SmartSeminars

SmartData@PoliTO center

Torino, 30 November 2018

Data Protection in the Age of AI: Going Beyond the GDPR

Alessandro Mantelero

Polytechnic University of Turin

HEALTH

IBM Watson Reportedly Recommended Cancer Treatments That Were 'Unsafe and Incorrect'

Jennings Brown
7/25/18 3:00pm
45.3K 27 5

Filed to: DOCTOR: THIS PRODUCT IS A PIECE OF SHIT...

Photo: Getty

Internal company documents from IBM show that medical experts working with the company's Watson supercomputer found "multiple examples of unsafe and incorrect treatment recommendations" when using the software, according to a report from Stat News.

Minority homebuyers face widespread statistical lending discrimination

November 15, 2018 by Laura Counts, University of California - Berkeley

ublic Domain

-to-face meetings between mortgage officers and homebuyers ha
cations and algorithms, but lending discrimination hasn't gone aw

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

▲ Amazon's automated hiring tool was found to be inadequate after penalizing the résumés of female candidates. Photograph: Brian Snyder/Reuters

Amazon's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

<|>

Data protection regulation: a long dialogue between engineers and law-makers

Mantelero, A., Vaciago, G. Legal Aspects of Information Science, Data Science and Big Data. In Dehmer, M., Emmert-Streib, F. (eds). *Frontiers in Data Science*. (CRC Press, 2017)

- Data protection as the response to the growing concern of citizens about the risk of computer-based social control by governments (and large corporations)
- Data protection as a counter-control over information

- Process-based approach
- Rights-based approach vs risks/benefits analysis
- Data protection in the framework of fundamental rights

Viktor Mayer-Schönberger, 'Generational development of data protection in Europe?' in Philip E. Agre and Marc Rotenberg (eds), Technology and privacy: The new landscape (MIT Press 1997)

< II >

The GDPR and its limitations

Mantelero, A. 2014. The future of consumer data protection in the E.U. Rethinking the “notice and consent” paradigm in the new era of predictive analytics. Computer Law and Security Review, 30 (6): 643-660

- The myth of the consent
- The myth of the predictable specific purposes
- The myth of data minimization
- The troubles of risk assessment (DPIA and uncertainty)

< III >

Going Beyond the GDPR

Mantelero, A. 2016. Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection. Computer Law and Security Review, 32 (2): 238-255

- Big data/AI: a new representation of our society
  - A categorical approach in BD/AI supported decision-making processes
  - From individual to group profiling: A variable geometry

- Data protection as an enabling right
  - The impact on rights and freedoms other than data protection

< IV >

The emerging of the ethical dimension

EU border 'lie detector' system criticised as pseudoscience

Technology that analyses facial expressions being trialled in Hungary, Greece and Latvia

▲ A Hungarian police officer stands guard at Serbia's border with Hungary near a makeshift camp for migrants. Photograph: Darko Vojinovic/AP

The EU has been accused of promoting pseudoscience after announcing plans for a "smart lie-detection system" at its busiest borders in an attempt to identify illegal migrants.

Privacy

The odd reality of life under China's all-seeing credit score system

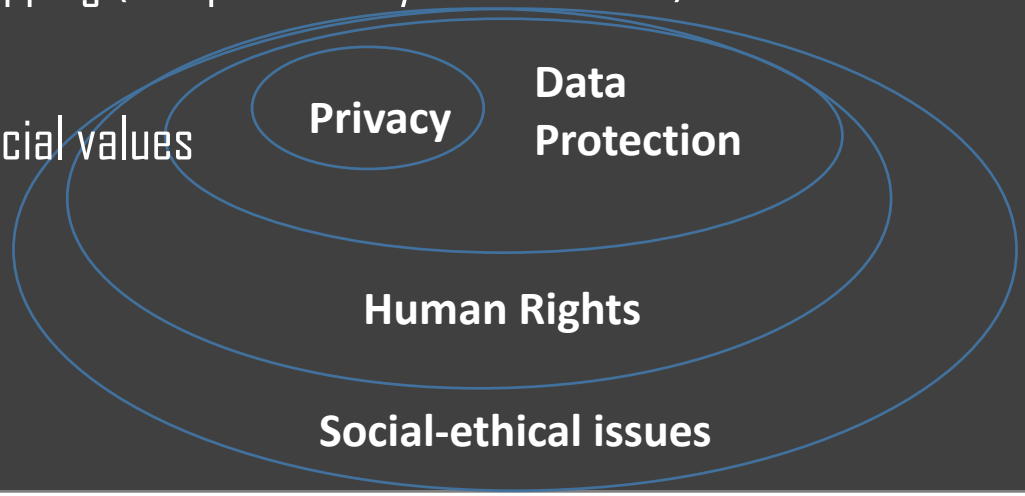
Looking for love? In China, a good credit score could get you access to exclusive singles

Credit **Kevin Hong**

- The challenges of BD/AI and the role of ethics
- Going beyond the individual dimension: the role of a participatory approach

- Ethics and law: avoiding an improper overlapping (complementary role of ethics)
- Ethics as a corporate branding
- Context-dependent nature of ethical and social values

- Excessive emphasis on transparency



< V >

The ongoing debate and the existing guidelines

Mantelero, A. 2017. Regulating Big Data. The guidelines of the Council of Europe in the Context of the European Data Protection Framework. Computer Law and Security Review, 33 (5): 584-602.

- The guidelines on big data adopted by the Council of Europe
- The draft of the CoE guidelines on AI
- European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))
- The EDPS initiative on data ethics
- The corporate and professional code of conducts

- The guidelines on big data adopted by the Council of Europe [2017]
  - "Adequately take into account the likely impact of the intended Big Data processing and its broader ethical and social implications"
  - Precautionary approach
  - Defining social and ethical values: An "architecture of values" based on three layers
  - Involvement of different stakeholders (role of DPAs)
  - Transparency of risk assessment
  - "minimise the presence of redundant or marginal data, avoid potential hidden data biases and the risk of discrimination or negative impact on the rights and fundamental freedoms of data subjects"

- (when technically feasible) test by-design solutions adopted on a limited amount of data by means of simulations
- By-design solutions to avoid as much as possible non-sensitive data being used to infer sensitive information
- Deign solution: Interfaces which simulates the effects of the use of data and its potential impact on the data subject; user-friendly technical ways for data subjects to react to unfair uses of data/withdraw consent
- Risk of re-identification
- Freedom not to rely on the result of the recommendations provided using Big Data
- Open Data: "carefully consider… open data policies"; assess the effects merging and mining open data sets
- Education

- The draft of the CoE guidelines on AI [October 2018]

  I. General guidance

  - 1. Responsibility towards individuals and society is the primary objective in AI development, taking the safeguard of human rights and fundamental freedoms, and in particular human dignity and the right to the protection of personal data, as the pre-eminent goal/an absolute pre-requisite.
  - 2. AI development and AI applications must adopt a fundamental rights-oriented perspective. This is even more important when AI applications are used in the context of decision-making processes.
  - 3. Data-centric AI development must be based on the principles of Convention 108. The key elements of this approach are: proportionality of data processing, responsibility, accountability, transparency, and risk management.
  - 4. A risk-aware approach is no barrier to innovation, but an enabler. Governments and citizens should therefore consider the risks of datafication and the potentially adverse implications of data-driven solutions.

General guidance

- 5. Individuals and communities should have the right to freely decide what role AI should play in shaping social dynamics, collective behaviour, and decisions affecting entire groups of individuals.
- 6. In line with the guidance on risk assessment provided in the Guidelines on Big Data adopted by the Council of Europe in 2017, a wider view of the possible outcomes of data processing should be adopted to consider the impact of data use not only on human rights and fundamental freedoms but also on collective social and ethical values.
- 7. AI development and AI applications shall not diminish or negatively affect data subjects' rights enshrined by Convention 108.

II. Guidance for developers

- 1. The Parties to the Convention actively encourage AI developers towards a value-oriented design of their products and services, which shall be consistent with the values expressed in Convention 108 and in the regulations of the Council of Europe.
- 2. AI developers shall assess the adverse consequences of AI applications on data subjects' human rights and fundamental freedoms, considering these effects and their distribution to adopt a precautionary approach based on risk prevention policies.
- 3. In developing AI applications, it is important to adopt a design paradigm that critically assesses the nature and amount of data used.
- 4. In all phases of the processing, including data collection and analysis stages, AI developers shall adopt a by-design approach to avoid potential unintentional and hidden data biases, and the risk of discrimination or negative impacts on the human rights and fundamental freedoms of data subjects.

- 5. The risk of de-contextualised data and de-contextualised algorithmic models should be adequately considered in developing and using AI applications
- 6. Independent committees of experts from a range of fields, as well as independent academic institutions, should be involved in AI development and use, to provide a valuable support in designing rights-based and socially-oriented AI and to contribute to detect potential bias. Such committees play an even more important role in areas where transparency and stakeholders' engagement are difficult to achieve, such as predictive justice, crime detection or predictive policing.
- 7. Participatory forms of risk assessment, based on the active engagement of the individuals and communities potentially affected by AI applications, should be adopted.
- 8. A participatory assessment of the far-reaching effects of algorithmic decision-making should drive controllers to adopt co-design solutions for developing AI applications, actively engaging the groups potentially affected by them.

- 9. When it is technically feasible, AI developers should design their products and services in a manner that safeguards users' freedom of choice over the use of AI and provides alternatives to AI-equipped devices and services.
- 10. The adopted design paradigms in AI development shall critically assess the nature and amount of data used, reducing redundant or marginal data, starting with a restricted amount of training data, and then monitoring the model's accuracy as it is fed with new data. The use of synthetic data can be considered as one of the possible solutions to minimise personal data processing.
- 11. Data subjects shall be entitled to be informed appropriately when they are interacting with an AI system, and to know the AI applications used and the logic underlying AI data processing operations, including the consequences of such a reasoning.

III. Guidance for policy makers

- 1. Public procurement procedures could impose specific duties of transparency, prior assessment and algorithm vigilance of AI systems to service providers.
- 2. Public trust in AI products and services could benefit from an increased AI developers' accountability and the adoption of risk assessment procedures.
- 3. Supervisory authorities and controllers should adopt forms of algorithm vigilance to better ensure compliance with data protection and human rights principles over the entire lifetime of AI applications.
- 4. AI applications may induce overconfidence among decision-makers on the reliable nature of the solutions provided. This decision-makers' attitude may be reinforced by the threat of potential liability for taking a decision other than the one suggested by AI systems. The autonomy of human intervention in decision-making processes and the freedom of human decision makers not to rely on the result of the recommendations provided using AI shall therefore be preserved.

- 5. When AI applications may significantly impact on the human rights and fundamental freedoms of data subjects, controllers should consult the supervisory authorities to seek advice to mitigate this potential adverse impact.
- 6. Since several countries have independent watchdogs for supervising specific sectors where AI applications operate or may operate, it is important to strengthen the mutual cooperation between these authorities and their cooperation with the supervisory authorities of Convention 108.
- 7. When committees of experts are created at company level, the Parties might empower supervisory authorities to scrutinise these committees when shortcomings in their independency, abilities or decisions affect data processing.
- 8. Policy makers should invest resources in digital literacy and education to increase data subjects' awareness and understanding of AI systems and their effects. They should also encourage professional training for AI developers to raise awareness and understating of the potential effects of AI on individuals and society.

- The EDPS initiative on data ethics
  - Contextualisation of several values
  - Overlapping between socio-ethical values and legal values (dignity, freedom, autonomy, solidarity, equality, democracy, justice, and trust)
  - Guidelines concerning regulated issues (intellectual property, market, property rights) or regulated contexts (health care, finance, micro-targeting)

- The EDPS initiative on data ethics

  Five 'directions' of thought and innovation
  - 1. The dignity of the person remains inviolable in the digital age
  - 2. Personhood and personal data are inseparable from one another
  - 3. Digital technologies risk weakening the foundation of democratic governance
  - 4. Digitised data processing risks fostering new forms of discrimination
  - 5. Data commoditisation risks shifting value from persons to personal data

Alessandro Mantelero

alessandro.mantelero@polito.it

@mantelero

Department of Management and Production Engineering (DIGEP)

Mantelero, A. 2018. AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. Computer Law and Security Review, 34 (4): 754-772 (open access)