Committee d'Experts sur les

intermédiaires internet

(MSI-NET)

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

12 septembre 2016

MSI-NET(2016)06

PROJET DE RAPPORT SUR LES DIMENSIONS DES DROITS DE L'HOMME DANS L'APPLICATION DES ALGORITHMES

RAPPORTEUR: BEN WAGNER

Directeur du Centre pour l'Internet et les droits de l'homme

Université européenne Viadrina, Francfort sur l'Oder

Allemagne

SOMMAIRE

SUMMARY AND MAIN CONCLUSIONS

[Summary will be inserted here once a final version of the study has been completed]

1. INTRODUCTION

À quelles informations avez-vous accès sur votre flux Facebook ? Qui est un criminel ou un terroriste ? Aurez-vous droit à une assurance-maladie ? Allons-nous vous donner un emploi ? Il fut un temps où des êtres humains répondaient à ces questions. Aujourd'hui, ce sont de plus en plus souvent des systèmes décisionnels automatisés. Ces systèmes sont source d'enjeux considérables, comme la défense des droits fondamentaux et de la dignité humaine face à des technologies en constante évolution et ce, non seulement dans chaque domaine d'action où ils sont utilisés, mais également dans l'ensemble des sociétés. Le droit à des élections libres, les droits des travailleurs, le droit à la vie, à la liberté d'expression, au respect de la vie privée et même l'État de droit subissent tous l'influence de ces mutations, comme le montrera l'étude suivante. Aussi, il n'est pas vraiment surprenant que les moyens de relever des défis associés aux « algorithmes » et aux prestataires intermédiaires de l'internet soient actuellement l'une des questions d'intérêt général les plus débattues.

Alors que les « logiciels mangent le monde » (Andreessen 2011), les êtres humains sont de plus en plus cernés par des systèmes techniques prenant des décisions « qu'ils ne comprennent pas et sur lesquelles ils n'ont aucun contrôle » (Groupe de travail « Article 29 » sur la protection des données, 2013). Si cette évolution peut être déconcertante, elle n'est pas forcément négative, mais plutôt un produit de cette phase particulière de la modernité dans laquelle de grandes quantités d'artefacts techniques gérés par logiciel sont fabriqués grâce à des progrès économiques et technologiques mondialisés. Ces « objets codés » (Kitchin et Dodge 2011) intègrent toutes sortes de capacités décisionnelles utiles aux décideurs de la politique publique : quels choix instantanés doit faire un véhicule piloté par logiciel s'il anticipe un accident ? Les algorithmes des entreprises de l'internet en situation de quasi-monopole ont-ils le pouvoir d'influer sur les élections ? Quels sont les droits des travailleurs ayant avec leur employeur des relations entièrement automatisées ? Y a-t-il plus de risques de préjugés sexistes, ethniques ou raciaux dans un système automatisé et quel degré de préjugés doit-on considérer comme acceptable ?

Bien qu'il ne soit pas facile de répondre à ces interrogations sur la politique publique, les décideurs peuvent et doivent prendre le temps d'essayer de relever ces défis. Dans le passé, nombre de décisions relatives à la manière d'élaborer ces types de logiciels étaient prises par des sociétés privées en fonction de cadres éthiques, juridiques et économiques jugés appropriés par ces mêmes sociétés. Le manque de politiques publiques fondées offrant un cadre réglementaire pour la prise de décision algorithmique (systèmes et processus) est indéniable. Pour autant, on ignore s'il est possible de mettre en place une réglementation efficace alors que de nombreuses technologies fondées sur les algorithmes n'en sont qu'à leurs balbutiements. Les questions soulevées par l'utilisation d'algorithmes dans le processus de prise de décision sont multiples et complexes et suscitent des inquiétudes quant à la qualité des données, au

respect de la vie privée et à la discrimination injustifiée. Le débat en est cependant à un stade si précoce qu'il est parfois difficile de comprendre quelles sont les fonctions réelles des algorithmes et les conséquences qui en découlent pour la société. Ceci ne doit néanmoins pas empêcher la tenue de débats plus fondamentaux sur les moyens de protéger les droits de l'homme dans un monde régi par des décisions algorithmiques en s'appuyant davantage sur les principes. Il faut recenser les différentes sources d'inquiétude et réfléchir à la façon dont un cadre normatif pourrait les apaiser.

Ceci est d'autant plus vrai que de nombreuses mesures de politique publique prises en ce domaine sont promptement imitées par d'autres États dans le monde, provoquant ainsi une avalanche de politiques, bonnes ou mauvaises. En outre, la plupart des débats sur les algorithmes sont moins axés sur les algorithmes eux-mêmes que sur le rôle de la technologie au sein de la société (Bucher 2016). Bien que les débats sur la technologie et la société soient à divers titres importants pour comprendre les algorithmes (Bijker et al. 2012), de nouveaux enjeux politiques se posent également lors des discussions sur les algorithmes associés à la prise de décision automatisée (ou semi-automatisée). Ces enjeux seront au cœur de la présente étude dont nous espérons qu'elle pourra éclairer un tant soit peu les dimensions des droits de l'homme dans l'application des algorithmes.

2. DEFINING ALGORITHMS AND AUTOMATED DATA PROCESSING TECHNIQUES

When looking at algorithms and the automated data processing techniques they engage in, it is important to be absolutely clear what types of algorithms are being discussed here. Rather than reinventing the wheel, this study will build on existing well-established definitions used by other authors, in particular the work of Tarleton Gillespie (2014), Nicholas Diakopoulos (2015) and Frank Pasquale (2015).

This definition used here 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) Thus it can be suggested that algorithms are "a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome." (Diakopoulos 2015:400)

However saying what algorithms *are* is not the same as defining which algorithms matter. For the purpose of this report it seems reasonable to limit the scope of the algorithms being discussed to those which are digital (Diakopoulos 2015) and are of "public relevance" (Gillespie 2014:168). Moreover in order to separate out the specific human rights dimensions of algorithms, this report will focus on algorithmic decision-making, i.e. when algorithms make decisions in an automated or semi-automated fashion.[1] This type of decision-making is often

[1] The distinction between 'semi-automated' and 'solely automated' is important and exists in various EU Directors on data protection such as the GDPR. For the purposes of this study we will – following similar work

subjective in that there is no obvious right or wrong answer but rather the judgement of a human being was previously used to make a subjective determination that is now being made by an automated system (Pasquale 2015:8).

Finally it should be noted that algorithms as discussed here do not exist meaningfully without interaction with human beings. They are deeply entangled with practice (Gillespie 2014:168) and the "promise of algorithmic objectivity" (Gillespie 2014:168), both of which serve to create the social and institutional conditions in which algorithms have effects on real life human beings. It is heavily misleading to claim the computing systems are or even can be neutral, rather technologies are deeply social constructs (Winner 1980, 1986) with considerable political implications (Denardis 2012). Thus when the 'computer says no'[2] the decision-making software in the computer is "biased but ambivalent" (McCarthy 2011:90), it has no meaning without a social system around it that gives this 'no' meaning. It is thus too simple to simply blame the algorithm and simply suggesting to stop using computers or computing is rarely a helpful alternative. Rather specific norms and values are embedded in and enmeshed with algorithms that need to be questioned, critiqued and challenged.

Thus is should be evident that many of the decision making processes around algorithms are relevant for scrutiny by policy makers. However for the purposes of this study this ascertain is not sufficient, it is also important to ask to whether algorithms have an impact on human rights?

The French Parliament certainly seems to think so. "On 26 January 2016, the French National Assembly voted 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). Beyond France there are numerous indications that algorithms do indeed have an impact on human rights. The longest and most sustained human rights debate on algorithms and automated data processing relates to the Right to Privacy (Tene and Polonetsky 2013). It is possible to find articles from more than 45 years ago which discuss infringements of the right to privacy (Sills 1970) associated with automated data processing. Moreover data protection regulation such as the EU's General Data Protection Regulation has also produced some of the key regulatory instruments for algorithms such as the "right to explanation" (Goodman and Flaxman 2016) or the right of access to "knowledge of the logic involved in any automatic processing of data concerning him" (EU Directive 95/46/EC). However one of the main challenges faced in this area is that data protection is often understood in an individual rather than a collective sense (Mantelero 2016), which suggests a false sense of agency for individuals. It can also be seen in this context that the European Data Protection

---

on this topic by the Council of Europe - not differentiate but consider both semi-automated and solely automated decision-making as relevant automated decision-making processes.

[2] See https://www.youtube.com/watch?v=AJQ3TM-p2QI for a full explanation of this phrase.

Supervisor (EDPS) appointed an Ethics Advisory Group to go beyond the boundaries of existing Data Protection law to search for a new Digital Ethics.[3]

Another human right that is evidently affected by the usage of algorithms is Freedom of Expression. The report of 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) suggests that "search engine algorithms dictate what users see and in what priority, and they may be manipulated to restrict or prioritize content" (Kaye 2016:7) and that "platforms deploy algorithmic predictions of user preferences and consequently guide the advertisements individuals might see, how their social media feeds are arranged and the order in which search results appear" (Kaye 2016:16).

Another key fundamental freedom that is frequently cited in relation to human rights is the right to Protection against Discrimination. Various discriminatory patterns arise around the usage of algorithms that are frequently suggested to violate this right (Caliskan-Islam, Bryson, and Narayanan 2016; Tufekci et al. 2015). There are also suggestions that certain forms of algorithmic decision-making lead to "social sorting" (Lyon 2003).

Beyond these three fundamental rights discussed above, there are numerous other areas in which human rights may be affected by algorithms. These include ensuring the rule of law (Pasquale 2015; Joerden 2015), the right to free elections (Bond et al. 2012), workers' rights (Irani 2015) and even the Right to Life (Asaro 2013). A similar elaboration could be made for almost any other human right, but suffice to say at this point that there are evidently human rights aspects to the usage of algorithms and that they are thus worthy of further study by policy makers to understand these aspects.

3. CHARACTERISTICS AND LIMITATIONS OF ALGORITHMS

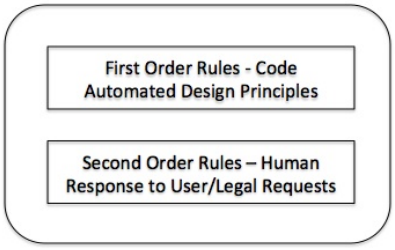
There are many different aspects that can be considered as key characteristics of algorithms that engage in automated data processing and (semi-)automated decision making. As a result this list cannot be exhaustive or predict all possible potential iterations of algorithms and their decision-making. Rather it attempts to provide an overview of the current characteristics and limitations traits of algorithms in 2016.

Automation is one of the core challenges associated with algorithmic decision-making. The ability of automated computing systems to make decisions about a growing number of situations previously decided by human-beings is a key characteristic of the practical implementation of algorithms. It is important to note in this context that human decision-making and algorithmic decision-making is fundamentally different (Spiekermann 2015). While algorithmic decision-making is increasingly adept at mimicking the decision making of human beings, important elements (such as discretion) of decision-making

[3] See https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/Ethics for further details.

processes cannot be automated and often get 'lost on the way' when human decision-making processes are automated (Spiekermann 2015). This is not to say that decision-making by humans or algorithms is necessarily better or worse, but rather that the two are fundamentally and categorically different, have different kinds of consequences and make different kinds of mistakes. While societies and governments around the world have considerable experience understanding human decision-making and its failures, we are only beginning to understand algorithmic decision-making and its flaws.

Another important aspect of algorithmic decision making is their perceived inflexibility in some areas and their incredible flexibility in others. Broadly speaking, algorithms are typically highly adaptive within the scope of their programming and are typically able to integrate considerable amounts of additional data without great difficulty. In order to simplify this statement the graph on the left may help to explain the adaptive nature of algorithms. Without any changes to the actual code the data processing (first order rules) of algorithms is typically more difficult to influence than the outputs of algorithms (second order rules) (Wagner 2016a). This distinction is important, as the first order elements of automated decision-making processes are often adaptive, and can adjust to second order rules very rapidly (Wagner 2016b).



One area where algorithms are typically very inflexible (where the computer says 'no') is in the area of procedural and deeper structural changes to their decision making, or first order rules as they are termed here. This characteristic can often be observed with organisations which implement algorithmic systems without an in-house capacity to change the software code themselves. Even if the outputs of the system are perceived as 'mistakes' it will often be easier and/or cheaper for such organisations to 'fix' the problem with second order rules rather than to change the algorithm itself. Problems about algorithmic flexibility or inflexibility have little to do with the actual algorithms themselves but instead are a product of how human beings implement and regulate algorithms.

In relation to flexibility another aspect of some algorithms design also needs to be considered, namely self-learning or adaptive algorithms. There algorithms have the ability to learn based on data they receive through machine learning techniques (Williamson 2016). This ability to learn new 'tricks' from the data they receive certainly makes algorithms more flexible in some regard, but also makes their output harder to predict (Gillespie 2016). This also has led some authors to suggest that many forms of algorithmic transparency, accountability and regulation are impossible because the programmers themselves are unable to predict or fully understand how the algorithm makes the decisions that it makes (Kroll 2016).

Finally many technologies based on algorithms use data mining and pattern recognition without "understanding" causal relationships (correlation instead of causation), which may lead to errors and raise concerns about data quality. Related to this, one key challenge linked to the implementation of algorithms is the frequent perception that they are able to create neutral and independent

predictions about future events. The hype around Google Flu trends in 2011 which later turned out to be completely unjustified as their prediction ability was far lower than claimed is just one example of the ongoing struggle with claims around predictive algorithms (Lazer et al. 2014; Lazer and Kennedy 2015). The challenge however is less to do with algorithms themselves and far more about how human beings 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. Rather than changing the algorithms in any way, it would be far more helpful to ensure more critical engagement in public debates about them.

4. EMPIRICAL DIMENSIONS: CURRENT AND FUTURE USES OF ALGORITHMS

The following set of cases provides a broad overview of the areas in which automated decision-making is currently employed in a way which raises relevant public policy questions. However this list is not exhaustive and is based primarily on those areas in which considerable public, academic or expert debate already exists and thus some preliminary comparisons and conclusions are possible. Areas that could have also been covered in this context include healthcare, mobility and many other areas that are not even being debated yet but may be in future.

A. CONTENT FILTERING AND WEBSITE BLOCKING

One of the key public policy debates about online content filtering is the extent to which content removal by social media platforms takes places manually, semi-automatically or automatically. While large social media platforms like Google or Facebook love to claim that all content taken down is previously reviewed by a human being (Buni and Chemaly 2016), there are evidently large parts of the process which are automated (Wagner 2016b). Given the importance of platforms like Google or Facebook, their centrality for many users 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.

Automation of the content removal process by social media platforms is particularly evident in the response times that different types of content receive and how content is prioritized, a process that is evidently automated. The same goes for the threshold of user complaints that are required before a piece of content is reviewed. While this varies for different kinds of content it is not 0 and there are strong suggestions that the complete responses of Facebook to user queries is automated for many types of content (Wagner 2016b; Zhang, Stalla-Bourdillon, and Gilbert 2016). Thus it is possible that many users complain about a specific type of content without an automated algorithm judging it relevant to ask a human operator to review the content. As a result, it is possible to claim that large parts of the content takedown process are automated and the even human content takedowns are at minimum semi-automated. Another example are 'upload filters,' which are used to scan for and automatically remove content considered copyright infringement or child sexual abuse images (McIntyre

2012). It has been suggested that similar algorithms could be used for extremist content (Toor 2016).

As noted by Tufekci et al. (2015) "The scale of the content on user-generated platforms and costs associated with human moderation are the reasons algorithmic processes appeal to platforms. Yet, given the crucial gate-keeping function played by these platforms, algorithms also introduce new complications rather than creating a simple solution." Thus many of the practices discussed above involved algorithmic decision-making and thus pose considerable challenges for the rights to Freedom of Expression and Privacy.

B. TERRORISM AND CRIME PREVENTION

There has been a considerable push for the usage of automated decision-making in the areas of national security and crime prevention. Following a string of violent attacks in the US and Europe in recent years, many European and US politicians have begun calling for online social media platforms to use their algorithms to identify terrorists (Rifkind 2014; Toor 2016). Many social media platforms are seemingly already using algorithms to identify extremist content and are being asked by governments to pass on the results generated by these algorithms to governments. One important example of this practice relates to British parliamentary investigation into the murder of the British soldier Lee Rigby in London 2013.

According to a report from the British Intelligence and Security Committee of Parliament the killers had, apparently, posted extremist content in an online social network that was flagged and removed, reportedly algorithmically. The Parliamentary report states that: "The Committee asked GCHQ about the processes by which companies hosting such platforms might close accounts. GCHQ explained that different Communications Service Providers (CSPs) use different systems. However, it appears that there are: various automated techniques for identifying accounts which they believe break their terms of service. They use these techniques to identify and disable accounts which they believe may be linked to child exploitation and to illegal acts such as inciting violence […] Such accounts are then automatically suspended." (Rifkind 2014)

In the US, the Obama administration has advocated for the use of 'hashes'[4] for the detection and automatic removal of extremist videos and images. Additionally, there have been proposals to modify search algorithms in order to "hide" websites that would incite and support extremism. The hash mechanism has been adopted by Facebook and YouTube for video content, however no information has been released over the presence of human input nor on the criteria adopted to establish which videos are "extremist". In Europe, while similar projects are under scrutiny, the Interpol has created a regional organization monitoring online extremist content called the "Internet Referral Unit". The system will be automated in the next few months with the

[4] Hashes are unique identifiers for pieces of internet content that are typically generated by an algorithm and simplify the identification process. For a further explanation of what hashes are and how they are used to regulate internet content see (McIntyre 2012).

introduction of the "Joint Referral Platform."[5] Notably the data on extremist online content that Europol is processing refers not just to Internet content which is illegal in Europe or one of its member states but also to material which violates the Terms of Service of an Internet intermediary. By contrast the Internet Referral Unit of the Netherlands have publicly stated that they have no interest in policing the Terms of Service of Internet Intermediaries and "don't do anything automated" (Lestrade 2016).

In a different vein, it should also be noted that automated recommender systems can also have problematic effects in regards to ideological or political content. Specifically, the programming of many online recommendation systems can create 'filter bubbles' - fully-automated echo chambers in which individuals only see pieces or information which confirm their own existing opinions - of extremist content (Bozdag 2013; Pariser 2011; Zuckerman 2013) which according to the results of recent research are relatively easy to stumble into and relatively hard to get out of (Salamatian 2014). These fully-automated echo chambers pose the danger of creating "ideological bubbles" (O'Callaghan et al. 2014) of online content. However other scholars like Borgesius et al. who argue that there is "there is little empirical evidence that warrants any worries about filter bubbles" (Zuiderveen Borgesius et al. 2016).

Moving from terrorism to crime prevention, the main policy debates around the usage of algorithms seem to be related to the concept of predictive policing. This approach - broadly framed – bases its analysis on historical patterns of crime to predict future patterns of crime beyond the ability of human beings. This has included developed automated systems which predict which individuals are likely to become involved in a crime (Perry 2013). Similarly, to the online content algorithms discussed above, there is considerable concern that these recommender systems for crime are likely to create echo chambers within which existing biases and prejudices are sedimented. Existing biases and prejudices related to for example racial or ethics groups are then not recognized by the police as their own biases, as they have become integrated into an automated computer program. As the computer program is perceived to be independent and neutral, these biases become 'normal' and part of everyday usage of a computer, rather than specific decisions of an individual which can be more readily questioned.

C. SEARCH ALGORITHMS AND SEARCH FUNCTIONALITY MORE GENERALLY

Search algorithms and search functionality more generally form a key aspect of the Internet. The ability to search the Internet would however be impossible without search algorithms that provide responses to user queries. Search algorithms can be horizontal or vertical in nature. Horizontal search algorithms are used for general search. For instance, the search engine providers Google and Bing employ horizontal search algorithms for general search; to help web

[5] Submission from Article 19.

users locate particular information from millions of web pages.[6] Vertical search algorithms (otherwise known as specialised search algorithms) are used to search *"a specific segment of online content"* (Verhaert, 2014, p. 266). Moreover most modern search tend to provide personalized results which adapt the search results to the algorithmically predicted preferences of their users (and this creates the potential filter bubble).

Search algorithms and search engines are likely to have a positive impact on the fundamental right to freedom of expression. As observed by the Committee of Ministers of the Council of Europe: "*Search 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.*"[7]

However search algorithms and search engines may also have a negative impact on freedom of expression. Content which is not indexed or not ranked highly by one of the few popular internet search engines may be less likely to reach a large audience. A search algorithm might also be biased towards certain types of content or content providers, thereby risking to affect related values such as media pluralism and diversity.[8] This can lead to considerable discrimination issues, both in regards to end-users, customers and societies as a whole.[9] A biased algorithm within a large quasi-monopolistic search engine that systematically discriminates one group in society based on their age, sexuality, race or gender would cause considerable problems not just for the individuals affected by these decisions, but for societies as a whole.

There are also concerns with search engines impact on the right to privacy and data protection. The specific dimensions of this impact relate to facilitating aggregation through gathering information about a specific individual, reducing practical obscurity by making it easier to find information about an individual, violating contextual integrity by violating norms about the distribution of information and reducing individual control over information disclosure as a whole.[10]

---

[6] Submission from Sophie Stalla-Bourdillon, Steffen Staab and Laura Carmichael.

[7] 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.

[8] Submission from Aleksandra Kuczerawy, Brendan van Alsenoy and Jef Ausloos.

[9] Submission from Sophie Stalla-Bourdillon, Steffen Staab and Laura Carmichael.

[10] Submission from Aleksandra Kuczerawy, Brendan van Alsenoy and Jef Ausloos.

D. SURVEILLANCE, ONLINE TRACKING, PROFILING AND "SOCIAL SORTING"

Algorithms play a role in online tracking and profiling of individuals whose browsing patterns are recorded on the basis of cookies and similar technologies such as digital fingerprinting and aggregated with search queries (search engines) and other data (eg social media tracking and data collection through apps on mobile devices) (Tene and Polonetsky 2012). One of the main applications of online tracking and profiling is targeted advertising based on the profile of presumed interests of the person tracked. However these profiles can also be used in the context of assessing a person's risk profile for the purpose of insurance or credit scoring (discussed further below) or more generally for differential pricing (offering different prices for the same goods or services to different consumers, based on their profile). Particular concerns arise from the use of data brokers who aggregate the information contained in personal profiles and this information may then be mined through the use of algorithms, which creates a risk of large-scale surveillance (dataveillance) by private entities and governments alike (Rubinstein, Lee, and Schwartz 2008). The major problem of using data from profiles for different purposes through algorithms is that the data loses its original context and this use is therefore likely to affect a person's informational self-determination and is likely to be prejudicial and/or discriminatory, as the data loses its contextual integrity (Nissenbaum 2004; Tene and Polonetsky 2012). Furthermore the use of algorithms on aggregated profile data may also increase undesirable social inequalities (for example power, status, wealth) (Tene and Polonetsky 2012). This has already been described as 'social sorting' (Lyon 2003).[11]

a. From a human rights point of view some of these concerns can be addressed through developing the right to privacy, but other concerns are not sufficiently captured by the right to privacy. Thus, some scholars have argued that from a normative point of view more conceptualisation of these concerns is required to develop the framework of normative principles further, for example as part of a fundamental right not to be unfairly discriminated against. Others have suggested that this right is already sufficiently covered by Article 14 of the European Convention on Human Rights on anti-discrimination which just needs to be applied more broadly.

E. INSURANCE AND CREDIT SCORING

The objective of insurance is to provide a degree of "financial protection" (David, 2015, p. 147) to the applicant(s). Insurance risk assessment is a formal statistical method utilised to assign applicants to appropriate insurance tariffs by considering the likelihood, frequency and cost of a potential claim (David, 2015, p. 148). Prior to the use of formal statistical methods, "subjective human assessment" was used for creditworthiness assessment (Hand & Henley, 1997, p. 530) and insurance; i.e. the examination of an individual's application on a case-by-case basis. However, the current credit scoring model faces a number of

[11] Submission from Julia Hornle.

criticisms – principally they lack transparency and assess a limited range of variables.[12]

In some instances, algorithmic credit scoring aims to focus beyond traditional variables and take advantage of additional types and amounts of data, such as social media data (Williams 2016) and browsing history to further enrich creditworthiness assessment (Holloman, 2014), (Shipley & Zhuo, 2016), (Clements, 2015). Credit scoring has a potentially broader remit than was initially intended; e.g. it is not only used by lenders but employers (The Editorial Board, 2013).[13] Many businesses traditionally working in the field of credit scoring have expanded their reach to become data brokers and identity management companies (for example providing age-verification information).

F. AUTOMATING THE WORKPLACE AND WORKERS RIGHTS

The workplace is another key area whether automated decision making has become increasingly common in recent hears. Algorithms are increasingly involved in decisions on both hiring and firing staff, staff organization and management as well as evaluating the professional contribution of individual staff members (Tufekci et al. 2015). These decision-making processes are by no means perfect when they are conducted by humans and there are numerous biases in hiring related to race (Bertrand and Mullainathan 2004) class and gender (Altonji and Blank 1999; Goldin and Rouse 1997) that have been demonstrated. With more and more companies moving towards algorithmic hiring (Rosenblat, Kneese, and others 2014) it is however highly problematic that the systems employed typically lack any transparency in the decisions they make, both in the hiring process and beyond. Moreover many of these automated decision-making processes are based on data from Internet intermediaries.

Another challenge is related to automated feedback loops which decide how employees should be managed and are typically linked to customer input (Kocher and Hensel 2016). By allowing the 'wisdom of the crowd' to make decisions about individuals employment is not only highly ethically questionable, it also limits the ability of workers to contest such decisions as they seem to be 'objective' measures of their performance. As individual employment platforms are "Transforming People into Human Computation" there are obvious questions to be asked about workers' rights, employee self-determination and how societies as a whole believe that human beings should be treated at the workplace.

G. CLOUD PROVIDERS AND DATA STORAGE

Another key aspect related to the usage of algorithms for automated data processing focusses on 'cloud' data storage. This refers to solutions whereby files

[12] Submission from Sophie Stalla-Bourdillon, Steffen Staab and Laura Carmichael.

[13] Submission from Sophie Stalla-Bourdillon, Steffen Staab and Laura Carmichael.

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 is also subject to being processed by algorithms while stored remotely in ways that would not be the case if the information is stored remotely. There are two places where such automated data processing can take place: (1) in transit to the remote network storage location and (2) on the remote servers where the data is stored. Importantly as modern operating systems are increasingly deeply enmeshes with 'cloud' i.e. remote services, it is increasingly difficult for users to ascertain to what extent they are using local or remote services. For example, the Siri service for voice interaction with users phones regularly interacts with and stores data on remote servers (Yamamoto et al. 2014) without this being evident let alone transparent to users (Article 29 Data Protection Working Party 2013).

The key question in regards to data in transit is whether it is sufficiently protected or not through technologies such as strong end-to-end encryption [Schulz and Hoboken forthcoming]. If data is encrypted automated data processing becomes a lot more difficult and in some cases completely impossible. If the data which is transferred to the cloud is not encrypted - as is surprisingly common even though this should not be the case – then whichever networks the data passes can analyze and even modify the data. Actors doing so range from intelligence services such as the U.S. NSA or the British GCHQ to more mundane peddlers of advertising such as Phorm and NebuAd (Ermert 2013; Greenwald and MacAskill 2013; Metz 2008; Williams 2008).

However this does not mean that cloud data is safe simply because it is encrypted in transit. It is also possible for data to be manipulated and analyzed on the servers where it is being stored. For example, Microsoft's cloud service 'SkyDrive' operates an "automated process designed to pull the trigger when it 'sees' certain content (such as nudity)," leading to a users account being terminated when such content is found (Clay 2012). This is particularly problematic for many users who "believe their images to be private" (Heckert 2011) are now confronted with an unwanted automated decisions being made based on their personal data.

H. ELECTIONS & IMPLICATIONS FOR DEMOCRACY

[Include cross-link and reference to CoE MSI-MED Report by Damian Tambini].

One important area that is often ignored when looking at the effects of automated data processing and algorithms relates to elections. Since the advent of the Internet it has been argued that online campaigning and social media were likely to change the way in which politics and elections work but it is only more recently that academic research has revealed the extent to which changes to online content platforms can 'tip' elections. More specifically researchers manipulated the Facebook platform without users knowledge during U.S. elections and were thus able to convince a statistically significant segment of the population to vote (Bond et al. 2012). There are strong indications that since then Facebook has been selling similar services to political parties around the world, with similar behaviour observed during the UK local elections in 2016 (Griffin 2016). Whether Facebook and similar quasi-monopolistic online platforms are using their power to influence human voting benevolently or not is

less the point than the fact that they – in principle – have the ability to massively influence elections.

At the same time Facebook is increasingly considered by scholars such as Helberger et al. to be acting as a "news editor [that] has editorial responsibility for its trending topics" (Helberger and Trilling 2016). This in turn begs the question of whether social media platforms should be considered Internet intermediaries or rather the editors of news websites.

I. INTERNET OF THINGS AND SMART CITIES

As discussed in the introduction the spread of programmed objects into all areas of society and human life is being increasingly common. This shift which is sometimes referred to as the 'Internet of Things' or discussed in the context of 'smart cities.' With increasing amount of automation and larger amounts of data that is typically stored by Internet intermediaries it is in theory possible to better tailor automated systems surrounding human beings to their needs. However it is an open question whether such automated systems are used in the interests of users or citizens, particularly when they are implemented in highly sensitive areas such as the medical sector (Bates et al. 2014).

As noted by Natali Helberger "there are possible challenges from the Internet of Things for the 'profiled consumer.' These challenges go beyond issues of privacy and data protection – which will continue to play a prominent role. In addition, the protection of contractual fairness, adequate information and autonomous and free choices comes to the fore" (Helberger 2016:22) Finally it is very common for the data collected by such services to be shared between different data brokers (Hoofnagle 2003), ensuring that the 'profiled consumer' becomes a 'profiled citizen.'

J. DIGITISATION OF PUBLIC SECTOR & GOVERNMENT SERVICES

Numerous government agencies and services are increasingly automating their decision-making with the use of algorithms. While it is heavily debated whether such systems can increase efficiency or not what is evident is that many of the systems pose considerable questions for transparency and accountability of public decision-making. This is particularly the case as these are government authorities which are typically held to a higher standard in their decision making than private sector or other non-governmental organisations.

Despite these standards there are strong suggestions that the public sector is employing automated-decision making in areas as diverse as social security, taxation, health care and the justice system. For example many courts in the United States use a computer program to assess the risk of repeat offending, which has been shown to be "biased against blacks" (Kirchner 2016). Another example relates to the practice of *Profiling the Unemployed in Poland* (Jędrzej Niklas, Karolina Sztandar-Sztanderska, and Katarzyna Szymielewicz 2015). During their analysis they identified several challenges which are broadly also true for the usage of algorithms in other areas of the public sector service delivery as well:

> *"1.Non-transparent rules of distributing public services […]*

2. Shortcomings of computer systems as a trigger for arbitrary decisions …]

3. Gap between declared goals and practice […]

4. System based on the 'presumption of guilt' […]

5. Categorization as a source of social stigma […]

6. Risk of discrimination" (Jędrzej Niklas et al. 2015:33–37)

Finally there are risked associated with outsourcing key government functions such as the provision of government benefits to the private sector. It has been argued in South Africa that operating such privatized government services while simultaneously engaging in competitive banking and insurance markets provides an inappropriate competitive advantage to companies who operate privatized government services (Elza Van Lingen 2016). Aside from the competition concerns there are evident concerns related to privacy and data protection that also arise from such arrangements, particularly as many of the organisations providing these kinds of services are Internet intermediaries.

5. ETHICAL & LEGAL & HUMAN RIGHTS DIMENSIONS

Many of the challenges discussed above touch upon, ethical, legal and human rights challenges. None of these challenges are easily or readily resolved, not should regulatory responses to algorithms and automation be taken lightly. Importantly many of the challenges related to automated data processing have historically been resolved by data protection legislation. Thus relevant innovative governance mechanisms such as the "right to explanation" (Goodman and Flaxman 2016) are also typically the product of data protection legislation. However it should be noted that there is a significant difference between the right to privacy and the data protection regulation, which is at the end still a governance mechanism to safeguard privacy and other rights. Moreover it should be evident from the previous analysis that the "challenges go beyond issues of privacy and data protection" (Helberger 2016:22) and thus cannot be resolved by recourse to data protection regulation alone.

Challenges around discrimination of content raises questions of competition law and discrimination of minorities, while the ability of algorithms to manipulate elections is a matter for electoral commissions and parliaments. There are also issues related to "protection of contractual fairness, adequate information and autonomous and free choices" (Helberger 2016:22). While the issues touched upon above are too broad to be covered by data protection regulation alone, nor should these regulatory mechanisms or the expertise of the data protection community be forgotten in the process of finding regulatory responses to algorithmic governance.

Importantly there is a danger that if misconstrued some of the statements or recommendations in this report could be used to regulate the development of algorithms or other software code. So to be absolutely clear it should be stated here that this should is not a reasonable outcome of this report and any such interpretation would be false. This is because interference with the right of individuals to research, develop and test would itself be a grave infringement of

human rights and in particular freedom of opinion, expression, thought and research. Aside from significant human rights impacts of doing so, limiting research and development of algorithms actually limits a better understanding of how algorithms operate and what effects they have. Importantly many of the public policy solutions can only that will be discussed below are only relevant for very large actors with a considerable quasi-monopolistic market share (Naughton 2016) and could well be harmful if implemented on implementers of algorithms of all shapes and sizes.

Finally, there are very fundamental legal and ethical surrounding the legal personhood of automated systems such as algorithms that cannot easily be resolved in this report. Particularly around questions such as liability and accountability these questions are important, not in a manner to exculpate those involved in developed and implementing such systems from responsibility but rather to acknowledge that accountability and liability is becoming increasingly complex with autonomous systems.

6. MECHANISMS OF GOVERNANCE, ACCOUNTABILITY & TRANSPARENCY

There is a frequently stated perception that the regulation of algorithms in automated systems is either impossible or extremely difficult. Such statements tend to ignore the numerous cases in which algorithms are already regulated before their implementation by government regulators or independent auditors. To provide just one example, the software and algorithms used in 'slot machines' in Australia and New Zealand are required by government regulation to be "fair, secure and auditable" (Woolley et al. 2013). As part of this process the developers of such machines are required to submit their algorithms to regulators before they can be presented to consumers. The same applies to the regulation of online gambling in the United Kingdom, where gambling equipment is controlled by a specific licensing regime.

The Australian/New Zealand Gaming Machine National Standard in its most recent revision 10.3 defines in extraordinary detail the technical specifications by which such machines can 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".[14] This is not to say that such mechanisms would definitely be appropriate for regulating Internet intermediaries, but rather that the claim that the regulation of algorithms is impossible is evidently false and that regulators should look to existing mechanisms already implemented.

[14] The Australian/New Zealand Gaming Machine National Standard which is available here: https://publications.qld.gov.au/dataset/a-nz-gaming-machine-national-standards

A. TRANSPARENCY

One of the main challenges frequently cited in regulating algorithms is that they seem like black boxes to both consumers and regulators (Pasquale 2015). As 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 question of accountability” (Tufekci et al. 2015:11). Thus there is a frequent and growing debate about algorithmic transparency in which algorithms could be provided to independent auditors, regulators or the general public (Diakopoulos 2015; Rosnay 2016).

As provision of entire algorithms to the public is typically considered unlikely, there is also a debate around the possibility of providing key subsets of information about the algorithms to the public, for example which variables are in use, the average values and standard deviations of the results produced or the amount and type of data being processed by the algorithm.

All of these measures aim to increase transparency of automated systems. This is obviously complicated by the frequent changes in the algorithms used, as Google for example changes its algorithm hundreds of times per year (Tufekci et al. 2015). There is also the frequently danger of manipulation and 'gaming' or algorithms if they were made public. At the same time the usage of machine learning complicates transparency to a point where provision of all of the source code of an algorithm may not even be sufficient, and instead there is a need for an actual explanation of how the results of an algorithm were produced. Initial steps to such a right can be drawn from the European General Data Protection Regulation (GDPR) including a right to explanation (Goodman and Flaxman 2016).

Private companies also regard their algorithm as their key trade secret and hence disclosure is unrealistic. Besides, in a decision of 28 January 2014, the German Federal Supreme Court (Bundesgerichtshof) rejected a claim for information concerning the credit agency's algorithm as it was a protected business secret but allowed a claim for information concerning the data used to calculate creditworthiness through the means of the algorithm. However where algorithms are used in decision-making which potentially prejudices the rights of individuals an oversight mechanism may ensure that the algorithm operates in a fair and sustainable manner. An example for this is section 28 b of the German Federal Law on Data Protection which provides that there has to be a scientifically proven mathematical-statistical process for the calculation of the probability of a specific behaviour of an individual before such an algorithm can be used for making a decision about a contract.

B. ACCOUNTABILITY

What accountability do individuals have for the algorithms they implement? This depends very much on the nature of the algorithms and their outputs. In many cases if the outputs are defamatory, infringe copyright or raise other legal concerns there are already governance mechanisms to ensure that these kinds of outputs are limited (Staab, Stalla-Bourdillon, and Carmichael 2016), with the case of Max Mosley taking action against Google just one of many examples (Stanley 2011). However such mechanisms typically only affect second order rules, i.e. changes to the outputs of algorithms. By contrasts there is a general lack of regulatory frameworks to influence first order rules and ensure that algorithms – in most cases – are actually producing results that uphold and protect fundamental values or basic ethical and societal principles.

However it has been suggested that "[t]echnologists think about trust and assurance for computer systems a bit different 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).

7. REGULATING ALGORITHMS DIRECTLY

As was discussed above, gambling is one area where the code of algorithms is regulated directly and required to conform to certain standards, but it is not the only area where this kind of regulation is being discussed. In the financial sector there is an ongoing debate about the regulation of high-speed trading algorithms as these are seen to have a strong potential destabilizing effect on the overall financial system. One of the leading social democrat politicians suggested in 2012 that financial trading "algorithms will have to undergo a stress test to ascertain its stability" (Steinbrück 2012).

One associated area where similar regulation has been threatened is in the area of online content regulation and Internet hotlines. Here the British Police special unit CEOP demanded 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 could be expected if states begin to define the content of algorithms on large online platforms.

Aside from direct regulatory mechanisms to influence the code of algorithms, indirect mechanisms to influence algorithms code could also be considered. These address the production process or the producers of algorithms and attempt to ensure they are aware of the legal challenges, ethical dilemmas and human rights concerns raised by automated decision-making. Another instrument to achieve such goals could be consist of standardized 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.[15]

In conclusion it should be noted that the approach to direct regulation of algorithms or software code should be pursued with extreme care. It is the regulatory approach that provides the most pitfalls and is most likely to wider problems. Notably the direct regulatory approach raises considerable concerns about freedom of opinion and expression, the right to privacy as well as Freedom of Profession. Moreover given the fact that regulators currently do not know that much about algorithms, greater steps towards transparency and accountability of algorithms would seem far more appropriate.

8. CONCLUSIONS ET RECOMMANDATIONS

Comprendre comment fonctionnent les systèmes décisionnels automatisés est très complexe et soulève de nombreuses questions en matière de politique publique. Même si aucune d'entre elles n'a de réponse simple, les décideurs ne doivent pas pour autant être dissuadés de s'y intéresser de plus près. Si ces problèmes sont aussi difficiles à apprécier, c'est parce qu'ils relèvent d'un domaine relativement nouveau et que trouver des solutions efficaces reste malaisé. Dans un premier temps, il paraît sensé de suggérer aux décideurs

[15] Submission from Markus Oermann, University of Hamburg.

politiques de chercher à en savoir plus sur la mise en œuvre des systèmes décisionnels automatisés dans leur pays respectif. Dans un deuxième temps, ils devraient tenter de s'assurer que la législation et les cadres juridiques en vigueur continuent d'être appliqués et opérants pour faire face aux difficultés suscitées par la prise de décisions automatisée dans de nombreux domaines. Enfin, il faudrait que les décideurs hésitent à limiter l'action des chercheurs ou d'autres personnes cherchant à comprendre comment fonctionnent les algorithmes, ou à concentrer leurs activités de réglementation sur des organisations possédant de faibles parts de marché. En effet, de nombreux problèmes concernant par exemple la liberté et l'équité des élections ou la modération des contenus en ligne sont uniquement dus à des marchés quasi monopolistiques.

Cependant, le MSI-NET a tenté d'élaborer des recommandations de base qui, nous l'espérons, contribueront à faire avancer le débat afin de concevoir des politiques publiques efficaces pour résoudre le problème de la prise de décision automatisée.

a. Les gouvernements devraient s'engager avec leurs propres organismes sectoriels de réglementation (compagnies d'assurances, organismes de crédit, banques, secteur du commerce électronique) à élaborer des normes et orientations sectorielles spécifiques pour s'assurer d'être en mesure de faire face aux problèmes posés par l'emploi des algorithmes dans la prise de décision automatisée et prendre en compte l'intérêt du consommateur.

b. Les gouvernements devraient envisager de proposer des voies de recours (systèmes de plaintes) aux consommateurs ayant été injustement lésés par une prise de décision automatisée.

c. Les gouvernements devraient veiller à ce que les consommateurs aient accès aux éléments clés des algorithmes pour leur permettre de prendre des décisions éclairées sur les services à utiliser.

d. Les gouvernements devraient s'assurer de la totale transparence de tous les systèmes décisionnels automatisés utilisés par une autorité publique ou tout autre organisme gouvernemental, et fournir à l'ensemble des parties concernées toutes les informations requises pour un examen et un contrôle complets de ces systèmes.

e. Les gouvernements ne devraient prendre aucune mesure limitant la possibilité de mener des recherches sur des systèmes décisionnels automatisés, de les concevoir ou de les comprendre.

f. Les gouvernements devraient veiller à ce que les particuliers et les organisations soient tenus responsables d'une utilisation incorrecte des systèmes décisionnels automatisés.

g. Les gouvernements devraient faire en sorte que des élections libres et régulières restent possibles.

h. Les gouvernements devraient s'abstenir d'imposer aux intermédiaires de l'internet l'obligation générale d'employer des techniques automatisées pour contrôler les informations qu'ils transmettent ou stockent.

i. Les intermédiaires de l'internet devraient faire preuve d'une transparence parfaite vis-à-vis des usagers en ce qui concerne le retrait et le blocage de contenus et leur faire savoir si et dans quelle mesure la prise de décision répondant aux demandes de retrait de contenus internet est automatisée.

j. Les utilisateurs devraient avoir la possibilité de contester le blocage et le filtrage de contenus.

9. BIBLIOGRAPHY

Altonji, JG and RM Blank. 1999. 'Race and Gender in the Labor Market'. *Handbook of Labor Economics*. 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).

Article 29 Data Protection Working Party. 2013. *Opinion 03/2013 on Purpose Limitation*. Brussels, Belgium: ARTICLE 29 DATA PROTECTION WORKING PARTY. Retrieved (ARTICLE 29 DATA PROTECTION WORKING PARTY).

Asaro, Peter. 2013. 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making'. *International Review of the Red Cross* 94(886):687–709.

Bates, David W., Suchi Saria, Lucila Ohno-Machado, Anand Shah, and Gabriel Escobar. 2014. 'Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients'. *Health Affairs* 33(7):1123–31.

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.

Bijker, Wiebe E., Thomas P. Hughes, Trevor Pinch, and Deborah G. Douglas. 2012. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. MIT Press.

Bond, Robert M. et al. 2012. 'A 61-Million-Person Experiment in Social Influence and Political Mobilization'. *Nature* 489(7415):295–98.

Bozdag, Engin. 2013. 'Bias in Algorithmic Filtering and Personalization'. *Ethics and Information Technology* 15(3):209–27.

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).

Caliskan-Islam, Aylin, Joanna Bryson, and Arvind Narayanan. 2016. 'A Story of Discrimination and Unfairness Implicit Bias Embedded in Language Models Aylin'. in *Security & Privacy Week 2016*. Darmstadt; Germany: TU Darmstadt.

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.

Elza Van Lingen. 2016. 'DA Refers Net1's Abuse of Advantage to Competition Commission'. *Democratic Alliance*. Retrieved 31 August 2016 (https://www.da.org.za/2016/02/da-refers-net1s-abuse-of-advantage-to-competition-commission/).

Ermert, Monika. 2013. *NSA-Abhörskandal PRISM: Internet-Austauschknoten Als Abhörziele*. Retrieved 19 August 2014 (http://www.heise.de/newsticker/meldung/NSA-Abhoerskandal-PRISM-Internet-Austauschknoten-als-Abhoerziele-1909604.html).

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.

Gillespie, Tarleton. 2016. 'Algorithm'. in *Digital Keywords: A Vocabulary of Information Society and Culture*, edited by B. Peters. Princeton University Press.

Goldin, Claudia and Cecilia Rouse. 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"'. Retrieved (http://arxiv.org/abs/1606.08813).

Greenwald, Glenn and Ewen MacAskill. 2013. 'NSA Prism Program Taps in to User Data of Apple, Google and Others'. Retrieved (http://www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data).

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).

Heckert, Marc. 2011. 'Wie Ein Handy-Fan von Wolke Sieben Fiel'. *Aachener Zeitung*. Retrieved 31 August 2016 (http://www.aachener-zeitung.de/news/digital/wie-ein-handy-fan-von-wolke-sieben-fiel-1.372632).

Helberger, Natali. 2016. *Profiling and Targeting Consumers in the Internet of Things – A New Challenge for Consumer Law*. Rochester, NY: Social Science Research Network. Retrieved 31 August 2016 (http://papers.ssrn.com/abstract=2728717).

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/).

Hoofnagle, Chris Jay. 2003. 'Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement'. *NCJ Int'l L. & Com. Reg.* 29:595.

Irani, L. 2015. 'Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk'. *South Atlantic Quarterly* 114(1):225–34.

Jędrzej Niklas, Karolina Sztandar-Sztanderska, and Katarzyna Szymielewicz. 2015. *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*. Warsaw, Poland: Panoptykon Foundation. Retrieved (https://en.panoptykon.org/articles/profiling-unemployed-poland-%E2%80%93-report).

Joerden, Jan C. 2015. 'Zum Einsatz von Algorithmen in Notstandslagen'. in *3. Würzburger Tagung zum Technikrecht*. Würzburg: Universität Würzburg.

Kaye, David. 2016. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression to the Thirty-Second Session of the Human Rights Council*. Geneva, Switzerland.

Kirchner, Julia Angwin Surya Mattu, Jeff Larson, Lauren. 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).

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

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.

Lestrade, Niels. 2016. 'Use of Surveillance to Counter VERLT – Challenges and Opportunities to Safeguard Online Freedom'.

Lyon, David. 2003. 'Surveillance as Social Sorting: Computer Codes and Mobile Bodies'. in *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*, edited by D. Lyon. New York: Routledge.

Mantelero, Alessandro. 2016. 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection'. *Computer Law and Security Review* 32(2):238–55.

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.

McIntyre, TJ. 2012. 'Child Abuse Images and Cleanfeeds: Assessing Internet Blocking Systems'. in *Research Handbook on Governance of the Internet*, edited by I. Brown. Cheltenham: Edward Elgar.

Metz, Cade. 2008. *Phorm Secretly Tracked Americans Too*. Retrieved (http://www.theregister.co.uk/2008/08/13/phorm_us_tests/print.html).

Naughton, John. 2016. 'Digital Dominance: Forget the "digital" Bit'. *Media Policy Project*. Retrieved 9 September 2016 (http://blogs.lse.ac.uk/mediapolicyproject/2016/07/12/digital-dominance-forget-the-digital-bit/).

Nissenbaum, Helen. 2004. 'Privacy as Contextual Integrity'. *Wash. L. Rev.* 79:119.

O'Callaghan, D., D. Greene, M. Conway, J. Carthy, and P. Cunningham. 2014. 'Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems'. *Social Science Computer Review* 33(4):459–78.

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.

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).

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).

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).

Stanley, JE. 2011. 'Max Mosley and the English Right to Privacy'. *Wash. U. Global Stud. L. Rev.* 10(3). Retrieved

(http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/wasglo10&section=25).

Steinbrück, Peer. 2012. *Vertrauen Zurückgewinnen: Ein Neuer Anlauf Zur Bändigung Der Finanzmärkte*. Berlin, Germany: Deutscher Bundestag - German Federal Parliament.

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).

Tene, O. and J. Polonetsky. 2013. 'Judged by the Tin Man: Individual Rights in the Age of Big Data'. *J. on Telecomm. & High Tech. L.*

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/).

Wagner, Ben. 2016a. 'Algorithmic Regulation and the Global Default: Shifting Norms in Internet Technology'. *Etikk I Praksis-Nordic Journal of Applied Ethics*.

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

Williams, Aime. 2016. 'How Facebook Can Affect Your Credit Score'. *Financial Times*, August 25. Retrieved 9 September 2016 (http://www.ft.com/cms/s/2/e8ccd7b8-6459-11e6-a08a-c7ac04ef00aa.html#axzz4JmmAbUyE).

Williams, Chris. 2008. *BT's Secret Phorm Trials: UK.gov Responds*. Retrieved (http://www.theregister.co.uk/2008/09/16/phorm_eu_berr/print.html).

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.

Yamamoto, Daisuke et al. 2014. 'Voice Interaction System with 3D-CG Virtual Agent for Stand-Alone Smartphones'. Pp. 323–30 in *Proceedings of the second international conference on Human-agent interaction*. ACM. Retrieved 31 August 2016 (http://dl.acm.org/citation.cfm?id=2658874).

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).

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).