the data protection and security of the data employed by the algorithm. Protection against attacks, access and data manipulation should be guaranteed and be part of the system's architecture, and the system must be tested prior to implementation, whereas security precautions need to be documented [F514777]. Additionally, records of how the algorithms use data should be kept. As highlighted in [F514735], "keeping track of how data is used can be thought as similar to a factory to keep detailed records of the raw materials and their sources (the data) it uses to make its products, as well as to keep detailed records of the finished products (the outputs of the AI system)". Another solution could be the introduction of regulatory sandboxes, which will enable companies to test innovative solutions and decision-makers to understand better the real-world implications and benefits of new technologies [F528880]. This could potentially address some existing issues such as fairness and bias. However, determining how much data should be retained to document a decision is also a challenge relevant to the security of AI tools, and thus guidance on the level and type of risk mitigations need to be included in the regulatory framework [F528867]. Also, regulating the use of anonymised personal data for societal benefit, regulation needs to ensure the right balance between protecting the privacy and enabling innovation and societal benefits [F528880].

5.5. Autonomy

Automated decision making (ADM) is one of the themes that stakeholders focus their attention. The main challenges relative to ADM are transparency and accountability in decision making [F514570] and the role of human oversight [F514589, F530513]. Several stakeholders highlight the importance and the usefulness of AI systems in developing personalised services and products, automating repetitive tasks, and providing accuracy for critical tasks, such as during complex surgeries [F514709]. In other words, AI might have an assistive role in cases where human oversight is important. In such cases, systems are not necessarily replacing human intelligence but rather supplementing it by providing additional information, whereby, for example, AI might enhance a physician's expertise and co-exist with their decision-making [709].

At the same time, however, they caution that human oversight might be an important part of managing the risks associated with the sole use of AI when it comes to ADM. However, there is currently a need for more informative approaches on when and how to conduct this type of intervention and how or when to implement ADM approaches [735]. Human intervention might be required because citizens may have particular preferences regarding how and to what extent they wish to interact with machines. As described in [734], different value systems across citizens might make them more or less willing to forgo efficiencies gained by AI for more human interactions, such as in healthcare provisioning. This might be also associated with the fact that individuals might quickly feel powerless and experience a loss of autonomy as the "intrusion into their lives once such AI application becomes ubiquitous" [724].

Such issues need to be considered from an ethical and a legal perspective, particularly in light of the GDPR requirements across the EU countries and potentially local regulations that qualify these further. It is highlighted that “[s]ome of the requirements risk undermining the viability of using

AI at all" [F530489]. As mentioned, supporting the 'human in the loop' requirement can decrease the overall performance of AI systems, both in speed and quality [F514710]. However, ethically and legally, it is seen as "essential that an individual who may be subject to a purely automated decision has the right and power to challenge the decision" [F514570] and that "[t]he practitioner should always be able to understand the reasoning of the machine in order to be able to ultimately arbitrate" [F530483]. If needed, they should be able to overrule the system [F530513]. This signals a need for outputs to be reproducible, generalisable and explainable whereby practitioners, such as medical doctors, are able “to understand clearly the power and reliability of AI in their context”[F514709].

It is noted that AI systems at one point in time may function according to clearly defined frameworks and contexts, following one's instructions, but at a later point, they may evolve or be used across contexts where its developers and designers "wouldn't have been able to reasonably foresee this evolution of the AI System. (…) Therefore, the entire approach and strategy with respect to AI needs to take into account things which we may not know or fully understand at this point, but which we could reasonably predict" [F514589]. A possible mitigation against such events would be including "peer review, approval and licencing *[sic]* obligations” on top of the assessment of algorithms, which would ensure safeguarding standards [F514570].

Finally, support should be given to developing clear rules on product safety and liability matters to ensure sufficient legal certainty for the development, market launch, operation and application/use of products and services. Even if these are conceived in a technologically neutral manner, they must, if necessary, be adapted to the specific properties of AI technology. However, it must be ensured that this development does not lead to overregulation, which would inhibit

innovation and stand in the way of the intentions for the future development of AI in Europe, described in the introduction of the White Paper [F516955].

5.6. Governance

As AI becomes widely adopted, measures and recommendations on its use need to be released. This requires the establishment of a permanent secretariat that could potentially coordinate, "fund and conduct continuous measurement, assessment, and "spot check" activities, which would provide valuable information for EU citizens, elected officials, and the assembled committee of experts" [F514735]. With regards to this governance approach, stakeholders note that it risks being "fragmentary and insufficient" because one group (High-Level Expert Group on Artificial Intelligence) is tasked with drafting ethics, policy and investment recommendations and another (European AI Alliance) and for providing input to from societal stakeholders [F514589]. While different approaches are proposed for governance at the EU level, several stakeholders believe there should be localised (sector-level or country-level) approaches. For example, one recommendation is that of establishing a European Artificial Intelligence Agency (EAIA), responsible for research, coordination, budgetary issues, which will ensure collaborations and maintain an overview of implementations and which will at the same time be comprised of sector-specific local departments, specialising in, e.g., healthcare, law, energy [F514589]. Others focus more on how any governance structure creates efficiencies without resulting in administrations with overlapping powers that could potentially hinder development [F530509]. Indeed, establishing a European reference point for AI implementations will provide citizens and organisations with readily accessible means of conduct and filing complaints and taking action [F514777]. The establishment of a permanent secretariat will also assist the European Commission

in speeding up the establishment of the regulatory framework and the generation of regulatory sandboxes through which companies can test their solutions [F514880].

Yet, in light of "potential private-public collaboration on AI tools" and applications of AI applications within high-risk sectors, stakeholders raise the issue of data protection, accountability, liability and responsibility. Regarding data protection, the relevant EU legal framework needs to be identified, and GDPR seems to be most often the unequivocal regulatory framework [F514570]. As far as accountability and liability are concerned, in sectors such as healthcare, where AI failures may be critical, stakeholders warn that it should be clear where, e.g., physicians' liability begins [F514709]. Considering the potential impact of AI in relation to inequalities and potential discrimination due to automated decision-making, stakeholders indicate that minority groups (such as persons with disabilities) should be invited to engage with design and development from the very first stages as well as to participate in oversight and accountability bodies and processes [F530506].

Regarding the potential labelling of high-risk AI applications and systems within the context of B2B, stakeholders seem divided. A large proportion of them queries the usefulness of such a scheme, noting that "[a] B2B relationship is by definition based on trust between partners (supplier - user) and based on mutual contractual agreements (specifications)"; thus, such labelling would only increase the administrative burden [F530447]. On the other hand, others stand clearly against potential labelling, suggesting that it should be left “to self-regulation and code of conduct by businesses” [F530509]. Nevertheless, the majority of stakeholders pinpoint the need to follow “existing, tried-and-tested regulatory frameworks supplemented by regulatory guidance, codes of practice and industry standards” [F528968], so that AI develops further, ensuring trusted and robust governance [F530447].

Regarding governance, the High-Level Expert Group, in particular, has issued seven key requirements about governance matters that relate to human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; and societal and environmental wellbeing. Stakeholders recommend that instead of thinking in terms of 'human in the loop', the guiding principle should be 'machine in the loop'. Similarly, they focus on methodological integrity for robustness and safety and highlight that data governance, in particular, should build on GDPR's article 5, which helps protect against potential violations of human rights, cautioning that unlawful bias cannot be technologically resolved. Finally, to address the transparency requirement beyond ADM, a recommendation relates to the actual technological solutions used and developed, whereby public administrations need to require open-source software during tender procedures and provide AI systems as free software [F515025], which support auditing.

Another theme related to governance links to how AI is defined and described. AI in the White Paper is described as perceiving and interacting with the environment through the acquisition and interpretation of data, based on which an entity can draw conclusions and proceed with decision making, particularly in complex situations. However, as noted by [F516955], this description fits many traditional information systems, such as autonomous vehicles, whereby non-AI technology (e.g., traditional driving dynamics control, advanced driver assistance systems using sensor technology and radars) is employed for assessing the surroundings of a vehicle and adapting its behaviour. As such, while the White Paper describes environmental perception, adaptivity, and autonomy as core properties of AI systems, traditional technologies may exhibit similar behaviour. These risks impose regulations on products and services that are already satisfactorily regulated

by existing frameworks. AI needs to be clearly defined and differentiated from other adaptive and automated systems to avoid this. Table 1 summarises the dimensions identified.

Table 1 – Dimensions and emerging key subthemes

Dimensions	Dimension Description	Key Subthemes	Indicative Quote
Ecosystems	Collaboration between different entities on AI	Enable Europe to speed up the development of AI solutions Eliminate the possibility that Member States implement their own regulatory frameworks Enhance collaboration among member states on AI for the exchange of information, best practices and provisioning of guidance Facilitate cooperation and involvement of public and private sector Activate government role in establishing and maintaining multi-stakeholder collaboration Collaborate to build necessary oversight tools and regulatory toolboxes Facilitate research collaborations Address global competition Encourage the involvement of global actors	*"The intention to create a forum for exchange of information and best practices, and to issue guidance and opinion is sensible. In light of this, we strongly encourage the Commission to identify and communicate mechanisms to maximum stakeholder participation". [146]*
Education & Upskilling	Digital literacy on AI	Upskill employees in specific industries Upskill the general working population to be able to utilise AI skills Enlist qualifications necessary to apply AI and deal with the associated changes Mitigate potential discrimination against the digitally excluded or the low-skilled Upskill to support the EU to catch up with "the other major powers" Enhance education curricula on AI	*"Providing citizens and the workforce with the qualifications necessary to allow them to apply AI and deal with the associated changes." [955]*
Liability & Accountability	Responsibility when AI tools are used	Create trust in AI Develop liability regime Derive rules on product safety, liability and accountability Define the entity or person with responsibility for the AI algorithm Eliminate the possibility of discrimination against certain communities or individuals	*"AI that is based on robust evidences; its use must be accountable, non-discriminatory" [709]*
Data availability, sufficiency & protection	Provisioning of good quality and quantity of data to be used by AI algorithms	Share of data & ensure data sufficiency for algorithms Address issues with data validity Provide incentives for data sharing and generating valid data Support local governments to provide open data Cultivate a positive culture towards the value of data sharing and availability for innovation Ensure data protection and data security Promote a by-design data protection Develop regulatory sandboxes to facilitate the testing of AI solutions	*"Increased data access and improved data quality" [851]*
Autonomy	The extent to which AI systems should be	Facilitate transparency and accountability in decision making Enable AI human oversight & AI's assistive role Define when to intervene and how Address how GDPR & local regulations that may restrict autonomy Address risk to overrule AI systems	*"Human agency and oversight (please think in terms of 'machine in the loop' instead of 'human in the loop')". [025]*

	autonomous in each case		
Governance	Continuous monitoring and assessment of the use of AI	Support coordination, research guidance, budgeting overview implementations, providing authorisations, filing complaints. Establish a permanent secretariat and/or agency dedicated to AI Account for the human Vs the machine in the loop	*"A supervisory authority should be created to be responsible for certification and for the accreditation of entities to audit the processes with a view to certification." [844]*

This section analysed stakeholders' perceptions involved in developing Europe's AI Strategy. The following section reflects on the findings aiming to construct a realities and desires framework and derive a research agenda for future researchers.

6. Discussion: Towards an AI Realities & Desires Framework

This paper analysed stakeholder responses to the EU's consultation process towards developing its AI strategy. Our analysis consolidated and thematically analysed stakeholder letters submitted in response to the consultation and was informed by relevant literature on responsibility and ethical aspects of AI design, development and use. In their letters, stakeholders not only expressed the reality as arising from their own understanding and experience with AI, but they also made suggestions on the desired state to be facilitated through the forthcoming strategy. The analysis of the stakeholders' perceptions revealed six distinct themes that stakeholders desire to be addressed by the upcoming AI Strategy: ecosystems, education and upskilling, liability and accountability, data availability, sufficiency and protection, autonomy, and governance. Table 2 presents our results, including the desires mentioned above and the corresponding today's realities, as assessed by our synthesis of the existing literature and stakeholder perceptions.

Table 2 – Framework of AI Strategy Realities & Desires

Dimensions	Reality	Desire
Ecosystems	Lack of an orchestrator between public, private sector and society	More active government role in establishing and maintaining multi-stakeholder collaboration
Education	Urgent need for upskilling	Upskill the existing workforce and prepare the upcoming workforce
Liability	Lack of liability and accountability for AI use	Clearly define liability and accountability boundaries
Data availability, sufficiency & protection	Need for secure access to large datasets, data sufficiency & security standards	Encourage and incentivise the development and maintenance of large datasets to feed AI tools and define standards for data sufficiency, security and data usage traceability

Autonomy	Fear of losing control	Balancing between human and machine autonomy
Governance	Lack of a central reference institution	Develop a central secretariat for AI in Europe

Our choice of the term 'desire' is inspired by studies on the sociology of expectations, whereby such expectations do not always materialise but influence policymaking (Van Lente, 2012; Kerr et al., 2020). This approach has been previously applied in cloud computing, whereby the technology was analysed in terms of technological and service desires and emphasised that businesses and the broader society experience difficulties in attaining the benefits of cloud computing (Venters and Whitley, 2012). Our study indicates similar findings. Although AI has proven to be quite enchanting for businesses and governments thanks to its potential for greater efficiency and accuracy and its ability to develop insights grounded on large datasets, in reality, many stakeholders are sceptical and put forward numerous concerns. Most of these relate to the responsible and ethical design and development of AI systems, and it is desired that the EU AI strategy will address these. As such, these expectations exert significant pressure on incorporating these considerations in the forthcoming document.

6.1. Implications for research

Our analysis has identified areas that require further attention by researchers. On the one hand, our findings show that policy recommendations, technical implementation and research go hand in hand. On the other hand, empirical research data can provide ideas for future policy recommendations, and research itself may provide solutions that will help the member states reach the desired state while being in line with certain policies and strategies. Beyond contributing to policymaking, the tension between desires and realities indicate areas for future research, whereby researchers can take future action and propose tangible solutions to the current realities associated

with the implementation, adoption and use of AI. Reflecting on the findings of this paper, we develop a set of research questions that could be employed in future research. These are listed in Table 3.

Table 3 - Research questions for each dimension of AI Strategy

Theme	Research questions
Ecosystems	● How can businesses contribute to elevating the trust in AI? ● How to facilitate productive dialogue in AI ecosystems? ● What regulatory toolboxes to implement to best address citizen's expectations towards AI? ● What is the value of large tech giants' participation in AI ecosystems? ● How to encourage each stakeholder group to participate in AI ecosystems?
Education	● What should be the content of future curricula to provide supplementary AI training to the future workforce? ● How to educate existing and future employees on promoting interdisciplinary research and being open to such collaborations? ● What are the upskilling needs of the existing workforce to boost AI knowledge and digital skills?
Liability	● How can liability be evaluated in the context of AI? ● How can liability be measured in the context of AI? ● How to define the proportion of liability and accountability associated with the software owner and developer? ● What should be the content on new regulations and frameworks on AI liability and accountability?
Data availability, sufficiency & protection	● What incentives would encourage existing businesses to share data? ● What funds are needed for supporting the development of new datasets? ● What technical implementations are necessary to facilitate data sharing across AI ecosystems? ● How should data be collected, processed and managed, considering the privacy rights? ● What KPIs to use for defining data sufficiency for different AI tools? ● How can specific indicators be used to inform the user about the tool's data sufficiency level? ● What rules and regulations are needed to ensure data security and protection?
Autonomy	● Which criteria to use to balance between human and AI algorithm autonomy? ● How to best accommodate individual preferences on how and to what extent human-machine interaction should be allowed? ● How to train citizens to feel less powerless when interacting with AI tools?
Governance	● What services will be offered by the permanent secretariat on AI?

In more detail, our results revealed that there is currently a lack of an organised ecosystem, and thus, in reality, public, private and civic society organisations and the society at large work in isolation, whereby each develops their approaches for addressing AI challenges. Stakeholders desire that a forthcoming EU AI Strategy will put governments and local authorities at the centre and allocate them more responsibility for building and maintaining collaborative AI ecosystems in their area. The orchestration of private organisations, including micro-businesses, SMEs, and

international organisations (including tech giants), could elevate but equally decrease trust in AI systems (O'Brien, Jørgensen & Hogan 2020). Furthermore, the involvement of citizens, civic society organisations and the wider society could potentially eliminate or moderate the concerns of the population regarding risk, whereby views on AI's fairness and usefulness (as also identified in (Araujo et al., 2018)) are considered and contribute towards AI regulation. Academics, researchers and policymakers also hold an important role in implementing this desire as additional research questions on managing and facilitating dialogue within AI ecosystems and on the role of larger and smaller companies in these ecosystems will arise.

Furthermore, in reality, most of the EU workforce currently lacks the necessary digital literacy for using AI (Engler, 2020). This indicates a need to upskill the existing workforce and train the future workforce. In particular, stakeholders desire that a forthcoming AI Strategy will specify the approaches for upskilling the different types of the workforce to ensure that they can understand and use AI tools in the frame of their work or even find new forms of employment in a case where their previous roles became obsolete. Additionally, the future workforce needs to be appropriately prepared to use the particular technology in the context of their roles. Finally, given the continuous evolution of the technology, lifelong learning planning is also desired. Thus, beyond identifying and defining curricula for teaching AI ethics to data scientists (e.g., Garzcarek & Steuer (2020), Goldsmith, Burton(2017)), a strategy needs also to specify how to educate the existing and forthcoming workforces on using AI. Researchers may also choose to contribute to this effort by exploring the content of such curricula for the upskilling and training of the workforce and cultivating interdisciplinary collaboration.

Liability and accountability of AI tools are, in reality, undefined. In particular, stakeholders expressed their concerns on who should be liable and accountable when using AI tools.

Stakeholders' desire for the new strategy is to explicitly address this matter by clearly identifying approaches addressing liability and accountability under different AI usage scenarios, ranging from healthcare use to law enforcement and surveillance. Identifying this desire complements existing research on this topic, highlighting the lack of liability guidance regarding ethical considerations toward AI's accountability and responsibility (e.g., Jobin et al., 2019; Hagendorff, 2020). Furthermore, to reach the desired state, metrics for measuring and evaluating liability in the context of AI are expected to define the liability balance between AI user and developer. Some of these considerations have been captured indirectly by earlier studies. For example, Collins et al. (2021) draw attention to the fact that the effectiveness of machine learning and algorithms needs to be considered in greater detail, what such 'effectiveness' may look like and how it can be measured, particularly in light of a resurgence of interest in these methodologies vis a vis their societal, organisational and personal impacts.

In reality, aspects related to data availability, sufficiency and protection do not foster the more comprehensive development or use of AI tools. In particular, as raised by the stakeholders, this is partly associated with the lack of access to necessary data for the functioning of AI algorithms. Along the same lines, in their letters, stakeholders expressed their concerns on how to define and control data sufficiency for AI algorithms and their concerns on data security and protection. The latter (data security) is also reflected in existing ethical considerations (Coeckelbergh, 2019). The desire concerning data availability is for the upcoming strategy to provide incentives and even to fund enterprises to develop and, where needed, share the necessary data. To date, data sharing across organisations and enterprises has been one of the most significant challenges toward digitalisation and further AI development (Wentworth, Christie, Harriss & Charalampous, 2021). As such, identifying certain incentives that will encourage researchers to develop the necessary

datasets or foster the sharing of data across multiple stakeholders could potentially help the EU reach the desired state.

Additionally, the development of technical implementations (e.g., tracking how data is being used by each AI tool) that facilitate secure data sharing and usage within AI ecosystems are welcome. However, the forthcoming strategy could include more specific guidelines and standards for determining data sufficiency. Future research may also contribute by identifying the metrics, and KPIs needed for determining data sufficiency in each case.

Autonomy and autonomous decision-making also figured prominently in stakeholders' letters. These emerge as a desire to be able to balance human and machine decision making and interaction. Earlier work has also identified this balance as crucial for justice and fairness perceptions. For example, Bankins et al. (2022) note that the use of AI in decision-making in areas such as human resource management contexts can have negative consequences concerning one's dignity and respectful treatment and can lead to the dehumanisation of such business processes. Our findings extend this conversation further: stakeholders desire that the new strategy will determine the conditions and the mechanisms for balancing between human and machine authority and delineating the circumstances where one or the other will need to be prioritised and/or respected. Research on ethics and responsible technology could contribute towards addressing the realities of this and identify the methods and the techniques for fulfilling stakeholders' desires for clear boundaries between the two approaches to decision making, as well as towards deriving usable approaches that can accommodate user preferences on the level of desired machine interaction (where relevant).

Finally, stakeholders highlight the lack of a central authority which will act as a common information point. Their desire is to establish a central secretariat dedicated to AI matters, such

that citizens and enterprises can interact directly and resolve any issues. The development of a central secretariat on AI could address the population's perception of AI having a negative impact on employees and society (e.g., Güngör (2020)), whereby the secretariat may function as a neutral party and investigate sensitive matters confidentially and impartially. In addition, such a secretariat could function in an explanatory and facilitating role, too: for example, it could foster collaboration and coordination among different entities (Jelinek, Wallach & Kerimi, 2021) and indicate the actions that can be taken in order to demonstrate and enhance the potential benefits of AI towards society and the working population. Future research on governance could explore the role of this new governance authority.

6.2. Implications for policy and practice

The contributions of our paper are manifold. Overall, the realities of the current state of the technology and the stakeholders' desires, as identified in this study, sketch the outline of Europe's future AI strategy. Importantly, AI reliability, trustworthiness and ethical considerations frame the dimensions and guide the desires of the AI strategy. As such, our first contribution is that our analysis of the identified themes through the lens of desires and realities can inform and guide policymaking, whereby we offer a detailed description of what are the current challenges and realities of the technology and how stakeholders expect these to be addressed and hopefully solved. Second, our findings can support practitioners and industry actors in approaching the technology, its design and application by providing a picture of current concerns. Third, we believe that focusing on addressing these in future designs will result in greater transparency, inclusivity and responsibility in the AI space, which will positively influence trust perceptions towards the technology and which in turn will contribute towards increasing the uptake and adoption of AI.

6.3. Limitations and Future Research

Our analysis has been based on the data captured through the open consultation on the white paper published by the European Commission. Although the open consultation also included a survey, our analysis reflects only the letters submitted by the stakeholders. We followed this approach as the survey was directed towards specific themes, whereas letters allowed stakeholders to discuss and raise themes that genuinely relate to their interests and openly highlight their views. Following up on the open consultation, the European Commission has recently submitted a proposal for the "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" (European Commission, 2021b). This document did not form part of our analysis; however, there are indications that the open consultation indeed informed this document. As such, besides further exploring our research questions (Table 3), future research could also investigate in more detail to what extent the EU has incorporated stakeholders' desires and viewpoints in this proposed regulation, and in the longer term, to what extent these are reflected in the EU AI Strategy. In addition, based on the proposed regulation, data availability, sufficiency and protection, and AI autonomy are prominent themes that future research should focus more on. Lastly, we note that we did not proceed with a classification of the stakeholders contributing to the consultation (e.g., academics, industry). As such, we did not explore aspects that relate to conflicting agendas and power, among others. Future studies should focus more on these and potentially conduct a stakeholder analysis (e.g., Pouloudi, Currie, Whitely (2016)) as part of a larger project to unpack in more detail the relationship between desires and interests and values and policymaking choices in addressing these or not. Concluding, we highlight that in this study, we followed the approach of analysing the empirical material through the lenses of desires and realities, espousing the

sociology of expectations. We consider that alternative approaches, such as a stakeholder analysis as indicated above, would potentially result in different themes and interpretations emerging through the data analysis and a framework more focused on the stakeholders themselves rather than their major expectations.

7. Conclusion

This paper has analysed AI in terms of the desires and realities associated with the technology's current status. This analysis was based on empirical data, i.e., letters collected by the EC in the frame of an open consultation call on its forthcoming AI strategy. By synthesising this data and the existing literature, we identified six dimensions of AI that need to be addressed by the forthcoming strategy. We then discussed these dimensions into a realities and desires framework and developed a series of research questions. We expect that this framework will be helpful for policymakers who seek to understand stakeholders' desires from a forthcoming AI strategy; for practitioners, by indicating areas that require further attention; and for researchers, by signalling areas in which challenges exist and thus future research could contribute towards bridging the gap.

Conflict of Interest

We confirm that there is no conflict of interested regarding this work.

References

Afiouni, R. (2019). Organisational learning in the rise of machine learning. *Proceedings of the International Conference on Information Systems, Munich, Germany*

Alkemade, F., & Suurs, R. A. (2012). Patterns of expectations for emerging sustainable technologies. *Technological Forecasting and Social Change, 79*(3), 448-456.

Araujo, T., Helberger, N., Kruikemeier, S., De Vreese, C. H.(2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35,* 611–623. https://doi.org/10.1007/s00146-019-00931-w

Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context. Information Systems Frontiers, 1-19. https://doi.org/10.1007/s10796-021-10223-8

Benbasat, I., & Zmud, R. W. (1999). Empirical Research in Information Systems: The Practice of Relevance. MIS Quarterly, 23(1), 3–16. https://doi.org/10.2307/249403

Blomqvist, K., Van Der Werff, L. (2020). Understanding organisational stakeholder trust expectations for AI-based services. *Proceedings of the Hawaii International Conference On System Sciences 2020 (HICSS-53), Hawaii, USA.*

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology analysis & strategic management, 18*(3-4), 285-298

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Coeckelbergh, M. (2019). Artificial intelligence: some ethical issues and regulatory challenges. *Technology and regulation (May)*, 31-34, https://doi.org/10.26116/techreg.2019.003.

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

Enholm, I.M., Papagiannidis, E., Mikalef, P., Krogstie, J. (2021). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10186-w

Engler, S. (2020) Lack of skills threatens digital transformation. Gartner. Accessed: October 15th 2021, Available at: https://www.gartner.com/smarterwithgartner/lack-of-skills-threatens-digital-transformation, last accessed on 13/10/2021

European Commission (2018a). *EU Member States sign up to cooperate on Artificial Intelligence.* Accessed: September 10th, 2021. Available at: https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence

European Commission (2018b) Artificial Intelligence for Europe. COM(2018) 237 Communication: https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

European Commission, (2018c). Coordinated Plan on the Development and Use of Artificial Intelligence Made in Europe – 2018. Accessed: September 10th 2021. Available at: https://knowledge4policy.ec.europa.eu/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

European Commission (2020). *White Paper on Artificial Intelligence: a European approach to excellence and trust.* Accessed: September 10th, 2021. Available at: https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission (2021a). *A European approach to Artificial intelligence*. Accessed June 15th 2021, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Commission (2021b). *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.* COM (2021) 206. Accessed: September 10th, 2021 https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

Firth-Butterfield, K. (2021) *Building an Organizational Approach to Responsible AI*, MIT Sloan Management Review, https://sloanreview.mit.edu/article/building-an-organizational-approach-to-

responsible-ai/?utm_source=newsletter&utm_medium=email&utm_content=use%20of%20AI&utm_campaign=Enews%20BOTW%2010/08/2021, last accessed 9/10/2021

Fleck J. (1982) Development and Establishment in Artificial Intelligence. In: Elias N., Martins H., Whitley R. (eds) Scientific Establishments and Hierarchies. Sociology of the Sciences a Yearbook, 6. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-7729-7_7

Kelley, P.G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., Newman, D. T., Woodruff, A. (2021). Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries. *In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*. Association for Computing Machinery, New York, NY, USA, 627–637. DOI:https://doi.org/10.1145/3461702.3462605

German Federal Government (2018). Artificial Intelligence Strategy. Accessed: October 8th, 2021. Available at: https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf.

Griva, A., Kotsopoulos, D., Karagiannaki, A., Zamani, E.D. (2021). What do growing early-stage digital start-ups look like? A mixed-methods approach. *International Journal of Information Management*.

Güngör, H. (2020). Creating Value with Artificial Intelligence: A Multi-stakeholder Perspective. *Journal of Creating Value, 6*(1), 72-85.

Jelinek, T., Wallach, W., & Kerimi, D. (2021). Policy brief: the creation of a G20 coordinating committee for the governance of artificial intelligence. *AI and Ethics, 1*(2), 141-150.

Kastner, W., Kofler, M. J. and Reinisch, C. (2010). Using AI to realise energy efficient yet comfortable smart homes, *In Proceedings of the 2010 IEEE International Workshop on Factory Communication Systems Proceedings*, 169-172, doi: 10.1109/WFCS.2010.5548612.

Karagiannis, I., Mavrogiannis, K., Soldatos, J., Drakoulis, D., Troiano, E., Polyviou,A. (2020). Information Sharing and Stakeholders' Collaboration for Stronger Security in Financial Sector Supply Chains: A Blockchain Approach. In Soldatos, J. (ed.), Philpot,J. (ed.), Giunta, G. (ed.), Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures, 76-93. Boston-Delft: now publishers.

Kerr, A., Barry, M., & Kelleher, J. D. (2020). Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. *Big Data & Society, 7*(1), 2053951720915939.

Melnychenko, O. (2020). Is Artificial Intelligence Ready to Assess an Enterprise's Financial Security? *Journal of. Risk Financial Management, 13*(9), 191. https://doi.org/10.3390/jrfm13090191

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. Information Systems Management, 39(1), 53-63.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualisation, measurement calibration, and empirical study on its impact on organisational creativity and firm performance. *Information & Management*, *58*(3), 103-434.

O'Brien, M. C., Jørgensen, R. F. and Hogan, B. F., Tech Giants: Human Rights Risks and Frameworks (December 15th, 2020). Accessed: October, 15th, 2021. Available at: https://ssrn.com/abstract=3768813

Pouloudi, N., Currie, W., & Whitley, E. A. (2016). Entangled stakeholder roles and perceptions in health information systems: a longitudinal study of the UK NHS N3 network. *Journal of the Association for Information Systems, 17*(2), 107-161.

Qian Sun, M., Medaglia, R. (2019).Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly, 36*(2), 368-383.

Rességuier A, Rodrigues R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*.10.1177/2053951720942541

Rizzi F.T., Pera A. (2020) Balancing tests as a tool to regulate artificial intelligence in the field of criminal law. In Special collection on artificial intelligence UNICRI. Accessed: October, 15th, 2021. Available at: http://www.unicri.it/node/3228.

Roe, M., Spanaki, K., Ioannou, A., Zamani, E. D., & Giannakis, M. (2022). Drivers and challenges of internet of things diffusion in smart stores: A field exploration. Technological Forecasting and Social Change, 178, 121593. Doi: https://doi.org/10.1016/j.techfore.2022.121593

Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, *19* (1), 61-86.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., ... & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualisation and operationalisation. *Quality & quantity*, *52*(4), 1893-1907

Shneidermam., B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, International *Journal of Human–Computer Interaction, 36*(6), 495-504.

Sigov, A., Ratkin, L., Ivanov, L.A. & Xu, L. D. (2022). Emerging Enabling Technologies for Industry 4.0 and Beyond. Information Systems Frontiers https://doi.org/10.1007/s10796-021-10213-w

Stahl, B. (2021) Artificial Intelligence for a Better Future. An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. Accessed: October, 15th, 2021. Available at: https://link.springer.com/book/10.1007%2F978-3-030-69978-9.

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhé Shaelou, S., Patel, A., Ryan, M., & Wright, D. (2021). Artificial intelligence for human flourishing – Beyond principles for machine learning. Journal of Business Research, 124, 374–388.

Stahl, B.C., Antoniou, J., Ryan, M., Macnish, K., Jiya, T.(2021). Organisational responses to the ethical issues of artificial intelligence. *AI & Society*. https://doi.org/10.1007/s00146-021-01148-6

Trocin, C., Mikalef, P., Papamitsiou, Z., Conboy, K. (2021). Responsible AI for Digital Health: a Synthesis and a Research Agenda. Information Systems Frontiers. https://doi.org/10.1007/s10796-021-10146-4

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, 10.1080/14494035.2020.1855800

Van Lente, H., Spitters, C., & Peine, A. (2013). Comparing technological hype cycles: Towards a theory. *Technological Forecasting and Social Change, 80*(8), 1615-1628.

Van Lente, H. (2012). Navigating foresight in a sea of expectations: Lessons from the sociology of expectations. *Technology Analysis & Strategic Management, 24*(8), 769-782.

Venters, W. and Whitley, E. (2012). A Critical Review of Cloud Computing: Researching Desires and Realities. *Journal of Information Technology*, 27(3) 179-197.

Wentworth, J., Christie, L., Harriss, L. and Charalampous, R. (2021). Energy Sector Digitalisation, POST note 655, UK Parliament POST, Accessed: October 15th, 2021, Available

at: https://researchbriefings.files.parliament.uk/documents/POST-PN-0655/POST-PN-0655.pdf last accessed 13/10/2021

Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences*, 376(2133).

Yu, K.H., Beam, A.L. & Kohane, I.S. (2018). Artificial intelligence in healthcare. *National Biomedical Engineering,* 2, 719–731, https://doi.org/10.1038/s41551-018-0305-z