Mapping the Public Debate on Ethical Concerns: Algorithms in Mainstream Media

Journal:	*Journal of Information, Communication & Ethics in Society*
Manuscript ID	JICES-04-2019-0039.R2
Manuscript Type:	Journal Paper
Keywords:	Ethics, Information ethics, Mass media, Algorithm, Artificial intelligence

Mapping the Public Debate on Ethical Concerns: Algorithms in Mainstream Media

Algorithms are in the mainstream media news on an almost daily basis. Their context is invariably artificial intelligence and machine learning decision making. In media articles, algorithms are described as powerful, autonomous actors that have a capability of producing actions that have consequences. Despite a tendency to deification, the prevailing critique of algorithms focusses on ethical concerns raised by decisions resulting from algorithmic processing. However, this paper proposes that the ethical concerns discussed are limited in scope and it is not clear which concerns dominate the debate. The research presented in this paper contributes the first systematic mapping study of articles appearing in leading UK national papers from the perspective of widely accepted ethical concerns. The UK context is important because of UK public policy initiatives around artificial intelligence. In academic literature, the key ethical concerns have been well documented and numerous models have been developed. To review the media content from the perspective of ethical concerns, this paper uses the synthesised conceptual map of ethical concerns developed by Mittelstad et. al. Given the widespread use of that framework, as evident through citations, this paper's contribution is also an important illustration and experiment using that conceptual map.

Keywords: Ethics, Algorithms, Machine Learning, Artificial Intelligence, Bias, Governance

1 Introduction

'Algorithms' are in the mainstream media (the so-called fourth estate (Newman et al., 2012)) news. Their description follows a predictable, almost algorithmically scripted drama of two acts that invariably conjures a deification of the algorithm (Bogost, 2015). In this seductive drama elucidated with much eloquence by Ziewitz (2016) we are first introduced to an algorithm as a powerful autonomous actor with a capability of producing actions that have consequence. Examples of consequences may be derived

from medicine, social media, policing or even employment. In the last act of this drama the autonomous, agency like characteristics of algorithms are enhanced further, through descriptions of bias, opacity, mistakes and corporate misdemeanour. The drama concludes by pointing out a need for governance, regulation and a new logic for ethics. In this manner, the public discourse in the mainstream media about algorithms unfolds. Recall one week in May 2018 in the UK, where the BBC reported on Amnesty International exposing a Metropolitan Police database as racially discriminatory; a facial recognition system used by South Wales police throwing up false positives and the data profiling company Cambridge Analytica under scrutiny by the Information Commissioner (Andrews, 2018).

One newspaper article demonstrates this drama in an almost perfect fashion. Reported in the Financial Times, a new artificial intelligence-based software tool called 'Annie' aims to relocate migrants to maximise their job prospects through a matching algorithm (Warrell, 2018). The system outperforms humans with early results showing an increase in the refugee employment rate. (Act 1: the setting and the consequence). As the article continues, it introduces fear of job losses : "People immediately think, 'You just want to get rid of all of us?'" and questions of unintended outcomes, bias, transparency become a dominant theme: "there can be a lack of transparency about how final conclusions are drawn". The final act predictably, identifies governance concerns such as the refugees having the right to know whether the future pattern of their lives is being dictated by a human or a machine.

The benefits of algorithms followed by the superficial relaying of ethical concerns and the pleas for governance is consistent with Gitlin's definition of framing as 'persistent patterns of cognition, interpretation and presentation of selection, emphasis and exclusion by which symbol handlers routinely organise discourse' (Gitlin,

1980, p7, cited in De Vreese (2005)). Characteristics such as identifiable conceptual linguistic features, a recognisable journalistic style, and reliable distinguishing from other types of frames are all visible (De Vreese, 2005).

Thus, the discussion on algorithms permeates the mainstream media. What is uncertain though, is how rich and informed is this debate and what is the nature of the debate? What are the primary concerns? Which moral values are at the core? Does security trump transparency? Unpicking the complexity of how ethics of usage of algorithms is examined is central to this overall discussion and is important for at least three reasons. Firstly, the debate is closely engaged with public policy. For example, the UK Industrial Strategy, places artificial intelligence as a core technology for addressing grand challenges (ageing population and transport infrastructure) (Government, 2018). Secondly, the Select Committee on Artificial Intelligence (AI) reported on key concerns of AI such as the explainability of decisions made through AI algorithms (House of Lords Select Committee, 2018). Thirdly, a joint report by the Royal Society and the British Academy detailed the necessary governance principles and requirements in a future data driven society (Royal Society, 2017). Hence, this paper sets out to explore the public debate from a United Kingdom (UK) perspective. The public debate, enacted in mainstream media, is important because as Diakopoulos (2013, p.2) notes: 'What we generally lack as a public is clarity about how algorithms exercise their power over us.' And further, 'new forms of algorithmic power … are reshaping how social and economic systems work.' (Kitchin, 2017).

Algorithms and their relevance in modern society has been examined in detail. Gillespie (2014) provides an analytical conceptual map of how algorithms are a key feature of the information ecosystem and the political ramifications arising from the role of algorithms. Such impacts include the choices made about what data is excluded from

an index, how algorithms determine what is relevant to us and as a consequence, how users reshape their practices. Collectively these impacts lead to the creation of a new knowledge logic that operates on presumptions and adjustments of norms and practices. With the construction of a new knowledge logic comes the need for additional governance in the use of algorithms in areas such as explainability, interpretability, and ethical auditing (Cath, 2018). Algorithms cannot be divorced from the data on which they are required to operate. Data for example is the basis of 'hypernudge' technologies and consequently requires its own regulation and governance (Yeung, 2017). Principles of data governance and use have been proposed in a joint report by the Royal Society and the British Academy that aim to:

- protect individual and collective rights and interests
- ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively
- seek out good practices and learn from success and failure
- enhance existing democratic governance (Royal Society, 2017).

Of note in this report, is the framework that details the multi-dimensional tradeoffs that explore social and ethical tensions that require choices from society if data-enabled technologies are to be widely adopted.

Much academic research has been published exploring the ethical dimensions of algorithms. Various frameworks for conceptualising the key ethical concerns have been reported (for example, Marda, 2018, Ziewitz, 2016, Pereira, 2018). In 2016, Mittelstadt et. al. (2016) conducted a systematic review in an attempt to map the ethical problems prompted by algorithmic decision making. This paper does not therefore present a background literature to ethics studies of algorithms. Their proposal comprises a conceptual map that consolidates themes emerging from the literature to a unifying

framework that can serve as an "organising structure based on how algorithms operate" (Mittelstadt, 2016) and can structure future discussion of how algorithms deal with ethical issues. Importantly, they recognise the prevalence of "algorithms in the news" and accordingly, they aim to map the conceptual dimensions of the ethics of usage of algorithms interpreted along public discourse lines. The framework contributes to more general approach that aims to translate principles into practices especially for algorithms embedded in machine learning technologies (Morley et al, 2019). Hence their framework provides a useful theoretical model for the analysis conducted in this paper. The framework is described in further detail in section 2.

To do this analysis, this paper reviews articles from UK newspaper articles from leading newspapers in the UK using the systematic mapping study guidelines developed in (Kitchenham, 2007, Peterson et al, 2015). The goal is to categorise contributions from UK newspapers along the ethical dimensions and using the conceptual tool described by Mittelstadt et al.

The remainder of the paper proceeds as follows. Section 2 outlines the theoretical framework utilised for the systematic review. Section 3 presents the research questions and the overarching research framework. Section 4 presents the results and analysis for the research questions. Section 5 outlines the principal limitations of this work and finally Section 6 concludes with a summary of the work and directions for future research.

2 Theoretical Framework

Before it is possible to map the ethics debate around algorithms in the newspapers, it is useful to have key terms defined. Following Knuth's characterisation of the key properties of an algorithm, there is broad agreement in the computer science discipline that an algorithm is a set of precisely defined steps for processing an initial set of given

objects (inputs) in order to transform them to a set of outputs such that the sequence of steps terminates in a finite time. As Mittelstadt et. al. points out, algorithms are only of interest in the public media if there is an implementation in a specific technology and the application of the technology is configured for a task (Mittelstadt et al, 2016, pg 2).

The algorithms that are in the news are also mostly concerned with those that make decisions based on complex rules whose detailed workings are difficult to predict or where the actionable insight cannot be explained. Like Mittelstadt et al, algorithms that implement mundane tasks such as "search and replace" are not of interest (Mittelstadt et al, 2016, p3). Nor are abstract descriptions of algorithms of interest. Also interesting is how the notion of algorithm is separated from its host system.

The conceptual mapping tool offered by Mittelstadt et al. is an organising structure for scaffolding discussions of ethical issues. They propose that ethical issues raised by algorithms making decisions, (and implicitly, with little or no human involvement[1]), are categorised into six types.

Theories in general can offer routes to analysis and prediction, explanation, prediction and prescription (a description of a method) (Gregor, 2006). Thus the six types of ethical concerns can be seen as a type of theory for classifying dimensions or characteristics and its primary use is for analysis. The six concerns arise from the use of algorithms, how algorithms process data to produce evidence that might motivate an action. Where an action might lead to a potential failure responsibility or accountability means that traceability is also of concern. The six are presented below briefly:

[1] Mittelstadt et al note that "algorithms augment or replace analysis and decision making by humans" (page 3).

Inconclusive evidence: Algorithms in the media draw conclusions from the data that is processed using inferential statistics and/or machine learning techniques. The conclusions are probable and there is uncertainty. Motivation of an action based on probable knowledge is thus an epistemic limitation.

Inscrutable evidence: A conclusion based on processed data creates a reasonable expectation that the connection between the data and conclusion should be accessible, intelligible and open to critique. Further, evaluations of the process leading to a conclusion, and whether the evidence produced is misguided, are observer-dependent.

Misguided Evidence: Conclusions and actions arising from them[2] can only be as reliable as the data on which they are based. Informally, computer science traditionally treats this is `garbage in, garbage out'.

Unfair outcomes: Action arising from a conclusion based on processed data should be assessed to criteria or ethical principles broadly understood as "fair" and not discriminatory even if the action is conclusive, scrutable and well-founded.

Transformative effects: Algorithms and their prevalence affect how we conceptualise the world in new ways perhaps even modifying people's behaviour and challenging their autonomy. For example, our book buying habits might change though personalisation and nudge technologies[3]. Personalisation also reduces the diversity of

[2] Actions arising from conclusions is added here as a supplementary claim following an earlier statement where Mittlestad et al note that their conceptual map takes into account a given outcome (henceforth conclusion) and the conclusion can then trigger an action. (Mittlestad et al, 2016, page 4).

[3] Personalisation techniques are inherently paradoxical. On one hand, providing only relevant information can aid decision making by reducing information overload. Yet the provision of

information encountered, a pre-condition for autonomy (Van den Hoven and Rooksby, E, 2008).

Traceability: Algorithms as software artefacts create further ethical challenges through design and availability of new technologies as well as complexity associated with the manipulation of large volumes of personal and other data. An implication is that the harm caused by software is difficult to debug. The nature of modern distributed computing techniques, dispersed development activities also means it is "rarely straight forward to to identify who should be held responsible for the harm caused" (Mittelstadt, 2016, page 5).

A critique of the Mittelstadt et al. paper (Mittelstadt et al., 2016) suggests these categories provide the bare conceptual structure for categorising and distinguishing epistemic, ethical and traceability concerns in descriptions of ethical problems related to the use of (implementations) of algorithms. Significantly, the conceptual map offered by Mittelstadt et al. (Figure 1 in their paper) does not identify or formally state potential relationships that might exist between these type structures. Possibly, this could be based on the approach taken in the systematic review but details of the review step are not provided and reading of the text does not suggest that it follows along the lines of accepted models of systematic review processes. However, for the purposes of this mapping study, the type structure is sufficient with which to explore mapping of the public debate of algorithms. The next section explains how the theoretical framework is applied.

---

the information is subjective and can be coercive in that only institutionally preferred actions are made possible.

3 Research Design

Literature reviews in Information Systems (IS) academic research have tended to be *ad hoc* or systemised using either a systematic mapping study approach or a systematic literature review process (Kitchenam, 2004) Mapping studies or *scoping studies* are used to provide an overview of a research area by a process of classification and counting of the outputs (Peterson, 2008, Peterson et al. 2015). Results and analysis are often used to provide a visual map of subject areas and is often coarse grained. More often than not, mapping studies aim to uncover research trends and to provide simple routes to comparison. There are now numerous examples in the literature and the recent mapping study of twenty-eight years of component-based software engineering is cited as an example for the purposes of illustration of how mapping studies reveal research trends (Vale et al., 2016).

In contrast to mapping studies, systematic literature review (SLR) approaches are detailed studies that drill down on a very specific area with the aim of aggregating evidence and are strongly focussed on methods and results. They lend themselves to meta studies of the data, often with the aim of establishing a hypothesis. In a software engineering context, SLRs are more appropriate for meta-analysis of studies that have either quantitative or empirical results (Kitchenam, 2007). For example, Walia and Carver use an SLR approach to review, identify and classify software requirements errors (Walia and Carver, 2009).

Given these differences and the main aim of establishing an overarching map of the ethical concerns of algorithms being debated in the media, it is more appropriate to use a mapping study approach.

3.1 Data sources

Two comprehensive data bases were candidates for use (LexisNexis Academic and Factiva) to conduct this media analysis. The former, however had limited functionality from the licensing arrangements available at the author's institution and was primarily used to provide a cross-check facility to ensure that the potentially relevant media were correctly located. Factiva provides a large online collection of newspaper and other media sources such as television and radio transcripts, web blog content and other sources. However, this research focussed only on mainstream newspapers in the UK. Previous social media research has demonstrated that traditional fourth estate media are key brokers of news across all types of social and news networks (Barn et al., 2017).

3.2 Search strategy

The Factiva electronic newspaper database was searched for relevant media reports that were published between the dates 21 September 2016 to 21st September 2018. Sources searched were the category Top UK newspapers (both print and online where available). These papers are: Financial Times, The Times (UK), The Daily Telegraph (UK), Daily Mail, The Independent, thesun.co.uk, express.co.uk, The Sunday Telegraph, The Sunday Times and The Guardian (including The Observer. A deliberate focus on the UK was chosen because of several key policy initiatives being enacted by the UK Government. These included the now established Turing Institute (Turing.ac.uk, 2019) the planned Centre for Data Ethics (HM Government, 2019) as well as the 2017 UK industrial strategy with its focus on Artificial Intelligence (HM Government, 2018).

The keywords used were kept deliberately simple to maximise the corpus of data. Keywords used were: "Algorithms AND Ethics". All newspaper articles (print or

online) found were included unless the exclusion criteria identified a duplicate, or the article was not about the type of algorithm described in section 2 earlier.

Figure 1 illustrates the initial search, the screening of the corpus of media reports and the total number of media reports available for analysis.

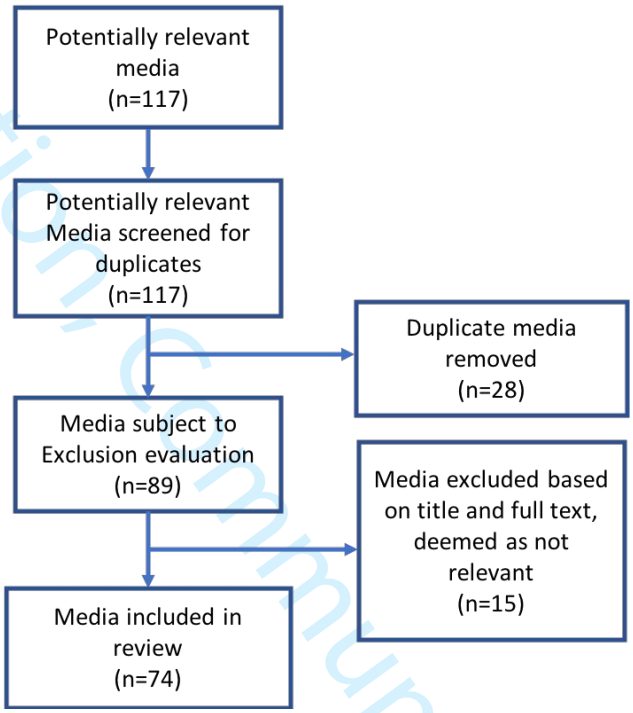

Figure 1: Screening flow diagram

3.3 Themes

Processing of the 74 screened media reports was carried out by close reading of text (Charon, 2017). The media reports are available from the author. The individual articles were allocated to ethical concern categories described in section 2. These categories described by Mittelstadt et. al. help in understanding and formulating a response to the central research question:

> What is the nature of the public debate around ethical concerns of the use of algorithms?

Through assignment to the ethical concern categories, it is possible to develop a systematic map of the key concerns being discussed in mainstream media as well as understanding the nature and distribution of the concerns.

Other data / categories were also collected but they are not reported in this paper as they relate to other research questions which are part of a separate study. Additionally, data generated by Factiva was also available but these data were based on the initial corpus of 117, i.e. before any screening had taken place.

Figure. 2a

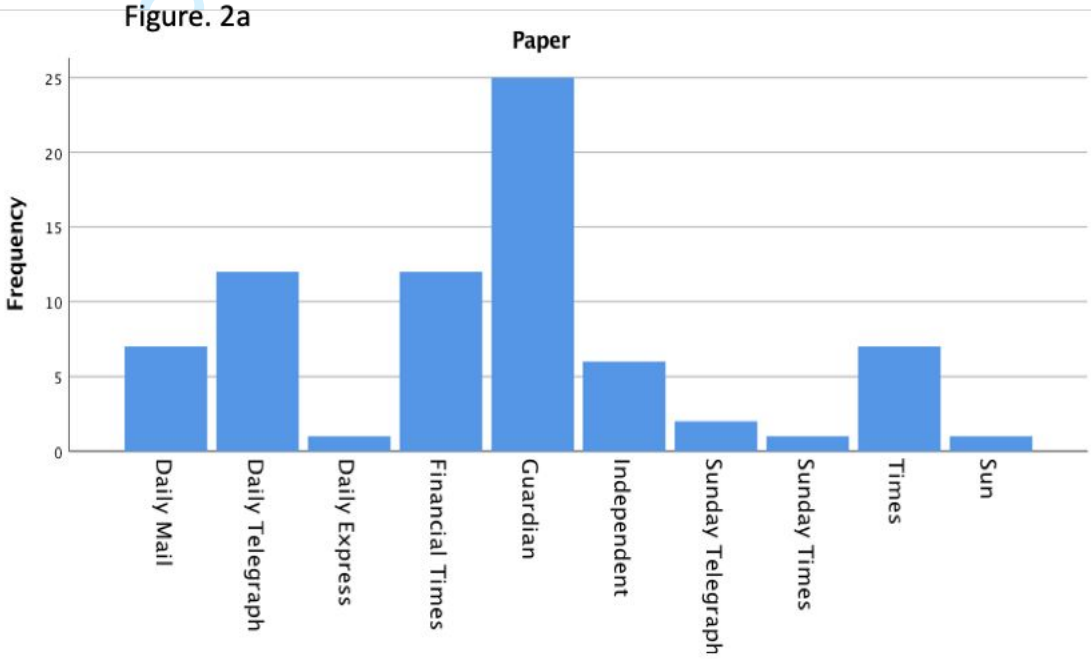

Figure. 2b

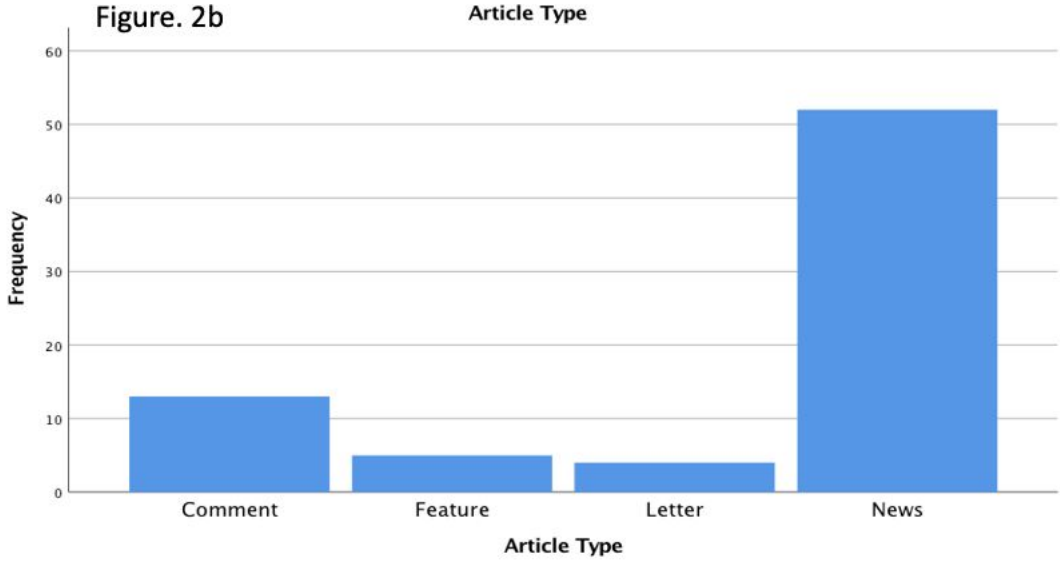

Figure 2: Overall summary

4 Results and Discussion

The media search in Factiva selected 117 reports that mentioned Algorithms and ethics. Following the screening for duplicates and removal of excluded items, 74 items were cleaned and exported to SPSS (IBM SPSS Statistics for Macintosh, Version 25.0) for descriptive analysis of the categorisations of ethical concerns. Results from the descriptive analysis are reported here in line with systematic mapping study requirements.

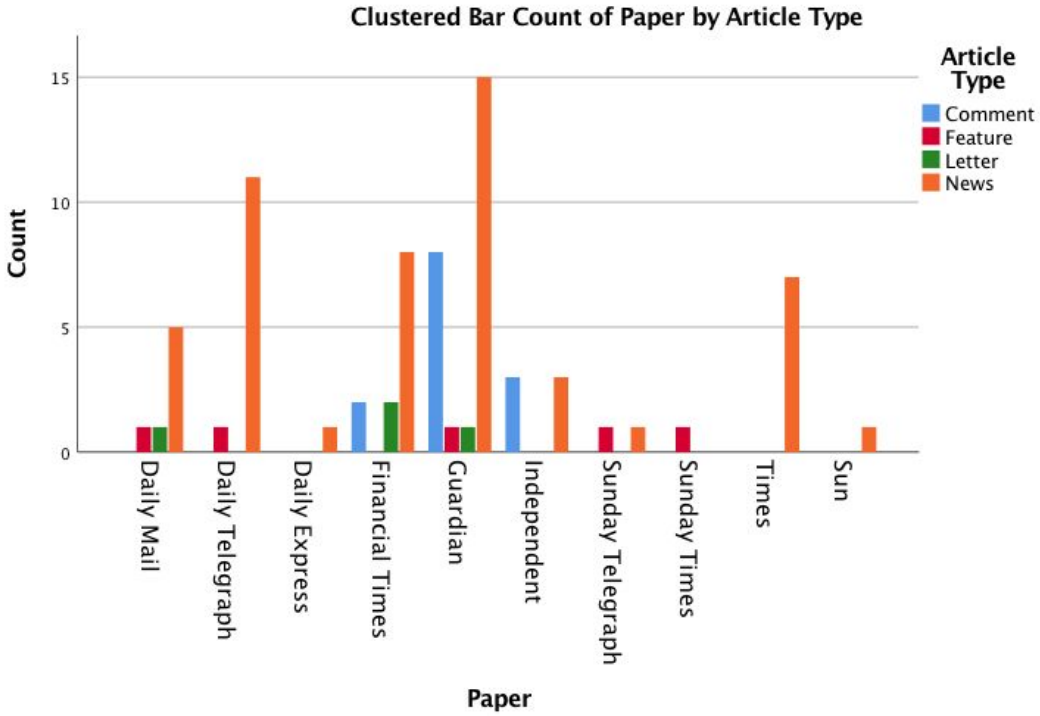

Figure 3: Types of Articles

4.1 Overall picture

Straightforward descriptive statistics were computed for the key variables, the Newspaper (The Guardian, Financial Times, etc.) and the type of article. During the close reading of the articles, various types of article were identified and are described

briefly here.

- Comment (termed opinion in some newspapers), pieces written by experts in the field or those with an interest in the domain.
- Feature articles were longer pieces going into depth into key issues and were often based on recent books.
- Letters were identified as a usefully separate type given that a letter could be treated as an opinion. They were kept separate as they could represent the lay reader.
- News articles are regular articles written by journalists and reported on the use of algorithms at large.

Figure 2 provides a summary of the overall picture of the articles. The nature of newspaper media in the UK does not present any surprises in the overview results given the general position of the newspapers with respect to political leanings (Smith, 2017). The (relatively libertarian) Guardian (ibid) published the most articles and given its approach to attracting commentary and opinion from outside normal journalism also had the most comment pieces (Figure 4B).

4.2 Ethical concerns

The conceptual map of ethical concerns proposed by Mittelstadt et. al provides a useful basis for structuring discussions. Following the close reading of the 74 media texts, the texts were assigned to one or more of the ethical concerns. The results in Figure 4 present the frequency count of the various concerns along two principal axes. First, (in Figure 4C) we can see that the two dominant themes are Unfair Outcomes (42%)

(essentially a discussion about potential bias and discrimination), and secondly, Inscrutable Evidence (27%), that is concerns that data and any conclusion are accessible, intelligible and open to critique such that algorithms are not behaving as black boxes. Other concerns were harder to delineate perhaps because unpicking the complexity of say Misguided Evidence or Transformative Effects is not straight forward.

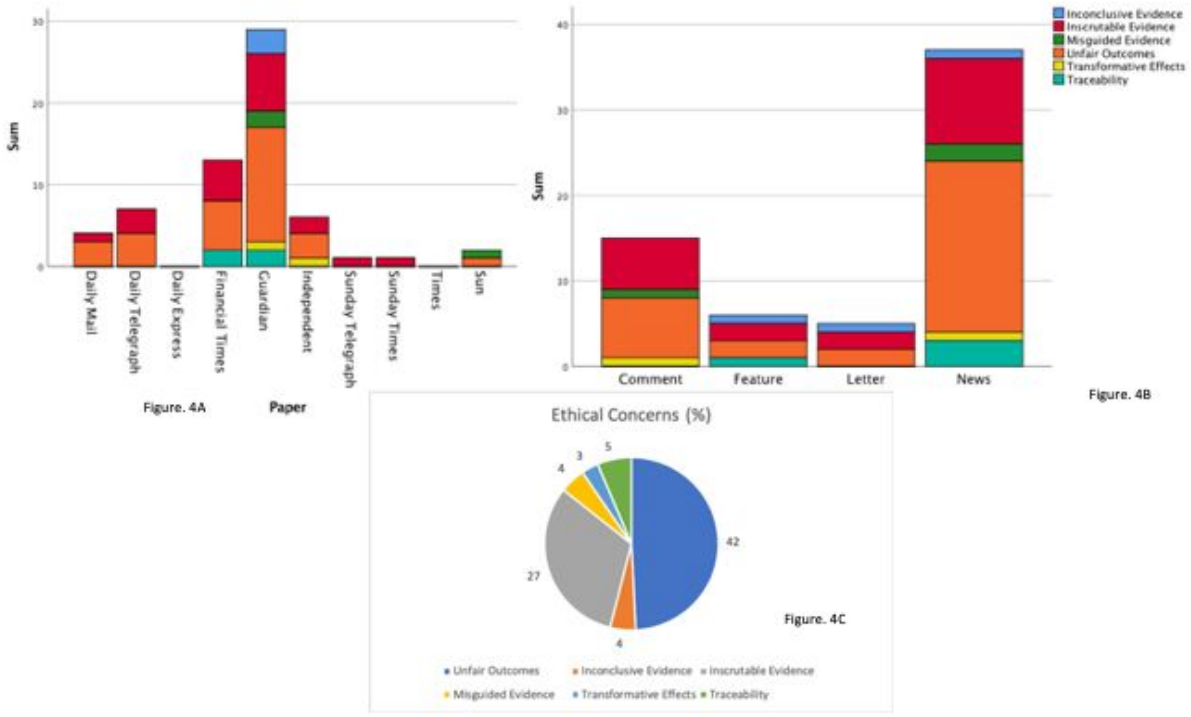

Figure 4: Ethical Concerns

The media reported examples of unfair outcomes along protected characteristics[4] (e.g. sex, race, religion or belief). Commonly described examples of the use of algorithms often listed the use of predictive analytics in policing and even sentencing (Zarsky, 2016). Most articles however did not go beyond an assertion of a risk of bias.

[4] It is against the law (Equality Act, 2010) to discriminate against someone because of a protected characteristic.

Indicating that the public debate is lacking sophistication even when reported in quality newspapers. For example:

> "Algorithms are displaying white male bias, and automation is decimating our jobs we have a lot to lose unless we (women) get involved....If a non-diverse workforce is creating them, they are more prone to be implanted with unexamined, undiscussed, often unconscious assumptions and biases about things such as race, gender and class." (Bartoletti, 2018).

The debate about algorithms is made more pertinent in the context of the use of algorithms to process large volumes of data to detect patterns and identify relationships across vast and distributed data sets (Floridi, 2012). The processes involved use machine learning techniques, that is, methodologies used to generate models for prediction and that are capable of defining or modifying decision-making rules autonomously (Van Otterlo, 2013).

Not all cases are about negative discrimination, one of the most positive sentiments articles revealed was that the use of machine learning (and indirectly, the algorithms embedded) helped remove Unfair Outcomes. The use of machine learning based chatbots to support the recruitment process suggests that neutralising discrimination is possible:

> ..."The fact it remains completely neutral with gender, age and race unidentified keeps things focused on what matters." (Warrell, 2018).

Inscrutable Evidence (27%) featured as the second biggest ethical concern. The core concern centres around the notion of the algorithm as a black box, where software does not allow a user to 'look under the bonnet' and allow independent verification of the inputs that have been entered in order to drive the output. This concern has been given much emphasis in the select committee report on AI (House of Lords Select

Committee, 2018) which recognised the need for technical transparency to address intelligibility requirements of decisions arising from algorithms. Their conclusion that "full technical transparency is difficult, and possibly even impossible" leads to an alternative proposition of "Explainability" whereby AI systems are developed in such a way that they can explain the information and logic used to arrive at their decisions. This remains an active area of research.

Examples of concerns of inscrutable evidence include the use of predictive analytics tools in social care settings prompted a letter reported in the Guardian:

> "...how can government maintain public trust in services and the use of public data, when councils are applying undisclosed algorithms to public data without our knowledge or consent, for unevaluated interventions, to screen their populations for a problem as serious as child maltreatment?" (Gilbert and Pearson, 2018).

Misguided Evidence (4%), the third component of Mittelstadt et al.'s epistemic concerns was reported primarily in the Guardian. For this type of ethical concern, data reliability is central in that conclusions and actions arising from conclusions can only be as reliable as the data in which they are based. The low numbers reported is surprising but perhaps reflects the lack of sophistication of the debate in the public media. The most significant reference to misguided evidence, was in fact, experts critiquing an algorithm that used facial recognition to detect sexual orientation.

> "...Todorov and two AI researchers at Google argued that Kosinski's algorithm could have been responding to patterns in people's makeup, beards or glasses, even the angle they held the camera at..." (Lewis, 2018)

The systematic mapping of the ethical concerns and their spread can be seen further in figure 4A and 4B where the quality newspapers and in particular Guardian

reporting most articles about ethical concerns centring on the epistemic issues of algorithms.

4.3 Relationships between ethical concerns

The epistemic and normative concerns and the traceability categories offered by Mittelstadt et al. are presented as discrete concepts and formal relationships between these concepts are not offered as part of the concept map. To this author, it is clear that there are complex relationships that might exist between these concepts that are not explicated or are simply alluded to in the narrative text. Hence, to support the mapping study exploration, a further result is presented - a conceptual model for ethical concerns that presents visually the Mittelstadt framework. In this model, the concept map is enhanced with proposed relationships delineating the types and associations between types. The derivation is based on the author's analysis of the Mittelstadt paper. A conceptual model form is chosen as it has the potential to enhance the framework and progress its development to act as a theory. The word "theory" suffers from both an over-use and a reluctance in its use by researchers. Weick comments that most theories that are labeled as theories are actually *approximate* theory in that they go some way to establishing a theory but fall short in some aspect such as: failing to sufficiently articulate relationships between variables/concepts contributing to the theory; or perhaps where ad hoc hypotheses are derived from limited observations (Weick, 1995).

Others have noted that scientific theories demonstrate analogous properties to conceptual models (Evermann and Mistry, 2008):

> "In system development, the purpose of the conceptual model is to describe the elements of the domain and their relationships (Mylopoulos, 1992). The conceptual model serves as the basis of understanding and problem solving within the domain.

It allows analysts to capture and communicate their understanding of a domain and the problems in the domain. In science, it is the role of theories to describe the constituents of the domain, their relationships and their behavior, and to serve as the means by which problems in a domain are specific and solved."

Given this analogy, it appears viable that a conceptual model defining constructs, relations, constraints and possibly behaviours (propositions) could be used as a representative form for describing and further elaborating this Ethical Concerns theory.

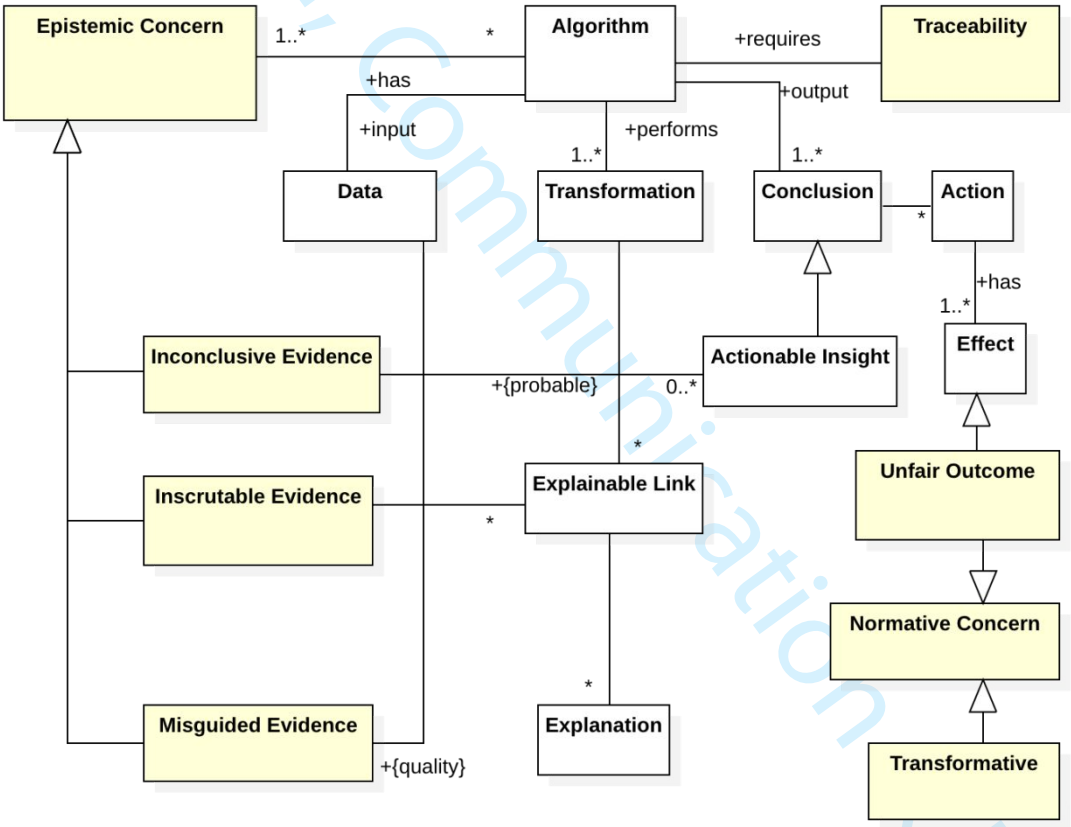

Figure 5. Ethical concerns meta model

As noted above, complex relationships that might exist between these concepts that are not explicated in the original proposal are made visible in Figure 5. The UML conceptual model formalises associations outlined in the descriptive prose in (Mittelstadt et al., 2016). The UML model also allows the formulation of constraints to

add further semantics to the diagram. When a correlation test between the various concepts was performed three significant relationships were identified. These are shown in figure 6. The strongest correlation (0.305**) was between Inconclusive Evidence and Misguided Evidence. In the Mittelstadt et al. model/theory, these concepts both exist within the Epistemic concerns structure and may therefore be implictly related. In the UML conceptual model shown in Figure 5, these concepts are both subtypes of the same supertype (EpistemicConcerns). The data supports that proposal. The two other significant correlations are Inconclusive Evidence and Traceability (0.254*) and Inconclusive Evidence and Unfair Outcomes (0.242*). These cross from the Epistemic to Normative Concerns. Here the relationships, while clearly strong, require lengthy navigation across the model. Both navigation constraints expressed using an English language variant of Object Constraint Language (Warmer and Kleppe, 1998) are offered below.

Context: Inconclusive Evidence: self.conclusion.algorithm.traceability
Context: Inconclusive Evidence: self.conclusion.action.unfairoutcome

The complexity created by the navigations across several concepts including multiplicity constraints offers an explanation as to why ethical concerns centre around lack of transparency of decision making and unfair outcomes arising from inconclusive evidence. Transparency is made challenging because of the number of types involved, the relationships traversed, and implication that Actionable Insights are type of Conclusion and Unfair Outcomes are a type of Effect. The UML model, by breaking down the relationship into sub-components presents, minimally, a handle on the complexity.

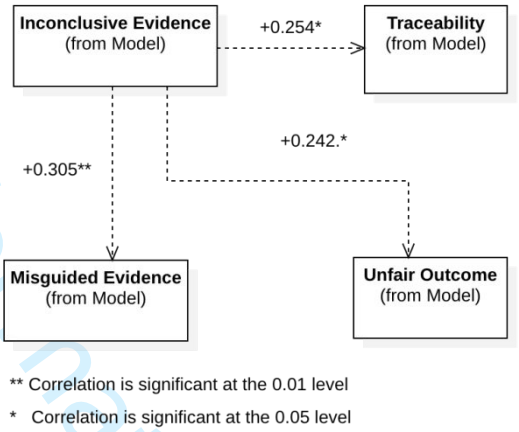

Figure 5 Correlations between ethical concerns

5 Limitations

The principal limitation of the results reported in the study centre around the systematic mapping of the media reports from the main UK national newspapers. Limitations can can be presented with respect to construct validity, reliability, internal and external validity as used by Engstrom and Runeson (2011).

Construct validity depends upon the research questions and whether all relevant sources have been taken into consideration. The scope of the mapping was very well defined, the main UK national newspapers as specified by one of the leading news aggregator search databases - Factiva. The key question on construct validity is a discussion on whether the ethical concerns framework used in this paper is sufficiently robust. There is a reliance and assumption that peer review, academic rigour and subsequent citation of the Mittelstadt et al. model offers some confidence regarding suitable construct validity. Categorisation of media with respect to each ethical concern is dependent upon close reading the text, but it is accepted that errors in allocating categories may be possible.

Reliability concerns the arrival of a similar conclusion through repeating the study by another researcher. The protocol definition phase and aspects of the execution phase are repeatable by other researchers. The Mittlstad et al. framework offers a base benchmark for using pre-defined categories offering a significant opportunity for building in reliability for future studies.

External validity is about generalisation about the results from this study. This paper does not draw any conclusions about mapping studies in general hence, external validity threats are not applicable here.

The scope of the mapping study could have been bigger. For example, a longer period could have been chosen. It is noted however, that even this period from 21 September 2016 to 21 September 2018 saw one of the biggest issues around the use of algorithms and ethical concerns explode with the news story concerning Cambridge Analytica and the harvesting of 50 million Facebook profiles for political analysis purposes (Graham-Harrison and Cadwalladr, 2018).

6 Conclusion

Algorithms are increasingly in the public eye through articles in the mainstream media, the so-called fourth estate. Their role, especially, in the context of artificial intelligence / machine learning is in decision making in all aspects of society. The key debate of the role of algorithms centres around ethical concerns such as the transparency of decision making and unfair outcomes arising from decision making. The work described here has made three principal contributions around this debate. Firstly, a systematic mapping study of leading UK national newspapers for one year has been conducted and a map of the key ethical concerns being debated has been produced. Secondly, the paper has provided experimental evaluation of a significant framework for evaluating ethical concerns. Finally, the research has contributed an initial UML conceptual model of this

framework thereby adding further depth and texture to the Mittelstadt et al. conceptual framework. In particular, the conceptual model exposes dependencies between the two principal elements of the original framework, the Epistemic and Normative concerns.

While the presentation of algorithm articles follow a media framing model, future research could explore more precisely the defining characteristics of algorithm framing in the media and therefore allow for more substantive critique of such articles.

Future studies could adopt the core approach taken in this paper by developing further mapping studies and perhaps systematic literature reviews of both public media and academic literature to explore specific relationships exposed in the UML conceptual model.

Acknowledgements

The comments from the two anonymous reviewers have been very helpful and constructive and the feedback is duly acknowledged.

7 References

Andrews, L., 2018. Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'. *Public Administration*.

House of Lords Select Committee, 2018. AI in the UK: ready, willing and able? *House of Lords*, *36*. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. [Accessed 10 Jun. 2019].

Barn, B S., Ravinder Barn, and Franco Raimondi (2018). "The BBC Broadcast of "India's Daughter": An Examination of the Interactions between Mainstream Media and Social Media". In: International Journal On Advances in Internet Technology 11.1 & 2.

Bartoletti, I (2018). Women must act now, or male-designed robots will take over our lives. Available at: https://www.theguardian.com/commentisfree/2018/mar/13/women-robots-ai-male-artificial-intelligence-automation [Accessed 10 Jun. 2019].

Bogost, I (2015). The Cathedral of Computation. Available at: https://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/ . [Accessed 10 Jun. 2019].

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080

De Vreese, C.H., 2005. News framing: Theory and typology. *Information design journal & document design*, *13*(1).

Charon, R., 2017. Close reading: the signature method of narrative medicine. *The principles and practice of narrative medicine*, pp.157-179.

Diakopoulos, N. (2013). Algorithmic accountability reporting: On the investigation of black boxes. A Tow/Knight Brief. Tow Center for Digital Journalism, Columbia Journalism School. Available at: https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2. [Accessed 10 Jun. 2019].

Engström, E. and Runeson, P., 2011. Software product line testing–a systematic mapping study. *Information and Software Technology*, *53*(1), pp.2-13.

Evermann, J. and Mistry, R., 2008. System Analysis as Scientific Inquiry. *AMCIS 2008 Proceedings*, p.154.

Floridi, L., 2012. Big data and their epistemological challenge. *Philosophy & Technology*, *25*(4), pp.435-437.

Gilbert, Ruth and Rachel Pearson (2018). Don't trust algorithms to predict child abuse risk. Available at: https://www.theguardian.com/technology/2018/sep/19/dont-trust-algorithms-to-predict-child-abuse-risk [Accessed 10 Jun. 2019].

Gillespie, T., 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, *167*, p.167.

Gitlin, T. (1980). The whole world is watching. Berkeley: University of California Press.

Graham-Harrison, E. and Cadwalladr, C. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. [online] the Guardian. Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [Accessed 13 Jul. 2019].

Government, HM., 2018. Industrial Strategy: building a Britain fit for the future. Available at: https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future. [Accessed 10 Jun. 2019].

Government, HM. (2019). *Centre for Data Ethics and Innovation Consultation*. [online] GOV.UK. Available at: https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation [Accessed 13 Jul. 2019].

Gregor, S., 2006. The nature of theory in information systems. *MIS quarterly*, pp.611-642.

Graham-Harrison, E. and Cadwalladr, C. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. [online] the Guardian. Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [Accessed 10 Jun. 2019].

Kitchin, R., 2017. Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), pp.14-29.

Kitchenham, B., 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*(2004), pp.1-26.

Kitchenham, B., 2007. *Guidelines for performing systematic literature reviews in software engineering* (Vol. 5). Technical report, Ver. 2.3 EBSE Technical Report. EBSE.

Lewis, Paul (2018). 'I was shocked it was so easy': meet the professor who says facial recognition can tell if you're gay. Available at: https://www.theguardian.com/technology/2018/jul/07/artificial-intelligence-can-tell-your-sexuality-politics-surveillance-paul-lewis . [Accessed 10 Jun. 2019].

Marda, V., 2018. Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), p.20180087.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L., 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), p.205.

Morley, J., Floridi, L., Kinsey, L. and Elhalal, A., 2019. From What to How. An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. *arXiv preprint arXiv:1905.06876*.

Newman, N., Dutton, W.H. and Blank, G., 2012. Social Media in the Changing Ecology of News: The Fourth and Fifth Estates in Britain. *International Journal of Internet Science*, *7*(1).

Pereira, R., Baranauskas, M.C.C. and Liu, K., 2018. An Essay on Human Values in HCI. *SBC Journal on Interactive Systems*, *9*(1), pp.4-16.

Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M., 2008, June. Systematic mapping studies in software engineering. In *Ease* (Vol. 8, pp. 68-77).

Petersen, K., Vakkalanka, S. and Kuzniarz, L., 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, pp.1-18.

Royal Society (2017). *Data management and use: Governance in the 21st century - a British Academy and Royal Society project | Royal Society*. [online] Royalsociety.org. Available at: https://royalsociety.org/topics-policy/projects/data-governance/ [Accessed 11 Jun. 2019].

Smith, Matthew (2017). How left or right-wing are the UK's newspapers? Available at: https://yougov.co.uk/topics/politics/articles-reports/2017/03/07/how-left-or-right-wing-are-uks-newspapers . [Accessed 10 Jun. 2019].

Turing.ac.uk. (2019). *The Alan Turing Institute*. [online] Available at: https://www.turing.ac.uk [Accessed 13 Jul. 2019].

Vale, T., Crnkovic, I., De Almeida, E.S., Neto, P.A.D.M.S., Cavalcanti, Y.C. and de Lemos Meira, S.R., 2016. Twenty-eight years of component-based software engineering. *Journal of Systems and Software*, *111*, pp.128-148.

Van den Hoven, J. and Rooksby, E., 2008. Distributive justice and the value of information: A (broadly) Rawlsian approach. *Information technology and moral philosophy*, *376*.

Van Otterlo, M., 2013. A machine learning view on profiling. *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology. Abingdon: Routledge*, pp.41-64.

Walia, G.S. and Carver, J.C., 2009. A systematic literature review to identify and classify software requirement errors. *Information and Software Technology*, *51*(7), pp.1087-1109.

Warmer, J.B. and Kleppe, A.G., 1998. The object constraint language: Precise modeling with uml (addison-wesley object technology series).

Warrell, Helen (2018). Just the job: how AI is helping build a better life for refugees. Available at: https://www.ft.com/content/9332fffc-ec57-11e8-89c8-d36339d835c0 [Accessed 10 Jun. 2019].

Weick, K.E., 1995. What theory is not, theorizing is. *Administrative science quarterly*, *40*(3), pp.385-390.

Zarsky, T., 2016. The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, *41*(1), pp.118-132.

Ziwitz, M., 2016. Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, *41*(1), pp.3-16.