AI observatory and society article

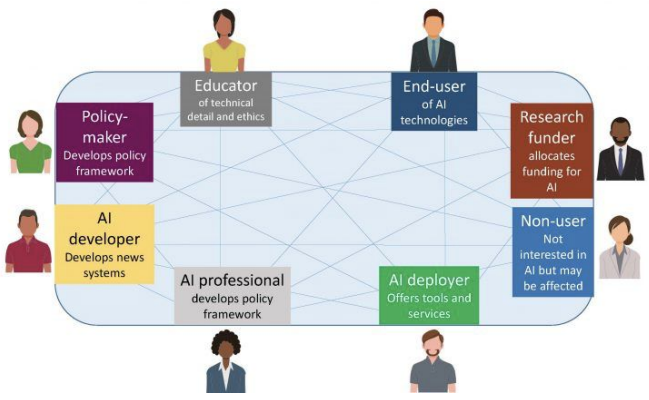
By Dr Nitika Bhalla and Prof Bernd Stahl

Smart information systems (SIS) have the potential to affect most aspects of modern society. The ideal scenario would be to create an environment where humans feel complimented and empowered by big data analytics and artificial intelligence (AI), however this is sometimes not the case, and therefore attracts significant public debate.

A key European funded project that brings ethical and human rights issues of these technologies to light is the SHERPA project ('Shaping the ethical dimensions of smart information systems (SIS) – a European perspective'). The aim of the project is to analyse and understand novel ways in which SIS systems can impact current ethical and human rights issues, and produce a set of recommendations that will help to enhance human flourishing and public confidence. The final phase of the project will advocate these recommendations to the EU and other policy and decision makers, so that these can be taken forward to improve current ethical, human rights and legal frameworks in the interest of public good.

SHERPA initially investigated 10 case studies of areas that employ SIS and a further 5 policy scenarios. The empirical studies were important to understand what are the main issues which concern the users of SIS. Simultaneously, a human rights analysis was carried out to understand current frameworks, and an analysis of technical aspects of cybersecurity vulnerabilities of SIS. One unique aspect of this project is that there is a continual working relationship with a broad range of stakeholders, who have helped to evaluate our findings via; interviews, an online survey, the Delphi study, focus groups and the SHERPA stakeholder board. In light of this, the project has developed a set of guidelines, (one for developers and one for users of SIS), undertaken an analysis of regulatory options and contributed to standardisation activities relating to AI and big data.

Based on the understanding of ethical and human rights issues from all our findings, SHERPA then developed a set of recommendations described below. (For a more detailed explanation please see our website).



SHERPA Recommendations:

The recommendations are inspired by the recognition that AI can be seen as an innovation ecosystem. In order for such an ecosystem to be conducive to human flourishing, it needs to be clearly defined, develop and maintain a relevant knowledge base and have an appropriate governance structure. The following recommendations help to achieve this.

- **Conceptual clarity - Recommendation: Use appropriate and clear definitions of AI and digital technology:**

  Use a concept of AI that points to the features of the technology that are ethically relevant, such as opacity (can hide bias) or automation (replaces jobs).

  SHERPA's work on case studies and scenarios has informed the categorisation of AI in terms of narrow AI (machine learning), converging socio-technical systems and artificial general intelligence.

- **AI Impact Assessment - Recommendation: Develop baseline model for AI impact assessments**

  There is no universally accepted baseline of what an AI impact assessment should entail. There is broad agreement that a risk-based approach to AI is appropriate. For such an approach to work, there must be guidance on how to define, measure, interpret and address relevant risks.

  SHERPA has provided accounts of ethical and human rights issues of AI in organisations as part of its case study research and likely future issues are discussed in the scenarios.

- **Ethics by Design - Recommendation: Promote Ethics by Design for researchers in EC-funded projects**

  There is a current lack of ethical awareness and guidance for researchers working on AI-assisted systems in EU framework programmes.

  SHERPA has produced an overview of ethical concerns arising in relation to AI. Furthermore, the project has created sets of guidelines for operationalisation of ethics by design for developers and for users of smart information.

These can be used as a foundation from which to develop guidelines for Horizon Europe research applicants and organisations.

- **Education on AI and Ethics - Recommendation: Create training and education pathways that include ethics and human rights in AI**

The ability to recognise and deal with ethical and human rights aspects of emerging digital technologies (such as AI) depends on awareness and understanding, which is currently lacking throughout the educational system. The importance of education was introduced by the SHERPA stakeholder board and was investigated further in the; Online Survey, the Delphi Study, case study research, the scenarios, analysis of the human rights aspects, the guidelines for operationalisation of Ethics by Design, analysis of the Stakeholders' Focus groups and Interviews.

- **Standardization - Recommendation: Include research findings on AI ethics in standardization**

When ethics and human rights are not taken into consideration for AI-related standards, there is a risk of serious negative implications for society, and a reduction of trust in AI.

SHERPA has contributed to the standard on health and wellness apps that includes an annex on ethics and SHERPA members are represented on national mirror committees of ISO/IEC JTC 1/SC 42 Artificial intelligence.

- **Security - Recommendation: Undertake security analysis for machine learning systems**

Most safeguards for ethics and human rights in AI rely on the integrity and reliability of the technical systems. However, these AI systems may be subject to novel attacks and security vulnerabilities. SHERPA has produced the deliverable on Technical Options and Interventions which discusses technical security of AI for ethical and human rights safeguards.

- Regulatory Framework - **Recommendation: Develop a regulatory framework and enforcement mechanisms for AI**

The protection of human rights and safety is a societal priority. The EU should develop a mandatory regulatory framework to ensure that AI systems are safe, and do not violate fundamental rights and ethical principles.

SHERPA has produced the deliverable D3.3 Report on regulatory options and has given feedback to the European Parliament on JURI report on a framework of ethical aspects of artificial intelligence, robotics and related technologies.

- **European Agency for AI - Recommendation: Establish an independent European Union Agency for AI**

The current regulatory landscape for AI in the EU is fragmented, and concerns have been raised regarding cooperation, coordination and consistent application of EU law.

The SHERPA Terms of Reference for a European Agency for AI developed using research and expert consultations, including interviews, a focus group, and feedback from the Stakeholder Board.

- **AI Ethics Officer - Recommendation: Establish role of AI (Ethics) Officer in organisations**

  There is no single role bearing responsibility for AI ethics concerns in all organisations using and developing AI-assisted systems. This could allow for unethical practices to develop and prosper. SHERPA has produced an overview of ethical and human rights concerns arising in relation to AI.

Moving forward:

SHERPA will now advocate these recommendations through communication and dissemination to EC policy makers. It is the aim of SHERPA to make a difference to the world of AI and society.

Finally, the project continues to collaborate with two other projects under H2020, namely SIENNA and PANELFIT. The collaboration is in the process of organising a final event with STOA (Panel for the future of Science and Technology) on 23rd March 2021, with a view to moving forward in the area of ethics, human right and legal aspects of ICT research and development.