SHERPA

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

Delphi Study – Round 1 Results

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

Introduction

This report presents the highlights of the first round of a Delphi Study undertaken by the EU project SHERPA. SHERPA investigates ethical and human rights issues in smart information systems, the combination of artificial intelligence and big data. The first round of the Delphi study consisted of a set of open questions that aimed to determine expert opinions on the following points:

- What do you think are the three most important ethical or human rights issues raised by AI and / or big data?
- Which current approaches, methods, or tools for addressing these issues are you aware of?
- What do you think are the pros and cons of these current approaches, methods, or tools?
- What would you propose to address such issues better?
- Which should be the top 3 criteria for society to select and prioritise the most appropriate measures?

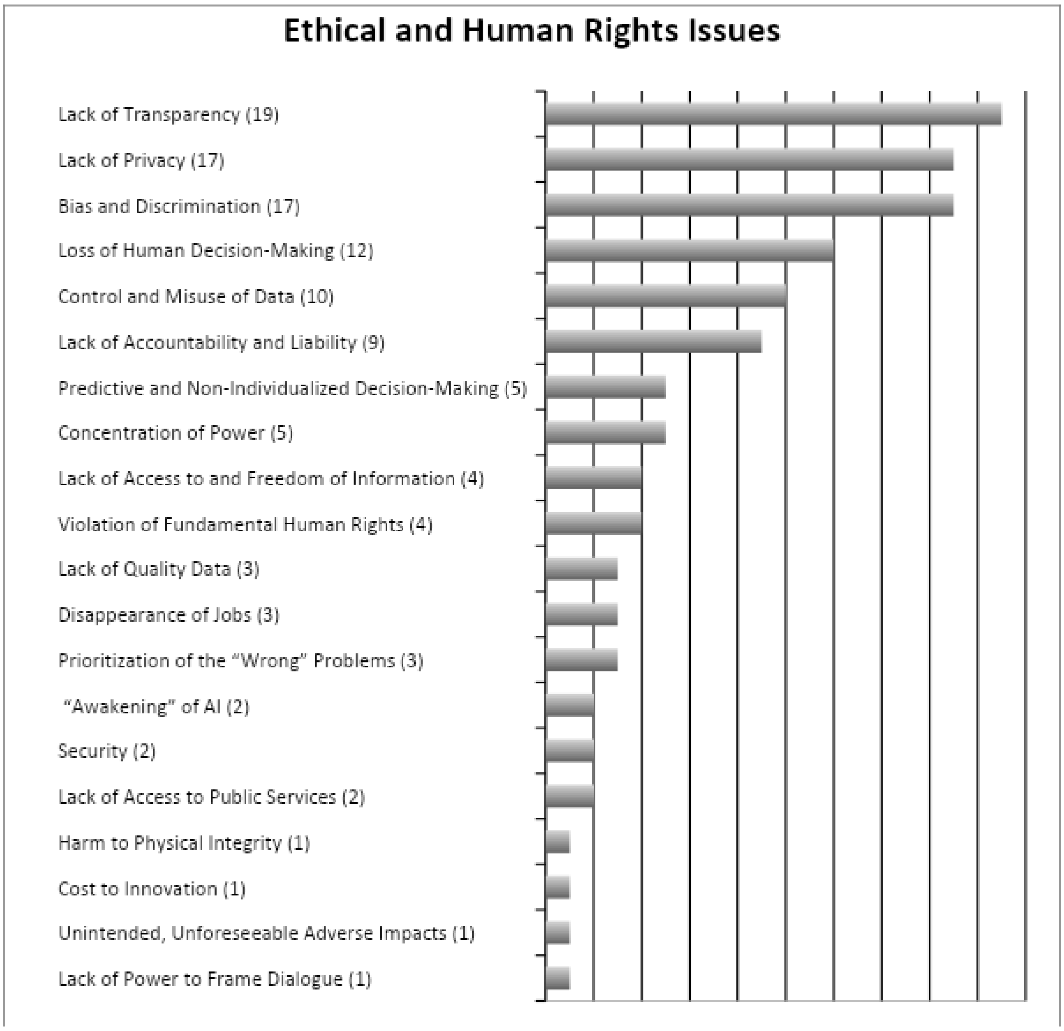
The survey was designed by the SHERPA consortium and, following initial review by consortium members, pilot tested by 27 consortium members and 38 external advisers. Following ethics approval by De Montfort University, the survey was sent out to 231 experts, 50% of who were women. We received 145 responses. Following review of the data and data cleansing, 41 responses contained sufficient information to warrant analysis. Responses were analysed within the question that was most relevant (e.g. response with *proposals* for regulatory measures was analysed in Q4, not where provided in Q2).

This report highlights key findings from each of the questions. It then summarises key insights and findings, including some interesting omissions. It then provides an outlook to the next round of the Delphi Study.

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

Summary of Responses

Question 1: What do you think are the three most important ethical or human rights issues raised by AI and / or big data?



There were 41 responses to Q1. One Q1 response was deemed more relevant to another question, and two responses to other questions were deemed more relevant to Q1. Therefore, a total of 42 responses were analysed under Q1.

Lack of transparency, lack of privacy, bias and discrimination, and loss of human decision-making were the most frequently mentioned concerns.

Lack of **transparency** was identified as a concern in that the average citizen does not understand how AI and data systems work, nor do they understand how decisions are made by SIS that affect them in their daily lives; a need for transparency (and explainability) about the source of data and the decision-making processes was clearly expressed. When discussing **privacy** concerns, respondents tended to focus on the vast amounts of personal data collected, specifically raising

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

concern about real-time surveillance. With **bias and discrimination**, respondents were concerned with built-in and entrenching bias caused by AI systems that reproduce bias, both of which may produce unfair and/or unequal decisions that violate the right to equality. Concerns about the impact of new technologies on humanity and the **value of human decision-making** were articulated in various ways, but most common was anxiety that humans were being taking "out of the loop" on critical decision-making as machine-intelligence is privileged, resulting in 'depersonalized' decisions and a perceived loss in human intellect.

Concerns about a lack of accountability and the misuse of personal data were also cited by a number of respondents. In regards to **accountability**, respondents called for a clear definition of legal responsibility for all actors, including AI systems. When discussing the **misuse of personal data**, respondents expressed specific concerns about abuse (e.g. mass surveillance), control, ownership, and commercialisation of data.

The remaining responses were each mentioned by only a couple respondents. Particularly notable was that **harm to physical integrity** was only mentioned once in relation to self-driving cars and autonomous weapons.

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

Question 2: Which current approaches, methods, or tools for addressing these issues are you aware of?

Current Measures	
Regulatory Measures	● Regulations (18) ● Public Register of Permissions to Use Data (1) ● Reporting Guidelines (1) ● Monitoring Mechanism (2)
Technical Measures	● Testing Algorithms on Diverse Subsets (1) ● Using Analytics Systems to Judge Whether Decisions Are Equal/Fair (1) ● Generative Adversarial Networks and Other Techniques for Deriving Explanations from Outcomes (1) ● More Open Data (2)
Other Measures	● Codes of Conduct (3) ● Education Campaigns (4) ● Employing 'Fairness' Officer or Ethics Board (3) ● Frameworks, Guidelines, and Toolkits (14) ● Grievance Mechanism (1) ● High-Level Expert Groups (6) ● Individual Action (2) ● International Framework (3) ● Investigative Journalism (3) ● NGO Coalitions (1) ● Open Letters (1) ● Public Policy Commitment (1) ● Self-Regulation (1) ● Stakeholder Dialogue and Scrutiny (3) ● Standardisation (3) ● Third-Party Testing and External Audits (2)

There were 36 responses to Q2. Two responses were deemed more relevant to another question. Therefore, a total of 34 responses were analysed under Q2.

There were very few responses identifying current approaches, methods, or tools at the **international level**. No respondent identified an international law instrument, and some noted the practical limitations of creating and implementing an international approach.

When identifying approaches, methods and tools at the **regional level**, all examples cited referred to the European Union, and most frequently to the GDPR.

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

At the **national government level**, the majority of responses referred to measures in Western Europe; only three responses concerned the United States and one concerned Hong Kong. National laws were the most frequently cited, but other specific examples cited included national policies and frameworks, and national education campaigns.

There was a greater variety of measures referenced that were developed by industry, NGOs, and civil society (including academia). A number of specific initiatives were included that had been created both by private-sector actors alone (e.g. Google) and in partnership with other stakeholders (e.g. Partnership for AI). It is worth noting in Q3 that there were no critiques of industry-driven initiatives like company codes of ethics or toolkits.

From **NGOs and civil society** (which includes academia), specific measures cited included educational tools, ethical guidance and frameworks, NGO coalitions, and an open letter signed by famous AI scientists and experts. One respondent cited a report by ETH Zurich[1] that found there are 84 projects and organisations working on AI issues, suggesting that there is a proliferation of frameworks, potentially leading to further confusion.

Lastly, some respondents mentioned the role of **journalists** to investigate and highlight concerns, and the role that **individuals** assume to protect themselves (e.g. disabling ads on personal devices).

[1] Anna Jobin, 'Ethics guidelines galore for AI – so now what?', ETH Zürich, 17 January 2020, https://ethz.ch/en/news-and-events/eth-news/news/2020/01/ethics-guidelines-galore-for-ai.html

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

Question 3: What do you think are the pros and cons of these current approaches, methods, or tools?

Pros

- Dialogue means we **learn from each other**
- Regulation has **power of enforcement**
- Transparency measures means **building ethics into the design**
- Education **enhances citizen/consumer power**
- Ethical Impact Assessments provide **clear methodology & tools**
- Standardisation has **objective set of criteria**
- Oversight **addresses human rights violations**

Cons

- **Lack of understanding** about roles & responsibilities
- **Risk of shifting burden of responsibility** to developers or consumers
- Measures are **too abstract**
- Creation & implementation is **resource intensive**
- Non-binding measures have **no enforcement**
- **No comprehensive approach**
- **Too complicated** to implement new ways of thinking
- Regulation has **limited application**
- Technology **development outpaces rule-making process**
- Measures **perceived as a hurdle**
- Measures are **public-sector focused**
- **Difficult to measure ethics objectively**
- Educational campaigns ineffective because **don't reach people who need it most**

There were 31 responses to Q3. Three responses to Q3 were deemed more relevant to another question. Therefore, a total of 28 responses were analysed under Q3.

There were far more cons mentioned than pros.

The 'pros' focused only on specific types of current measures; for example, one 'pro' of regulation cited was the power of enforcement. Other 'pros' mentioned referred to stakeholder dialogue, transparency efforts, ethical impact assessments, standardisation, and oversight mechanisms.

In contrast, nearly half of respondents identified at least one 'con' of existing measures; there were both general critiques and critiques specific to individual types of measures. A common general critique was that **key players do not understand their responsibilities,** and therefore do not appreciate the potential impact of their work. One respondent refused to put all the blame on developers, calling out an "apathetic set of consumers." Two other notable critiques were that current measures are **too abstract** to be effective and **resource intensive** to create and implement. This was one of the only three times that costs were mentioned by respondents.

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

In addition to general ‘cons’, respondents also evaluated the limitations of specific current measures. Regulations were the most frequently mentioned, with critiques ranging from their **limited scope of application** to the fear that they **hamper innovation** or **contribute to compliance-only setting**. Multiple respondents also noted that disconnect between the rapid development of new technologies and the **slow speed of policy-making processes**. Other ‘cons’ included **long and overly complex guidance** and **hard to measure objectives**.

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

Question 4: What would you propose to address such issues better?

Proposed Measures	
Regulatory Measures	● Regulations (13) ● Public Register of Permissions to Use Data (1) ● Reporting Guidelines (1) ● Monitoring Mechanism (2)
Technical Measures	● More Open Data (1) ● Use of AI to Protect Data (1) ● Improve Control of Data (1) ● Easily-Explained Algorithms (1) ● Comprehensive AI Example Sets (1) ● Retaining Possibility of Human Override (1)
Other Measures	● Citizen Juries (1) ● Codes of Conduct (1) ● Education Campaigns (11) ● Employing 'Fairness' Officer or Ethics Board (2) ● Ethical Mindset (1) ● Exchange of Best Practices (1) ● Frameworks, Guidelines, and Toolkits (2) ● Grievance Mechanism (1) ● High-Level Expert Groups (1) ● Individual Action (1) ● International Framework (3) ● More Open Source Tools (1) ● Retaining 'Unsmart' Products and Services (1) ● Stakeholder Dialogue and Scrutiny (5) ● Standardisation (1) ● Third-Party Testing and External Audits (2)

There were 30 responses to Q4. Four responses to other questions were deemed more relevant to Q4. Therefore, a total of 34 responses were analysed under Q4.

Regulatory measures were the most frequently proposed, with regulations being the most common. There were not any general themes for regulation, as each respondent proposed something unique (e.g. 'smart mix' of regulatory initiatives; legislation for transparent AI; and recognition of right to work).

International or regional agreements were mentioned only a few times, which could be seen as either realistic given the difficulty of creating such agreements, or an unfortunate reflection that

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

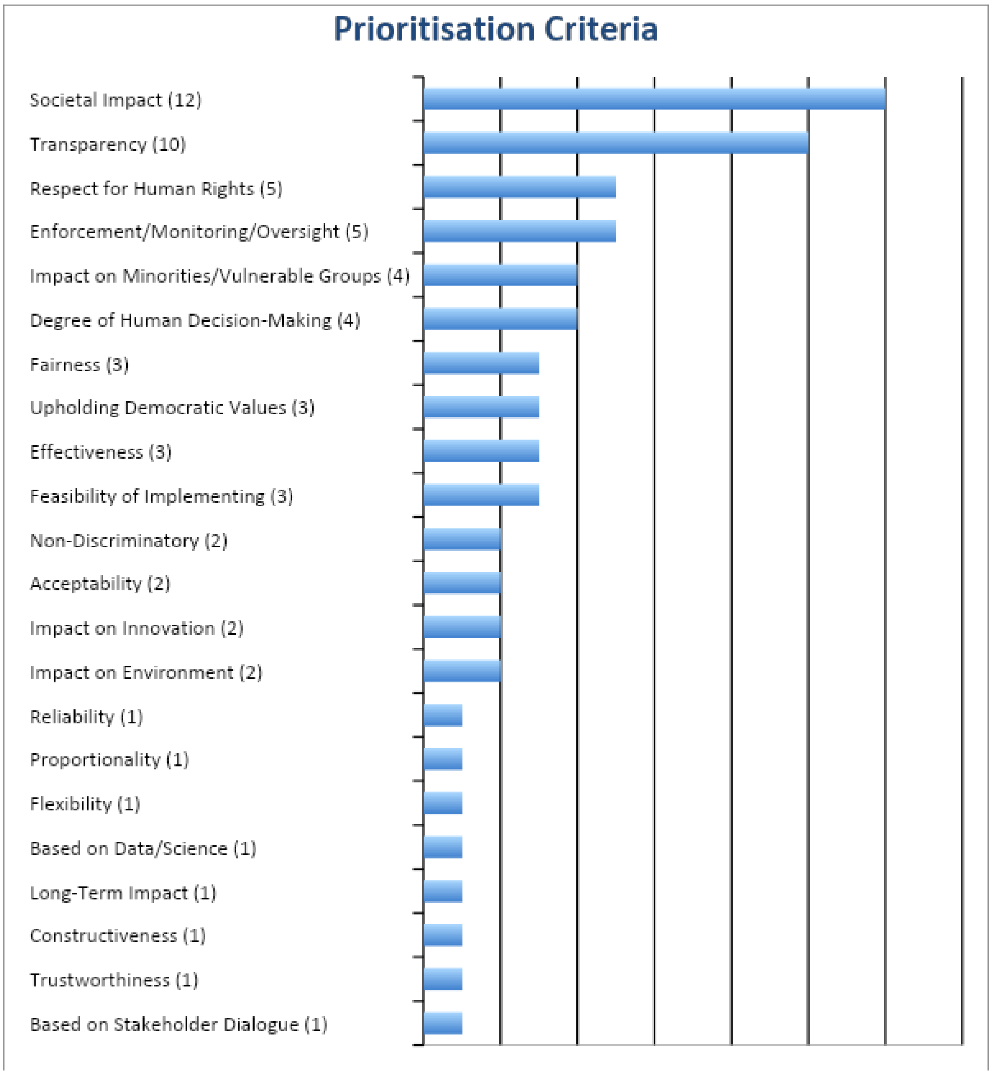
international agreement is extremely unlikely. These responses were, however, consistent with Q2 and Q3, which focused on regulatory measures at the regional and national level.

Additionally, respondents proposed a broad range of other measures, including technical measures, encouraging collaboration among stakeholders, developing differentiated toolkits, and implementing third-party auditing. One respondent proposed creating ‘citizen juries,’ which is a novel idea that could be a means to encourage stakeholder dialogue.

Many respondents also proposed developing educational and awareness campaigns for all stakeholders at all levels, including children, students, developers and professionals, politicians and government officials, and members of the public generally.

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

Question 5: What should be the top 3 criteria for society to select and prioritise the most appropriate measures?



There were 31 responses to Q5.

About half of respondents identified criteria that should guide the development of new technologies. While not a direct response to the question, the responses provide valuable insight into the types of issues and concerns that should be prioritised when developing and implementing appropriate measures. For example, measures could be developed in such a way that the two most frequently mentioned issues – societal impact and transparency – are addressed. These issues cited were consistent with the concerns raised in Q1 about lack of transparency and loss of human decision-making.

More traditional criteria for evaluating measures – like costs, feasibility, and effectiveness – were mentioned only a few times. In fact, monetary cost was only mentioned once, and the time to develop and implement a measure was not mentioned at all. Additionally, the oft-cited concern that regulation would stifle innovation was only mentioned twice.

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

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

General Insights

The most prominent issues running throughout responses to all questions were concerns about a **lack of transparency and human decision-making**. These concerns were articulated strongly in Q1 (most important ethical or human rights issues) and Q5 (criteria for developing new measures), and were common underlying themes in responses to other questions. While **privacy** and **discrimination** were top-ranking concerns in Q1, they were not mentioned often in responses to other questions.

Regulation was the most frequently discussed 'approach, methods, or tool,' both in terms of criticisms (Q3) and potential solutions (Q4). Given that most respondents are Europe-based, most of the specific examples cited, including the GDPR, were in Europe. Most of the proposed regulatory solutions focused on the regional and national level; international solutions were rarely mentioned in a positive light, suggesting that the respondents **do not view an international approach as the most effective**.

There was also a lot of discussion of other approaches, methods and tools. **Ethical frameworks and toolkits in particular came up frequently**, but there seemed to be some tension between those who found them useful and those who believe they create a confusing hurdle. Perhaps this is due to the fact that there is already a number of projects and organisations that have put forward guidelines, frameworks, and toolkits.

The cost of developing governing measures was rarely mentioned. A few respondents referenced the resources needed to develop governing measures, but only one respondent specifically cited the financial costs and no respondent discussed the time needed to develop and implement new measures.

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

Open Questions and Next Steps

There were a number of notable omissions in the Round 1 responses.

The focus of responses was on existing measures in Europe, though that is far from the only place these important conversations are taking place. Future research may reveal valuable insights and examples of on-going initiatives in other regions. This would be particularly critical to the success of any potential measures developed at the international level or in partnership with countries outside the EU.

In regards to ethical and human rights concerns, respondents were primarily focused on the immediate issues impacting end-users in Europe. However, as a result, a number of related ethical and human rights issues were not referenced. For example, there was no discussion of the ethical and human rights abuses suffered by those extracting the resources and manufacturing the devices that enable SIS technologies to function; the long-term impact to physical and physiological health from using SIS devices and technologies; or the environmental impact of the manufacture, storage, and disposal of the devices that enable SIS.

Only one respondent discussed concerns related to physical integrity, despite the potential injury (or death) that could be caused by technologies like self-driving cars and autonomous weapons. Additionally, security, reliability, and trustworthiness scored very low. A potential explanation might be respondents are not currently as likely to suffer direct physical harm, and the concerns are therefore less immediate. Regardless, the possible impacts of these technologies on our physical bodies are not only relevant, but potentially severe.

In regards to current and proposed measures, there was a lack of reference to existing human rights law and mechanisms. There are many on-going discussions at the international level concerning whether and how to apply existing human rights frameworks to emerging technologies (see, e.g. UN Human Rights Council and Office of the High Commissioner for Human Rights). Those discussions could inform the development of mechanisms at the regional and national level, particularly in countries that have existing international human rights legal obligations.

Lastly, there was no specific mention of the creation of a new regulator or regulatory body. While respondents did mention the need for oversight and monitoring bodies, responses were generally vague about the structure and responsibilities of those bodies.

Although not specifically mentioned in the responses to Round 1, these omitted issues and proposals have been included in Round 2 as a means to gauge whether the omissions were intentional.

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