MSI-NET

Committee of experts on internet intermediaries

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

6 October 2017 **MSI-NET (2017)06**

MSI-NET 4th meeting
18-19 September 2017
(Strasbourg, Agora, Room G05)

Meeting report

1. The meeting was opened by the Chair of the MSI-NET, Prof. Wolfgang Schulz. Jan Kleijssen, Council of Europe Director of the Information Society and Action Against Crime Department then welcomed members and participants, commended them for the work carried out so far and invited them to actively engage in the finalisation of the two deliverables during this last meeting. Mr Kleijssen recalled that during the 12th meeting of the CDMSI (Steering Committee on Media and Information Society), which took place in June 2017, delegates had been informed by the MSI-NET's Vice-Chair about the progress made on both texts, and had agreed to launch public consultations on the draft recommendation over the summer. In the course of the consultations, 23 sets of detailed comments were received from representatives of member states, conventional committees, civil society and from academia. Mr Kleijssen further informed the MSI-NET members and participants of the ongoing implementation of the Internet Governance Strategy as well as of recent developments in the context of the Council of Europe initiative to create a platform to foster the dialogue between the member states and internet companies to improve respect for human rights, democracy and the rule of law online.

Mr. Kleijssen further highlighted the Conference on 13 October in Vienna, co-organised with the OSCE Chairmanship (Austria) and the Chairmanship of the Council of Europe Committee of Ministers (Czech Republic) with the title "The roles and responsibilities of internet intermediaries" where the MSI-NET would be prominently represented and which would be a unique opportunity to draw attention to the Committee's work and the draft recommendation. Wishing the members and participants a fruitful debate on the highly relevant topics on their agenda, he also encouraged them to aim for concise and clear texts which could provide guidance to member states.

2. The agenda Appendix 1) was adopted without any changes. The list of participants appears in Appendix 2. The gender distribution of the 30 participants was 12 women (40%) and 18 men (60%).

Conclusions and decisions

3. With respect to the *draft Committee of Ministers recommendation on the roles and responsibilities of internet intermediaries*, the MSI-NET reviewed the revised version of the document, as presented by the Rapporteur and Secretariat, which incorporated the comments received in the course of the public consultations. The MSI-NET agreed with most amendments as presented and further enhanced the text by:

a) clarifying standards for expeditious procedures for the restriction of content without judicial order;

b) providing more detail to the paragraphs on notice-based procedures in both chapters of the guidelines (states obligations and intermediary responsibilities);

c) introducing more specific guarantees related to judicial and non-judicial review and complaints mechanisms established by intermediaries, and

d) adding language related to the responsibility of large intermediaries to develop standards and codes of ethics aimed at the prevention of abusive language and imagery, hatred and incitement to violence.

A number of additional specific observations, comments and proposals for changes to the text were made **and the draft recommendation finalised for approval by the CDMSI in its 13th meeting** (5 – 8 December 2017) (Appendix 3).

4. The MSI-NET further discussed parts IV and V of the *draft study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*, as had been submitted by the Rapporteur at the end of August. Members and participants made a number of additional comments, suggestions and proposals for changes which were agreed to be inserted for **the final version of the study** (Appendix 4). It was agreed to conclude the final chapter with concrete proposals for action by member states, and to suggest that the Council of Europe should conduct further work on the topics of the study, with a view to developing future standard-setting instruments for guidance to member states.

Any other business

6. The Secretariat was tasked to prepare a draft meeting report to be sent to the Chair and the Vice-Chair for consideration. Thereafter, the draft report will be sent to the MSI-NET with a deadline of five full working days allowing for comments. In the absence of comments, the report will be deemed finalised and will be transmitted to the CDMSI for information, together with the finalised versions of the two deliverables. Therefore, it is not considered necessary to produce an abridged report of meeting.

APPENDIX 1

AGENDA[1]

1. **Opening of the meeting**

2. **Adoption of the agenda**

3. **Information by the Secretariat**

4. **Finalisation of the third revised draft recommendation by the Committee of Ministers on the roles and responsibilities of internet intermediaries**
   doc MSI-NET(2016)05rev3

5. **Finalisation of the second revised draft study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications**
   doc MSI-NET(2016)06rev2

6. **Any other business**

MSI-NET Terms of Reference

[1] As it appears in document MSI-NET(2017)05

APPENDIX 2

LIST OF PARTICIPANTS

COMMITTEE MEMBERS

Mr Bertrand de la CHAPELLE – Co-founder and Director of the Internet & Jurisdiction, France

Ms Julia HÖRNLE – Professor of Internet Law, Queen Mary University of London

Ms Tanja KERŠEVAN-SMOKVINA – Principal Advisor to Director General, Agency for Communication Networks and Services – Slovenia (Gender Equality Rapporteur)

Mr Matthias KETTEMANN – Postdoc Fellow, Cluster of Excellence "Normative Orders" University of Frankfurt/Main (Rapporteur Recommendation)

Ms Dörte NIELANDT**,** Division VI A3 (Legal framework for digital services, media industry), Federal Ministry for Economic Affairs and Energy – Germany

Mr Arseny NEDYAK – Deputy Director, Department of Media State Policy, Ministry of Telecommunication, Russian Federation

Mr Pēteris PODVINSKIS – Ministry of Foreign Affairs, International Organisations Directorate, Department for Public Policy related to Internet – Latvia

Mr Thomas SCHNEIDER – Deputy Director of International Affairs, International Information Society Coordinator, Federal Department of the Environment, Transport, Energy and Communication DETEC, Federal Office of Communications (OFCOM) - Switzerland

Mr Wolfgang SCHULZ – Professor, Faculty of Law, University of Hamburg / Hans-Bredow-Institut (Chair)

Ms Sophie STALLA-BOURDILLON – Associate Professor in Information Technology / Intellectual Property Law, Director of ILAWS, Southampton Law School University of Southampton

Ms Karmen TURK – Trinity Tallinn – Estonia (Vice-Chair)

Mr Dirk VOORHOOF – Lecturer European Media Law, UCPH (Copenhagen University) / Professor at Ghent University / member of the CMPF Scientific Committee Centre for Media Pluralism and Press Freedom

Mr Benjamin WAGNER – Assistant Professor, Institute for Management Information Systems, Vienna University of Economics and Business (Rapporteur Study)

COUNCIL OF EUROPE MEMBER STATES

AUSTRIA - Mr Gerhard HOLLEY, Federal Chancellery, constitutional office

CZECH REPUBLIC – Mr Jakub SVAB, Media and Audio-vision Department, Ministry of Culture

TURKEY - Mr İrfan Dündar ERENTÜRK, Media Specialist, Radio and Television Supreme Council (RTÜK) Ankara

OBSERVERS

EUROPEAN UNION - AGENCY FOR FUNDAMENTAL RIGHTS (FRA) *Apologized*

EUROPEAN COMMISSION - DG CONNECT - Ms Irene ROCHE LAGUNA, Legal officer, DG for Communications Networks, Content & Technology

EUROPEAN AUDIOVISUAL OBSERVATORY - Ms Maja CAPPELLO, Head of Legal Information Department (apologised)

EBU / EUROPEAN BROADCASTING UNION - Mr Michael WAGNER, Head of Media Law and Communications, Legal Department

REPRESENTATIVES OF CIVIL SOCIETY, ACADEMIC COMMUNITIES AND THE PRIVATE SECTOR

Ms Christina ANGELOPOULOS, Centre for Intellectual Property and Information Law (CIPIL), University of Cambridge, United Kingdom

Mr Allon BAR, Independent Consultant

Mr Giancarlo FROSIO - Centre for International Intellectual Property Studies (CEIPI) - University of Strasbourg, France

Ms Gabrielle GUILLEMIN, Article 19 (18.09.2017)

Mr Martin HUSOVEC, Assistant Professor; Institute for Law, Technology and Society, Tilburg, Netherlands

Ms Catherine KENT - Essex University (apologized)

Ms Aleksandra KUCZERAWY, Legal Researcher, Centre for IT & IP Law – iMinds, Univeristy Leuven, Belgium

Mr Joe McNAMEE, Executive director, European Digital Rights (EDRi), Brussels, Belgium

NON-MEMBER STATES

MOROCCO

Ms Chanaz El AKRICHI, Head of Cooperation division, Ministry of Communication

Ms Meriem KHATOURI, Director for Media Studies and Development, Ministry of Communication

Mr El Mahdi AROUSSI IDRISSI, Director Legal Affairs, The High Authority for Audio-visual Communication (HACA) RABAT, MAROC

SECRETARIAT

Mr Jan KLEIJSSEN, Director, Directorate of Information Society and Action against Crime

Mr Patrick PENNINCKX, Head of Information Society Department (apologised)

Ms Silvia GRUNDMANN, Head of Media and Internet Division, Information Society Department

Ms Charlotte ALTENHÖNER-DION, Secretary of the MSI-NET Committee, Media and Internet Division, Information Society Department

Ms Francesca MONTAGNA, Administrator, Media and Internet Division, Information Society Department

Ms Elisabeth MAETZ, Assistant, Media and Internet Division, Information Society Department

INTERPRETERS / INTERPRETES

Mr Grégoire DEVICTOR, Mr Jean-Jacques PEDUSSAUD, Mr Nicolas GUITTONNEAU

APPENDIX 3

FINAL DRAFT [2]

DRAFT RECOMMENDATION OF THE COMMITTEE OF MINISTERS ON THE ROLES AND RESPONSIBILITIES OF INTERNET INTERMEDIARIES RESPONSIBILITIES

Preamble

1. In line with the jurisprudence of the European Court of Human Rights (hereinafter "the Court"), Council of Europe member states have the obligation to secure to everyone within their jurisdiction the rights and freedoms contained in the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5, hereinafter "the Convention"), both offline and online. Access to the internet is a precondition for the exercise of Convention rights and freedoms on the Internet.

2. By enhancing the public's ability to seek, receive and impart information without interference and regardless of frontiers, the internet plays a particularly important role with respect to the right to freedom of expression. It also enables the exercise of other rights protected by the Convention and its Protocols, such as the right to freedom of assembly and association, the right to education, access to knowledge and culture, as well as participation in public and political debate and in democratic governance.

3. The protection of privacy and personal data is a foundation for the enjoyment and exercise of most of the rights and freedoms guaranteed in the Convention. However, the internet has facilitated an increase of privacy-related risks and infringements and has spurred the spread of certain forms of harassment, hatred and incitement to violence, in particular on the basis of gender, race and religion, which remain under-reported and rarely remedied or prosecuted. Moreover, the rise of the internet and related technological developments have triggered substantial challenges for the maintenance of public order and national security, for crime prevention and law enforcement, as well as for the protection of the rights of others, including intellectual property rights**.**

4. A wide, diverse and rapidly evolving range of actors, commonly referred to as internet intermediaries, facilitate interactions between natural and legal persons on the internet by offering and performing a variety of functions and services. Some connect users to the internet, enable the processing of information and data, or host web-based services, including for user-generated content. Others aggregate information and enable searches, and give access to, host and index content and services designed and/or operated by third parties. Some facilitate the sale of goods and services, including audio-visual services, and enable other commercial transactions, including payments.

[2] As contained in document MSI-NET(2016)05rev4 dated 19 September 2017

5. Intermediaries may carry out several functions in parallel. They may also moderate and rank content, including through automated processing of personal data, and may thereby exert forms of control which influence users' access to information online in ways comparable to media, or they may perform other functions that resemble those of publishers. Intermediary services may also be offered by traditional media, for instance, when space for user-generated content is offered on their platforms. The regulatory framework governing the intermediary function is without prejudice to the frameworks that are applicable to the other functions offered by the same entity.

6. The rule of law is a prerequisite for the protection and promotion of the exercise of human rights and for pluralistic and participatory democracy. Member states have the negative obligation to refrain from violating the right to freedom of expression and other human rights in the digital environment. They also have a positive obligation to protect human rights and to create an enabling and safe environment for everyone to participate in the public debate and to express their opinions and ideas without fear, including those that offend, shock, or disturb the state or any sector of the population. This positive obligation to ensure the exercise and enjoyment of rights and freedoms includes, due to the horizontal effects of human rights, the protection of individuals from actions of private parties by ensuring compliance with relevant legal and regulatory frameworks. It is further indispensable that due process guarantees are in place and access to effective remedies is facilitated vis-à-vis both states and intermediaries with respect to the services in question.

7. It is further essential to support initiatives promoting media and information literacy skills for accessing and managing the digital space. Such efforts should be implemented through various means, including formal and non-formal education, with a view to promoting the effective and equal enjoyment of the rights enshrined in the Convention by everyone without discrimination of any kind. Given the particularly high number of young and child users of the internet, the importance of empowering, protecting, and supporting children in their safe access to rights in the digital environment must be acknowledged throughout. To this end, sustained engagement is required to enhance skills among children, parents and educators on how to deal with an information and communications environment that provides access to degrading content of a sexual or violent nature which might be harmful.

8. The regulatory framework governing the services provided by or through intermediaries is diverse, multi-layered and continuously evolving. States are confronted with the complex challenge of regulating an environment in which private actors fulfil a crucial role in providing services with significant public service value. The task of regulation is further complicated by the global nature of the internet networks and services, by the diversity of intermediaries, by the volume of internet communication, and by the speed at which it is produced and processed. Owing to the fact that intermediaries operate or are used across many countries, including in a cloud-computing context, their actions may further have effects under several, sometimes conflicting, laws of different jurisdictions.

9. Internet intermediaries also develop their own rules, usually in form of terms of service or community standards that often contain content restriction policies. Moreover, intermediaries collect, generate, retain and process a wealth of information and data from and about users. These activities may interfere with, among other rights, the users' rights to privacy and freedom of expression. Effective reporting and complaints mechanisms may

be lacking, be insufficiently transparent and efficient, or be provided only through automated processes.

10. In line with the UN Guiding Principles on Business and Human Rights and the Protect, Respect and Remedy Framework, intermediaries should respect the human rights of their users and affected parties in all their actions. This includes the responsibility to act in compliance with applicable laws and regulatory frameworks. Owing to the multi-functionality of intermediaries, their corresponding duties and responsibilities and their protection under law, must be determined with respect to the specific services and functions that are performed.

11. A variety of network effects and mergers have led to the existence of fewer, larger entities that dominate the market in a manner that may jeopardise the opportunities for smaller intermediaries or start-ups and places them in positions of influence or even control of principal modes of public communication. The power of such intermediaries as protagonists of online expression makes it imperative to clarify their role and impact on human rights as well as their corresponding duties and responsibilities.

12. Against this background and in order to provide guidance to all relevant actors who are faced with the complex task of protecting and respecting human rights in the digital environment, the Committee of Ministers, under the terms of Article 15.b of the Statute of the Council of Europe, recommends that member states:

- implement the Guidelines included in this recommendation when developing and implementing legislative frameworks relating to internet intermediaries in line with their obligations under the Convention, the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108, hereinafter "Convention 108"), the Convention on Cybercrime (ETS No. 185, "the Budapest Convention"), the Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (ETS No. 201, "the Lanzarote Convention"), and the Convention on Preventing and Combating Violence against Women and Domestic Violence (ETS No. 210, "the Istanbul Convention) and promote them in international and regional forums that deal with the roles and responsibilities of internet intermediaries;

- take all necessary measures to ensure that internet intermediaries fulfill their responsibilities to respect human rights in line with the UN Guiding Principles on Business and Human Rights and the Recommendation CM/Rec (2016)3 of the Committee of Ministers to member states on human rights and business;

- in implementing the Guidelines, take due account of Committee of Ministers Recommendation 2016/5 on internet freedom; Recommendation 2016/1 on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality; Recommendation 2015/6 on the free, trans-boundary flow of information on the internet; Recommendation 2014/6 on a Guide to human rights for internet users; Recommendation 2013/1 on gender equality and media; Recommendation 2012/3 on the protection of human rights with respect to search engines; Recommendation 2012/4 on the protection of human rights with respect to social networking services; Recommendation 2011/7 on a new notion of media; Recommendation 2010/13 on the protection of individuals with regard to automatic processing of personal data in the context

of profiling; Recommendation 2007/16 on measures to promote the public service value of the internet; the 2017 Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data; the 2008 Guidelines for the cooperation between law enforcement and internet service providers against cybercrime, and the Human Rights guidelines for internet service providers, developed in 2008 by the Council of Europe in co-operation with the European Internet Service Providers Association which, as far as the responsibilities of internet service providers are concerned, are reinforced by this Recommendation.

- engage in a regular, inclusive and transparent dialogue with all relevant stakeholders, including from the private sector, public service media, civil society, education establishments and academia, with a view to sharing and discussing information and promoting the responsible use of emerging technological developments related to internet intermediaries that impact the exercise and enjoyment of human rights and related legal and policy issues;

- encourage and promote the implementation of effective age and gender-sensitive media and information literacy programmes to enable adults, young people and children to enjoy the benefits and minimise the exposure to risks of the online communications environment, in cooperation with all relevant stakeholders, including from the private sector, public service media, civil society, education establishments and academia.

Guidelines on the protection and promotion of human rights and fundamental freedoms with regard to internet intermediaries

1 – Duties and obligations of states

1.1 LEGALITY

1.1.1. Any request, demand or other action by public authorities addressed to internet intermediaries that interferes with human rights and fundamental freedoms must be prescribed by law and must constitute a necessary and proportionate measure in a democratic society. All powers of public authorities in relation to internet intermediaries must be prescribed by law and exercised within the limits conferred by law. States should not use informal means to circumvent the guarantees offered by formal legal proceedings.

1.1.2. Laws, regulations and policies applicable to internet intermediaries, regardless of their objective or scope of application, including commercial and non-commercial activities, shall effectively safeguard human rights and fundamental freedoms, and shall maintain adequate guarantees against arbitrary application in practice.

1.1.3. States shall not seek to absolve themselves from their ultimate obligation to protect human rights and fundamental freedoms in the digital environment. All regulatory frameworks, including self- or co-regulatory approaches, must include effective oversight mechanisms to comply with that obligation and must be accompanied by appropriate legal redress opportunities.

1.1.4. The process of enacting legislation or regulations applicable to internet intermediaries should be transparent and inclusive. States should regularly consult with all relevant stakeholders with a view to ensuring that an appropriate balance is struck between the public interest, the interests of the users and affected parties, and the interest of the intermediary. Before adopting legislation or regulations, states should conduct human rights impact assessments to understand potential negative impacts on human rights in order to prevent or mitigate these.

1.1.5. States shall ensure that legislation, regulation, and policies related to internet intermediaries are interpreted, applied and enforced without discrimination, also taking into account multiple and intersecting forms of discrimination. The prohibition of discrimination may in some instances require special measures to address specific needs or correct existing inequalities. States should further take into account the substantial differences in size, function and organisational structure of intermediaries when developing, interpreting and applying the legislative framework in order to prevent possible discriminatory effects.

1.1.6. States should ensure that legislation, regulation and policies relating to internet intermediaries are effectively implementable and enforceable and that they do not unduly restrict the operation and free flow of trans-border communication.

1.2. LEGAL CERTAINTY AND TRANSPARENCY

1.2.1. Any legislation applicable to internet intermediaries and to their relations with states and users must be accessible and predictable. All laws should be clear and sufficiently precise to enable intermediaries, users and affected parties to regulate their conduct. The laws should create a safe and enabling online environment for private communications and public debate and should comply with relevant international standards.

1.2.2. Any legislation must include clear limits to the powers, discretionary or non-discretionary, granted to public authorities in relation to internet intermediaries, particularly when exercised by the executive branch and specifically by law enforcement. The law must indicate the scope of such discretion to protect against arbitrary application.

1.2.3. States should make publicly available, in a timely and regular manner, comprehensive information on the number, nature and legal basis of restrictions of human rights, such as regarding content restrictions or disclosure of personal data, that they have applied in a certain period through requests addressed to intermediaries, including those based on international mutual legal assistance treaties, and on actions taken as a result of those requests. States should require intermediaries to disclose clear (easily accessible and machine-readable) and meaningful information about interferences with the exercise of rights and freedoms in the digital environment, whether based on court or administrative orders, private complainants' requests, or enforcement of their own content restriction policies.

1.2.4. With a view to avoiding legal uncertainty and conflicts of laws, states should commit to cooperating with each other and with all relevant stakeholders in cases where different laws apply, and should support the development of common approaches and jurisdictional principles, including through appropriate non-state forums.

1.3. SAFEGUARDS FOR FREEDOM OF EXPRESSION

1.3.1. Any request, demand or other action by public authorities addressed to internet intermediaries to restrict access (including blocking or removal of content), or any other measure that interferes with the right to freedom of expression, must be prescribed by law, pursue one of the legitimate aims foreseen in Article 10 of the Convention, be necessary in a democratic society and proportionate to the aim pursued. State authorities must carefully evaluate possible, including unintended, impacts of any restrictions before and after applying them, while seeking to apply the least intrusive measure necessary to meet the policy objective.

1.3.2. State authorities should obtain an order by a judicial authority or other independent administrative authority whose decisions are subject to judicial review when demanding intermediaries to restrict access to content. All exceptions must also be clearly prescribed by law, pursue one of the legitimate aims foreseen in Article 10, be necessary in a democratic society and proportionate to the aim pursued.

1.3.3. When internet intermediaries restrict access to third-party content, state authorities should ensure that intermediaries provide effective redress mechanisms and adhere to due process guarantees. When intermediaries remove content based on their own terms of service, state authorities should not consider this as a form of control that makes them liable for the third-party content they give access to.

1.3.4. State authorities should consider the adoption of appropriate legislation to prevent strategic lawsuits against public participation (SLAPP) or abusive and vexatious litigation against users, content providers and intermediaries which is intended to curtail the right to freedom of expression.

1.3.5. State authorities should not directly or indirectly impose a general obligation on intermediaries to monitor content which they merely give access to, or which they transmit or store, be it by automated means or not. When addressing any request to internet intermediaries or promoting, alone or with other states or international organisations, co-regulatory approaches by internet intermediaries, state authorities should avoid any action that may lead to general content monitoring. They should further consider that any content monitoring performed through automated means is unable to assess context properly. All co-regulatory approaches must comply with rule of law and transparency safeguards.

1.3.6. The imposition of disproportionate sanctions on intermediaries for non-compliance with regulatory frameworks is likely to lead to restriction of lawful content. It therefore has a chilling effect for the right to freedom of expression. In addition, content monitoring risks interfering with user's enjoyment of their right to privacy.

1.3.7. States should ensure in law and in practice that intermediaries are not held liable for third-party content, to which they merely give access to or which they transmit or store. State authorities may hold intermediaries co-responsible with respect to content that they store, if they do not act expeditiously to restrict access to content or services as soon as they become aware of their illegal nature, including through notice-based procedures. State authorities should ensure that notice-based procedures are not designed in a manner that incentivises the take-down of legal content, such as through inappropriately short timeframes. Notices should contain sufficient information for intermediaries to act upon. Notices submitted by states should be based on their own assessment of the illegality of the notified content. Content restrictions should allow notice of such restriction as early as possible to the content producer/issuer, unless this interferes with ongoing law enforcement activities. Information should also be made available to users seeking access to the content, in accordance with applicable data protection laws.

1.3.8. In order to ensure that content identical to that which has previously been determined to be illegal by a judicial authority or other independent administrative authority whose decisions are subject to judicial review, is effectively prevented from being accessed, states should co-operate closely with intermediaries to secure the restriction of such content in line with the principles of legality, necessity and proportionality. Such restrictions should not prevent the legitimate use of identical or similar content in other contexts.

1.3.9. In cases where the function of intermediaries consists of producing or managing content available on their platforms or where intermediaries perform curatorial or editorial-like functions, including through operation of algorithms, state authorities should apply an approach that is graduated and differentiated, in line with Recommendation CM/Rec(2011)7 of the Committee of Ministers to member states on a new notion of media. States should determine corresponding levels of protection as well as duties and responsibilities according to the role that intermediaries play in content production and dissemination processes, while paying due attention to their obligation to protect and promote pluralism and diversity in the online distribution of content.

1.3.10. When determining the applicable duties and responsibilities of intermediaries who are engaged in curatorial or editorial-like functions, including the production and dissemination of content, states should encourage appropriate self-regulatory or the development of co-regulatory mechanisms, taking due account of the extent that their action may negatively affect pluralism and diversity of online content, as well as the ability of the intermediary to provide services of public value, such as platforms for public discourse and democratic debate, as protected by Article 10 of the Convention.

1.4. SAFEGUARDS FOR PRIVACY AND DATA PROTECTION

1.4.1. Any demand or request by state authorities addressed to internet intermediaries to access, collect or intercept personal data of their users, including for criminal justice purposes, or any other measure which interferes with the right to privacy, must be prescribed by law, must pursue one of the legitimate aims foreseen in Article 8 of the Convention and Article 9 of Convention 108, and must be used only when it is necessary and proportionate in a democratic society to the aim pursued. The protection of the right to privacy and data protection extends to devices used to access the internet or store data.

1.4.2. State authorities should ensure that their legal frameworks and the ensuing policies and practices of intermediaries uphold the principles of data processing (lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage time limitations, and data security, including integrity and confidentiality,) and guarantee the rights of the data subject in full compliance with Convention 108, providing also for the oversight of an independent authority within the meaning of Article 1 of the Additional Protocol concerning Supervisory Authorities and Trans-border Data Flows.

1.4.3. State authorities should protect the right to confidentiality of all private communications facilitated by internet intermediaries, extending to the content of the communication as well as metadata, and should ensure that appropriate levels of data protection and respect for privacy are also guaranteed in situations of trans-border data flows.

1.4.4. Surveillance measures undertaken by states, whether in co-operation with internet intermediaries or not, must be targeted, precisely defined, and must comply with Article 8 of the Convention as well as Article 9 of Convention 108. They must in

particular be mandated by law, necessary in a democratic society and proportionate to the aim pursued, and they must include sufficient oversight, procedural safeguards and redress mechanisms. All surveillance must be authorised by a judicial authority or other independent administrative authority whose decisions are subject to judicial review.

1.4.5. State authorities should ensure that appropriate complementary safeguards, such as explicit consent of the data subject, apply to the automatic processing of special categories of data as defined in Article 6 of Convention 108.

1.5. ACCESS TO AN EFFECTIVE REMEDY

1.5.1. States should guarantee accessible and effective judicial and non-judicial procedures that ensure the impartial review of all claims of violations of Convention rights in the digital environment, such as the right to privacy, right to freedom of expression, or the right not to be discriminated against, in compliance with Article 6 of the Convention.

1.5.2. States should guarantee an effective remedy for all violations of human rights and fundamental freedoms by internet intermediaries, in compliance with Article 13 of the Convention. They should further ensure that intermediaries provide access to prompt, transparent and effective reviews of user or affected party grievances and alleged terms of service violations, and provide for effective remedies. These may include various forms, such as restoration of content, apology, rectification and damages. Judicial review must remain available, when internal and alternative dispute settlement mechanisms prove insufficient or where the affected parties opt for judicial redress or appeal.

1.5.3. States should proactively seek to reduce all legal, practical or other relevant barriers that could lead to a denial of access to an effective remedy for grievances of users, affected parties and internet intermediaries.

1.5.4. States should support age- and gender-sensitive media and information literacy promotion activities to ensure that all users are effectively made aware of their rights and freedoms, in particular regarding their right to access to an effective remedy vis-à-vis both state authorities and internet intermediaries. The promotion of media and information literacy should encompass education about the rights of all stakeholders, including other users and affected parties.

2 - Responsibilities of internet intermediaries with regard to human rights and fundamental freedoms

2.1. RESPECT FOR HUMAN RIGHTS AND FUNDAMENTAL FREEDOMS

2.1.1. Internet intermediaries should in all their actions respect the internationally recognised human rights and fundamental freedoms of their users and of other parties who are affected by their activities. This responsibility, in line with the UN Guiding Principles on Business and Human Rights, exists independently of the states' ability or willingness to fulfil their own human rights obligations.

2.1.2. The responsibility of intermediaries to respect human rights and to employ adequate measures applies regardless of their size, sector, operational context, ownership structure, or nature. The scale and complexity of the means through which intermediaries meet their responsibilities may vary, however, taking into account the severity of the possible human rights impact of the services provided by the intermediary. The higher the impact and the potential damage to the objects of legal protection and the higher the value of the services for the exercise of human rights, the greater the precautions that the intermediary must employ when developing and applying policies, community standards, and codes of ethics aiming, notably, at the prevention of the spread of abusive language and imagery, of hatred and of incitement to violence.

2.1.3. Any interference by intermediaries with the free and open flow of information and data should be based on clear and transparent policies and must be limited to specific legitimate purposes, such as to restrict access to content that has been determined as unlawful by a judicial authority or other independent administrative authority whose decisions are subject to judicial review, or in accordance with their own content restriction policies or codes of ethics.

2.1.4 Internet intermediaries should carry out regular due diligence assessments of their compliance with the responsibility to respect human rights and fundamental freedoms and with their applicable duties. To this end, they should conduct assessments of the direct and indirect human rights impacts of their current and possible future policies, products and services on users and affected parties, and ensure appropriate follow-up to these assessments by acting upon the findings, and monitoring and evaluating the effectiveness of identified responses. Intermediaries should conduct these assessments as openly as possible and encourage active user engagement. In all their actions they should be mindful of the public service value of the services they deliver and should seek to avoid and mitigate any adverse effects on the effective exercise of rights by their users or affected parties.

2.1.5. Internet intermediaries should seek to provide their products and services without discrimination. They should seek to ensure that their actions do not have direct or indirect discriminatory effects or harmful impacts on their users or other parties affected by their actions, including on those who have special needs or disabilities or may face structural inequalities in their access to human rights. Intermediaries should further take reasonable and proportionate measures to ensure that their

terms of service agreements, community standards and codes of ethics are applied and enforced consistently and in compliance with applicable due process safeguards. The prohibition of discrimination may under certain circumstances require that intermediaries make special provisions for certain users or groups of users in order to correct existing inequalities.

2.2. TRANSPARENCY AND ACCOUNTABILITY

2.2.1. Internet intermediaries should ensure that all terms of service agreements and policies specifying the rights of users and all other standards and practices for content moderation and the processing and disclosure of user data are publicly available in clear, plain language and accessible formats. When operating globally, intermediaries should translate such documents into the languages that their users and affected parties understand. Users should be notified in advance of all changes in relevant policies regarding their terms of service and operating conditions as applicable and without delay, and in formats that they can easily access and understand, including explanatory guides.

2.2.2. The process of developing and applying terms of service agreements, community standards and content restriction policies should be transparent, accountable and inclusive. Intermediaries should seek to collaborate and negotiate with consumer associations, human rights advocates, and other organisations representing the interests of users and affected parties, as well as with data protection authorities before adopting and modifying their policies. Intermediaries should seek to empower their users to engage in processes of evaluating, reviewing and revising, where appropriate, intermediaries' policies and practices.

2.2.3. Internet intermediaries should clearly and transparently provide meaningful public information about the operation of automated data processing techniques in the performance of their functions, including the operation of algorithms that facilitate searches based on user profiling or the distribution of algorithmically selected and personalised content, such as news. This should include information on which data is being processed, which criteria are used, and for what purpose the processing takes place.

2.2.4. Intermediaries should regularly publish transparency reports that provide clear (easily accessible and machine-readable) and meaningful information about all interference and all requests for such interference with the free and open flow of information and ideas and related to requests for data access and preservation, whether based on court orders, international mutual legal assistance treaties, private complainant's requests or enforcement of their own content restriction policies.

2.3. CONTENT MODERATION

2.3.1. Internet intermediaries should respect the rights of users to receive and impart information, opinions and ideas. They should not on a general basis monitor content to which they merely give access, or which they transmit or store, as a result of a state order or request. Any measures taken to restrict access (including

blocking or removal of content) as a result of a state order or request must be necessary and be implemented through the least restrictive means, following a careful assessment of their effectiveness and proportionality to the legitimate aim pursued.

2.3.2. When restricting access to content in line with their own content restriction policies, intermediaries should do so in a transparent and non-discriminatory manner. All content restrictions must be performed by the least restrictive technical means and must be only as broad and maintained for as long as strictly necessary to avoid the collateral restriction or removal of legal content.

2.3.3. Any restriction of content must be limited in scope to the precise remit of the order or request and should be accompanied with information to the public, explaining which content has been restricted and on what legal basis. Notice should also be given to the user and, as appropriate, other affected parties, including information on procedural safeguards, opportunities for adversarial procedures for both parties as appropriate, and available redress mechanisms.

2.3.4. All staff of intermediaries who are engaged in content moderation should be given adequate initial and on-going training on the applicable laws, international human rights standards, their relationship with the intermediaries' terms of service and their internal standards, as well as on the action to be taken in case of conflict. Such training may be provided internally or externally, including through intermediary associations, and should in its scope correspond to the importance of the intermediaries' role and the impact that their actions may have on the ability of users to exercise their freedom of expression. Staff should further be provided with appropriate working conditions. This includes the allocation of sufficient time for deciding on the legality of content and opportunities to seek professional support and qualified legal advice where necessary.

2.3.5. Given that automated means of content identification used to prevent the reappearance of specific items of previously restricted content have limited ability to assess context, intermediaries should carefully assess the human rights impact of automated content management, and should ensure human review where appropriate. They should take into account the risk of over- and under-blocking as a result of inexact algorithmic systems, and the effect this may have on the services that they provide for public debate. Restrictions of access to identical content should not prevent the legitimate use of such content in other contexts.

2.3.6. In cases where content is restricted by intermediaries in line with their own content restriction policies because it contains an indication of a serious crime under international law, restriction must be accompanied by adequate measures to ensure that evidence is retained for effective criminal law investigations. If intermediaries have specific knowledge of such restricted content, they should report this to a law enforcement authority without undue delay.

2.4. USE OF PERSONAL DATA

2.4.1. Intermediaries should not disclose personal data unless required by law or requested to do so by a judicial authority or other independent administrative authority whose decisions are subject to judicial review that has determined that the disclosure is consistent with applicable laws and standards, necessary in a democratic society and proportionate to the legitimate aim pursued.

2.4.2. Internet intermediaries should limit the processing of personal data from users to what is necessary in the context of a clearly defined purpose, which is explicitly communicated to all users in a proactive manner. The processing, including collection, retention, aggregation, linking or sharing of personal data must be based on the free, specific, informed and unambiguous consent of the user, with respect to a specific purpose, or on another legitimate basis laid down by law, as prescribed by Convention 108. Complementary safeguards, such as explicit consent, should be applied to the automatic processing of special categories of data, as defined in Article 6 of Convention 108.

2.4.3. Intermediaries should minimise the processing of personal data in light of the purposes for which they are processed. 'Privacy by default' and 'privacy by design' principles should be applied at all stages with a view to prevent or minimise the risk of interference with the rights and fundamental freedoms of users. User data should only be aggregated and migrated across multiple devices or services following the free, specific, informed and unambiguous consent of users. Users should have the option of using a service without agreeing to such combining of their data.

2.4.4. Users have the right to access their personal data and to obtain correction of it, and they should be informed about it in clear and plain language. They should further be informed clearly about the conditions under which they may exercise the right to data erasure, to object to the processing of data, and to withdraw consent provided for the processing of personal data, following which all processing based on the consent of the user should be terminated.

2.4.5. Intermediaries should act in line with applicable legal conditions and safeguards regardless of where the collection of data has occurred and including with respect to trans-border data flows.

2.4.6. Any tracking and profiling of users by intermediaries should be fully transparent towards users. In order to protect their users' identity, internet intermediaries should not employ profiling and digital tracking techniques that infringe on the user's exercise of human rights. Intermediaries should seek to protect their users from tracking and profiling by third parties. Adequately trained staff should oversee all matters related to the disclosure of user data to third parties in line with the intermediaries' responsibilities and duties under international personal data protection and privacy standards. A person subjected to a decision that is taken on the basis of profiling or affected by legal consequences stemming from that decision, should be able to object to that decision.

2.5. ACCESS TO AN EFFECTIVE REMEDY

2.5.1. Internet intermediaries should make available – online and offline – effective remedies and dispute resolution systems that provide prompt and direct redress in cases of user, content provider and affected party grievances. While the complaint mechanisms and their procedural implementation may vary with the size, impact and role of the internet intermediary, all remedies must allow for an impartial and independent review of the alleged violation. These should - depending on the violation in question - include inquiry, explanation, reply, correction, apology, reinstatement, reconnection and compensation.

2.5.2. Complaint mechanisms, including notice-based procedures, should comply with due process safeguards and should be accessible, equitable, rights-compatible, affordable and transparent. They should further include in-built safeguards to avoid conflicts of interest when the company is directly administering the mechanism, for example, by involving oversight structures. Complaints mechanisms should be handled without unwarranted delays and should not negatively impact the opportunities for complainants to seek recourse through national, including judicial, review mechanisms.

2.5.3. Intermediaries should ensure that all users and other parties affected by their actions have full and easy access to transparent information in clear and easily understandable language about applicable complaints mechanisms, the various stages of the procedure, indicative time frames, and expected outcomes.

2.5.4. Intermediaries should not include in their terms of service waivers of rights or hindrances to the effective access to remedies, such as mandatory jurisdiction outside of a user's country of residence or non-derogable arbitration clauses.

2.5.5. Intermediaries should seek to provide access to alternative review mechanisms that can facilitate the resolution of disputes that may arise between users. Intermediaries should not, however, make alternative dispute mechanisms obligatory as the only means of dispute resolution.

2.5.6. Intermediaries should engage in dialogue with consumer associations, human rights advocates and other organisations representing the interests of users and affected parties, as well as with data protection authorities, to ensure that their complaint mechanisms are designed, implemented, and evaluated through participatory processes. They should further regularly analyse the frequency, patterns and causes of complaints received in order to learn lessons for improving their policies, procedures and practices and for preventing future grievances.

2.5.7. Intermediaries should engage in and promote targeted age- and gender-sensitive efforts to promote the awareness of all users of their rights and freedoms in the digital environment, both vis-à-vis states and intermediaries, including in particular information about applicable complaints mechanisms and procedures. The promotion of media and information literacy should encompass education about the rights of all stakeholders, including other users and affected parties.

* * *

APPENDIX 4

FINAL[3] DRAFT STUDY ON THE HUMAN RIGHTS DIMENSIONS OF AUTOMATED DATA PROCESSING TECHNIQUES (IN PARTICULAR ALGORITHMS) AND POSSIBLE REGULATORY IMPLICATIONS

[3] As contained in document MSI-NET(2016)06rev3 dated 6 October 2017

In the terms of reference for the Steering Committee on Media and Information Society (CDMSI) for the biennium 2016 – 2017, the Committee of Ministers asked the CDMSI to "undertake work to study the human rights dimensions of automated data processing techniques (in particular algorithms) and their possible regulatory implications" and approved the committee of experts on internet intermediaries (MSI-NET) as a subordinate structure to facilitate the work of the CDMSI. In its first meeting on 17-18 March 2016, the expert committee decided to appoint Benjamin Wagner as rapporteur for the study, while other members of the MSI-NET expressed the wish to support the rapporteur in a small working group.

COMPOSITION OF THE MSI-NET

Wolfgang SCHULZ – Professor, Faculty of Law, University of Hamburg / Hans-Bredow-Institut (Chair)

Karmen TURK – Partner at Trinity Tallinn – Estonia (Vice-Chair)

Bertrand de la CHAPELLE – Co-founder/Director of the Internet & Jurisdiction - France

Julia HÖRNLE – Professor of Internet Law, Queen Mary University of London

Tanja KERŠEVAN-SMOKVINA – Principal Advisor to Director General, Agency for Communication Networks and Services - Slovenia

Matthias KETTEMANN – Postdoc Fellow, Cluster of Excellence "Normative Orders" University of Frankfurt/Main

Dörte NIELANDT **-** Division VI A3 (Legal framework for digital services, media industry), Federal Ministry for Economic Affairs and Energy – Germany

Arseny NEDYAK – Deputy Director, Department of Media State Policy, Ministry of Telecommunication - Russian Federation

Pēteris PODVINSKIS – Ministry of Foreign Affairs, International Organisations Directorate, Department for Public Policy related to Internet – Latvia

Thomas SCHNEIDER – Deputy Director of International Affairs, International Information Society Coordinator, Federal Department of the Environment, Transport, Energy and Communication DETEC, Federal Office of Communications (OFCOM) - Switzerland

Sophie STALLA-BOURDILLON – Associate Professor in Information Technology / Intellectual Property Law, Director of ILAWS, Southampton Law School University of Southampton

Dirk VOORHOOF – Lecturer European Media Law, UCPH (Copenhagen University) / Professor at Ghent University / Member of the CMPF Scientific Committee Centre for Media Pluralism and Press Freedom

Benjamin WAGNER – Assistant Professor, Institute for Management Information Systems, Vienna University of Economics and Business

Table of Contents

I. INTRODUCTION

What information is made available to users on their Facebook newsfeeds? On what basis is a person's risk profile determined and what profiles provide best chances for obtaining health insurance, or employment, or for being regarded a potential criminal or terrorist? Automated data processing techniques, such as algorithms, do not only enable internet users to seek and access information, they are also increasingly used in decision-making processes, that were previously entirely in the remit of human beings. Algorithms may be used to prepare human decisions or to take them immediately through automated means. In fact, boundaries between human and automated decision-making are often blurred, resulting in the notion of 'quasi- or semi-automated decision-making'.

The use of algorithms raises considerable challenges not only for the specific policy area in which they are operated, but also for society as a whole. How to safeguard human rights and human dignity in the face of rapidly changing technologies? The right to life, the right to fair trial and the presumption of innocence, the right to privacy and freedom of expression, workers' rights, the right to free elections, even the rule of law itself are all impacted. Responding to challenges associated with 'algorithms' used by the public and private sector, in particular by internet platforms is currently one of the most hotly debated questions.

There is an increasing perception that "software is eating the world" (Andreessen 2011), as human beings feel that they have no control over and do not understand the technical systems that surround them. While disconcerting, it is not always negative. It is a by-product of this phase of modern life in which globalised economic and technological developments produce large numbers of software-driven technical artefacts and "coded objects" (Kitchin and Dodge 2011) embed key human rights relevant decision-making capacities. Which split-second choices should a software-driven vehicle make if it knows it is going to crash? Is racial, ethnic or gender bias more likely or less likely in an automated system? Are societal inequalities merely replicated or amplified through automated data processing techniques?

Historically, private companies decided how to develop software in line with the economic, legal and ethical frameworks they deemed appropriate. While there are emerging frameworks for the development of systems and processes that lead to algorithmic decision-making or for the implementation thereof, they are still at an early stage and do usually not explicitly address human rights concerns. In fact, it is uncertain whether and to what extent existing legal concepts can adequately capture the ethical challenges posed by algorithms. Moreover, it is unclear whether a normative framework regarding the use of algorithms or an effective regulation of automated data processing techniques is even feasible as many technologies based on algorithms are still in their infancy and a greater understanding of their societal implications is needed. Issues arising from use of algorithms as part of the decision-making process are manifold and complex. At the same time, the debate about algorithms and their possible consequences for individuals, groups and societies is at an early stage. This should not, however, prevent efforts towards understanding what algorithms actually do, which consequences for society flow from them and how possible human rights concerns could be addressed.

This report identifies a number of human rights concerns triggered by the increasing role of algorithms in decision-making. Depending on the types of functions performed by algorithms and the level of abstraction and complexity of the automated processing that is used, their impact on the exercise of human rights will vary. Who is responsible when human rights are infringed based on algorithmically-prepared decisions? The person who programmed the algorithm, the operator of the algorithm, or the human being who implemented the decision? Is there a difference between such a decision and a human-made decision? What effects does it have on the way in which human rights are exercised and guaranteed in accordance with well-established human rights standards, including rule of law principles and judiciary processes?

Challenges related to the human rights impact of algorithms and automated data processing techniques are bound to grow as related systems are becoming increasingly complex and interact with each other's outputs in ways that become progressively impenetrable to the human mind. This report does not intend to comprehensively address all aspects related to the human rights impacts of algorithms but rather seeks to map out some of the main current concerns from the Council of Europe's human rights perspective, and to look at possible regulatory options that member states may consider to minimise adverse effects, or to promote good practices. A number of related themes will require more detailed research to more systematically assess their challenges and potential from a human rights point of view, including questions related to big data processing, machine learning, artificial intelligence and the Internet of things.

II. THE SCOPE OF THE REPORT

When assessing automated data processing techniques and the algorithms they use, it is important to be clear on what types of algorithms are being discussed. This study will build on existing well-established definitions, in particular the work of Tarleton Gillespie (2014), Nicholas Diakopoulos (2015) and Frank Pasquale (2015). It is further important to keep in mind that the term 'algorithm' is applied widely and has a varied set of meanings, depending on whether it is used in the computer science community, among mathematicians and information technologists, in communication and cultural media studies or in public, including political and social, discourse. Mapping out the human rights dimensions of algorithms must also consider the divergence between formal definitions of algorithms and the popular usage of the term. In fact, many of the debates about algorithms focus less on algorithms themselves and more broadly on the role of technology in society (Bucher 2016).

The report's basic approach starts from Tarleton Gillespie's assumption that "algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations. The procedures name both a problem and the steps by which it should be solved." (Gillespie 2014:167) Algorithms are thus perceived as "a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome" (Diakopoulos 2015:400).

This report will not discuss algorithms that automate manufacturing processes or perform other such routine tasks. Rather, it seems reasonable to limit the discussion to algorithms that are digital and affect the public at large, thus focussing mainly on algorithmic decision-making that has implications for human rights. Without being exhaustive or aiming to predict all potential properties of algorithms and their decision-making in the future, the following characteristics of algorithms that engage in automated data processing and (semi-)automated decision making are considered key issues from a human rights perspective for this report: automation, data analysis, and adaptability. In addition, algorithms and data processing techniques are produced by human beings and operated by human beings. Their implications can therefore not be understood without acknowledgement of the social constructs that exist around them.

1. AUTOMATION

Automation is one of the core characteristics associated with algorithmic decision-making. The ability of automated computing systems to replace human beings in a growing number of situations is a key characteristic of the practical implementation of algorithms. The reasons for replacing human beings with automated computing systems can be usually traced back to issues of large-scale data processing, speed, volume and scale of decision-making, and in many cases to expectations of lower error rates compared to human beings. Automated decision-making algorithms are used across a variety of domains, from simplistic models that help online service providers to carry out operations on behalf of their users (Kim et al., 2014) to more complex profiling algorithms (Hildebrandt, 2008) that filter systems for personalised content. Automated, algorithmic decision-making is usually difficult to predict for a human being and its logic will be difficult to explain after the fact.

2. DATA ANALYSIS

Data analysis algorithms are applied to large amounts of data to find patterns of correlation within datasets without necessarily making a statement on causation (Grindrod, 2014). Their use of data mining and pattern recognition without "understanding" their correlation or causal relationships may lead to errors and raise concerns about data quality. These algorithms replicate the functions previously performed by human beings but involve a quantitatively and qualitatively different decision-making logic to much larger amounts of data input.

It is noteworthy that effects of automated decision-making can be framed as interplay of the applied analytics (based on algorithms) and the data sets used. An assessment of human rights impacts should take both elements into account since, to take an example, bias may be hidden in the data set and thus not found by analysing the algorithm itself. When assessing the human rights impacts of algorithms, it further must be considered that designers of algorithmic systems have varying levels of discretion when deciding, for instance, what training data to use or how to respond to false positives, and that the power of the operator of the algorithm may lie in his or her knowledge of the structure of the data set, rather than in insight into the exact workings of the algorithms.

3. ADAPTABILITY

Adaptability is demonstrated in self-learning algorithms that use data to develop novel patterns and knowledge, and to generate new decision-making rules through machine learning techniques (Williamson 2016). By adopting various learning styles, algorithms model problems based on data sets and produce new solutions that may be impossible for a human being to grasp. Essentially through constant trial and error techniques, algorithms detect patterns in existing data, identify similar patterns in future data and make data driven predictions.

Machine learning techniques are used, among others, in search engines that auto-correct spelling mistakes, as well as in more complex fields, such as fraud prevention, risk analysis, advancement of insight into customer behaviour and enhancement of medical science.

The predictability of an algorithm's outcome by the operator is important when considering its accountability and the design of adequate governance structures. The progress of "deep learning" technologies may lead to more systems that cannot be understood by using the mental model of mechanical machines. There is considerable debate in the academic community about the degree to which such systems can be made intelligible to human beings and what consequences such intelligibility could have.[4]

4. SOCIAL CONSTRUCTS AROUND ALGORITHMS

While algorithmic decision-making is increasingly adept at replacing human decision-making, important elements (such as discretion) of decision-making processes cannot be automated and often become lost when human decision-making processes are automated (Spiekermann 2015). Without judging their respective "quality", decision-making processes by humans and by algorithms are fundamentally and categorically different, make different mistakes, and might have different outcomes and therefore consequences. While society and governments have considerable experience understanding human decision-making and its failures, they are only beginning to understand the flaws, limitations and boundaries of algorithmic decision-making. One key challenge is the frequent perception that algorithms are able to create neutral, non-discriminatory and independent predictions about future events. The frenzy surrounding the operation of Google Flu trends in 2011, which later turned out to be unjustified as its prediction ability was far lower than had been claimed, is one example of the on-going struggle with assertions regarding the accuracy of predictive algorithms (Lazer et al. 2014; Lazer and Kennedy 2015). This challenge, however, relates less to algorithms as a tool and more to their design as well as human perception and interpretation of their implementation and results. Thus, the key to promoting human rights compliance in the use of algorithms may be to understand what algorithms can and cannot achieve and not to let their use be dictated merely by considerations of efficiency or effectiveness alone.

---

[4] See, for example, Yuan Stevens, 'The Promises and Perils of Artificial Intelligence: Why Human Rights and the Rule of Law Matter', https://medium.com/@ystvns/the-promises-and-perils-of-artificial-intelligence-why-human-rights-norms-and-the-rule-of-law-40c57338e806, September 5, 2017.

Traditionally, developers have programmed algorithms by hand "to process and transform input data into a desired output, based on specified calculations" (Gillespie 2014). With technological evolution, however, the socio-technical systems like algorithms are becoming increasingly opaque. This is not technically necessary, but rather a frequent design choice leading to algorithmic systems whose inner workings cannot be made transparent or accountable to the outside world. Even when a human being formally takes a decision, for instance the decision to remove certain content from a social media platform (see below 3.), the human being may often be led to 'rubber stamp' an algorithmically prepared decision, not having the time, context or skills to make an adequate decision in the individual case. Thus, while it may seem logical to draw a distinction between fully automated decision-making and semi-automated decision-making, in practice the boundaries between the two are blurred. In neither case will a human being be able to provide a reasoned argument why a certain decision needed to be taken in the specific case. This has repercussions for the right of the concerned individual to seek an effective remedy against a human rights violation (see below 5.).

It should be noted that algorithms as discussed here do not exist meaningfully without interaction with human beings. Mathematic or computational constructs do not by themselves have adverse human rights impacts but their implementation and application to human interaction does. Technologies – in their application to human interaction - are deeply social constructs (Winner 1980, 1986) with considerable political implications (Denardis 2012). While a decision-making software, for example, may be "biased but ambivalent" (McCarthy 2011:90), it has no meaning without a social system around it which provides meaning and impact.

It is thus too simple to blame the algorithm or to suggest to no longer resort to computers or computing. Rather, it is the social construct and the specific norms and values embedded in algorithms that need to be questioned, critiqued and challenged. Indeed, it is not the algorithms themselves but the decision-making processes around algorithms that must be scrutinised in terms of how they affect human rights.

The question whether the quality of decisions with respect to human rights differs between those taken by human and those taken by or based on algorithmic calculation can only be answered if we know how human decision-making functions. There is evidence that it is special (Tversky and Kahneman 1974) as regards the use of tacit knowledge and tacit norms (Schulz and Dankert 2016). This, to take an example, enables humans to notice exceptional cases where the application of a rule is not appropriate even though the case falls within its scope. The increasing importance of algorithms in decision-making calls for a better understanding of the design and characteristics of decision making procedures.

III. IMPACTS OF ALGORITHMS ON HUMAN RIGHTS

Reservations against algorithms and automated data processing techniques usually point to their opacity and unpredictability.[5] Beyond these general concerns, however, there is an increasing awareness that specific human rights are particularly affected. These are referenced below with practical examples as to how and why the use of algorithms may lead to rights violations or may otherwise undermine the effective enjoyment of these human rights.

1. FAIR TRIAL AND DUE PROCESS

The trend towards using automated processing techniques and algorithms in crime prevention and the criminal justice system is growing. Indeed, there may be some benefits in such use as massive data sets may be processed more speedily or flight risks assessed more accurately. Moreover, the use of automated processing techniques for the determination of the length of a prison sentence may allow more even approaches to comparable cases. Yet, growing national security concerns have led to ever more ambitious applications of new technologies. Following a string of terrorist attacks in the US and Europe, politicians have called for online social media platforms to use their algorithms to identify potential terrorists and to take action accordingly (Rifkind 2014; Toor 2016). Some such platforms are already using algorithms to identify accounts that generate extremist content. Apart from the significant impact such application of algorithms has for the freedom of expression (see below 3.), it also raises concerns for fair trial standards contained in Article 6 of the ECHR, notably the presumption of innocence, the right to be informed promptly of the cause and nature of an accusation, the right to a fair hearing and the right to defend oneself in person. Concerns may also arise with respect to Article 5 of the ECHR, which protects against arbitrary deprivation of liberty, and Article 7 (no punishment without law). In the field of crime prevention, the main policy debates regarding the use of algorithms relate to predictive policing. This approach goes beyond the ability of human beings to draw conclusions from past offences to predict possible future patterns of crime. It includes developed automated systems that predict which individuals are likely to become involved in a crime (Perry 2013), or are likely to become repeat offenders and therefore require more severe sentencing.[6] It also includes systems meant to predict where crime is likely to take place at a given time which are then used for prioritising police time for investigations and arrests. Such approaches may be highly prejudicial in terms of ethnic and racial backgrounds and therefore require scrupulous oversight and appropriate safeguards. Often the systems are based on existing police

[5]See Tim O'Reilly, "The great question of the 21st century: Whose black box do you trust?", 13 September 2016, available at: https://www.linkedin.com/pulse/great-question-21st-century-whose-black-box-do-you-trust-tim-o-reilly?trk=eml-b2_content_ecosystem_digest-hero-22-null&midToken=AQGexvwxq0Q3iQ&fromEmail=fromEmail&ut=2SrYDZ8lkCS7o1 (last visited on 25 September 2017).

[6] See also Article 19, *Algorithms and Automated Decision-Making in the Content of Crime Prevention: A Briefing paper,* 2016.

databases that intentionally or unintentionally reflect systemic biases.[7] Depending on how crimes are recorded, which crimes are selected to be included within the analysis and which analytical tools are used, predictive algorithms may thus contribute to prejudicial decision-making and discriminatory outcomes.

In addition, considerable concerns exist that the operation of such assessments in the context of crime prevention is likely to create echo chambers within which pre-existing prejudice may be further cemented. Bias or prejudice related, for example, to racial or ethnic background, may not be recognised as such by the police when integrated into an automated computer program that is deemed independent and neutral (see also 6.). As a result, bias may become standardised and may then be less likely to be identified and questioned as such. While it is unclear how prevalent such decisions created by algorithms are in the criminal justice system generally, the mere potential of their use raises serious concerns with regard to Article 6 of the ECHR and the principle of equality of arms and adversarial proceedings as established by the European Court of Human Rights.[8]

Furthermore, algorithms are increasingly used in the context of the civil and criminal justice systems where artificial intelligence is being developed to eventually support or replace decision-making by human judges. Such systems are currently being tested to identify decision outcomes with a view to detect patters in complex judicial decision-making. Thus far, the reliable prediction rate is relatively low at 79%. It is therefore considered premature at the current time to imagine such systems replacing judges.[9] Nevertheless, it is suggested that such systems can support or assist judges (and lawyers).[10] Given the pressure of high caseloads and insufficient resources from which most judiciaries suffer, there is a danger that support systems based on artificial intelligence are inappropriately used by judges to "delegate" decisions to technological systems that were not developed for that purpose and are perceived as being more 'objective' even when this is not the case. Great care should therefore be taken to assess what such systems can deliver and under what conditions that may be used in order not to jeopardise the right to a fair trial. This is particularly the case when such systems are introduced mandatorily, as is the case for parole decisions in the United States. Concerns about judicial bias around parole decisions have led to the mandatory introduction of software to predict the likelihood of offenders reoffending in

---

[7] See, for example, William Issac, Kristian Lum Kristian Lum and William Isaac (2016), To predict and serve? Significance, October 10, 2016, The Royal Statistical Society, available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/epdf (last visited on 25 September 2017).

[8] See, for instance, in *Jespers v. Belgium*, 15 October 1980, no 8404/78, *Salduz v. Turkey*, 17 November 2008, no 36391/02 and *Blokhin v. Russia*, 13 April 2016, no 47152/06.

[9] Nikolaos Altreas et al "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective" *PeerJ Computer Science* Open Access (Published 24. October 2016), available at https://peerj.com/articles/cs-93.pdf at p.2; see also The Law society gazette, Monidipa Fouzder, "Artificial Intelligence mimics judicial reasoning", 22 June 2016, available at: https://www.lawgazette.co.uk/law/artificial-intelligence-mimics-judicial-reasoning/5056017.article (last visited on 25 September 2017).

[10] Ibid.

many U.S. states.[11] However independent investigation of this software suggests that the "software used […] to predict future criminals […] is biased against blacks" (Angwin, Mattu, and Kirchner 2016).

2. PRIVACY AND DATA PROTECTION

The longest and most sustained human rights debate on automated data processing and algorithms relates to the right to privacy.[12] Algorithms facilitate the collection, processing and repurposing of vast amounts of data and images. This may have serious consequences on the enjoyment of the right to private and family life, including the right to data protection, as guaranteed in Article 8 of the ECHR. Algorithms are used in online tracking and profiling of individuals whose browsing patterns are recorded by "cookies"[13] and similar technologies such as digital fingerprinting, aggregated with search queries (search engines/virtual assistants). Moreover, behavioural data is processed from smart devices, such as location and other sensor data through apps on mobile devices (Tene and Polonetsky 2012), raising increasing challenges for privacy and data protection.

Applications of online tracking and profiling are also used in targeted advertising based on the profile of a person's presumed interests. Here, user consent is an important regulatory concern. Research at Berkeley in 2012 established, for instance, that the use of privacy-invasive tracking technologies that cannot be observed by users (such as digital fingerprinting and behavioural data generated by sensors) has increased following the greater awareness of consumers and their growing practice of deleting or disabling cookies as part of the "do-not-track" choice settings in internet browsers.[14] Moreover, extensive data processing through the use of algorithms may aggravate infringements of other rights, as personal data is used to target individuals, such as in the context of insurance or employment applications.

One particular challenge of algorithmic processing of personal data is the generation of new data. When a data subject shares a few discrete pieces of data, it is often possible for those data to be merged, generating second and even third generations of data about the individual. Two innocuous pieces of data, when assessed in comparison with a much larger data set can "breed" and generate "baby data", the nature of which can be entirely unpredictable for the data subject. This raises major issues for the notions of consent,

---

[11] See GCN, Kevin McCaney, "Prisons turn to analytics software for parole decisions", 1 November 2013, available at https://gcn.com/articles/2013/11/01/prison-analytics-software.aspx (last visited on 25 September 2017).

[12] See Sills 1970.

[13] A cookie is a small amount of data generated by a website and saved by the web browser with the purpose to remember information about the user, similar to a preference file created by a software application. While cookies may serve many functions, their most common purpose is to store login information for a specific site. Cookies are also used to store user preferences for a specific site. For example, a search engine may store search settings in a cookie.

[14] CJ Hoofnagle "Behavioural Advertising: The Offer You Cannot Refuse" (2012) 6 Harvard Policy & Law Review 273-296

transparency and personal autonomy. Research from Cambridge and Stanford Universities illustrate the scale of the challenge.[15]

Efforts are ongoing to modernise the 1981 Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention ETS 108) in line with the technological evolution, and to further define the rights of the data subject with respect to the implications for privacy of contemporary tools for data collection, processing, repurposing and profiling. Article 8 of the draft modernised Convention establishes the explicit right of every individual not to be subjected to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration; the right to obtain knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her; and to object at any time, on grounds relating to his or her situation, and to the processing of personal data concerning him or her, unless the controller demonstrates legitimate grounds for the processing which override his or her interests or rights and fundamental freedoms. The modernisation proposals further aim to provide complementary safeguards as regards transparency (Article 7bis) and the need for an examination of the likely impact of data processing on the rights and fundamental freedoms of the person prior to commencing such processing (Article 8bis).[16]

The "Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big data" [17] recently adopted by the Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data provide a general framework to apply appropriate policies and measures to continue to make effective the data protection principles in the context of Big Data.

Data protection regulatory frameworks at EU level, such as the General Data Protection Regulation of April 2016 (Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data), which will apply as of May 2018, also establish standards for the use of algorithms in data collection, including possibly a limited right to information or even a "right to explanation" (Goodman and Flaxman 2016) with respect to decision-making processes – although the

[15] See Stanford news, "New Stanford research finds computers are better judges of personality than friends and family", available at: http://news.stanford.edu/2015/01/12/personality-computer-knows-011215/ (last visited on 25 September 2017).

[16]See the Draft modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data, September 2016, available at: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a616c (last visited on 25 September 2017).

[17] Council of Europe, Guidelines on the Protection of Individuals with regard to the Processing of Personal Data in a World of Big Data, 17 January 2017, available at: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806f06d0 (last visited on 25 September 2017).

exact scope of this right to explanation is heavily contested,[18] (Wachter, Mittelstadt, and Floridi 2016) - as well as the right to access to "knowledge of the logic involved in any automatic processing of data concerning him."[19]

Particular concerns arise from the use of data brokers who aggregate the information contained in personal profiles. Profiling, in itself, means extrapolation of data available on the internet through processes of automated information gathering and subsequent construction and application of profiles. Profiling techniques can benefit individuals and society by, for instance, leading to better market segmentation or permitting an analysis of risks and fraud. Yet, there are also important concerns about the usage of the technique. The Council of Europe Recommendation on Profiling[20] addresses the risk that profiles attributed to a data subject make it possible to generate *new* data, including through data aggregation. This information may then be mined through the use of algorithms, which creates a risk of large-scale surveillance ("data-veillance") by private entities and governments alike (Rubinstein, Lee, and Schwartz 2008). This view is echoed by the United Nations Human Rights Council, which on 22 March 2017 noted with concern "that automatic processing of personal data for individual profiling may lead to discrimination or decisions that otherwise have the potential to affect the enjoyment of human rights, including economic, social and cultural rights."[21]

The main concern of using data from profiles for different purposes through algorithms is that the data loses its original context. Repurposing of data is likely to affect a person's informational self-determination. Search engines may have a similar effect on the right to privacy and data protection as they also facilitate the aggregation of data about a specific individual.

The use of data from profiles, including those established based on data collected by search algorithms and search engines, directly affects the right to a person's informational self-determination. The data subject will usually not be aware of the profiling itself and of the subsequent repurposing of data beyond its original context, making it easier to find information by reducing the practical obscurity of anonymous data. In addition, the results obtained through search algorithms may be incomplete, inaccurate or out-dated, thereby

---

[18] See Wachter, Mittelstadt and Floridi, 2016. See also Lilian Edvards and Michael Veale, 2017, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855 (last visited on 06 October 2017).

[19] See for further details European Data Protection Supervisor, "ethics", webpage, available at: https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/Ethics (last visited on 25 September 2017). Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data provides a framework for the processing of data in the course of actions that do not fall under Community Law, such as judicial cooperation in criminal matters and police cooperation.

[20] Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling.

[21] U.N. Human Rights Council Resolution on the Right to Privacy in the Digital Age, U.N. Doc. A/HRC/34/7, 23 Mar. 2017, para.2

placing individuals in a distorted light, which may be prejudicial.[22] Such profiles may have particularly serious consequences for children and their future. Finally, there is increasing evidence that data is harvested in order to gain behavioural insights that can be used to target voters and – ultimately – even manipulate elections (see below 8.).[23]

Another key aspect related to the usage of algorithms for automated data processing focusses on 'cloud' data storage. This refers to solutions whereby files and other data are no longer stored on local storage but are stored remotely on servers accessible via the Internet. However, by virtue of engaging in non-local storage practices, the data of users may be processed by algorithms while stored remotely in intrusive ways that would not usually be practiced. Such automated data processing can take place in two places: (1) in transit to the remote network storage location and (2) on the remote servers where the data is stored. It may be increasingly difficult for users to ascertain whether they are using local or remote services, as modern operating systems are gradually becoming more deeply enmeshed with 'cloud' remote services. With regard to data in transit, it may therefore be difficult to determine whether it is sufficiently protected through technologies such as strong end-to-end encryption, and whether it is not manipulated in some form.[24]

3. FREEDOM OF EXPRESSION

The operation of algorithms and data processing techniques has tremendous effects on the right to freedom of expression, which includes the right to receive and impart information. While the positive effects of search algorithms and search engines for the human right to freedom of expression has been repeatedly referred to,[25] their potential for harming the freedom of information and freedom of expression of individuals, groups and whole segments of societies is now increasingly discussed.[26] Concerns arise not only with respect

---

[22] See Solove (2006), p. 547. As regards data processing in the course of judicial cooperation in criminal matters and police cooperation, which do not fall under Community Law, Directive (EU) 2016/680 *on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data* establish data protection safeguards.

[23] See also The Guardian, "The great British Brexit robberty: how our democracy was hijacked", 7 May 2017, available at: https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy (last visited on 25 September 2017).

[24] For example, Microsoft's cloud service 'SkyDrive' operates an automated process designed to remove certain content (such as nudity). See Clay 2012.

[25] See, for instance, Council of Europe, Recommendation of the Committee of Ministers to member States on the protection of human rights with regard to search engines, CM/Rec(2012)3, Adopted by the Committee of Ministers on 4 April 2012 at the 1139th meeting of the Ministers' Deputies, paragraph 1, available at https://wcd.coe.int/ViewDoc.jsp?id=1929429 (last visited on 25 September 2017), observing that s*earch engines "enable a worldwide public to seek, receive and impart information and ideas and other content in particular to acquire knowledge, engage in debate and participate in democratic processes."*

[26] See, for instance, the 2016 Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, to the Thirty-second session of the Human Rights Council (A/HRC/32/ 38), pointing out that "search engine algorithms dictate what users see and in what priority, and they may be manipulated to restrict or prioritise content".

to the individual right to freedom of expression but also with respect to the inherent aim of Article 10 of creating an enabling environment for pluralist public debate that is equally accessible and inclusive to all. Moreover, the privacy and data protection concerns raised above can significantly impede on individuals' ability to freely express themselves.

Search engines act as crucial gatekeepers for human beings who wish to seek, receive or impart information. Content which is not indexed or ranked highly by an Internet search engine is less likely to reach a large audience or to be seen at all. As a result, the use of algorithms may lead to fragmentation of the public sphere and to the creation of "echo chambers" that favour only certain types of news outlets, thereby enhancing levels of polarisation in society which can seriously jeopardise social cohesion.[27] A search algorithm might also be biased towards certain types of content or content providers, thereby risking affecting related values such as media pluralism and diversity.[28] This is particularly the case in the context of dominant online search engines (Pasquale 2016).

The algorithmic predictions of user preferences deployed by social media platforms guide not only what advertisements individuals might see, but they also personalise search results and dictate the way how social media feeds, including newsfeeds, are arranged. Given the size of platforms such as Google or Facebook, their centrality for many experience of the internet as a quasi-public sphere (York 2010) and their ability to massively amplify certain voices (Bucher 2012), this is by no means a trivial matter. On the contrary, the personalisation of information that users receive based on their predicted preferences and interests can create "filter bubbles" and may substantially compromise the freedom of expression, which includes the right to information. While filter bubbles and echo chambers are a plausible and therefore a widely-discussed concept, it should be noted that the empirical evidence for their existence in Europe is mixed (Nguyen et al. 2014; Zuiderveen Borgesius et al. 2016). Individuals usually inform themselves by using a repertoire of sources, not just via social media or internet searches.

According to Article 10 of the ECHR, any measure that blocks access to content through filtering or removal of content must be prescribed by law, pursue one of the legitimate aims foreseen in Article 10.2, and must be necessary in a democratic society. In line with the jurisprudence of the European Court of Human Rights, any restriction of the freedom of

[27] [27] See also Arstecnica, Roheeni Saxena, "The social media "echo chamber" is real", available at https://arstechnica.com/science/2017/03/the-social-media-echo-chamber-is-real/ (last visited on 25 September 2017).

[28] According to the UNESCO's World Trends in Freedom of Expression and Media Development Publication, internet technologies have enabled many more voices to be heard. While the lack of gender-disaggregated statistics thus far prevents a better understanding of the gender-specific impacts of algorithms controlled search tools on the exercise of the right to freedom of expression, it appears that regional and gender patterns of communications are replicated also in this new volume of voices; see UNESCO's World Trends in Freedom of Expression and Media Development Publication at: http://www.unesco.org/new/en/world-media-trends (last visited on 25 September 2017).

expression must correspond to a "pressing social need" and be proportionate to the legitimate aim(s) pursued.[29]

However, content removal on social media platforms often takes place through semi-automated or automated processes. Algorithms are widely used for content filtering and content removal processes (Urban, Karaganis, and Schofield 2016), including on social media platforms, directly impacting on the freedom of expression and raising rule of law concerns (questions of legality, legitimacy and proportionality). While large social media platforms like Google or Facebook have frequently claimed that human beings remove all content (Buni and Chemaly 2016), large parts of the process are automated (Wagner 2016b) and based on semi-automated processes. According to a report from the British Intelligence and Security Committee of Parliament, [30] various automated techniques exist for identifying content believed to break the terms of service of the respective provider, be it because of extremist content, child exploitation or illegal acts such as the incitement to violence. These techniques may also be used to disable or automatically suspend user accounts (Rifkind 2014). A particular challenge in this context is that intermediaries are encouraged to remove this content voluntarily, without clear legal basis. This lack of a legal basis for 'voluntary' automated content removal makes it even more difficult to ensure that basic legal guarantees such as accountability, transparency or due process are upheld (Fernández Pérez 2017).

In the US, the Obama administration has advocated for the use of automated detection and removal of extremist videos and images.[31] Additionally, there have been proposals to modify search algorithms in order to "hide" websites that would incite and support extremism. The automated filtering mechanism for extremist videos has been adopted by Facebook and YouTube for videos. However, no information has been released about the process or about the criteria adopted to establish which videos are "extremist" or show

[29] In *Yildirim v. Turkey*, 18 March 2013, no 3111/10, the European Court of Human Rights has emphasised that "the dangers inherent in prior restraints are such that they call for the most careful scrutiny on the part of the Court, (..) for news is a perishable commodity and to delay its publication, even for a short period, may well deprive it of all its value and interest". Therefore blocking access to the internet or removal of online content requires a legal framework, "ensuring both tight control over the scope of bans and effective judicial review to prevent any abuse of power (..) In that regard, the judicial review of such a measure, based on a weighing-up of the competing interests at stake and designed to strike a balance between them, is inconceivable without a framework establishing precise and specific rules regarding the application of preventive restrictions on freedom of expression".

[30] See UK Intelligence and Security Committee of Parliament report, Privacy and Security: A modern and transparent legal framework, March 2015, available at: http://isc.independent.gov.uk/committee-reports/special-reports (last visited on 25 September 2017).

[31]See Article 19, "Algorithms and automated decision-making in the context of crime prevention", 2 December 2016, available at: https://www.article19.org/resources.php/resource/38579/en/algorithms-and-automated-decision-making-in-the-context-of-crime-prevention (last visited on 25 September 2017).

"clearly illegal content"'[32] In the wake of reports from The Times of London and The Wall Street Journal that ads were appearing on YouTube videos that espoused "extremism" and "hate speech", YouTube reacted with a tighter use of its algorithm operated to detect "not advertiser-friendly" content, which has reportedly affected independent media outlets, including comedians, political commentators and experts.[33]

Similar initiatives have been developed in Europe, where intermediary service providers, in response to public and political pressure, have committed themselves to actively counter online hate speech through automated techniques that detect and delete all illegal content. While not disputing the necessity to effectively confront hate speech, such arrangements have been criticised for delegating law enforcement responsibilities from state to private companies, for creating the risk of excessive interference with the right to freedom of expression, and for their lack of compliance with the principles of legality, proportionality, and due process. Requiring intermediaries to restrict access to content based on vague notions such as "extremism" obliges them to monitor all flows of communication and data online in order to be able to detect what may be illegal content. It therefore goes against the established principle that there should be no monitoring obligation for intermediaries, which is enshrined in EU-law and in relevant Council of Europe policy guidelines.[34] Due to the significant chilling effect that such monitoring has on the freedom of expression, this principle is also reiterated in the draft recommendation on the roles and responsibilities of internet intermediaries prepared by the Council of Europe's Committee of Experts on Internet Intermediaries in September 2017.[35]

Moreover, by ordering the intermediary to decide itself what to remove as "extremist" and what not, the public authority passes the choice of tools and measures onto a private party, which can then implement solutions (such as content removal or restriction) that the public authorities themselves could not legally prescribe. Public-private partnerships may thus allow public actors "to impose regulations on expression that could fail to pass constitutional muster" (Mueller 2010:213) in contravention of rule of law standards. Moreover, these kinds of demands by public institutions of private actors lead to overbroad and automated monitoring and filtering of content.

---

[32] See Reuters, Joseph Menn, Dustin Volz, Exclusive: Google, "Facebook quietly move toward automatic blocking of extremist videos", available at: http://www.reuters.com/article/us-internet-extremism-video-exclusive-idUSKCN0ZB00M (last visited on 25 September 2017).

[33] See The New York Times, Amanda Hess, "How YouTube's Shifting Algorithms Hurt Independent Media", 17 April 2017, available at: https://www.nytimes.com/2017/04/17/arts/youtube-broadcasters-algorithm-ads.html?_r=0 (last visited on 25 September 2017).

[34] See Article 15 of Directive 2000/31/EC of 8 June 2000 ("Directive on electronic commerce"), and Principle 6 *on limited liability of service providers for Internet content* of the Council of Europe Declaration on freedom of communication on the Internet of 28 May 2003.

[35] See Draft Recommendation of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries, finalized by the MSI-NET on 19 September 2017, at https://rm.coe.int/draft-recommendation-on-internet-intermediaries-version-4/1680759e67.

The Europol Internet Referral Unit had, one year after its launch in July 2015, assessed and processed 11.000 messages containing violent extremist content materials across 31 online platforms in eight languages, reportedly leading to the removal of 91.4% of the total content from the platforms.[36] Steps have reportedly been taken to automate this system with the introduction of the Joint Referral Platform announced in April 2016.[37]

While the imperative of acting decisively against the spread of hate messages and the incitement to racially-motivated offences is indisputable, such practices raise considerable concerns related to foreseeability and legality of interferences with the freedom of expression. Notably the data on extremist online content that Europol is processing refers not just to content that is illegal in Council of Europe Member States, but also to material that violates the terms of service of an internet intermediary. Moreover, in many situations extremist content or material inciting violence is difficult to identify, even for a trained human being, because of the complexity of disentangling factors such as cultural context and humor. Algorithms are today not capable of detecting irony or critical analysis. The filtering of speech to eliminate harmful content through algorithms therefore faces a high risk of over-blocking and removing speech that is not only harmless but can contribute positively to the public debate. According to the European Court of Human Rights, Article 10 also protects shocking, offensive or disturbing content.[38] Algorithmic blocking, filtering or removal of content may thus have a significant adverse impact on legitimate content. The already highly prevalent dilemma of large amounts of legal content being removed because of the terms of service of internet platforms is further exacerbated by the pressure placed on them to actively filter according to vague notions such as "extremist", "hate speech" or "clearly illegal content". According to the European Court of Human Rights, any obligation to filter or remove certain types of comments by users from online platforms puts an "excessive and impracticable" burden on the operators and risks to oblige them to install a monitoring system "capable of undermining the right to impart information on the internet."

---

[36] See Europol Internet Referral Unit One Year On, Press release, 22 July 2016, available at: https://www.europol.europa.eu/newsroom/news/europol-internet-referral-unit-one-year (last visited on 25 September 2017).

[37] See EC Communication from the Commission to the European Parliament, The European Council And The Council delivering on the European Agenda on Security to fight against terrorism and pave the way towards an effective and genuine Security Union, available at: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/policies/european-agenda-security/legislative-documents/docs/20160420/communication_eas_progress_since_april_2015_en.pdf (last visited on 25 September 2017). See also Article 19, *Algorithms and Automated Decision-Making in the Context of Crime Prevention: A briefing paper,* 2016.

[38] As demonstrated not only in jurisprudence of domestic courts, but also in the case-law of the European Court of Human Rights, the exercise of qualifying speech as (illegal) hate speech is delicate. Several judgments of the Court concerning the question whether certain speech could or should be qualified as criminal hate speech resulted in divided votes, such as e.g. in *I.A. v. Turkey*, 13 September 2005, no 42571/98; *Lindon, Otchakovsky-Laurens and July v. France*, 22 October 2007, no 21279/02 and no 36448/02; *Féret v. Belgium*, 16 July 2009, no 15615/07 and *Perinçek v. Switzerland*, 15 October 2015, no 27510/08. See also *Vejdeland and others v. Sweden*, 9 February 2012, no 1813/07.

[39] The Venice Commission has equally called for efforts to strengthen human rights safeguards and to avoid excessive burdens being placed on providers of electronic communication networks and systems.[40]

Public concern in Europe and the U.S. has grown following the U.S. elections in 2016 with respect to the dissemination of misinformation via fabricated, intentionally false and misleading news (so-called ”fake news”), including through automated techniques and on social media platforms, thereby possibly having significant influence over democratic decision-making processes (see also below 8.).[41] As a result, there have been renewed calls for traditional media responsibility standards to be applied to social media platforms. Some scholars have likened Facebook to be acting as a “news editor [that] has editorial responsibility for its trending topics” (Helberger and Trilling 2016). The question follows, whether social media platforms, through their algorithms that rank and curate third-party submissions, exert a form of editorial control traditionally performed by media professionals and therefore engage specific media responsibilities.[42]

4. FREEDOM OF ASSEMBLY AND ASSOCIATION

The internet and in particular social networking services are vital tools for the exercise and enjoyment of the right to freedom of assembly and association, offering great possibilities for enhancing the potential for participation of individuals in political, social and cultural life. [43] The freedom of individuals to use internet platforms, such as social media, to establish associations and to organise themselves for purposes of peaceful assembly, including

---

[39] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, 2 February 2016, no 22947/13.

[40] See Joint Opinion of the Venice Commission, the Directorate of information society and action against crime and of the Directorate of Human Rights (DHR) of the Directorate General of Human Rights and Rule of Law (DGI) of the Council of Europe on the Draft Law n° 281 amending and completing Moldovan Legislation on the so-called "Mandate of security", adopted by the Venice Commission at its 110th Plenary Session (Venice, 10-11 March 2017), available at: http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2017)009-e (last visited on 25 September 2017).

[41] See for example The Power of Big Data and Psychographics, available at: https://www.youtube.com/watch?v=n8Dd5aVXLCc (last visited on 25 September 2017) or Das Magazin, Hannes Grassegger und Mikael Krogerus, “Ich habe nur gezeigt, dass es die Bombe gibt“, no 48, 3 December 2016, available at https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/, (last visited on 25 September 2017) although the exact role of the techniques used by Cambridge Analytica and others during the U.S. elections is heavily disputed.

[42]See also Reuters institute, Emma Goodman, “Editors vs algorithms: who do you want choosing your news?”, available at: http://reutersinstitute.politics.ox.ac.uk/news/editors-vs-algorithms-who-do-you-want-choosing-your-news (last visited on 25 September 2017), and the Code of Conduct on countering illegal hate speech online, of 31 May 2016 between the EU and Facebook, Microsoft, Twitter and You Tube. See also The Guardian, “2016: the year Facebook became the bad guy”, available at: https://www.theguardian.com/technology/2016/dec/12/facebook-2016-problems-fake-news-censorship (last visited on 25 September 2017).

[43] See Recommendation CM/Rec(2012)4 of the Committee of Ministers to member States on the protection of human rights with regard to social networking services.

protest, in line with Article 11 of the ECHR has equally been emphasised.[44] Around the globe, social media and their algorithmically advanced dissemination and networking potential have been suggested to play a prominent role in organising and motivating activists and protestors.[45]

In line with Article 11, any restriction to the right to freedom of peaceful assembly and to freedom of association must be prescribed by law, pursue a legitimate aim and be necessary in a democratic society.

The operation of algorithms on social media platforms and the vast amount of personally identifiable information on individuals that is available may of course also be used to track and identify human beings and may lead to the automatic sorting out of certain individuals or groups from calls for assemblies, which could have a significant negative impact on the freedom of assembly. Profiling and crowd control of protesters does not just take place on the internet, but also extends to off-line data-based crowd control methods. Theoretically, algorithms used to predict possible conflict and protest situations could also be used as pre-emptive tool to prevent demonstrations or protests by arresting certain individuals before they even gather.[46]

5. EFFECTIVE REMEDY

Article 13 of the ECHR stipulates that everyone, whose rights have been infringed upon shall have an effective remedy before a national authority. The available remedy should be effective in practice and in law. States must therefore ensure that individuals have access to judicial or other procedures that can impartially decide on their claims concerning violations of human rights online, including effective non-judicial mechanisms, administrative or other means for seeking remedy such as through national human rights institutions. As primary responsible entity for all rights contained in the ECHR, states must take appropriate steps to protect against human rights violations, including by private-sector actors, and must ensure that those affected have access to an effective remedy. This includes ensuring that private-sector actors respect human rights throughout their operations, in particular by establishing effective complaint mechanisms that promptly remedy the grievances of individuals.

Automated decision-making processes lend themselves to particular challenges for individuals' ability to obtain effective remedy. These include the opaqueness of the decision itself, its basis, and whether the individuals have consented to the use of their data in

---

[44] See Recommendation CM/Rec(2016)5 of the Committee of Ministers to member States on Internet freedom and Recommendation CM/Rec(2014)6 of the Committee of Ministers to member States on a Guide to human rights for Internet users.

[45] See, among others, Pablo Barberá and Megan Metzger, "Tweeting the Revolution: Social Media Use and the #Euromaidan Protests", available at: http://www.huffingtonpost.com/pablo-barbera/tweeting-the-revolution-s_b_4831104.html ((last visited on 25 September 2017). See also Zeynep Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest, Yale University Press, 2017.*

[46] See Tim de Chant, "The Inevitability of Predicting the Future", available at: http://www.pbs.org/wgbh/nova/next/tech/predicting-the-future/ (last visited on 25 September 2017).

making this decision, or are even aware of the decision affecting them. The difficulty in assigning responsibility for the decision also complicates individuals' understanding of whom to turn to address the decision. The nature of decisions being made automatic, without or with little human input, and with a primacy placed on efficiency rather than human-contextual thinking, means that there is an even larger burden on the organisations employing such systems to provide affected individuals with a way to obtain remedy.

The wide variety of sectors in which automated decision-making systems are employed can have serious repercussions on human rights, whether related to health treatments, job opportunities, predictive policing or otherwise, rendering the capability to obtain effective remedy in each of these even more essential.

An increasing number of companies, especially larger ones, use algorithms and automated data processing techniques for running their complaints procedures. In the context of automated content removal processes on social media platforms (see above 3.), the use of algorithms is particularly evident in the responses that different types of content receive and how content is prioritised, a process that is evidently automated. The same is true for the threshold of user complaints that is required before a piece of content is reviewed. There are strong suggestions that the complete response systems of internet platforms such as Facebook, Google or Microsoft to user queries are automated for many types of inquiries and complaints (Wagner 2016b; Zhang, Stalla-Bourdillon, and Gilbert 2016). Often, many users will need to complain about a specific type of content before an automated algorithm identifies it as relevant to be referred to a human operator for content review. These operators are reported to be working often under considerable time pressure and with minimal instructions, in line with internal "deletion rules".[47]

The right to an effective remedy implies the right to a reasoned and individual decision. Historically, all such decisions have been taken by human beings who, in the exercise of their functions, based on comprehensive training and in line with the applicable decision-making processes, have been granted a margin of discretion. In principle, it is a judge, government minister or administrative official's task to decide, in accordance with the criteria and case-law developed by the Court, how the balancing of individual rights, such as the freedom of expression and the protection from violence or the protection of the rights of others, shall be put into practice. The decision must be based on a careful analysis of the specific context, taking into consideration the "chilling effect" that the interference may entail and considering the proportionality of the interference. Today however, it is increasingly algorithmic data processing techniques that are preparing and influencing decision-making in complaints procedures.

In addition, serious concerns exist as to whether automatic response processes to complaints constitute an effective remedy. While the famous removal of a YouTube video on a European Parliament debate related to torture was reinstated after only few hours, following an MEP complaint, who even received a public apology from Google, there are

[47] See Süddeutsche Zeitung, Till Krause and Hannes Grassegger, "Inside Facebook", available at: http://international.sueddeutsche.de/post/154513473995/inside-facebook (last visited on 25 September 2017).

considerable doubts as to whether all complaints are treated with such attentiveness.[48] Rather, algorithms often obscure access to a reasoned explanation as to why certain steps were taken in a particular case.

In all cases, the right to an effective remedy demands that access to an escalated system of dispute resolution is provided. While the first step may be operated through automated means, there must be a possibility to complain against the outcome to a higher internal review mechanism. If the complainant is not satisfied with the outcome, he must have the possibility to challenge it through judicial remedy, in line with Article 13 of the European Convention.[49] However there is some suggestions that a judicial redress mechanism alone is insufficient and that there is a need for government "supervision of collaborative negotiations between consumers and corporations" (Loo, 2016).

With respect to the right to privacy, automated techniques and algorithms facilitate forms of secret surveillance and "data-veillance" that are impossible for the affected individual to know about. The European Court of Human Rights has underlined that the absence of notification at any point undermines the effectiveness of remedies against such measures.[50]

6. PROHIBITION OF DISCRIMINATION

Another key human right that is frequently cited in relation to the operation of algorithms and other automated processing techniques is the right to enjoy all human rights and fundamental freedoms without discrimination.

In terms of speed and volume of data processed, algorithmic decision-making can have considerable advantages over certain types of human decision-making. However, algorithms may well have inbuilt biases that may be hard to detect and/or correct (Sandvig et al. 2016). This is particularly the case when individual variables in big data algorithms serve as 'proxies' for protected categories such as race, gender or age. An algorithm may choose to discriminate against a group of users which correlates to 80%, 90%, 95% or even 99% with a variable such as race, gender or age, without doing so 100% of the time.

Search algorithms and search engines by definition do not treat all information equally. While processes used to select and index information may be applied consistently, the search results will typically be ranked according to perceived relevance. Accordingly, different items of information will receive different degrees of visibility depending on which factors are taken into account by the ranking algorithm (see also 3).[51] As a result of data aggregation and profiling, search algorithms and search engines rank the advertisement of

---

[48] See Marietje Schaake, "When You Tube took down my video", available at: https://www.marietjeschaake.eu/en/when-youtube-took-down-my-video (last visited on 25 September 2017).

[49] See, among others, *O'Keefe v. Ireland*, 28 January 2014, no 35810/09.

[50] See *Roman Zakharov v. Russia*, 4 December 2015, no 47143/06.

[51] The algorithm may also – deliberately or not – be impacted by a variety of external factors, which may relate to business models, legal constraints (e.g. copyright) or other contextual factors.

smaller companies that are registered in less affluent neighbourhoods lower than those of large entities, which may put them at a commercial disadvantage. Search engines and search algorithms also do not treat all users equally. Different users may be presented with different results, on the basis of behavioural or other profiles, including personal risk profiles that may be developed for the purpose of insurance or credit scoring or more generally for differential pricing, i.e., offering different prices for the same goods or services to different consumers based on their profile (see above 2.).[52]

A biased algorithm that systematically discriminates one group in society, for example based on their age, sexual orientation, race, gender or socio-economic standing, may raise considerable concerns not just in terms of the access to rights of the individual end-users or customers affected by these decisions, but also for society as a whole. Some authors have even suggested that online services which use personalised rating systems are inherently likely to lead to discriminatory practices (Rosenblat et al. 2016). It can be argued as a result that individuals should have the right to view an 'unbiased' and not personally targeted version of their search results. This can be seen as a way for an individual to exit their own 'filter bubble' and see an untargeted version of the search content, social media timeline or other internet-based service or product that they are using. In theory, algorithms could be useful tools to reduce bias in places where it is common, such as in hiring processes. Yet, experts have warned that automation and machine learning have the potential to reinforce existing biases because, unlike humans, algorithms may be unequipped to consciously counteract learned biases. [53]

One potentially helpful consideration to discern whether algorithms promote or prevent discriminatory treatment is to refer to the legal distinction between direct and indirect discrimination. Direct discrimination occurs where a decision-maker bases her decision directly on criteria or factors which are regarded as unlawful (such as race, ethnicity, religion, gender, sexual orientation, age, or disability). Frequently these unlawful biases are made sub-consciously and on the basis of information which is external to the dataset which *should* form the basis of the decision-making (for example, an interviewer noticing the age or racial origin of the person standing in front of her). Arguably algorithm-based systems are better at excluding such direct biases. Indirect discrimination occurs where a certain characteristic or factor occurs more frequently in the population groups against whom it is unlawful to discriminate (such as a person with a certain racial or ethnic background living in a certain geographical area; women having fewer pensionable years because of career breaks). Since algorithmic decision-making systems may be based on correlation between data sets and efficiency considerations, there is a danger that such systems perpetuate or

[52] See also relevant provisions in the EU Regulation 2016/679 related to profiling and automated data processing and the rights of the data subject.

[53] See, for instance, The Guardian, "AI programs exhibit racial and gender biases, research reveals", available at: https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals (last visited on 25 September 2017); and The Guardian, "How algorithms rule our working lives", available at: https://www.theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives (last visited on 25 September 2017).

exacerbate indirect discrimination through stereotyping. Indirect discrimination is only present where differential treatment cannot be justified.

When using algorithmic decision-making systems it is therefore important to seek to prevent unjustified differential treatments and to design systems accordingly. In particular, differential treatment will be unjustified and unlawful where it relies on biased data to generate a risk assessment. In that case, the decision itself is not directly but indirectly discriminatory, as it relies on data and information which may be, for instance, racially biased. An example for this is where the criminal system uses risk assessment tools to decide whether a person should be granted bail. This system generates risk profiles that are based on police data, such as the number or re-arrests for the same offence. The fact of re-arrests, however, may be the consequence of direct discrimination (racial bias).[54] If algorithmic decision-making systems are based on previous human decisions, it is likely that the same biases which potentially undermine the human decision-making are replicated and multiplied in the algorithmic decision-making systems, only that they are then more difficult to identify and correct.

7. SOCIAL RIGHTS AND ACCESS TO PUBLIC SERVICES

The workplace is another key area where automated decision-making has become increasingly common in recent years. Algorithms may be involved in decisions on both hiring and firing staff, staff organisation and management, as well as the individual evaluations of employees. Automated feedback loops, sometimes linked to customer input, may decide over the performance evaluation of staff (Kocher and Hensel 2016). These decision-making processes are by no means perfect when humans conduct them. Bias related to race (Bertrand and Mullainathan 2004) class and gender (Altonji and Blank 1999; Goldin and Rouse 1997) has been demonstrated repeatedly in human resources management practices and processes. With more and more companies moving towards algorithmic recruitment methods (Rosenblat, Kneese, and others 2014), however, new concerns related to the lack of transparency in the decisions they make, both in the hiring process and beyond, have been raised. Moreover many of these automated decision-making processes are based on data received via internet platforms. Allowing the 'wisdom of the crowd' to make decisions about individuals' employment is not only highly questionable from an ethical point of view, it also limits the ability of workers to contest such decisions as they seem to be an 'objective' measures of their performance (Tufekci et al. 2015).

[54] See Laurel Eckhouse, "Big data may be reinforcing racial bias in the criminal justice system", available at: https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d_story.html?utm_term=.720084735d73 (last visited on 25 September 2017); and ProPublica, Angwin, Julia, Surya Mattu, and Lauren Kirchner, "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks", 2016, available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (last visited on 25 September 2017).

As individual employment platforms are "transforming people into Human Computation," (Irani 2015:227) questions arise about workers' rights, employee self-determination and how societies as a whole believe that human beings should be treated at the workplace.[55] Notably the increased automation in the workplace also raises considerable challenges in relation to privacy rights (Hendrickx and van Bever 2013) of employees and how they can be safeguarded in the workplace. As more and more systems are automated and more and more data is collected at the workplace, employees' rights under Article 8 are in danger even if they are not directly targeted by general data collection measures (see above 2.) Finally, there are additional challenges related to the use of algorithms by both public and private sector organisations to monitor staff communications or to conduct internal "rankings" of employees that may not be part of the formal evaluation process but possibly more decisive with respect to individual career opportunities. Such practices are typically employed to ensure that staff represent well either a company or a bureaucracy and have evident implications for the freedom of expression of the employees (Voorhoof and Humblet 2013) and their human rights under Article 10 of the Convention (see above 3.).

Government agencies and services are increasingly automating their decision-making with the use of algorithms (van Haastert 2016). While it is heavily debated whether such systems can increase efficiency or not, what is evident is that the operation of such systems poses considerable questions for transparency and accountability of public decision-making, which must be held to a higher standard than the private or non-profit sector. At present the public sector in Europe is employing automated decision-making in areas as diverse as social security, taxation, health care and the justice system (van Haastert 2016; Tufekci et al. 2015). There is considerable danger of social sorting in medical data as algorithms can sort out specific citizen groups or human profiles, thereby possibly preventing their access to social services. Another example relates to the practice of profiling the unemployed, which was analysed by researchers in an effort to assess the social and political implications of algorithmic decision-making associated with social benefits (Jędrzej Niklas, Karolina Sztandar-Sztanderska, and Katarzyna Szymielewicz 2015). This analysis identified several challenges which are relevant also for the use of algorithms in other areas of public sector service delivery, such as non-transparent and algorithmic rules being applied in the distribution of public services and computational shortcomings triggering arbitrary decisions, for instance, with respect to receipt of social benefits.

8. THE RIGHT TO FREE ELECTIONS

The operation of algorithms and automated recommender systems that may create 'filter bubbles' - fully-automated echo chambers in which individuals only see pieces of information that confirm their own opinions or match their profile (Bozdag 2013; Pariser 2011; Zuckerman 2013) - can have momentous effects for democratic processes in society. While the actual impact of 'filter bubbles' and targeted misinformation on the formation of

[55] See F. Dorssemont, K. Lörcher and I. Schömann (eds.), *The European Convention of Human Rights and the Employment Relation,* Hart Publishing, Oxford, 2013.

political opinion is difficult to determine accurately,[56] fully-automated echo chambers pose the danger of creating "ideological bubbles" (O'Callaghan et al. 2015), that may be relatively easy to enter but hard to exit (Salamatian 2014). This may have crucial effects in particular in the context of elections.

While it has been argued since the advent of the internet that online campaigning and social media networks were likely to change the way in which politics and elections were run, it is only more recently that academic research has revealed the extent to which the curation and manipulation of online content on social media platforms may 'tip' elections. During U.S. elections, researchers reportedly manipulated the Facebook platform to influence users voting behaviour by telling them how their friends had said they had voted, without users' knowledge, and were able to convince a statistically significant segment of the population to vote in the congressional mid-term elections on 2 November 2010 (Bond et al. 2012). [57] There are strong indications that since then Facebook has been selling related political advertising services to political parties around the world, with similar behaviour observed during the UK local elections in 2016 (Griffin 2016). Whether Facebook and similar dominant online platforms may (deliberately or not) use their power to influence human voting or not is less the point than the fact that they – in principle – have the ability to influence elections.

Recent research suggests that elections may be won not by the candidates with the best political argument, but by those who use the most efficient technology to manipulate voters, sometimes emotionally and irrationally.[58] While this may not be an altogether new phenomenon, it has certainly increased in scale and effect, leading to a shift in paradigm that could jeopardise democracy itself. Data that is inconspicuously amassed, harvested and stored through algorithmic technologies has been likened to the new "currency of power", as it can directly be employed for the micro-targeting of voters, possibly with decisive effects on elections. Indeed, less-known candidates may not have the means to afford the

[56] See Nguyen, Tien T., Pik-Mai Hui, F.Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. 'Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity'. Pp. 677–686 in Proceedings of the 23rd International Conference on World Wide Web, WWW '14. New York, NY, USA: ACM (available at http://doi.acm.org/10.1145/2566486.2568012) and Zuiderveen Borgesius, Frederik J. et al. 2016. 'Should We Worry About Filter Bubbles?' Internet Policy Review. Journal on Internet Regulation 5(1). Retrieved 1 September 2016, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758126.

[57] In an experiment, Facebook researchers showed a graphic to some users in their news feed, indicating how many of their friends had voted that day and providing a button to click that they had voted as well. Users who were prompted with news of their friends' voting turned out to be 0.39% more likely to vote than the others, and their decision had a further effect on the voting behavior of their friends. The researchers concluded that their single message on Facebook, strategically delivered, increased turnout directly by 60,000 voters, and thanks to the ripple effect, ultimately caused an additional 340,-000 votes to be cast (amongst an overall 82 million) that day. See Jonathan Zittrain, *Engineering an election,* Harvard Law Review Forum Vol. 127, 335 – 339 (2014).

[58] See also The Guardian, "The great British Brexit robberty: how our democracy was hijacked", 7 May 2017, available at: https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy (last visited on 25 September 2017), arguing that the Brexit referendum was decided in the end by some 600,000 votes, just over 1% of the total of registered voters, which had been targeted by a firm that "introduced mass data-harvesting to its psychological warfare techniques", bringing together "psychology, propaganda and technology in this powerful new way".

most effective manipulation technologies that help predict voter preferences.[59] While political advertising on TV is nowadays regulated and there are impartiality requirements imposed on public broadcasters, no such equivalents exist for the use of algorithmic predictions of preferences and voter behavior that may have equally if not more powerful an impact on voters.

In this context, the particular role played by social bots in shaping the political and public debate leading up to elections has been discussed in particular in the context of the 2016 US elections and the Brexit referendum. Social bots are algorithmically controlled accounts that emulate the activity of human users but operate at much higher pace (*e.g.*, automatically producing content or engaging in social interactions), while successfully keeping their artificial identity undisclosed. Research into the extent to which the presence of social media bots affected political discussion around the 2016 U.S. Presidential election suggests that it can negatively affect democratic political discussion rather than improving it, which in turn can potentially alter public opinion and endanger the integrity of the election process.[60] The right to free elections, as established by Article 3 of Protocol 1 has been acknowledged by the European Court of Human Rights as "fundamental principle in a truly democratic political regime." Importantly, and as noted in the Feasibility study on the use of Internet in elections by the Committee of Experts on Media Pluralism and Transparency of Media Ownership (MSI-MED) at the Council of Europe, regulatory challenges related to elections are not due to the rise of intermediaries but rather a lack of adequate regulation. As the study notes the "most fundamental, pernicious, and simultaneously difficult to detect implication of the shift to social media is not the rising power of intermediaries but the inability of regulation to level the playing field for political contest and limit the role of money in elections".[61]

9. OTHER POSSIBLE IMPACTS

The above list of specific human rights that may be impacted through the use of automated processing techniques and algorithms is not exhaustive. It rather aims to project the most obviously implicated rights that are to a stronger or lesser degree already in the public discussion. Human rights and fundamental freedoms are universal, indivisible, inter-dependent and inter-related. As a result, all human rights and fundamental freedom are potentially impacted by the use of algorithmic technologies. Given its limited scope, this study has not engaged in a discussion of the right to life in the context of smart weapons and algorithmically operated drones, or in the context of health and related research. It has further not explored the possible effects that the systematisation of views and opinions

[59] Hannes Grassegger & Mikael Krogerus, "The Data That Turned the World Upside Down", available at: https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win (last visited on 25 September 2017).

[60] Bessi, Alessandro, and Emilio Ferrara. 2016. Social bots distort the 2016 U.S. Presidential election online discussion, FIRST MONDAY, Volume 21, no 11, 7 November 2016, available at: http://journals.uic.edu/ojs/index.php/fm/article/view/7090/5653 (last visited on 25 September 2017).

[61] See Feasibility Study on the use of internet in electoral campaigns (MSI-MED(2016)10rev (ONCE PUBLIC).

through algorithms may have on the right to hold opinions and on the right to freedom of thought, conscience and religion.

Indeed, the increasing use of automation and algorithmic decision-making in all spheres of public and private life is threatening to disrupt the very concept of human rights as protective shields against state interference. The traditional asymmetry of power and information between state structures and human beings is shifting towards an asymmetry of power and information between operators of algorithms (who may be public or private) and those who are acted upon and governed.

IV. REGULATORY IMPLICATIONS OF THE USE OF AUTOMATED DATA PROCESSING TECHNIQUES AND ALGORITHMS

There is growing concern at the political and public level globally regarding the increased use of algorithms and automated processing techniques and their considerable impact on the exercise of human rights. As a result, calls are being made to introduce tighter control and regulation.[62]

Already, there are numerous cases where governments and independent auditors engage in some form of regulation of algorithmic development, usually before operation is commenced. The software and data processing systems, including algorithms, used in 'slot machines' in Australia and New Zealand must, by government regulation, be "fair, secure and auditable" (Woolley et al. 2013). Developers of such machines are required to submit their algorithmic systems to regulators before they can be presented to consumers. The Australian/New Zealand Gaming Machine National Standard in its most recent revision 10.3 defines in extraordinary technical detail how such machines should operate. For example the "Nominal Standard Deviation (NSD) of a game must be no greater than 15" and "the hashing algorithm for the verification of gaming equipment software, firmware and PSDs is the HMAC-SHA1 algorithm".[63] Gambling equipment in the United Kingdom is controlled by a specific licensing regime and, at EU level, regulatory technical standards have been adopted specifying the organisational requirements of investment firms engaged in algorithmic trading.[64] Section 28b of the German Federal Law on Data Protection provides that there has to be a scientifically proven mathematical-statistical process for the calculation of the

[62] See, for instance, the vote on 26 January 2016 in the French National Assembly for a new Bill on digital rights. The Bill includes provisions relating to algorithmic transparency and the duty of 'loyalty', or fairness, of online platforms and algorithmic decision-making" (Rosnay 2016).

[63] The Australian/New Zealand Gaming Machine National Standard which is available at the following link: https://publications.qld.gov.au/dataset/a-nz-gaming-machine-national-standards (last visited on 25 September 2017).

[64] See http://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160719-rts-6_en.pdf (last visited on 25 September 2017).

probability of a specific behaviour of an individual before such an algorithm can be used for making a decision about a contract.[65]

Such licensing systems for algorithms that are used in certain sectors resemble the quality control and assurance schemes employed in the production and manufacturing industry. They are prepared by relevant experts who know and control the respective quality standards in the given field. It is doubtful, however, to what extent such regulatory methods can be exported to the multiple, evolving spheres of public and private life in which automated data processing techniques and algorithms are operated. The British Police Child Exploitation and Online Protection Centre demanded, for instance, that their 'Facebook button' be provided by default to all Internet users (Wagner 2016b). While this attempt to pressure Facebook into changing its default code on the British Facebook website was unsuccessful, it suggests what kind of regulatory responses may be expected if states seek to define the functioning of algorithms on large online platforms.

Fundamental legal and ethical questions surround the legal personhood of automated systems such as algorithms that cannot easily be resolved in this report. While not wishing to exculpate those involved in development, programming and implementation of autonomous systems, it must be acknowledged that automation, vast data analysis and adaptability and self-learning create considerable challenges for accountability of algorithmic decisions. In February 16, 2017 European Parliament adopted resolution calling European Commission to develop legislative proposal for Civil Law Rules on Robotics. Such proposal is expected to address, amongst other things, general principles concerning the development of robotics and artificial intelligence for civil use, ethical principles, liability issues, intellectual property rights and the flow of data, safety, security and other issues.[66]

Historically, challenges related to automated data processing have been addressed through data protection legislation. Today, relevant and innovative approaches such as the introduction of a limited "right to explanation" (Goodman and Flaxman 2016; Wachter et al. 2016) and other rights of internet users are also the product of data protection legislation. However, there is a significant difference between the right to privacy and data protection regulation, which is in the end still a governance mechanism to safeguard privacy and personal data protection rights. Importantly, privacy, as the exercise of other human rights, requires effective enforcement. Some of the greatest challenges in the area of data protection come from a lack of willingness to provide sufficient resources to data protection authorities (DPAs). While it is clear that the challenges around discrimination of content or the manipulation of elections go beyond privacy and data protection and raise fundamental questions on a large set of issues, the expertise of the data protection community may well

---

[65] See German Federal Law on Data Protection, promulgated on 14 January 2003 (Federal Law Gazette I p. 66), and amended by Article 1 of the Act of 14 August 2009 (Federal Law Gazette I p. 2814), available at: https://www.gesetze-im-internet.de/bdsg_1990/__28b.html (last visited on 25 September 2017).

[66] See the European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN&language=EN#BKMD-11 (last visited on 25 September 2017).

be drawn from when attempting to identify suitable regulatory responses to algorithmic governance.

It has been suggested that "[t]echnologists think about trust and assurance for computer systems a bit differently from policymakers, seeking strong formal guarantees or trustworthy digital evidence that a system works as it is intended to or complies with a rule or policy objective rather than simple assurances that a piece of software acts in a certain way" (Kroll et al. 2016).

This in turn feeds into the wider debate on auditing of algorithms by which 'zero knowledge proofs' could conceivably be generated by algorithms to demonstrate that they conform to certain properties, without the individual engaging in the proof being able to see the actual algorithm (Kroll 2016). Beyond zero knowledge proofs, new types of technical accountability may be able to support common human notions of trust and accountability. They could therefore be used in the future as supportive technological approaches for establishing trust, transparency, and accountability.

One major regulatory challenge relating to the use of automated processing techniques and algorithms comes from the strategy adopted by states in some cases to regulate the activities of Internet intermediaries knowingly relying upon automated means rather than end users, which raise transparency, accountability and human rights issues. This is particularly the case in the field of content regulation.

As attempts at regulation may not only in themselves raise human rights concerns but may also be problematic in the sense that regulators may not have developed sufficiently comprehensive expertise to formulate standards that reflect not only the technological and engineering perspectives but also legal and ethical considerations, efforts towards promoting greater transparency and accountability surrounding the use of algorithms seem more appropriate initial steps than direct regulation.[67] Such standards would also need to be combined with high-level technology neutral regulations.

While regulatory restraint is therefore warranted at this stage of the implementation of algorithms and automated processing techniques, their implications for human rights (Section III) and ethical considerations must be carefully examined. In particular, the current academic discourse has centred on concepts such as human autonomy and individual agency, both related to the right to privacy (Section 2) and informational self-determination but not congruent with privacy. Therefore, autonomy and agency should be considered separately. They refer to the human capability to set one's own goals and the human capability to make decisions and exercise discretion and as such may conflict with the use of algorithms and automated processing techniques. This may mean that human rights may have to be extended or reinterpreted to protect individual autonomy and agency.

[67] For further examples see Chapter 5 of Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

1. TRANSPARENCY

Algorithms are often viewed as black boxes by both consumers and regulators alike (Pasquale 2015). Demands for more algorithmic transparency have thus been growing in public and political debate,[68] including government requests to companies regarding algorithms that should be reviewed by independent auditors, regulators or the general public (Diakopoulos 2015; Rosnay 2016) before their implementation.[69]

Importantly, these challenges exist not just for professionals who develop algorithms but also for other groups such as 'data scientists' who use them. Different levels of abstraction and complexity prompt distinct challenges of opacity and transparency. It has been frequently argued that much of the usage of algorithms in machine learning takes places without "understanding" causal relationships (correlation instead of causation), which may lead to bias and errors and raise concerns about data quality (O'Neil 2016). The challenge, however, relates also to the way human beings use, perceive and interpret their results. The belief that computer algorithms produce neutral unbiased results (Chun 2006) without any form of politics (Denardis 2008) is at the heart of this problem. Accordingly, it would be more helpful to ensure more critical engagement in public debates about algorithms than to attempt to change them.

The provision of entire algorithms or the underlying software code to the public is an unlikely solution in this context, as private companies regard their algorithm as key proprietary software that is protected.[70] However, there may be a possibility of demanding that key subsets of information about the algorithms be provided to the public, for example which variables are in use, which goals the algorithms are being optimised for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm.

Key in this context is not the provision of all data imaginable, but rather the notion of "effective transparency". The underlying goal of increasing transparency must actually be met by the data disclosed, which implies that the demand for 'more data' may not always

---

[68] Angela Merkel, for instance, has called on major internet platforms to divulge information on their algorithms as internet users had a right to know on what basis the information they received via search engines was channeled to them. See The Guardian, "Angela Merkel: internet search engines are 'distorting perception'", available at: https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception (last visited on 25 September 2017).

[69] See also Tufekci et al note: "a common ethical concern about algorithmic decision-making is the opaque nature of many algorithms. When algorithms are employed to make straightforward decisions, such as in the case of medical diagnostics or aviation, a lack of transparency raises important questions of accountability" (Tufekci et al. 2015:11).

[70] In a decision of 28 January 2014, the German Federal Supreme Court (Bundesgerichtshof) rejected a claim for information concerning a credit agency's algorithm as it was a protected business secret. It, however, allowed a claim for information concerning the data used to calculate creditworthiness through the means of the algorithm; see German Federal Supreme Court (Bundesgerichtshof), Judgment, 28 January 2014 Az. VI ZR 156/13, available at: https://openjur.de/u/677956.html.

be helpful but may, in the worst case, even serve to counteract the goal of enhancing transparency.

Effective transparency of automated systems is complicated, however, by the frequent changes in the algorithms that are used. Google, for example, changes its algorithm hundreds of times per year (Tufekci et al. 2015). There is also the danger of manipulation and 'gaming' of algorithms if they are made public. Moreover, machine learning techniques complicate transparency to a point where provision of all of the source codes of an algorithm may not even be sufficient. Instead, there is a need for an actual explanation of how the results of an algorithm were produced. Since algorithms may actively obscure that a consequential decision is taken, transparency promotion measures may also be targeted at the decision-making process itself, given that algorithms cannot meaningfully be studied outside of their social and organisational context.

Transparency enhancement measures, finally, may not only facilitate scrutiny by the public but also by independent experts, commissions or specialised agencies which, in turn, may support efforts to promote compliance with consumer protection and human rights standards.

2. ACCOUNTABILITY

Accountability is the principle that a person who is legally or politically responsible for harm has to provide some form of justification or compensation. However someone can only be accountable if they have a degree of control in the sense that they have facilitated or caused the harm or are in a position to prevent or mitigate it. Legally, accountability manifests itself through the concept of liability to provide a remedy (such as damages). The law usually imposes liability on the person who is in a position to prevent harm or mitigate a risk (for example through insurance). The allocation of accountability for algorithmic decision-making is complicated by the fact that frequently it is not clear who has the necessary degree of control to be imputed with legal or political accountability.

One aspect here is that the developer of algorithmic tools may not know their precise future use and implementation. The person(s) implementing the algorithmic tools for applications may, in turn, not fully understand how the algorithmic tools operate. Are those developing and programming the algorithm to be held accountable? Some authors have suggested that algorithmic accountability and regulation are impossible because the programmers themselves are unable to predict or fully understand how the algorithm takes the decisions that it makes (Kroll 2016). Another avenue to explore is whether existing product liability regulation should be extended to include software? Or are rather the public or private actors to be held accountable who purchase the algorithm and introduce it into their services, even without understanding its operation?

The governance failure in the automobile emissions scandal also exemplifies the wider challenge of enhancing accountability of algorithms across numerous different sectors. Whether in the criminal justice, social media, healthcare, insurance or banking sector, to name just a few examples, each area will need specific regulatory responses to ensure greater transparency and accountability of automated data-processing and algorithmic decision-making systems. Algorithmic accountability must further be safeguarded through

due process and the rule of law. Effective redress mechanisms for individuals whose rights were infringed by automated decision-making systems are also essential.

Such an approach places a challenging duty on the operators of algorithms and automated data processing techniques, whether public or private, to ensure basic standards of human rights. These fundamental principles cannot be offset with arguments of possible greater efficiency of opaque technological systems (Wagner 2016a). Similar issues arise in relation to private actors who employ algorithms and automated data processing techniques in their operations, particularly when they are market-dominant. Owing to the size and scale of their activities, they deliver services with important public service value which, in turn, may also have an important impact on the enjoyment of human rights.

The accountability of individuals or companies with respect to the algorithms they implement depends very much on the nature of the algorithms and their outputs. In some cases, if the outputs are defamatory, infringe copyright or raise other legal concerns, existing governance mechanisms ensure that these kinds of outputs are limited (Staab, Stalla-Bourdillon, and Carmichael 2016). However, such mechanisms typically only regard the outputs of algorithms, but not the algorithms themselves. In fact, there is a general lack of regulatory frameworks that ensure that algorithms, in the first place, are programmed to produce results that uphold and protect fundamental values or basic ethical and societal principles.

This touches upon fundamental ethical questions with respect to the operation of automated data processing techniques and algorithms that are not addressed in this study. How can normative values be reflected in an automated system? Some of the ethical discussions surrounding the self-driving car provide an insight into the complexity of the challenge: how should the algorithm decide in the hypothetical situation where a likely accident may either threaten the life of a young child or the life of an elderly person? Does the number of lives possibly at stake play a role? What are "right" or "wrong" decisions in such a situation, and with what legal consequence? Who is held accountable in case a "wrong" decision is taken?

3. ETHICAL FRAMEWORKS AND IMPROVED RISK ASSESSMENT

Aside from direct regulatory mechanisms to influence the code of algorithms, indirect mechanisms to influence algorithm codes could also be considered. These address the production process or the producers of algorithms and attempt to ensure that they are aware of the legal challenges, ethical dilemmas and human rights concerns that arise from automated data-processing and decision-making techniques. An instrument to achieve such goals could consist of standardised professional ethics or forms of licensing system for data engineers and algorithm designers similar to those that exist for professions like doctors, lawyers or architects. Another suggestion frequently made is that existing mechanisms for the management and development processes of software could be improved (Spiekermann 2015). This may particularly concern agile software development techniques where modularity, temporality and capture pose considerable challenges for privacy (Gürses and Hoboken 2017) as well as other human rights (Mannaro 2008). As the use of algorithms in decision-making potentially prejudices the rights of individuals, additional oversight

mechanisms could contribute to ensuring that the algorithm operates in a fair and sustainable manner.

In order to assess and understand the human rights risks involved with operating automated decision making systems, companies can exercise human right due diligence. This can take the form of human rights impact assessments, investigating the concrete and potential impacts on individuals that the employment of these systems may have, whether direct or indirect, and preventing or mitigating harms identified in these assessments.

There are examples of emerging standards by industry associations such as the IEEE (Institute of Electrical and Electronics Engineers) on algorithms, transparency, privacy, bias and more broadly on ethical system design and the Internet Engineering Task Force (IETF):

- IEEE P7000: Model Process for Addressing Ethical Concerns During System Design
- IEEE P7001: Transparency of Autonomous Systems
- IEEE P7002: Data Privacy Process
- IEEE P7003: Algorithmic Bias Considerations
- IETF Research into Human Rights Protocol Considerations draft

Other examples of relevant industry frameworks that could support greater levels of human rights compliance include the FAT-ML (Fairness, Accountability, Transparency in machine learning) principles for more accountable algorithms.[71]

That we see a frequent use of the word "ethics" in connection with algorithms among experts but also in the public debate may be an indicator for a tactical move by some actors who want to avoid strict regulation by pointing to non-formal normative concepts. It may, however, also point to the need for deeper reflection about the interplay of different types of norms and the role and responsibility of various actors in order to shape the governance structure for algorithmic decision-making and "ethics" as a new set of applicable meta-norms.

V. MAIN FINDINGS AND CONCLUSIONS

The notion of 'algorithmic processing and decision-making' is diversely interpreted and understood in legal, technological, and social science circles, and again differently amongst the public. In addition, the field is comparatively new. Awareness of impacts for the exercise of human rights and for broader societal development has grown only recently and is yet to translate into a wider and inclusive public policy debate on possible regulatory implications.

The authors of this study acknowledge that there is far too little information available to make well-founded decisions on this topic and thus considerable additional research and analysis is required, including with respect to the characteristics of human decision-making processes. As decision-making processes by human beings are not necessarily "better" than but simply different automated decision-making systems, different kinds or bias, risk or error are likely to develop in automated decision-making. Thus it needs to be openly

[71] See http://www.fatml.org/resources/principles-for-accountable-algorithms (last visited on 25 September 2017).

discussed what criteria should be developed to measure the quality of automated-decision making.

It is highly welcome that there is increasing research on these topics. However, academic research on its own is insufficient. It is essential to ensure that members of professional (technological, engineering, legal, media, philosophical and ethical) communities engage in discussions and debates that must also include the general public. In order to promote active engagement of human beings and a lively public debate about an issue that affects all human beings and communities, adequate media and information literacy promotion activities should be organised to facilitate the empowerment of the public to critically understand and deal with the logic and operation of algorithms. Notably, public entities and governments must have access to sufficiently comprehensive information to properly understand algorithmic decision-making systems that are already deeply embedded in societies across the world. To provide just one concrete example of this problem, the automobile emissions scandal demonstrates what can happen when a small piece of frequently used software is widely implemented without adequate independent regulatory scrutiny. It is undesirable from a human rights perspective that there are powerful publicly-relevant algorithmic systems that lack a meaningful form of public scrutiny. The application of a human rights framework is crucial because it goes beyond just ensuring transparency and accountability, as it ensures that all rights are effectively considered in automated decision-making systems such as algorithms. This is no simple task and will require a combination of further developing industry standards which put human beings and human rights at the centre of the technology design process, and effective regulatory measures to ensure that when industry standards fail governments are able to step in to promote and protect human rights.

Human beings have a right to effectively scrutinise the decisions made by public authorities. Issues related to algorithmic governance and/or regulation are public policy prerogatives and should not be left to private actors alone. While these may engage in voluntary measures to promote transparency and accountability within their operations, and while they have a duty of care towards their users and the responsibility to respect human rights, the task of devising comprehensive and effective mechanisms for ensuring algorithmic accountability lies on the states. This is crucial not only because of the important impact of automated data processing techniques and algorithms on the exercise and enjoyment of human rights, but also because of their capacity to expand, reinforce and redistribute power, authority and resources in society.

Importantly, there may be areas of societal and human interaction where algorithmic decision-making systems are not appropriate. Automated data processing and decision-making systems should not be relied upon heavily to promote societal development or resolve complex new challenges for future generations, as this is likely to do more harm than good. Therefore it is critical to ensure that in key areas where automation is not appropriate from a human rights perspective, it does not take place.

It is the view of the authors of this study that the public debate on the multiple human rights dimensions of algorithms is lagging behind technological evolution and must be strengthened rapidly to ensure that the human rights and interests of individuals are

effectively and sustainably safeguarded in line with the values laid down in the European Convention and other international treaties. The use of algorithms and other automated data processing techniques can potentially have positive and negative impacts on the exercise and enjoyment of human rights. It must be the aim of policy makers to ensure that these technologies are used in line with the principle of the "primacy of the human being",[72] and that our increasingly technology-driven societies are designed - first and foremost - with the effective exercise and enjoyment of the rights of all human beings in mind.

In consequence, this study comes to the <u>following conclusions</u>:

1. Public entities and independent non-state actors should initiate and support research that helps to better understand and respond to the human rights, ethical and legal implications of algorithmic decision-making. Therefore, they should support and engage with trans-disciplinary, problem-orientated and evidence-based research, as well as the exchange of best practices.

2. Public entities should be held responsible for the decisions they take based on algorithmic processes. The adoption of mechanisms should be encouraged that enable redress for individuals that are negatively impacted by algorithmically informed decisions. Human rights impact assessments should be conducted before making use of algorithmic decision-making in all areas of public administration.

3. Technological developments should be monitored closely and reviewed for potential negative impacts, with particular attention paid to the use of algorithmic processing techniques during elections and election campaigns. Effective responses to such negative impacts could include experimental regulatory approaches on how best to protect rights of others and guarantee regulatory goals, provided they are accompanied with systematic monitoring of their effects.

4. Public awareness and discourse are crucially important. All available means should be used to inform and engage the general public so that users are empowered to critically understand and deal with the logic and operation of algorithms. This can include but is not limited to information and media literacy campaigns. Institutions using algorithmic processes should be encouraged to provide easily accessible explanations with respect to the procedures followed by the algorithms and to how

[72] See also *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*, Report by the Rathenau Institut commissioned and funded by the Parliamentary Assembly of the Council of Europe, adopted by PACE on 28 April 2017.

decisions are made. Industries that develop the analytical systems used in algorithmic decision-making and data collection processes have a particular responsibility to create awareness and understanding, including with respect to the possible biases that may be induced by the design and use of algorithms.

5. Certification and auditing mechanisms for automated data processing techniques such as algorithms should be developed to ensure their compliance with human rights. Public entities and non-state actors should encourage and promote the further development of human rights by design and ethical-by-design approaches and the adoption of stronger risk-assessment approaches in the development of software.

6. States should not impose a general obligation on internet intermediaries to use automated techniques to monitor information that they transmit, store or give access to, as such monitoring infringes on users' privacy and has a chilling effect on the freedom of expression.

7. Public entities should engage with their own sector-regulators (insurance, credit reference agencies, banks, e-commerce and others) to develop specific standards and guidelines to ensure that they are able to respond to the challenges of the use of automated decision-making through algorithms and taking into account the interests of consumers and the general public.

8. Considering the complexity of the field, awareness of the general public – important as it is – will not suffice. There is an evident need for additional institutional arrangements. Therefore, public entities should initiate and support the creation of networks and spaces for all relevant stakeholders to analyse and assess different forms of algorithmic decision-making. All relevant stakeholders should engage in such an endeavour.

9. The Council of Europe as the continent's leading human rights organisation is the appropriate venue to further explore the impacts on the effective exercise of human rights of the increasing use of automated data processing and decision-making systems (in particular algorithms) in public and private spheres. It should continue its endeavours in this regard with a view to developing appropriate standards-setting instruments for guidance to member states.

BIBLIOGRAPHY

Altonji, JG and RM Blank. 1999. 'Race and Gender in the Labor Market'. Pp. 3143–3259 in Handbook of labor economics. Elsevier B.V. Retrieved (http://www.sciencedirect.com/science/article/pii/S1573446399300390).

Andreessen, Marc. 2011. 'Why Software Is Eating The World'. Wall Street Journal, August 20. Retrieved 1 September 2016 (http://www.wsj.com/articles/SB10001424053111903480904576512250915629460).

Angwin, Julia, Surya Mattu, and Lauren Kirchner. 2016. 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.' ProPublica. Retrieved 31 August 2016 (https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing).

Bertrand, Marianne and Sendhil Mullainathan. 2004. 'Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination'. The American Economic Review 94(4):991–1013.

Bond, Robert M. et al. 2012. 'A 61-Million-Person Experiment in Social Influence and Political Mobilization'. Nature 489(7415):295–298.

Bozdag, Engin. 2013. 'Bias in Algorithmic Filtering and Personalization'. Ethics and Information Technology 15(3):209–227.

Bucher, Taina. 2012. 'Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook'. New Media & Society 1461444812440159.

Bucher, Taina. 2016. 'The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms'. *Information, Communication & Society* 1–15.

Buni, Catherine and Soraya Chemaly. 2016. 'The Secret Rules of the Internet'. *The Verge*. Retrieved 9 September 2016 (http://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech).

Chun, Wendy Hui Kyong. 2006. *Control and Freedom : Power and Paranoia in the Age of Fiber Optics*. Cambridge Mass.: MIT Press.

Clay, Kelly. 2012. 'Is Microsoft Spying On SkyDrive Users?' Forbes. Retrieved 31 August 2016 (http://www.forbes.com/sites/kellyclay/2012/07/19/is-microsoft-spying-on-skydrive-users/).

Denardis, Laura. 2008. 'Architecting Civil Liberties'. in *Global Internet Governance Academic Network Annual Meeting*. Hyderabad (Andra Pradesh), India: GIGANET. Retrieved (http://worldcat.org/oclc/619234880/viewonline).

Denardis, Laura. 2012. 'Hidden Levers of Internet Control'. *Information, Communication & Society* (September):37–41.

Diakopoulos, Nicholas. 2015. 'Algorithmic Accountability'. *Digital Journalism* 3(3):398–415.

Edwards, Lilian and Veale, Michael. 2017. 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For', Duke Law & Technology Review (Forthcoming), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855.

Fernández Pérez, Maryant. 2017. 'Parliamentarians Encourage Online Platforms to Censor Legal Content'. *EDRi*. Retrieved 7 June 2017 (https://edri.org/parliamentarians-encourage-online-platforms-to-censor-legal-content/).

Gillespie, Tarleton. 2014. 'The Relevance of Algorithms'. Pp. 167–94 in *Media technologies: Essays on communication, materiality, and society*, edited by T. Gillespie, P. J. Boczkowski, and K. A. Foot. Cambridge Mass.: MIT Press.

Goldin, Claudia and Rouse, Cecilia. 1997. *Orchestrating Impartiality: The Impact Of' blind' auditions on Female Musicians*. National bureau of economic research. Retrieved 9 September 2016 (http://www.nber.org/papers/w5903).

Goodman, Bryce and Seth Flaxman. 2016. 'European Union Regulations on Algorithmic Decision-Making and a Right to Explanation'. in *2016 ICML Workshop on Human Interpretability in Machine Learning*. New York, NY: ArXiv e-prints.

Griffin, Andrew. 2016. 'How Facebook Is Manipulating You to Vote'. *The Independent*. Retrieved 31 August 2016 (http://www.independent.co.uk/life-style/gadgets-and-tech/news/uk-elections-2016-how-facebook-is-manipulating-you-to-vote-a7015196.html).

Grindrod, Peter. 2014. Mathematical Underpinnings of Analytics: Theory and Applications for Data Science in Customer-Facing Industries. Oxford: Oxford Univ. Press.

Gürses, Seda and Joris Hoboken. 2017. 'Privacy After the Agile Turn'. in *The Cambridge Handbook of Consumer Privacy*, edited by Selinger. Retrieved (https://osf.io/ufdvb/).

van Haastert, Hugo. 2016. 'Government as a Platform: Public Values in the Age of Big Data'. Oxford Internet Institute.

Helberger, Natali and Damian Trilling. 2016. 'Facebook Is a News Editor: The Real Issues to Be Concerned about'. *Media Policy Project*. Retrieved 9 September 2016 (http://blogs.lse.ac.uk/mediapolicyproject/2016/05/26/facebook-is-a-news-editor-the-real-issues-to-be-concerned-about/).

Hendrickx, Frank and Aline van Bever. 2013. 'Article 8 ECHR: Judicial Patterns of Employment Privacy Protection'. Pp. 183–208 in *The European Convention on Human Rights and the Employment Relation*, edited by F. Dorssemont, K. Lörcher, and I. Schömann. Oxford: Hart Publishing.

Hildebrandt, Mireille and Serge Gutwirth. 2008. 'General Introduction and Overview'. Pp. 1–13 in Profiling the European Citizen. Springer, Dordrecht. Retrieved 26 September 2017 (https://link.springer.com/chapter/10.1007/978-1-4020-6914-7_1).

Hoofnagle Chris Jay "Behavioural Advertising: The Offer You Cannot Refuse" (2012) 6 Harvard Policy & Law Review 273-296.

Irani, L. 2015. 'Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk'. *South Atlantic Quarterly* 114(1):225–234.

Jędrzej Niklas, Karolina Sztandar-Sztanderska, and Katarzyna Szymielewicz. 2015.Warsaw, Poland: Panoptykon Foundation. Retrieved (https://en.panoptykon.org/articles/profiling-unemployed-poland-%E2%80%93-report).

Kim H, Giacomin J and Macredie R (2014) A qualitative study of stakeholders' perspectives on the social network service environment. International Journal of Human– Computer Interaction 30(12): 965–976.

Kitchin, R. and M. Dodge. 2011. *Code/Space Software and Everyday Life*.

Kocher, Eva and Isabell Hensel. 2016. 'Herausforderungen Des Arbeitsrechts Durch Digitale Plattformen – Ein Neuer Koordinationsmodus von Erwerbsarbeit'. *Neue Zeitschrift Für Arbeitsrecht* (16/2016):984–89.

Kroll, Joshua A. et al. 2016. 'Accountable Algorithms'. Retrieved 1 September 2016 (http://balkin.blogspot.co.at/2016/03/accountable-algorithms.html).

Kroll, Joshua A. 2016. 'Accountable Algorithms (A Provocation)'. *Media Policy Project*. Retrieved 9 September 2016 (http://blogs.lse.ac.uk/mediapolicyproject/2016/02/10/accountable-algorithms-a-provocation/).

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. 'The Parable of Google Flu: Traps in Big Data Analysis'. Science 343(6176):1203–5.

Lazer, David and Ryan Kennedy. 2015. What We Can Learn from the Epic Failure of Google Flu Trends.

Loo, Van. 2016. *The Corporation as Courthouse*. Rochester, NY: Social Science Research Network. Retrieved 7 June 2017 (https://papers.ssrn.com/abstract=2872096).

Mannaro, Katiuscia. 2008. 'Adopting Agile Methodologies in Distributed Software Development'. Universita` degli Studi di Cagliari, Cagliari. Italy. Retrieved (http://le.uwpress.org/content/87/2/284.short).

McCarthy, Daniel R. 2011. 'Open Networks and the Open Door: American Foreign Policy and the Narration of the Internet'. *Foreign Policy Analysis* 7(1):89–111.

Mueller, Milton. 2010. *Networks and States: The Global Politics of Internet Governance*. MIT Press.

Nguyen, Tien T., Pik-Mai Hui, F.Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. 'Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity'. Pp. 677–686 in *Proceedings of the 23rd International Conference on World Wide Web*, *WWW '14*. New York, NY, USA: ACM. Retrieved (http://doi.acm.org/10.1145/2566486.2568012).

Nikolaos Altreas et al "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective" PeerJ Computer Science Open Access (Published 24. October 2016)

O'Callaghan, D., D. Greene, M. Conway, J. Carthy, and P. Cunningham. 2015. 'Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems'. *Social Science Computer Review Social Science Computer Review* 33(4):459–78.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pasquale, Frank A. 2016. *Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power*. Rochester, NY: Social Science Research Network. Retrieved 7 June 2017 (https://papers.ssrn.com/abstract=2779270).

Perry, Walt L. 2013. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Rand Corporation. Retrieved 9 September 2016 (https://books.google.com/books?hl=en&lr=&id=ZdstAQAAQBAJ&oi=fnd&pg=PP1&dq=Perry,+Walter,+and+Brian+McInnis.+2013.+Predictive+Policing:+The+Role+of+Crime+Forecasting+in+Law+Enforcement+Operations+Santa+Monica,+CA:+RAND.&ots=924yNa6Vct&sig=N3HnEi1FBr9YyMXV77GsgPbovYc).

Rifkind, Malcolm. 2014. *Report on the Intelligence Relating to the Murder of Fusilier Lee Rigby*.

Rosenblat, Alex, Tamara Kneese, and others. 2014. 'Networked Employment Discrimination'. *Open Society Foundations' Future of Work Commissioned Research Papers*. Retrieved 9 September 2016 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2543507).

Rosenblat, Alex, Karen EC Levy, Solon Barocas, and Tim Hwang. 2016. 'Discriminating Tastes: Customer Ratings as Vehicles for Bias'. Retrieved 7 June 2017 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2858946).

Rosnay, Mélanie Dulong de. 2016. 'Algorithmic Transparency and Platform Loyalty or Fairness in the French Digital Republic Bill'. *Media Policy Project*. Retrieved 1 September 2016 (http://blogs.lse.ac.uk/mediapolicyproject/2016/04/22/algorithmic-transparency-and-platform-loyalty-or-fairness-in-the-french-digital-republic-bill/).

Rubinstein, Ira, Ronald D. Lee, and Paul M. Schwartz. 2008. *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*. Rochester, NY: Social Science Research Network. Retrieved 9 September 2016 (http://papers.ssrn.com/abstract=1116728).

Salamatian, Kavé. 2014. 'From Big Data to Banality of Evil'. Retrieved 9 September 2016 (https://www.oximity.com/article/Vortrag-Big-Data-und-Ethik-1).

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2016. 'When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software'. *International Journal of Communication* 10:19.

Schulz, Wolfgang and Kevin Dankert. 2016. ''Governance by Things' as a Challenge to Regulation by Law'. *Internet Policy Review* 5(2).

Sills, Arthur J. 1970. 'Automated Data Processing and the Issue of Privacy'. *Seton Hall Law Review* 1.

Spiekermann, Sarah. 2015. *Ethical IT Innovation: A Value-Based System Design Approach*. CRC Press.

Staab, Steffen, Sophie Stalla-Bourdillon, and Laura Carmichael. 2016. 'Observing and Recommending from a Social Web with Biases'. *arXiv Preprint arXiv:1604.07180*. Retrieved 9 September 2016 (http://arxiv.org/abs/1604.07180).

Tene, Omer and Jules Polonetsky. 2012. 'To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising'. Retrieved 9 September 2016 (http://conservancy.umn.edu/handle/11299/155947).

Toor, Amar. 2016. 'Automated Systems Fight ISIS Propaganda, but at What Cost?' *The Verge*. Retrieved 9 September 2016 (http://www.theverge.com/2016/9/6/12811680/isis-propaganda-algorithm-facebook-twitter-google).

Tufekci, Zeynep, Jillian C. York, Ben Wagner, and Frederike Kaltheuner. 2015. *The Ethics of Algorithms: From Radical Content to Self-Driving Cars*. Berlin, Germany: European University Viadrina. Retrieved (https://cihr.eu/publication-the-ethics-of-algorithms/).

Tufekci, Zeynep, Twitter and Tear Gas: The Power and Fragility of Networked Protest, Yale University Press, 2017.

Tversky, Amos and Daniel Kahneman. 1974. 'Judgment under Uncertainty: Heuristics and Biases'. *Science* 185(4157):1124–31.

Urban, Jennifer M., Joe Karaganis, and Brianna L. Schofield. 2016. 'Notice and Takedown in Everyday Practice'. Available at SSRN 2755628. Retrieved 28 October 2016 (http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2755628).

Voorhoof, Dirk and P. Humblet, eds. 2013. 'The Right to Freedom of Expression in the Workplace under Article 10 ECHR'. Pp. 183–208 in *The European Convention on Human Rights and the Employment Relation*. Oxford: Hart Publishing.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2016. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. Rochester, NY: Social Science Research Network. Retrieved 7 June 2017 (https://papers.ssrn.com/abstract=2903469).

Wagner, Ben. 2016a. 'Efficiency vs. Accountability?' Bureau de Helling. Retrieved 10 March 2017 (https://bureaudehelling.nl/artikel-tijdschrift/efficiency-vs-accountability).

Wagner, Ben. 2016b. Global Free Expression: Governing the Boundaries of Internet Content. Cham, Switzerland: Springer International Publishing.

Williamson, Ben. 2016. 'Computing Brains: Learning Algorithms and Neurocomputation in the Smart City'. Information, Communication & Society 0(0):1–19.

Winner, L. 1980. 'Do Artifacts Have Politics?' Daedalus.

Winner, L. 1986. 'The Whale and the Reactor: A Search for Limits in an Age of High Technology'.

Woolley, Richard, Charles Livingstone, Kevin Harrigan, and Angela Rintoul. 2013. 'House Edge: Hold Percentage and the Cost of EGM Gambling'. International Gambling Studies 13(3):388–402.

York, Jillian C. 2010. 'Policing Content in the Quasi-Public Sphere'. Boston, MA: Open Net Initiative Bulletin. Berkman Center. Harvard University.

Zhang, Pei, Sophie Stalla-Bourdillon, and Lester Gilbert. 2016. 'A Content-Linking-Context Model For "notice-and-Take-Down" procedures'. Pp. 161–65 in. ACM Press. Retrieved 9 September 2016 (http://dl.acm.org/citation.cfm?doid=2908131.2908171).

Zittrain, Jonathan, Engineering an election, Harvard Law Review Forum Vol. 127, 335 – 339 (2014).

Zuckerman, Ethan. 2013. Digital Cosmopolitans: Why We Think the Internet Connects Us, Why It Doesn't, and How to Rewire It. W. W. Norton & Company.

Zuiderveen Borgesius, Frederik J. et al. 2016. 'Should We Worry About Filter Bubbles?' Internet Policy Review. Journal on Internet Regulation 5(1). Retrieved 1 September 2016 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758126).

REFERENCES

I) Council of Europe's instruments

European Convention on Human Rights (ETS no 5).

Automatic Processing of Personal Data Convention (ETS no 108).

Declaration on freedom of communication on the Internet of 28 May 2003.

Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling, adopted on 23 November 2010.

Recommendation CM/Rec(2012)3 of the Committee of Ministers to member States on the protection of human rights with regard to search engines, adopted on 4 April 2012.

Recommendation CM/Rec(2012)4 of the Committee of Ministers to member States on the protection of human rights with regard to social networking services, adopted on 4 April 2012.

Recommendation CM/Rec(2014)6 of the Committee of Ministers to member States on a Guide to human rights for Internet users, adopted on 26 April 2014.

Recommendation CM/Rec(2016)5 of the Committee of Ministers to member States on Internet freedom, adopted on 13 April 2016.

Guidelines on the Protection of Individuals with regard to the Processing of Personal Data in a World of Big Data, 17 January 2017.

Draft Recommendation of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries, finalized by the MSI-NET on 19 September 2017, available at https://rm.coe.int/draft-recommendation-on-internet-intermediaries-version-4/1680759e67.

II) European Union Instruments

Directive 2000/31/EC of 8 June 2000 of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ("Directive on electronic commerce").

Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data establish data protection safeguards.

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//EN&language=EN#BKMD-11.

III) Newspapers and online articles

The great question of the 21st century: Whose black box do you trust?, available at: https://www.linkedin.com/pulse/great-question-21st-century-whose-black-box-do-you-trust-tim-o-reilly?trk=eml-b2_content_ecosystem_digest-hero-22-null&midToken=AQGexvwxq0Q3iQ&fromEmail=fromEmail&ut=2SrYDZ8lkCS7o1.

Article 19, "Algorithms and Automated Decision-Making in the Content of Crime Prevention: A Briefing paper", 2016.

Das Magazin, Hannes Grassegger und Mikael Krogerus, "Ich habe nur gezeigt, dass es die Bombe gibt", no 48, 3 December 2016, available at https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/.

Reuters institute, Emma Goodman, "Editors vs algorithms: who do you want choosing your news?", available at: http://reutersinstitute.politics.ox.ac.uk/news/editors-vs-algorithms-who-do-you-want-choosing-your-news.

GCN, Kevin McCaney, "Prisons turn to analytics software for parole decisions", 1 November 2013, available at: https://gcn.com/articles/2013/11/01/prison-analytics-software.aspx.

Stanford news, New Stanford research finds computers are better judges of personality than friends and family, available at: http://news.stanford.edu/2015/01/12/personality-computer-knows-011215/.

The Guardian, "The great British Brexit robberty: how our democracy was hijacked", 7 May 2017, available at: https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy.

Roheeni Saxena, arstecnica, "The social media "echo chamber" is real", available at: https://arstechnica.com/science/2017/03/the-social-media-echo-chamber-is-real/.

Article 19, "Algorithms and automated decision-making in the context of crime prevention", 2 December 2016, available at: https://www.article19.org/resources.php/resource/38579/en/algorithms-and-automated-decision-making-in-the-context-of-crime-prevention.

Joseph Menn, Dustin Volz, Reuters, Exclusive: Google, "Facebook quietly move toward automatic blocking of extremist videos", available at: http://www.reuters.com/article/us-internet-extremism-video-exclusive-idUSKCN0ZB00M.

The Guardian, "2016: the year Facebook became the bad guy", available at: https://www.theguardian.com/technology/2016/dec/12/facebook-2016-problems-fake-news-censorship.

Pablo Barberá and Megan Metzger, "Tweeting the Revolution: Social Media Use and the #Euromaidan Protests", available at: http://www.huffingtonpost.com/pablo-barbera/tweeting-the-revolution-s_b_4831104.html.

Tim de Chant, "The Inevitability of Predicting the Future", available at: http://www.pbs.org/wgbh/nova/next/tech/predicting-the-future/.

Till Krause and Hannes Grassegger, Süddeutsche Zeitung, "Inside Facebook", available at: http://international.sueddeutsche.de/post/154513473995/inside-facebook.

Marietje Schaake, "When You Tube took down my video", available at: https://www.marietjeschaake.eu/en/when-youtube-took-down-my-video.

The Guardian, "AI programs exhibit racial and gender biases, research reveals", available at: https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals.

The Guardian, "How algorithms rule our working lives", available at: https://www.theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives.

Laurel Eckhouse, "Big data may be reinforcing racial bias in the criminal justice system", available at: https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d_story.html?utm_term=.720084735d73.

ProPublica, "Machine Bias", available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Hannes Grassegger & Mikael Krogerus, "The Data That Turned the World Upside Down", available at: https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win.

Alessandro Bessi and Emilio Ferrara, "Social bots distort the 2016 U.S. Presidential election online discussion", FIRST MONDAY, Volume 21, Number 11, 7 November 2016, available at: http://journals.uic.edu/ojs/index.php/fm/article/view/7090/5653.

The Guardian, "Angela Merkel: internet search engines are 'distorting perception'", available at: https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception.

IV) Miscellaneous

UNESCO's World Trends in Freedom of Expression and Media Development Publication, 2014, available at: http://www.unesco.org/new/en/world-media-trends.

UK Intelligence and Security Committee of Parliament report, Privacy and Security: A modern and transparent legal framework, March 2015, available at: http://isc.independent.gov.uk/committee-reports/special-reports.

EC Communication From The Commission To The European Parliament, The European Council And The Council delivering on the European Agenda on Security to fight against terrorism and pave the way towards an effective and genuine Security Union, available at: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/policies/european-agenda-security/legislative-documents/docs/20160420/communication_eas_progress_since_april_2015_en.pdf.

2016 Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, to the Thirty-second session of the Human Rights Council (A/HRC/32/ 38).

Joint Opinion of the Venice Commission, the Directorate of information society and action against crime and of the Directorate of Human Rights (DHR) of the Directorate General of Human Rights and Rule of Law (DGI) of the Council of Europe on the Draft Law n° 281 amending and completing Moldovan Legislation on the so-called "Mandate of security", adopted by the Venice Commission at its 110th Plenary Session (Venice, 10-11 March 2017) http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2017)009-e.

U.N. Human Rights Council Resolution on the Right to Privacy in the Digital Age, U.N. Doc. A/HRC/34/7, 23 Mar. 2017.

Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality, Report by the Rathenau Institut commissioned and funded by the Parliamentary Assembly of the Council of Europe, adopted by PACE on 28 April 2017.

* * *