



SHERPA

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

Delphi Study – Round 2 Results

03 September 2020

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



Introduction

This summary presents the key insights of the second round of a Delphi study undertaken by the EU-funded project SHERPA, which is investigating the ethical and human rights issues in smart information systems (SIS) (the combination of artificial intelligence (AI) and big data).

The first round (R1) of this Delphi study (October 2019-January 2020) consisted of five open questions to elicit participant views on **ethical and human rights issues** related to SIS, the **current and emerging approaches** to address these issues, and **criteria for prioritisation** of the most appropriate measures. The SHERPA team then analysed all responses and synthesised them into a summary report circulated to participants for comments and feedback. That report provided the key starting point for the design of questions in round 2 (R2). The purpose of R2 (March-June 2020) was to prioritise issues and mitigation measures identified during R1 and other complementary work undertaken in SHERPA. R2 consisted of four sets of questions, asking participants to rate issues and potential measures across three criteria on a scale of 1 (low) to 5 (high).

- Question 1 asked respondents to rate a list of **ethical and human rights issues** in terms of reach, significance, and attention.
- Question 2 asked respondents to rate a list of **potential regulatory measures** in terms of desirability, feasibility, and probability.
- Question 3 asked respondents to rate a list of **potential technical measures** in terms of desirability, feasibility, and probability.
- Question 4 asked respondents to rate a list of **other potential measures** in terms of desirability, feasibility, and probability.

Twenty-six responses contained sufficient information to warrant analysis.

Key Insights

Ethical and Human Rights Issues

In R1, the most prominent issues were a *lack of transparency*, *lack of human decision-making*, *lack of privacy*, and *bias and discrimination*. In R2, *lack of privacy* and *bias and discrimination* continued to be issues of high concern, along with *misuse of personal data*, *lack of access to (and limitations on) freedom of information*, and *impacts on democracy*. Other key concerns (not in the top five) included *violation of human rights for end-users*, *loss of freedom and individual autonomy*, *impacts on power relations* (political and economic), *lack of transparency and trust*, *potential for criminal and malicious use*, and *disappearance of jobs*. While not cited in R1, the *environmental impact of SIS* was also among the key concerns in R2.

Lack of human decision-making and *human contact* were not key concerns, scoring mid-low to low across the criteria. As in R1, *harm to physical integrity* was also a low-rated concern; it scored much lower in reach and significance than attention. *“Awakening” of AI* also scored high in attention but was the lowest in both reach and significance.



Two issues that scored lower than anticipated were *unintended, unforeseeable adverse impacts* and *lack of accountability and liability*. Moving forward with the prioritisation of governance measures, both issues seem to be of critical importance, particularly if determining high-risk applications for SIS.

Top Five Issues	Bottom Five Issues
<ul style="list-style-type: none"> • Misuse of Personal Data (4.05) • Lack of Privacy (3.96) • Lack of Access to and Freedom of Information (3.85) • Bias and Discrimination (3.80) • Impact on Democracy (3.80) 	<ul style="list-style-type: none"> • “Awakening” of AI (3.01) • Violation of Fundamental Human Rights in Supply-Chain (3.00) • Integrity (2.99) • Cost to Innovation (2.89) • Prioritization of the “Wrong” Problems (2.81)

Table 1: Top five and bottom five scoring potential regulatory measures based on overall scores.

Potential Governance Measures

The governance of SIS requires a smart mix of instruments that will address ethical and human rights concerns. To better understand how different types of measures (regulatory, technical, and other) should be prioritised, all 52 potential governance measures were analysed together.

Average Ratings of Potential Measures				
	Desirability	Feasibility	Probability	Overall
Highest	Technical Measures (4.34)	Other Measures (3.85)	Other Measures (3.49)	Other Measures (3.82)
	Other Measures (4.12)	Technical Measures (3.56)	Technical Measures (3.29)	Technical Measures (3.73)
Lowest	Regulatory Measures (3.70)	Regulatory Measures (3.31)	Regulatory Measures (2.79)	Regulatory Measures (3.26)

Table 2: Average ratings of potential measures (regulatory, technical and other) on desirability, feasibility, probability, and overall, from highest to lowest.

Regulatory measures were the lowest scoring in all criteria and overall, with no regulatory measures making into the top fifteen measures overall. More promising were **technical measures**, which were the most desirable on average. However, only three technical measures were in the top fifteen measures overall. Most promising were **other measures**, which scored highest on feasibility, probability and overall, and which were twelve of the top fifteen measures overall.



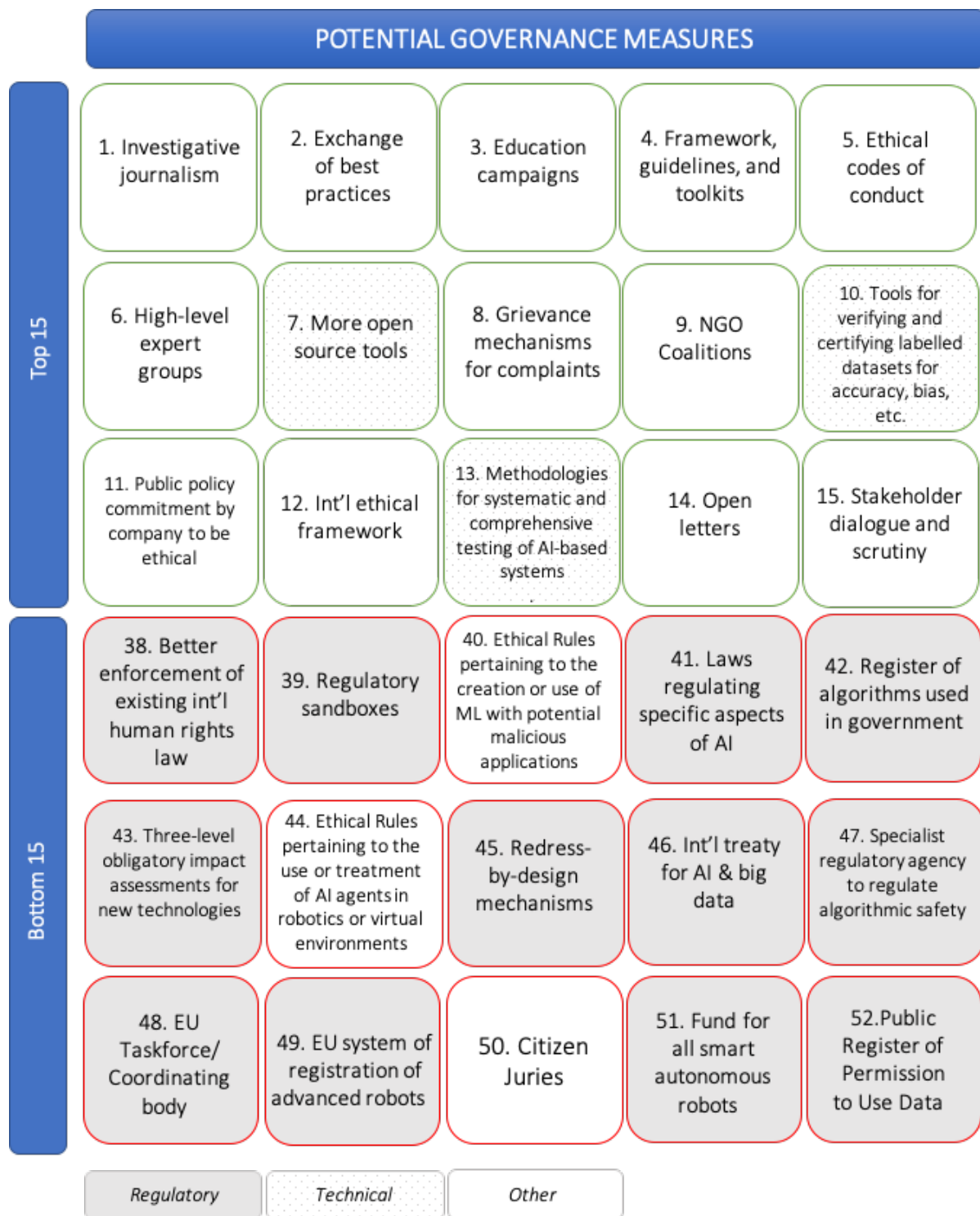


Figure 1: Top fifteen and bottom fifteen potential governance measures (from complete list of 52 potential measures) based on average overall scores (average of desirability, feasibility, and probability scores).



Regulatory Measures

In R1, regulation was the most frequently cited example of a possible ‘approach, method, or tool’ to address the ethical and human rights concerns associated with SIS. However, in R2, most potential regulatory measures scored low, both in absolute terms and relative to other types of potential measures. No regulatory measure was in the top fifteen measures, and twelve regulatory measures were in the bottom fifteen potential measures. This was because potential regulatory measures received the lowest average scores in all three criteria and overall. For the overall scores, no regulatory measures scored in the very high (4.5-5) or high (4-4.49) range. All of the top five regulatory measures (below) scored in the mid-high (3.5-3.99) range, which was lower than the top scoring technical and other measures. More significantly, potential regulatory measures had the highest percentage of measures scoring in the mid-low (3-3.49) to low (2-2.99) range for all three criteria. This was particularly true of probability, where 95% of measures scored low. Within potential regulatory measures, the majority (16 of 18) were rated more desirable than feasible or probable.

Top Five Regulatory Measures	Bottom Five Regulatory Measures
<ul style="list-style-type: none"> • Legislative framework for independent and effective oversight of human rights compliance (3.70) • Algorithmic impact assessments (3.65) • National independent cross-sector advisory body (3.59) • Binding Framework Convention (3.51) • Reporting Guidelines (3.50) 	<ul style="list-style-type: none"> • Specialist regulatory agency to regulate algorithmic safety (3.07) • EU Taskforce/Coordinating (3.06) • EU system of registration of advanced robots (2.85) • Funds for all smart autonomous robots (2.75) • Public Register of Permission to Use Data (2.71)

Table 3: Top five and bottom five scoring potential regulatory measures based on overall scores.

Technical Measures

In R1, technical measures were rarely mentioned. However, in R2, technical measures scored relatively high, particularly in regard to desirability; all technical measures were very high (4.5-5) or high (4-4.49) for desirability. However, with lower average scores in feasibility and probability, only three technical measures were in the top fifteen measures. For the overall scores, all technical measures scored in the mid-high range (3.5-3.99). All potential technical measures were rated more desirable, then feasible, then probable.

Top Three Technical Measures	Bottom Three Technical Measures
<ul style="list-style-type: none"> • Tools for verifying & certifying labelled datasets for accuracy, bias & other important properties (3.95) • Methodologies for systematic & comprehensive testing of AI-based systems (3.90) • Techniques for providing explanations for output of AI models (3.87) 	<ul style="list-style-type: none"> • Reputation information about publicly available services based on machine learning models (3.63) • Tools capable of identifying synthetically created or manipulated content (3.58) • AI-as-a-service (3.52)

Table 4: Top three and bottom three scoring potential technical measures based on overall scores. As there were only eight potential technical measures, only the top and bottom three were highlighted.



Other Measures

In R1, respondents cited a broad range of other measures. In R2, these other potential measures scored high, both in absolute terms and relative to the other two categories of measures. Twelve of the top fifteen measures were other measures. This was because other measures received the highest average scores in feasibility, probability, and overall. For the overall scores, seven measures were in the very high (4.5-5) range, sixteen in the high (4-4.49) range, and two in the mid-high (3.5-3.99) range. The only measure to score in the mid-low (3-3.49) range overall was citizen juries. The majority of measures (23 of 26) scored more desirable than feasible or probable.

Top Five Other Measures	Bottom Five Other Measures
<ul style="list-style-type: none">• Investigative Journalism (4.48)• Exchange of Best Practices (4.43)• Education Campaigns (4.19)• Framework, Guidelines, and Toolkits (4.14)• Ethical Codes of Conduct (4.11)	<ul style="list-style-type: none">• Self-Regulation by Company (3.58)• Retaining “Unsmart” Products and Services (3.54)• Ethical Rules pertaining to the creation or use of machine learning models with potential malicious applications (3.40)• Ethical Rules pertaining to the use or treatment of AI agents in robotics or virtual environments (3.22)• Citizen Juries (2.82)

Table 5: Top five and bottom five scoring other potential measures based on overall scores.

Next Steps

The SHERPA project is currently developing its final recommendations on the ethical and human rights aspects of SIS. Round 3 of this Delphi study will help us refine and prioritise these recommendations. It will consist of identifying the most important potential governance measures for immediate action.

