

Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Culture

Luke Stark and Anna Lauren Hoffmann

05.02.19

Peer-Reviewed By: Anonymous and Dr. Brian Beaton

Clusters: Data, Infrastructure

Article DOI: 10.22148/16.036

Journal ISSN: 2371-4549

Cite: Luke Stark and Anna Lauren Hoffmann, "Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Culture," Journal of Cultural Analytics. May 2, 2019.

A growing list of high-profile controversies involving the social impacts of artificial intelligence systems (AI), digital data collection and algorithmic analysis have forced difficult conversations around the ethics of data-intensive digital technologies and so-called "big data" research.¹² These incidents are directly relevant to newly coalescing cultures of "data science," an emergent field which seeks both to interpret and capitalize on the creation, collection, and processing of knowledge through large collections of digital data, often in conjunction with particular techniques like machine learning (ML).¹ The long list of recent public controversies, as Brian Beaton observes, lays bare data science's extant lack of direction regarding professional ethics or values.²

¹Brian Beaton, Amelia Acker, Lauren Di Monte, Shivrang Setlur, Tonia Sutherland, and Sarah E Tracy, "Debating Data Science," *Radical History Review* 2017, no. 127 (March 2, 2017): 133-48. doi:10.1215/01636545-3690918.

²Brian Beaton, "How to Respond to Data Science: Early Data Criticism by Lionel Trilling," *Information & Culture: a Journal of History* 51, no. 3 (August 2016): 352-72. doi:10.7560/IC51303.

Various groups have ventured into this conceptual gap to advance principles or guidelines for doing ethical data science. In 2017, for example, digital organization Data for Democracy (D4D)—in partnership with media conglomerate Bloomberg and open source platform BrightHive—announced an effort to crowdsource a code of ethics for data scientists, seeking to “define values and priorities for overall ethical behavior, in order to guide a data scientist in being a thoughtful, responsible agent of change.”³ This recent focus on developing ethical codes for data science (alongside related areas such as AI/ML) suggests the field seeks to address its social impacts through discourses and processes of professional consolidation.⁴ Yet such efforts raise further questions. What kind of work counts as “data science” in the first place? What are its aims and historical precursors? And what, if any, baseline ethical commitments bind disparately situated researchers, analysts, and (of course) professional data scientists?

As data science seeks to constitute itself as a professional field, these questions will continue to lurk in the background of efforts to articulate and codify data science’s ethical commitments. In this paper, we examine one dimension of this conceptual terrain: the relevance and resonance of extant codes of professional and research ethics in—and beyond—domains related to computing, information science, and data analytics. The paper proceeds in four parts. In the first, we draw on work in the history and sociology of professional ethics to establish the (often fraught) relationship between ethical codes, professionalization, and moral responsibility, with a focus on ethics codes in domains conventionally allied with data science such as computing and information science. Second, we expand the domain of potentially relevant professional analogues by reviewing and drawing inspiration from many of the popular metaphors for “data”—and by extension, the work of data scientists—proliferating in both popular and academic settings.

In the third and fourth sections, we employ discourse analysis to assess two sets of professional ethics codes: one set rooted in conventionally allied domains like computing; the other derived from those professions evoked by the metaphors associated with data and data science. Through this discursive analysis, we develop a number of insights regarding the challenges and opportunities of developing a code of ethics for data science, and by extension the ambivalent role of

³“Code of Ethics” *Data for Democracy*, accessed June 2, 2018, <http://datafordemocracy.org/projects/ethics.html>.

⁴For an analysis of ethics statements in the AI context, see Daniel Greene, Anna Lauren Hoffmann, and Luke Stark (forthcoming), “Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning,” 52nd Hawaii International Conference on Systems Science.

data science as a profession within a broader tapestry of “data cultures.”⁵ With multiple metaphors for what “big data” is and can do, and multiple potentially relevant ethics codes scattered across different domains, data ethics is a field in flux. Collectively, these conversations represent an effort to better grapple with the consequences of the language we use for understanding and working with data—“big” or otherwise—today, and how our discourses around data cultures shape their material, cultural, and political impact.

Ethics Codes As/And Professional Cultures

Codes of ethics are longstanding mechanisms through which groups of experts have sought to define themselves and their priorities as “professionals,” or recognized members of particular professions.⁶ High-status professions such as doctors, lawyers and engineers pioneered professional ethics codes as early as the eighteenth century. Today, many other groups of experts—from foresters to firefighters to astrologers—have enacted these codes as signals, sometimes merely aspirational, of their professional status. Ongoing historical and sociological work in this area has shown such codes can and have served a range of valuable functions, from educating professionals and instilling positive group norms to setting benchmarks against which unethical behavior may be censured.⁹¹⁰ Most important for present purposes, a profession’s ethical code offers insight into how it collectively understands and seeks to modulate the distribution of obligations between individual practitioners, other individuals or stakeholders, and society more broadly.⁷

In his survey of the history of professional ethics codes, Andrew Abbot demonstrates the ubiquity of such codes as a key element of industrial state/social formation, representing perhaps “the most concrete cultural form in which professions acknowledge their societal obligations.”⁸ These codes do not arise in theoretical or conceptual vacuums. As Jacob Metcalf writes of research ethics, they represent “hard-won responses to major disruptions, especially medical and behavioral research scandals.”⁹ In addition, ethics codes are enforced most often when

⁵Kath Albury et al., “Data Cultures of Mobile Dating and Hook-Up Apps: Emerging Issues for Critical Social Science Research,” *Big Data & Society* 4, no. 2 (July 3, 2017): 205395171772095-11, 135, doi:10.1177/2053951717720950.

⁶Harold L. Wilensky, “The Professionalization of Everyone?,” *American Journal of Sociology* 70, no. 2 (September 1964): 137-58.

⁷Abbott, “Professional Ethics,” 856.

⁸Abbott, “Professional Ethics.”

⁹Jacob Metcalf, “Ethics Codes: History, Context, and Challenges,” *Council for Big Data, Ethics, and Society*, November 9, 2014.

ethical infractions or controversies are highly visible. As Abbot notes, “general public service obligations are extremely important as claims but extremely vague as rules.”¹⁰ Mark Frankel likewise notes a relationship between “professions’ pursuit of autonomy,¹¹ and the public’s demand for accountability.”¹² In view of this tension between visibility and vagueness, ethics codes often elide granular attention to professional activities, relying instead on informal everyday rules over which individual practitioners have some (albeit limited) control.

Fundamental to these debates is the question of whether a professional ethics code should be a fine-grained “how-to” manual or a set of broad principles to be inculcated into an individual. These questions are particularly evident in the history of engineering codes and engineering ethics.¹³ For engineers, there remains a tension in the inherent definition of a profession as a group of experts explicitly set to serve the general public, and a profession’s own interests as a particular group.¹⁴ Extant conversations surrounding codes of ethics for computing mirror these disagreements.¹⁵ As Metcalf observes, the current surge of practitioner and public interest in ethics of data science was paralleled by a flurry of work in the early 1990s by major international professional organizations—including the Association of Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE)—to draft and implement ethics codes.¹⁶ Metcalf suggests ethics codes are often reactive, instituted in response to crises of professional confidence.¹⁷ Further, Effy Oz’s assessment of ethics codes from four professional associations involving computing,¹⁸ including the ACM, argued the lack of prioritization around moral obligations to various groups, common to professional ethics codes more generally, is especially pronounced in professional computing codes.¹⁹ As Frankel warns, ethics codes that fail to engage with broader social values and expectations risk becoming mere “political tools” for signaling moral virtue to society at large.

¹⁰Metcalf, “Ethics Codes: History, Context, and Challenges,” 863.

¹¹Frankel, “Professional Codes: Why, How, and with What Impact?”

¹²Frankel, “Professional Codes: Why, How, and with What Impact?” 109.

¹³Heinz C. Luegenbiehl, “Codes of Ethics and the Moral Education of Engineers,” *Business Professional Ethics Journal* 2, no. 4 (1983): 41–61.

¹⁴Michael Davis, “Thinking Like an Engineer: the Place of a Code of Ethics in the Practice of a Profession,” *Philosophy & Public Affairs* 20, no. 2 (1991): 150–67.

¹⁵Effy Oz, “Ethical Standards for Computer Professionals: a Comparative Analysis of Four Major Codes,” *Journal of Business Ethics* 12, no. 9 (September 1993): 709–26.

¹⁶Metcalf, “Ethics Codes: History, Context, and Challenges.”

¹⁷Metcalf, “Ethics Codes: History, Context, and Challenges.”

¹⁸Oz, “Ethical Standards for Computer Professionals: a Comparative Analysis of Four Major Codes.”

¹⁹Don Gotterbarn and James H Moor, “Virtual Decisions: Video Game Ethics, Just Consequentialism, and Ethics on the Fly,” *SIGCAS Computers and Society* 39, no. 3 (December 2009): 27–42.

Data & Data Science Metaphors

If professional codes of ethics are often “hard-won responses to major disruptions,” we should attend to the nature of the “disruption” in question. Doing so points towards previously underappreciated or overlooked ethical domains—in this case, domains which help us better come to terms with the rise of the data scientist and an epistemological shift in how we produce, understand, and act on knowledge about the world.²⁰ Here, we identify additional domains of ethical consideration relevant to data ethics by turning to the metaphors we use to talk about data itself.

The metaphors we deploy to make sense of new tools and technologies serve the dual purpose of highlighting the novel by reference to the familiar, while also obscuring or abstracting away from some features of a given technology or practice²¹—as Teun A. van Dijk describes,²² metaphors “are powerful means to make abstract mental models more concrete.”²³ In this way, metaphors—Rowan Wilken notes²⁴—are never innocent; they “always influence and shape the meanings that are generated by, and the meanings which accumulate around, a given metaphor.”²⁵ For example, as Dawn Nafus describes in the domain of data visualization,²⁶ the idea that data wants to be freed—itsself an offshoot of the earlier claim, “information wants to be free”²⁷—masks the labor expended in “freeing” data, especially in cases where data, in the words of her research subjects, are “stuck” or “disloyal.”

Cornelius Puschmann and Jean Burgess describe two predominating groups of

²⁰Rob Kitchin, “Big Data, New Epistemologies and Paradigm Shifts,” *Big Data & Society* 1, no. 1 (April 2014): 205395171452848-12, doi:10.1177/2053951714528481.

²¹Mark Nunes, “Baudrillard in Cyberspace: Internet, Virtuality, and Postmodernity,” *Style* 29 (1995): 314-27; Sally Wyatt, “Danger! Metaphors at Work in Economics, Geophysics, and the Internet,” *Science, Technology, & Human Values* 29, no. 2 (August 18, 2016): 242-61, doi:10.1177/0162243903261947; Rowan Wilken, “An Exploratory Comparative Analysis of the Use of Metaphors in Writing on the Internet and Mobile Phones,” *Social Semiotics* 23, no. 5 (November 2013): 632-47, <https://towardsdatascience.com/big-data-metaphors-we-live-by-98d3fa44ebf8>.

²²as Teun A. van Dijk, “Principles of Critical Discourse Analysis,” *Discourse & Society* 4, no. 2 (1993): 249-83; “Critical Discourse Analysis,” in *The Handbook of Discourse Analysis*, (West Sussex, UK: Wiley, 2015), 466-85.

²³van Dijk, “Critical Discourse Analysis,” 473.

²⁴Rowan Wilken, “An Exploratory Comparative Analysis of the Use of Metaphors in Writing on the Internet and Mobile Phones,” *Social Semiotics* 23, no. 5 (November 2013): 632-47, doi:10.1080/10350330.2012.738999.

²⁵Wilken, “An Exploratory Comparative Analysis,” 642.

²⁶Dawn Nafus, “Stuck Data, Dead Data, and Disloyal Data: the Stops and Starts in Making Numbers Into Social Practices,” *Distinktion: Journal of Social Theory* 15, no. 2 (June 17, 2014): 208-22, doi:10.1080/1600910X.2014.920266.

²⁷Stewart Brand, *The Media Lab*, (New York: Viking Penguin, 1987), 202.

metaphors around contemporary descriptions of data: 1) data as a natural force to be controlled and 2) data as a resource to be consumed.²⁸ As a resource, the authors note data are analogized to “food and fuel,” staple materials which “must be consumed to exist and to move forward rather than being consciously used.”²⁹ As a force of nature, the authors observe data are often described as being in a liquid state: “allusion to water [or oil]” they write, “supports the notion that data is all at once essential, valuable, difficult to control, and ubiquitous.”³⁰ From data lakes, rivers, and oceans to data floods, deluges, and tsunamis, these metaphors position data as something massive and volatile while also necessary to support human life.³¹ As Deborah Lupton notes, however, liquid metaphors also tacitly work to forestall ethical or regulatory interventions by positioning data as ubiquitous, uncontrollable, and resistant to transparency or accountability.³²

Both discursive strains—data as force and resource—point toward additional metaphors rooted in industrial production. As Sara M. Watson observes, many of our metaphors for data in the “knowledge economy” reference older industrial occupations.³³ Efforts to think through or propose strategies for managing data, in particular, rely on appeals to industrial imagery.³⁴ Descriptions of data as “toxic” or “radioactive” evoke images of massive nuclear facilities and radioactive waste management experts.³⁵

Curiously, as Tim Hwang and Karen Levy point out, often “people are nowhere to be found” in this landscape of data metaphors:³⁶ the excess of environmental

²⁸Cornelius Puschmann and Jean Burgess, “Big Data, Big Questions| Metaphors of Big Data,” *International Journal of Communication* 8 (2014).

²⁹Puschmann and Burgess, “Big Data, Big Questions,” 1700.

³⁰Puschmann and Burgess, “Big Data, Big Questions,” 1699.

³¹The now-ubiquitous metaphor of “cloud” computing has similar resonances, though it captures data as vaporous—somewhere between an invisible gas and a suffocating liquid.

³²Deborah Lupton, “Swimming or Drowning in the Data Ocean? Thoughts on the Metaphors of Big Data,” *The Sociological Life*, October 29, 2013.

³³Sara M. Watson, “Metaphors of Big Data,” *DIS Magazine*, May 28, 2016.

³⁴Kailash Awati and Simon Buckingham Shum, “Big Data Metaphors We Live by,” *Towards Data Science*, May 14, 2015.

³⁵Cory Doctorow, “Why Personal Data Is Like Nuclear Waste,” *The Guardian*, January 15, 2008.) Metaphors referencing more quotidian forms of waste also point toward the similarly quotidian forms of work involved in data science: are those who work with data “rock stars” or “data janitors”—or both? (Lilly Irani, “Justice for ‘Data Janitors,’” *Public Books*, January 15, 2015.

³⁶There is one notable—and noxious—analogy which does foreground people: “big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it....” Posted to Facebook in 2013 by Dan Ariely, Professor at Duke University, the phrase has since been popularized in numerous blog posts, news articles, and talks. We have deliberately omitted this analogy, however, as we find it more problematic than insightful. Specifically, we do not wish to contribute to the sexualization of technological work—a process * often serves to exclude or marginalize women in STEM—or engage in the belittling of young persons’ sexual development. While other work may find constructive connections between

metaphors for data “do us a disservice by masking the human behaviors, relationships, and communications that make up all that data we’re streaming and mining.”³⁷ By analogizing digital data as a part of the “natural world”—the latter term itself a techno-scientific reification and justification of centuries of imperial and settler colonial exploitation³⁸—the status of data as a record of human activity is doubly occluded.³⁹

Method and Analysis

To explore the relationships between data science, data metaphors, and professional ethics, we undertook a comparative discourse analysis of two sets of text-based data: 1) the text of ethics codes conventionally accepted as relevant to data science today and 2) the text of ethics codes from professions associated with popular or dominant data metaphors. Codes for both sets were accessed through the Center for the Study of Ethics in the Professions, which indexes more than 2500 codes and guidelines from over 1500 organizations, dating from 1887 to the present.⁴⁰ For our first set of codes, we drew from dominant professional associations in information and computing fields, namely the Association of Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE). We also identified major professional associations in the areas of mathematics and statistics, like the American Statistical Association. This set of codes provide a baseline of historically salient professional and ethical concerns in domains which—at least on the surface—directly inform the development, training, and backgrounds of today’s data scientists.⁴¹

For the second set, we selected codes that 1) spoke directly to our identified metaphors/concepts and 2) were of sufficient length and substance to effectively compare with the well-developed codes in our first set. For example, if data is often framed as a natural resource, then we considered professions engaged in the

sexual ethics and data ethics broadly (see, for example, work on the ethics of consent: Una Lee and Dann Toliver, *Building Consentful Tech*, Toronto, ON: And Also Too, 2017.), we do not find this particular analogy productive.

³⁷Hwang and Levy, “‘The Cloud’ and Other Dangerous Metaphors.”

³⁸Kavita Philip, “Nature, Culture Capital, Empire,” *Capitalism Nature Socialism* 18, no. 1 (March 2007): 5-12, doi:10.1080/10455750601164584.

³⁹Irina Raicu, “Metaphors of Big Data,” *Recode*, February 26, 2016.

⁴⁰The CSEP’s Ethics Code Collection is available at <http://ethicscodescollection.org>. For further information on the CSEP’s collection and update policy, please see <http://ethics.iit.edu/ecodes/node/5421>.

⁴¹Jeff Hammerbacher, “Information Platforms and the Rise of the Data Scientist,” in *Beautiful Data: the Stories Behind Elegant Data Solutions*, ed. Toby Segaran and Jeff Hammerbacher, (Sebastopol, CA, 2009), 73-84.

extraction and stewardship of resources, as with forestry; if big data is considered “radioactive,” then we looked to professionals tasked with the storage, management, and safekeeping of radioactive or toxic materials. Data collection was further informed by our own knowledge of terms and metaphors, derived from our own experiences working with and educating data scientists (for example, commonly employed terms like “cleaning,” — i.e. processing—data) and studying the ethics and rhetoric of data-intensive platforms more broadly.⁴² We also drew on discussions of predictive analytics equating practices of data-driven forecasting and prediction with the work of astrologers.⁴³ Finally, inspired by descriptions of online data as the material traces humans leave behind (akin to, say, dead skin cells), we drew on the ethics of funeral directors and morticians charged with overseeing and caring for human remains.

In total, we selected 20 ethics codes for analysis. These codes ranged in length from approximately 330 to 4300 words, with a mean length of approximately 1300 words and a median length of approximately 800 words. Many of the codes were undated, with those with dates ranging from 1992 to 2016 (see Appendix A for details). We used the qualitative research software ATLAS.ti to aid in manual qualitative coding of the corpus according to principles of critical discourse analysis (CDA).⁴⁴ We focused, in particular, on CDA’s commitment to mark both the *social actors* and *social actions* explicit and implicit in a given text.⁴⁵ As professional codes aim to regulate (to varying degrees) the practice of professionals in a given domain, we were interested first in identifying those actors (individual or institutional) and actions (encouraged or prohibited) conceived of as relevant by a given code. In addition, we also identified and labeled references to any explicit value or virtue (i.e., “fairness,” “honesty,” “integrity”), allowing us to connect specific values to specific actors or practices. As such, our analysis was “directed,” as our coding scheme was derived both before (the actor/action frame adopted prior to coding) and during (values identified while coding) the analysis.⁴⁶

⁴² Oliver L. Haimson and Anna Lauren Hoffmann, “Constructing and Enforcing ‘Authentic’ Identity Online: Facebook, Real Names, and Non-Normative Identities,” *First Monday* 21, no. 6 (June 6, 2016), doi: <http://dx.doi.org/10.5210/fm.v21i6.6791>; Anna Lauren Hoffmann, Nicholas Proferes, and Michael Zimmer, “‘Making the World More Open and Connected’: Mark Zuckerberg and the Discursive Construction of Facebook and Its Users,” *New Media & Society* 12, no. 1 (January 13, 2017): 146144481666078-20, doi:10.1177/1461444816660784.

⁴³ Maude Standish, “Data Is the New Astrology,” *The Huffington Post*, April 25, 2013; Kate Crawford, “Asking the Oracle,” in *Laura Poitras: Astro Noise*, (New York/New Haven, CT, 2016), 138-53.

⁴⁴ Theo van Leeuwen, *Discourse and Practice: New Tools for Critical Discourse Analysis* (Oxford Studies in Sociolinguistics), Oxford: Oxford University Press, 2008.

⁴⁵ van Dijk, “Critical Discourse Analysis,” 466.

⁴⁶ H. F. Hsieh, “Three Approaches to Qualitative Content Analysis,” *Qualitative Health Research* 15, no. 9 (November 1, 2005): 1277-88, doi:10.1177/1049732305276687.

Analysis

Despite the broad range of sampled fields and subject areas, ethics codes in the corpus demonstrated considerable thematic overlap. Much of this overlap is a matter of genre—ethics codes are recognizable as such precisely because they share certain schematic and substantive markers.⁴⁷ For example, prohibitions on maintaining conflicts of interest were common across all codes, regardless of professional domain. In the following, we focus less on thematic overlap and instead on *social actors* and *social actions* explicit or implicit within the codes, noting 1) where the codes from different domains diverge and 2) and how each relates to representative virtues or other forms of ethical behavior.

Social Actors

Professional ethics codes often position the individual practitioner as the primary locus of ethical responsibility.⁴⁸ Individuals are asked to take the brunt of ethical conflict and adjudicate between potentially conflicting or incommensurate values.⁴⁹ The codes we examined followed this pattern, foregrounding the individual practitioner as a primary social actor through an emphasis on personal responsibility. Only rarely were codes extended to cover responsibilities of others, especially those above, around, or below an individual professional. References to individual professional competence also appear frequently across both sets of codes: individual professionals are directed to take on work or perform tasks only within their specific domain (or sub-domain) of expertise, and to not falsify or misrepresent their credentials.

What did vary, however, was the scope and range of social actors considered relevant to a professional context. The absence of considerations for third parties in most of the codes was notable, especially for professions where a code's stated professional ethical commitments could easily conflict with—or had evident impacts on—third parties. While a number of the sampled codes state commitments both to the general public and to the interests of the employer or

⁴⁷ Gaumnitz and Lere, "A Classification Scheme for Codes of Business Ethics."

⁴⁸ Abbot, "Professional Ethics."

⁴⁹ Ethics codes thus perform for professionals in the abstract what the interaction design of many digital platforms does for gig workers in practice: figure their subjects as "moral liability sponges" or "moral crumple zones." See Alex Rosenblat and Luke Stark, "Algorithmic Labor and Information Asymmetries: a Case Study of Uber's Drivers," *The International Journal of Communication* 10 (July 27, 2016): 3758-84; and M. C. Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction" (We Robot 2016) (March 20, 2016). We Robot 2016 Working Paper. Available at SSRN: <https://ssrn.com/abstract=2757236> or <http://dx.doi.org/10.2139/ssrn.2757236>

profession more narrowly, the morality of a profession's or an employer's motives are not scrutinized, and the individual has no guidance on how to navigate moral conflicts. One exception was the Association of Computer Machinery's ethics code, which contained explicit provisions around whistleblowing—likely inspired by the historical relationship between computing, state and corporate surveillance, and privacy harms.⁵⁰

Other actors—sometimes people, other times objects, spaces, or ideals—are evoked across these codes through what we refer to as “responsibility to” language. At any given time, an individual professional may have (per their relevant code) a responsibility to: specific other actors (clients, colleagues, users, employees, or employers); general others (the public, society, or simply “others”); a political body (nation, country); specific objects or ideals (technology, truth, knowledge, the profession); or abstract spaces or sites (nature). The particular discursive arrangement of these other actors hints at a given code's imagined scope. Those addressing only interpersonal or work relationships with colleagues, employers, or employees could be understood as relatively narrow in scope, suggesting the body responsible for developing the code either 1) did not imagine the profession as playing a broad social role or 2) wanted to actively avoid offering moral guidance on issues beyond the immediate workplace.

Beyond colleagues and employer/employee relationships, some of the surveyed codes cited specific responsibilities to users (presumably of a particular system and not of technology generally). This focus was, unsurprisingly, a common theme in computing and information codes, where broader ethical commitments were often treated as high-level abstractions and more precise attention was paid to intra-professional considerations. Accordingly, social issues and problems of the public good were often reduced to their most technocratic features. For example, the Code of Ethics of the American Society for Information Society & Technology (ASIS&T) begins by urging “its members to be ever aware of the social, economic, cultural, and political impacts of their actions or inaction,” yet proceeds to detail explicit responsibilities to employers, clients, and system users—and not society more broadly.

Other information and computing codes also evoked such “general others” briefly. The IEEE Code of Ethics merely implores professionals “to accept responsibility in making decisions consistent with the safety, health, and welfare

⁵⁰David Lyon, ed., *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*, (New York: Routledge, 2003); Herman Tavani, “Privacy Enhancing Technologies as a Panacea for Online Privacy Concerns: Some Ethical Considerations,” *Journal of Information Ethics* 9, no. 2 (2000); Helen Nissenbaum, “Respecting Context to Protect Privacy: Why Meaning Matters,” *Science and Engineering Ethics* 109, no. 4 (July 11, 2015): 1–22, doi:10.1007/s11948-015-9674-9.

of the public, and to disclose promptly factors that might endanger the public or the environment.” The Association of Computing Machinery (ACM) was again an exception, with its code going into detail regarding “general moral imperatives” and their specific features. For example, the broad stated obligation to “contribute to society and human well-being” is followed by the somewhat more precise imperative to “protect fundamental human rights” and “respect the diversity of all cultures.” While still vague, this explication does offer some insight into how the ACM conceives of “human well-being.”

In contrast to information and computing codes, codes related to statistics and statistical work consistently foregrounded professional responsibilities to the public or society. The Statistical Society of Canada lists responsibilities to society first while the Royal Statistical Society (UK) singles out “an overriding responsibility to the public good.” The American Statistical Association (ASA) goes the furthest in making statisticians’ responsibility towards society explicit, averring that “the discipline of statistics links the capacity to observe with the ability to gather evidence and make decisions, providing a foundation for building a more informed society.” The ASA also details a set of commitments aiming to limit or reduce the chance an individual practitioner will be placed in a morally tenuous position by a third party, for example by “[striving] to protect the professional freedom and responsibility of statistical practitioners,” especially in cases where “analyses...are known or anticipated to have tangible physical, financial, or psychological impacts.” Further, it asks employers to “recognize that the results of valid statistical studies cannot be guaranteed to conform to the expectations or desires of those commissioning the study or the statistical practitioner(s).” While the ASA code does not stipulate further ethical commitments the employer may have, it does to some degree acknowledge and incorporate the how statistical work interacts with corporate or other institutional processes.

In addition, these statistics-based ethics codes incorporate ethical concerns from the broader history and trajectory of research ethics. The ASA code specifically details an obligation to seek consent, especially for secondary or indirect uses of data. It also gestures towards concerns of justice in research ethics, noting, “statistical descriptions of groups may carry risks of stereotypes and stigmatization.” In response, “statisticians should contemplate, and be sensitive to, the manner in which information is framed so as to avoid disproportionate harms to vulnerable groups.” Given these codes are intended to inform statisticians’ research practices, it is perhaps unsurprising the treatment of research subjects would feature prominently as objects of ethical concern - but makes it all the more notable that data scientists, who as Beaton notes share a lineage with both statisticians and

psychologists, have yet to fully embrace such language.⁵¹

Codes associated with data metaphors often suggested responsibility to broadly construed spaces, sites, or ideals through the language of stewardship. For instance, the Society of American Foresters defines the forestry profession as “[serving] society by fostering stewardship of the world’s forests,” while the North American Nature Photography Association articulates its ethics through principles of stewardship of and non-interference with various natural habitats. Professions charged with sanitation, cleanliness, or dealing with hazardous waste also address environmental concerns. For Certified Hazardous Materials Managers, the “primary responsibility is to protect the public and the environment.” For the International Sanitary Supply Association (ISSA), public health is paramount—all commercial considerations are, according to their code of ethics, secondary to this broader concern.

In contrast to stewardship-based or environmental models, which tend to refer to objects in the material world, computing and statistics codes tend to foreground a specific value or ideal and then work to situate professional ethical responsibilities in service of this higher ideal. In statistics codes, for example, “truth” and “objective knowledge” emerge as both quasi-deontological ends-in-themselves and contested categories professionals must strive to realize in their work. Section A of the American Statistical Association specifically addresses statisticians’ obligations to reduce, mitigate against, or eliminate biases—either on the part of the professional or the contracting party—which might skew research results. An ethical statistician, the Code asserts, “uses methodology and data that are relevant and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results.” The unethical statistician, then, is one who manipulates data or skews findings in ways that are self-serving or aims to deliberately mislead or manipulate others.

Finally, there was also variation in the codes with regard to the positioning of ethics relative to a given professional. In a few cases, ethics was construed not as a set of commitments, but as constitutive of professional identity. For example, the code of ethics for the Association of Information Technology Professionals (AITP) states the standards set out by the code “are not objectives to be strived for, they are rules that no true professional will violate.” Though seemingly minor, this difference has consequences for professional identity and regulation: ethics as professional commitment implies one can remain a “professional” even when acting in ways potentially construed as unethical, while, conversely, ethics

⁵¹ Brian Beaton, “How to Respond to Data Science: Early Data Criticism by Lionel Trilling,” *Information & Culture: a Journal of History* 51, no. 3 (August 2016): 352–72. doi:10.7560/IC51303.

as constitutive of professional identity implies that a breach of ethics simultaneously disqualifies one's status as a professional.

Values and Social Actions

Where the previous section focused on actors and objects, this section focuses on specific values and actions captured or implied in the surveyed ethics codes. Not only do ethics codes represent an effort to define and delimit an ideal ethical professional, they also work to articulate, set, and scope out the sorts of actions and values endemic to a particular profession. These actions often perform justificatory work (e.g., positioning certain actions as socially useful helps justify a profession's existence) or work to stave off certain types of attention (e.g., clearly stated prohibitions on certain actions may help deflect public or regulatory scrutiny). Additionally, stated values reveal something about the anticipated impact of a given action—positioning something as potentially discriminatory, for example, demonstrates an awareness of a profession's political impact. In this way, references to specific values or virtues offer insight into the ethical tenor of actions an individual might be expected to undertake or account for in their capacity as a professional.

Within the inventory of values laid out in the codes we analyzed, some of the starkest divisions came between computing and statistics codes, and those codes associated with data metaphors. A dominant theme across all analyzed codes was a proscription against what we term “biopolitical harms”: environmental harms, and the health and safety of populations.⁵² This emphasis suggests a parallel to the preponderance of resource-based metaphors for data—namely a sense that, in the abstract, it is easy for professionals to agree to protecting (or perhaps, managing) the natural world and natural resources, as part of what Gabriel Abend terms “the moral background” to professionalism itself.⁵³ In contrast, computing codes and statistics codes often foreground political issues of privacy and freedom of speech, as well as conceptions of data as confidential and requiring safeguarding. This attention to privacy, security, and speech grows out of a longer-standing focus on these areas during the development of information technology and computing

⁵²Patricia Ticineto Clough and Craig Willse, “Gendered Security/National Security: Political Branding and Population Racism,” March 7, 2013; Michel Foucault Collège de France, *The Birth of Biopolitics: Lectures at the Collège De France, 1978-1979*, 1st ed., (New York: Picador, 2010); Matthew B Sparke, “A Neoliberal Nexus: Economy, Security and the Biopolitics of Citizenship on the Border,” *Political Geography* no. 25, no. 25 (2006): 151-80, doi:10.1016/j.polgeo.2005.10.002.

⁵³Gabriel Abend, *The Moral Background* (Princeton, NJ and Oxford: Princeton University Press, 2014).

ethics.⁵⁴ Save for limited references to confidentiality in the codes of The Wildlife Society and Society of American Foresters, privacy and speech considerations are absent in our second set of codes.

Given established historical connections between information technology, communication, and privacy, the emphasis on the latter in computer science codes makes sense. However, it also points to one of the starkest differences between the conventional codes and those codes rooted in our data metaphors: the former tends to foreground abstract users, rights, and values, while many of the metaphor-based codes explicitly revolve around material stewardship, public health, and bodily safety. For example, the Code of Ethics for Hazardous Materials Managers explicitly states that professionals' "primary responsibility is to protect the public and the environment," while "the interests of individual clients and employers must be secondary." In a different way, the code of ethics for The Wildlife Society places "research and scientific management of wildlife species, their environments, and stakeholders" as primary and specifically directs professionals to educate broadly construed others on these topics. This contrast between abstract values in the conventional codes and material stewardship in the metaphor-based codes lends support to the idea that conventional codes' outsized focus on issues like privacy and speech—themselves heavily informed by the technical rather than social affordances of computers—has limited their ability to account for a wider range of potential social, political, and environmental harms.⁵⁵ A focus on natural resources as a site for both conservation and potential harm also helps professionals avoid the social complexities and challenges of adjudicating their responsibilities to society as comprised of other people, instead of a material background.

Finally, a number of the sampled codes formally prohibit discrimination, and some others contain explicit lists of those classes of people deserving of particular protection. However, only the code of ethics for the National Funeral Directors Association made an explicit connection between the mission of the profession

⁵⁴Tavani, "The State of Computer Ethics as a Philosophical Field of Inquiry;" Frances S. Grodzinsky and Herman T. Tavani, "Applying the 'Contextual Integrity' Model of Privacy to Personal Blogs in the Blogosphere," *International Journal of Internet Research Ethics* 3 (December 30, 2010): 38-47; Lucas D. Introna, "Maintaining the Reversibility of Foldings: Making the Ethics (Politics) of Information Technology Visible," *Ethics and Information Technology* 9, no. 1 (December 21, 2006): 11-25, doi:10.1007/s10676-006-9133-z.

⁵⁵Literature on the concept of "securitization" supports this view: when topics of public debate are securitized, they are removed from the realm of political contestation and deemed to be technical matters addressable by experts. See Lene Hansen and Helen Nissenbaum, "Digital Disaster, Cyber Security, and the Copenhagen School," *International Studies Quarterly* 53 (2009): 1155-75; Helen Nissenbaum, "Where Computer Security Meets National Security," *Ethics and Information Technology* 7 (2005): 61-73, doi:10.1007/s10676-005-4582-3.

and a specific vulnerable class of people.⁵⁶ Specifically, people of low socioeconomic standing or limited financial means are cited as deserving particular consideration in view of funeral directors' mission to "to provide families with meaningful end-of-life services at the highest levels of excellence and integrity."⁵⁷ This commitment is explicitly accounted for in a prohibition against withholding services (like delaying the embalming process) or the body of a loved one (from release to a family or other legally recognized party) until payment for services has been received. While clearly rooted in a concern over extortionate business practices and hucksterism—as well as the time sensitive nature of some features of morticians' work—there is, perhaps a lesson for data scientists here in how to think about specific power dynamics at play. For example, a code of ethics for data scientists might be careful to not make certain kinds of data or informational transparency contingent on an ability to pay or by coercing users to give up more personal data in the process.

Discussion & Recommendations

Our analyses of professional ethics codes from both computer science and from metaphorically related fields suggest several broad conclusions relevant to conversations about data in the public sphere, and to data science as an emerging profession.

First, one of the chief values common to codes from across our sample was fairness. In the context of computer science ethics, fairness is a hot topic: computer scientists interested in fairness, accountability, transparency and other values in machine learning and artificial intelligence have already begun to interrogate how fairness might be translated into computational metrics for evaluating algorithms and systems.⁵⁸ A parallel movement in science and technology studies and information science has focused on the need to articulate fairness not only as an allocative metric internal to digital systems, but also as social one addressing systematic discrimination, bias, and representational harms.⁵⁹ In the context

⁵⁶It is worth noting that The Wildlife Society does address unspecified "stakeholders" in the well-being of certain wildlife species. These stakeholders could easily include oppressed individuals or groups deserving of particular consideration, especially native or indigenous peoples. However, these groups are not specifically named in the code.

⁵⁷"About NFDA," *National Funeral Directors Association*, June 9, 2018.

⁵⁸Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104 (2016): 671-732. doi:10.15779/Z38BG31; Chelsea Barabas, Madars Virza, Karthik Dinakar, Joichi Ito, and Jonathan Zittrain, "Interventions Over Predictions - Reframing the Ethical Debate for Actuarial Risk Assessment," *Proceedings of Machine Learning Research* 81:1-15, 2018.

⁵⁹Anna Lauren Hoffmann, "Data Violence and How Bad Engineering Choices Can Damage Soci-

of professional ethics codes, however, fairness takes on yet another meaning: as a sign of interpersonal trust within a profession, and a signal for professional frankness and honesty.

The strong emphasis on fairness within the conventional codes we sampled may in fact be connected to the technical definitions of fairness increasingly prominent in computer science literature: technical systems are perceived by computer scientists and data scientists as objects which are the fruit of expert processes of collaboration and communication, and so those same expert processes would, in theory, be able to similarly produce technical definitions of values—such as fairness. The strong emphasis on computational solutions to values questions has much to do with common areas of professional familiarity, expertise and technical language among participants. Fairness is a complicated concept in non-computational contexts as well as computational ones. We argue, however, a reemphasis on fairness as a social value (in line with honesty, frankness, and “good faith”) would help reorient data science and related professions towards their broader social role.

The professional codes we sampled—from CS and from other fields—are also highly focused on technical and professional credibility. This second insight is related to our first point above: professional fairness and honesty are integral to technical credibility and recognition of expertise. However, scholarly work in the sociology of trust that observes trust as a social construct is only partially focused on technical credibility, or whether a person or institution is able to perform the task they claim they will do.⁶⁰ While the voluminous literature on trust is too broad to more than briefly sample here, the other key aspect of social trust is benevolence: the notion that a trusted person or organization has the best interests of the trustor at heart.⁶¹ The emphasis across professional codes for computer science is credibility, and less so benevolence.

Given an increasing scholarly and public awareness of the social impacts of big data, ML, and AI, the lack of attention to social benevolence in the conventional

ety,” *Medium* (Member Feature Story), 30 April 2018, available at <https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4>

⁶⁰Linda D. Molm, Nobuyuki Takahashi, and Gretchen Peterson, “Risk and Trust in Social Exchange: an Experimental Test of a Classical Proposition,” *The American Journal of Sociology* 105, no. 5 (March 2000): 1396-1427; Susan P. Shapiro, “The Social Control of Impersonal Trust,” *American Journal of Sociology* 93, no. 3 (November 1987): 623-58; Guido Möllering, “The Trust/Control Duality,” *International Sociology* 20, no. 3 (June 29, 2016): 283-305, doi:10.1177/0268580905055478.

⁶¹Matthew K. O. Lee and Efraim Turban, “A Trust Model for Consumer Internet Shopping,” *International Journal of Electronic Commerce* 6, no. 1 (October 1, 2001): 75-91, doi:10.2307/27751003?ref=search-gateway:65426dfa915042907ae6237cf88297e1; David Gefen, Izak Benbasat, and Paula Pavlou, “A Research Agenda for Trust in Online Environments,” *Journal of Management Information Systems* 24, no. 4 (May 12, 2008): 275-86, doi:10.2753/MIS0742-1222240411.

codes is notable and in need of remedy. Here, insights rooted in our data metaphors stand to be of particular use. For certain professions concerned with, for example, sanitation or the handling of hazardous materials, benevolence (often in the form public or environmental safety) is paramount—it works as an overarching frame for judgments of professional duty and competence. Accordingly, we should ask critical questions about how broadly or narrowly “competence” in particular data scientific projects should be construed. For example, while a data scientist might have expertise in particular computational and statistical methods, they may know very little—in the social scientific sense—about the particular communities or behaviors captured in a given dataset. Interpreting competence in view of social benevolence would require some knowledge of particular communities, behaviors, or broader social and political forces (or at least a requirement to contract or seek out collaborations with other professionals or community members who do possess such expertise). In certain contexts, it may even be useful to follow the lead of morticians and funeral directors in codifying historically-salient social or political groups deserving of particular consideration. Though not represented in our surveyed codes, the activist slogan “nothing about us without us” (notably espoused by disability activists in the United States) may be a useful starting point for reasoning about data scientists’ social obligations.⁶²

In addition, ethical commitments addressing both the employer and the employed, as well as possible conflicts between those ethical commitments, are particularly relevant for thinking through the ethics of data science within large corporate settings. Economic incentives and ethical commitments to the privacy, security and labor rights of both data subjects and employees often come into conflict⁶³—for example, when data or scientific work might contribute to the development of military weapons and targeting systems. How to harmonize broad social commitments with both individual professional practice and corporate structures in the data science context—especially since, given the heavy reliance on gig or crowd labor in data science, it remains an open question just who the individual professional figured across the codes we surveyed might be—is a necessary subject for further research and activism.⁶⁴ Related conversations around the ethics and governance of AI provide an instructive parallel case.⁶⁵ Data gov-

⁶²James I. Charlton, *Nothing About Us Without Us: Disability Oppression and Empowerment*, (Berkeley and Los Angeles: University of California Press, 1998).

⁶³Noëmi Manders-Huits and Michael Zimmer, “Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities,” *International Review of Information Ethics*, January 30, 2009, 1-8.

⁶⁴Kate Conger, “Google Plans Not to Renew Its Contract for Project Maven, a Controversial Pentagon Drone AI Imaging Program,” *Gizmodo*, June 1, 2018.

⁶⁵Daniel Greene, Anna Lauren Hoffmann, and Luke Stark, “Better, Nicer, Clearer, Fairer: A Crit-

ernance through the lens of human rights is one possible approach in this vein, though the suitability and effectiveness of human rights discourse in mobilizing good corporate governance and change for social justice is an open question.⁶⁶

A related insight drawn from our analysis concerns the notion of the fiduciary, and their “fiduciary duty.” A fiduciary is a trustee, and a fiduciary duty is the legal obligation of one party to act in the best interest of another. Doctors and lawyers, for instance, have fiduciary duties to their patients and clients—so do other professionals such as wealth managers and morticians. Legal scholars Jack Balkin and Jonathan Zittrain have recently developed the concept of the “information fiduciary,”⁶⁷ entailing data scientists and the companies they work for—like Facebook and Google—accepting “the duty to use personal data in ways that don’t betray end users and harm them.”⁶⁸

We see several points from our analysis pertaining to the information fiduciary model. One is that for the most part, professions with fiduciary duties protect personal data as a secondary result of their primary duty of care: to the health of a patient, to (in theory) the law and justice, to the bodies of the dead. For example, doctors and morticians, more so than lawyers, are intimately engaged in caring for us as embodied agents. Standards for trusting doctors and morticians are thus high: we imagine a doctor, on the average, to be honest—both credible and benevolent—in a way we may not (alas) imagine a computer scientist.

Data scientists and their employers are generally presented as being in the business of collecting and analyzing data as a primary goal—but this is of course an erroneous view. Data scientists and computer programmers invariably work in and across various domains and their work has, in many cases, obvious material consequences—from precision medicine to criminal sentencing. Here, the emphasis on material stewardship found in metaphor-based codes of ethics, as opposed to the valorization of abstract ideals found in many computer science codes, shows how the material objects of a profession’s attention can be obscured by a profession’s conceptual axioms. Data scientists as a profession are generally associated with abstraction and disembodiment, with numbers aggregated and immaterial

ical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning,” Hawaii International Conference on System Sciences (HICSS 2019), Maui, HI.

⁶⁶Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights and Dignity*, Data & Society Research Institute, (10 October 2018), available at <https://datasociety.net/output/governing-artificial-intelligence/>; Ron Dudai, “Human Rights in the Populist Era: Mourn then (Re)Organize,” *Journal of Human Rights Practice* 9, no. 1 (February 2017): 16–21. <https://doi.org/10.1093/jhuman/hux005>

⁶⁷Jack M. Balkin, “Information Fiduciaries and the First Amendment,” *UC Davis Law Review* 49, no. 4 (April 2016): 1183–1234.

⁶⁸Jack M. Balkin and Jonathan Zittrain, “A Grand Bargain to Make Tech Companies Trustworthy,” *The Atlantic*, October 3, 2016.

in “the cloud.” Yet the real-world consequences on individual embodied humans of data science can be real, visceral, and devastating. Likewise, discourses of data as a force, a resource, and as an industrial product erase the human subjects both of digital data collection and accumulation and its effects; yet they also lack the traditions of stewardship and responsibility, which however imperfectly typify the professional discourses in those fields.

Analogizing digital data as “natural” without stewardship discourses implicitly signals data - and the living people it involves - are open for rank exploitation. As Arvind Narayanan argues, “it’s not enough to ask if code executes correctly. We also need to ask if it makes society better or worse.”⁶⁹ The seeming abstraction of data is in some ways belied by the material metaphors used to describe it: water, oil, and ore. Yet as Hwang and Levy note, these metaphors, though material, deflect attention from the fact human data are produced and have impacts on human beings. We heartily concur with Rebecca Lemov: “‘Big data is people!’”⁷⁰ As such, we argue data scientists should be held to fiduciary standards closer to those of doctors, morticians, or even hazardous waste managers—they are professionals dealing with the human bodies and populations in ways that demand a high degree of both credibility and benevolence. Professional ethics in data science should include broader, more ecological thinking about the role of data, computing, and statistical analyses in their social context and real-world applications.

One final area where professional ethics codes related to data metaphors differed from those in computer science was around the question of professional sanction. For instance, the ethical code of the International Society of Petroleum Engineers has an explicit clause encouraging practical mitigating action, and also states in the code itself the mechanism through which members can report violations to the professional body. If data is indeed the new oil, the IEEE and other data science codes should be at least as explicit as that of petroleum engineers in listing consequences for violations of the code and articulating how those violations can be reported.

As noted above, there is one professional group thematically related to data science whose professional ethics codes demonstrate a relatively high degree of sensitivity to a variety of actors, societal values and broad-based social responsibility: statisticians. Given the centrality of statistical analysis to data science, we suggest data scientists could do much worse than having a full-throated professional engagement with statistical codes of conduct, and with the legacy of statistics

⁶⁹Doug Hulette, “Patrolling the Intersection of Computers and People,” *Princeton University Computer Science*, October 2, 2017.

⁷⁰Rebecca Lemov, “‘Big data is people!’” *Aeon Magazine*, June 16, 2016.

as a tool in social science research. Like data scientists, statisticians work with many other professions, but nonetheless have articulated their social obligations. Future scholarship engaging statisticians around the concept of “fiduciary duty” would be valuable—a statistical fiduciary duty of care is one conceptual way to understand the professional duties of data scientists and others engaging in data analysis.

Conclusion

Our analytic goals in juxtaposing ethics codes in computer science and related fields alongside codes related to data science metaphors are threefold. As scholars working at the intersection of information science, science and technology studies (STS), and the philosophy and ethics of technology, we have an intellectual interest in tracing the conceptual connections and sociological themes of computer and data science, the professions which to a large degree shape the technologies and social practices around them which we study. Ethics codes are a micro-level instantiation of broader structural and institutional values and debates. These codes represent “the transformation of social practices into discourses about social practices”;⁷¹ they are not absolute or perfect representations of these discussions but sites where certain results of certain political, professional, and material struggles are stabilized and put to work in the world.

However, we are also concerned with two other, practical normative outcomes: a statement of the kinds of ethos we as critical scholars want to see permeate these emerging data cultures; and a concomitant sense of how data scientists might consider changing existing or nascent professional ethics codes in data science in light of our findings and our broader normative commitments.

While many of our proposals are spelled out in the Discussion section above, we make two more general recommendations for data scientists and those interested in professional ethics for data science. The first recommendation is to take “ethics” as a starting point, not as an end. Codes of ethics increasingly serve to demarcate data culture as a domain of experts, and conversations around professional ethics in data science and related fields such as ML/AI are a necessary but absolutely insufficient condition for the kinds of progressive, just and equitable social outcomes we seek for the world. Testifying before the US Congress in 2003 about another new and disruptive technical field, nanotechnology, STS scholar Langdon Winner noted a historical tendency, as he saw it, “for those who

⁷¹Theo van Leeuwen, *Discourse and Practice: New Tools for Critical Discourse Analysis* (Oxford Studies in Sociolinguistics), Oxford: Oxford University Press, 2008, 105.

conduct research about the ethical dimensions of emerging technology to gravitate toward the more comfortable, even trivial questions involved, avoiding issues that might become a focus of conflict.”⁷² Digital technologies are powerful tools: for data scientists to insufficiently engage with the rich social contexts and disparate, sometimes conflicting human realities of their use is a lapse of a core ethical obligation, courage, in itself. To limit conversations about the societal impacts and obligations of data science solely to professional ethics is a mistake we are keen to have all parties involved avoid.

Our second recommendation concerns the relationship between professional ethics codes in data science and data science pedagogy. The ethical norms of a profession emerge out of every stage of that profession’s training process. As such, we advocate forcefully for data science education to address not only the professional ethics questions posed by extant professional codes, but also the societal questions posed by the metaphors through which the profession, and discourse more broadly, understands data. These metaphors can serve as what Katie Shilton⁷³ terms “values levers,” prompting novel conversations about data science’s impacts and responsibilities to society at large. Innovative work on broadening the terms of data science education is already underway.⁷⁴ Data science and data scientists would benefit from expanding their collaborations with interdisciplinary work in STS, information studies, and media studies to further reap the benefits of engaging with values, ethics, and norms at every stage of their work.⁷⁵

Finally, our ethical commitments as authors are grounded in a desire for data justice,⁷⁶ design justice,⁷⁷ as well as the recognition of historical injustices in emerging data cultures and in society more broadly. As Kate Crawford has put it, data

⁷²Langdon Winner, Testimony to the Committee on Science of the U.S. House of Representatives on The Societal Implications of Nanotechnology, 9 April 2003, available at <http://homepages.rpi.edu/~winner/testimony.htm>.

⁷³Katie Shilton, “Values Levers: Building Ethics Into Design,” *Science, Technology, & Human Values* 38, no. 3 (May 2013): 374–97, doi:10.1177/0162243912436985.

⁷⁴Michael Skirpan et al., “Ethics Education in Context,” (the 49th ACM Technical Symposium, New York, New York, USA: ACM Press, 2018), 940–45, doi:10.1145/3159450.3159573; [Author, 2018]; Costanza-Chock, “Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice.”

⁷⁵Mary Flanagan and Helen Nissenbaum, *Values at Play in Digital Games*, (Cambridge, MA: The MIT Press, 2014); Phoebe Sengers et al., “Reflective Design,” (Proceedings of the 4th Decennial Aarhus Conference, Aarhus, Denmark, 2005), 49–58; Batya Friedman, Peter H. Kahn, and Alan Bornring, “Value Sensitive Design and Information Systems,” in *Human-Computer Interaction in Management Information Systems: Foundations*, ed. B. Schneiderman, Ping Zhang, and D. Galletta, (New York: M.E. Sharpe, Inc., 2006), 348–72.

⁷⁶Hoffmann, “Data Violence.”

⁷⁷Sasha Costanza-Chock, “Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice,” 2018, 1–14, doi:10.21606/dma.2017.679.

ethics needs to ask, “What kind of world do we want to live in?”⁷⁸We are explicit about these normative commitments as a way to conceptually tax related ethics codes and our data metaphors alike. With data-driven online platforms and digital systems already a potential source of bias and discrimination,⁷⁹the processual ethics common to professional codes need to be supplanted by a more explicit set of norms around data cultures as spaces for equality and justice—within and beyond a code of ethics for data scientists.

Appendix: List of Professional Ethics Codes Consulted

Organization	Date Last Updated	Word Count
American Federation of Astrologers (AFA) Code of Ethics	No date listed	356
American Mathematical Society Policy Statement on Ethical Guidelines	2005	1729
American Society for Photogrammetry and Remote Sensing Code of Ethics	No date listed	658
American Statistical Association (ASA) Ethical Guidelines for Statistical Practice	2016	3478
Association for Computing Machinery (ACM) Code of Ethics and Professional Conduct	1992	3459
Association for Information Science and Technology (ASIS&T) Professional Guidelines	1992	531
Association for Information Systems (AIS) Code of Research Conduct	2013	4323
Association of Information Technology Professionals (AITP) Code of Ethics & Standards of Conduct	No date listed	790
Information Systems Security Association (ISSA) Code of Ethics	No date listed	332
Institute for Certification of Computing Professionals (ICCP) Code of Ethics	No date listed	2145
Institute of Electrical and Electronics Engineers (IEEE) Code of Ethics	No date listed	429
Institute of Hazardous Materials Management Code of Ethics and Professional Conduct	No date listed	895
International Society of Petroleum Engineers Code of Ethics	2013	561
National Funeral Directors Association Code of Professional Conduct	2008	1530
North American Nature Photography Association	No date listed	465
Royal Statistical Society	No date listed	1718
Society of American Foresters Code of Ethics	2000	833
Statistical Society of Canada Code of Ethical Statistical Practice	No date listed	766
The Wildlife Society Code of Ethics	No date listed	1486
The Worldwide Cleaning Industry Association (ISSA) Member Code of Ethics	2005	358



Unless otherwise specified, all work in this journal is licensed under a Creative Commons Attribution 4.0 International License.

⁷⁸Kate Crawford, “AI Now: Social and Political Questions for Artificial Intelligence,” (University of Washington, March 6, 2018), available at: <https://www.youtube.com/watch?v=a2IT7gWBfaE>.
⁷⁹Batya Friedman and Helen Nissenbaum, “Bias in Computer Systems,” *ACM Transactions on Information Systems* 14, no. 3 (September 5, 1996): 330-47; Joy Buolamwini and Timnit Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,” *Proceedings of Machine Learning Research* 81 (2018): 1-15.