CENTRE FOR INTERNET AND HUMAN RIGHTS

https://cihr.eu

The Ethics of Algorithms:
from radical content to self-driving cars

Final Draft Background Paper[1]

1) INTRODUCTION

A new kind of object, intermediator, gate-keeper and more has risen: the algorithm, or the code that operates increasingly ubiquitous computational objects and governs digital environments.

Computer chips and other forms of computation are not new; however, the increasing integration of digital connectivity in everyday life; the rise of massive amounts of datasets with personal, financial and other kinds of information, and the rise in objects that have embedded chips have combined to create a new environment. This environment has been shaped by three developments: Advances especially in machine learning which allow artificial intelligence, with the help of big data, to perform tasks that were outside its reach just a few years ago; the rise of powerful platforms online such a Google, Amazon or Facebook that mediate social, political, personal and commercial interactions for billions of people and act as powerful gatekeepers; and the incorporation of algorithmic capabilities to other areas of decision-making ranging from hiring, firing and employment to healthcare, advertising, to finance and many others.

In sum, algorithms are increasingly used to make decisions for us, about us, or with us. They are progressively capable and pervasive. They are now either main or auxiliary tools, or even sole decision-makers, in areas of life that either did not exist more than a decade ago (what updates and news should you be shown from your social network, as in Facebook's Newsfeed) to traditional areas where decisions used to be made primarily via human judgment, such as health-care and employment. Significantly, algorithms are rapidly encroaching into "subjective" decision-making where there is no right or wrong answer, or even a good definition of what a "right" answer would look like without much transparency, accountability or even a mapping out of the issues. The speed of technological developments, corporate and government incentives have overtaken and overshadowed the urgently needed discussion of ethics and accountability of this new decision-making infrastructure.

The concerns that often bring us to thinking about algorithms are both historic and mundane: fairness, discrimination and power. Algorithms, and all complex computational systems, however, operate in ways that are a new category of objects compared with other institutions, persons or objects that have not been probed for such concerns.

In this report, we provide some of the key areas that require further probing, research and discussion, and should be taken up by policy-makers, civic actors, citizens and everyone concerned about the ethical, legal and policy frameworks in the 21st century which can no

---

[1] This paper was prepared by the Centre for Internet & Human Rights at European University Viadrina. For further questions about the paper please contact bwagner@europa-uni.de or office@cihr.eu.

longer be discussed without incorporating questions of computation. We will begin by defining algorithms, in particular those that demand ethical scrutiny. We will proceed by illustrating three characteristics of algorithms with cases from a wide variety of fields. In the final section, we will address three regulatory responses that have discussed in response to the challenges posed by algorithmic decision-making.

This background paper is the result of a two-day conference on "The Ethics of Algorithms", held in Berlin on March 9 and 10, 2015. The event was jointly organised by the Centre for Internet and Human Rights and the Technical University Berlin, with the support of the Dutch Ministry of Foreign Affairs. The results presented in this paper will feed into the discussions at the Global Conference on Cyberspace, which will take place in the Hague on 16 and 17 April 2015.

2) ALGORITHMS: SUBJECT & SCOPE OF THE REPORT

Algorithms can be used to refer to any computer code - from the simplest to the most complex - that carries out some set of instructions. As such, algorithms are essential to the way computers process data. While we often think about code, when we think about algorithms, in fact, algorithms do not necessarily require any code: the instructions for carrying out long-divisions, for example, are an algorithm. As Tarleton Gillespie (2014, p. 167) put it: "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." In other words, algorithms are "a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome." (Diakopoulos, 2014)

Algorithms can have a number of functions. A Tow Center report on the role of algorithms in the news ecology identifies four practical uses: prioritization, classification, association and filtering (Diakopoulos, 2014). When algorithms prioritize, they bring "attention to certain things, at the expense of others." Classification is defined as decisions that "involve categorizing a particular entity as a constituent of a given class by looking at any number of that entity's features." Association marks relationships, while filtering is the act of excluding information (or other inputs or outputs depending on the system).

Gillespie (2014) also has a categorization scheme of algorithmic functions. Though these are a different level of analyses, they are also useful since every functional purpose served by an algorithm invites questions of ethics, accountability and responsibility. According to Gillespie, algorithms are characterised by the following functions:

> 1: Patterns of inclusion: the choices behind what makes it into an index in the first place, what is excluded, and how data is made algorithm ready
>
> 2. Cycles of anticipation: the implications of algorithm providers' attempts to thoroughly know and predict their users, and how the conclusions they draw can matter
>
> 3. The evaluation of relevance: the criteria by which algorithms determine what is relevant, how those criteria are obscured from us, and how they enact political choices about appropriate and legitimate knowledge
>
> 4. The promise of algorithmic objectivity: the way the technical character of the algorithm is positioned as an assurance of impartiality, and how that claim is maintained in the face of controversy

5. Entanglement with practice: how users reshape their practices to suit the algorithms they depend on, and how they can turn algorithms into terrains for political contest, sometimes even to interrogate the politics of the algorithm itself

6. The production of calculated publics: how the algorithmic presentation of publics back to themselves shape a public's sense of itself, and who is best positioned to benefit from that knowledge. (Gillespie, 2014).

3) ALGORITHMS OF CONCERN

The above definition of algorithms and their function shows, that not all algorithms raise ethical concerns. Algorithms of concern in this report are subset of all algorithms in operation, primarily those that are digital (Diakopoulos, 2014) and are of "public relevance" (Tarleton, 2014). In the following, we would like to argue that it is three attributes that make algorithms demand ethical scrutiny: complexity and opacity, gatekeeping functions, and subjective decision-making. While the reasons we think about the ethical dimension of algorithms is mundane and historic, such as questions of fairness, discrimination and power, it is these three attributes that make algorithms a new category of objects that have not been probed for such concerns.

Complexity and Opacity:

Sufficiently complex algorithms are often practically inscrutable to the outside observers, and can even be hard for humans to comprehend, even if their source code was shared with competent observers. This complexity adds to their power, but also points to a fundamental truth: where there is this much choice, there are values, biases and potential discrimination built in, which are not always readily visible or available.

A computer algorithm functions the same way as any other algorithm, by taking inputs, breaking a task into its constituent parts, undertaking all those tasks one by one according to a prescribed instruction set, and then producing an output. Yet, once enough inputs and steps are added, the resulting branching tree of program flow means that algorithms can play out in a very large number of possible ways. Algorithms are not "alive" in that they are computational machines, but they act with complex "agency"—a term that is sometimes used in social theory is "actant" which acknowledges their dynamic, interactive, complex and purposive modality. Algorithms perform complex calculations, that follow many potential steps along the way and can consist of thousands or even millions of individual data points. The range of potential inputs, plus the range of internal steps and branching, and dynamic interaction with the environment along the way, means that algorithms act in a way similar to living things: it is not easy to always understand or predict what they do, and how, even if we have a broad understanding of how they work. Algorithms that are complex and opaque can make it difficult to understand their processes or intervene in their effects.

Cases: Facebook Newsfeed

In online platforms, the problem is furthermore compounded by the fact that the underlying algorithms are frequently deployed invisibly, that is to say, in such a way that users are often unaware of their existence (Eslami et al., 2015). To make things even worse in terms of transparency or accountability, many algorithms are truly opaque in the sense that their source code, principles of functioning and their basic operations are truly secret. Proprietary

commercial algorithms are used by corporations in a variety of settings for functions ranging from regulating online platforms to Wall Street and many other industries (Pasquale, 2015).

Opaque algorithms have become so widespread that Frank Pasquale refers to "black box" society when examining the role of such algorithms especially in areas of finance, credit and search engine rankings (Citron & Pasquale, 2014; Pasquale, 2015). The most obvious example for complex and opaque algorithms is Facebook's newsfeed, the primary means that its billion-plus user base accesses the updates posted by family, friends, acquaintances, civic or corporate pages, and which is responsible for up to 40 percent of traffic publishers receive (Benton, 2015). The Newsfeed is curated by an algorithm that decides what to show and what to hold back. The algorithm is tweaked, weekly, by a team of researchers headed by a 26 year old-engineer who said that they take "thousands and thousands" metrics into consideration (Somaiya, 2014). However, in a recent study, 62 percent of students in a high level school were not even aware of the existence of this algorithm, let alone its inner workings (Eslami et al., 2015). Despite its considerable clout, reach and everyday engagement, crucial algorithms such as that of Facebook remains opaque to all of its users, and its effects are likely too complex to predict, even to Facebook's own engineers.

As this example illustrates, algorithms that are sufficiently complex are often practically inscrutable to outside observers, even though they inevitably have values, biases and potential discrimination built in. The complex and opaque nature of many algorithmic operations is often overlooked in calls to regulate them, or to hold them accountable.

2. Gatekeeping function:

Another function that makes algorithms significant for ethical considerations is their role as gatekeepers. Increasingly, algorithms decide what gets attention, and what is ignored; and even what gets published at all, and what is censored. While we are familiar with cases like search, or the curation of social media timelines, their role expands to areas such as hiring and geolocation-specific content.

Case: Platforms and Tensions of Gatekeeping & Jurisdictional Tensions

In the above section, we already gave examples from Facebook, which is effectively a gatekeeper for news in the online sphere. Google search ranking is another obvious example that much research, and regulatory efforts, has been focused on (Grimmelmann, 2008, 2013; Hazan, 2013; Powles & Chaparro, 2015; Timberg, 2013). What has not gained the same amount of attention, however, is the way in which Google's "answer boxes"—that is, framed top-level search results that automatically pull the "best" answer to a search query from third-party sources—are responsive to location-based data. For example, when the phrase "What time is it in London?" is queried on Google, the resulting answer presented in the "answer boxes" will be based on the time in London, England unless the user is within a certain proximity to London, Ontario. While this may seem straightforward with regards to simple statements of fact, it becomes more complicated when more complex subjective questions are involved than deciding the local time.

An excellent illustration of that problem is the case of contested national territories, with different interpretations of legitimacy and legality depending on one's jurisdiction. One's online view of a particular location on Google or Bing's maps is frequently dependent on the physical location from which one is accessing them (Yanofsky, 2014a, 2014b). Thus, the manner in which certain contested borders—such as those around Kashmir—are shown, will differ as such: "Google Maps will show Kashmir to be part of Pakistan to users in Pakistan, in line with

Pakistani national legislation, but show Kashmir to be part of India to users in India, in line with Indian legislation, yet show the territory to be contested in the rest of the world." (Wagner, 2013) Automated algorithms create a different geopolitical reality for different users. Thereby, the way that information is presented by social media companies can affect how users perceive physical and geopolitical reality on such platforms. By (having to) deal with conflicting national jurisdictions, platforms act as gatekeepers.

Case: Hiring

Algorithms also play gatekeeping roles beyond digital platforms, and these should not be overlooked. For example, algorithms, rather than managers, are increasingly taking a significant part in hiring decisions (as well as firing). Hiring and employment is a significant gate-keeping step that holds important consequences for both individuals and society. Discrimination in hiring, firing and similar areas has the potential to create lifelong effects. People's first few jobs, for example, tend to have long-term effects on their lifelong earnings, and the jobs that they will be able to attain later in life. As such, hiring and firing is among the most important gatekeeping functions in society.

Hiring (and firing) are important case studies in studying the ethics of algorithms as they demonstrate that the answer cannot simply be "humans do it better" since research going back decades also demonstrates that human managers display significant biases in hiring. For example, people tend to hire from their own social class, race and gender (Altonji and Blank 1999). Interviewing is a fraught area, as interviews that last hours or even days have not been shown to be a good indicator for future job performance. People whose names sound African-American have been repeatedly found to experience great disadvantages in the job market in the United States, with African-American resumes with college degrees and no criminal background being passed over for applicants with white-sounding names, less education and a criminal background (Bertrand & Mullainathan, 2004). Women, for example, rarely used to be picked for the highest-prestige symphony orchestras in open auditions. However, after increasing complaints, curtains were employed in which the judges only heard the instrument being played, but did not see the musician. This resulted in many more women passing the auditions (Goldin & Rouse, 2000). Clearly human hiring systems are far from perfect.

Currently, more and more companies are moving to algorithmic hiring, which has turned out to be more accurate, in some measurable dimensions, than human gatekeeping for jobs (Rosenblat, Kneese, and boyd, 2014). As with all such systems, the system's power partly comes from the increasingly large amounts of personal data, as well as future performance measurements that are fed into such a system. In fact, many algorithms are becoming more and "learning" algorithms. This means that the algorithm does not need to be told what the rules are, but instead is merely supplied with inputs and desired outputs.

Through increasingly advanced machine learning systems, such algorithms can fine-tune themselves, learning over time what works and what does not in complex ways. Hence, a hiring algorithm only requires large amounts of data from each hire, which are later matched with performance and turnover metrics. Such a machine learning system will not just be too complex to understand because the code is complicated, but it will also be doing things that the programmers did not program directly.

While this looks like a rosy scenario—replacing human gate-keeping with more meritorious, and successful algorithmic gatekeeping, already the few insights we have about proprietary algorithms raises quick alarms. For example, such computational hiring systems have found that

low commute time corresponds to low turnover (Walker, 2012). In other words, not hiring from neighborhoods that are either far away, or have substandard public transportation, is good for business. Obviously, in many countries, especially residentially, heavily-segregated ones like the United States, such hiring patterns would tend to further depress employment among the poor and minorities, which would create even worse conditions in such neighborhoods. Such worsening conditions would likely lead to even lower job performance, as stressful lives, lack of child-care support and other dimensions of poor neighborhoods tend to also create conditions for high job turnover. Since algorithms would stop hiring from such neighborhoods, this would create a downward spiral of discrimination, unemployment, and further worsening conditions. This is obviously bad public policy and may well include racial and ethnic discrimination in countries where neighborhoods are ethnically or racially segregated. All these downsides of hiring algorithms are not an argument for why human hiring is perfect, or lacks discrimination, but a clear reminder that even when algorithms are replacing faulty systems, and solving some problems, they often introduce new, sometimes unanticipated ones.

3. Algorithmic Subjective decision-making

A lot of public debates about algorithms and automation focus on cases where there is a right answer, and the crucial question is whether the algorithm can figure this answer out as well as —or better—than a human. Can IBM's artificial intelligence computer Watson beat Jeopardy super-champ Ken Jennings? Can automated systems fly planes or drive cars? Can computers understand people talking in plain language without us having to translate it into computer-readable formats? Can algorithms predict heart attacks as well as doctors can? Algorithms where there are right answers, like medical diagnostics or flying planes, raise many important issues of accountability, employment and role of human skills in our world, but the context for evaluating them is quite different than algorithms without checkable right answers.

What we are concerned here are cases where there is no right answer, but only judgment, sensibility and values. As algorithms move from playing chess to making matches in online dating and selecting news to read, they are now answering a whole new category of questions. What is important? What is relevant? What is love? What is worthy of attention and what should be ignored, or suppressed? What should you watch or read? Who is a threat to public safety, and who is not? Who should be allowed to fly? Who should you date? It is these types of subjective decisions that are increasingly and quietly being turned over to algorithms, and that we need to discuss, on their own.

Case: Recommender Systems and Terrorist or Extremist Content

One such example is the use of algorithms to assist in the regulation and removal of online content. Following the attack on the offices of French satirical publication *Charlie Hebdo*, as well as the continued horrors perpetrated by Islamic State, or ISIS, numerous governments have called upon social media companies to be proactive in moderating "terrorist" content, in order to weed out material that is either deemed supportive of or aiding in the recruitment of terrorists.

While the automated flagging of so-called terrorist content has gained a lot of attention lately, the practice has already played a part in the investigation into the murder of the British soldier Lee Rigby in London 2013. According to a report from the 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.

From the report:

> 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...*430
>
> Such accounts are then automatically suspended. (Rifkind, 2014)

The fact that social media companies had flagged extremist content by the killers, led to complaints by intelligence agencies that the company in question had not warned them (Rifkind, 2014). The company's position appears to be that they could not possibly report all algorithmic removals of content to the government. Nonetheless, this case shows that government agencies are clearly interested in automated flagging of "terrorist" or "extremist" content. There is already research that purports to model and predict "a terrorist organization's probable actions." (Xue, Wang, & Zhang, 2011). The use of algorithms to identify such content is likely to gain momentum as interest in this topic increases.

This practice raises a number of important concerns: how and by whom should terrorism be defined? Most social media companies regulate, but do not define, terrorist content. As the above example illustrates, "terrorism" in these contexts is determined by lawmakers, and their definition are used to underpin corporate policy and, which is carried out or assisted by algorithms.

Anecdotal evidence and media reporting about the application of these rules seem to suggest that many platform policies on this topic are guided by United States law, which designates certain groups as "foreign terrorist organizations" and stipulates that anyone who provides "material support" to those groups be penalized. (U.S. Department Of State, 2015; Legal Information Institute, 2015) Interpretation of this law in the context of online service providers varies.

In addition to the definition of "extremism", "terrorism" or "extremist content", a further concern lies in the practice of flagging or removing such users or content. Facebook's "Community Standards" state that "dangerous organizations"—including terrorist groups and groups engaged in organized criminal activity—may not have a presence on the platform. However, Facebook's ability to proactively and algorithmically remove such content seems minimal, despite media reports. The company is instead reliant on its users to "flag" content that they suspect of violating the rules (Crawford, & Gillespie, 2014). The content is then reviewed by staff and, if determined to violate the "Community Standards", removed from the platform.

Some of the effects of algorithmic content-removal can be seen more directly in Facebook's "community pages". In 2010, Facebook introduced community pages, which pulled information from Wikipedia's API (Park, 2010). The intent behind the addition was to ensure that if a user searched Facebook for a particular page or content, they would not be met with a blank set of results. There is then, in theory, a Facebook page to match every Wikipedia page in English.

A 2014 search for certain groups, however, found that Facebook had removed the Wikipedia text content from certain pages, mainly those representing US-designated terrorist organizations. Examples of blank Wikipedia pages included several that contained information on such groups

—including “Hezbollah Movement in Iraq,” Hezbollah’s official page, the “Jewish Defense League,” and “Islamic Jihad Union”—as well as several others, such as “Nazism” and “Neo-Nazism.” The page representing the Mojahedin-e-Khalq (MEK)—a group that was until 2009 listed by the EU and until 2012 by the Department of State—was also blank. Somewhat troublingly, the page for “Islamic State of Iraq and Levant” was blank, but nevertheless labelled a “country” by Facebook’s internal tagging system (York, 2014).

This example shows that the reliance on algorithms to remove content raises the additional concern of potential for unintended effects, either by removing content that might have been considered acceptable had it been reviewed by human eyes, or vice versa, by not removing content that a human might flag as unacceptable. 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.

Finally, a further complication comes in the form of recommender systems in many platforms, which algorithmically show a person what else they may be interested in. This raises additional issues in the case of extremist or terrorist content as the recommender system can keep “recommending” further extremist material once a user has watched just one, helping strengthen the creation of “ideological bubbles” (O’Callaghan, Greene, Conway, Carthy, & Cunningham, 2014). The role of algorithmic recommender systems in creating filter bubbles is an important topic that has already been explored in other contexts but is specifically implicated in the questions of terrorism, extremism or violent content online (Bozdag, 2013; Flaxman, Goel, & Rao, n.d.; Pariser, 2012; Zuckerman, 2014).

Case predictive policing

Another example for the use of algorithms in subjective decision-making is predictive policing. In early 2014, the Chicago Police Department made national headlines in the US for visiting residents that were considered to be most likely involved in violent crime (Robinson et al.). The selection of individuals, which were not necessarily under investigation, was guided by a computer-generated “Heat List”; an algorithm that seeks to predict future involvement in violent crime. While officials have given some indication of how this algorithm operated - it takes arrest records into account - there is no public description of the algorithm’s operation or input. A Freedom of Information Act request to see the names on the list was denied.

Whether some systems prove to be effective in reducing crime or not, algorithmic prediction of future crime behaviour is problematic on a number of grounds. The primary concern is that such automated systems may create an echo chamber, or self-fulfilling prophecy. Areas or communities that are more heavily policed increase the likelihood that crime will also be detected, as more police means more opportunities to observe residents’ activities. At the same time, the focus on particular areas or communities decreases the intensity of policing elsewhere. Moreover "predictions“ are just that—predictions with no guarantee they are right, with both false negatives (missing future crime) and false positives (identifying potential crime where there will be none). If officials act on incorrect predictions, they may create unjustified guilt-by associations. In addition to the negative effects of being interrogated, or being under close observation prior to being under investigation, increased police presence may create further complications in areas that suffer from systematic biases in the application of law.

Case: trending

We have already discussed the role of algorithms in content regulation and the compassion of Facebook newsfeed. A further place where algorithms make subjective-decisions is in the case of trending topics on Twitter.

Before discussions about the subjective nature of Facebook's news feeds in the cases of Ferguson, (Tufekci, 2014), there had been a controversy on whether Occupy movement's hashtags were suppressed on Twitter (Gillespie, 2011). Even though the Occupy movement has been widely covered in national and international media, curiously it had not trended nationally, causing some to suspect "censorship". However, for any topic to become "trending" on Twitter requires much more than simply a large number of mentions. Twitter employs an algorithm, called term-frequency, inverse-document-frequency, to detect trending topics. In essence, this algorithm "weeds out" frequent terms as background noise. This means that things like the word "the" or "Monday", or popular topics like "Justin Beiber" don't constantly trend on Twitter. The trending topics algorithm also takes into account the speed with which something is spiking, how connected the users that talk about the topic are, the amount of unique content or retweets, and whether the hashtag has trended before (Gillespie, 2011). Defining that any particular hashtag is trending is relevant, because it creates a feedback loop. Once a topic is "trending", it is presented to all Twitter users who have chosen that particular area (topics trend by geography) as as column. This has the potential to gain momentum and attract more readers or participants to the topic that is trending.

Indeed, such attention is quite valuable, and part of Twitter's business model: corporations may purchase a space in trending topics column. However, the algorithms that define social media trends are making a very political and subjective decision, since there is no single, objective, uncontested assessment of the public's complete traces, and the question of "who claims to know the mind of the public, and how they claim to know it" remains relevant when the curatorial decision is not made by a person or an editor, but by an algorithm.

In the case of Occupy Wall Street movement, what had happened was that the movement had grown over time instead of spiking in a spectacular fashion all at once, a spike. The 'slowish' growth of the hashtag over time had caused the algorithm to weed it out as background noise, hence causing it not to trend. In a curious correspondence, Twitter's algorithm rewarded short-term attention, similar to the way the evening news on TV may cover one-time, outrageous events over chronic problems, the way a single homeless man's tragic death may garner more coverage than the chronic problem of homelessness.

This case of trending algorithms directly corresponds to the function of algorithms to produce calculated publics (Gillespie, 2014). Here, algorithms are used to create a version of the public and present it back to itself, thereby shaping the public's sense of itself.

Case: internet of things

The power of subjective algorithms is getting more pervasive as digital technologies are used in more and more areas of our lives, and rapidly spread into our offline environments through sensors and the so-called "internet of things" – in which a lot of previously inert devices are expected soon to start becoming networked, take inputs and potentially make decisions for us. It's not much of an issue when it is the refrigerator deciding that the house needs more milk, and having it delivered, but the thorny issues have the potential to arise quickly.

Should your door call the police if it is opened by someone without your tracking device? What if your walls hear the sounds of a fight or a child crying too many days in a row? Who owns the data your nightstand and your reading device knows about you? Who is responsible for the

security of this data? Should your insurance company get a peek into what your fridge knows about you? Should a transportation company punish get to learn about—and maybe fire—drivers whose coffee drinking habits it does not approve of?

These are not all hypothetical. Hospitals are purchasing consumer data to plug into the algorithms that will predict who is more likely to get sick soon. The chief officer responsible for analytics at Carolinas Care, a hospital chain, defended the practice by saying that "consumer spending" can give a more complete picture than doctors get in an office visit (Pettypiece, 2014). Health-care institutions are also algorithmically predicting which patients are likely to be "high-cost" and which ones may experience medical complications (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014). Obviously, such data can be of use if deployed with the interests of the patient in mind. But will it? And what happens if the algorithm is wrong? Or who is accountable if there is an erroneous decision? How will we even know? Increasing digitization of things and spread of sensors will only raise more questions in this arena.

From extremist content regulation and the curation of newsfeeds, to predictive policing and the internet of things; algorithms are making decisions for which there is no simple answer, only judgment, sensibility and values. In this section, we have started by discussion the ethical challenges of algorithms that are complex, opaque, and serve gate-keeping functions. These functions pose new challenges, but these are compounded in cases where the algorithmic output has no agreed upon metric or right answer which can be used to calibrate the effects. If computers are making these judgments, which are considered highly subjective when humans make them, what does it mean for them to be right or wrong? Whose interests do they serve? What are their inputs? What about the cases where they are invisible to users? Who oversees all of this? These questions and more remain urgently in need of exploration.

4) Regulatory responses

Algorithms are becoming ever more capable and used to make decisions for us, about us or with us. In this report, we have argued that in particular algorithms that are opaque, act as gatekeepers and make subjective decisions require ethical scrutiny.

While the scope of ethical challenges is very broad, the above mentioned cases raise the question how those creating complex algorithms of concern, policy-makers, civic actors, citizens and everyone concerned about the ethical, legal and policy frameworks in the 21st century should address these challenges. In the following, we would like to briefly address three responses that have been proposed and discussed on this matter.

1. Algorithmic transparency and notification

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. Whenever black-box algorithms make inherently subjective decisions, the concern becomes that the algorithm contains implicit or explicit biases. If an algorithm is opaque, it becomes impossible for outsiders to understand the rationale behind any particular outcome, or when algorithms are misused. An example for the latter is the controversy surrounding search bias. The Federal Trade Commission conducted an exhaustive investigation of Google's internet search practices during 2011 and 2012, concluding that search practices were not, "on balance, demonstrably anticompetitive" (Federal Trade Commission, 2015), while the European Commission continues an antitrust investigation. Could more transparent be a solution to the problems opaque algorithms cause? And what exactly would transparency imply?

As we have illustrated for the case of hiring algorithms, making complex algorithms transparent can be extremely challenging. In 2011 alone, Google made 538 changes to its search algorithms (Google, 2012), each preceded by numerous tests, run on a small percentage of search users. Merely publishing the source code of an algorithm is not sufficient. Machine learning systems will inevitably make decisions the programmers did not program directly. Hence, it has been suggested that transparency is required at multiple dimensions of algorithmic decision-making (Shadoan, 2014): data inputs, control surfaces, algorithm steps and internal state, assumptions and models the algorithm uses, as well as justification for outputs produced. In other words, complete transparency requires that we can explain why any particular outcome was produced.

This is furthermore complicated by the fact that there are often good reasons why complex algorithms operate in an opaque matter. Public access to algorithms such as search, social media feeds, or even online dating sites makes these considerably more vulnerable to manipulation and spam; and complete documentation takes time and money. Another important question is whom algorithms should be transparent to? Public access to the source code of an algorithm does not guarantee scrutiny; and, especially if there is a lack of alternative services, does not necessarily help consumers and citizens.

A very different form of transparency is notification. Consumers can demand for control over their personal information that feeds into algorithms which might have a considerable effect on their lives. This includes the rights to correct information and demand their personal information to be excluded from the database of data vendors.

2. Algorithmic accountability

Due to the ever-changing, complex and opaque nature of algorithms, a major civil responsibility consists in understanding how exactly any particular algorithm works. Journalists, academics and activists have done this by reverse engineering algorithms that deserve ethical scrutiny.

Nicholas Diakopoulos (2013), for instance, has established the algorithmic rules according to which search engines censor autocomplete on sexual and violent search terms. Any automated and opaque censorship of this kind raises the important of how exactly boundaries and editorial criteria are defined and how they differ among search engines. At Harvard, Latanya Sweeney (2013) has investigated how online advertisement can be biased by the racial association of names used as queries. Within journalism, various authors have experimented with the function and behaviour that influence the Facebook newsfeed algorithm (Honan, 2014; Morgan, 2014)

While such investigations might not be able to establish why it is exactly an algorithm creates any particular outcome, they are the essential precondition for the public scrutiny of algorithms. On the individual level, Sarah Watson (2014) has made a similar argument with regards to the uncanny experience of all-too personal targeted advertisement. Causal explanations that link our digital experiences with the data they are based upon can empower individuals to better understand how the algorithms around them are influencing their life-worlds.

3. Governments directly regulating an algorithm

Transparency, or reverse-engineering algorithms may create public awareness, but should some algorithms be regulated more directly? One area where algorithmic regulation has been discussed in is finance. Due to the increasing use of automated high-speed trading systems and their potentially destabilising effect on financial markets, regulators have begun to demand both transparency over high-speed trading algorithms and the ability to modify these algorithms if they are considered unstable (Steinbrück 2012). Such regulation has for instance been suggested by Peer Steinbrück, the social democratic (SPD) candidate for the German chancellor in 2013. With regards to high-speed financial trading systems a policy paper suggested:

> "The core of an effective regulation needs to be a public certification system not just for trading companies, but for the trading algorithms themselves. This certification system will first analyse the algorithm based on its trading strategy: dangerous trading strategies must be banned! Moreover the algorithms will have to undergo a stress test to ascertain its stability" (Steinbrück 2012).

Another prominent example for the desire to regulate algorithms directly is the case of search. Although public regulators have not yet been able to influence Google algorithms, the on-going FTC and EU investigation into 'search neutrality' revolves precisely around this question (Manne and Wright 2012; Kanter and Streitfeld, 2012). Can regulators require Google to force its algorithm to act in certain ways towards certain competing sites? This would in effect require public regulators to have access to the algorithm, employ individuals capable of understanding its properties and be able to modify it effectively in the interests of the public. Such regulation also assumes that it is possible to objectively predict the responses of certain algorithms to different types of situations and that the assessment of such algorithms is more effective than the assessment of the people developing it. Even defining what would be in the "public interest" is a complex and contested question, exactly because there is no right answer to how Google—or any other search engine—should rank its results.

Directly regulating specific algorithms is characterised by a number of problems. As we have argued above, complex algorithms are automated systems that have moved beyond a simplistic variable based response, but are instead many are trying to respond to their surrounding through based on a 'machine learning' process or other kinds of complex computations. Given this complexity, second order rules are occasionally introduced in order to modify the prima facie responses of the system.

In this section, we have discussed three popular forms of regulatory responses that are most commonly called for to tackle the ethical challenges of algorithms: transparency, accountability and direct regulation. As algorithms are spreading to ever more parts of society, the consequent ethical challenges become more diverse as well. Especially when algorithms are involved in inherently subjective decision-making, it is important that the underlying processes are subject to some form of scrutiny. While in some cases, direct regulation, or complete and public transparency is necessary, there is no one-size-fits-all regulatory response. To enable scrutiny requires new practices from industry and technologists, but also calls for consumer protection and more direct regulation, where appropriate.

5) Conclusions

The increasing importance of algorithms raises a whole host of ethical quandaries, some of which involve issues of regulatory policies. As computational devices become more and more widespread in society, and as smart, networked devices are embedded in more of our objects—through the internet of things—this will all become central to our lives. Moreover, as is the case with most technologies, as such devices become mundane, there will be a tendency to treat them as "black boxes", and their operations as just the way things are. Like all transitions, this is a crucial historical period to grapple with all the implications raised by our new technologies.

Algorithms raise many questions. How does free speech operate, if algorithms are deciding what is allowed to be published, or what acquires attention and visibility? Now that machine learning can, opaquely, pick out all sorts of subgroups beyond ordinary demographics, how do we deal with questions of discrimination and equal opportunity? If private, corporate actors are increasingly deploying these algorithms, within their business models, how do we deal with questions of accountability and due process in the public sphere? What are our rights, as citizens in this networked world, to notification and transparency? And finally, how do conceptualize, understand and exercise human agency in a new world in which machines, too, are acting with agency and purpose? These questions and more deserve urgent and deep attention by everyone concerned about the future shape of human society.

REFERENCES

Altonji, JG, and RM Blank. (1999). "Race and Gender in the Labor Market." *Handbook of labor economics*.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. Health Affairs, 33(7), 1123–1131. doi:10.1377/hlthaff.2014.0041

Benton, J. (2015, March 25). Facebook wants to be the new World Wide Web, and news orgs are apparently on board. Nieman Lab. Retrieved from http://www.niemanlab.org/2014/12/how-much-of-your-news-sites-search-traffic-comes-from-google-news-probably-5-to-25-percent/

Bertrand, M., & Mullainathan, S. (2004). Are Emily And Greg More Employable Than Lakisha And Jamal? A Field Experiment On Labor Market Discrimination. American Economic Review, (94), 991-1013.

Bozdag, E. (2013). Bias in Algorithmic Filtering and Personalization. *Ethics and Information Technology*, *15*(3), 209–227. http://doi.org/10.1007/s10676-013-9321-6

Citron, D. K., & Pasquale, F. A. (2014). The Scored Society: Due Process for Automated Predictions. Washington Law Review, 89(1). Retrieved from http://digitalcommons.law.umaryland.edu/fac_pubs/1431/

Crawford, K., & Gillespie, T. (2014). What is a flag for? Social media reporting tools and the vocabulary of complaint. New Media & Society. doi:10.1177/1461444814543163

Diakopoulos, N. (2014). Algorithmic Accountability. Digital Journalism, 1–18. doi: 10.1080/21670811.2014.976411

Diakopoulos, N. (2013, August 2). Sex, Violence, and Autocomplete Algorithms. Retrieved from www.slate.com/articles/technology/future_tense/2013/08/words_banned_from_bing_and_google_s_autocomplete_algorithms.html

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., … Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in the news feed. In Proceedings of the 33rd Annual SIGCHI Conference on Human Factors in Computing Systems. Retrieved from http://social.cs.uiuc.edu/papers/pdfs/Eslami_Algorithms_CHI15.pdf

Federal Trade Commission. (2015, March 25). Statement of Chairwoman Edith Ramirez, and Commissioners Julie Brill and Maureen K. Ohlhausen regarding the Google Investigation. Retrieved April 29, 2015, from https://www.ftc.gov/news-events/press-releases/2015/03/statement-chairwoman-edith-ramirez-commissioners-julie-brill

Flaxman, S. R., Goel, S., & Rao, J. M. (n.d.). Filter Bubbles, Echo Chambers, and Online News Consumption↤. Retrieved from https://5harad.com/papers/bubbles.pdf

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), Media technologies: Essays on communication, materiality, and society (p. 167). Cambridge, MA: MIT Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=zeK2AgAAQBAJ&oi=fnd&pg=PA167&dq=the+relevance+of+algorithms&ots=GmiFR0Z2Ak&sig=ndjEc6rrPmmS2tzZdI1P4S1EG1I

Gillespie, T. (2011). Can an Algorithm be wrong? Culture Digitally. Retrieved from: http://culturedigitally.org/2011/10/can-an-algorithm-be-wrong/

Grimmelmann, J. (2008). The Google Dilemma. *New York Law School Law Review*, *53*, 939.

Goldin, C., & Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, 90, 715-741.

Google. (2012). Algorithms. Retrieved March 29, 2014, from http://www.google.com/intl/en/insidesearch/howsearchworks/algorithms.html

Hazan, J. G. (2013). Stop Being Evil: A Proposal for Unbiased Google Search. *Michigan Law Review*, *111*(5), 789+.

Honan, M. (2014, November 8). I liked everything I saw on Facebook for two days. Here's what it did to me. *WIRED*. Retrieved from: http://www.wired.com/2014/08/i-liked-everything-i-saw-on-facebook-for-two-days-heres-what-it-did-to-me/

Kanter, J., Streitfeld, D. (2012, Mai 21). Europe Weighs Antitrust Case Aigainst Google, Urging Search Changes. *New York Times.* Retrieved from http://www.nytimes.com/2012/05/22/business/global/europe-warns-google-over-antitrust.html

Legal Information Institute. (2015.) 18 U.S. Code § 2339A - Providing material support to terrorists. Retrieved from: https://www.law.cornell.edu/uscode/text/18/2339A

Manne, GA, and JD Wright. (2012). If Search Neutrality Is the Answer, What's the Question. *ICLE Antitrust & Consumer Protection Program White Paper Series.*

Morgan, E. (2014, August 14). I Quit Liking Things On Facebook for Two Weeks. Here's How It Changed My View of Humanity. *Medium*. Retrieved March 29, 2015, from https://medium.com/@schmutzie/i-quit-liking-things-on-facebook-for-two-weeks-heres-how-it-changed-my-view-of-humanity-29b5102abaxe

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2014). Down the (White) Rabbit Hole The Extreme Right and Online Recommender Systems. *Social Science Computer Review*, 0894439314555329. http://doi.org/10.1177/0894439314555329

Pariser, E. (2012). *The Filter Bubble: What the Internet is Hiding from You*. New York: Penguin Press.

Park, J. (2010). Wikipedia on new Facebook community pages. *Creative Commons*. Retrieved from: http://creativecommons.org/weblog/entry/21721

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

Pettypiece, S. (2014, July 3). Hospitals Are Mining Patients' Credit Card Data to Predict Who Will Get Sick. *Bloomberg View*. Retrieved from http://www.bloomberg.com/bw/articles/2014-07-03/hospitals-are-mining-patients-credit-card-data-to-predict-who-will-get-sick

Powles, J., & Chaparro, E. (2015, February 18). How Google Determined Our Right to Be Forgotten. *Guardian*. Retrieved from http://www.theguardian.com/technology/2015/feb/18/the-right-be-forgotten-google-search

Rifkind, M. (2014). Report on the intelligence relating to the murder of Fusilier Lee Rigby [Woolwich Report}. *Intelligence and Security Committee of Parliament*.

Robinson, D., Yu, H., & Rieke, A. (2014). Civil Rights, Big Data, and Our Algorithmic Future. *Robinson + Yu*. Retrieved from http://bigdata.fairness.io/

Rosenblat, A., Kneese, T., boyd, d. (2014). Networked Employment Discrimination. *Data & Society Working Paper.* Retrieved from http://www.datasociety.net/pubs/fow/EmploymentDiscrimination.pdf

Shadoan, R. (2014, July 11). Why Algorithm Transparency is Vital to the Future of Thinking. *Akashic Labs*. Retrieved from http://www.akashiclabs.com/why-algorithm-transparency-is-vital-to-the-future-of-thinking/

Steinbrück, P. (2012). *Vertrauen Zurückgewinnen: Ein Neuer Anlauf Zur Bändigung Der Finanzmärkte*. Berlin, Germany.

Somaiya, R. (2014, October 26). How Facebook Is Changing the Way Its Users Consume Journalism. *New York Times*. Retrieved from http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-consume-journalism.html

Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Communications of the ACM*, 56(5), 44-54.

Timberg, C. (2013, March 31). Could Google Pick the Next President? *The Washington Post*, p. OUTLOOK; Pg. B04.

Tufekci, Z. (2014). What Happens to #Ferguson Affects Ferguson. *Medium*. Retrieved from: https://medium.com/message/ferguson-is-also-a-net-neutrality-issue-6d2f3db51eb0

U.S. Department Of State. (2015). Foreign Terrorist Organizations. http://www.state.gov/j/ct/rls/other/des/123085.htm

Wagner, B. (2013). Governing Internet Expression: The International and Transnational Politics of Freedom of Expression. *European University Institute*.

Walker, J. (2012, September 20). Meet the New Boss: Big Data. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB10000872396390443890304578006252019616768

Yanofsky, D. (2014a). Here are the 32 countries Google Maps won't draw borders around. *Quartz*. Retrieved from: http://qz.com/218675/here-are-the-32-countries-google-maps- wont-draw-borders-around/

Yanofsky, D. (2014b). See how borders change on Google Maps depending on where you view them. *Quartz*. Retrieved from: http://qz.com/224821/see-how-borders-change-on-google-maps-depending-on-where-you-view-them/

York, J. (2014). Let's Talk about Terrorism and Facebook. http://jilliancyork.com/2014/09/15/ets-talk-about-terrorism-and-facebook/.

Zuckerman, E. (2014). Digital Cosmopolitans: Why We Think the Internet Connects Us, Why It Doesn't, and How to Rewire It (1 edition). New York: W. W. Norton & Company.