

# Tackling bias in AI and promoting responsible research and innovation: Insights from discussions with different stakeholders

Kalypso Iordanou<sup>1</sup> and Josephina Antoniou<sup>2</sup>  
School of Sciences, University of Central Lancashire (UCLan) Cyprus

## Abstract

Responsible research and innovation is critical when it comes to AI and Big Data technologies, and their integration in society. Engaging different stakeholders in the dialog on how to address the ethical challenges of AI and tackle AI bias, is fundamental for finding viable solutions that would ensure the alignment of AI with core ethical values and would support individual and societal wellbeing. The chapter presents data from two qualitative studies that took place a few years apart, each under the umbrella of a different European project, namely the EU Horizon 2020 COMPASS project (2016-2019) and the EU Horizon 2020 SHERPA project (2018-2021). The aim of these studies was to capture different stakeholder views on AI, Big Data and Responsible Research and Innovation (RRI), with a particular focus on the dimension of ethics. While the COMPASS data collection focused on the responsible use of emerging technologies in specific sectors, i.e. healthcare and nanotechnology, the SHERPA data collection focused on the specific technologies themselves, i.e. AI and Big Data. Thirty Industry leaders, from the healthcare and nanotechnology sectors across Europe, participated in in-depth interviews, in the context of COMPASS. For SHERPA, six focus groups (N = 49, in total) were conducted that focused on discussing proposed guidelines for the responsible development and use of such technologies as well as requirements for AI regulation in Europe. The interviews and focus groups featured a broad range of stakeholders from different sectors, such as technology experts and stakeholders, industry leaders, technology practitioners and social scientists. The projects aimed to capture different perspectives and varying expertise related to the aims of the projects. The wide range of perspectives is important to capture in order to understand how the

---

<sup>1</sup> Kalypso Iordanou holds a PhD from Columbia University, New York, USA

<sup>2</sup> Josephina Antoniou holds a PhD from University of Cyprus, Cyprus.

implementation of smart information systems in different sectors impact society and business. Interestingly enough, juxtaposing the results from the two data collection phases revealed some important common recommendations, highlighting the role of Education and Regulation, in moving forward with responsible and ethical development and operational use of emerging technologies such as AI and Big Data.

Keywords: Responsible Research and Innovation, AI, Big Data, education, regulation, ethics, society

## Introduction

Biases in the information provided by AI algorithms, perpetuating gender and racial stereotypes (Otterbacher, 2018), coupled with human's tendency to process information through lenses protecting their initial beliefs and biases (Iordanou et al., 2020), favor one-sided thinking, extremism, and fanaticism, often leveraged in racist ways, and seldom to indict white supremacy (Cave & Dihal, 2020) and Western imperialism, with deleterious effects on democracy and the societal wellbeing. Several ethical issues have been acknowledged arising from AI and Big Data in the literature, including exploitation of behavioural biases, deception, and addiction generation to maximize profit (Costa & Halpern, 2019), manipulation (Helbing et al., 2019), spread of misinformation, hate speech and conspiracy theories (Scheufele & Krause, 2019). The algorithmic filtering, which refers to prioritizing the selection, sequence, and the visibility of posts (Bucher, 2012), is viewed as a reinforcer of individuals' worldviews, biases and polarization (Loader, Vromen, & Xenos 2016; Gillespie, 2018; Helbing et al., 2019). In addition, the emerging technologies have been accused for encapsulating the worldviews and biases of their creators (Broussard, 2018; Noble, 2018) or the data they rely on (Cave, 2019). The result of algorithmic bias or bias in Big Data is the replication of biases, stereotypes and biased decisions (Pols & Spahn, 2015). The lack of transparency and accountability (Wachter, Mittelstadt & Floridi, 2017) enabled surveillance capitalism to flourish, putting democracy under attack (Christodoulou & Iordanou, 2021).

*“Surveillance capitalists now hold the answers to each question, though we never elected them to govern... They claim the authority to decide who knows by asserting ownership rights over our personal information and defend that authority with the power to control critical information systems and infrastructures.”* (Zuboff, 2021).

What can we do to deal with the challenges coming along with the rapid development of technology – AI and Big Data? The answer cannot be to move backwards, turning our back to emerging technologies. The answer might be to focus on promoting the development of *responsible* technology, that will be human-centered, not discriminating between races, cultural backgrounds, etc. and will promote individual and societal wellbeing, a topic that we turn to next.

## Responsible Innovation

European Commission, acknowledging the importance of responsible innovation, proposed the term RRI, to describe Responsible Research and Innovation, a science policy framework that aims to describe scientific and technological research in a way that takes into account potential impacts on society and the environment. The European Commission subsequently funded a number of projects to examine how RRI can be promoted – an example constitutes the project COMPASS, one of the two projects that the present chapter focuses on. RRI, is a multi-dimensional concept, involving *public engagement, gender equality, science education, open access, ethics, and governance* (Soraker, et. al., 2017). The RRI *Ethics* pillar is particularly interesting for this paper, especially in relation to the impact of AI and Big Data related technologies in industry and society. “The need for ethical considerations in the development of intelligent interactive systems is becoming one of the main influential areas of research in the last few years” (Dignum, 2018, p. 1).

Given that a considerable amount of research is conducted in the industry sector, any efforts to promote RRI which exclude industry are condemned to fail (Iordanou, 2019). RRI encourages an approach towards innovation where “societal actors and innovators become mutually responsive to each other with a view to the acceptability, sustainability and societal desirability of the innovation process” (Von Schomberg, 2012). However, as the effort to intertwine scientific excellence and society, through the implementation of responsible practices, has been increasing, so have the observed challenges.

At the core of European Commission's concept of RRI, a concept primarily tailored to be applicable within the European context, (from the Responsible Research and Innovation report (2012) published under the Science in Society initiative), is the inclusion of different stakeholders, working together, for finding solutions to the complex issues arising from emerging technologies and the successful implementation of RRI.

*RRI means that societal actors work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of European Society [...] an ambitious challenge for the creation of a Research and Innovation policy driven by the needs of society and engaging all societal actors via inclusive participatory approaches.*

In the present work, we reach out to different stakeholders – experts in AI technology – employing focus groups and individual interviews, to examine their views on the ethical dimensions of AI/Big Data and how to promote responsible research and innovation (RRI). The findings presented here are part of larger studies pursued in the projects COMPASS and SHERPA, both funded by European Commission.

The paper is organized as follows: The Methodology of the two studies is presented followed by the Results of both studies, identifying common themes between the two studies. The paper continues with a discussion of the common recommendations emerging from both data collection phases, focusing on the implications for education, which needs to focus more on promoting critical, responsible thinking, and on the aspects to be considered for successful AI regulation. Implications for promoting RRI and addressing the ethical issues of AI are also discussed.

## Methodological Approach

The paper is a comparative study between the data collected from two separate qualitative studies that took place under the umbrella of two different European projects, the EU Horizon 2020 COMPASS project (2016-2019) and the EU Horizon 2020 SHERPA project (2018-2021). The aim of both studies was to capture the opinions of technology experts and other stakeholders, on

the application of RRI, and in particular the ethical impact of developing and using emerging technologies, e.g. AI.

The COMPASS project was an EU-funded project that supported Small and Medium-sized Enterprises (SMEs) from emerging technology industries to manage their research, development and innovation activities in a responsible and inclusive manner in a number of ways including tools and services tailored to SME needs. SHERPA was an EU project that, in collaboration with a broad range of stakeholders, investigated, analyzed and synthesized the ways in which AI and Big Data analytics can impact ethics and human rights in society, and made recommendations to advocate the most desirable and sustainable solutions.

Overall, the COMPASS data collection focused on the responsible use of emerging technologies in specific sectors, i.e. healthcare and nanotechnology, and the SHERPA data collection focused on the specific technologies themselves, i.e. AI and Big Data.

## Participants

### Individual Interviews

Thirty key industry representatives, 18 from the ICT healthcare sector and 12 from the nanotechnology sector, participated in individual interviews in the context of the COMPASS project. Purposeful sampling, and in particular maximum variation sampling, aiming to capture a wide range of perspectives relevant to the specific study, was conducted to identify prospect interviewees. Specifically, all the partners in the project COMPASS were asked to propose prospect interviewees from their network, which were reviewed by the researchers and included in the sample, making sure that a diverse and representative sample from healthcare and nontechnology industry was secured. The participants were key industry representatives, mostly CEOs in their companies, from different countries across Europe, 10 were from the UK, 7 were from Austria, 5 were from Spain, 4 were from Cyprus, 2 were from Belgium, 1 participant was from Italy, 1 participant was from Slovenia and another one was from Switzerland.

## Focus Groups

Forty-nine individuals participated in six focus groups for the SHERPA data collection. Four focus groups were pursued on Guidelines for AI developers promoting Ethics-by-design, developed in the context of the project. Two of the focus groups on Guidelines were pursued in Cyprus with a range of stakeholders – AI designers, AI users in the industry and social researchers from NGOs –, and the other two focus groups were pursued in the UK with members of the British Computer Society. The remaining two focus groups were pursued on Regulatory Options, where the idea of an AI regulator was discussed with the stakeholders, mostly policy and legal experts, but also technology experts, again for AI and Big Data systems, one in Cyprus and the other virtually, due to the pandemic with different participants from a number of countries.

## Instruments

Both for the interviews and for the focus groups, interview protocols were developed by one of the authors for the COMPASS interviews and by partners in the SHERPA project. In COMPASS project, participants were asked to explain what they considered as the main barriers and challenges for adopting RRI in healthcare and nanotechnology, as well as what are their recommendations for successful implementation of RRI. In SHERPA, for the Guidelines focus groups, participants were asked to provide feedback on the Guidelines developed in the context of SHERPA for the consideration of ethics during development and use of AI and Big Data technologies. The proposed guidelines are available on the SHERPA project website (<https://www.project-sherpa.eu/guidelines/>). For the Regulatory options, the focus groups participants were asked to comment on potential requirements for a new AI regulator at EU level.

## Procedure

Thirty in-depth interviews were conducted by three researchers within the COMPASS project. The task of the interviewers was one of probing for further details or asking for clarification when necessary; the interviews proceeded as a conversation rather than a question-and answer session. Each interview lasted approximately 25-60 minutes and were conducted in person, via Skype or telephone. The procedure for the focus groups conducted within the SHERPA project was similar. The duration of the focus groups was on average between 60-90 minutes.

## Data Analysis

Although the projects were completed within ethically approved guidelines from the EU, the data collection phases received additional ethical approvals from the Cyprus National Bioethics Committee. The data obtained during the focus groups was analysed using thematic analysis (Braun & Clarke, 2006). Analysis of the transcripts was undertaken centrally, led by one of the authors, to ensure consistency. A stepped process of thematic coding was utilized. An inductive approach was used. The first stage of open coding was followed by a further stage of thematic coding during which emerging themes were compared and contrasted and gradually refined. After discussion between the authors, the themes were finalized.

## Results Overview

Given the particular needs for RRI in industry, and in particular the specific needs identified in the healthcare and nanotechnology sectors, the COMPASS results explored, through interviews, a deeper understanding of what key industry stakeholders from these sectors considered as the main challenges for RRI. The results also provided insight into their recommendations towards alleviating those challenges. The SHERPA results explored, through a number of focus groups with technology-related stakeholders, how to successfully move forward with the development and operational use of AI and Big Data in various sectors.

## Challenges Identified

According to results from COMPASS, companies within the nanotechnology sector are expected to promote and support a high level of innovation that results in positive societal impact, as it is a sector “impacting modern social life and economies” (Galatsis et. al., 2015). The sector has been transformed by information technology. This transformation was caused by the rapidly growing technological sector, e.g., introduction of new technologies, and has resulted in challenging social issues that need to be addressed, such as job losses, or gender equality and diversity in technical sectors. There has been an eminent need for responsible practices to address these social issues, in addition to many others that fall under the umbrella of responsibility, such as environmental issues, and new policies and regulation.

Furthermore, COMPASS identifies similar regulatory needs in the sector of healthcare. Technological innovation must be regulated within this sector, especially with regards to “the changing relationship between the private and public sector in the use of human genomics and personal medical information” (Martin & Hollin, 2014). The relationship is transforming into a collaborative one, offering a better foundation for responsible practices in the private sector that will be encouraged by the public sector. Martin and Hollin (2014) recognize that the healthcare sector is moving towards this direction since “throughout the 2000s a series of UK and EU public policy initiatives were taken to promote innovation and growth of the [...] commercial development of biotechnology in particular”.

In SHERPA, the issue of governance linked to responsibility, is also highlighted in the discussion that took place during the focus groups. In fact, it is suggested by participants that the context must be clarified in terms of who is responsible for the design of appropriate guidelines in order for information governance to be aligned to the business governance structures, since it is “the business that sets the ethical principles against which an ethical system must be designed” (FG-GUID-1). This does not refer to the business governing the ethics, but that the business should include ways within its structure to check the ethics.

The concept of information governance should be evident in any guidelines, clearly indicating the responsible parties, both in guidelines for users, including business management, but also guidelines for developers, such that “information governance is outside the IT department”. (FG-GUID-1)

*IT build the car, but the business needs to drive it and describe it. Yes, it's been a hard job to get a corporate organisation to accept that they drive privacy. [...] The whole thing needed to be moved into the corporate context. Once we started getting information governance, that's when we started succeeding in that. (FG-GUID-1)*

Another challenge identified particularly in the context of the COMPASS project are societal *values and attitudes*. Almost half of the interviewees asked, mentioned that culture's, including scientists', values and attitudes is one of the major barriers for adoption of RRI. At the same time, they identify reflection and reconsideration of the values that scientists and society prioritize, as a promising way for promoting responsible research and innovation, as discussed in the following



section. They mentioned that societies value fast money and there is no consideration of the consequences of the attempts for fast money. The following interviewee has described nicely this barrier:

*I think that the Vantage Points that are listed in philosophy and from philosophy to culture and from the cultural to the policy and from the policy to the strategy, from the strategy to tactics, from the tactics to logistics and from logistics to tasks. So, you don't go to the task directly if you don't have the philosophy and the culture before. I think that if your philosophy and your culture is to get as much money as you can, as fast as you can and the world will be fixed by God, then there is nothing to be done. I think it is the cultural thing before you start to innovate. (N1)*

Another interviewee noted that this attitude towards fast profit is not limited to industry, but has also affected universities, which depend to some extent on external agencies for funding.

*There is traditional business culture where research is not appreciated because you have to make money and all the business, culture around the university, investors, spin offs and start-ups don't want to invest into anything which is not fast profit. It is quite a low cost, a low quality business culture we have now in our world, in US, in Europe. That's the problem. They don't ask you do things well. They ask you to cheat and invent and make artificial things and get plenty of money from investors. The economical culture is quite low, happy, low cost, not very useful. (H1)*

“Those starting the social responsible innovation are already socially responsible scientists or stakeholders” another interviewee mentioned to highlight that social responsibility precedes social innovation, i.e., the innovation will follow where there is a culture of responsibility.

The attitudes towards change and innovation by small family business were also mentioned by some interviewees as a barrier to the adoption of RRI. As the following interviewee mentioned, even if the director of the business is not part of the family business, “culture is beyond everything”.

*Many of the SMEs we've got in XXX are very innovative because they come from academia so they have technology breakthroughs so they really want to share even open innovations so*

*they are not afraid. The bigger or the more traditional the company then it changes a lot. Because they are more afraid of risks, they are less willing to change things the way they've been doing it for years... the CEO is not from the family but at the end of the day the culture is beyond everything. So in my experience, generally speaking, these companies are more reluctant in trying new things and they have their part lines they know what they want to do and if something is interesting okay but they don't want to try fancy things as they say. (H1)*

In the project SHERPA, a project that explored how AI and Big Data analytics can impact ethics and human rights, participants identified the challenge for policy makers to take into consideration public values. The need for inclusive policy making, taking into consideration a wide range of citizens, and human rights was underlined in the FGs:

*An EU AI regulator could support national regulators in ensuring that the algorithms used safeguard human rights and other ethical republic values. (FG, REG-1)*

## Recommendations from Stakeholders

In the following sections, we juxtapose the recommendations highlighted in the discussions from COMPASS and SHERPA. The recommendations are organised by their core theme. In particular, data collected from the COMPASS interviews and the SHERPA focus groups seemed to follow two separate thematic axes, relating to the discussion of challenges and possible recommendations: education and regulation.

### Education

The majority of the COMPASS interviewees acknowledged the role of education in promoting RRI. As the interviewee below mentioned, in order to promote RRI we need to develop responsible scientists and citizens first.

*I don't think that RRI program will make people more responsible. To make people more responsible you have to go deeper into society. You need education. And this is not a problem*

*of innovation responsibility. It is the problem of society. Not responsible societies will not make responsible innovation! They will make stupid innovation. Stupid societies make stupid innovations, stupid decisions, and stupid referendums. So education should go first and once you have responsible societies you will get responsible innovations. (N1)*

Through education people need to reconsider and change values, especially values of fast profit.

*Other people like it bigger, stronger, faster and cheaper and they don't care about consequences. That's the problem. That's an education problem. We will use our overwhelming future benefits to make the education. The bigger, faster and stronger that was 2000 years ago in Roman times, we should get away from that. The atomic bomb and all that. We don't need it anymore. The philosophy of going faster, bigger and stronger and more powerful and cheaper and bigger and bigger and more energy more energy more energy. (N1)*

Educating citizens to be responsible, and informing them particularly about RRI, could substantially affect the application of RRI, because citizens will prefer products that have been developed following RRI (Thompson & Kumar, 2021) and will put social pressure on companies forcing them to adopt RRI. Here are some of the quotes provided by the interviewees:

*If this is a market which is usually driven by customer needs, if there is an awareness even at the level of customer ..., for example, components that are developed with an RRI behind the development, the companies will actually use this as a marketing tool.*

Some COMPASS interviewees proposed even to focus on scientists' education, while others suggested to include RRI in the mainstream education in order to educate all citizens. This links to the public opinion collected in SHERPA. For SHERPA, having guidelines for educational purposes, for developers and users of AI is discussed as “a way of really educating people and giving the information to people so it can really make sense to them” (FG-GUID-2).

When developing and using smart information systems, including AI and Big Data technologies, bias tends to become an issue that developers and users need to consider. Aiming for transparency can be a way to address bias. Allowing for algorithmic or implementation transparency increases the likelihood of identifying any biases in the software design and development. The focus groups participants highlighted the link between transparency and bias:

*'And it comes down to who's programming, how cognitive computing starts? I mean who is behind the AI? How did they train the AI to give advice? So, you need that background of the AI machine as well to see if it's biased or not.'* (FG-GUID-3)

Lack of necessary knowledge extends beyond lack of technical knowledge, but must also include a common understanding of ethics and their implementation. It can be argued that identifying what is ethical may vary between individuals. This is an even more challenging task at the company level:

*'everyone's definition of what is ethical or where to draw the line will differ. And as it was pointed out, companies do not have the incentive, and being or maybe not being that ethical, it makes more financial sense. And it's the same with governments as well. One country might have completely different definition of what they want to peruse for their gains as compared to the other countries.'* (FG-GUID-3)

Individual certification can also be a strong motivator for developers. It acts as evidence of relevant professional training and adds confidence in the developer's inter-disciplinary knowledge and skills:

*'For individuals, if there is a certification process or something, it's an interest for them to follow the training because somehow it can display or they can show that they have this and that, which maybe one day will appear in vacancies and say, well we would like this person to have instead of CCNA and whatever, a certification of ethical implementations.'* (FG-GUID-2)

The lack of education of what it means to responsibly develop technology, i.e. what to consider, and how to responsibly use technology is the basic need in moving forward, according to SHERPA focus groups participants:

*'At the basis of this it's education, especially with kids. Because kids grow up with Google and we've grown up with Google and the kids are [relying] on our personal assistance on phones and stuff like that, which is AI and big data. So, it depends on how they learn. You need to somehow give them the education and the logic to not always believe what they hear, or they see on smart phones, let's say, or through software.'* (FG-GUID-3)

And also:

*'Because, let's say we have now smart phones with our personal assistant on it. Everyone has the personal assistant. The personal assistant knows all our contacts, our emails, text messages, our social media activity. Each personal assistant perhaps in the future will be able to somehow tell you what to do at some point. And you will trust that. So, education should start from now on this. Like make people be careful when they start using personal assistants.'* (FG-GUID-2)

Education can also be a way to mitigate negative impacts of technology caused by misuse of technology. Threats to individuals, to the society and the environment will exist but educating citizens, especially the younger generation, in responsible use, can act as a counter-measure:

*I think the most important is education of the younger generation. ...I think the society should be active so that to introduce some more policies for controlling all technologies related to AI. But I think that as the society improves and progress, you cannot stop the evolution, the progress and from my point of view there is a threat and we need to see how we handle it. For me the most important threat is the misuse of AI, to use AI for damaging of society, of a person of a whole world. And we can see every day, we hear in the news that there was a robot that was cleaning the pollution in the ocean. Why not this robot go and make some huge damage in the ocean?* (FG-GUID-2)

## Regulation

When it comes to regulation of emerging technologies and particularly of AI, the opinion of SHERPA focus group participants is that education of the regulators themselves is crucial, i.e. there is a need for technologically- and ethically-educated politicians, since “in order for politicians to vote the correct, the right laws, they should have the education.” (FG-GUID-2) Moreover, and according to SHERPA results, a potential appointment of an EU regulator for AI would need to ensure to educate or train regulatory bodies at national levels so that they can apply the necessary guidelines, even where there is lack of expertise:

*I think one of the problems currently with many regulators is that they lack the knowledge and the expertise to cover a wide range of different algorithms and developments etc. So, I can imagine that an EU AI regulator could support national regulators in ensuring that the algorithms used safeguard human rights and other ethical republic values. (FG-REG-1)*

In COMPASS interviews, the majority of the interviewees discussed regulatory options. Specifically, 70%, suggested changes in the legislation in order to force the companies to use standards. They also mentioned that the European Commission and local governments should support RRI, either by having RRI as a criterion for funding EU projects or by offering direct support to companies to adopt RRI.

*You can make a law to make them share [...] I mean it's law or legislation that's the best way to do it and the second way is with money it can be direct money or it can be in tax with action or whatever at the end of the day for the company the money is the money coming in or less money going out. But for me [...] you either push them to do so, so you say that's the way it's going to be now that's the legislation so you have to do it so that's very effective but is hard for politicians to do because sometimes, the legislation, people are not really happy with it so they don't know if with new elections the legislation is still going to be there. But I think sometimes the government should be a little bit braver doing that. (H1)*

Some interviewees suggested particularly for Academia to set compliance to RRI standards as a criterion for receiving funding.

*I think that Academia needs another funding agency foundation [...] we can trust. These people could start asking for responsible innovation. You don't get funded if you don't have responsible innovation plan, you don't get the scholarship if you don't have responsible innovation plan. [...] that will force people to think about it and engage resources. (N1)*

Other interviewees, identifying that a substantial amount of research is conducted outside the academia, suggested to set RRI compliance as a criterion for funding industry projects as well.

*There is part of research going in industry, like in (SMEs), which is funded by public money and there should also be a push from the public funding towards RRI because today in our case, for example, we submit a project for research and innovation. We are not judged based on RRI.*

*[...] If you can get this as a policy of the European Commission and the local governments that it has to be. [...] So if you include this as one of the criteria... and it has specific weight and judgment, this will be motivating because, at least in Europe, many, many SMEs depend on the public funding for the big part of their research activities. (N2)*

Another interviewee identified that regulation should be applied globally; the reason for this is to allow for a higher impact, and thus the regulation would eventually be more efficient.

*It [regulation] should be looked at from a global perspective rather than a national perspective. (N10)*

Similar to the COMPASS results, the SHERPA findings from the focus groups are not much different in terms of recommendations for regulation. The focus groups participants discuss the significance of national participation in the regulation but further comment on requirements for a potential EU regulator that could consider important regulation aspects at an EU level (in addition to the national regulation) for each member state. Effectively applying the measures for responsible AI and Big Data development and operational use, requires that national authorities to action, however, the EU regulator must be able to coordinate such action:

*I think at a national level we have legislation, we have authorities that maintain laws and I suppose that they can be strengthened at an EU level and they have to take advantage and strengthen the national levels instead of passing by. (FG-REG-1)*

Nevertheless, the participants agree on the importance of having an EU regulator that can coordinate the different stakeholders, including professional differences, national differences, cultural differences, differences in the use of technology, etc.

*A European regulator is important and essential because, again, (the regulator) could be very focused on how this technology is going to impact on a broader level the European dynamics and then of course each government to decide for themselves if and how they want to implement that technology. I think the regulation should still be coming from a European level and then move on to a national level. (FG-REG-1)*

So having a regulatory body, having a policy or law, can be a strong motivator. Given that the primary step is to develop guidelines for the development and use of AI and other emerging technologies, “if the guidelines are legally binding or they have to be there, so if this were more of a policy or law, then people would go through this” (FG-GUID-2). The idea of using some type of legal enforcement, where the threat of punishment is more real, is a way to get things started. The idea of accountability is highlighted as the core motivator:

*It really helps when someone gets in trouble to be able to go to a board. Because they are looking at it as a business risk. They're not looking at it as a moral, ethical issue in the end. They're saying, "What would we lose in terms of market penetration? What would we lose in terms of customers compared to the low probability of a fine or publicity?" (FG-GUID-4)*

## Discussion

The present study aimed to examine different stakeholders' views on the challenges of applying responsible research and innovation (RRI) and in particular ethics in the development and operational use of emerging technologies, focusing on AI. Results identified some barriers and challenges for the adoption of RRI. A prominent challenge identified by key industry stakeholders are *societal values* which place emphasis on fast profit without considering the consequences of one's actions, e.g., transnational racial capitalism. To address those challenges participants emphasized the role of education towards individualized responsibility, for developing the sense individuals' agency – that they are in control of their actions – for developing responsible researchers, business leaders, consumers and citizens. Recommendations also involved changes in the regulation of these technologies and suggestions about the context of any guidelines of how to move forward, i.e. that guidelines must be clarified in terms of who is responsible for the design of these guidelines.

Interviewees viewed the values of innovators, researchers, and citizens, in general, as a major barrier for responsible research and innovation. The interviewees identified this as a barrier both for innovators, scientists, business leaders who pursue research and develop products but also for citizens, who are the consumers of those products. Research malpractice, driven by individualist values, can have detrimental consequences on societies, where public money is wasted without



providing solutions to problems, such as health problems (see for example Kolata, 2018). On the other hand, citizens' individualistic values also have consequences on responsible research and practice. As one of the interviewees noted, "When you use your money, you are making a political act; you are giving the money to someone, or someone else. You are ... closing companies". Other interviewees mentioned that citizens can put pressure on companies to be aligned with responsible research and innovation through social media. Interviewees' views on the role of individuals' values on the adoption of RRI are in line with research values, which has provided evidence of a relation between individuals' values and their behaviour (Miles, 2015). For example, the value of collectivism predicts environmental behaviour, particularly green purchase (Kim, 2011).

Differences in values and practices between different countries affect competitiveness. Interviewees expressed the concern that adherence to RRI may involve extra cost which will result in more expensive products which will be less competitive compared to products who have been constructed by companies who do not adopt RRI. This concern confirms previous concerns about competitiveness that have been reported in the literature as a barrier for the adoption of RRI (Martinuzzi et. al., 2018). A possible solution to this problem has been provided by the interviewees themselves, who suggested changes in the *legislation* system in order to enforce all the companies, worldwide, to follow the same standards, not specifying how these standards will be agreed upon, but being cautious not to have these standards be dictated. Participants in the focus groups have also extensively discussed regulatory options, and in particular, the presence of an EU regulator for AI as a solution. In this way, companies' concerns that compliance with RRI may result in more expensive and therefore less competitive products, compared with the products of companies which are not compliant with RRI, will be alleviated.

The need of regulation for emerging technologies, to have a more transparent, accountable and inclusive policy making towards emerging technologies, was one of the major suggestions of focus groups participants, as well. The regulation of such technologies, implies the need for ethics-aligned technologies, aligned with La Fors, Custers, and Keymolen's (2019) proposal for emerging technologies to take into consideration the following values: human welfare (the fair treatment of citizens in various contexts, e.g., employment, schooling, travelling), autonomy (awareness, free will and free choice), non-maleficence (transparency in order to avoid harm), justice (fair, preventative law enforcement practices), accountability (know who is responsible

for data sharing), trustworthiness (for technology operators and big-data based conclusions), privacy (safeguard from easy identifiability), dignity (avoid discrimination and stigmatization), solidarity (avoid the prioritization of commercial interests and root for mutual support) and environmental welfare (safeguard against direct and indirect threats to the environment). Furthermore, our findings show, based on a variety of stakeholders' views, that development should indeed consider the integration of values, and that any regulation should take this into consideration. Dignum (2018) proposed that algorithms need "to integrate societal, legal and moral values into technological developments in AI, at all stages of development (analysis, design, construction, deployment and evaluation)" (p. 1). Dignum also highlighted the need for making AI reasoning more inclusive, to "weigh the respective priorities of values held by different stakeholders in various multicultural contexts; explain its reasoning; and guarantee transparency" (2018, p. 1).

The other major recommendation offered by all participants is to promote RRI and address ethical considerations of AI and Big Data through *education*. Some interviewees mentioned that education towards RRI should be embedded in all levels of the educational system, starting from the elementary school. Participants highlighted particularly the role of higher education in promoting RRI and the ethical dimensions of AI, which is consistent with recommendations in the literature (Zembylas, 2021). The interviewees acknowledged that there is an imperative need to support the development of individuals' critical thinking skills, particularly their ability to take multiple perspectives into account, not focusing only on the economic aspect or personal benefit, and place greater consideration on the possible long-term, collective consequences for society, of their decisions and actions. Interviewees expressed the expectation for education to undertake this important task, viewing education as the nursery of future scientists, business leaders, policy makers, consumers and citizens, whose decisions and behavior will determine whether research and innovation is pursued in a more responsible way, respecting one another and our environment. If we wish future business leaders to adopt responsible research and innovation procedures in their business, future politicians and policy makers to make more responsible decisions and promote legislation supporting RRI, future teachers to teach their students starting from pre-K to university, to think critically and responsibly, future journalists to be cautious towards not replicating fake scientific news, and citizens to support responsible research and innovation through their actions

as voters and consumers, we need to pay closer attention on efforts promoting RRI throughout the education system and lifelong learning. The concerns and recommendations of stakeholders are in line with other voices in academia calling for a reform in education and the need to place greater emphasis on promoting critical thinking skills, including consideration of the ethical dimensions of AI (Kuhn, 2005; Zembylas, 2021). Some intervention programs appear promising in supporting the development of students' critical thinking skills at different levels, from primary to higher education (Iordanou & Kuhn, 2020; Iordanou & Rapanta, 2021), including greater reflective thinking (Iordanou, 2022a) and consideration of the ethical dimensions of an issue (Iordanou, 2022b). Future research should examine the effectiveness of those interventions, based on discussion and reflection (Iordanou & Rapanta, 2021), on supporting particularly deeper thinking on ethical issues of AI and Big Data and more responsible behaviour. The objective of promoting more collective values, promoting human and environmental wellbeing, is more challenging, yet an important one that future research should address. Furthermore, more exploratory studies, such as this one are needed, involving a larger and more diverse sample, involving participants from different domains not covered in this study.

Finally, the present study shows that launching an open dialog with multiple stakeholders, as the studies discussed in the present chapter employed, to understand different perspectives on how to promote RRI and deal with the ethical challenges of AI, is a promising endeavor to deal with this complex issue. Addressing the ethical challenges of emerging technologies is a complex issue, involving multiple stakeholders and we should address it as such. Engaging different stakeholders in the dialog on how to address the ethical challenges of AI, could increase mutual understanding, commitment and willingness to work all together for designing a more human-centered AI, promoting individual and societal wellbeing.

## References

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

Broussard M (2018) Artificial Unintelligence: How Computers Misunderstand the World. Cambridge: MIT Press.

Bucher, T. (2012) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook, *New Media and Society*, 14(7), 1164–1180.

Cave, S., 2019, “To Save Us from a Kafkaesque Future, We Must Democratise AI”, *The Guardian*, 04 January 2019.

Cave, S., & Dihal, K. (2020). The Whiteness of AI. *Philos. Technol.* 33, 685–703.  
<https://doi.org/10.1007/s13347-020-00415-6>

Costa, E., & Halpern, D. (2019). The behavioural science of online harm and manipulation, and what to do about it. The Behavioural Insights Team.

Gillespie, T. (2018) Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, New Haven, CT: Yale University Press.

Christodoulou, E., & Iordanou, K. (2021). Democracy Under Attack: Challenges of Addressing Ethical Issues of AI and Big Data for More Democratic Digital Media and Societies. *Frontiers in Political Science*, 71.

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue, *Ethics and Information Technology*, 20, 1-3.

European Commission (2012). Responsible Research and Innovation: Europe’s ability to respond to societal challenges, Research and Innovation Policy. DOI: 10.2777/11739.

- Galatsis, et. al. (2015). Nanoelectronics Research Gaps and Recommendations, IEEE Society and Technology Magazine, 21-30.
- Helbing, D., Caron, & Helbing. (2019). Towards digital enlightenment (pp. 73-98). Cham, Switzerland: Springer International Publishing.
- Iordanou, K. (2019). Involving patients in research? Responsible research and innovation in small- and medium-sized European health care enterprises. *Cambridge Quarterly of Healthcare Ethics*. 28 (1), 144 – 52. <https://doi.org/10.1017/S0963180118000488>.
- Iordanou, K. (2022a). Supporting strategic and meta-strategic development of argument skill: the role of reflection. *Metacognition and Learning*. Doi:10.1007/s11409-021-09289-1
- Iordanou, K., (2022b). Supporting critical thinking through engagement in dialogic argumentation: Taking multiple considerations into account when reasoning about genetically modified food. In M, Jimenez-Aleixandre, B., Puig Mauriz. *Critical Thinking in Biology & Environmental Education*. Springer.
- Iordanou, K., Kendeou, P., & Zembylas, M. (2020). Examining my-side bias during and after reading controversial historical accounts. *Metacognition and Learning*, 15(3), 319-342. <https://doi.org/10.1007/s11409-020-09240-w>
- Iordanou, K., & Kuhn, D. (2020). Contemplating the opposition: Does a personal touch matter?. *Discourse processes*, 57(4), 343-359.
- Iordanou, K., & Rapanta, C. (2021). " Argue with Me": A Method for Developing Argument Skills. *Frontiers in Psychology*, 12, 359.
- Kim, Y. (2011). Understanding green purchase: The influence of collectivism, personal values and environmental attitudes, and the moderating effect of perceived consumer effectiveness. *Seoul Journal of Business*, 17(1) 65-92.
- Kuhn, D. (2005). *Education for thinking*. Harvard University Press.

- La Fors, K., Custers, B., & Keymolen, E. (2019). Reassessing values for emerging big data technologies: Integrating design-based and application-based approaches. *Ethics and Information Technology*, 21(3), 209-226.
- Loader, B.D., Vromen, A., and Xenos, M.A. (2016) Performing for the young networked citizen? Celebrity politics, social networking and the political engagement of young people, *Media, Culture and Society*, 38(3), 400–419.
- Martin, P., & Hollin, G. (2014). A new model of innovation in biomedicine? [http://nuffieldbioethics.org/wp-content/uploads/A-New-Model-of-Innovation\\_web.pdf](http://nuffieldbioethics.org/wp-content/uploads/A-New-Model-of-Innovation_web.pdf).
- Martinuzzi, A., Blok, V., Brem, A., Stahl, B., & Schönherr, N. (2018). Responsible research and innovation in industry—Challenges, insights and perspectives. *Sustainability*, 10(3), 702.
- Kolata, G. (2018, Oct. 29). He promised to restore damaged hearts. harvard says his lab fabricated research. *New York Times*.
- Miles, A. (2015). The (re) genesis of values: Examining the importance of values for action. *American Sociological Review*, 80(4), 680-704.
- Noble S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.
- Otterbacher, J. (2018, September). Addressing social bias in information retrieval. In *International Conference of the Cross-Language Evaluation Forum for European Languages* (pp. 121-127). Springer, Cham.
- Pols, A. J. K., & Spahn, A. (2015). Design for the values of democracy and justice. In *Handbook of Ethics, Values and Technology Design* (pp. 335-363). Springer.
- Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 116(16), 7662-7669.

Soraker, J. H., et. al. (2017). Models of RRI in Industry, Responsible Industry Deliverable D3.3, 7th Framework Programme.

Thompson C. J, & Kumar A. (2021). Beyond consumer responsabilization: Slow Food's actually existing neoliberalism. *Journal of Consumer Culture*. 2021;21(2):317-336.

doi:[10.1177/1469540518818632](https://doi.org/10.1177/1469540518818632)

Von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. In *Technikfolgen abschätzen lehren* (pp. 39-61): Springer.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6), ean6080.

Zembylas, M. (2021). A decolonial approach to AI in higher education teaching and learning: strategies for undoing the ethics of digital neocolonialism. *Learning, Media and Technology*, 1-13.

Zuboff, S. (January 29, 2021). "The Coup We Are Not Talking About". *The New York Times*. Retrieved January 22, 2022.