Article

Algorithms, ontology, and social progress

Andrew Iliadis [iD]
Temple University, USA

Abstract
Recently, media and communication researchers have shown an increasing interest in critical data studies and ways to utilize data for social progress. In this commentary, I highlight several useful contributions in the International Panel on Social Progress (IPSP) report toward identifying key data justice issues, before suggesting extra focus on algorithmic discrimination and implicit bias. Following my assessment of the IPSP's report, I emphasize the importance of two emerging media and communication areas – applied ontology and semantic technology – that impact internet users daily, yet receive limited attention from critical data researchers. I illustrate two examples to show how applied ontologies and semantic technologies impact social processes by engaging in the hierarchization of social relations and entities, a practice that will become more common as the Internet changes states towards a 'smarter' version of itself.

Introduction

Pick any area of the internet and its power dynamics can be observed from several levels of technical (Bratton, 2016), legal (Braman, 2006) and social (Isin and Ruppert, 2015) abstraction. On the technical level, new instances of data connectivity or blockages impact internet users daily, creating re-ontologized (Floridi, 2013) realities that manifest as opportunities or challenges. The growth of data-driven apps and platforms (Gillespie, 2010), ubiquitous computing and internet of things (Greengard, 2015), and smart cities (Kitchin et al., 2017) extend datafication and dataveillance (Van Dijck, 2014) beyond the web and social media, interconnecting layers of legacy infrastructures with newer innovations and trends. Existing as a nebulous network of networks (Noam, 2001),

Corresponding author:
Andrew Iliadis, Department of Media Studies and Production, Temple University, 2020 N. 13th St., Philadelphia, PA 19122, USA.
Email: andrew.iliadis@temple.edu

the internet's pathways and perennial ability to stack, repurpose or erode technological infrastructure create a shifting ground upon which users must constantly renegotiate their positionalities, rights and abilities to interact with one another.

At the technical level, critical data researchers have recently focused their attention on social justice issues and the social implications of networked data science and data-driven processes on the internet (Acker, 2015; Boyd and Crawford, 2012; Couldry and Powell, 2014; Crawford et al., 2014; Dalton and Thatcher, 2014; Dencik et al., 2016; Hoffmann, 2017; Iliadis and Russo, 2016; Johnson, 2014; Kitchin, 2014a; 2014b; Kitchin and Lauriault, 2014; Markham, 2013; Mayernik and Acker, 2018; Milan and van der Velden, 2016; Neff et al., 2017; Posner and Klein, 2017; Schrock, 2017; Steinmann et al., 2016; Taylor, 2017). That data can be used to impede or improve social progress is one of the underlying assumptions of recent critical data research. How might data processes create more just and fair societies through the internet while holding improper use of power to account? How can critical data studies contribute to social progress and social justice, and what are the necessary tools and frameworks that can enable positive outcomes and public understanding? As noted by Michael and Lupton (2015), public understanding of how big data are collected and curated on the internet requires 'engagement with new modes of knowledge production and circulation, new academic literatures and new ways of thinking about data and data practices' (p. 110). Part of the challenge of creating successful campaigns for the public understanding of big data includes reassessing the policies, best practices, criteria and standards through which we asses social progress, including those of big data research (Metcalf and Crawford, 2016).

In this commentary, I highlight several useful contributions in the International Panel on Social Progress (IPSP) report toward identifying key data justice issues, before suggesting extra focus on algorithmic discrimination and implicit bias. Following my assessment of the IPSP's report, I emphasize the importance of two emerging media and communication areas – applied ontology and semantic technology – that impact internet users daily, yet receive limited attention from critical data researchers. I illustrate two examples to show how applied ontologies and semantic technologies impact social processes by engaging in the hierarchization of social relations and entities, a practice that will become more common as the internet changes states towards a 'smarter' version of itself.

Social progress

New advances in internet technologies provide moments for reflection, particularly when they concern social relations and structures. The IPSP's newly updated report (Couldry et al., 2017) on pressing social issues in the 21st century features a comprehensive chapter (hereafter, 'the report') on media and communication technologies that covers, among other topics, issues related to social media, algorithms and big data. In a world where media and communication technologies are growing rapidly, the updated chapter is welcome, and the case studies contained therein are timely and relevant. The IPSP's critical engagement with the Social Progress Index (SPI) moves beyond the narrow criteria for assessment of media and communication outlined in the SPI (mobile telephone subscriptions, internet users and press freedom), highlighting and expanding the criteria for gauging social progress to include recent data-driven developments.

More specifically, the report addresses the shift from so-called Web 1.0 to Web 2.0 technologies (Barassi and Treré, 2012). Where early instantiations of the internet mainly included static and hyperlinked material to be read, the report describes Web 2.0 as offering interactivity and writing between users on the internet as a defining feature, spread through social media. Highlighting this trend toward datafication, dataveillance and the potential exploitation of internet users by companies like Facebook and Google, the entire report is divided into eight sections. The first half positions media as boosting cultural complexity and describes social justice issues and the need for media reform, providing a history of pre-internet media including several case studies from China, Russia, Sweden and South Africa, before focusing on transnational governance and the importance of journalism and public knowledge for democracy. The second half of the report addresses networked communications and possibilities for citizenship, including case studies in China and East Asia, describes struggles for social justice through the democratization of media and presents two case studies on Facebook Free Basics in India (Yim et al., 2016) and Brazil's Marco Civil (Hoskins, 2017). The report ends by discussing struggles for social justice through media, their affordances and constraints.

Algorithmic discrimination and implicit bias

Every day, algorithms contribute to what Couldry and Hepp (2017) refer to as the deep mediatization of reality – our worlds are becoming steeped in algorithmic mediation given the ways we rely on algorithms for everything from news consumption (Diakopoulos and Koliska, 2016) to shopping and advertising (Gal and Elkin-Koren, 2017; Sinclair, 2016). Algorithms are media technologies that are changing decision-making practices in multiple areas of life and activity with increasing relevance (Gillespie, 2014), and the IPSP report highlights several areas related to algorithms for assessing social progress.

Among wider concerns about net neutrality and internet freedom, the IPSP report contains warnings about predicative algorithms' use in potentially discriminatory operations. Concerns include algorithmic filtering to remove or delist content that infringes copyright, the creation of filter bubbles in search results and newsfeeds (Pariser, 2011), the opacity of algorithms as trade or state secrets hidden from the public (Pasquale, 2015), and algorithms' ability to identify segments of the population according to granular demographics, raising concerns that targeting may lead to discrimination in sales, employment and other areas (Barocas and Selbst, 2016). The report further discusses the need for algorithmic transparency and accountability, considering evidence that algorithms can be used to manipulate users, as in the Facebook emotion manipulation study (Kramer et al., 2016) – though some researchers have argued that there are limitations to the transparency ideal when it does not take into consideration user understanding (Ananny and Crawford, 2016).

There is a growing body of literature on the ethics of algorithms (Ananny, 2016; Kraemer et al., 2010; Mittelstadt et al., 2016; Neyland, 2016) and extra material on the application of algorithmic media in areas such as prison sentencing, the employee hiring process, local policing, college and university admissions, and parole decisions would be welcome in the report. The report includes computational media as an important factor in the assessment of social progress. If algorithms contribute to mediatization and the

datafication of social processes, then their use in the above areas suggests that algorithmic mediation impacts realms of social decision-making traditionally governed by human judges and decision-makers. Proponents of algorithms often cite algorithms' alleged ability to make bias-free decisions that are more accurate than those of their human counterparts. Recent studies have even reported that algorithms are more accurate than humans at detecting sexual orientation (Wang and Kosinski, 2017), disguised faces (Singh et al., 2017) and criminality using face images (Wu and Zhang, 2016). Such studies raise serious ethical concerns and present claims based on inconclusive, inscrutable or misguided evidence (Mittelstadt et al., 2016). In the construction phase, algorithms are trained to sort data according to a set of rules. What those rules and data are, how they are labelled and the values that are ascribed to them matter in combating algorithmic discrimination and implicit bias.

Media and communication researchers who are interested in algorithms and social progress must also engage data-intensive technologies in newly mediated spheres. Data themselves can reflect human bias or be uneven in their collection, may produce feedback loops in applications like predicative policing (Ferguson, 2017), as well as miss subgroups of individuals based on hidden properties. Lack of transparency and accountability in data construction practices prevent the public from meaningfully engaging new, algorithmically mediated aspects of life. Algorithms may impact society in data-intensive contexts such as web advertising (Sweeney, 2013) and predictive sentencing guidelines (Angwin et al., 2016). Yet, data-based decision-making has always had the potential to discriminate in unintended, a priori ways. Algorithms do not eliminate implicit bias and can amplify it if training data are themselves biased, as current research has shown (Levendowski, 2017).

The IPSP report includes an action plan and toolkit with recommendations for increasing social progress in algorithms, including regulating the use of algorithms for marketing or surveillance purposes, leading public conversations about filtering and predictive algorithms and, importantly, demanding transparency and accountability of data collection and filtering. These recommendations introduce important action items for the IPSP while updating the SPI's criteria for assessment. The calls are in line with past critical datafication and dataveillance movements like Do Not Track (Fomenkova, 2012), which, beginning as an update to the US federal Do Not Call registry, focused on privacy solutions to internet tacking by online advertising companies through cookie-disabling browser extensions and browser header notifications. Similarly, the Privacy by Design (Cavoukian, 2012) and Value Sensitive Design (Davis and Nathan, 2015) movements have called for privacy measures to be built into technological systems, their maintenance and policies.

More information on data collection, curation and circulation before their utilization in algorithmic systems would be welcome in the IPSP's report. In May 2016, the Obama administration released *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (The White House, 2016). The report outlines several opportunities and challenges that algorithms provide related to access to credit, employment, higher education and criminal justice. Each of these challenges relates to discrimination or implicit bias, but the first section of the report addresses the primary issue of selecting inputs (data) for an algorithm. Poorly selected, incomplete, incorrect or outdated data, selection

bias, and unintentional perpetuation and promotion of historical biases are described as some of the main challenges. What data points count as personalization in the algorithm and what count as discrimination? Are the data labelled correctly, and what if they are inaccurate?

Applied ontology and semantic technology

One area of media and communications left unaddressed by the IPSP's report is applied (computational) ontology (Arp et al., 2015; Pease, 2011). Ontologies help integrate disparate or unorganized data to produce meaning, sort of 'like a thesaurus, a finite set of terms, organized as a hierarchy that can be used to provide a value for an element. Additionally, this includes a set of rules for action, often in the form of software algorithms' (Pomerantz, 2015). Modern applied ontologies are an outgrowth of early artificial intelligence research in expert systems (Hayes-Roth et al., 1983) and knowledge representation (Sowa, 1999). Today, ontologies are used in many data-driven media technologies like virtual assistants and social media platforms (Tecuci et al., 2016). As the Internet continues to mature and smart devices depend on some level of semantic engineering, applied ontologies will continue to become a key feature of individuals' everyday interactions with the internet. Applied ontologies increasingly dictate data labelling and data flows and can have significant impact on social processes, including research in industry and academia. Take, for example, the notion that 'commitment to a particular ontology will influence one's epistemology and the attendant research methodology and protocol' (Arneson, 2009: 696). Ontology work infuses popular data-products and research processes in data-intensive science, yet there are several potential problems related to data-driven ontology practice and multifaceted approaches to studying applied ontology work.

The following questions represent just some of the emerging social concerns in data-based ontology research and practice:

- Is there evidence that ontologies typify semantic logics or biases?
- Do ontologies enforce exclusionary criteria in categorizing data about social entities and, if so, is there a need to make such criteria explicit?
- What types of domain-specific data do ontologies organize?
- How are ontologies practically applied in scientific and social contexts?
- What does ontology work look like in research projects at multiple scales?
- How may ontologies assist with access to big data troves and algorithmic processes that collect, collate and make decisions based on them?
- What methodologies are currently deployed by ontologists and internet researchers and what policies, regulations, and standards apply to them?

Some of these questions have been addressed by Science and Technology Studies researchers (Bowker et al., 2010; Bowker and Star, 1999; Edwards et al., 2011; Millerand and Bowker, 2009; Ribes and Bowker, 2009; Ribes and Polk, 2015; Woolgar and Lezaun, 2013) and Library and Information Science researchers (Almeida, 2013; Borgman, 2010, 2011, 2015; Borgman et al., 2012; Fonseca, 2007; Fonseca and Martin, 2007) interested

in data labelling and sharing in the context of scientists solving data-sharing problems in science teams. Building and extending this work to data justice and social progress issues involves also looking at how ontology is connected to data gathering, data modelling, databases, metadata and how the use of these and other tools like application programming interfaces (Qiu, 2017) impact civil society through public facing ontology-driven apps and technologies. Drawing on the work of Gitelman (2008), Srinivasan (2012) offers one such approach by asking how we might include computational ontology in our discussion of 'ethical questions about the sovereignty of diverse knowledge, and whether the voices of emerging users should be ignored or empowered' (p. 205).

What happens, for example, when social entities and relations must be represented as digital objects (Hui, 2012; Kallinikos et al., 2010)? This is a modern update to an old problem, one we have seen in critical scholarship on the history of the census, statistics and, more recently, big data (Beer, 2016; Hacking, 1982, 1991). The question 'who counts?' can be read as a double articulation – who is doing the counting and who deserves to be counted? Applied ontologies are an update to those problems, complicated by semantics ('who counts *what*?'). Currently employed in areas as diverse as municipal administration, virtual personal assistants, business and logistics, and military intelligence gathering, applied ontologies that deal with social entities and relations necessitate what Couldry and Kallinikos (2017) describe as a 'new ontology of the social' (p. 153). Computational ontologies encourage the datafication of social entities and relations by constructing social ontologies (Searle, 2006) that are used to provide labels for data in organized, semantic structures. Once completed, heterogeneous data can be combined and analysed in ways that were not possible when they retained their own idiosyncratic labels, and computations can be executed to extract new information.

Two applied ontology examples will help make the connection to data justice issues clearer. Computational ontologies generally exist on two levels – upper level and domain specific. Upper level ontologies are highly formalized and used by various industries to help integrate heterogeneous data from different points of origin. A good example of this would the Basic Formal Ontology (BFO), which offers a strict set of rules to follow if administrators would like to integrate different datasets. Currently, the BFO is successful in the field of bioinformatics, and the Open Biomedical and Biological Foundry (OBO Foundry) uses the BFO as a standard for inclusion (Smith et al., 2007). Various research teams around the world submit their BFO-specified data to the OBO Foundry, and if those data are found to be properly labelled according to BFO principles, the dataset is included in the Foundry. The benefit of this practice is that research data from vastly different areas of science can now be analysed alongside one another because they are now using the same data-labelling standards. Data from the domain-specific Human Disease Ontology can now be analysed alongside the Gene Ontology because they both adhere to BFO principles.

Now, the IPSP states near the end of their report that national security services engage in data collection and processing, and that they often share the results with one another to circumvent restrictions that might apply to data collection and processing conducted within territorial boundaries. As we have seen from the Snowden revelations, this is true, and computational ontology is one way that governments achieve this (Lyon, 2014). For example, the United States Armed Forces uses the same BFO as the OBO Foundry to

achieve horizontal integration of war-fighter intelligence data (Smith et al., 2012). In both cases, the BFO includes social ontology categories and relations, but these are populated in different ways, depending on who is doing the data annotation, and this might present potential problems. Individuals who are the subjects of data may never know that their data are being integrated with other databases, thus lacking consent. There are also, to my knowledge, no ethics reviews conducted on ontology research. Administrative mislabelling of social data categories may be exceedingly difficult to remove once the data are combined for use in an open database. There are also consequences such as potential grouping, correlation and guilt by association that may adversely affect data subjects once cross-sector data harmonization becomes more widespread. What will happen when several surveillance databases of various sizes are harmonized or integrated through an applied computational ontology – including, perhaps, one innocuous database that contains your name?

Applied ontologies are also used in popular social media apps like Facebook and virtual personal assistants like Siri. On the front-end interface, Facebook presents categories that the user can choose from – for example, the way gender is included in social media design (Bivens and Haimson, 2016). Users see only an outward-facing part of proprietary ontologies and generally the public does not know how companies have organized their metadata, or with whom they might be shared. The popular virtual personal assistant Siri, referred to by its creators as 'an ontology-driven application for the masses' (Cheyer and Gruber, 2010), uses something they call an 'Active Ontology' to build and run applications (Cheyer and Guzzoni, 2006). Siri's Active Ontology takes data provided from various services that are orchestrated (apps like Yelp and Instagram), runs them through domains and task modules (essentially, designated categories like restaurant and movies), and connects them to intelligent user interfaces (what the user sees or hears). The Active Ontology connects and organizes the disparate data provided by apps (via APIs) so Siri can relate semantic or smart information to the user. There are other elements, of course, but the Active Ontology is the conduit through which data from the other components are passed and shaped. Thomas Gruber, one of the pioneers of computational ontology and a key developer of the Siri assistant, states that 'Siri is building on the ecosystem of APIs, which are better if they declare the meaning of the data in and out via ontologies. That is the original purpose of ontologies-as-specification that I promoted in the 1990s – to help specify how to interact with these agents via knowledge-level APIs' (Spivack, 2010).

Ontology-driven media technologies like Google and Facebook's graphs, popular semantic web standards like the World Wide Web Consortium's (W3C) Web Ontology Language (OWL) and virtual personal assistants like Siri, Cortana, Alexa and Bixby provide unique opportunities. But, like any technology, they can impede social progress if, during their development, designers are not also attentive to data justice issues. Ontologies present truly unique problems – they are not only a matter of quantification, but also a matter of meaning. What counts as a restaurant in Siri's Active Ontology? How are social entities and relations defined in OWL? What languages do ontologies recognize? The IPSP's report updates the SPI and covers social progress issues in the shift

from Web 1.0 to Web 2.0 technologies, but we are already well on our way to Web 3.0 – a semantic, intelligent internet that presumes to know what we mean.

Conclusion

As the Internet continues to mature, we will have to contend with semantic technologies that presume to know what we want or what we are looking for. Such technologies depend on formalized and standardized vocabularies and will produce new realities for social engagement and activity. But they also bring with them all the old concerns that are expressed each time new or emerging technologies are brought into the mainstream and utilized by the masses. More attention will have to be paid to emerging technologies like applied ontologies, semantic technologies, and how they prioritize and create hierarchies of social entities and relations. The IPSP report ends by stating that the new media landscape is marked by the logic of data extraction and data stimulation, where unfamiliar forms of domination and exclusion are emerging, and this is certainly true of where we are heading. Part of capital's power comes from its ability to tag, name, identify and quantify even the seemingly unquantifiable things in our world. If this is the case, the logic of computational ontologies – while extremely beneficial for things like cancer research and virtual science – lend themselves to the type of calculations that will always privilege transactions over translations.

ORCID iD

Andrew Iliadis (iD) https://orcid.org/0000-0002-8345-6251

References

Acker A (2015) Toward a hermeneutics of data. *IEEE Annals of the History of Computing* 37(3): 70–75.

Almeida MB (2013) Revisiting ontologies: A necessary clarification. *Journal of the Association for Information Science and Technology* 64(8): 1682–1693.

Ananny M (2016) Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values* 41(1): 93–117.

Ananny M and Crawford K (2016) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20: 973–989.

Angwin J, Larson J, Mattu S, et al. (2016) Machine bias. *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 1 December 2017).

Arneson P (2009) Ontology. In: Littlejohn SW and Foss KA (eds) *Encyclopedia of Communication Theory*. Los Angeles, CA: Sage, pp.695–698.

Arp R, Smith B and Spear A (2015) *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: The MIT Press.

Barassi V and Treré E (2012) Does Web 3.0 come after Web 2.0? Deconstructing theoretical assumptions through practice. *New Media & Society* 14(8): 1269–1285.

Barocas S and Selbst A (2016) Big data's disparate impact. *California Law Review* 104: 671–732.

Beer D (2016) How should we do the history of big data? *Big Data & Society* 3(1): 1–10.

Bivens R and Haimson OL (2016) Baking gender into social media design: How platforms shape categories for users and advertisers. *Social Media + Society* 2(4).

Borgman CL (2010) *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: The MIT Press.

Borgman CL (2011) The conundrum of sharing research data. *Journal of the Association for Information Science and Technology* 63(6): 1059–1078.

Borgman CL (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: The MIT Press.

Borgman CL, Wallis JC and Mayernik MS (2012) Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work* 21: 485–523.

Bowker G and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.

Bowker G, Baker K and Millerand F (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrup L and Allen M (eds) *International Handbook of Internet Research*. Dordrecht: Springer, pp.97–117.

Boyd D and Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.

Braman S (2006) *Change of State: Information, Policy, and Power*. Cambridge, MA: The MIT Press.

Bratton B (2016) *The Stack: On Software and Sovereignty*. Cambridge, MA: The MIT Press.

Cavoukian A (2012) Privacy by design. *IEEE Technology and Society Magazine* 31(4): 18–19.

Cheyer A and Gruber T (2010) Siri: A virtual personal assistant. Available at: http://ontolog.cim3.net/file/resource/presentation/Siri_20100225/Siri–An-Ontology-driven-Application-for-the-Masses–AdamCheyer-TomGruber_20100225.pdf (accessed 1 December 2017).

Cheyer A and Guzzoni D (2006) Method and apparatus for building an intelligent automated assistant US 8677377 B2. Available at: https://www.google.ca/patents/US8677377 (accessed 1 December 2017).

Couldry N and Hepp A (2017) *The Mediated Construction of Reality*. Cambridge: Polity Press.

Couldry N and Kallinikos J (2017) Ontology. In: Burgess J, Marwick A and Poell T (eds) *The SAGE Handbook of Social Media*. London: Sage, pp.146–159.

Couldry N and Powell A (2014) Big data from the bottom up. *Big Data & Society* 1(2): 1–5.

Couldry N, Rodriguez C, Bolin G, et al. (2017) *Media and communications*. Paris: International Panel on Social Progress.

Crawford K, Miltner K and Gray ML (2014) Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication* 8: 1663–1672.

Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data'.. *Society & Space*. Available at: http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/ (accessed 6 September 2016).

Davis J and Nathan LP (2015) Value sensitive design: Applications, adaptations, and critiques. In: Van den Hoven J, Vermaas PE and van de Poel I (eds) *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*. Dordrecht: Springer, pp.11–40.

Dencik L, Hintz A and Cable J (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society* 3(2).

Diakopoulos N and Koliska M (2016) Algorithmic transparency in the news media. *Digital Journalism* 5(7): 809–828.

Edwards PN, Mayernik MS, Batcheller AL, et al. (2011) Science friction: Data, metadata, and collaboration. *Social Studies of Science* 41(5): 667–690.

Ferguson A (2017) *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: New York University Press.

Floridi L (2013) *The Ethics of Information*. Oxford: Oxford University Press.

Fomenkova GI (2012) For your eyes only? A 'Do Not Track' proposal. *Information & Communications Technology Law* 21(1): 33–52.

Fonseca F (2007) The double role of ontologies in information science research. *Journal of the Association for Information Science and Technology* 58(6): 786–793.

Fonseca F and Martin J (2007) Learning the differences between ontologies and conceptual schemas through ontology-driven information systems. *Journal of the Association for Information Systems* 8(2): 129–142.

Gal MS and Elkin-Koren N (2017) Algorithmic consumers. *Harvard Journal of Law & Technology* 30(2): 309–353.

Gillespie T (2010) The politics of 'platforms'. *New Media & Society* 12(3): 347–364.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: The MIT Press, pp.167–194.

Gitelman L (2008) *Always Already New: Media, History, and the Data of Culture*. Cambridge, MA: The MIT Press.

Greengard S (2015) *The Internet of Things*. Cambridge, MA: The MIT Press.

Hacking I (1982) Biopower and the avalanche of printed numbers. *Humanities in Society* 5: 279–295.

Hacking I (1991) How should we do the history of statistics?. In: Burchill G, Gordon C and Miller P (eds) *The Foucault Effect: Studies in Governmentality*. Chicago, IL: The University of Chicago Press, pp.181–195.

Hayes-Roth F, Waterman DA and Lenat DB (1983) *Building Expert Systems*. Reading, MA: Addison-Wesley.

Hoffmann AL (2017) Making data valuable: Political, economic, and conceptual bases of big data. *Philosophy & Technology* 31(2): 209–212.

Hoskins G (2017) Draft once; deploy everywhere? Contextualizing digital law and Brazil's marco civil da internet. *Television & New Media* 19(5): 431–447.

Hui Y (2012) What is a digital object? *Metaphilosophy* 43(4): 380–395.

Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2).

Isin E and Ruppert E (2015) *Being Digital Citizens*. London: Rowman & Littlefield.

Johnson J (2014) From open data to information justice. *Ethics and Information Technology* 16(4): 263–274.

Kallinikos J, Aaltonen AV and Marton A (2010) A theory of digital objects. *First Monday* 15(6). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564 (accessed 1 December 2017).

Kitchin R (2014a) Big data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1).

Kitchin R (2014b) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles, CA: Sage.

Kitchin R and Lauriault TP (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. The programmable city working paper 2. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112 (accessed 6 September 2016).

Kitchin R, Lauriault TP and McArdle G (2017) Data and the city. In: Kitchin R, Lauriault TP and McArdle G (eds) *Data and the City*. Abingdon: Routledge, pp.1–14.

Kraemer F, Overveld K and Peterson M (2010) Is there an ethics of algorithms? *Ethics and Information Technology* 13(3): 251–260.

Kramer ADI, Guillory JE and Hancock JT (2016) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24): 8788–8790.

Levendowski A (2017) How copyright law can fix artificial intelligence's implicit bias problem. *Washington Law Review*. Available at: https://ssrn.com/abstract=3024938 (accessed 1 December 2017).

Lyon D (2014) Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society* 1(2).

Markham A (2013) Undermining 'data': A critical examination of a core term in scientific inquiry. *First Monday* 18(10).

Mayernik MS and Acker A (2018) Tracing the traces: The critical role of metadata within networked communications. *Journal of the Association for Information Science and Technology* 69: 177–180.

Metcalf J and Crawford K (2016) Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society* 3(1).

Michael M and Lupton D (2015) Toward a manifesto for the 'public understanding of big data'. *Public Understanding of Science* 25(1): 104–116.

Milan S and van der Velden L (2016) The alternative epistemologies of data activism. *Digital Culture & Society* 2(2): 57–74.

Millerand F and Bowker G (2009) Metadata standards: Trajectories and enactment in the life of an ontology. In: Lampland M and Star SL (eds) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca, NY: Cornell University Press, pp.149–165.

Mittelstadt BD, Allo P, Taddeo M, et al. (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* 3(2).

Neff G, Tanweer A, Fiore-Gartland B, et al. (2017) Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data* 5(2): 85–97.

Neyland D (2016) Bearing account-able witness to the ethical algorithmic system. *Science, Technology, & Human Values* 41(1): 50–76.

Noam E (2001) *Interconnecting the Network of Networks*. Cambridge, MA: The MIT Press.

Pariser E (2011) *The Filter Bubble: What the Internet is Hiding from You*. New York: Penguin Books.

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

Pease A (2011) *Ontology: A Practical Guide*. Angwin, CA: Articulate Software Press.

Pomerantz J (2015) *Metadata*. Cambridge, MA: The MIT Press.

Posner M and Klein L (2017) Editor's introduction: Data as media. *Feminist Media Histories* 3(3): 1–8.

Qiu Y (2017) The openness of open application programming interfaces. *Information, Communication & Society* 20(11): 1720–1736.

Ribes D and Bowker G (2009) Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* 19: 199–217.

Ribes D and Polk JB (2015) Organizing for ontological change: The kernel of an AIDS research infrastructure. *Social Studies of Science* 45(2): 214–241.

Schrock A (2017) What communication can contribute to data studies: Three lenses on communication and data. *International Journal of Communication* 11: 701–709.

Searle J (2006) Social ontology: Some basic principles. *Anthropological Theory* 6(1): 12–29.

Sinclair J (2016) Advertising and media in the age of the algorithm. *International Journal of Communication* 10: 3522–3535.

Singh A, Patil D, Reddy GM, et al. (2017) Disguised face identification (DFI) with facial keypoints using spatial fusion convolutional network. Available at: https://arxiv.org/abs/1708.09317 (accessed 1 December 2017).

Smith B (2004) Ontology. In: Floridi F (ed.) *The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Blackwell, pp.155–166.

Smith B, Ashburner M, Rosse C, et al. (2007) The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25: 1251–1255.

Smith B, Malyuta T, Mandrick WS, et al. (2012) Horizontal integration of warfighter intelligence data: A shared semantic resource for the intelligence community. In: *Proceedings of the seventh international conference on semantic technologies for intelligence, defense, and security* (eds PGC Costa and KB Laskey), Fairfax, VA, 23–25 October.

Sowa JF (1999) *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole.

Spivack N (2010) How Siri works—Interview with Tom Gruber, CTO of SIRI. Available at: www.novaspivack.com/technology/how-hisiri-works-interview-with-tom-gruber-cto-of-siri (accessed 1 December 2017).

Srinivasan R (2012) Re-thinking the cultural codes of new media: The question concerning ontology. *New Media & Society* 15(2): 203–223.

Steinmann M, Matei SA and Collmann J (2016) A theoretical framework for ethical reflection in big data research. In: Collmann J and Matei SA (eds) *Ethical Reasoning in Big Data: An Exploratory Analysis*. Cham: Springer, pp.11–27.

Sweeney L (2013) Discrimination in online ad delivery. *Communications of the ACM* 56(5): 44–54.

Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4(2).

Tecuci G, Marcu D, Boicu M, et al. (2016) *Knowledge Engineering: Building Cognitive Assistants for Evidence-Based Reasoning*. Cambridge: Cambridge University Press.

The White House (2016) Big data: A report on algorithmic systems, opportunity, and civil rights. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (accessed 1 December 2017).

Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

Wang Y and Kosinski M (2017) Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Available at: psyarxiv.com/hv28a (accessed 1 December 2017).

Woolgar S and Lezaun J (2013) The wrong bin bag: A turn to ontology in science and technology studies? *Social Studies of Science* 43(3): 321–340.

Wu X and Zhang X (2016) Automated inference on criminality using face images. Available at: https://arxiv.org/abs/1611.04135 (accessed 1 December 2017).

Yim M, Gomez R and Carter M (2016) Facebook 'free basics': For or against community development? *Journal of Community Informatics* 12(2): 217–225.

Author biography

Andrew Iliadis is an Assistant Professor in the Department of Media Studies and Production at Temple University.