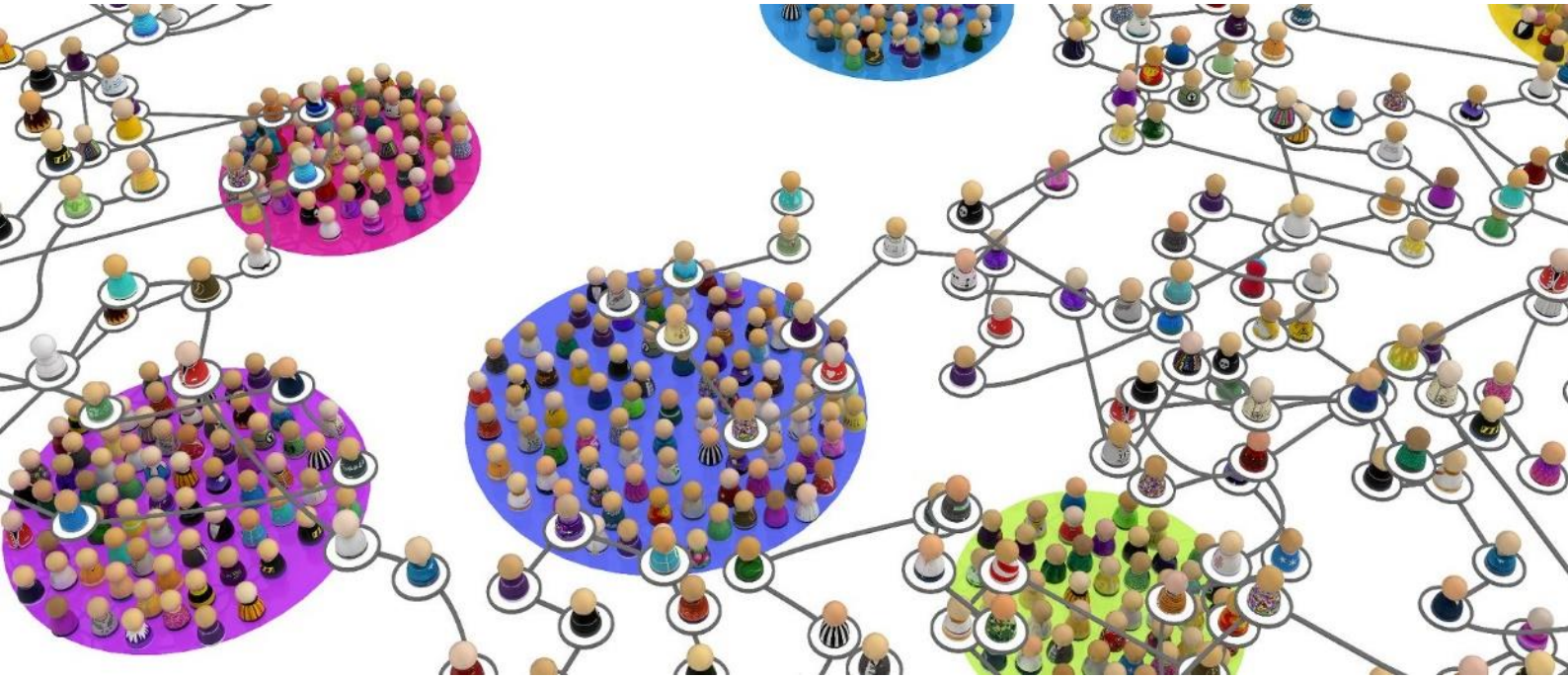
SHERPA

Evaluation Report

Kalypso Iordanou, Eleni Christodoulou and Josephina Antoniou



Deliverable 4.2. - 26 August 2020

This project has received funding from the
European Union's Horizon 2020 Research and Innovation Programme
Under Grant Agreement no. 786641

Document Control

Deliverable	D 4.2. Evaluation Report
WP/Task Related	WP 4
Delivery Date	Month 28
Dissemination Level	PU
Lead Partner	UCLanCY
Contributors	
Reviewers	
Abstract	
Key Words	Ethics; Human Rights; Big Data; AI; focus groups; thematic analysis;

Revision History

Version	Date	Author(s)	Reviewer(s)	Notes
0.1	30 July 2020	Kalypso Iordanou and Eleni Christodoulou	Doris Schroeder	First Draft
0.2	14 August 2020	Kalypso Iordanou and Eleni Christodoulou		Second Draft
0.3	16 August 2020	Josephina Antoniou	Doris Schroeder	Third draft
0.4	17 August 2020	Kalypso Iordanou, Eleni Christodoulou and Josephina Antoniou	Bernd Stahl	Fourth draft
0.5	26 August 2020	Kalypso Iordanou, Eleni Christodoulou and Josephina Antoniou		Fifth draft

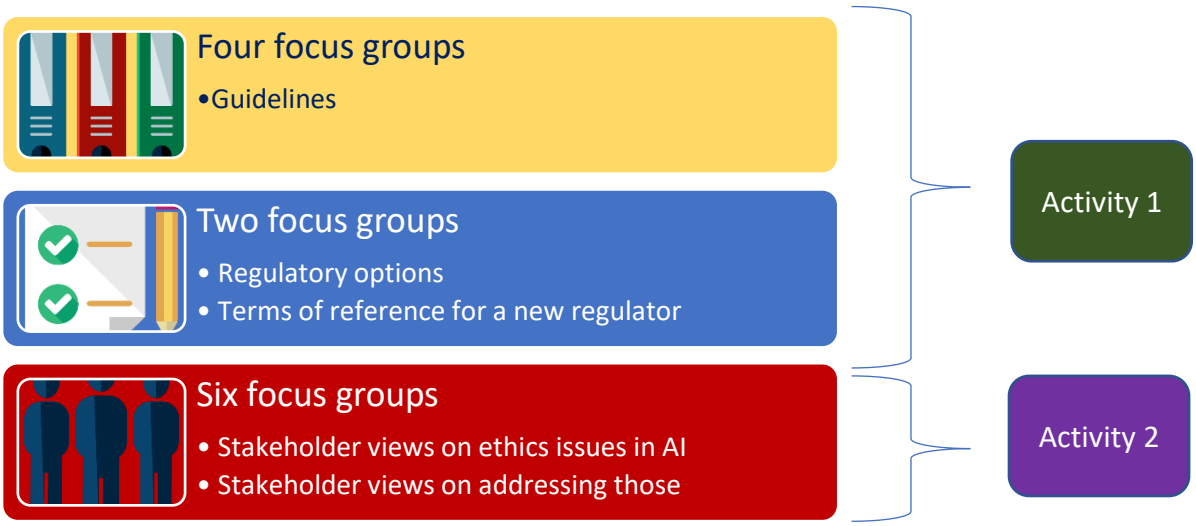
This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Contents

Executive Summary

Twelve Focus Groups with different stakeholders across Europe were pursued in the context of SHERPA's Evaluation and Validation Strategy (T4.1). The aim of the focus groups was twofold. Firstly, to gain stakeholders' views regarding the recommendations that have been developed in the SHERPA project, particularly regarding the set of Guidelines for user and developers (T3.2), the Regulatory Options (T3.3.) and the terms of reference for new regulator (T3.6). Secondly, to obtain an in-depth understanding of stakeholders' views regarding what they consider as the main ethical issues that come out of Artificial Intelligence (AI) and big data, on the way those ethical issues are currently addressed and their suggestions on how those ethical issues can be addressed for efficiently in the future. The distribution of focus groups was as follows:

Figure 1. Activities involved in Task 4.2.



The findings of the FGs on stakeholders' views about SHERPA products, will provide internal feedback to the project and can serve as a means for corrective action if necessary. The findings of the Exploratory FGs will contribute towards SHERPA's objective for development of a set of recommendations for the responsible development of SIS.

Thematic analysis was used to analyse all the FGs. In terms of discussing the proposed Development and Use Guidelines, the Guidelines FGs highlighted a number of positive aspects, but simultaneously identified some issues and challenges to consider for successfully implementing the proposed Guidelines. Regarding the Regulatory FGs, the analysis highlighted specific topics such as what needs to be regulated and how such regulation should take place, concluding with a number of proposed areas that a potential regulator should focus on.

Exploratory FGs revealed that the main ethical issues identified by stakeholders in relation to big data and AI involve loss of: autonomy, privacy, trust, accountability; manipulation of users; lack of transparency and adequate information for users, algorithmic bias and threats to human rights. To address those ethical issues, education and public engagement were identified as key factors that we should invest more in.

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Explanation
AHR	Aequitas Human Rights NGO
AI	Artificial Intelligence
BCS	British Computer Society
DMU	De Montfort University
ETHICOMP	International Conference on the Ethical and Social issues in Information and Communication Technologies
Eurec	European Network of Research Ethics Committees
FG	Focus Group
GDPR	General Data Protection Regulation

IEEE	The Institute of Electrical and Electronics Engineers
NEN	The Royal Netherlands Standardization Institute
QRCA	Quality Requirements Conformity Assessment
SB	Stakeholder Board
SIS	Smart Information Systems
UCLanCY	University of Central Lancashire - Cyprus
UKAIS	UK Academy for Information Systems’

1. Introduction

1.1. Background and objectives

The aim of the focus groups is to gain stakeholders' views and suggestions regarding the recommendations of the SHERPA project, in order to contribute to the development of best practices regarding the design and use of Smart Information Systems (SIS) for the benefit of stakeholders (e.g. industry, policy, funding bodies, research, civil society and the public). In addition, focus groups aim to obtain stakeholders' views on novel proposals for the responsible development and use of SIS.

To achieve these objectives, ten focus groups (FG) were planned, with 10 members of the main stakeholder groups participating in each group. In the Consortium meeting in Cyprus in October 2019, it was decided that the FGs should focus on three thematic areas:

a) Guidelines (Short version of two sets of Guidelines for users and developers - D3.2)

b) Regulatory Options (T3.3) & Terms of reference for new regulator (T3.6)

c) Exploratory FG on Ethical issues on AI/Big data

The first two types of FGs aimed to evaluate the major products and recommendations developed in the context of the SHERPA project, namely the development of guidelines for users and developers and providing recommendations for regulatory options and the terms of reference of a potential new regulator. The third type of FGs aimed to explore stakeholders' views on ethical issues in relation to AI/Big Data, using open-ended questions, thereby contributing to the formation of recommendations.

Further, it was decided in the October 2019 consortium meeting that two of the FGs, one focusing on Guidelines and another on Regulatory Options, would involve two sets of FGs with the same individuals. The first set would explore the overall set of recommendations and initial reactions from different groups of stakeholders. Then participants would be invited to experiment with the recommendations in their social setting and collect broader feedback from their colleagues in their organisations. In the second set of FGs, the same participants who took part in the 1st set of FGs would be invited to provide feedback received from their colleagues and their own experimentation with the Guidelines or Regulatory Options recommendations. This would allow for the development of a set of targeted recommendations which will be disseminated, communicated and put forward for implementation by the main stakeholder groups.

The precise content of the FGs was decided after receiving feedback from Task Leaders of other tasks in the project, particularly from T3.2. Guidelines for research and innovation of SIS, T3.6. Regulatory Options, T3.3. TOR for regulator, T3.4. Assessment of standardization potential and T3.4. Prioritization and finalization of recommendations. The results of the FGs analysis will feed back to those tasks.

The report is structured as follows. We begin by providing the methodology, including information about participants and the data collection process, before moving on to the findings of the data analysis. The findings section consists of three main parts. The first part presents the results of the Exploratory FGs, the second part presents the results of the Guidelines FGs while the third part presents the results of the Regulatory options/New Regulator.

2. Methodology

2.1. Ethics Approval and Data Management

The Task Leader applied and secured ethical clearance from the Cyprus National Bioethics Committee (see Appendix A). Each partner who organized a focus group made sure that all the required measures were taken in order to be compliant with the ethics guidelines for conducting research in the country where the focus group took place. Prior to the FGs the interviewees' written consent was secured, using the information sheet (Appendix B), adapting it first to the aims of each FG, and the consent form (Appendix C) that had been prepared by the task leader. The transcripts of the interviews were anonymous. Each partner uploaded the transcripts and the consent forms for the interviews pursued, in the project's drive, where only SHERPA participants have access.

2.2. Data Collection

2.2.1. How many FGs were conducted and by whom?

The distribution of FGs among partners was decided in the Consortium Meeting in Cyprus in October 2019, based on partners' person months on this task and their interest and expertise. Table 1 presents the distribution of FGs among partners. A training workshop was held by the task leader on October 11th 2019 at UCLan Cyprus, in the context of the SHERPA consortium meeting. The aim of the workshop was to train the FG coordinators to ensure consistency among the focus group sessions.

Overall 12 FGs were pursued between January 15th and June 26th 2020. In some cases, when we had more interest than we could accommodate in one group, we pursued 2 FGs. As a result, we have 12 FGs, instead of the ten from our initial plan.

Guidelines FGs

Four FGs were pursued on Guidelines. UCLan Cyprus organized two sets of FGs on Guidelines. The participants who attended the 1st set were invited back in the second set. Given the limited response rate (n = 2), more participants were invited to participate in the second FG. A substantial part of the 2nd set of FG was devoted to exploratory questions, because the two participants who attended the 1st set of FG didn't have much feedback to report regarding the Guidelines. In order not to lose any material, the information from the 2nd set of FGs was coded both under exploratory and guidelines analysis (we refer to this particular FG as "hybrid"). DMU pursued two FGs on Guidelines, with different individuals participating in each FG. The participants in the FGs organized by the DMU were members of the British Computer Society (BCS).

Regulatory Options/TOR new Regulator FGs

Two FGs were pursued on Regulatory Options and TOR of new Regulator by AHR. All the participants from the 1st FG were invited to take part in the 2nd FG. As no participant was able or willing to return, others were recruited from the Stakeholder member list and attended the 2nd FG.

Exploratory FGs

Six FGs were pursued focusing on exploring the ethical issues of AI and Big data. DMU organized three exploratory FGs, while UCLan Cyprus, EUREC and NEN organized one exploratory FG each.

2.2.2. Who participated?

114 individuals participated in the FGs. About half of the participants, 46%, were women (based on the available information). Table 1 shows the number of participants per FG.

Table 1. Information about Focus Groups

	Organizer	Place	Date	Number of Participants
		Guidelines		
1	UCLanCY	Cyprus	Jan-16	8
2	UCLanCY (hybrid*)	Cyprus	Mar-05	4
3	DMU	UK (BCS)	Mar-05	13
4	DMU	UK (BCS)	Mar-05	13
		Regulatory Options		
5	AHR	Cyprus	Feb-24	6
6	AHR	Online	Jun-26	5
		Exploratory		
7	NEN	Netherlands	Jan-15	10
8	EUREC	Online (SBM)	Mar-23	19
9	UCLanCY	Online	Apr-09	9
10	DMU	Online (UKAIS)	Apr-09	12
11	DMU	Online (ETHICOMP)	May-05	7
12	DMU	Online (ETHICOMP)	May-05	8
	Overall			114

*In this FG Guidelines and Exploratory questions were discussed, therefore we refer to this FG as hybrid.

2.2.3. How were participants recruited?

The task leader suggested the inclusion criteria, namely diversity in stakeholder groups, gender balance and ethnic diversity. The participants should also be professionally involved in some aspect of relevance to the FGs. The consortium partners discussed with the task leader their plans for recruiting participants to make sure that the recruitment criteria were met.

Seven FGs took place in the context of relevant conferences or experts' meetings. Participants were recruited based on their intention to participate in the relevant conferences. When the conferences or meetings took place virtually or were cancelled, due to the COVID-19 pandemic, the FGs took place virtually.

- One Exploratory FG organized by NEN took place as a special session after the Standardisation Group meeting which discussed the content of a possible standard on quality and reliability of health and wellness apps. The FG was attended by different stakeholders: patient representatives, healthcare providers, data scientists, and app developers.
- Another Exploratory FG organized by EUREC, took place in the context of SHERPA's Stakeholder Board meeting in April 2020.
- A third Exploratory FG, organized by DMU, took place in the context of the UK Academy for Information Systems' (UKAIS) conference, which is attended by researchers and practitioners of information systems in the UK.
- Two exploratory FGs took place in the context of ETHICOMP 2020 - International Conference on the Ethical and Social issues in Information and Communication Technologies - organized by the Centre for Computing and Social Responsibility.
- Finally, two FGs on Guidelines, organized by DMU, took place in the context of the British Computing Society's annual meeting in London.

For the other 5 FGs, the organizers recruited participants from the Stakeholder member list or by reaching out to prospect participants from their contact list, based on their expertise. In addition, experts with a highly relevant profile were recruited after searching on the web.

2.2.4. How were FGs conducted?

Half of the FGs (6), were conducted face-to-face in three different countries (the UK, the Netherlands and Cyprus), while the other half were pursued virtually given the restrictions imposed by the Covid-19 pandemic. In some cases the FGs took place in participants' native languages. Therefore, 1 FG was pursued in Dutch (organized by NEN), another one was pursued in Greek (organized by UCLan CY) and the rest were pursued in English. The duration of the FGs was on average between 60-90 minutes. The FGs were recorded and transcribed. In the case of the Dutch FG (NEN), there was a summary of the discussion offered in English, whilst the one conducted in Greek was transcribed in full and the relevant parts used in the report were translated by the authors of this report.

2.2.5. What was the content of the questions?

For the Guidelines FGs the Task Leader of Task 3.2. (UT - "Develop guidelines for research and innovation in and with SIS") provided the FG questions. There were two sets of questions focusing on the set of guidelines for developers or the set of guidelines of users. Appendix D presents the list of questions that were used for the Guidelines FGs.

For the Regulatory Options/TOR of new Regulator, the Task Leader of T3.3. (TRI - "Explore regulatory options") and T3.6 ("Propose terms of reference for a new regulator for SIS") provided the questions. Appendix E shows the questions on Regulatory options and Appendix F includes the questions for terms of references of new regulator. Both sets of questions were used in each of the two FGs which focused on Regulatory options and TOR of new Regulator.

For the Exploratory FGs, the T4.2. ("Stakeholder evaluation and validation") Task Leader developed the questions, in collaboration with the Co-ordinator of the project, Prof. Bernd Stahl. Although there was some degree of flexibility in the FG discussions, the majority of FGs followed the same questions. These revolved around 3 core aspects:

- the main ethical issues that come out of AI and big data and their relation to human rights;
- the nature and limitations of current efforts to address these ethical issues;
- and suggestions of activities that should be undertaken in the future to deal with ethical issues that have not been yet adequately addressed.

We combined broad and exploratory questions with asking for number ranges (e.g. identify 3-5 ethical issues) or a specific number (e.g. suggest three activities) so as to have manageable discussions and data. The questions posed to the participants did not offer specific options but rather were open-ended in order for the views of the participants to emerge rather than for the facilitator of the FGs to impose their own views or influence the discussion. Table 2 presents the questions that were used in the Exploratory FGs.

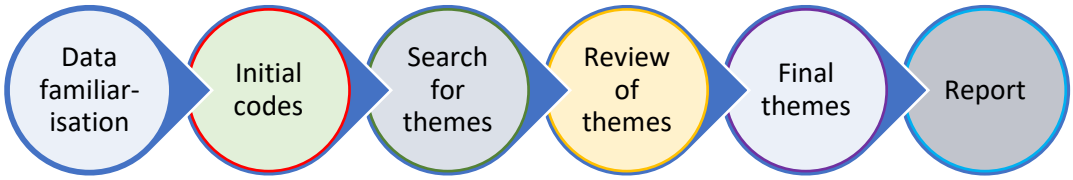
Table 2. Questions for Exploratory focus group discussions

1. What are the (3-5 main) ethical issues that come out of AI and big data?
2. How do those ethics issues relate with Human Rights?
3. How are those ethical issues currently addressed?
4. What are the limitations of the current efforts addressing those ethical issues?
5. What ethical issues haven't been addressed so far?
6. What are the three most important activities that should be undertaken to deal with the ethical issues that haven't been adequately addressed yet?

2.3. Data Analysis

The data obtained during the FGs was analysed using thematic analysis and in particular the framework provided by Braun and Clarke (2006). We followed the six stages of thematic analysis (2006, p.87): (1) Initial data familiarisation; (2) Generation of initial codes; (3) Search for themes; (4) Review of themes in relation to coded extracts; (5) Definition and final naming of themes; (6) Production of the report (see Figure 2).

Figure 2. Six Stages of Thematic Analysis



Thematic analysis can be defined as: 'a method for identifying, analysing and reporting patterns (themes) within data. It minimally organizes and describes [the] data set in (rich) detail' (2006, p.79). Although thematic analysis always involves a degree of interpretation (Boyatzis, 1998), due to the main objective of the FGs, which was to let the participants' opinions and expertise inform the findings, we took a data driven, inductive approach in which codes and themes were generated from the data (open coding) rather than having a codebook prepared in advance of the data analysis (deductive approach).

Given the large number of FGs analysed and to ensure better organisation, classification and interpretation of the data, the analysis was performed using Nvivo, a qualitative data analysis computer software (Version 12). One researcher was involved in the data analysis for stages 1-3 in order to ensure consistency across coding of the data. After this the researcher shared the resulting codebook with the task leader and discussions and brainstorming led to completion of stages 4-6 of thematic analysis (see Figure 2) in close collaboration with each other.

In total, for the 7 Exploratory Focus Groups 460 codes emerged (of different hierarchies, 'parent', 'child', 'grandchild' etc.). The coding of text segments was not mutually exclusive, for instance one paragraph could include up to 7-8 different codes. In a handful of instances, there were overlapping 'children codes' which could be used in two different ways, so these sub-codes were copied so their information could be potentially used in future theme creation in either topic. For the coding process, we followed closely the instructions and advice of Braun and Clarke (2006); Bryman (2008) and Charmaz (2004), namely:

a) Code as thoroughly as possible as 'you never know what might be interesting later' (2006, p.90). This point is also suggested by Bryman (2008) as well as Charmaz (2004) who suggest a line by line coding and for researchers not to be alarmed by the proliferation of codes; this is important so as not to lose any detail or potential interpretation of the data.
b) Code the data extracts in an inclusive way (with the surrounding text) so 'that the context is not lost'.
c) The same text can be coded for several different codes (and then themes).
d) Keep in mind that contradictions, inconsistencies and tensions are an inevitable part of the data and the researcher should not feel that they should smoothen these out or ignore them.
e) Retain accounts that depart from the dominant themes (e.g. in one FG when one participant said they do think education is important but should not be a priority).

In our report, the aim is to be as representative as possible of the diverse opinions expressed whilst also keeping in mind that we want our outputs to be accessible and useful to policy-makers and maximise impact.

We took a strong qualitative focus rather than a quantitative one: we identified patterns of meaning expressed and included indications when this occurred often in discussions (prevalent) whilst also including useful inputs even if they were only mentioned once. For example, 'companies driven more by profit and practical issues than by ethics' was a prevalent code across the FGs, but a suggestion to raise public awareness of ethical issues through story-telling using soap operas was suggested by a participant only once. Nevertheless, both pieces of data were included regardless of the difference in numerical terms. In line with qualitative analysis, the focus was on answering the 'what, how and why' questions, the kinds of issues raised, what these mean for improving our understanding of ethics and human rights related to SIS, and the insights and experiences offered by experts and stakeholders, rather than how many times issues were mentioned.

3. Analysis of Findings: Guidelines FGs

The Focus Groups that focused on discussing the proposed guidelines for the development and use of technology, featured a variety of stakeholders from different sectors, such as academia, small to medium software development companies, larger organisations e.g. from the banking sector, etc. All the stakeholders had expertise with technology research initiatives and projects, not only as software developers, but also as designers, project managers and legal consultants, and non-technology researchers (e.g. social researchers) of innovative technological projects.

The proposed guidelines are available on the SHERPA project website (https://www.project-sherpa.eu/guidelines/). The guidelines are published in two flavours, (i) the development guidelines, which deal with how developers can construct an ethical AI or Big Data system, and (ii) the use guidelines, which deal with how to ethically use an AI or Big Data system, especially adapted to governane and management of organisations that use these technologies as part of their services.

There was a consensus among the stakeholders that the Guidelines, both for SIS Use and SIS Development, **successfully and comprehensively addressed relevant human rights** issues:

> 'I actually found it amazing. For me, maybe because this is what I do, I was going through the rights in my head … and everything was addressed. […] Everything was interlinked, everything on the rights for me was addressed in a very good way and on many levels.' (UCLan Cyprus, 1st Wave)

Positive commentary further addressed the comprehensiveness of the "Guidelines" documents as **compared to other similar documents** that the participants had experience with in the past:

> ‘So compared to some other guidelines, I think it was much less general, more step by step and divided into the different stages of data processing etc.’ (UCLan Cyprus, 1st wave)

> ‘I think it’s very comprehensive and I especially like the fact point two - merging databases aspect. I think this for me this is the most interesting thing, because it’s something that I do in my work. […] But I think this is very comprehensive and I think you have covered every single thing here.’ (UCLan Cyprus, 1st wave)

The **clarity** of the documents was also highlighted by participants, expressing how they felt comfortable with the content of the documents even when **outside of their area of expertise**:

> ‘For me that I have absolutely no experience of technical stuff, no understanding, […], I understood everything very clearly and it was amazing. So, I think even for people who have no technical knowledge, there is a way of getting information.’ (UCLan Cyprus, 1st wave)

It was noted that there is a **gap** in terms of guidelines for development since, often, such guidelines are not dealt with by developers but by project management:

> ‘I think developers quite often whinge about getting the standards and guidelines that are at an abstract level, they don’t need to care about because it would have been dealt with at management levels.’ (DMU, BCS 2nd wave)

In fact, it is suggested by participants that the context must be clarified in terms of **who is responsible** for the design of these guidelines in order for information governance to be aligned with business governance, such that ‘Information governance is outside the IT department’ (DMU, BCS, 2nd wave):

> ‘And it feels to me that, at the beginning of those guidelines, there needs to be a context where it’s clear that it’s the business that sets the ethical principles against which an ethical system must be designed.’ (DMU, BCS 2nd wave)

Examples of successfully driving the implementation of privacy already, by controlling the process at the business level were documented. The concept of **information governance** should be evident in the guidelines, both for users, including business management, but also for development:

> ‘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.’ (DMU, BCS 2nd wave)

However, the potential of **using the guidelines for educational purposes**, for developers and users that would not normally come across such documentation, is mentioned:

> ‘So, there is a way of really educating people and giving the information to people so it can really make sense to them.’ (UCLan Cyprus, 1st wave)

The following sections organise the findings into: (i) a discussion of important ethical and human rights issues that the stakeholders identified when going over the guidelines, (ii) limitations and challenges for implementation of the guidelines (iii) proposals for moving towards a successful implementation of the proposed guidelines.

3.1. Ethical and Human Rights Issues related to the proposed guidelines

3.1.1. Aim for transparency to address bias

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 participants highlighted, primarily, the significance of being able to **understand how technology works**:

> '… but one big thing, I would relate is explainability. To what extent a human can actually understand what the machine has done or has been designed to…' (UCLan Cyprus, 1st Wave).

The **link between transparency and bias** is apparent to participants and its importance is voiced:

> '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.'(UCLan Cyprus II hybrid)

Moreover, the participants feel that the guidelines have adequately considered such concerns:

> 'I like the way actually it was handled in the guidelines, on how you can have the bias on the input, when you design when the assumptions that you are making already you know, they have a bias in the development stage' (UCLan Cyprus, 1st Wave)

Nevertheless, the stakeholders raise **concerns on existing bias**, which compromises the online experience of the average user, focusing on the example of a specific type of bias, known as confirmation bias:

> 'one of the problems right now is confirmation bias on the web. So, the more you search the more pumped up you get about your own opinions. And this is because the most search engines, […], from my little experience, the artificial intelligence behind it is based on collaborative filtering algorithms. […] So, people of same preferences and the same output, same online behaviour…will get similar results.' (UCLan Cyprus, 1st Wave)

Additional types of bias stem from incomplete datasets can results in biased statements or decisions relevant to **gender, race, ethnicity, disability, etc.**

> '-…gender bias or something like that, any thoughts about did it work with people of colour as well as white people and all of that sort of thing.
> - Disability as well.' (DMU, BCS, 1st wave)

The idea of having a narrow view of the world because of bias in technology, can become a serious issue given the significance of the information, resulting in **incorrect information** and conclusions:

> 'we need to have that choice, we need to have that information to at least know that we are aware of what's happening. For example, as were talking I was thinking about the search

> engines because I work a lot on gender rights, and I was doing this thing today about sex workers. It only gave me the feminist view and then I just had to find something else and I had to search with other terms … And I just realised because they know [what] I search for, they know what I work with, and I was getting the same discourse. (UCLan Cyprus 1st wave)

Developers must consider **social effects** of their work and that a way that the guidelines can be useful.

> 'I think that's the sort of thing that you should be telling people to look out for in the document. I think that's reasonable. I think telling people to look out for whether what they're developing is going to put somebody out of work is a more contentious thing because if they're working for Tesla or something or even working for an RPA software company or something, it's automation. (DMU, BCS 1st wave)

Finally, the participants emphasize **that guidelines for developers may act as a solution** in attempts of eliminating bias:

> 'And I can see the advantage in reminding developers, for example, when they're building test bases to be able to explain, "Where is there bias?"' (DMU, BCS, 2nd wave)

3.1.2. Informed consent and consideration of human rights

Access to data poses access risks, as the system is by definition connected to the Internet in order to collect or generate such data. Informed consent of data collection or data sharing, especially where data is personal or even sensitive, is crucial. Technology design must consider the provision of informed consent input. Informed consent should, as implied by the term, be informed, i.e. providing the user all the necessary information necessary to make decision. The users have **a right to be able to understand** what they are consenting to. Linked to that, users should be able to maintain their **right of safety** and their **right not to be discriminated against**:

> 'So, if you don't have the right information and it's not clear to you, if it's confusing, or it's in a grey area, then that violates your right of safety in all aspects. You know, you compromise your safety in measures that need to be taken. The right to not be discriminated, which is also very important. I mean for me is on all levels that are related to the whole spectrum of human rights. (UCLan Cyprus, 1st wave)

Reinforcing the idea of the *right information* being available, participants focus on how **consent may compromise autonomy**, for example when you consent to forms of surveillance supported by SIS:

> 'But actually, at the end of the day, you are taking away autonomy and that's what they want to do because it's assisting in a different way to help the people live more independent lives. […] thinking about consent and autonomy and the difficulty of balancing. I think it's far more complicated and nuanced.' (DMU, BCS 2nd wave)

Thus, the necessary information must be considered during the design process. Participants consider **existing codes of ethics** since these often include relevant elements. An example of the code of ethics that security professionals must sign before receiving relevant security certificates is discussed:

> 'It says you keep any confidential information you gain from your activities confidential; you protect the intellectual property of the people you are interacting with; being honest and

forthright about your services; you don't use software that it's unethical or illegal; not being deceptive and lot of other things' (UCLan Cyprus II hybrid)

The **ideal**, however, may **conflict**, with practicalities of preparing such consent forms for SIS uses, such as **national or local** legal acts:

'I'm thinking about the health and social care act, where it says that if the NHS can use any of our data for the benefit of the NHS', that's it. So, we are not entitled to give informed consent when the NHS sells our data.' (DMU, BCS 2nd wave)

3.1.3. Issues caused by large amounts of data available to SIS

Even if the process of ensuring informed consent for technology users can be challenging, it is crucial in dealing with the large amounts of data available in SIS. Emerging technologies such as AI and Big Data use enormous amounts of data collected, manipulated and stored on the Internet. The necessity of using such amounts of data are also the cause of vulnerabilities for such systems, in terms of **compromising user privacy, and user security**:

'I think the biggest issue with such systems is the vast volume of information they collect on an individual, which makes it dangerous on every aspect. Like if it doesn't breach, they have all that information, the way information is used, like for promoting something or… It is such a big volume of information which makes it very hard to handle and very dangerous as well so…' (UCLan Cyprus, 1st Wave)

The large amounts of data that can be shared pose a lot of questions, with regards to red lines around what to share and who to share data with. In particular, particular areas of practice have very **different requirements for different types of data**:

'Basically, to share or not to share when you're dealing with medical information. It's a similar sort of question in terms of the ethics. It's not a machine learning type thing. It's all flowcharted, so it's an ethical or a moral question.' (DMU, BCS 2nd wave)

As in the example above, the impact of collecting and sharing medical data needs to be considered, and Information Technology professionals are not accustomed to making technical decisions based on ethical impact. Participants once more highlight **information governance**, such that decision-making is not the responsibility of the IT developers or supporting personnel:

'… about the user guidelines, that it seemed to be held within the confines of the IT department, which is not a department that's going to be thinking in that way, necessarily, or building up a body of knowledge or understanding about the decision making process.' (DMU, BCS 2nd wave)

Nevertheless, **control of data** once it is part of the Internet is not so simple, especially easily accessible data, such as data generated by social media. When different technology products access such data there should be a way to respect consent to sharing and **data ownership**. The guidelines mention similar concerns, which are picked up by participants, and the issue of **accountability** is highlighted:

'If somebody violates my consent to data sharing or whatever GDPR has put in place or whatever, the extent to which I can say something about it is I think very little. When a democratic process is being flood by artificial intelligence, we can call it whatever, I would say

> manipulation of data on social media and everything, and basically what are the mechanisms for having people accountable for violating those principles.' (UCLan Cyprus, 1st wave)

The complexity of how data can affect ethics and human rights through the development and use of SIS goes beyond control of data. The grey areas between being able to enforce **legal and ethical** guidelines, leave room for human rights violations. An example of how this complexity can affect the **right to be forgotten**, and how enhanced development guidelines can help in this, are elaborated:

> 'I think it's a different discussion on what is legal and what is ethical and where is the grey area, where companies can be unethical but being 100% legal. […] I was thinking about non-consensual pornography, when intimate pictures and videos and information about people gets leaked in soft porn sites, without their consent. […] So, if there is no way of that information being deleted. Because at that point it's really non-consensual. Somebody else leaked it that had access to it. So I think things are just with technology and with rights and with ethics and with legal, they are so multidimensional and so complex that there is no one way and it would be very difficult I think to find one way that we can accommodate all. […] I think it's very important that the data is deleted in this case because there is a lot of young people that we really need to protect when the information gets leaked on that level.' (UCLan Cyprus, 1st wave)

3.2. Challenges for implementing the Guidelines

The participants in the different Guidelines Focus Groups quickly addressed the issue of feasibility, and raised important limitations and challenges for implementing the proposed guidelines:

> 'My only concern on this is how feasible this is going to be. I mean if it would be feasible, I think it would be amazing.' (UCLan Cyprus, 1st Wave)

Oftentimes, it is difficult to even decide on what is ethical behaviour in order to move onto proposing ethical decision making. The red lines will need to be defined:

> 'I mean it's not illegal to work for a company that makes weapons. Some people think it's ethical and some don't. It's the same thing really as to working for an automation company that in effect is going to put people out of work because this company can process let's say invoices hundred times the speed of a person. Is that ethical? It's really difficult area.' (DMU, BCS 1st wave)

One of the main challenges concerned the potential for **informed decision-makers**. In fact, decision makers involve more than simply technology experts. Decision makers could be educators, policy-makers, social workers, lawyers, and other technology users. **Lack of relevant knowledge** could compromise their decision-making:

> 'I always say as an example, we've all seen Mark Zuckerberg, how he played the congress men after the Cambridge Analytica. He went to the Congress to justify what happened and he was talking in a different language than what the congressmen could understand. And that was a huge gap, that was a huge problem. I mean the congressmen should have been able to talk his language. And that's where I think it's a major gap. Politicians who… they vote on laws and they vote on how things should work, they are far behind on how technology takes over in culture.' (UCLan Cyprus 1st wave)

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.' (UCLan Cyprus II hybrid)

In fact, participants highlight the importance of **technologically- and ethically-educated politicians** as key stakeholders to apply the *User Guidelines*:

> 'In order for politicians to vote the correct, the right laws, they should have the education.' (UCLan Cyprus, 1st wave)

Similar understanding should be designed and implemented within the software itself, and SIS **developers should be able to approach their design** on a similar *educated* basis. This is especially urgent for AI design, according to the participants, because of the capabilities of such technology:

> 'To political representation and to equal representation as we've seen that in some cases seems that some opinions are not treated equally by artificial intelligence. (UCLan Cyprus, 1st wave)

> 'Because I see AI, is a kind of a weapon; you can shoot. Or nuclear power...and Iran is using it… So, it's another technology, it's there and you cannot stop the progress, the evolution. So, what you can only do is if humans try to use it in the correct way.' (UCLan Cyprus, 1st wave)

Next, there is **the challenge of time**, to educate developers and users. Even if high-level requirements are clear to all stakeholders, becoming familiar with the necessary detail for decision-making purposes is far from a simple task:

> 'It's a lot of terminology that falls under ethics, human rights etc. that for a technician… s/he is not familiar to these terms. They need some time to familiarize themselves and learn about these terms of course if they have to apply them. But also, I agree to how feasible it is. Because I noticed that in every product development process, you addressed almost all of those. So, in reality, in a real-life project, in a business project I don't think that you have time or the resources to address all these, unless you have the smartest system to do that.' (UCLan Cyprus, 1st wave)

Given the need for training and time, the challenge becomes greater for **smaller companies**, which will need to find the **resources** to go through with such processes:

> 'These things I don't think are applied for small companies. This whole document applies to big companies. Small companies cannot deal with data, they cannot deal with AI. These things are what drives marketing and what drives decision making.' (UCLan Cyprus, 1st wave)

The final challenge that participants repeatedly discuss, is the **motivation** of implementing these guidelines. The challenge of motivating both users and developers to incorporate additional tasks into

their routines needs to be addressed. Better **public awareness** to motivate individuals, or even **legally enforcing such implementation**, seem to be suggested approaches:

> 'Just developing the guidelines without any push, there is no effect. I mean it's good to have guidelines if people are using the guidelines. So, unless there is awareness about the guidelines, and unless there is awareness for the people so that they want to check the guidelines and look through them and so forth, they cannot be effective.' (UCLan Cyprus II hybrid)

> '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.' (UCLan Cyprus, 1st wave)

The idea of using some type of legal enforcement, where the threat of punishment is more real, is a way to get things started. **Accountability can be a strong 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?' (DMU, BCS 2nd wave)

3.3. Moving towards successful implementation of the Guidelines

As participants across the board seem to agree on the significance of implementing the proposed guidelines, as well as on the challenges that such implementation will face, several approaches towards a successful implementation were discussed.

Given the doubts of technology users, an element of value in the guidelines documents is the continuous testing '*whenever possible'* to ensure transparency. This is also a suggestion for future implementation, to enhance the process of **testing**:

> 'I mean the document here talks about trust in several places. It says, "If possible through testing," it says. I suppose someone could argue that if you can't test it then there are grave doubts about whether you should have it at all. Transparency is often mentioned. I think testing wherever possible.' (DMU, BCS 1st wave)

A technological solution could be the incorporation of ethics and human rights considerations, **in specific data tools, often used by developers**, such as data analytics tools, which have been quite popularly used by companies, including SMEs and larger corporations:

> 'I think it's a matter of incorporating ethics in these circles. In the cognitive computing, in the AI and Big Data analytics. It's something that's not there yet I think.' (UCLan Cyprus II hybrid)

Another way for companies to stay competitive and employ responsible development and use of technology according to the proposed guidelines, is by, in doing so, satisfying relevant standards. **Standardising the guidelines** such that they satisfy legal requirements would be a strong motivator for implementing them.

> ‘Any given company of a reasonable size is going to realise it’s way easier to standardise it because it reduces their legal risk, basically, and also the risk of them just simply getting a system that isn’t fit for purpose because somebody just didn’t understand the framework that they were operating in.’ (DMU BCS 2nd wave)

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.’ (UCLan Cyprus, 1st wave)

The above solutions have an element of enforcing the guidelines. The reason is that existing developers and stakeholders of the information society, do not have the necessary knowledge to comprehend the significant long term advantages of making a decision towards responsible development and use of SIS. Oftentimes, it feels like an unnecessary use of time and resources. This is a result of lack of education of how to responsibly develop technology, i.e. what to consider, but also of how to responsibly use technology. **Education** can act as a tool to improve such perceptions:

> ‘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 more relied 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.’ (UCLan Cyprus II hybrid)

> ‘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.’ (UCLan Cyprus, 1st wave)

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. This is the only hope that I see…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?’ (UCLan Cyprus, 1st wave)

Regarding improving the document itself towards being a useful guide for users and developers, the participants suggest to increase the number of specific examples and use cases that are mentioned, and follow the structure of specific sections in the document:

> 'The examples were really helpful when you say well you want, for example, do this, you know in the case of health care, this would be that maybe you would do this, or you would do that. And this specific example in many cases they were extremely useful to get the point. I would say as many as you can put, it's actually helpful.' (UCLan Cyprus, 1st wave)

> 'Section four was the section where you had the examples, where you had this kind of more operational... So, this is where I found the examples that I was talking about, that were I think really useful.' (UCLan Cyprus, 1st wave)

> 'I think the section five is one of the most useful and one of the most concrete sections. (UCLan Cyprus, 1st wave)

Finally, the participants discuss the need to regulate efforts for responsible implementation of these guidelines at the EU level. As such, they imply that there is a need for an **EU regulatory body**, which is the topic of discussion at the regulatory Focus Groups, which are analysed in the next section.

> 'It has to come from the EU and then be enforced locally. I'm not sure if they are including ethics in e-privacy regulation. It should be here by now, but they keep revising it. It will have similar effect to GDPR, but it will be more technical. It won't be just about cookies. It will be more about privacy and security in systems.' (UCLan Cyprus II hybrid)

We summarise the input collected in the Guidelines FGs in Table 3.

Table 3. A summary of feedback from the Guidelines FGs

General theme	Positive Elements	Elements to improve	Future work suggestions
Clarity of purpose	Clear description of how different ethics and human rights elements are considered	Better definition of what are the red lines on ethical vs unethical behaviour and decision-making	Work on standardisation or certification options for the guidelines
Comprehensive content	Comprehensive guidelines in terms of addressing human rights	Further consider the complexities of informed consent in the documents as well as the complexities due to big data	Exploitation through incorporation in appropriate curricula
Compared to existing documentation and common practice	The documents positively compare to similar guidelines documents	Address the gap identified relevant to a clear information governance (as is common practice)	Enhance the documents themselves with more examples and case studies
As an educational instrument	The documents can be used for educational purposes	Further consider education as a way to bridge knowledge gaps for stakeholders	Explore options for training existing professionals

Related to bias and technology monitoring	Critical issues of transparency to address bias are considered	Technology should be able to be educated as well, e.g. eliminated bias in terms of not just gender, but race, ethnicity, disability, etc.	Contribute to a push for an EU Regulator that can promote and monitor the implementation of the Guidelines
Guidelines and Societal impact	Social Context is introduced	Elaborate on social context to better address overall impacts	Improve public awareness of these issues
Related to trust	Frequent testing to ensure that the users trust technology	Address the view that responsibility may be viewed as a business risk for technology providers	Consider accountability in implementation and use, revisit information governance

4. Analysis of Findings: Regulatory Options/TOR new Regulator FGs

The current section presents the analysis for the regulatory Focus Groups, where participants are asked to comment on an EU regulator and propose corresponding potential approaches. Thus, the analysis primarily collects the regulatory issues identified in the corresponding Focus Groups, and then proceeds to highlight the aspects that, according to the participants' feedback, could be the basis for successful EU regulation.

The Regulatory Focus groups invited mainly policy makers, civil society and legal scholars to take part in discussions that explored a variety of topics, including the following:

- Identification of the fields of AI/Big Data that could mostly benefit from stricter regulation
- Identification of promising international options
- Identification of the most promising EU level options
- Identification of the most promising national options
- Identification of the most promising cross-over options
- Identification of immediate regulatory actions necessary
- Discussion of the need for any international bans
- Discussion of reaching a balance between benefits and risks
- Discussion of "smart mixing" options
- Discussion of how the law can help vulnerable groups and populations
- Discussion of critical future developments

The output from the regulatory analysis should contribute as input into the WP3 tasks, T3.3 and T3.6.

4.1. Regulatory issues for responsible SIS

4.1.1. What to regulate: consideration of important issues

It is challenging to decide on what to regulate, given the inter-disciplinary nature of the impact of SIS. Primarily, the **technology itself must be regulated** in terms of requirements and design. This will not be done from scratch, given existing work on security for example, and other responsibility aspects that technological development considers:

> 'what kind of requirements do we want that AI systems have? What kind of properties? What kind of requirements should we put on them? Like for example, protect human rights, including security, fairness and everything. I think EU has done already a lot of work on that. The other aspect is how shall we make stakeholders to meet these requirements?' (AHR, Online)

One such **existing framework**, is the **RRI** framework, identified by participants as a starting point for supporting a responsible technology development and use:

> 'I've been thinking about […] the framework of RRI – Responsible Research and Innovation, and that basically aims to look into how a new technology could be developed and used as well as an assessing potential or actual use of impact.' (AHR, Online)

However, not all technology is the same. The EU regulator should work with a specific definition of technology, e.g. when referring to AI, in terms of different types of AI, used for different purposes. Appropriate definitions can facilitate the **categorisation of risk** from different types of technology:

> 'I think EU should come up with criteria and also the categorization of different types of AI systems. How many are they? Can we categorize them into ten and then specify the risk factors for each type of AI? We have to go to the level where we can specify things … Because there is difference between robotic systems and a software system that makes decision support for example.' (AHR, Online)

Moving on from technology, the question of how to regulate use of these systems by different types of **industrial and societal stakeholders** needs to be addressed. Reaching out to different stakeholders, must be considered by the regulator so that everyone has a common understanding. Therefore, the **public awareness** aspect comes into play as something to act on.

> 'raising awareness about risks and benefits is an action towards satisfaction of the requirements rather than people, both the users and the developers, are aware of these risks and benefits. It helps them to understand the requirements and then related laws.' (AHR, Online)

In all cases, the regulator must be able to **monitor the development and use of SIS**, as well as the **impact on society and the environment**, using appropriate mechanisms. In response to a question about monitoring by propagating the responsibility to national level, participants agree:

> 'I mean there should be a very good collaboration between the national and the EU level. Also, nationals should take part in the EU commission, when EU makes the regulation. I mean EU is not independent. EU is all the national things as well. So, all nationals should be present there when making these regulations. But afterwards, the implementation and the inspection will be mostly at a national level…' (AHR, Online)

4.1.2. How to regulate: consideration of possible actions

Following, the participants' suggestion on the significance of national participation, the EU regulator could consider the above aspects at an EU level but also at a national level for each member state. Effectively applying the measures for responsible SIS development and use, requires that **national authorities must take action**. 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.' (AHR, Online)

A related task of the EU regulator would be **to educate or train** regulatory bodies at national levels so that they can **apply the necessary guidelines**, even where there is lack of expertise:

> 'Another strength is that 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.' (AHR, Online)

An **inter-disciplinary dialogue** must be a periodic action of an EU regulator. In particular, 'initiating a dialogue between IT-people (who know the technical aspects) with people (who know about regulations), in order to decide what they can achieve technically and which level they want to reach on a regulatory level.' (AHR, Cyprus)

Another strong tool is **funding**; funding of the regulator as well as funding by the regulator must be considered, including R&D funding for SIS development:

> 'R&D funding is influencing the development of the technology … It's like interesting pharmaceuticals right? … How can a regulator be interested in funding something?' (AHR, Online)

The coordination of funding, however, should ultimately not be the responsibility of the regulator:

> 'I can see a regulator being one of the partners in a project on such thing, like privacy enhancing where other types of research projects, but coordinating the R & D funding I think would be better suited at a different body than AI regulator.' (AHR, Online)

4.2. Proposed aspects for a successful EU regulator

4.2.1. Reaching a balance between technology benefits versus risks

Technology is rapidly evolving and society is usually playing catch-up. However, it is important that there is a **balance** in what the technology can do from a technological point of view, and what it should be allowed to do from a societal point of view. One of the areas where the benefits and risks must carefully find a balance, is that of **data control**, e.g. data generation, data sharing, etc.:

> 'It is important to safeguard the free flow of data, as personal data are protected. The commercial data are protected by the powerful ones themselves but with AI, those who collect data on a massive scale, they control the whole market.' (AHR, Cyprus)

The participants also note the challenge of addressing SIS development and use **outside the EU**, in order to reach a balance between EU and international regulation:

> ‘On an international level, there are numerous/different opinions and perceptions. A convention is not recommended because it will take time.’ (AHR, Cyprus)

Also, the quantification of benefits and risks can be a challenge even within the EU. Participants suggest **an impact assessment:**

> ‘In order to minimize risks/problems which do not derive from the design we must do impact assessment and find out what creates the manifested problems in order to try to minimize the negative effects/outcomes.’ (AHR, Cyprus)

Finally, it is important to **consider future development**, when quantifying benefits and risks, and thus the regulator would benefit from related expertise, for example within a stakeholders’ advisory unit:

> ‘Decentralised technology measures have to be taken into consideration – e.g. decentralised technologies/exchanges/blockchains/cryptocurrencies.’ (AHR, Cyprus)

4.2.2. Moving towards a ‘smart mixing’ solution

Both regulatory Focus Groups, address the issue of ‘**smart mixing**’, i.e. employing solutions from across the board. Specifically, this approach considers the use of a combination of technical and legal instruments, as well as ethical and social standards and guidelines, to offer a regulatory approach that is complementary, agile, and flexible.

> **‘**decision makers to be aware / educated on AI and to be able to understand in depth what AI is, what it is not and what it is doing.’ (AHR, Cyprus)

Given the combination of **different solutions from different sectors**, the new regulator can even make use of **existing approaches**, enhancing them and combining them:

> ‘There are some regulations already at the EU level, isn’t it? So, are we thinking about how to extent them, how to reshape them? I mean is it suggestions for further work on AI regulations?’ (AHR, Online)

For example, **promoting and developing responsible SIS** are two aspects that the regulator should be able to apply and monitor:

> ‘One is the EU promoting AI and Big Data as a new innovative market where it should take a lead as compared to the rest of the world, and on the other hand the EU is also wanting to take a lead in the responsible development of AI and Big Data. I imagine that AI regulator should perhaps combine these two.’ (AHR, Online)

‘Smart mixing’ may be the only viable approach according to participants, because of the **variety of stakeholders** that must be considered:

> ‘but it’s a very big job for EU to come up with all these criteria and all these, you know, specified at level that is actionable, that is usable by all the stakeholders. (AHR, Online)

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, that person 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. (AHR, Online)

Table 4 provides a summary of the feedback received from the regulatory FGs.

Table 4. A summary of feedback from the Regulatory FGs

For a successful regulator	What to regulate	How to Regulate
Smart Mixing: technical and legal instruments, ethical and social standards	Use of existing frameworks	Ensure application of necessary guidelines
Take advantage of existing regulation	Impact on society and the environment	Coordinate local authorities
Consider the variety of stakeholders	Stakeholders variety, e.g. industrial, societal, etc.	Educate/Train non-experts
Propagate action to the national levels	Public awareness	Promote an inter-disciplinary dialogue
Continuously monitor	Evolving technology itself, including understanding different types of the same technology, e.g. types of AI	Participate in Funding Management

5. Analysis of Findings: Exploratory FGs[1]

The purpose of the Exploratory FGs was to gain an in-depth understanding of stakeholders' views regarding what they consider as the main ethical issues that come out of AI and Big Data, on the challenges and limitations of current ways of addressing these issues, and their suggestions on how these ethical issues can be best addressed in the future. Reflecting the nature of qualitative research, the discussions were extensive and the data was overlapping and 'messy'; sometimes discussions spanned across questions or diverged from them, and did not always follow the indications in terms of number of items in the answers. Nevertheless, the data was rich, informative and achieved the purpose of the exploratory focus groups i.e.to explore stakeholders' views and suggestions on ethical issues in relation to AI and Big Data. As such all the views presented here are data driven, reflecting the positions, opinions and arguments of the participants.

A core aspect in the FG discussions was related to the collection, storage and use of data. Therefore, it is an issue which cuts across several of the themes discussed below. Also, some terms like 'transparency' were used in various ways by different speakers. For example, some emphasised

[1] Disclaimer: the views and opinions presented here are those of the participants of the Focus Groups and not of the authors. They are presented here as the results of the analysis of the data and do not in any way imply endorsement by the authors.

transparency in terms of having access to information collected by the government. In other words, they highlighted the importance of 'openness', while others focused on transparency in terms of data collection with the end point being the ability of the public to keep their data 'closed', secure and private. For others, transparency meant companies being explicit about, for instance, the way they derived their algorithms, the intended purpose of their app, or the business model that they are using.

As a contextual aspect, it is important to note that the majority of FGs took place during the Covid-19 pandemic and so this topic was present, to varying degrees, in 5 out of 7 FG discussions. These references are included in the report when they were relevant to the main themes that emerged.

5.1. Ethical Issues related to Big Data and AI

5.1.1. Loss of autonomy, human decision-making and control: 'a very, very dangerous direction'

A prominent concern for participants regarding Big Data and AI was the weakening or loss of autonomy, the ability to act and make choices without others making decisions for you:

> 'the core ethical issue, is this idea of the **undermining of autonomy**…when organisations or societies – or even individuals – start to accept the principle that a third party can, and should, make decisions for them, you're heading in a **very, very dangerous direction**. I think everything flows from that' (DMU-UKAIS).

This concern was presented as a 'major' ethical issue, related to how AI and Big data shift autonomy and agency (which was linked to power and responsibility) away from the individual to institutions. Ultimately, some political or commercial institute is 'making decisions for the individual' (DMU-UKAIS).

One participant gave the example of **Cambridge Analytica**[2] and the level of influence they had in the results of elections. This was presented as a breach to one's 'right to make a decision without any influence, especially when **we don't know that we are being influenced**, that's very very dangerous' (ETHICOMP I). Participants emphasised that there is a need for an 'absolute respect for the autonomy or the **freedom** of human beings' (ETHICOMP I).

There was reference also to the role of **governments** that played an active part in preventing citizens from preserving their autonomy in relation to decisions taken related to data collection as well as protecting their anonymity online:

> 'most of the national governments – and, regrettably, led by the Western governments – are also working towards the fact that they're denying us to have secure ways of communication or preventing our data to be collected. For example, the whole aspect of secure Internet is being undermined; the sharing of the keys – security keys – that is being undermined.' (DMU-UKAIS).

[2] The Facebook–Cambridge Analytica data breach refers to the harvesting of millions of Facebook users' personal data by Cambridge Analytica, without their consent, mostly for political advertising. Data collection began in 2014 and was officially disclosed by a former Cambridge Analytica employee in 2018. It is the largest known leak in Facebook history. Further information can be found here: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

*This was not presented as something passive or that happened by accident, but rather part of a **deliberate act of deception** which denied people the ability to resist, to have agency.*

This **deliberate act of deception** was seen as: 'a movement actually preventing us from denying those who are collecting data on us or allowing us to retain our privacy, retaining our anonymity' (DMU-UKAIS). Whereas in a lot of ethical issues there is a right to opt out or to object, in Big Data and AI, this was 'becoming increasingly difficult' (DMU-UKAIS). There was, it was argued, a shift from bottom-up to top-down control that made participants question whether people would be able to keep agency or are 'going to be subject to those who have very cleverly, really, **stolen the virtual world**' (DMU-UKAIS).

Going beyond loss of agency to third parties and institutions, there was a prevalent concern that AI and Big data delegates human agency to **automatic decision-making** which 'most of the times we don't know how it works. We don't know which data is being used to train an algorithm and so on' (ETHICOMP II). This was seen as a major issue impacting not only individual freedoms but also society as a whole. The issue here was that humans are replaced by computers and that algorithms drive decisions; the problem is 'the automaticity of the decision. And the fact that you have delegation of the decision to a machine' (ETHICOMP II).

Concerns regarding loss of choice and individuals being 'controlled' by SIS, went beyond political and health issues and extended also to more everyday cultural experiences, in terms of, for instance, one's choice of book or the restaurant one visits being dependent or influenced by the recommendations of the search engine (ETHICOMP II). Data collection becomes an ethical issue when it results in 'sharing [an] individual's life', which we often naively perceive as a 'gift'. The argument here was that people view these recommendations as simply making life easier, but ultimately, 'our choice [is] being taken away from us, we always depend on what is suggested' therefore losing one's freedom from control. (ETHICOMP I).

It was also noted that the ethical issue here is not just in terms of the data itself being collected but also the **lack of a critical perspective** when it comes to **decision-making** processes with long-term implications. There is a need for more emphasis on human values when taking decisions. As one participant observed:

> 'It's not just big data and AI, but the very process of **making decisions** based on data, large or small sets, and using that to inform the future, **without a critical and humanistic interpretation**, is fundamentally problematic' (DMU-UKAIS).

On a more specific level there were debates regarding autonomy and particular applications. In an FG discussion on **health and wellness apps**, for example, there was disagreement as to whether autonomy is sufficiently or should be addressed by the health and wellness apps (NEN). Some indicated that users have the autonomy to select the app of their choice and this is where the autonomy lies; others suggested that it should be necessary for users to be able to adjust settings to personal preferences, **choose** targets or strategies within the app etc. Transparency about the

'intended use' was seen as important, so that if a user did not like the app, or if the app did not offer what one was looking for another app could be chosen.

The need for the user to have a **choice**, the ability to have this 'flexibility', was related directly to ethics and the ethical responsibility of the company that develops the specific app:

> 'it's different when you have a **choice**. If I can customize something, it's my choice to customize it and see only what I want to see. But I need to have a choice... **Ethically,** the company that creates the app should give me the flexibility to see the full data and if I **choose** not to see the full data, and just selectively choose within the app "only show me this information", then I think it's my **responsibility** as a user' (UCLan Cyprus II hybrid).

In other words, once the company provides this ethical option, then the responsibility and power shifts to the user. But without this option, and full understanding of the consequences, the individual user does not have adequate options or the choice of whether they 'want to participate or not participate' (UCLan Cyprus Exp).

There was also a debate related to autonomy vis-à-vis **the Covid-19 pandemic** that was happening at the time of data collection. The participants raised questions and the need for a **'political debate'** to be had on whether it is ethically justifiable to 'give up some of our freedoms for the greater [good]'. For example, one participant implied that this may be a necessary thing to do given the current context with 'the demise of liberalist economics or capitalism as we know it' and a move towards community principles and socialism. In this case, the participant argued, speed - in search for a solution to the pandemic and data contributing fast to epidemiological models - may be prioritised over getting consent from those supplying their data (DMU-UKAIS). Others disagreed and pointed to the fact that a physical lock down is temporary whereas the collection of data in a virtual space may be a much more long-term project; as such 'it **constrains** your future actions in a way that being locked down for a period of time, and then that lockdown stops, doesn't' (DMU-UKAIS).

5.1.2. Loss of privacy: monitoring and surveillance

Loss of privacy, or 'the overexposure of data' as one participant put it, was also discussed as a crucial ethical issue across all FGs. This was related to data collection that enabled **monitoring and surveillance** - with or without informed consent - via tracking devices such as medical equipment, parents putting GPS trackers in children's bags, smartwatches, smartphones, search engines, fitness trackers, assistant applications such as Alexa or Siri, and home security systems.

Privacy often went hand in hand with references to **security** in the sense of data being 'secure' from third parties' unwarranted invasion of privacy, access to information or hacking. For example, one participant noted how children are at risk when parents naively put trackers in their school bags without taking into account the security implications if hackers gain access to their children's data, device and location: 'I know a lot of parents now put GPS trackers in their bags, but they can be hacked, or the data can be hacked, if not the device itself' (UCLan Cyprus II hybrid).

The right to privacy was expressed as a fundamental **human right** and the loss of it was presented by one participant as having spill-over effects on other human rights as well:

> 'democratic citizens have a right to, deeply important, right to privacy and if that right is compromised it's not simply your own free will that's at stake, it's the entire range of human rights, democratic rights such as equality, freedom of expression, you name it' (ETHICOMP I).

Participants also raised questions in relation to the duration (of time) that data are kept and potential negative consequences on an individual's human rights, their career as well as their future prospects in life. The specific ethical point here was the **long-term impact** of the **loss of privacy** on an individual and how this could constrain his/her future actions. Questions raised included:

- What does it mean for an individual when he/she provides this data for a common good?
- How could this affect him/her as an individual?
- How are data maintained, managed, stored, removed, deleted etc.? (UCLan Cyprus Exp)

In particular, participants argued that data collection and tracking through smartphones in the midst of the pandemic (to be able to monitor Covid-19 cases) should concern us in terms of the individual impact this loss of privacy may have in the long-term, potentially leading to **stigma and stereotyping** (UCLan Cyprus Exp and ETHICOMP I), for instance of those who were tested positive or who came into contact with a confirmed case.

Privacy seemed to be particularly emphasized when it came to **health-related** matters, with the pandemic giving additional weight and urgency to this discussion. For instance, in one FG that discussed health and wellness apps, privacy emerged as the most important ethical issue. There was also a concern from users of such apps that some parties such as health insurance companies could gain access to their data (NEN).

Not being informed, not knowing or understanding how this data was being used, as well as not giving explicit or informed consent was linked to **lack of transparency** in terms of data collection and use. One participant gave the example of a nurse working for the National Health Service (NHS) in the UK who was ignorant of how her nurse visits during the pandemic were being tracked:

> 'She didn't know that because she was using Google Maps with … GPS …, location services, that all her information, which is supposed to be **secure** for the NHS, is actually out there, knowing every single visit she's made and where she's been, because she's actually using a mobile phone, with Google Maps, to do her visits around the community.' (DMU-UKAIS)

The possible consequence of this was the exposure of private, sensitive health-related information outside of the NHS, and therefore the loss of 'security' of data.

Part of the problem was linked back to monitoring and surveillance and the role of both the public and Big Tech[3]. On the one hand, it was argued, the public - in the current context of the pandemic - has been forced to 'accept to be monitored for the common good'. This, according to participants, presents an unprecedented ethical challenge especially when research from Social Psychology and Organisation Studies has shown that 'the very fact of surveillance when we know we are being surveilled, changes the way that we behave' (ETHICOMP I). On the other hand, we have Big Tech that are not acknowledging the negative implications on human rights of this surveillance. According to one participant, in today's world of social movements, protests and 'expanding democratic rights', such ethical concerns and relevant calls for change could even potentially lead to, or even necessitate, violent unrest:

[3] The phrase 'Big Tech' was used by many participants; this term refers to the biggest, most dominant companies in the information technology industry.

> 'the tech giants tend to tell us that we shouldn't worry about **surveillance**. That if we're not doing anything wrong, you know, you have nothing to hide then what's the problem and part of the problem is democracy and **expanding democratic rights** whether it's the civil rights of people of colour, or if it's women or the environment now, that **always requires protest** and it always requires descent, sometimes violent unfortunately (ETHICOMP I).

Another ethical issue that was related to loss of privacy was the **monetisation** of data with or without transparency (and consent). Participants noted that often people were not aware of the fact that 'data is money and they are giving it away in everything they do' (ETHICOMP I). On the one hand, this is an ethical issue because companies are making profits at the expense of the loss of privacy of the individual, with disagreement being expressed by participants as to whether explicit consent made this less of an ethical issue or not. On the other hand, one issue that all participants agreed on was that monetisation was definitely unethical when this was done without the user's explicit consent or at least with their knowledge. There was a particular concern when organisations made profits by selling data that a health app generates, for example. (NEN).

5.1.3. Prioritisation of financial over ethical interests: big tech companies manipulating users

Loss of privacy, autonomy and decision-making was often related to the way in which data collection and harvesting (including biometrics, social media predictions, emotion prediction) was used by companies to direct human behaviour in a way that one could describe as **manipulation**. The Cambridge Analytica scandal was seen as an example of manipulating people based on data collection of what people seemed to be interested in. Companies choose 'how to phrase things, how to say things [so] that it's more believable to people' (UCLan Cyprus II hybrid).

Participants noted that there is a **responsibility** of companies to be completely transparent about data collection and present users 'the full data' and 'full picture' and 'not just a part of it', depending on what companies thought people would prefer to see (UCLan Cyprus II hybrid). However, there was also a realistic observation of the inevitable tension this raised between ethical and financial interests:

> 'there's an issue in which asking the big tech companies like Facebook, Amazon, Netflix, Google, Apple... asking them to restrict the data that they gather about us or the amount of surveillance that they undertake. It's like asking Harold T Ford to make each car by hand, you know... their whole modus operandi is based on data harvesting because the more they know about us, **the more predictable we are and therefore the more they can nudge us in the directions they want us to go which makes them more revenue'** (ETHICOMP I).

One other observer made a direct link of this type of company behaviour to the loss of free will and abuse of human rights (related to the loss of autonomy discussed in 5.1.1.):

> 'the philosophy behind the notion of human rights which presupposes the existence of free will and if, as human beings, we have free will, that presupposes the possibility to do otherwise….**there is an economic imperative driving many of those systems to render us more predictable, to nudge us, in certain directions, for the sake of the profit motive**. Our scope of possibilities, and what we are able to choose to do, becomes narrowed and I think there is a real case to be made to suggest that **ginormous global digital monopolies** are a threat to our free will, which thereby threatens all of our rights as human beings' (ETHICOMP I).

The prioritisation of financial interests – making money – over ethical issues was prevalent across the data sets and included not just Big Tech but also the Volkswagen scandal or Boeing's lack of transparency and manipulation of the Federal Aviation Administration.[4] Participants noted that even when there were legal standards and structures in place, companies were driven by profit and a 'top-down, do what the hell you're told' culture'. There was a discussion of this issue in the context of Shoshana Zuboff's argument that we are living in an age of **"surveillance capitalism",** the commodification of personal data for profit. This also linked to limitations in terms of what could possibly be done to address these issues (discussed in more detail in section 5.2.): 'It's almost a **utopia** to say, "We're going to, as the UN, create this regulation," because the authorities that create it will not create regulations that damage their objectives.' (DMU-UKAIS)

5.1.4. Lack of (access to) information and knowledge

In the ethical discussions of this theme, prevalent across all FGs, the 'knowledge keepers' were seen as intentionally trying to deceive the public and hence as the source of the problem. **Intentional deception** took various forms:

- terms and conditions written in small fonts so that readers avoid to read them;
- inaccessible 'legalese terms';
- text that is ambiguously written or expressed in a complicated way in order to deceive;
- avoiding discussions of issues that may awaken the public as to the role of Big Tech in, for example, spreading fake news.

According to one participant the knowledge of people of what Big Data is, its risks and the way it can be compromised is 'the biggest ethical issue'.

Part of the problem that exacerbates lack of transparency and loss of privacy and autonomy was traced down to the '**invisible'** nature of data collection, which made it difficult for users to fully grasp the data use, monitoring and surveillance that was made possible through SIS: 'I don't think there's a general understanding yet, really, of how sophisticated this type of data collection actually is and how invisible it is', noted one participant while another noted that it is also the 'data itself [that] is invisible and hard to understand for people' (DMU-UKAIS).

It was noted that it is important for the public to know about **ownership of data** and how data is sourced i.e. is it coming from the company itself or are the companies getting data from others and are these providers of data informed about this, have they provided their consent for data being used in various ways and analyses. (UCLan Cyprus Exp)? This formed part of a wider argument that it was important ethically for companies to have a 'user experience perspective' (explicitly referred to as such in DMU-UKAIS and UCLan Cyprus Exp).

For people to be able to **make a choice, to have agency** (see section 5.1.1.) rather than things being presented as inevitable, a prerequisite is needed, namely: honest and visible, transparent discussions that would allow users to make **informed decisions**. Access to knowledge and adequate information

[4] The Volkswagen scandal refers to the German car producer cheating emissions tests in the US. For further information see here: https://www.bbc.com/news/business-34324772 . The Boeing 737 Max was grounded in March 2019 after two planes previously crashed killing 346 people on board. For further information on why the company was criticised see here: https://www.cnbc.com/2020/01/09/boeing-releases-communications-on-737-max-simulators-it-calls-completely-unacceptable.html

was seen as a form of power, and the lack of it was seen as an **abuse of human rights**: 'the main issue with human rights…is the issue [of] giant companies gathering data without any you know, anyone saying you shouldn't be doing it' (ETHICOMP I). Without adequate information of the risks and consequences, the public is not able to resist, to protest against their violations of human rights or other ethical concerns or make companies accountable for their actions. This lack of information is also why, according to one participant, some users choose convenience over privacy: 'because they don't understand the consequences. Until it happens to them' (UCLan Cyprus II hybrid).

Secondly, there was a reference to **transparency** in terms of how the algorithm operates, for instance of a search engine: 'Even in Google search or any kind of search. We need transparency. We don't know the algorithm, right? We don't have information of how they present information to us' (UCLan Cyprus II hybrid). As one other participant noted, 'the average person does not know what is behind this algorithm, why it collects, what it collects, and where it will channel the information' (UCLan Cyprus Exp). Therefore, a crucial ethical issue was the inability of the public to be able to grasp and interpret complex aspects of algorithms due to the **lack of knowledge** (provided or of the user) or due to information presented in an **inaccessible** way.

> 'Who knows when someone presses on the 'accept' button, what happens to their information or how it is used? I do not believe that there is out in the public a good concept of what is happening out there with the use of their information. That is, let's say do people know that Google builds profiles based on your clicks and preferences? Which of us has looked at our Google profile in recent months? I have looked at it, I look at it regularly and it is always wrong of course (UCLan Cyprus Exp).

Some argued that invisibility was not inevitable and users could gain better understanding of how data was collected and used once they were provided with adequate information and when this was conveyed in an accessible way. One participant noted the role of **language** and being careful not to present algorithms as if they are a 'black box' which sounds like 'black magic': 'it's not magic. It's just numbers organised in ways that we can understand if people look at them properly' (DMU-UKAIS). This was also consistent with another counter-argument that pointed out that not all machine learning models are difficult to explain; there are certain models that can be explained, in terms of how they made their decision (ETHICOMP I).

5.1.5. Biased, inaccurate data and Algorithmic Bias

In addition to inadequate access to information, one important ethical issue was related to the accuracy and reliability of information.

Firstly, it was noted that AI systems (e.g. used in Facebook or Google) can amplify the spread of **misinformation and fake news** - with users interacting and sharing this information - and Big Tech have an ethical responsibility to prevent or at least control this (UCLan Cyprus II hybrid). Instead however, what they were doing according to one participant is deliberately avoiding the discussion of misinformation and fake news precisely because it is their own economic model of clickbait[5] that is 'the origin' of fake news: 'the more people click on news, the more they can gain information'

[5] Clickbait refers to content on the internet, usually in a form of false advertisement, that is presented in a form of a link, in a way that entices readers to follow the link and consume the data (read, view or listen to the content). It is typically criticised for doing this in a way that is deceptive, misleading and intentionally exaggerated to attract attention.

(ETHICOMP II). Again, a specific reference was made to the spread of fake news during the pandemic: 'now with the Coronavirus, people are sharing and sharing and sharing and it just spreads'(UCLan Cyprus II hybrid). It is important to note here that there was not always a consensus on the ethical responsibility of Big Tech. For example, whereas some participants clearly blamed search engines for deception, and for intentionally presenting data in a way that **spreads panic** during the current pandemic, others pointed out that Google, for instance, does not claim that it offers the most popular website as the first one; in fact, it is 'whoever pays Google more' that is going to come first on the list (UCLan Cyprus II hybrid). The argument was that it is a business model based on clearly marked advertisements and so ethically they are **not deceiving** the user. Still, this did not resolve the issue of data accuracy and reliability, related to lack of transparency of how the algorithm operates (see also section 5.1.4): 'But again, we don't know the algorithm. How that page was picked up? I think it has a lot to do with the social awareness. What you get from the internet is not always correct, not always the best or reliable answer' (UCLan Cyprus II hybrid).

Secondly, a significant aspect that emerged was in relation to algorithmic bias and decision-making being delegated to **machines instead of humans**. This occurs because machines and algorithms 'give more importance to the degree of evidence for our already confirmed biases'; this sets up a chain reaction of confirmation bias which focuses on numbers instead of humans, easily 'frames out' any related ethical questions and issues to do with human rights and values (ETHICOMP II). A core ethical issue related to dealing with algorithmic bias was approaching the ethical issue as if solving a 'mathematical puzzle' with 'mathematical certainty' rather than taking a holistic, cross-disciplinary approach that places individuals at the centre, both in terms of how they affect the decision but also in terms of the consequences of algorithmic decision-making on humans (ETHICOMP II). One example of the latter given in the FGs is biased algorithms that decide how long a person will spend in jail based on potential recidivism (i.e. the probability of a convicted criminal to reoffend). This in turn, it was argued, is based for instance on one's **ethnicity** or when one first came into contact with a police officer. The latter becomes particularly problematic when - to use an example given by a participant - an immigrant has in his record getting caught when crossing a particular border with his parent when he was an infant, and so in the future if they commit a crime, they will get more time in jail than someone who was not the child of an immigrant, simply because they came into contact with a police officer early in their life.

Further questions were raised as to how 'the predictive algorithms that people are using...further **erode human interest'** for instance with the financial market use of algorithms that has for decades focused solely on return on investment , therefore prioritising profit instead of human values. One participant also highlighted that one of the origins of the ethical issues is not only society's reliance on data, but the naïve 'built-in assumptions' that the more data the better without questioning if data is fundamentally and ethically good per se (ETHICOMP II). One other related example involved biased data sets that do not ensure equality or fairness. One participant gave the example of Brazil where judge sentences, she argued, tend to be **biased against women** (ETHICOMP I).

It was also argued that in a democratic society where there is rule of law, one should have the ability and the **right to object**, or to reject and contest an accusation. However, as one participant put it, one can contest evidence or go against reports of witnesses, but one 'can't contest algorithmic procedures', therefore a fundamental human right is stripped away due to the nature of algorithms being incontestable (ETHICOMP I). Ultimately, human behaviour was seen as managed and manipulated with little human agency; humans are 'treated as Skinner rats or Skinner pigeons in Skinner cages' (ETHICOMP I).

Thirdly, accuracy and reliability of data was seen as an important ethical issue especially when it came to **policy formulation** and its **impact on the public** (UCLan Cyprus Exp). The nature of Big Data was such that there is sometimes unreliability and inaccuracy due to a dependence on past cases, 'some kind of reduction' that made the results rather conservative (ETHICOMP II). There was also reference to huge negative impacts of statistical errors on the public, with an example given from the UK's trade statistics (DMU-UKAIS) but no further information was given due to the participant being bound by legislation to not disclose details.

Inaccurate and biased data was also seen as having a negative impact in the so-called 'migration crisis'. The argument presented by participants was that AI and data analysis is often driven by certain political agendas which spread **fear and panic**, with implications not only in terms of immigration policy but also on the views and behavior of both the prospective **immigrants** and the current residents of the host country. On the one hand, it was mentioned that there is false hope given to the immigrants that the new country they will enter is 'going to be a heaven' which is unrealistic. On the other hand, immigrants sometimes get a distorted picture that presents all locals as racists 'because they take all the bad examples and they are exaggerated' (UCLan Cyprus II hybrid). In terms of countries who are receiving the immigrants, the argument was that locals often are influenced in a negative, biased way, for instance by the use of certain language that exaggerates the risks associated with immigrants entering the country, emotionally manipulates people to induce fear, and that takes one negative possibility or example and spins it so as to make it seem as the norm. Again, the participants related this back to AI systems - for instance by search engines or social media - as well as the data behind the systems used, that 'make this problem even bigger because if someone starts a rumour it spreads faster and to many more people because of AI' (UCLan Cyprus II hybrid)..

5.1.6. Human jobs replaced by machines and dangers of machines and apps

A core aspect of this ethical issue was related to algorithmic decision-making. Building on the theme already discussed above in terms of the loss of human-decision making (see section 5.1.1.), there was apprehension expressed in terms of the dangers and risks that come with **machine domination**. Algorithms and humanoids could go out of control and have unintended but also unpredictable consequences.

'The decision making process that goes into developing the algorithms in the first place is embedded in those algorithms. In an Aristotelian kind of way. And then beyond that, the **algorithms then can start developing beyond the designer's intention**, behaviours and processes that have significant ethical impacts on stakeholders' (ETHICOMP II).

In addition, humanoids were discussed and presented as constituting 'a big challenge'. This was a new, fast-expanding area that is currently left unregulated. Humanoids reflected the 'full application of AI' and there was an '**ethical responsibility** of not being deceptive to humans' (UCLan Cyprus II hybrid). This was especially the case with **vulnerable** groups of people that could get emotionally attached to machines mimicking human behaviour without realising fully that they are not actually humans and so do not have any feelings:

> 'Humanoids are robots looking like people. So, especially with vulnerable groups, this is very dangerous because they make it attach and they don't realize that it's a machine and there are no feelings there because AI technology, programmers and so forth, they try to mimic human behaviour as much as possible and they try to make it as human, but in reality it's a

machine. So, there are people who do not have the capacity or maybe children who, you know, are not so aware of that and they get attached to machines' (UCLan Cyprus II hybrid).

The issue of liability was also illustrated using an example from accidents in self-driving machines. It was noted that already at a testing phase *Uber* had a deadly accident. The ethical concern raised was how will a decision be made in terms of liability in case there is an accident (UCLan Cyprus Exp)?

Moreover, it was argued that more attention should be given to the ethical issue emerging when **human jobs** are lost and human expertise is replaced by machines. Given current discussions on how AI might learn from experts and replace humans, there should be, it was argued, more emphasis on how individual experts can be compensated and be able to cope as 'they lose their expertness, when it is transferred to artificial intelligence and then it can be copied so' (ETHICOMP I).

Finally, there was a question of where does one draw the line and set certain limits on expansion or development so as to avoid undesirable consequences? 'When you have machines use AI and the machines can train itself(sic) on new things, where do you put the ethical line on how much they are allowed to expand?' (UCLan Cyprus II hybrid).

In terms of apps, it was noted that an ethical issue that was not mentioned in the SHERPA overview of ethical principles is the potential ethical implications of **'dual use apps'** e.g. applications that could be used for both health and military purposes (NEN). There were also calls to avoid generalisations and acknowledge that ethical issues may vary depending on the specific nature or purpose of an app. For example, ethical issues related to an app for companies to do personalised marketing 'will be very different from a medical application that will support the doctor making decisions about treating a patient. So we generalize the AI discussion, but the factors are different for each application' (UCLan Cyprus Exp).

5.1.7. Loss of access to services

This particular theme was connected to loss of access to services or software as well as job prospects. The most basic example was that if one does not accept cookies, then one cannot visit some websites and is therefore **denied access** to certain information and services. As one participant put it, choosing to opt out has become 'virtually impossible' now when compared to 5 years ago, adding that today, if one decides not to have a smartphone, for instance, one will not be able to have easy access to online services such as e-banking, look online for a job etc (DMU-UKAIS).

One participant pointed out their concern regarding the 'massive move away from cash' during the current pandemic which had the potential to exclude people from active participation and membership in a society especially when there was no alternative to pay unless one had a card or a smart phone (DMU-UKAIS). As one other participant put it, 'if you want to be part of this society and if you want to use all the devices and applications that everybody is using', you have no choice but to accept and comply with what a company wants (UCLan Cyprus II hybrid).

Even if one improves the understanding of how these technologies work and ways to go around data collection without consent, then people will not 'be able to use the full functionality of the software' that they are trying to use (DMU-UKAIS). Although one person noted the existence of GDPR regulation which meant that most websites gave you the option to uncheck certain cookies, the counter-argument was that this requires specific knowledge on the topic and 'an average person has no idea what cookies are. They are going to just press okay and move forward' (UCLan Cyprus II hybrid).

5.1.8. Loss of trust

The extent to which people can **trust AI systems** was raised by participants of the FGs. Trust was seen as important from the user side, aided by 'explainability and transparency' – this referred to the responsibility not only of the users to be educated but also of the creators/developers of technology to provide accessible information regarding their technology (UCLan Cyprus Exp). The lack of trustworthy computing services by Microsoft in recent years was mentioned by one participant, calling their actions 'treacherous' (ETHICOMP II). The response of another participant gave both arguments and counter-arguments:

> 'In many, many occasions, we have to trust the systems. For example, if you are in a vehicle which is driven … by a computer, in such case, you should trust the system, otherwise, you will never enter this car, this vehicle, which is driven by a computer. On the other hand, the computer system can make wrong decisions. So where is the balance between how much do we trust and should we trust the systems, the software systems that are making the decisions instead of us (ETHICOMP II).'

Another participant warned of the danger of using trust in a negative way - 'as a whip' - that used labels such as 'untrustworthy' or 'bad person' or phrases such as 'I'm not going to buy your product'. Trust he argued should be used **constructively**, not as a gatekeeper or a stick, but as a carrot that provided encouragement and incentives, awarding positive behaviour rather than seeking to punish negative behaviour (ETHICOMP II).

The role of trust was mentioned as an ethical issue that depended partly on membership of different **generations** e.g. Millennials[6] were found to stop using services when their trust was broken by a particular SIS, whereas individuals belonging to Generation Z were not really affected by trust issues; they cared less about privacy and 'more about the quality of the experience that they had' (DMU-UKAIS).

5.1.9. Lack of accountability

The ethical issue here was traced to the **lack of accountability**; or even a lack of fear of punishment as a deterrent for unethical behaviour and data manipulation for political or commercial reasons. Referring to this one participant noted:

> 'Accountability is the key that is not adequately addressed yet. We have Cambridge Analytica, but the Chief Executive didn't go to prison. We have other people who are actually manipulating data for political and commercial reasons, but nothing happens. They get fined by a miniscule amount of money, so, therefore, accountability is not adequate' (DMU-UKAIS).

Other specific examples mentioned were regarding individuals heading Big Tech such as Mark Zuckerberg or Bill Gates. The former was fined by the EU, but this 'doesn't mean much to him'. Both it was argued had no interest in the ethical consequences, but rather focused solely on the technological aspect (ETHICOMP II).

[6]Millennials tends to refer to those born between 1981 and 1996. Generation Z refers to those born from 1997 onward. For further information see here: https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/

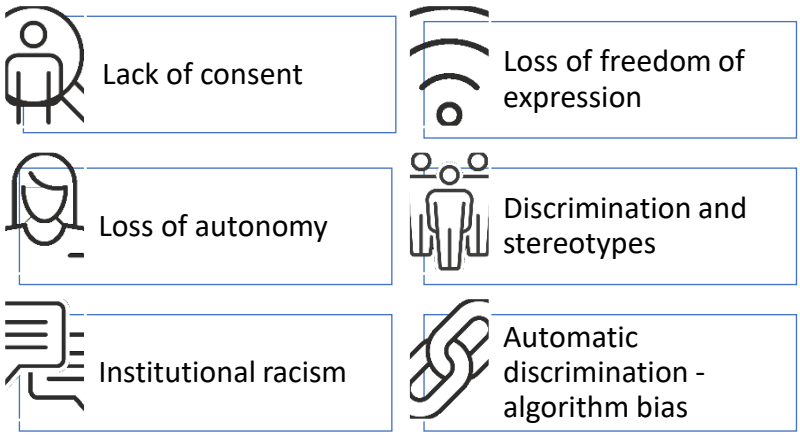
A participant from the US on the other hand noted that he believed the EU had done 'a lot' with creating 'reasonable constraints and then holding people accountable' but lamented the lack of adequate and similar efforts in the US (ETHICOMP II). There was a naïve assumption, the participant noted, that people are inherently ethical and always want to do good.

5.1.10. Threats to (and violations of) human rights

The topic of **human rights** was one that spanned across the focus group discussions and intersected strongly with other themes included in the sections above, such as access to information and knowledge, loss of privacy, loss of autonomy, biased data etc. This section will therefore provide an overview of the ways in which human rights were discussed vis-à-vis ethical issues of Big Data and AI.

Human rights were presented as closely related to one another e.g. equality linked to freedom of expression linked to data protection rights, and also strongly linked with an ethical approach as part of a working democracy. For example, **surveillance** was seen as a violation of the right to privacy, and this was connected with the loss of freedom of expression and the right to be free. There was a recognition that in today's environment of expanding democratic rights and bottom-up power (e.g. anti-racist protests or climate change protests), human rights have inevitably come to the fore and therefore deserve more emphasis in discussions related to SIS in order to identify how they are threatened, how they are violated and what people can do to prevent this. There was an emphasis on ensuring that AI systems are available to all and that **everyone benefits** rather than worsening inequalities (SB Meeting). Violations of human rights discussed in the FGs included inequality, lack of consent, loss of freedom of expression, loss of autonomy, discrimination and stereotypes based on ethnicity, colour or gender, institutional racism, and 'automatic discrimination' (referring to biased algorithms and decision-making by machines) (see Figure 3).

Figure 3. Human rights violations discussed in the FGs



There was also a recognition that 'not all countries are democracies' and that non-democratic countries may have difficulties with adopting human rights (DMU-UKAIS). Yet, there was also an acknowledgement that human rights violations take place in so-called democracies. One example referred to was the Cambridge Analytica scandal where data was manipulated for political gain in a democratic state, as noted earlier. An association was made with Big Tech - which due to their financial power enabled by data collection – are becoming a political force that constitutes a threat to democracy (ETHICOMP II)). Manipulation of data for either financial or political gain was seen as an abuse of human rights and one participant made the case that:

> '**human rights are always at a threat** even in the democracies that we live, let alone the countries that do not enjoy democracy' and will always inevitably 'be abused' (DMU-UKAIS).

The Cambridge Analytica scandal was also seen as breaching the right to make a decision free from external unknown influence (see section 5.1.1). For other participants, lack of access to information, in this case lack of transparency in terms of government policies and behaviour, was also viewed as 'invad[ing] our democratic rights' (DMU-UKAIS).

A counter-argument to the 'data collection as a threat to human rights' position was put forward by one participant, who noted that 'being on the record somewhere is not always negative for human rights' especially when it comes to gender discrimination (DMU-UKAIS). The point made here was that in some instances, e.g. data on health and safety or medical data on women's body sizes are sometimes excluded and so lack of data collection may in itself prevent the protection of human rights. This could also happen, for example, when human trafficking takes place and women are not protected because they are not 'on the record'.

When it came to who violated human rights, there were various actors identified with the most prominent ones being so called Big Tech and producers of AI-based technologies or state actors (nationalist governments, state leaders). On the one hand, it is these two actors who for political or financial reasons/gains threaten human rights, and on the other hand, it is the user of these technologies that often has a very low awareness of these ethical aspects related to SIS 'so they are not demanding from the producers, protecting their rights and addressing those ethical issues, so the producers don't' either (ETHICOMP I). In this sense, human rights are only going to be protected when social awareness about their violations is strengthened by the consumers/users themselves.

One example of the ways in which freedom of thought, belief and expression were affected by SIS was offered with reference to the government -in this case reference was made to China - that can **'switch you out of society…'** as a result of monitoring individuals' social media accounts (DMU-UKAIS). Another example was given in relation to Brazil with one participant noting that the current president is acting like 'a dictator', suspending all legislation related to data protection, privacy and freedom of information (ETHICOMP I).

Another argument put forward was that in the context of AI, often the ways in which human rights are violated are less explicit than other forms of human rights abuses, and therefore more difficult to identify. For example, as we saw in section 5.1.5., biases in algorithms that may lead to unfair treatment in **justice systems**, are a technical issue that may not be easy for the general public to comprehend, to visualise and protest against. Nor is algorithmic decision-making always transparent and contestable, therefore affecting in legal terms the ability of one's right to due process and a fair trial:

> 'the concerns I was raising about the transparency of algorithms…[i]n legal terms those affect directly I think what lawyers would call our rights to due process, that if we are accused of a certain crimes or suspected of planning certain things like terrorist actions, we still have a right of due process before that accusation can move much forward. So that would be a very specific link with the legal aspect of human rights (ETHICOMP I).'

Moreover, an ethical question was raised in terms of how one defines what counts as a breach or support of human rights when it comes to AI systems (SB Meeting) given that this is still a relatively new field with little regulation.

In a particular focus group, though equality was acknowledged as important, it was argued that equality is *not* the same as **inclusivity** and that certain app developers, for instance, are driven by business models that have specific target groups. Such developers are not always interested in making their applications user friendly or accessible to all and the fact that this could potentially exclude people with disabilities (e.g. deaf, blind) or the elderly was not viewed by the participants as unacceptable (NEN).

5.2. Challenges and Limitations of Current Efforts to Address Ethical Issues

When discussing **current efforts** to address ethical issues of SIS, participants agreed that the nature of technology was such that it had now become an inevitable part of our lives and 'opting out' of it whilst remaining an active part of society had become virtually impossible. It was seen as an inevitable issue affecting everyone, but also one in which people were becoming increasingly dependent on: 'I don't know if there is any turning back at this point in the level of technology that we all depend on so much' (ETHICOMP I) noted one individual while another sought to emphasise how it has become enmeshed into our everyday experiences. 'It is deeply embedded in our lives…let's bear in mind that Google uses an algorithm in order to provide us with information and we all use Google in our daily lives to find information from the age of 5, and how much this has influenced various phases of our life' (UCLan Cyprus Exp). It is against this backdrop – of the **dominant, fast-paced and inevitable nature** of AI and Big Data – that current efforts to address ethical issues in SIS were discussed.

This section offers a snapshot into the discussions regarding firstly, the challenges and secondly, the limitations of current efforts to address ethical issues of Big Data and AI. Current efforts and initiatives were sometimes discussed in general terms, while at times reference to specific initiatives, regulations or legislation were made. Many participants also expressed the opinion that there is already 'a lot of discussion' of ethical issues and there are already 'lots of regulations regarding ethical issues in science' (such as the ACM Code of Ethics) (ETHICOMP I) including several journal publishers and funding bodies that employ their own ethical guidelines.

'There's some appetite' noted one participant, referring to the increasing interest among people to delve more deeply into the ethics of AI (SB Meeting). At the same time it is important to remember that the FG participants were selected precisely because of their expertise and so it is more likely that they are involved in ethics debates and are more knowledgeable than other members of the public. Besides, as was noted a prominent ethical concern in SIS was the lack of sufficient awareness and understanding of the public of these ethical issues. As one participant remarked, even though the majority of people 'are becoming more technologically literate' and the pandemic caused a 'huge technology learning curve', most people are not 'truly aware of what goes on behind the machine' (DMU-UKAIS).

Challenges

5.2.1. Different perspectives depending on gender, age, occupation and other factors

Several participants across the FGs noted that it is a challenge to agree on what constitutes an ethical framework given the different and sometimes conflicting perspectives and assumptions on this topic. 'We are speaking a different language' noted one participant when she referred to discussions she had as an academic/technology specialist with legislators. Various factors and considerations were

mentioned that affected people's interpretations and how seriously they viewed ethical issues. Figure 4 lists the various factors identified by participants and Figure 5 the range of considerations depending on one's occupation, position, status etc.

Figure 4. Factors identified as affecting perspectives on ethical issues

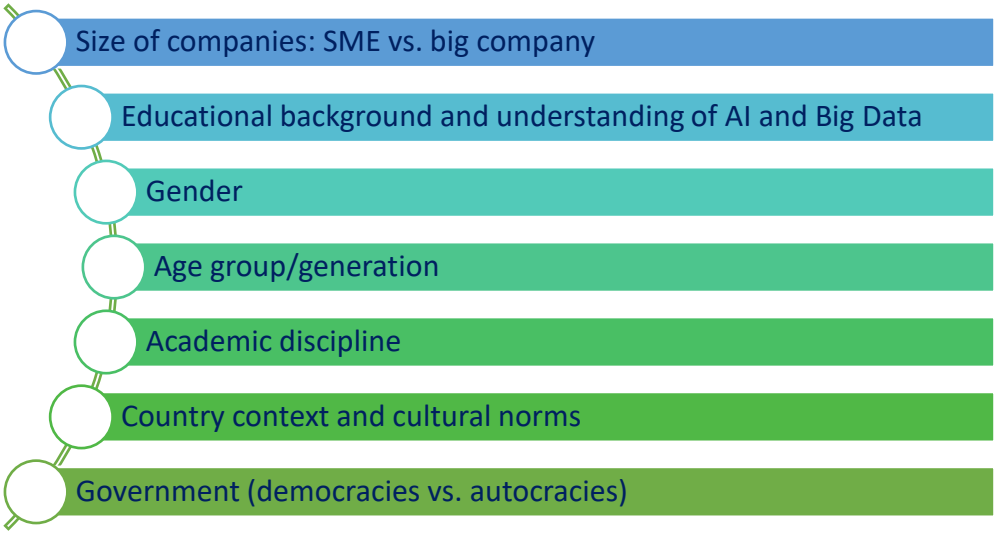
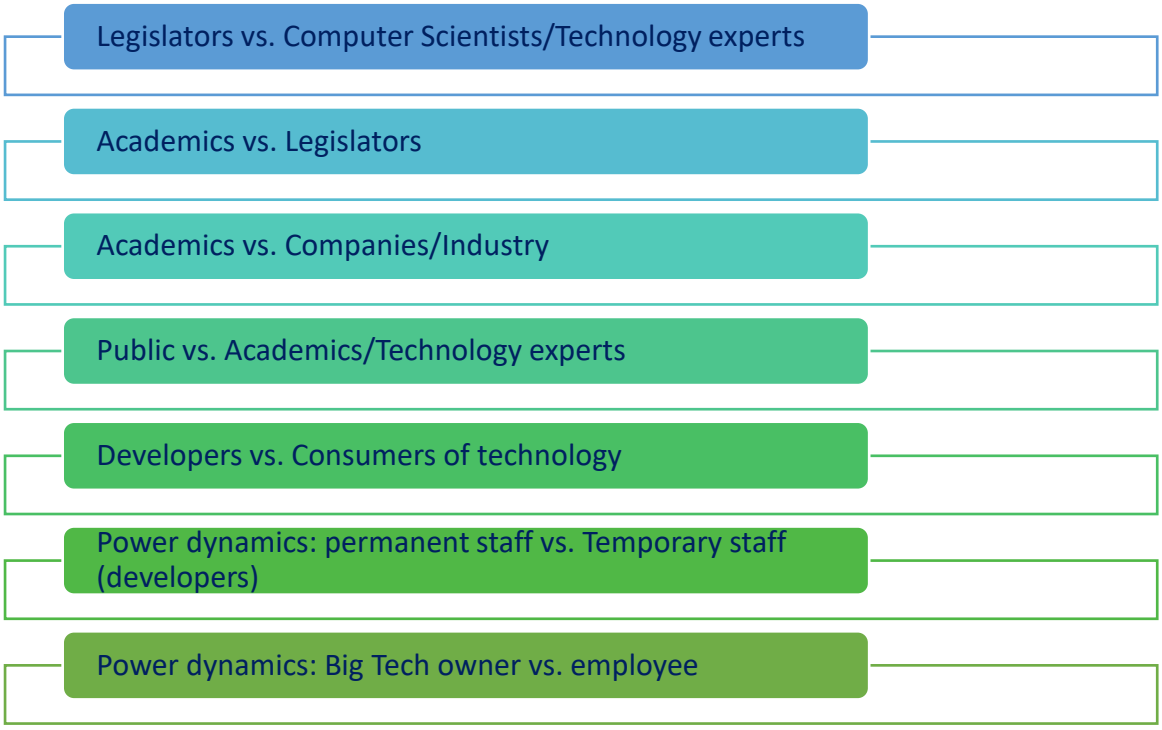


Figure 5. Range of considerations affecting perspectives on ethical issues



5.2.2. Democracies vs. Autocracies

A second challenge that FG participants noted emerged when attempts are made to address ethical issues is the rise of **nationalism, populism, hate speech and weak democracies**. As we saw in previous sections, there were various references to both specific scandals but also country contexts such as the US, the UK, China and Brazil, where human rights were viewed as under threat or violated with the public having little agency to resist or reverse the negative consequences. Participants spoke of **governments** intentionally and actively denying people data privacy and anonymity, with surveillance that not only manipulated their behaviour but was done by both companies and governments in such a way that the public often would not realize it was something problematic. As one participant put it:

> 'when we leave giant companies gathering data and when these data is used to influence the population, we can have a population that sees a leader suspending this kind of law and doesn't think that would be a problem because they think that in the same way that companies and people that benefit from it like us to think, that if I have nothing to hide why would I not want to be surveilled '(ETHICOMP I).

One participant spoke about a seemingly disintegrating world were basic institutions such as the World Health Organisation or the European Union were being challenged and in this 'dangerous' context it was crucial to overcome these forces of fragmentation and protect human rights:

> We see from President Trump's take on life that he wants to, effectively, fragment the World Health Organisation as we speak; the British perspective to fragment Europe, which is going on in association with that. While the United Nations is out there, as well, charities are in meltdown. Therefore, **this is a very critical time to retain some semblance of regulation and human rights on a global scale, due to the rise of nationalism, really, and blame games** that are going on (DMU-UKAIS).

Within this context one other participant juxtaposed two opposing forces: academic theory on the one hand with projects such as SHERPA trying to produce a positive impact, and populist debates on AI and Big Data on the other hand, which are channelled through mainstream media.

5.2.3. Lack of critical thinking and self-reflection on ethical aspects

The challenge mentioned here was lack of ethical **self-reflection and critical thinking** in order to be able to recognise biases and individual responsibility. There seems to be a simplistic binary of good vs. evil which makes it difficult for developers to be ethically reflective and pro-actively engage with ethical issues. As one participant pointed out:

> 'most of us think we're ethical and we operate with a very bad ethical premise that says I'm a good person and evil is caused by evil people. I'm not an evil person. So I don't have to worry about it. So when I write the algorithm, I'm a good software engineer. I don't even have to question this. I'm doing a fine job' (ETHICOMP II).

A related point was made in terms of being aware of one's own subjective **cultural norms** that may affect one's decisions. Not only was this critical awareness often missing, but crucially, the nature of software is such that it is difficult to change these underlying norms once the system is built:

> 'the cultural norms that we have, but don't even realise we have, that we use in order to make decisions about what's right and wrong in context. It's very difficult for any software system, even a really... advanced one, to transcend its current context. It's locked in to however it was framed, in whatever social norms were in place amongst the developers at the time it was built' (ETHICOMP II).

5.2.4. Ethics washing and the difficulty of educating companies and developers: a 'niche society'

The **education of companies** in relation to ethical issues was seen as a challenging task, not just in terms of lacking financial incentives but also ethical incentives or lack of an organizational culture of ethics in a company. 'You need to convince the managers' noted one participant (UCLan Cyprus II hybrid) who agreed with another participant who suggested that such issues may not 'affect them in any way. So, they don't have to care about it. That's why it's harder for them to apply it anyway'. Referring to **developers**, one participant noted that ethical issues are 'a bit too much' but left a window of small hope that perhaps finding the right way to communicate such issues may have potential in the future.

Another point raised regarding AI developers by some participants was that they usually 'represent only a niche of society, a particular **niche society**' and they do not always have the required pluralistic, diverse and broad perspective so as to build 'inclusive technologies' (ETHICOMP II). Agreeing with this point one participant who referred to himself as 'an old white guy' seconded this opinion arguing that narrow points of view of developers extend into and are reflected in the software:

> 'I'm **niche market** and I do the photo recognition software and I'm an old white guy. So the only people I recognise are white males with beards. And that happens in the **software**, we know it's happened and we've **framed out the ethics**' (ETHICOMP II).

Some IT companies were seen as giving empty promises or overpromising but not delivering (DMU-UKAIS). Another point raised was that companies tended to present the final end, a seemingly positive end point as a means to justify unethical means, thereby absolving themselves of the 'moral consequences' of their decisions. One participant referred to **'ethics washing'**, that certain large corporations merely want to give the impression that they care because they have a product to sell and if it looks ethical or they say it is ethical that will help their sales, even when it is not actually ethical (ETHICOMP I and II). There was however an admitted point that some companies may be genuinely interested in ethical issues and an appreciation of ethical frameworks such as the IEEE.

The problem that emerged here was that even when designers have the will to make ethically suitable designs, it is often difficult to provide them with 'concrete guidance' due to the complex nature of SIS (ETHICOMP I). Even regulations did not always provide specific directions on how to deal with data governance, corporate responsibility and other ethical issues (UCLan Cyprus Exp).

5.2.5. Challenge of users themselves choosing convenience over privacy/other ethical concerns

This challenge was related to the end-user. The argument was that even when users to some extent knew about certain data collection breaching privacy, some still chose access to a service and getting 'the job done' rather than paying attention to the ethical issues at stake. Although, as mentioned above, there was an acknowledgement that public awareness of ethical issues is growing, a prominent argument was that people 'love technology' (ETHICOMP I) and tend to avoid taking serious action in response to ethical concerns 'until something bad happens to you, personally, or on a larger scale' (DMU-UKAIS). In other words, Especially when it came to Generation Z (in contrast to Millennials) 'they didn't care so much for one error or one incident with an error in terms of privacy. They cared more about the quality of the experience that they had' (DMU-UKAIS). Giving the example of smart home devices such as Alexa or Siri, one participant remarked that when having the dilemma of convenience versus privacy or security- for instance, having the application to be constantly listening

to your discussions so that it responds when you call it versus having to press a button to activate it - then '[a]lmost every user chooses the convenience over privacy' (UCLan Cyprus II hybrid).

5.2.6. Not enough pressure on AI developers/producers to address ethical issues: 'no punishment for the bad actors'

One other challenge identified when trying to address ethical issues of SIS is the **lack of pressure** on AI developers to take responsibility and adequately address ethical issues. One participant who said that they had substantial experience with producers of AI-based technologies and solutions stated that because they are not really interested in addressing these issues and especially when this would require more expenses, their approach was one that tried to ascertain 'what is the minimum we have to make to be according to the law and not to address the issues really in full'. In other words, this approach was along similar lines to what one other participant called 'a checklist approach', merely to be able to tick the legal boxes in a superficial way that ensured the companies were allowed to operate by law even if the ethical issues were essentially left unaddressed or under-addressed.

Big Tech companies like Facebook were mentioned as an example of how companies manage to 'get away with things' (DMU-UKAIS) when malpractice has occurred, despite laws and regulations and this indicated a strong limitation or even failure of existing efforts to address ethical issues. This is related to the theme of the 'lack of accountability' discussed in section 5.1.9. above. As one participant put it, it often seems like 'there is **no punishment for the bad actors'**, no deterrent to prevent them from unethical practices (ETHICOMP II).

Another reason provided for the lack of pressure was the fact that Big Tech have managed to have 'minimum regulatory intrusion' because they leverage their **financial and political power** to successfully 'lobby the legislators that are supposed to be regulating them' in the first place (ETHICOMP I).

5.2.7. Mismatch between business models/market needs and ethical needs

The discussion here revolved around health and wellness apps and specifically that whereas people may have multiple health issues and would require an app that addressed these interrelationships, the market usually offers apps that focus only on one health issue (NEN).This was viewed as a mismatch between what the market offers and what people may need.

It was also pointed out by some participants that addressing ethical issues would potentially increase the costs for developing apps and reduce the likelihood of especially the smaller companies finding regulations useful (NEN). Following this rationale, a couple of participants noted that there was a 'vacuum' in terms of regulations in Big Data and AI that needed to be filled, but at the same time a stark reminder that 'we cannot regulate everything' (UCLan Cyprus Exp).

Limitations

The limitations of current efforts to address ethical issues in SIS often produced general comments from the FG participants on the lack of legislation/regulation or specific feedback on particular legislation/regulatory frameworks.

Ethical policies, legal systems and regulations are often too slow to emerge and cannot keep up with the fast pace of technology, according to some participants of the FGs. It was argued that: 'companies are already struggling with the GDPR. If we talk about global companies, then it's even more of a

struggle because, … Just like every other technology, the technology advances and then the policies start to follow.' (DMU-UKAIS). The range of tools available to mitigate the ethical consequences of Big Data and AI is limited. There are laws but these do not go far enough, offering effective data protection. In addition, there is also the problem of who is going to do the monitoring and ensure that people or companies comply with the regulation (UCLan Cyprus Exp)?

Again, with reference to **GDPR**, one participant argued that despite the general positive aspects of GDPR which they really liked, what is lacking from it is group privacy protection that goes beyond individual data protection and looks at 'how the data are being merged, are being collected, and so this kind of a connection between people…the protection is not strong enough there' (DMU-UKAIS). This was seen as particularly important in today's context with 'new organisations' which are 'booming and using D&A (Data and Analytics) just to do many things'(DMU-UKAIS).

Another 'major limitation' with the GDPR was identified as its inability to 'cope with blockchain' as well as its negative effect on innovation because of the restrictions it puts on companies (DMU-UKAIS). Despite GDPR, consumers are still monitored and their privacy is not protected, so it is not as effective as it could be (ETHICOMP I). GDPR has 'not even touched the surface' of issues related to data ownership, how data is sourced, maintained, managed, removed etc (UCLan Cyprus Exp).

As another limitation, it was noted that ethical issues were not being mentioned or dealt with in certain regulations, for instance, **ISO27001** (UCLan Cyprus II hybrid).

It was noted that despite very encouraging efforts (such as GDPR, high-level discussions, and IEEE ethics by design), which do refer to the two pillars of a) 'virtue ethics approach' and b) 'deontological terms', there was still a long way to go until these are adequately carried into actual policy or laws. The first approach (virtue ethics) is a holistic one that emphasizes the importance of **human flourishing**, of having good lives and going beyond focusing only on the profit of corporations. The second approach (deontology) constitutes one of 'absolute respect for the autonomy or the freedom of human beings' (ETHICOMP I).

Another limitation put forward, this time specifically for the human flourishing approach, was that it may not gather much interest in the UK, because an empirical view is more dominant there: policymakers and organizations think in terms of remedying harms and take a more goal-oriented, empirical approach, whereas the concept of human flourishing is more holistic (SB Meeting).

Legislation issues: Participants commented that legislation sometimes merely leads to **lengthy legal battles** and protects big companies rather than consumers especially as big companies lobby legislators and authorities: 'Yes, we have, with **GDPR**, these massive fines, but then all I can see that that leads to is a protracted legal battle' (DMU-UKAIS). Other limitations of such efforts were related to the fact that pushing new legislation through parliament is a lengthy and **slow process** (unlike the fast-paced nature of technological advances). Participants also made observations regarding **power dynamics** – 'it all comes back to politics and power' – that ultimately meant legislation ends up protecting big companies rather than consumers. Examples were made of companies that lobby legislators and specific cases such as IBM were mentioned where they just 'sat it out and made things very difficult for a period of years until the case was dropped' (DMU-UKAIS). Another example of a limitation regarding legislation was the Volkswagen scandal (and relevant study) that showed how even when there are legal structures in place, companies are still able to behave illegally, prioritising **profit over ethics**.

One participant was quite critical of what he called the '**checklist'** approach to ethical issues. Firstly, he was critical of the AI community in its approach to ethics, arguing that it 'thinks it is inventing ethics'

and that organisations writing ethical standards are currently doing so without looking at previous efforts in other areas of ethics. They are therefore lacking context and not trying to learn from past mistakes. Secondly, explaining what he meant by the 'checklist approach' he criticised what he saw as a very mechanical, superficial way of approaching ethical issues:

> 'they're producing a standard, a checklist, a thing that you do as if, "I do this, this, this and this, my AI will be OK"...if you have a compliance checklist, what happens, at least in companies, is that **checking the box is the consideration rather than the ethical impact of what you're doing.** So did I conduct a test?...So I get to check this box and I'm done, not a question of how it impacts others or raising other kinds of questions, but just did I do this kind of test? Yes. Have I got a comment in the code? Yes. And it's not a question about its ethical impact…And if you do this, you're doing good AI. So did you test that you coded properly that your programme doesn't crash? Yes, I did. Did you check that if people try and use it, they'll move their hand too fast and will get carpal tunnel syndrome? **Well, no, that's an ethical issue. I don't have to do that and I don't have to deal with this.** (ETHICOMP II)

The participant was also critical of the language used i.e. '**codes'** of ethics, which were treated as checklists, rather than explaining *why* certain **values** are important and *why* programmers should care and deal with these aspects. This he argued is a limitation as ethical codes are currently being 'treated as constraints rather than opportunities for goodness' and as 'gates'. In other words, they are not used in a constructive way but as a something that people fear they need to comply with or else face repercussions. An exception to this, he argued, was the ACM Code of Ethics[7] which instead of constraining the way that computing professionals could operate, focused on opportunities and responsibilities for improving society and working with stakeholders.

Thirdly, and related to the above, the participant argued that this 'checklist approach' is a limitation that can be found in recent **EU regulations/codes of ethics** that were released in late 2019. Again, this approach focused on producing quality software rather than on how to best support and improve society and stakeholders.

Limited ethical focus of **QRCA** (Quality Requirements Conformity Assessment): The standard on quality and reliability criteria has an underlying questionnaire of some 100 questions but only 2 out of 100 questions address 'ethics'. It was also found that ethical issues such as human agency, diversity and accountability were not sufficiently addressed so far (NEN).

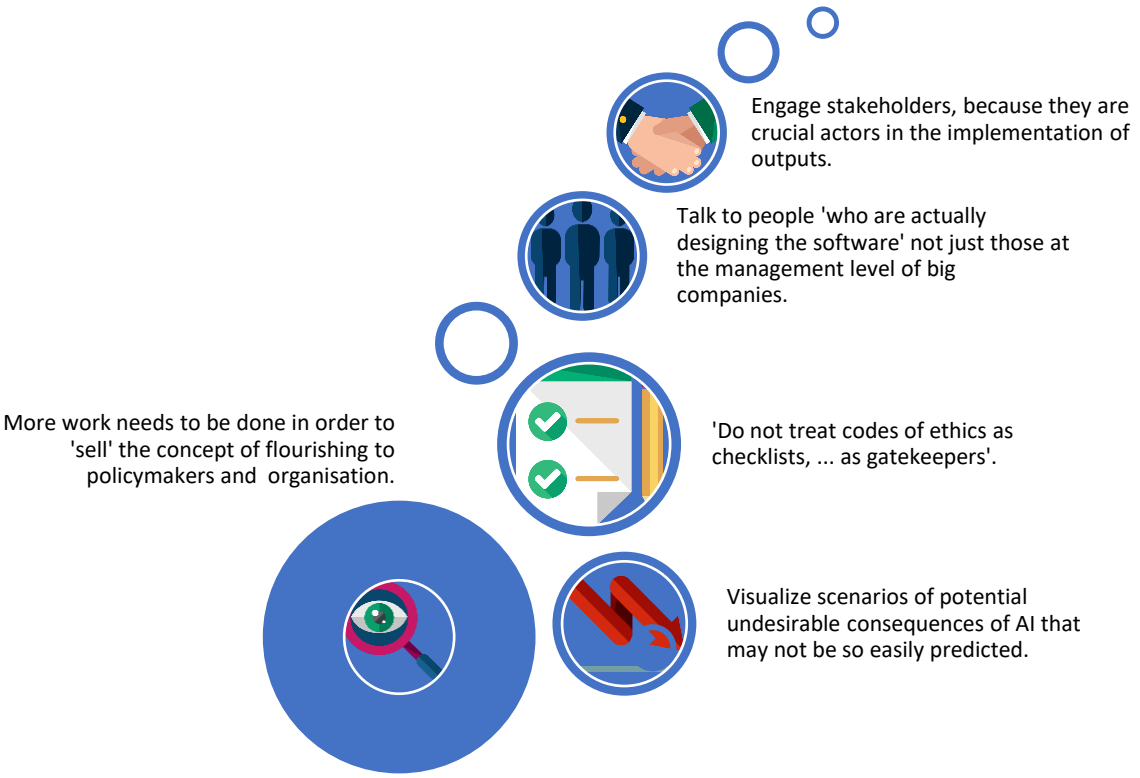
5.3. Future Suggestions for Dealing with Ethical Issues

This section presents an overview of important activities that were suggested by the participants in the FGs for dealing with the ethical issues of Big Data and AI. Due to the nature of offering suggestions to address a certain issue, participants' answers often had an intersection with discussions that involved 'challenges' or 'limitations' or identification of ethical issues. Therefore, it is advised that the readers of this section should also keep in mind the discussions outlined above (cross-references have also been made to guide the reader). As education was the most prominent suggestion across and within the FGs, it is presented in a more elaborate way. Although there was no specific question in the

[7] The ACM Code of Ethics 'is a collection of principles and guidelines designed to help computing professionals make ethically responsible decisions'.
https://www.eurekalert.org/pub_releases/2018-07/afcm-wlc071718.php

Exploratory FGs on the SHERPA project, some participants offered their suggestions and feedback for the project which are included in Figure 6.

Figure 6. Suggestions from participants regarding the SHERPA project



5.3.1. Raise public awareness and engagement on ethical issues related to AI

FG participants acknowledged the need to raise public awareness about how AI works and ethical issues related with AI. To achieve this objective participants suggested to have ethical discussions and debates in ways that are **accessible** to the public, such as media, TV, and global forums and debates, in addition to the discussion that should take place among experts in the academic sector and policy makers. The suggestion included the creation and encouragement of inclusive and multidisciplinary forums and debates, involving a wide cross-section of people, (with ongoing dialogue as technology is constantly evolving) to discuss ethical issues related with AI. The rationale behind this was that such debates enabled exploration of risks and consequences related SIS (SB Meeting). Users would become both more **engaged** and **empowered** in order to take responsibility and become more informed especially in terms of protecting their human rights. One participant called for a 'media enlightenment':

> 'a kind of new **media literacy** if you want to call it that, but one that helps highlight people are not stupid for the most part, that we are capable of becoming better informed and I've heard people indicate that they are curious enough about this, if it could somehow be explained to them in ways that are **accessible and intelligible**, then it's possible to **engage citizens** more fully than we have done before. So, some kind of active you could almost call it a **media**

> **enlightenment where we recognise that we as consumers have to take more responsibility for understanding the technology that we are using**' (ETHICOMP I).

Participants suggested more **bottom-up campaigns** that called for more ethical approaches to AI. Parallels were drawn with social movements for climate change; just as it was 'the voice of the people that pushed companies to do something that didn't immediately bring profit but could actually have customers like them more because they were environmental', then this could also take place in the context of AI. One participant suggested **fear** as a means for motivating the public, especially the younger generation, to be more proactive, again drawing parallels with environmental causes:

> 'when they [younger generation] post their private information and things like that, and when you explain to them you get the answer "I don't care. It's okay for people to see my information". And when you explain that it's dangerous, they say "It's fine, it's okay". So, the mentality of the young generation, it's very different. But then when it comes to environmental changes, we are going back to fear because a lot of fear was created about the environment... Because of the fear, a lot of people started being more conscious about it and started complaining, and then you started having **movements**. So, unless you **create a similar fear with ethics**, so that more people can be **motivated** to follow…' UCLan Cyprus II hybrid)

User **empowerment**, it was suggested, could happen through better access to information such as how data is collected, how long it is stored, what the implications of this are for the user, for instance for their long-term job prospects (UCLan Cyprus Exp). In addition, users should be more aware about information that may be inaccurate, biased or not objective, especially when it comes to interfering with election outcomes that have a direct implication on the way democracy in their country operates (UCLan Cyprus II hybrid). As we already saw in section 5.1.5., biases in algorithms that may lead to unfair treatment in **justice systems** may not be easy for the general public to comprehend, to visualise and protest against. So a recommendation is to provide more accessible ways to this information so that citizens have more human agency and the ability to appeal or challenge decisions about them. In other words, it is important for safeguards to be put in place that not only ensure algorithmic bias is identified and rectified (see section 5.1.10.) but that protect a citizen's right to due process and a fair trial if one is suspected of certain crimes (ETHICOMP I).

Related to empowerment and 'enlightenment' through better information was the perception amongst some participants that there was **intentional deception** (see section 5.1.4.) that was taking place against the public. Therefore, there were specific suggestions for users to be more aware of the **power dynamics and power asymmetries** involved that may be working against their human rights, and the fact that the 'virtual universe that we now live in…is much more hierarchical than most people believe' (DMU-UKAIS). The rationale behind this suggestion was that being more aware of the intentional deception taking place, would motivate the public to protest and resist it and work collectively to stop this deception.

5.3.2. Transparency and accountability of companies

Responsibility for better access to information and knowledge, was presented not only as belonging to the user but more importantly as that of governments and Big Tech. Participants argued that producers of AI and policy-makers have the moral responsibility to not only inform users of possible ethical issues but also of decisions that are affecting them. This was specifically related to the suggestion of having more **transparency** and **explainability**, informing the public on **how algorithms**

work and how decisions where taken and giving the users more **choice** over what data is used and how. It is important, they argued, to communicate complex aspects of algorithms in an accessible way that improves the **trust** of users (UCLan Cyprus Exp). Companies have a responsibility to present to the public 'the full data' and not just choose selectively what they think the user wants to see and present a skewed picture (UCLan Cyprus II hybrid) (see also section 5.1.3). In cases of change of data use, the users should be informed in a clear manner. One participant also suggested that there should be a specific activity for '**identifying the provenance of AI.** How do you get from an idea and some programming, through to something happening out there in the world?' (DMU-UKAIS). The rationale here was that tracing the origins of AI would enhance transparency and make more explicit the human agency involved.

Participants also stressed the need to be careful with the **language** used in the communication with the public in a way that does not present things in a more complicated way than they are, driving people away:

> 'We need to be just careful of how we use these terms, to make sure that we're being… We're not unwittingly giving the impression that it is, basically... 'Black box' sounds a bit like 'black magic', doesn't it? … It's not magic. It's just numbers organised in ways that we can understand if people look at them properly'(DMU-UKAIS).

FG participants also stressed the need to increase **accountability** of companies (see section 5.1.9), which would also act as a way of improving **responsible innovation**. Beyond the recommendations to enhance accountability via legislation and standardisation, which are discussed below, some specific recommendations were given in relation to how companies operate.

Firstly, one participant also recommended that there should be more **agency** given by companies to their **employees**; they should have more of a say in the ethical impacts of AI and what they are asked to produce:

> 'It's really important to, especially when a company is creating something using artificial intelligence that is going to have an impact on the people who are working there, I think it's very important to **foster a form of resistance, using dialogue or empowerment**. I don't know. It depends on… There are different ways, depending on the kind of organisation, but I think it's very important to foster agency within organisations' (DMU-UKAIS).

Secondly, some participants suggested the creation of a committee that will provide permission, check ethics and provide **ethics certification**. This certification could then be used as an **advertisement** of the company and thus act as an incentive for the companies to behave more responsibly. It was noted however, that even with having such certification, there should be a long-term mechanism to regularly check and for the public to be able to make complaints and report abuses (UCLan Cyprus II hybrid).

Thirdly, other participants suggested to have a dedicated person in each company to ensure that reminders and incentives for data protection become part of a company's culture. Participants spoke of the need to have 'somebody there to remind them of it', 'to push it', someone 'dedicated' to this role and 'mention it every time' especially when it came to project management (UCLan Cyprus II hybrid). Finally, some participants proposed creating fear of punishment, using for example fines (enforced through regulatory frameworks), as a way of forcing companies and developers to be transparent and also more accountable for their actions.

5.3.3. Further legislation and regulation

Further development, strengthening and enforcement of legislation and regulations, going beyond GDPR, was another suggestion proposed by several FG participants. They saw the need to ensure respect of the human rights of consumers and **enforce** this into **legislation** with liabilities and consequences if they are breached. This could 'protect' people from 'the tendency of these tech giants to simply gather as much data about us as possible' (ETHICOMP I). It was also suggested that the process of creating legislation ensures that issues related to AI get more public attention and in that way it is viewed as a **'public issue'** that needs to be regulated as part of public policy rather than as an individual concern:

> 'the only way that we could protect it [data collection and violations of human rights] is by bringing it out as **a public issue**, just like Coronavirus: as a public issue. Let the whole society and the world, be it Europe, be it United Kingdom, be it whatever, we have to speak about it as a matter of global thing that we need to **regulate**. Regulation means, pure and simple, passing **laws in parliaments** and **making sure that violators are treated like any other violators of human rights**. In a sense, it is **no longer a matter of an individual**, going forward. It is something that we have to treat it as suspect, and that suspect and matter has to be treated by **legislation** to make sure that nobody can abuse it. Obviously, by legislation, by the mere fact of legislation, it becomes… It gets under **spotlight'** (DMU-UKAIS).

One participant also identified the need for legislators to act fast while another recommended the provision of training to legislators of human rights in order to make them more aware of the intersection of SIS with possible abuses of these rights.

Participants spoke also of the need to have a **new ethics regulator** that should 'come from the EU and then be enforced locally' (UCLan Cyprus II hybrid) and that needed to keep up to date with fast technological developments (Eurec SB). There was a sense of **urgency** for the need to have a body that could enforce certain regulations 'because the issue of AI and Big Data is going to grow big in coming years' and there was a need to 'control data…before it goes in all directions' (DMU-UKAIS). They proposed the creation of an **independent authority** in order to avoid manipulation of state authorities and legislators by big companies (one participant referred to the case of Boeing manipulating the Federal Aviation Authority as an example of failed regulation (DMU-UKAIS)). Other relevant recommendations were, namely, the creation of an **Institutional Review Board or a Bioethics Committee,** while there was also a suggestion for a larger international body, such as an **organisation under the United Nations** which will seek to find commonalities and reach consensus among countries, as well as enforce protocols involved in data collection and use.

In a particular FG (Eurec SB) that had a more extensive discussion on **regulators** the following were suggested as important considerations:

- a new regulator should not be inhibitive; it should be enabling rather than disabling
- a new regulator could foster better understanding of the benefits and risks of AI
- some saw scope for further regulatory options and tools but not for a specific regulatory body
- potential regulatory conflicts between regulators should be avoided
- no single regulator could address widely varying regulatory demands between e.g. medical and military uses of AI
- be aware of potential exclusions of others as a result of national government regulations

In another FG, one other participant spoke of the need for regulation of AI to be used as a means of **quality control:** just like other things are retracted from the market due to malfunctions, so should AI,

which we also 'consume' in a way (UCLan Cyprus Exp). Another suggestion involved investing in **standardisation**. According to a participant this is a good way of bringing stakeholders together but also of resolving conflicts under the condition that power dynamics across actors are well balanced (ETHICOMP II).

Finally, one controversial recommendation was the creation of **'legal personhood'** to algorithms. The participant recommending this, suggested it could be done in the same way that companies have legal personhood:

> 'If you can start to assign liability to algorithms and how they operate, maybe that might be a possible framework or a possible macro-level term for integrating some of these issues, moving forward. As I said, then the analogy would be companies have legal personhood. Why would it necessarily be the case that an algorithm wouldn't have **legal personhood, which in turn would then need the algorithm has to have insurance associated with it and so on?** I would have thought it's one of the things that I think would be a big-picture idea. Obviously, the mechanics, that, would vary from country to country and jurisdiction to jurisdiction, but as a big-picture item I wonder: is that a potential way forward for looking at some of these issues?' (DMU-UKAIS)

One participant expressed clear disagreement to this suggestion, however. The **counter-argument** was that given that an algorithm 'is based on the intellect and thoughts of the people who have decided how to write it', the **responsibility** should go back to those who created it, who had the 'freedom' to make these decisions in the first place. Otherwise, this is 'again reducing **human agency**' and not making it clear to firstly, the industry that they cannot shift responsibility for decisions and mistakes on 'the machine' and secondly to the public that they have some responsibility to 'understand what's going on, through education' (DMU-UKAIS).

5.3.4. Education: who, what and where?

Education, in different forms and levels, was a prominent theme across the FGs. Some suggestions were on a more general level such as highlighting the need for education of children on the dangers and ethical issues of AI, while others gave specific examples of, for instance, which types of **competencies** should be developed. Table 5 presents the diverse suggestions that were given by participants in relation to the role of education in addressing ethical issues.

Recommendations for more education spanned across a range of actors and target groups: **developers and data scientists, school-age children and adolescents, adults from the general public, legislators of human rights, journalists as well as teachers and academics.** So it included both those on the receiving end of education (children), educators themselves, as well as policy-makers.

There was a particular emphasis on adult's responsibility to **educate children**: it was seen that education 'should start from a young age and the child should be educated about the dangers hidden behind AI, Big Data and ethics' (UCLan Cyprus Exp). This was, it was argued, especially significant given firstly, the **vulnerable** age of children and secondly, the fact that nowadays they are **constantly being exposed** to technology. A third reason given was this education of children meant raising a generation that would be able to **use and develop technology more responsible and ethically in the future**:

> 'education should include a **responsible** online presence in covering all of these aspects, including… You know, they will be the future users, consumers and the generator of new technology. So, if they have the ethical aspect and responsible use of the technology itself in

> their mind since the very beginning, I think they will be better developers of technology and of course consumers' (UCLan Cyprus II hybrid),

In line with 5.3.1. education of the **end user** who is an ordinary member of the public (with no background in technology or SIS) was also a very prominent recommendation. It was suggested that the user should be able to have a full understanding of the kind of consent that she/he gives and the way his/her data are collected and used. The user 'should have the means to be educated properly' noted one participant while another talked about the importance of 'user empowerment through education)' (UCLan Cyprus Exp). In this context, the space for learning was referred to as **'adult education'**, and the elderly were particularly mentioned by a couple of participants. It was argued that they are the ones without previous experience compared to other adults and so more 'ad hoc' educational initiatives were recommended in order for them to become familiar with concepts such as user privacy, trust and security (UCLan Cyprus Exp).

We should go beyond regulations, suggested one participant and urged for **a new way of approaching ethical issues** and to 'think out of the box', referring to a fundamental **revision of the education system** both in and outside the school setting: 'they should become **embedded** into our lives like other topics in the curriculum. Like you learn Maths and History' beginning from the home, where ethical issues in AI should be discussed with children from an early age (UCLan Cyprus Exp). Beyond the recommendation of integrating ethics in **formal school curricula**, it was also suggested that parents have a responsibility to home school their children on these topics, in other words, education should also take place in the space of **non-formal learning**. In addition, one participant pointed out that 'the school cannot work on its own without the parents' and unless the parents are educated themselves, then they will not be able to pass this information to their children. Therefore, **education of parents** can also be seen as one recommendation for policy-makers to consider.

Education at higher education institutions, such as **universities**, was another recommendation, with the suggestion being that ethics should be embedded across all relevant topics taught such as modules on AI, modules on Big Data etc. One participant emphasised the importance of doing so especially 'for **computer science** related topics…[i]n terms of project and core courses of ethics and artificial intelligence' (ETHICOMP I). The academic sector was seen as having a responsibility to integrate ethical and legal matters related to AI and Big Data in the **education of students**. Moreover, it had a task to teach students themselves how to be reflective and behave responsibly. One lecturer reflected on his own teaching practices, making recommendations to other colleagues to follow this path as well:

> 'I have a few suggestions so me as an educator and lecturer, I try to enforce these kind of reflections on my students especially when I'm teaching artificial intelligence that they are **responsible** for what they write as in code, in the code sense. So, I think that is something that I particularly do, and I tried to talk to my peers and my colleagues that they should be doing the same kind of **reflection** as well in academia in general' (ETHICOMP I).

One interesting suggestion was given by two male participants who disagreed with the importance given to universities as spaces for education: 'Education doesn't happen in universities. It happens through **soap operas'** one of them noted (DMU-UKAIS). Another participant agreed and argued that the best way to educate people was through **stories** (story-telling) that would be more accessible to the public than 'regulations, or principles, or academic papers'.

Another space and medium for education was identified as being education via the **media**. It was argued that **educational campaigns** should be initiated from **government** or other **organisations** 'for

informing the public about new technologies and the extent to which we should trust them' (UCLan Cyprus Exp).

In terms of what this education should consist of, participants were limited to mentioning certain skills and competencies such as: **critical thinking, digital competencies, media literacy** and other necessary 'tools' or 'techniques' to be able to cross-check and verify information especially in the current context of misinformation and fake news:

> 'the kids need to have awareness and **critical thinking so they can decide on their own what to trust and how much to believe**. I mean you can take the information, you can filter, you can cross-check. So, they can learn other techniques just to **verify** that the information they receive is right' (UCLan Cyprus II hybrid).

One other suggestion was to have more education on the social-psychological aspects of online behavior which differ in terms of dynamics from the 'traditional' face-to-face encounters. The emphasis here was on more education on ethical issues related to problematic online behaviours such as cyber-bullying and trolling (UCLan Cyprus Exp).

It was also mentioned that users should be able to have a better understanding of the decision-making process that takes place in SIS and a better appreciation of the kind of data collection that they are consenting to when they are online.

Table 5. Suggestions on the role of education in addressing ethical issues

Education:
• of developers and of companies
• of the public
• children and students at schools: critical thinking and digital competencies; ethics should be included in the formal curriculum in schools and embedded across all topics
• research in education of children so that they can use technology more responsibly and ethically in the future
• home education and schooling on these topics
• education and regulation go hand in hand
• education at universities: including students and academics to be more reflective and responsible
• adult and life-long learning
• education of SMEs and the creative sector, for instance, through the use of sandboxes where regulators give innovators the chance to see how a particular algorithm is functioning
• embed AI ethics into training of data scientists and lawyers
• media literacy and education in an accessible way (read balanced press and watch documentaries)
• education through story-telling (narratives and soap operas)
• it is the responsibility of the user to an extent to become educated
• education has limits as the public is driven by trends, peers and the 'need to fit in'

A specific recommendation involved **further research** not only on how to educate children in ethical issues related to AI and Big Data, but also research that **measures 'how educated children are' on 'how to use technology'** in the context of ethical issues in Big Data and AI (UCLan Cyprus Exp). This participant also suggested that **comparative research** in relevant educational policy matters could be useful, so that **good practices** can be identified in some countries that are ahead in terms of their children having the necessary competencies, and adopted from other countries. In this way there should be 'a dialogue between countries, discussing the education policies that are implemented' (UCLan Cyprus Exp). Another recommendation related to research and education was on the need to investigate more how adults, especially **the elderly**, could be properly educated in order for them to realise the ethical issues of SIS that they may be interacting with (UCLan Cyprus Exp).

Some participants were eager to point out however that education is not a panacea; it also has certain limits and challenges (see also section 5.2.4). The public is often driven by emotions, trends, peers, the need to fit in and sometimes even access to knowledge is not enough to change behaviours. This was, it was argued, especially the case with teenagers:

> 'You know when you are fifteen-years-old and everybody is using TikTok, and no matter what people are telling you on how to be careful, you are still going to use it **because everybody does**. It's the same with Facebook…systems need to be foolproof because the users are always going to be **driven by their feelings and the need to fit in** and stuff like that' (UCLan Cyprus II hybrid).

Finally, it was noted that education could not work on its own without regulation but also vice versa, regulations could not work properly if education was missing. In other words, **education and regulation** were seen as two fundamental recommendations that went hand in hand (UCLan Cyprus Exp).

5.3.5. Keep in mind vulnerable groups, ensure fairness and avoid discriminations

Participants stressed the need to ensure that SIS are available to all and that everyone benefits, ensuring **equality** and **fairness**. They suggested placing special emphasis on the protection of **vulnerable groups** (e.g. children) and most at risk groups (e.g. black people) in society and how they could be discriminated against (see also section 5.1.5 and 5.1.6). As one participant put it, it is very important 'to review who is most vulnerable and who is most at risk, rather than take a universal approach' (DMU-UKAIS) when addressing ethical issues of AI. 'The people that are shouting the loudest about this may not be the ones that are really most at risk' emphasised a participant who made a call to 'remember the vulnerable' (DMU-UKAIS) and to ensure that they are 'protected' (SB Meeting). **Human rights** should be protected (see also section 5.1.10) and viewed not just as an issue affecting individuals but a societal, collective issue, participants argued.

5.3.6. Researchers: Involve different stakeholders in research projects, be concise and proactive

The involvement and engagement of stakeholders in research projects was another suggestion proposed in FGs. The rationale provided for this suggestion was that **stakeholders are crucial actors in the implementation of outputs and the success of such projects** is partly dependent on stakeholders' acceptance of these outputs, therefore stakeholders should be part of research projects.

One participant suggested in particular, to conduct an **empirical study** of all AI regulation of the European Parliament and analyse how stakeholders feed into the policymaking process (SB Meeting). There was also a recommendation for academics to collaborate with governments and act as advisors to them if they have this opportunity (ETHICOMP I). This would ensure governments have better access to their expertise and knowledge.

Another participant advised that research projects, like SHERPA, should not just talk to the management of big companies but also engage with 'the people who are actually designing the software' and discuss 'issues that they have to contend with'. Researchers, the participant argued, should refrain from labelling companies or developers as 'untrustworthy' or 'a bad person' and 'I'm not going to buy your product' but instead try to reach those who are making decisions and help them to **recognise** the ethical issues involved so as to have more responsible research and innovation (ETHICOMP II).

Participants also encouraged scientists to appreciate the importance of language used when discussing ethical issues and ensure that it is **accessible and clear.** This also included research projects that are guiding designers to make ethically sustainable designs – one participant noted how there is an inherent tension between ethical issues and concrete guidance and acknowledged how difficult it is to give clear, concrete guidance on ethical issues, but suggested that this is something researchers should pay more attention to (ETHICOMP I).

Finally, a proposition was that researchers should be proactive and not just react to negative situations but try to adequately prepare for them by looking at 'scenarios of unintended consequences' (DMU-UKAIS). **Modelling of various scenarios** with unanticipated or unintended consequences as well as better risk assessment could mean that the world is better prepared to deal with these problems once they arise.

5.3.7. Adopt a cross-disciplinary and global perspective

There was a suggestion by several participants to adopt a **pluralistic, inclusive and cross-disciplinary** approach that takes into account diverse political, professional, cultural and economic contexts. Participants proposed to encourage discourse that breaks 'silos of conversation' (ETHICOMP II) to have a broader understanding of the issues at stake and avoid the current 'niche' society that focuses only on the mathematical, automated, technical aspects that developers are concerned with on the one hand, and having predominantly white male people making decisions on the other (see also section 5.2.4).

Participants also proposed trying to reach a (global) consensus on what exactly constitutes a breach of human rights in the context of AI and Big Data. Related to this was the proposal for having **global collaborations** with various stakeholders like companies and policymakers in order to ensure that data is stored in a lawful, fair and ethical way (DMU-UKAIS). AI, argued one participant, which is an even 'bigger issue' than conventional technologies, needs to perform better than the latter in terms of having a truly global collaboration amongst all stakeholders: 'we really need a **global collaboration**, not just one party or a few parties working on it' (DMU-UKAIS). Projects, it was argued, should therefore run across 'different countries and cultures' (ETHICOMP I) and not just have a national approach.

5.3.8. General (theoretical) suggestions on how to approach ethical issues of SIS

Participants sometimes gave more general suggestions on how they thought ethical issues should be approached. One participant – who acknowledged that this sounded 'very idealistic' but nevertheless 'an important activity' – proposed to engage in a **better and deeper reflection** on the ethical issues through an imagination exercise: imagining another way of innovating and how 'a better democratic control over innovation' could possibly look like (ETHICOMP II). The same participant urged for a better understanding of the **social and environmental impacts** of Big Data and AI and to imagine 'that we don't have to innovate and to automate on everything'. This was consistent with one other participant who suggested that we should learn from the Boeing 787 Max scandal and 'perhaps, to seek more **simple technology**. 'We get to a complexity that is unnecessary', they noted (DMU-UKAIS).

There was also a call from one participant for policy discussions to focus more on **human flourishing** and deontological terms. One other participant strongly recommended to avoid superficial '**checklist approaches'** (see also 'Limitations' section above) and focus more on explaining why values are important for improving society. This could be done by firstly, ensuring that new efforts do not try to reinvent the wheel or write ethical standards without looking at previous efforts in other areas of ethics and learn from their mistakes. Secondly, mechanical, superficial ways of approaching ethical issues should be avoided and instead programmers should not be considered with 'checking the box' but rather really deeply reflecting on the ethical implications of their actions (ETHICOMP II). So there should be more an emphasis on *why* certain **values** are important and *why* programmers should care and deal with these aspects rather on creating 'ethical codes' as such. This is also consistent with a more **constructive, positive approach**, the participant argued, that focused on opportunities and responsibilities for improving society rather than using fear of punishment as a tool for inhibiting unethical behaviour.

6. Conclusion

Ten main themes regarding ethical issues emerged from the analysis of the Exploratory FGs. The ten main ethical issues identified by the diverse stakeholders included in the FGs were: loss of autonomy and human agency; loss of privacy due to data collection, monitoring and surveillance; Big Tech manipulating users for financial or political gain; lack of sufficient information given to, and knowledge of, users; inaccurate data and algorithmic bias; loss of human jobs; loss of access to services when denying data collection; loss of trust; lack of accountability as well as human rights violations.

The findings from the Exploratory FGs suggest that there is still a long way to go for policy-makers wanting to improve the way ethical issues related to AI and Big Data are addressed. Although participants acknowledged that there are already regulations, legislation as well as improvements and initiatives related, for instance, to transparency and data protection, the overwhelming tone of the findings was one that looked at the glass 'half empty' instead of 'half full'. In other words, it was not a matter of *if* changes are needed, but *how* these changes would materialise and *how much* needed to be done. Participants argued that unless changes are done fast, collaboratively in an inclusive, multidisciplinary, global way, the implications for humanity could be disastrous.

Participants also identified numerous challenges that emerged when trying to address ethical issues such as the rise of nationalism, populism, hate speech and weak democracies or the sheer diversity of perspectives on ethical issues depending on a variety of factors such as gender, age, occupation etc.

Limitations of existing efforts in terms of both specific legislation as well as weaknesses of general approaches to deal with ethical issues were also presented. Finally, the report offers eight suggestions that emerged from the analysis of the FGs, on how to better address ethical issues in the future. Suggestions included raising public awareness and engagement, increasing companies' transparency and accountability as well as extensive educational reforms that would increase the provision of education across different target groups, ages, and levels of society.

Several conflicts, tensions and dilemmas can be seen in the arguments made, some of which revealed unsolved inherent tensions and the complexity of the situation - which perhaps explains why relatively little progress has been made regarding ethical issues, when compared to how fast technology has advanced in the world of SIS. It is important for researchers and policy-makers to be aware of these conflicts and dilemmas when carving out future recommendations and policies related to SIS. Firstly, there was on the one hand, the need to act fast, to act urgently, before automated systems got out of control. There was a strong call to protect humans, especially the vulnerable, from discrimination, bias and human rights violations associated with algorithmic bias. On the other hand, on many occasions, the recommendations offered, such as long-life learning, inclusion of ethical issues in school curricula, a global entity that could act as a regulator etc., require care, caution and extensive periods of time. In many cases, the participants also acknowledged that we don't know yet *how* to achieve some objectives, such as promoting values, and further research is required. So we have urgency vs. activities that are time-consuming, especially as they are treading novel areas and a lot of 'trial and error' will be involved before getting things to an optimum scenario. Moreover, in a lot of ethical issues identified, there was an identification of the problem, a suggested solution, but the lack of specific expertise on how exactly this solution would materialise in practice.

A second tension is related to the innate nature of ethical issues. Ethical issues involve how humans approach moral values and principles, which govern human activities and behaviour. These are not always clear-cut objective aspects, it is not always simple to distinguish between what is 'right' from 'wrong' and what is acceptable to one may be entirely incompatible with another's moral compass. When this was applied to ethical issues related to SIS, there was on the one hand the need to have clear, concrete guidance offered to designers but on the other hand an appreciation of the difficulty of giving particular directions to designers when there is still a lack of consensus as to 'how far' ethical concerns could or should drive policy-makers. This was where the need for a more global consensus on what exactly constitutes ethical and unethical practices, on what exactly counts as responsible and irresponsible innovation became evident. Part of the complexity was also traced back to the different approach required, for instance for a health application versus an application used for military purposes. Policy-makers, could, hence have different approaches for different areas and purposes of SIS.

Thirdly, one tension that requires careful consideration by policy-makers is related to the 'convenience versus privacy/security' dilemma. In a fast-paced world, in one where knowledge was both difficult to access but also difficult to understand, as well as time-consuming, participants noted how users were sometimes naively assuming that ethical behaviour of companies or individuals is the norm failing to see issues of fraud, manipulation or surveillance. Or, they were, for the interests of practicality, choosing to ignore issues that they may later end up regretting, related long-term data collection that could jeopardise their privacy, safety as well as security.

Fourthly, and perhaps more crucially, there was a strong tension that was raised by participants between ethical aspects and financial interests. This was not just related to an inevitable tension between, for instance market needs and customer needs, or what is realistically to be expected from a company that ultimately needs to make profit to survive. More significantly, this was presented as

a ‘Catch 22’, given that companies would not themselves try to alter the type of behaviour that formed the ‘whole modus operandi’ and financial success of their company to begin with. Specifically, this was linked to the lack of motivation for companies to be transparent about data harvesting related to manipulative behaviour that nudged people in particular directions as this was how they were essentially making money.

One interesting and surprising result is the sheer extent of negative language that is used to describe the activities of companies, engineers, developers etc. This is prevalent across the FGs and can be clearly seen when looking at some of the quotes from the participants. It may not be too far-stretched to argue that the language used was not far from that used when describing dynamics prevalent in a state of dictatorship: loss of autonomy, loss of choice and free will, loss of control, loss of power, lack of knowledge, inequalities, discrimination, bias, human rights abuses, intended deception and so on, were all words used to describe the status quo of ethical issues and SIS. Regardless of whether this is accurate or not, the choice of language by itself denotes an alarming situation, a discursive gap that policy-makers may consider how to constructively bridge. There is an important voice of stakeholders ringing the alarm regarding a context which they argue is one of ‘ethics washing’, of ‘surveillance capitalism’ of a ‘top-down culture’ and this should not be taken lightly. Further research could perhaps also be done in terms of why exactly there was this quite critical approach, delving more specifically into individual experiences. A future deliverable of the SHERPA project (Deliverable D2.2.) involves an analysis of the findings from exploratory interviews that seek to address the individual aspect more as participants during one-to-one interviews have an opportunity to express their personal experiences and opinions in more detail than when part of a larger group.

Indeed, perhaps the strong general consensus over the negative dynamics of the situation especially when referring to Big Tech – the claim that they are intentionally deceiving people, that they are manipulating not just behaviours of the users but also leveraging their political power to successfully policy-makers and legislators – emerges from the FGs particularly because of the nature of the method used. Focus Groups are known for providing a dynamic forum where individuals can form a sort of solidarity as well as be empowered to speak out about sensitive issues (Pierce, 2008).

At the same time, here lies another strength of FGs: the fact that it involves more than one person means that there is an ability to also showcase disagreement. Although disagreement emerged relatively rarely in the FGs, there were some instances that provided a clear illustration of differing views. For example, as detailed in the empirical sections of the report, some had differing evaluations of the usefulness of legislation (leading to lengthy battles that disproportionately favour the big companies), a couple of participants were reluctant to the dominant consensus of using more accountability and fear of punishment and liability as a constructive strategy, or there was disagreement with the suggestion of giving a type of ‘legal personhood’ to algorithms.

The FGs also clearly illustrated the importance of having a common language, for people to mean the same things when they use certain words. Going beyond the use of clear and accessible language which was a prevalent finding, the point here refers to the use of the same word by participants to denote different things. For example, the term ‘transparency’ was used by some to describe having access to information collected by the government. For others, transparency meant companies being explicit about, data collection, data storage and use, the way they derived their algorithms, the intended purpose of their app, or the business model that they were using. Effective communication is something that researchers and policy-makers may want to consider when articulating recommendations, legislation, guidelines etc.

In terms of the specific findings and recommendations that resulted from the analysis of the Guidelines FGs and the Regulatory FGs, these can be divided into two main types. Firstly, a set of general recommendation relating to the SHERPA Guidelines documents and, secondly, relating to suggestions for consideration in the Terms of Reference for the EU Regulator.

Regarding the SHERPA Guidelines documents, these are clear comprehensive documents about development and use of SIS-related technologies, but could additionally focus on their potential to address issues of trust and bias more specifically. In addition, the documents could be used as a set of guidelines for standardisation of practice, or even as a set of guidelines that could be used for educational purposes. Education seems to be a main recommendation evident in all types of FGs.

Regarding the Regulatory FGs, the analysis recommends that Terms of Reference for an EU Regulator should consider approaching regulation of such SIS-related technologies using a smart-mixing approach, i.e. by considering technical and legal instruments as well as ethical and social standards. The importance of considering the variety of stakeholders, who should be involved, is highlighted. Moreover, it is recommended that action is propagated to the national levels, that existing regulation should be taken advantage of and that there should be continuous monitoring of the technology, the funding opportunities but also the application of the proposed guidelines.

Finally, perhaps the most novel aspect of the FGs was an unintended one: the timing of the FGs coincided with the 2020 pandemic onset and this provides a rich pool of information regarding stakeholders' views related to a global health issue and SIS. Extensive, rich and diverse discussions related to the pandemic emerged, with participants recognising that it presented an unprecedented ethical challenge for policy-makers. The pandemic gave additional weight and urgency, in particular to health-related matters related to privacy. The majority of the comments aimed at pointing out the negative implications of the pandemic for users, especially in the long-term. Discussions were related to whether it is ethically justifiable or not in search for a solution to the pandemic and dealing with it to give up autonomy, privacy and data concerns. Some for example pointed out that data collection and tracking through applications should concern us in terms of the long-term implications, the potential stigma and stereotyping that users would experience, while others emphasised the way the algorithms helped spread fear, panic, misinformation and fake news during the pandemic.

To conclude with, the fact that different stakeholders mostly expressed concerns that AI violates human rights, rather than making comments on how AI promotes human rights and wellbeing, is a matter of concern. Undoubtedly, substantial measures need to be taken to effectively deal with and prevent exploitation of AI and Big Data that ultimately manipulates human beings. These measures need to be both reactive i.e. to mitigate and reverse the consequences of existing practices as well as proactive i.e. to prevent such practices from happening in the first place, in particular through educating individuals on how to develop and use AI and Big Data for promoting individual and societal wellbeing.

7. References

Braun, Virginia and Victoria Clarke (2006) Using thematic analysis in psychology, *Qualitative Research in Psychology*, 3:2, 77-101

Bryman, Alan (2008) *Social Research Methods*, 3rd edition, Oxford: Oxford University Press

Charmaz, Kathy (2004) 'Grounded Theory', in M.S. Lewis-Beck, A. Bryman, and T.F.Liao (eds.), *The Sage Encyclopedia of Social Science Research Methods*, Thousand Oaks, California: Sage

Pierce, Roger (2008) *Research Methods in Politics: A Practical Guide*, London: SAGE Publications

8. Appendices

REPUBLIC OF CYPRUS

CYPRUS NATIONAL BIOETHICS COMMITTEE

Ref.: ΕΕΒΚ ΕΠ 2018.01.108
Tel: 22809038/039
Fax: 22353878

June 28th, 2018

Dr Kalypso Iordanous
Associate Professor
UCLan Cyprus
University Avenue 12-14
Pyla
7080 Larnaca

Dear Dr Iordanous,

Application for bioethical review for the research entitled:
«Shaping the ethical dimension of information technologies – a European Perspective (SHERPA)»

The Cyprus National Bioethics Committee (CNBC) has reviewed your application for ethical approval for the project outlined above submitted on the 21st of June 2018. From the review of the documents you have submitted, your research proposal is deemed to meet the requirements of the Law Providing for the Establishment and Function of the National Bioethics Committee (No. 150 (I) / 2001 -2010) and does not necessitate a full bioethical review from the CNBC.

2. Kindly note that approval is granted provided that the following conditions apply:

a) conduct of the research is strictly in accordance with the proposal submitted and granted ethics approval, including any amendments made to the proposal required by the CNBC,
b) inform CNBC immediately of any complaints or other issues in relating to the project which may warrant review of the ethical approval of the project,
c) before implementing any amendments to the proposal as approved, request a new approval by CNBC,
d) provide a follow up report on the progress of the program every 6 months from the approval date,
e) provide a final report upon completion of the program,
f) inform us in writing in case the project is discontinued.

.../2

Appendix B

SHERPA - Shaping the ethical dimensions of smart information systems (SIS) – a European perspective

Task 4.2 – Stakeholder evaluation and validation

Information Sheet

Please take some time to read this information and ask questions if anything is unclear.

Contact details can be found at the end of this document.

What is the purpose of this study?

The SHERPA project investigates, analyses and synthesises our understanding of the ways in which smart information systems (SIS; the combination of artificial intelligence and big data analytics) impact ethics and human rights issues. The project aims to develop novel ways of understanding and addressing SIS challenges. The focus groups aim to explore stakeholders' views regarding the recommendations that have been developed in the project thus far, with the objective to improve those recommendations.

Who is organising this research?

The research for this study is being undertaken by the EU-funded SHERPA project (SHERPA is the acronym for 'Shaping the ethical dimensions of information technologies – a European perspective' (https://www.project-sherpa.eu)

A Research Ethics Committee has reviewed and approved this research.

Why have I been chosen?

The project aims to conduct 2 waves of 5 focus groups each with several stakeholder categories – e.g., representatives from policy, science, industry, civil society, politics, media, academia etc. The aim of the 1st wave of interviews is to explore initial reactions form stakeholders regarding the overall set of recommendations. Focus group participants will be asked to use an action plan to take the recommendations back to their constituents and collect broader feedback.

Do I have to take part?

Participation in this study is voluntary and you may ask any questions before agreeing to participate. If you agree to participate, you will be asked to sign a consent form. However, at any time, you are free to withdraw from the study and if you choose to withdraw, we will not ask you to give any reasons.

What will happen to me if I take part?

If you agree to take part in this study you will participate in a focus group in person, with other 9 stakeholders where you will discuss the recommendations that have been developed from the SHERPA consortium regarding the development and use of SIS[KI1] . The time and place that the focus groups will take place will be determined in coordination with all the participants in order to find the most convenient time and place for all attendees. You will be asked to use an action plan to take the recommendations back to your organization and collect broader feedback. We will be asked to participate in a second focus-group meeting, though participation in this is optional. During the second round of focus groups, you will be asked to provide specific suggestions concerning the formulation and implementation of the recommendations. [KI2]

What are the possible benefits of participating?

The study aims to develop proposals for the responsible use of SIS. In addition to helping the SHERPA project, advanced analysis will be carried out on the stakeholders' focus groups discussions to which you contribute, which may raise issues that your organisation would like to know about and take steps to remedy.

What are the possible risks of taking part?

There are no risks in taking part in this study. At any time during the interview you can choose to withdraw. You may also choose to withdraw your data from being used in the project at any time until 1st July 2020.

How will my interview/focus group data be used?

The focus group will combine quantitative and qualitative elements and will be designed and analysed by SHERPA project partners. The recording of the focus groups may be transcribed by parties outside of the consortium. If this happens, the transcription company will delete the recording and transcription after the transcription is approved. On the consent form we will ask you to confirm that you are happy for the SHERPA consortium to use and quote from your interview. Any such use will be anonymous unless you indicate otherwise on the consent form. Information which will identify your organisation will also be kept out of publications unless otherwise indicated on the consent form.

What will happen to the results of the project?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be identified in any reports or publications and your name and other personal information will be anonymised unless you indicate otherwise on the consent form.

What happens to the focus group data collected during the study?

The focus groups discussion will be transcribed by the interviewers or a designated, approved third-party agency. If we use a third-party transcription service, we will ensure that there is a signed data processing agreement in place. The audio files will be deleted, once the analysis of the focus groups data is complete.

What happens at the end of the project?

You may request a summary of the research findings by contacting Kalypso Iordanou, University of Central Lancashire Cyprus (KIordanou@uclan.ac.uk).

What about use of the data in future research?

If you agree to participate in this project, the research may be used by other researchers and regulatory authorities for future research. The transcript will be kept for five years after the publication of the findings of the study.

Who is funding the research?

This research is funded by the European Commission under grant no. 786641.

What should I do if I have any concerns or complaints?

If you have any concerns about the project, please speak to the researcher, who should acknowledge your concerns within ten (10) working days and give you an indication of how your concern will be addressed. If you remain unhappy or wish to make a formal complaint, please contact Dr Kalypso Iordanou, KIordanou@uclan.ac.uk.

Fair Processing Statement

The information collected will be processed in accordance with the provisions of the EU *General Data Protection Regulation* (*GDPR*)

Appendix C

Project SHERPA – Consent form

Issue	**Respondent's initials**
I have read the information presented in the information letter	
I have had the opportunity to ask any questions related to this study, and received satisfactory answers to my questions, and any additional details I wanted.	
I am also aware that excerpts from the focus group meeting may be included in publications to come from this research. Quotations will be kept anonymous unless I give specific permission to the contrary (below).	
I give permission for my name to be associated with excerpts from the focus group which may be included in publications to come from this research.	
I give permission for my organisation to be identified in any final publications produced by SHERPA.	
I give permission for the focus group to be recorded using audio recording equipment (if necessary).	
I understand that relevant sections of the data collected during the study may be looked at by individuals from or a project partner from SHERPA. I give permission for these individuals to have access to my responses.	
I understand that the audio recording may be given to a transcription service company to transcribe. I give permission for these organisations to have access to my audio files for transcription purposes.	

With full knowledge of all foregoing, I agree to participate in this study.

I agree to being contacted again by the researchers if my responses give rise to interesting findings or cross references.

□ No □ Yes

If yes, my preferred method of being contacted is:

□ Telephone: ……..

□ Email: ……

□ Other: ……..

Participant Name		Consent taken by	
Participant Signature		Signature	
Date		Date	

Appendix D

Questions for Guidelines Focus Groups

1. You have now read two guidelines, one for use and one for development. Although these guidelines often overlap (e.g., because we sometimes want to protect end-users by requiring developers to adapt their systems), they are supposed to provide different guidance when appropriate. Reflecting on that, do you see any reasons for revisions?

2. The guidelines are supposed to be easy for practitioners to read, understand and apply. Do you see any need for adjustments because of a risk of misunderstanding, conflations, or ambiguous language, either because the guidance is not clear enough or because it includes too much jargon?

3. The guidelines are supposed to be engaging, which is always a problem for a relatively long documents of instructions. How would you judge the guidelines with respect to engagement?

4. What is your impression of the use of graphics (tables, figures, pictures) in the document? Should any changes be made, if so, in what way and why?

5. If you have experience with many other similar documents, how do you compare these guidelines to other guidelines with respect to: understandability, engagement, and usefulness?

Questions on specific parts:

6. What is your evaluation of the "Introduction"?
    1. Does it cover what it needs to cover? Is anything missing?
    2. Does it give a good introduction to the guidelines?
    3. Is it engaging?
    4. Are the language and length appropriate?
    5. Does your impression vary between the two guidelines?

7. What is your evaluation of the "High-level requirement section"?
    1. Does the section make an important contribution to the rest of the guidelines?
    2. Is the language appropriate (understandable, no jargon, engaging)?
    3. Are the different high-level requirements and their sub-requirements sufficiently well explained/motivated?
    4. Are the language and length appropriate?
    5. Should something be removed or added?
    6. Does your impression vary between the two guidelines?

8. What is your evaluation of section 3 (i.e., models/methods for development/governance)?
    1. Is it well-adapted for practitioners?
    2. Is it suitable for your organization?
    3. Does it contribute to the overall guidelines?

4. Is the language appropriate (understandable, no jargon, engaging)?
5. Is it too long or too short?
6. Should something be removed or added?
7. Does your impression vary between the two guidelines?

9. What is your overall evaluation of section 4 (the ethical operational requirement)? For each sub-section:
    1. Is it well-adapted for practitioners?
    2. Is the language appropriate (understandable, no jargon, engaging)?
    3. Can it be properly applied?
    4. Is it too long or too short?
    5. Is there something that needs to be changed?
    6. Are there important issues not covered?
    7. Do the guidelines require something they should not require?
    8. Are the proposals linked to the correct phases of development/management and governance?
    9. Does your impression vary between the two guidelines?

10. What is your evaluation of section 5 (special topics)?
    1. Is it well-adapted for practitioners?
    2. Does it contribute to the overall guidelines?
    3. Is the language appropriate (understandable, no jargon, engaging)?
    4. Is it too long or too short?
    5. Should something be removed or added?
    6. Does your impression vary between the two guidelines?

Appendix E

Focus group Questions on Regulatory options (T3.3.)

Goal: Solicit feedback on selected regulatory proposals for AI and big data

Target attendees: policy makers, civil society, legal scholars.

Input: D3.3/policy brief/exec summary of D3.3/extract of relevant info

Questions for discussion:

1. What are three high-risk, high-human rights impact AI/big data fields and/or applications that could benefit from stricter regulation?

2. Of the following international options, which three do you find most promising? Why?

    - Moratorium on lethal autonomous weapons systems
    - Binding Framework Convention for AI
    - Legislative framework for independent and effective oversight
    - Legal for human rights impact assessments on AI systems
    - Convention on human rights in the robot age
    - CEPEJ European Ethical Charter
    - International Artificial Intelligence Organization
    - Global legal AI and/or robotics observatory

3. Of the following EU-level options, which three do you find most promising? Why?
    - EU-level special list of robot rights
    - Adoption of common Union definitions
    - Creating electronic personhood status for autonomous systems
    - Establishment of a comprehensive Union system of registration of advanced robots
    - General fund for all smart autonomous robots

- Mandatory consumer protection impact assessment
- EU Taskforce of field specific regulators for AI/big data
- Algorithmic Impact Assessments under the GDPR
- Voluntary/mandatory certification of algorithmic decision systems

4. Of the following national options, which three do you find most promising? Why?
    - DEEP FAKES Accountability Act (US)
    - Algorithmic Accountability Act (US)
    - Canadian Directive on Automated Decision-Making
    - US Food and Drug Administration regulation of adaptive AI/ML technology
    - New statutory duty of care for online harms
    - Redress by design mechanisms for AI
    - Register of algorithms used in government
    - Digital Authority (UK)
    - Independent cross-sector advisory body (CDEI)
    - FDA for algorithms (US)
    - US Federal Trade Commission to regulate robotics

5. Of the following cross-over options, which one do you find most promising? Why?
    - Using anti-trust regulations to break up big tech and appoint regulators
    - Three-level obligatory impact assessments for new technologies
    - Regulatory sandboxes

6. What immediate regulatory actions are necessitated at the:
    1. International level

2. EU-level
3. National level

7. Should there be an international ban on the development/use of lethal autonomous weapons systems?

8. How can we strike a balance between enabling beneficial AI and risk mitigation? What will support this?

9. One key recommendation for AI and big data regulation emerging from SHERPA results is "smart mixing for good results" – is this feasible? If yes, how can this be achieved? Smart mixing refers to using a good combination of instruments, i.e., technical, standards, law and ethical that will offer complementarity, agility and flexibility needed to address the challenges of AI.

10. How can the law further support super-secure AI where it has high likelihood and high severity of risk and impact on rights and freedoms of individuals, especially vulnerable populations – children, minorities and the elderly?

11. What critical future developments need consideration in discussions/actions on the regulation of AI and big data?

Should we also consider given the current developments:

Should there be a ban on the use of facial recognition in public places? Why?

Appendix F

Focus group questions on Terms of reference for new/bespoke regulator (T3.6)

Prepared by TRI

Objective: explore the feasibility of a bespoke new regulator for AI and big data at the EU and/or Member State levels and what its terms of reference should include
Target audience: experts including regulators, policy-makers and other relevant stakeholders. *Please ensure equal male/female ratios.*
Location: online/F2F
Dates: TBD

Background:
Should there be a new/bespoke regulator for AI and big data at the EU and/or national levels are questions the SHERPA project is currently deliberating. There are pulls and pushes to the creation of new regulators/regulatory bodies at the international, EU and national level. At the EU-level, the European Parliament request to the European Commission to consider the designation of a European Agency for robotics and artificial intelligence to provide the technical, ethical and regulatory expertise needed to support the relevant public actors, at Union and Member State level, in their efforts to ensure a timely, ethical and well-informed response to the new opportunities and challenges, in particular those of a cross-border nature, arising from technological developments in robotics, such as in the transport sector was not taken up by the European Commission. At the national level, new bodies have been created in countries such as the UK (e.g., the Centre for Data Ethics and Innovation which is tasked with connecting policymakers, industry, civil society, and the public to develop the right governance regime for data-driven technologies) are in the process of being set up (Regulatory Horizons Council to co-ordinate policy and regulation in areas of rapid technological advances in the UK) or proposed (e.g., an FDA for algorithms, calls in Netherlands for a national algorithm watchdog, Digital authority to co-ordinate regulators in the digital world).
SHERPA invites your inputs and feedback.

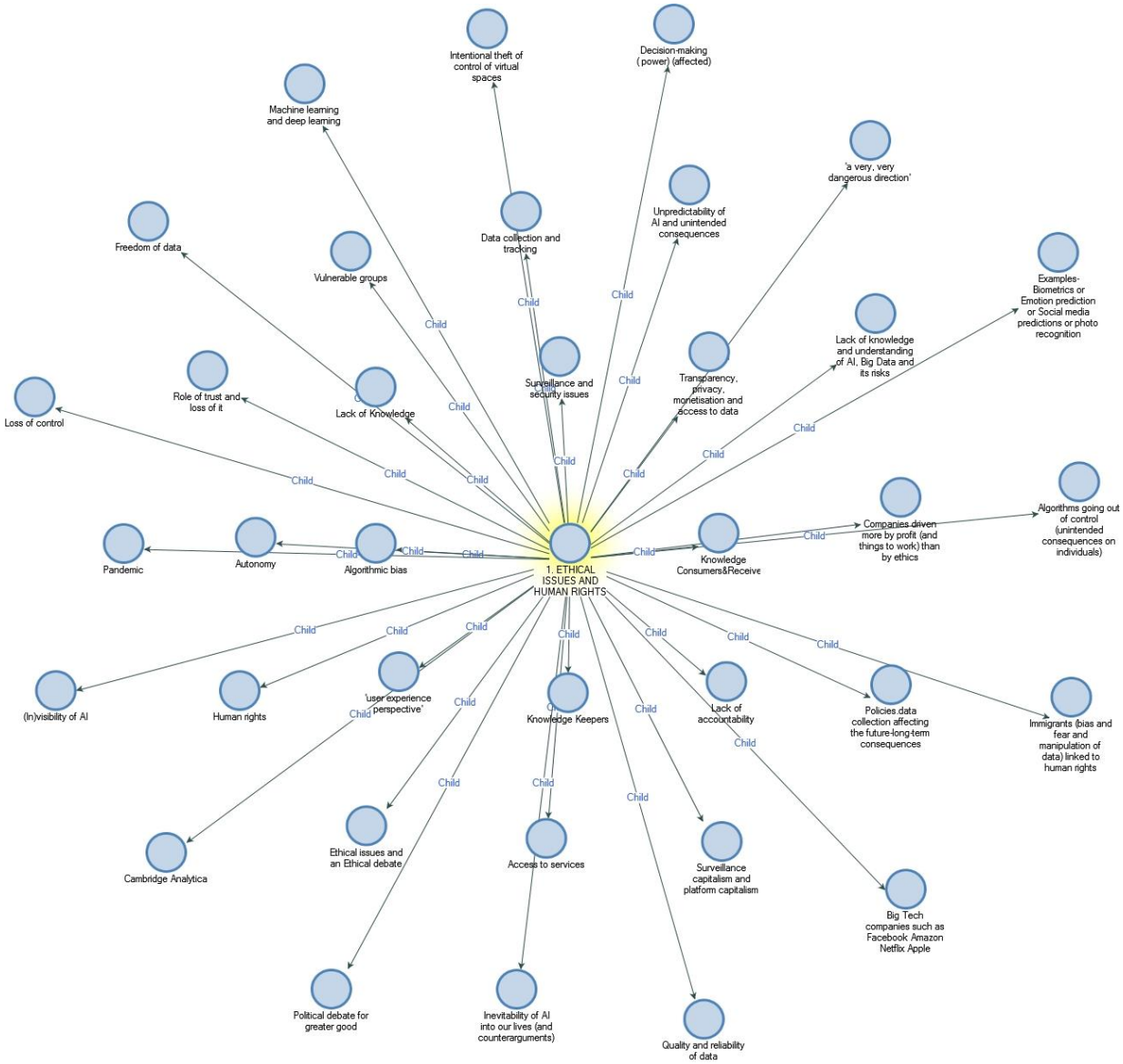
Questions for discussion (moderator to adapt and use)*:*

1. Do we need a new/bespoke regulator for AI and big data at the EU level? (yes, why; no, why; undecided)
2. Do we need a new regulator for AI and big data at the Member State level? (yes, why; no, why; undecided)
3. Why do you think a new regulator might be necessary? What gap would it address?
4. What type of regulator should this be ? (field-specific/general? Independent watchdog? Licensing body/authority? Inspectorate? Public sector/private sector/general? Professional regulator? Professional conduct authority? An EU regulators network? Supervisory agency? Statutory registration board; Commissioner; AI and big data standards agency; AI fundamental right protection agency? EU/national task force/Digital Authority)
5. What would/should it regulate? E.g., use of autonomous weapons? Human rights? Algorithms ? use/implementation?
6. What would be its legal basis?
7. What should its functions and tasks be?
8. What powers should it have?
9. What would be its role and responsibilities?
10. How should it be constituted? Who should its members be?
11. What should its conduct provisions be?

12. How would it operate? Discuss operation and procedural rules.
13. How would it be governed? How would it be funded? To whom would it report? e.g., the European Parliament?
14. How often should its terms of reference be reviewed?
15. What would be some challenges and barriers to its success? (creation and implementation – political will? regulatory creep/mission drift? Funding? Capacity, lack of independence, lack of teeth, competing priorities and conflicts; regulatory capture)
16. How could these be overcome?
17. Any other comments/considerations that need to be taken into account.

Appendix G

Visualisation of child codes of ethical Issues related to Big Data and AI (section 5.1.)



Visualisation of codes of suggestions in relation to education of the public (section 5.3.4)