



# Algorithmic governance

**Christian Katzenbach**

*The evolving digital society, Alexander von Humboldt Institute for Internet and Society, Berlin, Germany, christian.katzenbach@hiig.de*

**Lena Ulbricht**

*Berlin Social Science Center (WZB), Germany*

Published on 29 Nov 2019 | DOI: 10.14763/2019.4.1424

**Abstract:** Algorithmic governance as a key concept in controversies around the emerging digital society highlights the idea that digital technologies produce social ordering in a specific way. Starting with the origins of the concept, this paper portrays different perspectives and objects of inquiry where algorithmic governance has gained prominence ranging from the public sector to labour management and ordering digital communication. Recurrent controversies across all sectors such as datafication and surveillance, bias, agency and transparency indicate that the concept of algorithmic governance allows to bring objects of inquiry and research fields that had not been related before into a joint conversation. Short case studies on predictive policy and automated content moderation show that algorithmic governance is multiple, contingent and contested. It takes different forms in different contexts and jurisdictions, and it is shaped by interests, power, and resistance.

**Keywords:** Transparency, Automation, Politicisation, Regulation, Social ordering, Governance, Predictive policing, Content moderation, Algorithmic governance

## Article information

**Received:** 26 Jun 2019 **Reviewed:** 20 Sep 2019 **Published:** 29 Nov 2019

**Licence:** Creative Commons Attribution 3.0 Germany

**Competing interests:** The author has declared that no competing interests exist that have influenced the text.

**URL:** <http://policyreview.info/concepts/algorithmic-governance>

**Citation:** Katzenbach, C. & Ulbricht, L. (2019). Algorithmic governance. *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1424

*This article belongs to Concepts of the digital society, a special section of Internet Policy Review guest-edited by Christian Katzenbach and Thomas Christian Bächle.*

## 1. INTRODUCTION

The concept of algorithmic governance has emerged over the last decade, but takes up an idea

that has been present for much longer: that digital technologies structure the social in particular ways. Engaging with the concept of algorithmic governance is complex, as many research fields are interested in the phenomenon, using different terms and having different *foci*. To inquire what constitutes algorithmic governance makes an important contribution to contemporary social theory by interrogating the role of algorithms and their ordering effect. We define algorithms as computer-based epistemic procedures which are particularly complex – although what is complex depends on the context. Algorithms shape procedures with their inherent mathematical logics and statistical practices. With that, the discourse around algorithmic governance often overlaps and intersects with debates about datafication (cf. Mejias & Couldry, 2019 as part of this special section) and artificial intelligence (AI). Yet, algorithms sometimes also operate on ‘small data’ and use calculus-based procedures that do not learn and that are not adaptive.

While governance is a contested term, we define its core as coordination between actors based on rules. Other than regulation, governance is not necessarily intentional and goal-directed (Black, 2001), it also includes unintentional coordination (Hofmann, Katzenbach, & Gollatz, 2016). Yet, governance excludes all forms of social ordering which are purely occasional and don’t rely on some sort of rule; governance implies a minimum degree of stability which is necessary for actors to develop expectations, which are a precondition for coordination (Hobbes, 1909). We choose the term algorithmic governance instead of algorithmic regulation because governance allows to account for the multiplicity of social ordering with regard to actors, mechanisms, structures, degrees of institutionalisation and distribution of authority. It deliberately embraces social ordering that is analytically and structurally decentralised and not state-centred. Thus, algorithmic governance better reflects the ambition of this article to widely scrutinise the ways in which algorithms create social order. In that sense, we focus on governance *by* algorithms instead of the governance *of* algorithms (Musiani, 2013; Just & Latzer, 2017). In sum, algorithmic governance is a form of social ordering that relies on coordination between actors, is based on rules and incorporates particularly complex computer-based epistemic procedures.

The relevance of dealing with algorithmic governance becomes evident with regard to competing narratives of what *changes* in governance when it makes use of algorithms: one narrative is for example that governance becomes more powerful, intrusive and pervasive. A different narrative stresses that governance becomes more inclusive, responsive, and allows for more social diversity, as we will highlight in the following chapters.

If considered broadly, the roots of this concept can be traced back to the history and sociology of science, technology and society. Technology has always both reflected and reorganised the social (Bijker & Law, 1992; Latour, 2005). From Socrates’ concerns with writing and literacy (Ong, 1982) via cybernetic’s radically interdisciplinary connection between technical, biological and social systems and their control (Wiener, 1948) via Jacques Ellul’s bureaucratic dystopia of a ‘technological society’ (1964) to Langdon Winner’s widely cited, yet contested ‘politics of artefacts’ (1980) – the idea that technology and artefacts somehow govern society and social interactions is a recurring theme. The more direct predecessor of algorithmic governance is Lawrence Lessig’s famous catchphrase ‘code is law’. Here, software code or, more generally, technical architectures are seen as one of four factors regulating social behaviour (next to law, market and social norms). Scholars have also conceptualised the institutional character of software and algorithms (Katzenbach, 2017, 2012; Napoli, 2013; Orwat et al., 2010). While Rouvroy and Berns used the term ‘gouvernance algorithmique’ in 2009, the first to conceptualise the term ‘algorithmic governance’ were Müller-Birn, Dobusch and Herbsleb

(2013), presenting it as a coordination mechanism opposed to 'social governance'.<sup>1</sup> The concept of 'algorithmic regulation' was introduced by US publisher Tim O'Reilly (2013), highlighting the efficiency of automatically governed spaces – but overlooking the depoliticisation of highly contested issues that comes with delegating them to technological solutions (Morozov, 2014). In contrast to the implicit technological determinism of these accounts, the interdisciplinary field of critical software studies has complicated – in the best sense – the intricate mutual dependencies of software and algorithms on the one hand, and social interactions and structures on the other (MacKenzie, 2006; Fuller, 2008; Berry, 2011; Kitchin & Dodge, 2011). This article sets out to provide a primer on the concept of algorithmic governance, including an overview on dominant perspectives and areas of interest (section 2), a presentation of recurrent controversies in this space (section 3), an analytical delineation of different types of algorithmic governance (section 4), and a short discussion of predictive policing and automated content moderation as illustrative case studies (section 5). We seek to steer clear of the deterministic impetus of the trajectory towards ever more automation, while taking seriously the turn to increasingly manage social spaces and interaction with algorithmic systems.

## 2. ALGORITHMIC GOVERNANCE: PERSPECTIVES AND OBJECTS OF INQUIRY

The notion of algorithmic governance is addressed and discussed in different contexts and disciplines. They share similar understandings about the importance of algorithms for social ordering, but choose different objects of inquiry. The choice of relevant sets of literature presented here has a focus on research in science and technology studies (STS), sociology, political science, communication and media studies, but includes research from relevant other disciplines interested in algorithmic governance, such as computer science, legal studies, economics, and philosophy.

Various closely related and overlapping research areas are interested in how algorithms contribute to re-organising and shifting social interactions and structures. In contrast to public debate, however, these scholars reject the notion of algorithms as independent, external forces that single-handedly rule our world. They complicate this techno-determinist picture by asserting the high relevance of algorithms (Gillespie, 2014), yet highlighting the economic, cultural, and political contexts that both shape the design of algorithms as well as accommodate their operation. Thus, empirical studies in this field typically focus on the social interactions under study and interrogate the role of algorithms and their ordering effect in these specific contexts (Kitchin, 2016; Seaver, 2017; Ziewitz, 2016). They share an interest in how data sets, mathematical models and calculative procedures pave the way for a new quality of social quantification and classification. The notions of 'algorithmic regulation' (Yeung, 2018) and 'algorithmic governance' (Just & Latzer, 2016; König, 2019) emanate from the field of regulation and governance research, mostly composed of scholars from legal science, political science, economy, and sociology. The relevant studies have had the effect of organising and stimulating research about algorithmic governance with a shared understanding of regulation – as “intentional attempts to manage risk or alter behavior in order to achieve some pre-specified goal” (Yeung, 2018). This focus on goal-directed, intentional interventions sets the stage for inquiries that are explicitly interested in algorithms as a form of government purposefully employed to regulate social contexts and alter the behaviour of individuals, for example in the treatment of citizens or the management of workers. Other approaches also study non-intentional forms of social ordering through and with algorithms.

A slightly different approach puts the *technical systems in the centre*, not the social structures and relations. Relevant studies, particularly in computer science aim to build and optimise algorithmic systems to solve specific social problems: detect contested content, deviant behaviour, and preferences or opinions – in short: they are building the very instruments that are often employed in algorithmic governance. The common goal in this approach usually is to effectively detect patterns in data, e.g., translating social context into computable processes (i.e., optimising detection). This research stream is seeking efficient, robust, fair and accountable ways to classify subjects and objects both into general categories (such as species) as well as into specific dimensions such as psychometric types, emotional states, credit worthiness, or political preferences (Schmidt and Wiegand, 2017; Binns et al., 2017). Producers and providers of algorithmically-fuelled services not only optimise the detection of patterns in existing data sets, but they often – in turn – also aim to optimise their systems to most effectively nudge user behaviour in a way that seeks to maximise organisational benefits (optimising behaviour). By systematically testing different versions of user screens or other features (A/B-testing) and applying user and behavioural analytics, companies continually work to direct user interactions more effectively towards more engagement and less friction (Guerses et al., 2018). It is, however, important to note that there is no clear line between the research that develops and optimises algorithmic governance and the research analysing its societal implications; they overlap and there are many studies that strive towards both aims. An example in case are studies about algorithmic bias, fairness and accountability that both conceptualise and test metrics (e.g., Waseem & Hovy, 2016). Another important area of research that is applied and critical are studies about ‘automation bias’, ‘machine bias’ or ‘over-reliance’ that study under which conditions human agents can take a truly autonomous decision (Lee & See, 2004; Parasuraman & Manzey, 2010).

One important domain of inquiry especially relevant to STS, communication and media studies is *digital communication and social media*. Scholars have been interested for more than a decade in how search engines and social media platforms organise and structure information that is available online and how this affects subjectivation (Couldry & Langer, 2005). Platforms prioritise certain types of content (typically based on metrics of ‘engagement’) – thus constituting a new dominant mode to ascribe relevance in society, complementing traditional journalistic routines. Platforms also deploy algorithms to regulate content by blocking or filtering speech, videos and photos that are deemed unacceptable or unlawful (Gillespie, 2018; Gorwa, 2019). With increasing scale and growing political pressure, platforms readily turn to technical solutions to address difficult platform governance puzzles such as hate speech, misinformation and copyright (Gorwa, Binns, & Katzenbach, 2019). Other areas under study that make use of automated content detection are plagiarism checks in teaching and academic writing (Introna, 2016) and sentiment analysis for commercial and political marketing (Tactical Tech, 2019).

*Public sector service provisions, citizen management and surveillance* constitute another key area of interest for algorithmic governance scholars. Especially political scientists and legal scholars investigate automated procedures for state service delivery and administrative decision-making. The ambition here is that algorithms potentially increase the efficiency and efficacy of state services, for example by rationalising bureaucratic decision-making, by targeting information and interventions to precise profiles or by choosing the best available policy options (OECD, 2015). Yet, their promises are heavily contested. Scholars have shown that the deployment of algorithmic systems in the public sector produced many non-intended and non-disclosed consequences (Veale & Brass, 2019; Dencik, Hintz, Redden, & Warne, 2018). Applying algorithmic tools in government often relies on new forms of population surveillance

and classification by state and corporate actors (Neyland & Möllers, 2017; Lupton, 2016, Bennett, 2017). The grounds for many projects of digital service provision and algorithm-based policy choice are systems of rating, scoring and predicting citizen behaviour, preference and opinion. These are used for the allocation of social benefits, to combat tax evasion and fraud, to inform jurisdiction, policing and terrorism prevention, border control, and migration management.

*Rating and scoring* are not only applied to citizens, but also to consumers, as valuation and quantification studies have pointed out with regard to credit scores (Avery, Brevoort, & Canner, 2012; Brevoort, Grimm, & Kambara, 2015; Fourcade & Healy, 2016, 2017). These studies point out how algorithm-based valuation practices shape markets and create stratification mechanisms that can superimpose social class and reconfigure power relations – often to the detriment of the poor and ‘underscored’ (Fourcade & Healy, 2017; Zarsky, 2014).

Governance through algorithms is also an important matter of concern for scholars studying the digital transformation of work, such as the sociology of labour and labour economics. The objects of study here are automated governance on labour platforms and *management of labour* within companies, for example through performance management and rating systems (Lee, Poltrock, Barkhuus, Borges, & Kellogg, 2017; Rosenblat, 2018). This research field is characterised by empirical case studies that enquire the implications of algorithmic management and workplace surveillance for workers’ income, autonomy, well-being, rights and social security; and for social inequality and welfare states (Wood, Graham, Lehdonvirta, & Hjorth, 2019). Related objects of inquiry are algorithmic systems of augmented reality, of speech recognition and assistance systems for task execution, training and quality control (Gerber & Krzywdzinski, 2019). Important economic sectors under study are logistics, industrial production, delivery and services. Other relevant areas of research focus on the algorithmic management of transportation and traffic, energy, waste and water, for example in ‘smart city’ projects.

Some scholars approach algorithmic governance on a meta-level as a form of decentralised *coordination and participation*. They stress its power to process a high number of inputs and thus to tackle a high degree of complexity. As a consequence, they see algorithmic governance as a mode of coordination that offers new opportunities for participation, social inclusiveness, diversity and democratic responsiveness (König, 2019; Schrape, 2019). There is abundant research about the possibilities that software can offer to improve political participation through online participation tools (Boulianne, 2015; Boulianne & Theocharis, 2018), such as electronic elections and petitions, social media communication and legislative crowdsourcing. In addition, countless algorithmic tools are being developed with the explicit aim to ‘hear more voices’ and to improve the relationship between users and platforms or citizens and political elites. However, algorithmic governance through participatory tools often remains hierarchical, with unequal power distribution (Kelty, 2017).

### 3. CONTROVERSIES AND CONCERNS

Across these different perspectives and sectors, there are recurring controversies and concerns that are regularly raised whenever the phenomenon of algorithmic governance is discussed. Looking at these controversies more closely, we can often detect a dialectic movement between positive and negative connotations.

## DATAFICATION AND SURVEILLANCE

The literature about algorithmic governance shows an ample consensus that big data, algorithms and artificial intelligence change societies' perspectives on populations and individuals. This is due to the 'data deluge', an increase and variety in data collected by digital devices, online trackers and the surveillance of spaces (Beer, 2019). 'Datafication' (cf. Mejias & Couldry, 2019 as part of this special section) also benefits from increasingly powerful infrastructures which enable more and faster data analysis, and societal norms that benefit quantification, classification and surveillance (Rieder & Simon, 2016). Research about algorithmic governance has nevertheless always been concerned with the many risks of datafication and surveillance. To surveil entire populations and to create detailed profiles about individuals on the basis of their 'data doubles' creates ample opportunities for social sorting, discrimination, state oppression and the manipulation of consumers and citizens (Lyon, 2014; Gandy, 2010). Unfettered surveillance poses danger to many civil and human rights, such as freedom of speech, freedom of assembly, and privacy, to name just a few.

## AGENCY AND AUTONOMY

The ubiquity of algorithms as governance tools has created concerns about the effects on human agency and autonomy (Hildebrandt, 2016) – a central concept of the Enlightenment and a key characteristic of the modern individual. While earlier approaches conceived of algorithms as either augmenting or reducing human agency, it has become clear that the interaction between human and machine agents is complex and needs more differentiation. While typologies and debates typically construct a binary distinction between humans-in-the-loop vs. humans-out-of-the-loop, this dichotomy does not hold for in-depth analyses of the manifold realities of human-computer-interaction (Gray & Suri, 2018). In addition, human agency cannot only be assessed with regard to machines, but also with regard to constraints posed by organisations and social norms (Caplan & boyd, 2018).

## TRANSPARENCY AND OPACITY

The assumed opacity of algorithms and algorithmic governance is a strong and lasting theme in the debate, routinely coupled with a call for more transparency (Kitchin, 2016; Pasquale, 2015). However, more recent arguments point out that access to computer code should not become a fetish: absolute transparency is often not possible nor desirable and not the solution to most of the problems related to algorithmic governance, such as fairness, manipulation, civility, etc. (Ananny & Crawford, 2017; Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). In addition, the implementation of social norms into code not only creates opacity, but also unveils norms and processes that were previously hidden. A case in point are controversies around scoring systems about unemployment risk, as deployed in Austria and Poland (AlgorithmWatch and Bertelsmann Stiftung, 2019), and credit worthiness (AlgorithmWatch, 2019). The public interest in algorithmic governance has motivated civil society actors and scholars to inquire the composition and rationality of algorithmic scoring and to question the underlying social values. Given this development, the current turn to algorithmic governance might indeed even be conducive to more transparency as software code, once disclosed, requires the articulation of underlying assumptions into explicit models.

## DE-POLITICISATION AND RE-POLITICISATION

In a similar logic, there is a vivid public debate about the de-politicising and re-politicising effects of algorithms. Algorithms have often been criticised as de-politicising due to their 'aura of objectivity and truth' (boyd & Crawford, 2012) and their promise to solve problems of social complexity by the sheer size of data and increased computing power (Kitchin, 2013; Morozov,



it comes to algorithmic regulation, but also an important general matter of public controversy (Hansen & Flyverbom, 2015). Only (a certain degree of) transparency opens up decision-making systems and their inscribed social norms to scrutiny, deliberation and change. Transparency is therefore an important element of democratic legitimacy. It is, however, important to note that the assessment of a given case of algorithmic governance will differ between software developers, the public and supervisory bodies. As already mentioned (cf. section 3), algorithmic governance systems push informational boundaries in comparison to previous governance constellations: they demand formalisation, thus social norms and organisational interests need to be explicated and translated into code – thus *potentially* increasing the share of socially observable information. Yet, *in practice*, algorithmic governance often comes with an actual decrease in socially intelligible and accessible information due to cognitive boundaries (intelligibility of machine learning) and systemic barriers (non-access to algorithms due to trade secrecy, security concerns and privacy protection) (Ananny & Crawford, 2017; Pasquale, 2015; Wachter, Mittelstadt, & Floridi, 2017).

The *degree of automation* matters greatly because the legitimacy of governance regimes relies on the responsibility and accountability of a human decision-maker in her role as a professional (a judge, a doctor, a journalist) and ethical subject. To focus on the degree of automation marks also the choice to problematise the complex interaction within socio-technical systems: algorithmic systems can leave more or less autonomy to human decision-makers. Here, we reduce the gradual scale of involvement to the binary distinction between fully automated systems where decisions are not checked by a human operator, and recommender systems where human operators execute or approve the decisions ('human-in-the-loop') (Christin, 2017; Kroes & Verbeek, 2014; Yeung, 2018).

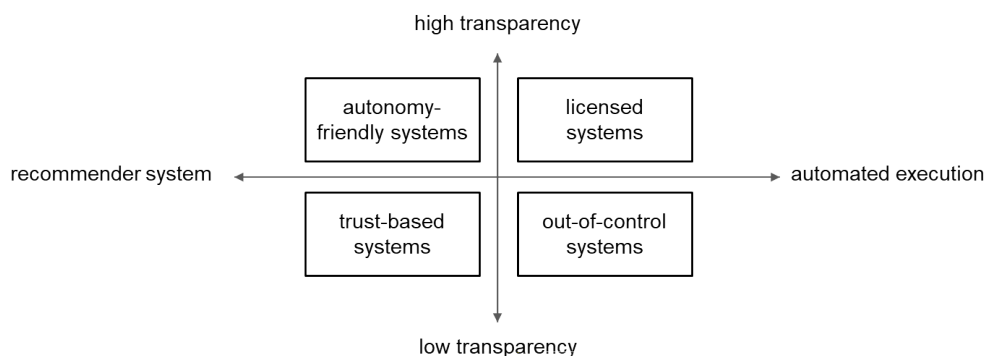


Figure 1: Types of algorithmic governance systems

The combination of both dimensions yields four, in the Weberian sense, ideal-types of algorithmic governance systems with different characteristics: 'autonomy-friendly systems' provide high transparency and leave decisions to humans; 'trust-based systems' operate with low transparency and human-decision-makers; 'licensed systems' combine high transparency with automated execution; and finally 'out-of-control systems' demonstrate low transparency and execute decisions in a fully-automated way.



## 5. ALGORITHMIC GOVERNANCE IN OPERATION: PREDICTIVE POLICING AND AUTOMATED CONTENT MODERATION

The four ideal-types can be found in the full range of sectors and domains that employ algorithmic governance systems today (for recent overviews see AlgorithmWatch and Bertelsmann Stiftung, 2019; Arora, 2019; Dencik, Hintz, Redden, & Warne, 2018; Tactical Tech, 2019). In order to illustrate algorithmic governance both in the public and private sector, and in platforms, we shortly present two prominent and contested cases: automated risk assessment for policing ('predictive policing') is among the most disseminated forms of public algorithmic governance in industrial countries; and automated content moderation on social media platforms belongs to the various ways in which private platforms use algorithmic governance on a global scale. The cases show that algorithmic governance is not one thing, but it takes different forms in different jurisdictions and contexts, and it is shaped by interests, power, and resistance. Algorithmic governance is multiple, contingent and contested.

### PREDICTIVE POLICING

Police authorities employ algorithmic governance by combining and analysing various data sources in order to assess crime risk and prevent crime (i.e., burglary, car theft, violent assault, etc.). This risk analysis addresses either individuals or geographic areas; some systems focus on perpetrators, others on potential victims. The results are predictions of risk that are mobilised to guide policing. Algorithmic governance can be directed towards the behaviour of citizens or of police officers. Typical actions are to assign increased police presence to geographic areas, to surveil potential perpetrators or to warn potential victims.

The degrees of *transparency* need to be assessed from two perspectives: with regard to the public and with regard to the organisation that uses predictive policing. In many cases, data collection, data analysis and governance measures lie in the responsibility of both police agencies and private companies, often in complex constellations (Egbert, 2019). Some projects rely on strictly crime-related data, other projects make use of additional data, such as data about weather, traffic, networks, consumption and online behaviour. In most cases, the software and its basic rationalities are not public. The same is true for the results of the analysis and their interpretation.<sup>3</sup> There is no predictive policing system that makes data and code available to the public, thus most applications in that space are *trust-based systems*. In some cases, such as in the German state of North Rhine-Westphalia, the software has been developed by the police. It is not public, but an *autonomy-friendly system* from the police's perspective. This relatively high degree of opacity is justified by the police with the argument that transparency would allow criminals to 'game the system' and render algorithmic governance ineffective. Opacity, however, hinders evaluations of the social effects of algorithmic governance in policing. Major public concerns are whether predictive policing reinforces illegitimate forms of discrimination, threatens social values and whether it is effective and efficient (Ulbricht, 2018).

With regard to the degree of *automation*, it is noteworthy that in most cases of algorithmic governance for policing the software is still designed as a recommender system: human operators receive a computer generated information or recommendation. It is their responsibility to make the final decision of whether to act and how. However, police officers have complained about the lack of discretion in deciding of where to patrol (Ratcliffe, Taylor, &

Fisher, 2019). Another concern is that police officers might not have the capacity to take an autonomous decision and to overrule the algorithmically generated recommendation (Brayne, 2017), effectively turning predictive policing into *out-of-control* or *licensed* systems of algorithmic governance. The massive number of research and pilot projects in this space indicate that in the near future, the degree of automation in predictive policing and border control governance will increase considerably.

## **AUTOMATED CONTENT MODERATION ON SOCIAL MEDIA PLATFORMS**

Another highly relevant and contested field of algorithmic governance in operation is the (partly) automated moderation and regulation of content on social media platforms. Two developments are driving the turn to AI and algorithms in this field (Gollatz, Beer, & Katzenbach, 2018): (a) The amount of communication and content circulating on these platforms is so massive that it is hard to imagine that human moderators could cope manually with all posts and other material, screening them for compliance with public law and platform rules. As platforms thrive to find solutions that scale with their global outreach, they have strong economic interests to find technical solutions. This is (b) reinforced by the growing political pressure on platforms to tackle issues of hate speech, misinformation and copyright violation on their sites – with regulation partly moving towards immediate platform liability for illegal content (Helberger, Pierson, & Poell, 2019). Thus, platforms develop, test and increasingly put into operation automated systems that aim to identify hate speech, match uploaded content with copyrighted works and tag disinformation campaigns (Gillespie 2018; Duarte, Llanso, & Loup, 2018).

With regard to *transparency*, platforms such as Facebook, YouTube and Twitter have long remained highly secretive about this process, the decision criteria, and the specific technologies and data in use. The increasing politicisation of content moderation, though, has pressured the companies to increase transparency in this space – with limited gains. Today, Facebook, for example, discloses the design of the general moderation process as well as the underlying decision criteria, but remains secretive about specifics of the process and detailed data on removals. <sup>4</sup> YouTube's system for blocking or monetising copyrighted content called ContentID provides a publicly accessible database of registered works. The high-level criteria for blocking content are communicated, yet critics argue that the system massively over-blocks legitimate content and that YouTube remains too secretive and unresponsive about the appeals process, including the exact criteria for delineating legitimate and illegitimate usage of copyrighted content (Erickson & Kretschmer, 2019; Klonick, 2018). The Global Internet Forum to Counter Terrorism (GIFCT), a joint effort by Facebook, Google, Twitter and Microsoft to combat the spread of terrorist content online, hosts a shared, but secretive database of known terrorist images, video, audio, and text.

With regard to *automation*, most systems in content moderation do not operate fully automated but most often flag contested content for human review – despite industry claims around efficiency of AI systems. For example, *Facebook* has technical hate speech classifiers in operation that evaluate apparently every uploaded post and flag items considered illegitimate for further human review. <sup>5</sup> In contrast, ContentID is generally operating fully automated, meaning that decisions are executed without routine human intervention: uploads that match registered content are either blocked, monetised by the rightsholder or tolerated according to the assumed right-holders' provisions. In the case of GIFCT, early press releases emphasised that “matching content will not be automatically removed” (Facebook Newsroom, 2016). However, the response of platforms to major incidents like the shooting in Christchurch, New Zealand, and to propaganda of major terrorist organisations such as ISIS and Al-Qaeda now

seems to indicate that certain matches of the GIFCT are executed and thus blocked automatically, without human moderators in the loop (out-of-control-systems) (Gorwa, Binns, & Katzenbach, 2019).

As these examples show, the binary classification of transparency and automation of a given system is not always easily drawn. Yet, until recently, most of these implementations of algorithmic governance could rightfully be considered out-of-control-systems. The recent political and discursive pressure has certainly pushed the companies towards more transparency, although in our evaluation this still does not qualify them as autonomy-friendly- or licensed"-systems as they still lack meaningful transparency.

## 6. CONCLUSION

The concept of algorithmic governance encapsulates a wide range of sociotechnical practices that order and regulate the social in specific ways ranging from predictive policing to the management of labour and content moderation. It is one benefit of the concept that it brings together these diverse sets of phenomena, discourses, and research fields, and thus contributes to the identification of key controversies and challenges of the emerging digital society. Bias and fairness, transparency and human agency are important issues that are to be addressed whenever algorithmic systems are deeply integrated into organisational processes, irrespective of the sector or specific application. Algorithmic governance has many faces: it is seen as ordering, regulation and behaviour modification, as a form of management, of optimisation and of participation. Depending on the research area it is characterised by inscrutability, the inscription of values and interests, by efficiency and effectiveness, by power asymmetry, by social inclusiveness, new exclusions, competition, responsiveness, participation, co-creation and overload. For most observers, governance becomes more powerful, intrusive and pervasive with algorithmisation and datafication. A different narrative stresses that governance becomes more inclusive, responsive, and allows for more social diversity.

And indeed, algorithmic governance is multiple. It does not follow a purely functional, teleological path thriving for ever-more optimisation. It is rather contingent on its social, economic and political context. The illustrative case studies on predictive policing and content moderation show that algorithmic governance can take very different forms, and it changes constantly – sometimes optimised for business interests, sometimes pressured by regulation and public controversies. The proposed ideal-types of algorithmic governance for means of evaluation constitute one way of assessing these systems against normative standards. We chose transparency and the degree of automation as key criteria, resulting in a spectrum of implementation ranging from out-of-control-systems to autonomy-friendly-systems – other criteria for evaluation could be the types of input data or of decision models. In any case, these structured and integrated ways of thinking about algorithmic governance might help us in the future to assess on more solid grounds which forms of algorithmic governance are legitimate and appropriate for which purpose and under which conditions – and where we might not want any form of algorithmic governance at all.

## REFERENCES

- Algorithm Watch. (2019). OpenSCHUFA: The campaign is over, the problems remain - what we expect from SCHUFA and Minister Barley. Retrieved from <https://openschufa.de/english/>
- Algorithm Watch & Bertelsmann Stiftung. (2019). *Automating Society. Taking Stock of Automated Decision-Making in the EU*. Retrieved from [https://algorithmwatch.org/wp-content/uploads/2019/02/Automating\\_Society\\_Report\\_2019.pdf](https://algorithmwatch.org/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf).
- Ananny, M., & Crawford, K. (2017). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 33(4), 973–989. doi:10.1177/1461444816676645
- Arora, P. (2019). Benign dataveillance? Examining novel data-driven governance systems in India and China. *First Monday*, 24(4). doi:10.5210/fm.v24i4.9840
- Avery, R. B., Brevoort, K. P., & Canner, G. (2012). Does Credit Scoring Produce a Disparate Impact? *Real Estate Economics*, 40(3), S65–S114. doi:10.1111/j.1540-6229.2012.00348.x
- Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104(3), 671–732. doi:10.15779/Z38BG31
- Beer, D. (2019). *The Data Gaze: Capitalism, Power and Perception*. SAGE Publications.
- Bennett, C. J. (2017). Voter databases, micro-targeting, and data protection law: Can political parties campaign in Europe as they do in North America? *International Data Privacy Law*, 6(4), 261–275. doi:10.1093/idpl/ipw021
- Berry, D. M. (2011). *The philosophy of software: code and mediation in the digital age*. Basingstoke, Hampshire; New York: Palgrave Macmillan. doi:10.1057/9780230306479
- Binns, R. Veale, M., Van Kleek, & Shadbolt, M. (2017) Like trainer, like bot? Inheritance of bias in algorithmic content moderation. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social Informatics*, (pp. 405–415). doi: 10.1007/978-3-319-67256-4\_32
- Bijker, W. E., & Law, J. (Eds.). (1992). *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: The MIT Press.
- Black, J. (2001). Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World. *Current Legal Problems*, 54(1), 103–146. doi:10.1093/clp/54.1.103
- Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society*, 18(5), 524–538. doi:10.1080/1369118X.2015.1008542
- Boulianne, S., & Theocharis, Y. (2018). Young People, Digital Media, and Engagement: A Meta-Analysis of Research. *Social Science Computer Review*, 19(1). doi:10.1177/0894439318814190
- boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–667. doi:10.1080/1369118X.2012.678878
- boyd, d., & Barocas, S. (2017). Engaging the Ethics of Data Science in Practice. *Communications of the ACM*, 60(11), 23–25. doi:10.1145/3144172

- Brayne, S. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review*, 82(5), 977–1008. doi:10.1177/0003122417725865
- Brevoort, K. P., Grimm, P., & Kambara, M. (2015). *Data Point: Credit Invisibles* [Research Report]. Washington DC: Consumer Financial Protection Bureau. Retrieved from <https://www.consumerfinance.gov/data-research/research-reports/data-point-credit-invisibles/>
- Caplan, R., & boyd, d. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society*, 5(1). doi:10.1177/2053951718757253
- Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, 4(2). doi:10.1177/2053951717718855
- Couldry, N., & Langer, A. I. (2005). Media Consumption and Public Connection: Toward a Typology of the Dispersed Citizen. *The Communication Review*, 8(2). 237–257. doi:10.1080/10714420590953325
- Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018). *Data scores as Governance: Investigating uses of citizen scoring in public services project report* [Project Report]. Cardiff University. Retrieved from Open Society Foundations website: <http://orca.cf.ac.uk/117517/>
- DeVito, M. A. (2017). From Editors to Algorithms. *Digital Journalism*, 5(6), 753–773. doi:10.1080/21670811.2016.1178592
- Duarte, N., Llanso, E., & Loup, A. (2018). *Mixed Messages? The Limits of Automated Social Media Content Analysis*. Report. Washington DC: Center for Democracy & Technology.
- Egbert, S. (2019). Predictive Policing and the Platformization of Police Work. *Surveillance & Society*, 17(1/2), 83–88. doi:10.24908/ss.v17i1/2.12920
- Ellul, J. (1964). *The technological society*. New York: Alfred A. Knopf.
- Erickson, K., & Kretschmer, M. (2018). “This Video is Unavailable”: Analyzing Copyright Takedown of User-Generated Content on YouTube. *JIPITEC*, 9(1). Retrieved from <http://www.jipitec.eu/issues/jipitec-9-1-2018/4680>
- Eyert, F., Irgmaier, F., & Ulbricht, L. (2018). Algorithmic social ordering: Towards a conceptual framework. In G. Getzinger (Ed.), *Critical Issues in Science, Technology and Society Studies* (pp. 48–57). Retrieved from <https://conference.aau.at/event/137/page/6>
- Facebook. (2016). Partnering to Help Curb Spread of Online Terrorist Content [Blog post]. Retrieved from Facebook Newsroom website <https://newsroom.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content>
- Fourcade, M., & Healy, K. (2016). Seeing like a market. *Socio-Economic Review*, 15(1), 9–29. doi:10.1093/ser/mww033
- Fourcade, M., & Healy, K. (2017). Categories All the Way Down. *Historical Social Research*, 42(1), 286–296. doi:10.12759/hsr.42.2017.1.286-296
- Fuller, M. (2008). *Software Studies: A Lexicon*. Cambridge, MA: The MIT Press.

doi:10.7551/mitpress/9780262062749.001.0001

Gandy, O. H. (2010). Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems. *Ethics and Information Technology*, 12(1), 29–42. doi:10.1007/s10676-009-9198-6

Gerber, C., & Krzywdzinski, M. (2019). Brave New Digital Work? New Forms of Performance Control in Crowdfunder. In: Steve P. Vallas & A. Kovalainen (Eds.), *Work and Labor in the Digital Age* (pp. 48–57). Bingley: Emerald Publishing. doi: 10.1108/S0277-283320190000033008

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven: Yale University Press.

Gillespie, T. (2014). The relevance of algorithms. In: T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies. Essays on Communication, Materiality, and Society* (pp. 167–193). Cambridge, MA: The MIT Press.

Gollatz, K., Beer, F., & Katzenbach, C. (2018). *The Turn to Artificial Intelligence in Governing Communication Online* [Workshop Report] Berlin: Alexander von Humboldt Institute for Internet and Society. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-59528-6>

Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, 22(6). doi:10.1080/1369118X.2019.1573914

Gorwa, R., Binns, R., & Katzenbach, C. (2019). Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance. *Big Data & Society*, forthcoming.

Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, 55(4), 1143–1185.

Hansen, H. K., & Flyverbom, M. (2014). The politics of transparency and the calibration of knowledge in the digital age. *Organization*, 22(6), 872–889. doi:10.1177/1350508414522315

Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, 34(1), 1–14. doi:10.1080/01972243.2017.1391913

Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology*. Cheltenham: Edward Elgar

Hobbes, T. (1909). *Hobbes's Leviathan: reprinted from the edition of 1651*. Oxford: Clarendon Press. Retrieved from <https://archive.org/details/hobbessleviathan00hobbuoft>

Hofmann, J., Katzenbach, C., & Gollatz, K. (2016). Between coordination and regulation: Finding the governance in Internet governance. *New Media & Society*, 19(9). doi:10.1177/1461444816639975

Introna, L. D. (2016). Algorithms, Governance, and Governmentality: On Governing Academic Writing. *Science, Technology, & Human Values*, 41(1), 17–49. doi:10.1177/0162243915587360

Jarke, J., & Gerhard, U. (2018). Using Probes for Sharing (Tacit) Knowing in Participatory Design: Facilitating Perspective Making and Perspective Taking. *i-com*, 17(2), 137–152. doi:10.1515/icom-2018-0014

Danaher, J. Hogan, M. J., Noone, C., Kennedy, R., Behan, B., de Paor, A. ... Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, 4(2). doi: 10.1177/2053951717726554

Just, N., & Latzer, M. (2016). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, 39(2), 238–258. doi:10.1177/0163443716643157

Katzenbach, C. (2017). *Die Regeln digitaler Kommunikation. Governance zwischen Norm, Diskurs und Technik* [The rules of digital communication. Governance between norm, discourse, and technology]. Wiesbaden: Springer VS. doi:10.1007/978-3-658-19337-9

Katzenbach, C. (2012). Technologies as Institutions: Rethinking the Role of Technology in Media Governance Constellations. In N. Just & M. Puppis (Eds.), *Trends in Communication Policy Research: New Theories, New Methods, New Subjects* (pp. 117–138). Bristol: Intellect.

Kelty, C. M. (2017). Too Much Democracy in All the Wrong Places: Toward a Grammar of Participation. *Current Anthropology*, 58(S15), S77–S90. doi:10.1086/688705

Kitchin, R., & Dodge, M. (2011). *Code/Space: Software in Everyday Life*. Cambridge, MA: The MIT Press.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267. doi:10.1177/2043820613513388

Kitchin, R. (2016). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29. doi:10.1080/1369118X.2016.1154087

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, 131, 1598–1670. Retrieved from <https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/>

König, P. D. (2019). Dissecting the Algorithmic Leviathan. On the Socio-Political Anatomy of Algorithmic Governance. *Philosophy & Technology*. doi:10.1007/s13347-019-00363-w

Kroes, P., & Verbeek, P.-P. (2014). Introduction: The Moral Status of Technical Artefacts. In P. Kroes & P.-P. Verbeek (Eds.), *Philosophy of Engineering and Technology. The Moral Status of Technical Artefacts* (pp. 1–9). Dordrecht: Springer. doi:10.1007/978-94-007-7914-3\_1

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford; New York: Oxford University Press.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50–80. doi:10.1518/hfes.46.1.50\_30392

- Lee, C. P., Poltrock, S., Barkhuus, L., Borges, M., & Kellogg, W. (Eds.) 2017. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*. New York: ACM Press.
- Lupton, D. (2016). Personal Data Practices in the Age of Lively Data. In J. Daniels, K. Gregory, and T. M. Cottom (Eds.). *Digital sociologies*. Bristol; Chicago: Policy Press (pp. 339–354).
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, 1(2). doi:10.1177/2053951714541861
- MacKenzie, D. A. (2006). *An engine, not a camera: How financial models shape markets*. Cambridge, MA: The MIT Press.
- Mejias, U. & Couldry, N. (2019) Datafication. *Internet Policy Review*, 8(4). doi:10.14763/2019.4.1428
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). doi:10.1177/2053951716679679
- Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. New York: Public Affairs.
- Musiani, F. (2013). Governance by algorithms. *Internet Policy Review*, 2(3). doi:10.14763/2013.3.188.
- Müller-Birn, C., Herbsleb, J., & Dobusch, L. (2013). Work-to-Rule: The Emergence of Algorithmic Governance in Wikipedia. *Proceedings of the 6th International Conference on Communities and Technologies*, 80–89. doi:10.1145/2482991.2482999
- Napoli, P. M. (2013). *The Algorithm as Institution: Toward a Theoretical Framework for Automated Media Production and Consumption* [Working Paper No. 26]. New York: McGannon Center, Fordham University. Retrieved from [https://fordham.bepress.com/mcgannon\\_working\\_papers/26](https://fordham.bepress.com/mcgannon_working_papers/26)
- Neyland, D. & Möllers, N. (2017). Algorithmic IF ... THEN rules and the conditions and consequences of power. *Information, Communication & Society*, 20(1), 45–62. doi:10.1080/1369118X.2016.1156141
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
- OECD. (2015). *Data-Driven Innovation: Big Data for Growth and Well-Being*. Paris: OECD Publishing. doi:10.1787/9789264229358-en .
- Ong, W. J. (1982). *Orality and literacy □: the technologizing of the word*. London; New York: Methuen.
- O'Reilly, T. (2013). Open Data and Algorithmic Regulation. In B. Goldstein & L. Dyson (Eds.), *Beyond transparency: Open data and the future of civic innovation* (pp. 289–300). San Francisco: Code for America Press.
- Orwat, C., Raabe, O., Buchmann, E., Anandasivam, A., Freytag, J.-C., Helberger, N., ... Werle, R. (2010). Software als Institution und ihre Gestaltbarkeit [Software as institution and its



designability]. *Informatik Spektrum*, 33(6), 626–633. doi:10.1007/s00287-009-0404-z

Parasuraman, R., Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors*, 52(3), 381–410. doi:10.1177/0018720810376055

Pasquale, F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Passig, K. (2017, November 23). Fünfzig Jahre Black Box [Fifty years black box]. *Merkur*. Retrieved from <https://www.merkur-zeitschrift.de/2017/11/23/fuenfzig-jahre-black-box/>

Ratcliffe, J. H., Taylor, R. B., & Fisher, R. (2019). Conflicts and congruencies between predictive policing and the patrol officer's craft. *Policing and Society*. doi:10.1080/10439463.2019.1577844

Rieder, G., & Simon, J. (2016). Datatrust: Or, The Political Quest for Numerical Evidence and the Epistemologies of Big Data. *Big Data & Society*, 3(1). doi:10.1177/2053951716649398

Rosenblat, A. (2018). *Uberland: How algorithms are rewriting the rules of work*. Oakland: University of California Press.

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. Valencia: Association for Computational Linguistics. doi:10.18653/v1/W17-1101

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2).doi:10.1177/2053951717738104.

Schrape, J.-F. (2019). The Promise of Technological Decentralization. A Brief Reconstruction. *Society*, 56(1), 31–37. doi:10.1007/s12115-018-00321-w

Schuepp, W. (2015, September 14). Achtung, bei Ihnen droht ein Einbruch [Attention, a burglary at yours is imminent]. *Tagesanzeiger Zürich*.

Suzor, N. P., West, S. M., Quodling, A., & York, J. (2019). What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation. *International Journal of Communication*, 13, 1526–1543. Retrieved from <https://ijoc.org/index.php/ijoc/article/view/9736>

Tactical Tech. (2019). Personal Data: Political Persuasion. Inside the Influence Industry. How it works. Retrieved from [https://ourdataourselves.tacticaltech.org/media/Personal-Data-Political-Persuasion-How-it-works\\_print-friendly.pdf](https://ourdataourselves.tacticaltech.org/media/Personal-Data-Political-Persuasion-How-it-works_print-friendly.pdf)

Ulbricht, L. (2018). When big data meet securitization. Algorithmic regulation with passenger name records. *European Journal for Security Research*, 3(2), 139–161. doi:10.1007/s41125-018-0030-3

Veale, M., & Brass, I. (2019). Administration by Algorithm? Public Management meets Public Sector Machine Learning. In K. Yeung & M. Lodge (Eds.), *Algorithmic Regulation* (pp. 121–149). Oxford: Oxford University Press.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated

Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76–99. doi: [10.1093/idpl/ipy005](https://doi.org/10.1093/idpl/ipy005)

Waseem, Z., & Hovy, D. (2016). Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. *Proceedings of the NAACL Student Research Workshop*, 88–93. doi:[10.18653/v1/N16-2013](https://doi.org/10.18653/v1/N16-2013).

Wiener, N. (1948). *Cybernetics: or control and communication in the animal and the machine*. Cambridge, MA: The MIT Press.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1), 121–136. Retrieved from <http://www.jstor.org/stable/20024652>

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment & Society: a Journal of the British Sociological Association*, 33(1), 56–75. doi:[10.1177/0950017018785616](https://doi.org/10.1177/0950017018785616)

Yeung, K. (2017). „Hypernudge“: Big Data as a mode of regulation by design. *Information Communication & Society*, 20(1), 118–136. doi:[10.1080/1369118X.2016.1186713](https://doi.org/10.1080/1369118X.2016.1186713)

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505–523. doi:[10.1111/rego.12158](https://doi.org/10.1111/rego.12158)

Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values*, 41(1), 3–16. doi:[10.1177/0162243915608948](https://doi.org/10.1177/0162243915608948)

Zarsky, T. Z. (2014). Understanding Discrimination in the Scored Society. *Washington Law Review*, 89(4). Retrieved from <https://digitalcommons.law.uw.edu/wlr/vol89/iss4/10/>

## FOOTNOTES

1. The context of the study are governance mechanisms in Wikipedia content production. The authors define social governance as the coordination that relies upon interpersonal communication and algorithmic governance as the coordination based on rules that are executed by algorithms (mostly bots) (Müller-Born et al., 2013, p. 3).
2. Other typologies are too granular for the generalising aim of this article and/or focus on sub-fields of algorithmic governance (Danaher et al., 2017), such as algorithmic selection (Just & Latzer, 2016), content moderation (Gorwa, Binns, & Katzenbach, 2019), and modes of regulation (Eyert, Irgmaier, & Ulbricht, 2018; Yeung, 2018).
3. An exception is the canton of Aargau in Switzerland that publishes its risk map (Schuepp, 2015).
4. Cf. Facebook’ Transparency Report for an example, <https://transparency.facebook.com>, and Suzor et al., 2019, for a critique.
5. Cf. Facebook Newsroom, ”Using Technology to Remove the Bad Stuff Before It’s Even Reported”, <https://perma.cc/VN5P-7VNU>.