ECIS 2022 Research Papers

ECIS 2022 Proceedings

6-18-2022

TOWARDS EUROPE'S AI STRATEGY: DESIRES & REALITIES

Ariana Polyviou
University of Nicosia, ariana.polyviou@gmail.com

Efpraxia Zamani
University of Sheffield, e.zamani@sheffield.ac.uk

Follow this and additional works at: https://aisel.aisnet.org/ecis2022_rp

TOWARDS EUROPE'S AI STRATEGY: DESIRES & REALITIES

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

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

Abstract:

Artificial Intelligence is possibly the most talked about topic among all emerging and disruptive technologies. Being under continuous development for the last 60 years or more, AI promises to provide benefits within a number of domains, including healthcare, finance security and energy, to name only a few. In this paper, we explore stakeholder perceptions with regards to the forthcoming European AI Strategy, as expressed during the European Commission's formal consultation process. We follow a qualitative inquiry to analyse our data and identify emerging themes. Our findings reveal six dimensions that are considered crucial for a responsible and ethical AI strategy: ecosystems; education; liability & accountability; data availability, data sufficiency & protection; and governance. We elaborate on these dimensions and develop a desires-realities framework. Our work contributes to a better understanding of stakeholder expectations from the AI strategy and holds important implications for research on addressing grey areas and overcoming the realities in the AI domain.

Keywords: Artificial Intelligence, Strategy, Ethics, technology regulation, responsible innovation.

1   Introduction

Artificial intelligence (AI) is expected to benefit organisations, citizens and the society as a whole and transform lives and processes for the better. For example, AI is expected to revolutionise disease diagnosis (Yu, Beam, & Kohane, 2018), to support the efforts towards energy efficiency (Kastner, Kofler, Reinisch, 2010), and to enhance the security of critical systems (Karagiannis et. al., 2020), to name only a few of the envisaged benefits. However, not all digital tools (e.g., dark web, fake news etc.) favor humanity and the same applies to AI, too. It is thus important to carefully craft strategies that support fail-safes and promote AI's wider deployment in a way that is supportive for both the economy and the society. In other words, it is crucial for the AI domain to crutinise its black boxed algorithms and that regulation is fit for purpose, so that academia and industry move beyond the risks of uncontrolled algorithms and work towards responsible and trustworthy AI.

Despite the extensive work undertaken in the AI domain, several AI definitions exist today, which is not conducive for discussing strategy and policies around AI. In this paper, we adopt the definition developed by the European's Commission (EC) and introduced through its 'White Paper on Artificial Intelligence: A European approach to excellence and trust' (European Commission, 2020): "AI is a collection of technologies that combines data, algorithms and computing power" (European

Commission, 2020, p.3). EC's White Paper indicates that AI is among its top priorities for its digital agenda. On the basis of its technological and industrial strengths and high-quality digital infrastructure, Europe has the potential to become an innovation leader in the data economy. To facilitate this, the EC has put together a list of actions included in its 'European approach to excellence in AI'. Further, with the view to identify Europe's vision and priorities on AI (European Commission, 2021a), the EC led a consultation on AI between February and June 2020, inviting relevant stakeholders to share their opinions and views on its recently published White Paper (European Commission, 2020). The White Paper explores the different policy options for addressing human and ethical implications of AI and for supporting the use of big data in innovation. More specifically, the White Paper consists of two building blocks:

    i)   a policy framework towards an ecosystem of excellence which will encapsulate the research and innovation community, Small-Medium Entreprises, partnerships with the private sector, adoption of AI by the private sector and collaboration with like-minded global players, and

    ii)   the 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.

Data arising from the public consultation and related activities has recently contributed to the publication of the European Commission's Proposal for a Regulation on Artificial Intelligence (AI) Systems (European Commission, 2021b).

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). Existing literature has considered the perceptions of individuals or employees of organizations towards AI (e.g., Spanaki et al., 2021). Yet, most often studies emphasize on identifying the general population's perception on AI or documenting advances in research and practice, rather than extracting views on how to facilitate AI for trust and excellence or for identifying the challenges around the implementation of AI (Enholm et al., 2021, Ulnicane et al., 2021). Our work complements existing literature through empirical data supplied by key stakeholders and enriches it by identifying pathways for future research and development. The contribution of this work is timely as there is currently increased focus to AI and AI-enabled information systems and the need of governments, national and international agencies and the industry as a whole to reflect on the needs and expectations of their stakeholders and those impacted by the technology when designing policies, products and services. 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 engaged in the process of expressing their views on how to reach AI excellence for the forthcoming European strategy. They accepted the EC's invitation to submit their letters via the consultation systems.

The objective of this paper is to shed light on the perceptions of European stakeholders on Europe's AI strategy. This is framed by the paper's research question: "What are the needs and expectations of stakeholders with regards to the EU's AI strategy?" Drawing on a qualitative approach, the paper analyses the perceptions of the stakeholders, and maps them into six dimensions: ecosystems; education; liability & accountability; data availability; data sufficiency & protection; and governance. We then reflect on the findings with regards to the realities and expectations from the AI Strategy.

The paper is structured as follows. Section 2 provides an overview of existing empirical studies on the perceptions of stakeholders. Section 3 describes the methodological approach employed for this study. Then, Section 4 presents the findings of our analysis as arising throughout the stakeholder letters. Section 5 discusses the findings with respect to existing literature and derives a framework mapping the

expectations and desires from Europe's AI strategy. In the last Section we summarize the key findings of this research.

2 Background

Over the recent years, researchers have focused on AI and approached this topic by sharing their perceptions and shedding light on key issues, such as reliability, trustworthiness and ethics.

Existing research has focused on deriving empirical data to explore the perceptions of the public towards AI. For example, in Kelley (2021), the author explored the perceptions of over ten thousand citizens in eight countries and six continents, aiming to shed light on the views of individuals on AI. The analysis identifies different sentiments associated with AI responses (exciting, useful, worrying, and futuristic). Similarly, in Güngör (2020), the author focused on examining the perceptions of professionals in Europe. Their research highlighted that there is a positive perception towards using AI for financial benefit and value creation. However, stakeholders perceive that employees and the society will be negatively impacted by the wide adoption of AI. In Araujo et al., (2020) the authors draw on a survey to explore the AI-based service trustworthiness in organizational stakeholders. Their results highlight that the attitudes are associated with concerns about risk, and mixed views on AI's fairness and usefulness. 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 being equal or superior for taking high-impact decisions. Other studies also explored the perceptions of stakeholders on the use of AI in specific niches domains, e.g., financial security (Melnychenko, 2020), public sector (Qian Sun & Medaglia 2019), health care sector (Blomqvist & Van Der Werff, 2020).

Existing research has reviewed different perspectives on 'what ethical AI is' and summarizes existing suggestions regarding principles and guidelines for ethical AI. For example, Jobin, Ienca, & Vayena, (2019) identify the converging views on ethical principles of AI. In particular, the paper identifies five distinct ethical principles including transparency, justice and fairness, non-maleficence, responsibility and privacy. The paper also highlights that despite the emergence of the five principles, there is notable divergence on how each of these principles is 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 the majority of ethical considerations are relevant to aspects of accountability, privacy or fairness. Others emphasize specific aspects of ethical AI, such as education and teaching AI to data scientists (e.g., Garzcarek & Steuer, 2020, Goldsmith & Burton, 2017), whereas discussions have been initiated on ethics washing as a means for avoiding or escaping governmental regulations (Rességuier & Rodrigues, 2020).

Literature also focuses on the design of systems that can facilitate reliability and trustworthiness by design. For example, in Shneidermam (2020), the author emphasizes human-centered AI and suggests a framework which can assist the balancing between human control and computer automation, identify situations in which full human or full machine control is required and minimize the risks arising by excessive human or machine control. In Güngör (2020), researchers introduce a framework which facilitates the interplay of human values, interpersonal dynamics and the socially situated nature of AI systems.

Our work complements existing literature and existing knowledge on public the perceptions on AI by shedding light on the realities of potential issues of using AI and desires on how these could be best addressed in a forthcoming AI strategy. Additonally, it adds to this stream of work highlighting the considerations of European stakeholders on these aspects and their implications of how to better

accommodate these considerations in a forthcoming AI Strategy in Europe. The next Section describes the methodological approach followed for this study.

3 Methodology

In this study, our aim is to explore how a human-centric AI strategy could be designed and developed. As such, we consider that a qualitative inquiry is appropriate because it enables a deep exploration and developing rich insights. Our study draws on secondary data, i.e., letters submitted by stakeholders during the European Commission's open consultation on AI. In particular, on 19 February 2020, the European Commission launched a Consultation on Artificial Intelligence. Citizens and stakeholders were invited to provide their feedback until 14 June 2020 with regards to their views on a proposed White Paper on Artificial Intelligence published earlier in 2020 (European Commission, 2020). Thereby to contribute on shaping the forthcoming AI strategy. Contributing participants included EU and non-EU citizens, Member States and other relevant stakeholders (including civil society, industry and academics) who provided their opinions, concerns and recommendations via uploading their letters on the Commission's online platform. Our initial dataset included 422 letters, out of which we excluded 92 (due to these not having been written in English, being very short, i.e., a paragraph long or shorter, and/or being research papers and other multimedia not directly responding to the consultation). Thus, only 330 letters were finally considered for our analysis. 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 largely the same, suggesting that we had achieved saturation in our analysis (Saunders et al., 2017). To distinguish each letter, in our analysis we employed three-digit number (e.g., [111]).

The letters were analyzed notably identifying the desires of the stakeholders for a human-centered AI strategy. In this choice, we were inspired by the Sociology of Expectations (Van Lente, 2012; Kerr et al., 2020), as applied by Venters and Whitley (2012) in their study on cloud computing. This approach allows us to develop a conceptual framework that guides us in intefying and synthesising the expectations (or desires) of both the EC and the stakeholders of the European AI strategy. As such, desires signify the expectations of stakeholders, i.e., what they hope that the ensuing strategy will incorporate and what it will achieve in its endeavour to facilitate AI research and innovation in the future. Such desires often shape the development and regulation of technologies, yet they are subject to the realities of the technology. In other words, the capabilities of AI influence stakeholder perceptions both in terms of what stakeholders' regard as possible and what they consider to be a desired state in the future. Therefore, this approach allows us to acknowledge both what is desired, but also the extent to which such desires are feasible (Venters and Whitley, 2012; Borup et al., 2006). Whether desires will be addressed will largely depend on a number of factors, such as conflicting agendas, the stakeholder's value and their relative weight within the context of the consultation and the emerging AI strategy (Alkemade and Suurs, 2012). Thus, adopting this approach allows us to elaborate on today's expectations, consider the realities of AI and in turn identify the challenges that need to be addressed towards reconciling expectations and realities.

Figure 1 – Data Analysis Approach

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). As illustrated in Figure 1, we began our analysis by familiarizing ourselves with the collated material to support the snuring coding process. This was followed up by initial coding phase where preliminary coding scheme was developed. After the initial coding phase, themes were identified. Codes were compared to existing literature, revised and additional sub-codes were identified. The themes were reviewed against our dataset and counter-examples for codes were also derived among the authors. Based on the output of this effort, the coding of the data was reviewed including an iterative process that involved reading, coding and questioning the validity of the codes. This led to a list of high-level thematic codes, which we then write up in our analysis as the major six dimensions. 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.

4   Analysis

This Section analyses the qualitative data arising from the stakeholders' letters. The findings are organized in six dimensions.

Ecosystems

In their letters, the stakeholders highlight the need for establishing a constant collaboration between government, businesses and the society. As discussed in [698], putting "societal interests and values at the center of its approach towards AI, it requires robust engagement and relationships between governments and many diverse actors from civil society". Indeed, supporting such a dialogue requires not only stakeholder involvement, but also political backing ([732]). Currently in Europe there is lack of stakeholder and civil society engagement and thus this is weakening the use of AI [698]. This prompts for the desire for a more active government role in establishing and maintaining multi-stakeholder collaboration. Governments could potentially also invest in building the necessary oversight tools and regulatory toolbox's to best address citizen's expectations and AI development and deployment [735]. Beyond the role of government in building ecosystems, to address global competition Europe also needs to encourage the involvement of large multinational companies in building such ecosystems. As noted in [725] Europe currently lacks large tech giants compared to its counterparts as the American and Chinese technological companies. However, without their involvement, Europe will miss the opportunity to build AI ecosystems that involve an integrated spectrum of actors which includes important market players.

Education

AI brings a lot of opportunities for decision support across various roles and industries. For example, in healthcare, AI can guide physicians and assist them to take mode informed, effective and confident decisions [709]. However, this also highlights that "AI systems have to be part of curricula in basic medical education, specialist training and continuing medical education to broaden knowledge" [709]. Beyond medical professions, any occupation that involved 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 [777]. Training needs also to enable individuals to perform checks and identify whether an objective can be achieved without a significant loss in quality through the use of a less complex algorithmic system that involves an easier to understand mode of operation. Of course, the use of AI in different industries, also poses challenges on the ability of the current working population to respond to the openings of the job market. While AI is expected to disrupt the job market, there is currently lack of digital literacy as well as skills that will enable citizens to secure employment and also ensure gender equality when it comes to job opportunities [729]. Alongside upskilling existing population on AI from a technical perspective, individuals need also to be educated on promoting

interdisciplinary exchange across tasks and remain open for anybody who might be interested or affected. This also involves strategies for improving EU's "communication and visibility to retain and attract talents" in this field [725]. Finally trade unions could also contribute to the establishment of upskilling strategies, clarifying the competences and training needed for the workplace of the future [753].

Liability & accountability

The wider use of AI require the guarantee for security and safety. This encapsulates the needs for liability and accountability. From the one hand as suggested by [713], this requires the addition of AI algorithms and software as products under the Product Liability Directive. It also includes the requirement for AI systems to facilitate accountability. Accountability requires the identification of an entity or a person who holds the responsibility associated to the tasks conducted using AI algorithm or piece of software. Although this in the context of AI, this is complex due to the fact that multiple entities, individuals or organizations may share the responsibility, the allocation of responsibility needs to be documented and made available to all parties involved [777]. 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" [732]. 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 [777].

Data availability, sufficiency and protection

The use of AI assumes the availability, sufficiency and protection of the data. Currently there is lack of sufficient data availability for AI tools. The European Commission could contribute by fostering and funding the creation of datasets. As a result, several actors in the AI ecosystem including developers and users could benefit from these datasets as a small number of users have the capacity to make useful training data available. As highlighted in [735], "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 incentivizing researchers and developers to create more representative AI applications via the use of these datasets". Along the same lines, as a large number of data are produced and collected in cities, local authorities could also contribute to making data available [729]. Also, facilitating a culture which will enable entities to utilize and understand the value of data and also generate incentives which will push public and private organizations to share data [880]. Thus, Europe could serve as a "single market for data needs to benefit all ecosystem players" [729]. This vision could be accompanied by a set of clear rules for how data is collected, processed and managed, considering the privacy rights [753]. This is especially applicable in cases where vulnerable parts of the population are involved (e.g., children) and where key actors involved are not only schools and teachers but also education platforms, data brokers etc. [778].

Beyond fostering the availability of data, the use of AI for decision making also requires the definition of the appropriate standards for data sufficiency. Drawing on the argumentation of [146] AI-based decision making is very useful in context of decision-making of benefit-risk evaluation such as for example an AI-based tool which would provide internal decision support on safety. However, it is important to define a priori through a set of standards and rules, how much data should be retained to document a decision at a point in time. Additionally, the reliability and robustness of AI also relies on ensure the data protection and security of the data employed by the algorithm. Thus, protection against attacks, access and data manipulation must be guaranteed and be part of the architecture of the system and system must be tested prior to implementation whereas security precautions need to be documented. [777]. Another requirement is to keep records of how data is being used by the algorithms. As highlighted in [735], 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)". This could potentially address some of the existing issues such as fairness and bias.

Autonomy

Beyond direct decision making, AI might also have an assistive role in cases in which human oversight is important. In such cases, systems are not necessarily human intelligence, but rather supplementing it by providing additional information. For example, AI might enhance physician's expertise and may co-existing with physician's decision-making [709]. In such cases human oversight might be an important part of managing the risks associated to sole use of AI. However, there currently need for more informative approaches on when and how to conduct this type of intervention [735]. Additionally, human intervention might be required because of citizens may have particular preferences of how and to what extent to interact with machines. As described in [734] different value systems across citizens might make them more or less willing to forego efficiencies gained by AI for more human interactions such as for example 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].

Governance

As AI becomes widely adopted, measures and recommendations on its use needs to be released. This requires the establishment of a permanent secretariat could which 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" [735]. Establishing a European reference point for AI implementations will provide to the citizens and organizations for easily accessible means of conduct as well as for filing complaints and taking actions [777]. The establishment of a permanent secretariat will also assist the European Commission in speeding up the establishment of the regulatory framework as well as the generation of regulatory sandboxes through which companies can test their solutions [880].

The content of the letters provided by the stakeholders converged towards six dimensions. These are: ecosystems; education; liability & accountability; data availability; data sufficiency & protection; and governance. Ecosystems refer on the synergy among different entities on topics related to AI. Education refers to the development of the necessary skills for using AI in different industries. Liability and accountability refer to assigning responsibility on the use of AI tools such as security and safety can be preserved. Liability and accountability refers to the need for defining the entity to be liable in each case AI algorithms or tools are being employed. Data availability, sufficiency and protection is relevant to ensuring the provisioning of the amount of data required for AI tools. Autonomy is concerning the extent to which in each case the use of AI tools will compromise human interaction. Finally, Governance is more relevant to the establishment of a procedure or an entity which will overlook the further development of AI and intervene in case further regulation or control is required. Table 1 summarizes each of the dimensions and provides an indicative quote for each case.

Dimensions	Description	Indicative Quote
Ecosystems	Collaboration between different entities on AI	*"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	Digital literacy 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	*"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	*"Increased data access and improved data quality" [851]*
Autonomy	Defining the extent to which AI systems should be autonomous in each case	*"Human agency and oversight (please think in terms of 'machine in the loop' instead of 'human in the loop')". [025]*
Governance	Continuous monitoring and assessment of the use of AI	*"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]*

Table 1 – Description for each dimension

This Section analyzed the perceptions of stakeholders involved in the development of Europe's AI Strategy. The next Section reflects on the findings aiming to construct a realities and desires framework for AI.

5   Towards an AI Realities & Desires Framework

Our analysis of the stakeholders' perceptions lead to the identification of six distinct dimensions for the upcoming AI Strategy. In their letters, stakeholders not only expressed the identified issues as arising by their own understanding and experience with AI, but they also made suggestions on the desired state to be made available through the forthcoming strategy. The desires and realities of each dimension is listed in Table 2. Our choice of the term 'desires and realities' (inspired by Venters & Whitley (2012) approach on analyzing cloud computing into technology and service desires back in 2021), emphasizes that currently businesses as well as the wider society are experiencing difficulties in exploiting AI and experience its benefits. While the allure of AI relates to greater efficiency and accuracy, in reality many stakeholders are skeptical and have numerous considerations which need to be addressed throughout the AI strategy.

Currently there is lack of organized ecosystems through which public and private organizations as well as the society can receive guidance on how make use and benefit by AI. Stakeholders desire that a forthcoming AI Strategy will put governments and local authorities at the center, thus giving them more responsibility on building and maintain collaborative AI ecosystems in their area. The involvement of the wider society could potentially eliminate the concerns of the population regarding risks, and as well as the views on AI's fairness and usefulness which are identified in (Araujo et al, 2020). Besides the role of policy making, researchers also hold an important role in implementing this desire as additional research questions on managing and facilitating dialogue within AI ecosystems as well as on the role of larger and smaller companies in these ecosystems.

Furthermore, the use of AI highlights the gap in terms of digital literacy, upskilling of existing and training of forthcoming workforce. The desire from the AI Strategy is to derive specific approaches for upskilling the different types of workforce to ensure that they are able to understand AI and use AI tools in the frame of their work or even find new forms of employment in case their previous roles became obsolete. Additionally, upcoming workforce needs to be appropriately prepared in order to be able to use technology in the context of their roles. Given the continuous evolvement of the technology life-

long learning planning is also necessary. Thus, beyond identifying and defining curricula for teaching AI ethics to data scientists (e.g., Garzcarek & Steuer (2020), Goldsmith, J., Burton (2017)), strategy needs also to specify how to educate the existing and forthcoming workforces on using AI. Researchers may also choose to contribute in this effort by exploring the content of such curricula for the upskilling and training of the workforce as well as for cultivating inter-disciplinary collaboration.

Liability and accountability was also raised by the stakeholders. In particular, stakeholders expressed their concerns on liability and accountability associated with the use of AI tools. In their letters, they expressed their desire for the new strategy to address this matter by clearly identifying approaches addressing liability and accountability under different AI usage scenarios. Existing research also identifies this in terms of ethical considerations towards AI's accountability and responsibility (e.g., (Hagendorff, 2020; Jobin, Ienca, & Vayena, 2019). Our paper thus complements the findings of existing research on these matters by shedding light on a list of desires which could potentially address accountability and liability through an AI strategy. Further to the policy, future research might also consider new approaches for measuring and evaluating reliability associated to the use of AI tools.

Data availability was another dimension arising by our analysis. As raised by the stakeholders this is associated with the lack of access to the data necessary for the functionality of AI algorithms. Their desire concerning data availability is for the upcoming strategy to providing incentives and even funding for enterprises to develop or even share the necessary data. Beyond policy makers, this provides useful directions for research. Future researchers may focus on providing technical implementations which can ease the secure data sharing and usage within AI ecosystems.

Data sufficiency and security was also raised by the stakeholders. In particular, in their letters they expressed the concern on how to define and control data sufficiency for AI algorithms. They expressed the desire for more specific guidelines and standards for determining data sufficiency. In addition, they stressed their concerns on data protection and expressed their desire to keep track on how data is being used by each AI tool. Security is also reflected in existing ethical considerations in (Hagendorff, 2020; Jobin et al., 2019). Our work extends our understanding of this matter by clarifying the stakeholders desired approaches for addressing such issues. Researchers may also contribute to identifying the metrics and key performance indicators (KPIs) needed for determining data sufficiency in each case.

Autonomy was an important concern identified, which regarded the ability to balance between human and machine decision making and interaction. In this respect, their desire from the strategy is to determine under which circumstances and how to balance between human and machine authority. Researchers could also potentially contribute to identifying methods for the better balancing between human and machine autonomy as well as deriving usable approaches accommodating user preferences on the level of desired machine interaction (if applicable).

Finally, stakeholders highlight the lack of a central authority which will act as a common information point. Their desire is to develop a central secretariat dedicated to AI matters, such that citizens as well as enterprises can interact directly and resolve any issues. The development of a central secretariat on AI, could potentially also contribute to addressing the population's perception on AI having a negative impact on employees and society (e.g., Güngör, (2020)) as it could potentially demonstrate how action can be taken to enhance the potential benefits of AI towards society and the working population. Future research on governance could be on exploring the role of this new governance authority.

Overall, AI reliability, trustworthiness and ethical considerations are framing the dimensions and desires associated with the AI strategy. Additionally, policy recommendation and research go hand in hand. Empirical data provide ideas for future policy recommendations whereas research may also provide

solutions for the implementation of certain policies and strategies. In sum, despite the considerations raised by the stakeholders, their desire for a Strategy which could potentially resolve the issues raised.

Dimensions	Reality	Desire
Ecosystems	Lack of an orchestrator between public, private sector and society	Activate government role in establishing and maintaining multi-stakeholder collaboration
Education	Urgent need for upskilling	Upskill existing workforce and prepare upcoming workforce
Liability & Accountability	Lack of liability and accountability for AI use	Define liability and accountability boundaries
Data availability, sufficiency & protection	Need for secure access to large datasets Lack of data sufficiency & security standards	Encourage and incentivize the development and maintenance of large datasets to feed AI tools Define standards on data sufficiency, security and data usage traceability
Autonomy	Fear of losing control	Balance between human and machine autonomy
Governance	Lack of a central reference institution	Develop a central secretariat for AI in Europe

Table 2 – Framework of AI Strategy Realities & Desires

6   Conclusion

In this paper, we analyzed stakeholder views and opinions regarding the upcoming European strategy for AI in terms of their desires and the realities associated with the technology's current status. This analysis was based on empirical data, i.e., stakeholder letters collected by the EC in the frame of an open consultation call on AI strategy. Through the synthesis of this data, we identified six dimensions of AI which need to be addressed by the forthcoming strategy: ecosystems; education; liability & accountability; data availability, sufficiency and protection; autonomy; and governance. We then analyzed these dimensions into a realities and desires framework. We foresee that this framework will be useful in aiding policy makers towards understanding stakeholder desires from a forthcoming AI strategy, but also in providing a sense of direction and suggestions for addressing grey areas and ways forward for overcoming the realities in the AI domain that currently function as obstacles for future responsible and ethical development of the technology. Similarly, for researchers, the framework provides a means of comparing the current and desired status relevant to the dimensions of AI. One of the limitations of this study is that it has focused on analyzing solely consultation letters and did not take into account the 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" recently released by the European Commission. This document did not form part of our analysis; however, there are indications that the open consultation indeed informed this document. Additonally, 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. We consider that additional research is needed to explore how an AI strategy could better address stakeholders' concerns and how cost–benefit calculations on the use of AI might be undertaken. Thus, further research may focus on the conflicting agendas and priorities of stakeholders and how these will shape and be depicted in the resulting European AI strategy. Finally,

this paper did not test this framework empirically. Future researchers may consider following a quantitative approach for validating the proposed framework.

References

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.

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

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.

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

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

Garzcarek, U., Steuer, D. (2020) . Approaching ethical guidelines for data scientists. *arXiv*, 1–18.

Goldsmith, J., Burton, E..(2017) Why teaching ethics to AI practitioners is important. *ACM SIGCAS Computers and Society*, 110–114.

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.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines,* 30, 99–120.

Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat Mach Intell* , 1, 389–399.

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.

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

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

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

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., ... & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *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.

Spanaki, K., Karafili, E., Despoudi, S. (2021). AI applications of data sharing in agriculture 4.0: A framework for role-based data access control. *International Journal of Information Management, 59*,102350.

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.

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. (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.

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