DECLARATION ON ETHICS AND DATA PROTECTION IN ARTIFICAL INTELLIGENCE

40th International Conference of Data Protection and Privacy Commissioners

Tuesday 23rd October 2018, Brussels

AUTHORS:

- Commission Nationale de l'Informatique et des Libertés (CNIL), France

- European Data Protection Supervisor (EDPS), European Union

- Garante per la protezione dei dati personali, Italy

CO-SPONSORS:

- Agencia de Acceso a la Información Pública, Argentina

- Commission d'accès à l'information, Québec, Canada

- Datatilsynet (Data Inspectorate), Norway

- Information Commissioner's Office (ICO), United Kingdom

- Préposé fédéral à la protection des données et à la transparence, Switzerland

- Data protection Authority, Belgium

- Privacy Commissioner for Personal Data, Hong-Kong

- Data protection Commission, Ireland

- Data Protection Office, Poland

- Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), Mexico

- National Authority for Data Protection and Freedom of Information, Hungary

- Federal Commissioner for Data Protection and Freedom of Information, Germany

- Office of the Privacy Commissioner (OPC), Canada

- National Privacy Commission, Philippines

The 40th International Conference of Data Protection and Privacy Commissioners:

Considering the initial discussion at the 38th International Conference of Data Protection and Privacy Commissioners in Marrakesh on Artificial intelligence, Robotics, Privacy and Data Protection;

Recognizing that artificial intelligence systems may bring significant benefits for users and society, including by: increasing the rapidity of processes and supporting decision-making; creating new ways to participate in democratic processes; improving efficiency in public sector and industry; achieving more equitable distribution of resources and opportunities; offering new methods and solutions in various fields such as public health, medical care, security, sustainable development, agriculture and transport; bringing new opportunities in scientific research and education and; providing individuals with more personalized services;

Taking into account the significant progress in certain areas of artificial intelligence, in particular regarding the processing of large amounts of information, the analysis and prediction of human behavior and characteristics, and in related fields such as robotics, computer vision and autonomous systems, likely to make significant progress in the near future;

Highlighting the rapid advancement of big data and artificial intelligence, notably machine learning, in particular with the development of deep learning technologies, allowing algorithms to solve complex operations leading to potential decisions, making however such processes more opaque;

Affirming that the respect of the rights to privacy and data protection are increasingly challenged by the development of artificial intelligence and that this development should be complemented by ethical and human rights considerations;

Considering that machine learning technologies in particular, and artificial intelligence systems in general, may rely on the processing of large sets of personal data for their development, potentially impacting data protection and privacy; also taking into account the potential risks induced by the current trend of market concentration in the field of artificial intelligence;

Recognizing the link between collections, uses and disclosures of personal information – the traditional sphere of privacy and data protection – on the one hand, and the direct impacts on human rights more broadly, most notably regarding discrimination and freedom of expression and information, and thus acknowledging the need for data protection and privacy authorities to think about human rights more broadly, and for data protection and privacy authorities to work with other authorities addressing human rights;

Pointing out that some data sets used to train machine learning-based and artificial intelligence systems have been found to contain inherent bias resulting in decisions which can unfairly discriminate against certain individuals or groups, potentially restricting the availability of certain services or content, and thus interfering with individuals' rights such as freedom of expression and information or resulting in the exclusion of people from certain aspects of personal, social, professional life;

Stressing that artificial intelligence powered systems whose decisions cannot be explained raise fundamental questions of accountability not only for privacy and data protection law but also liability in the event of errors and harm;

Noting that many stakeholders in the field of artificial intelligence have expressed their concerns about the risks of malicious use of artificial intelligence, as well as the risks related to privacy, data protection and

human dignity, pointing out for example that the development of artificial intelligence in combination with mass surveillance raises concerns about their possible use to curtail fundamental rights and freedoms;

Highlighting that those risks and challenges may affect individuals and society, and that the extent and nature of potential consequences are currently uncertain;

Emphasising the importance of trust, since strong data protection and privacy safeguards help to build individuals' trust in how their data is processed, which encourages data sharing and thereby promotes innovation;

Taking the view that the current challenges triggered by the development of artificial intelligence and machine learning systems reinforce the need for the adoption of an international approach and standards, in order to ensure the promotion and protection of human rights in all digital developments at international level;

Reaffirming the commitment of data protection authorities and the Conference of Data Protection and Privacy Commissioners to uphold data protection and privacy principles in adapting to this evolving environment, notably by engaging resources and developing new skills in order to be prepared for future changes.

The **40ᵗʰ International Conference of Data Protection and Privacy Commissioners** considers that any creation, development and use of artificial intelligence systems shall fully respect human rights, particularly the rights to the protection of personal data and to privacy, as well as human dignity, non-discrimination and fundamental values, and shall provide solutions to allow individuals to maintain control and understanding of artificial intelligence systems.

The Conference therefore endorses the following guiding principles, as its core values to preserve human rights in the development of artificial intelligence:

1. Artificial intelligence and machine learning technologies should be designed, developed and used in respect of fundamental human rights and in accordance with the **fairness principle**, in particular by:

   a. Considering individuals' reasonable expectations by ensuring that the use of artificial intelligence systems remains consistent with their original purposes, and that the data are used in a way that is not incompatible with the original purpose of their collection,

   b. taking into consideration not only the impact that the use of artificial intelligence may have on the individual, but also the collective impact on groups and on society at large,

   c. ensuring that artificial intelligence systems are developed in a way that facilitates human development and does not obstruct or endanger it, thus recognizing the need for delineation and boundaries on certain uses,

2. **Continued attention and vigilance,** as well as accountability, for the potential effects and consequences of, artificial intelligence systems should be ensured, in particular by:

   a. promoting accountability of all relevant stakeholders to individuals, supervisory authorities and other third parties as appropriate, including through the realization of audit, continuous monitoring and impact assessment of artificial intelligence systems, and periodic review of oversight mechanisms;

   b. fostering collective and joint responsibility, involving the whole chain of actors and stakeholders, for example with the development of collaborative standards and the sharing of best practices,

   c. investing in awareness raising, education, research and training in order to ensure a good level of information on and understanding of artificial intelligence and its potential effects in society, and

   d. establishing demonstrable governance processes for all relevant actors, such as relying on trusted third parties or the setting up of independent ethics committees,

3. Artificial intelligence **systems transparency and intelligibility** should be improved, with the objective of effective implementation, in particular by:

   a. investing in public and private scientific research on explainable artificial intelligence,

   b. promoting transparency, intelligibility and reachability, for instance through the development of innovative ways of communication, taking into account the different levels of transparency and information required for each relevant audience,

   c. making organizations' practices more transparent, notably by promoting algorithmic transparency and the auditability of systems, while ensuring meaningfulness of the information provided, and

   d. guaranteeing the right to informational self-determination, notably by ensuring that individuals are always informed appropriately when they are interacting directly with an artificial intelligence system or when they provide personal data to be processed by such systems,

   e. providing adequate information on the purpose and effects of artificial intelligence systems in order to verify continuous alignment with expectation of individuals and to enable overall human control on such systems.

4. As part of an overall "ethics by design" approach, artificial intelligence systems should be **designed and developed responsibly**, by applying the principles of **privacy by default and privacy by design**, in particular by:

   a. implementing technical and organizational measures and procedures – proportional to the type of system that is developed – to ensure that data subjects' privacy and personal data are respected, both when determining the means of the processing and at the moment of data processing,

   b. assessing and documenting the expected impacts on individuals and society at the beginning of an artificial intelligence project and for relevant developments during its entire life cycle, and

   c. identifying specific requirements for ethical and fair use of the systems and for respecting human rights as part of the development and operations of any artificial intelligence system,

5. **Empowerment of every individual** should be promoted, and the exercise of individuals' rights should be encouraged, as well as the creation of opportunities for public engagement, in particular by:

   a. respecting data protection and privacy rights, including where applicable the right to information, the right to access, the right to object to processing and the right to erasure, and promoting those rights through education and awareness campaigns,
   b. respecting related rights including freedom of expression and information, as well as non-discrimination,
   c. recognizing that the right to object or appeal applies to technologies that influence personal development or opinions and guaranteeing, where applicable, individuals' right not to be subject to a decision based solely on automated processing if it significantly affects them and, where not applicable, guaranteeing individuals' right to challenge such decision,
   d. using the capabilities of artificial intelligence systems to foster an equal empowerment and enhance public engagement, for example through adaptable interfaces and accessible tools.

6. Unlawful **biases or discriminations** that may result from the use of data in artificial intelligence should be reduced and mitigated, including by:

   a. ensuring the respect of international legal instruments on human rights and non-discrimination,
   b. investing in research into technical ways to identify, address and mitigate biases,
   c. taking reasonable steps to ensure the personal data and information used in automated decision making is accurate, up-to-date and as complete as possible, and
   d. elaborating specific guidance and principles in addressing biases and discrimination, and promoting individuals' and stakeholders' awareness.

Taking into consideration the principles above, the 40th International Conference of Data Protection and Privacy Commissioners calls for **common governance principles on artificial intelligence** to be established, fostering concerted international efforts in this field, in order to ensure that its development and use take place in accordance with ethics and human values, and respect human dignity. These common governance principles must be able to tackle the challenges raised by the rapid evolutions of artificial intelligence technologies, on the basis of a multi-stakeholder approach in order to address all cross-sectoral issues at stake. They must take place at an international level since the development of artificial intelligence is a trans-border phenomenon and may affect all humanity. The Conference should be involved in this international effort, working with and supporting general and sectoral authorities in other fields such as competition, market and consumer regulation.

The 40th International Conference of Data Protection and Privacy Commissioners therefore establishes, as a contribution to a future common governance at the international level, and in order to further elaborate guidance to accompany the principles on Ethics and Data Protection in Artificial Intelligence, a **permanent working group** addressing the challenges of artificial intelligence development. This **working group on Ethics and Data Protection in Artificial Intelligence** will be in charge of promoting understanding of and respect for

Declaration on Ethics and Data Protection in Artificial Intelligence

the principles of the present resolution, by all relevant parties involved in the development of artificial intelligence systems, including governments and public authorities, standardization bodies, artificial intelligence systems designers, providers and researchers, companies, citizens and end users of artificial intelligence systems. The working group on Ethics and Data Protection in Artificial Intelligence shall take into account the work carried out by other working groups of the Conference and shall report regularly on its activities to the Conference. The Conference thus endeavors to proactively support an active public debate on digital ethics aiming at the creation of a strong ethical culture and personal awareness in this field.

- The present declaration will be open for public consultation -