

Introduction: Contemporary Culture After Data Science

Richard Jean So

Richard Jean So, McGill University

ABSTRACT

This essay introduces the articles found in this special joint issue of Post-45 and the Journal of Cultural Analytics, "Post-45 by the Numbers." It foregrounds the affordances of computational methods for the study of contemporary culture and literature, and offers brief summaries of each of the eight essays that follow in this special issue.

The computational study of contemporary culture is experiencing something of a renaissance across the university. Computer scientists are writing algorithms to identify the emotional arcs of novels.¹ Sociologists are building statistical models to analyze why certain works of visual art resonate more than others.² Electrical engineers have scraped tens of thousands of book reviews from the online website Goodreads.com to parse why some types of stories drive readers to talk to each other, and what they talk about.³ Evolutionary biologists and cognitive scientists have adapted models from information retrieval to study tens of thousands of popular Western songs in order to understand cultural change and the evolution of cultural taste.⁴

Contemporary culture's unprecedented *volume* and *accessibility* — particularly as born-digital artifacts — has largely driven this movement. Each year, approximately 600,000 novels are published in the United States.⁵ In 2019, more than 500 scripted television shows were streamed or broadcasted.⁶ Currently, the fan fiction website, An Archive of our Own (AO3), hosts six million works of fiction written across 40,000 fandoms.⁷ More than fifty billion images have been uploaded to the social media platform Instagram.⁸ Much of this contemporary culture exists as forms of *data*, scrapeable and accessible to researchers. This vast accessibility has coincided with the rise of powerful new computational algorithms designed to parse and analyze large amounts of text- and images-as-data. In the past decade, contemporary culture and data science have emerged as unlikely, yet increasingly, well-suited partners.

The affordances of digitized data and new computational methods have already begun to leave their mark on a number of disciplines, from political science to psychology, facilitating a series of disciplinary breakthroughs. The opportunity to study culture at scale and through the lens of data has proven particularly compelling. A large and growing number of colleagues at the university, far removed from cultural-literary studies, have shown an intense enthusiasm for this new field. And this enthusiasm is part of a broader trend in which quantitative scholars have introduced quantitative methods, or used them in creative ways, to study topics otherwise seen as resistant to such methods. For example, Avidit Acharya, Matthew Blackwell, and Maya Sen have used new datasets to study the history of slavery and its lingering (and as they show, underestimated) effects on contemporary voting patterns in the Deep South.⁹

Yet, for the most part, humanities departments — departments of English and literature, in particular — have been far slower to accept new data-driven and quantitative approaches as part of their overall methodological toolkit. Excluding a small group of digital humanists, humanities scholars of contemporary culture have shown at best a minor interest in using the affordances of data science, particularly born-digital cultural data, to study the vast amount of culture being produced today. What has and continues to excite academic colleagues from outside the humanities — what constitutes a rapidly growing and vibrant subfield of “cultural analytics” — has failed to achieve mainstream influence within the humanistic fields ostensibly most committed and best positioned to study, appreciate, and make sense of contemporary culture.¹⁰

If anything, even the most minimal recent attempts to introduce computational methods into literary and cultural studies have been met with hostility. Timothy Brennan writes: “Rather than a revolution, the digital humanities is a wedge separating the humanities from its reason to exist — namely, to think against prevailing norms . . . The ‘results’ of DH, then, are not entirely illusory. They have turned many humanists into establishment curators and made critical thought a form of planned obsolescence.”¹¹ It is not enough to ignore the work of our colleagues, spread out across the university, interested in studying culture with data. That interest needs to be policed and ultimately, resisted. The literature scholar alone stands against this establishment — biology, computer science, statistics, political science, economics, cognitive science, psychology, sociology, communications — where apparently, “critical thought” does not occur.

More generously, there are defensible reasons for this resistance. Literary and cultural studies scholars trained in qualitative methods strongly believe in the value

of these methods, and they have good reasons to do so. For one, close reading and historical analysis allow for an understanding of the particularity of individuals and individual works of art that new forms of data aggregation and analysis precisely threaten to erase. *Data science*, understood in its commercial context, is the project of using algorithms to classify individuals by consumer preference. More broadly, journalists, politicians, and social scientists — not just literary and cultural studies scholars — worry about the impact that such computational algorithms, as well as the platforms that support them, such as Facebook and Twitter, have on public discourse and democracy.¹²

This social crisis is felt to be echoed in the recent rise of data science programs, as well as the decline of humanities departments, at North American universities. Humanities research, particularly historically or theoretically motivated work, is being devalued compared to applied research whose impact can be more easily quantified in terms of public policy or scientific innovation. Over the past decade, the entire humanities academic job market has been eviscerated, with no relief in sight. Humanities graduate students cannot get academic jobs. Against this backdrop, the encroachment of data and scientific methods into the humanities may seem like the next inevitable step in assimilating cultural studies into a wider neoliberal scheme.

The problem with this perspective, however, is that it views the meeting between the humanities and the sciences entirely as a threat, rather than as an opportunity. It also discerns the two as deeply polarized, antithetical opposites, rather than as increasingly porous and open to exchange. The “neoliberal critique of the digital humanities,” in my view, has been greatly overstated and not borne out in reality.¹³ Research in the computational humanities has not reduced cultural criticism to robotic bean-counting, erasing ideological critique or historicism. Scholarship from Lauren Klein, Ted Underwood, Andrew Piper, and Katherine Bode have actively sought to integrate digital and computational methods with perspectives from gender studies, critical race theory, deconstruction, and book history.¹⁴ There is a fundamental difference in using these methods for instrumental reasons, such as consumer research, and using them to deepen our understanding of social and cultural problems. Fields like political science and economics are dominated by statistics now, but in ways that have allowed them to critically study social problems, such as the role that digital environments and tools play in magnifying social inequality.

If we can’t understand the distinction between using quantitative methods for instrumental versus critical-scholarly ends, then we will never be able to take

advantage of their affordances to enrich our understanding of the socio-cultural issues that we care about.

At the same time, institutionally, we have no evidence that data science programs seek to abolish the humanities. If anything, we see the opposite: an array of institutions, from the University of Washington to Emory, have partnered with humanities departments to hire tenure track faculty. We can choose to sit this one out as a form of aggrieved critical defiance. Or we can train humanities graduate students who are eligible for these positions. The expansion of data science at the university can directly contribute to the growth of the humanities.

Moreover, Brennan's claim (echoed in a recent piece by Nan Z. Da) that cultural studies has an unchanging, essential "reason to exist," whether valorizing the purported ineffable complexity of literature or "thinking against prevailing norms," and that the introduction of quantitative methods violates this purpose, itself does not stand up to historical scrutiny.¹⁵ Rachel Buurma and Laura Heffernan have written a new history of the literary studies discipline, focusing on the classroom. Their history contradicts the belief that the introduction of quantitative methods into literary and cultural studies represents a recent and alien phenomenon, dominated by white men in cahoots with big tech. That belief "melts away when we look at the earlier twentieth-century women professors, both on and off the tenure track, who used classrooms as the original supercomputers."¹⁶ A long-standing historical tension between qualitative and quantitative methods — from I.A. Richard's "science of criticism" to Janice Radway's reader response theory — has been both deeply generative and constitutive for the literary studies discipline.¹⁷

At the same time, Matthew Handelman has written a new history of critical theory, focusing on the Frankfurt School, that similarly debunks the belief that the critical study of culture must always oppose quantitative reasoning and evidence. He recovers a cohort of thinkers adjacent to Theodore Adorno and Max Horkheimer, such as Gershom Scholem, whose "theories of aesthetics, messianism, and cultural critique borrow ideas from mathematical logic, infinitesimal calculus, and geometry to theorize art and culture that strive to reveal, and potentially counter the contradictions of modern society."¹⁸ Invoking critical theory as a bludgeon against quantitative methods diminishes its power. It does not sustain or expand it. There is a more capacious version of theory, one compatible with data science, that enriches its legacy. It is one that "can help us confront and intervene in our digital and increasingly mathematical present."¹⁹

This last point is especially important. Contemporary culture is deeply implicated and saturated by data. Few aspects of the production and reception of the arts today

are not touched, in some way, by algorithms, data science, and computational processes. Much of what we consume, whether Internet-based streaming television, Instagram photos, or physical books, comes to us already quantified and shaped by non-human, algorithmic decision-making. For most people who produce, sell, market, and evaluate art for a living, contemporary culture is in part a form of data, which itself requires data-driven tools to further propagate.

For the most part, media and cultural studies scholars have developed a critical and resistance-based posture towards this growing imbrication of data and culture. Earlier, I cited important work by scholars such as Wendy Hui Kyong Chun, Cathy O’Neil and Safiya Noble, which analyzes and critiques the harmful impacts of algorithms on society.²⁰ This research program will expand in the years to come. However, the discipline courts a peril in prioritizing this form of work to the exclusion of all applied data-driven cultural research, including research that uses data methods precisely to expose and critique the deleterious effects of data itself.

We risk neglecting methods and tools that can enrich our understanding of our materials and the scholarly problems we work on. We also risk isolating ourselves from the rest of the university and its community of scholars in the computational social sciences and sciences. Overall, we risk being unable to contribute to important data-driven research focused on combating the major problems of our time, whether as a function of our autonomous discipline, or through interdisciplinary collaboration with quantitative researchers.

Disciplines outside of the humanities feel profound excitement for the affordances of data and data science. New scholarly paradigms are being constructed. That enthusiasm is so strong that they have started to peer over their disciplinary borders, finding a particularly compelling opportunity to analyze contemporary culture. Most cultural studies scholars do not share this enthusiasm. The problem, as Ted Underwood observes, is that this opportunity sounds like a great opportunity but a great one for someone else.²¹ It doesn’t have to be this way.

The essays collected for this special joint issue of *Post45: Peer Reviewed* and the *Journal of Cultural Analytics* offer a vision of contemporary cultural studies when we do take advantage of this opportunity. Together, they present a new dispensation in the development of the digital humanities and cultural analytics. Specifically, these eight articles *do not*

- Make the polemical case for the validity of using computers to study culture.
- Discuss, debate, or treat the “digital humanities” as a meta-theoretical academic topic.

- Provide methodological primers on computational tools, like topic modeling.

Rather, they utilize data and data science methods, often in concert with traditional humanistic methods, such as close reading and historical analysis, to study contemporary culture, literature, and media in order to make original scholarly insights and contributions. The essays do not engage in a “methods war” discourse. They take as a given that quantitative and computational methods represent a valid mode of analysis for the study of culture and literature. They assume the reader has a basic familiarity with now staple computational methods in DH, such as classification or topic modeling, or assume that readers are capable of familiarizing themselves. With these assumptions in place, the essays spend the bulk of their space on more interesting tasks, such as producing new interpretations of contemporary fiction or creating novel analytical frameworks to study the circulation of short textual content or Internet memes.

Some of the essays mobilize a computational apparatus in order to provide an empirical foundation to long-standing qualitative intuitions or observations regarding cultural history. Michelle Moravec and Kent Chang revisit a cohort of US feminist bestsellers from the 1970s to parse how and why they helped to articulate the idea of “feminism” in this period. Nicholas Kelly, Nicole White, and Loren Glass analyze the effect that the University of Iowa’s creative writing program has had on the geographical imagination of postwar American fiction. The goal of this work is not to overturn decades’ worth of humanistic scholarship, but rather, to offer a stronger quantitative baseline for their claims. And in doing so, this scholarship allows us to ask a new set of questions, such as: *When exactly did the geographical imagination of postwar American fiction shift? Has it attenuated in the past decades? If so, why?*

Other essays use data and data science to develop a set of claims that more directly challenge or reorient the assumptions of a subfield of contemporary cultural studies, such as Asian American Studies or poetics. Dan Sinykin and Edwin Roland use classification models to find a fundamental, under-studied distinction between fiction published by “Big 5” conglomerate firms and fiction published by smaller indie and nonprofit publishers. Long Le-Khac and Kate Hao perform a quantitative meta-analysis of the rise of Asian American literary studies and its forms of literary attention to find a series of under-appreciated or unnoticed inequalities. James Lee and Ankit Basnet employ social network analysis to study the contemporary field of American poetry at scale — particularly how it has been organized by anthologies and audio archives — and make a series of novel discoveries as to what precisely has animated this field.

A final group of essays turns its attention to the Internet, and the ways in which the web has engendered a unique set of affordances for the expression and spread of culture, particularly text-based culture. Here, they find that born-digital content demands born-digital methods; for a cultural field that immanently exists at scale, methods that work at the level of a single tweet or Instagram photo will necessarily miss something very important. Scholars at the Stanford Literary Lab study the increasingly ubiquitous category of “voice” within online cultural criticism; Melanie Walsh and Maria Antoniak analyze the popular category of “the classic” on web-based reader platforms, such as Goodreads.com; Tess McNulty probes the recent rise of the “sentimental anecdote” as a new cultural form, both on the Internet and elsewhere in popular culture. Overall, these scholars find a new world of user-driven cultural production, organized by a set of otherwise unfamiliar (at least to academics) vocabularies, which require not only born-digital methods to make sense of at scale, but also, as a consequence, new analytical concepts and paradigms.

Together, the contributors believe, as I do, that the best way to make the case for digital methods for cultural analysis is to prove that digital methods can generate useful and original insights and discoveries for cultural analysis. Computational methods will only cease to appear alien to contemporary cultural studies when enough contemporary cultural studies scholars actually use them. To this end, these essays also mean to inspire and set the stage for future scholarship. Together, they ask: what would it mean to see, along with so many of our non-humanities colleagues across the university, new computational methods as an opportunity for cultural studies, rather than an existential threat? We wonder: perhaps it’s time to let some light in.

References

¹ O-Joun Lee and Jason J. Jung, “Story embedding: Learning distributed representations of stories based on character networks,” *Artificial Intelligence* 281 (2020): 5070-5074.

² John W. Mohr et al., *Measuring Culture* (New York: Columbia University Press, 2020).

³ Shadi Shahsavari et al., “An Automated Pipeline for Character and Relationship Extraction from Readers Literary Book Reviews on Goodreads.com,” *WebSci '20: 12th ACM Conference on Web Science* (July 2020): 277-286.

⁴ Matthias Mauch, Robert M. MacCallum, Mark Levy, and Armand M. Leroi, “The evolution of popular music: USA 1960 – 2010,” *Royal Society Open Science* 2, no. 5 (2015).

⁵ A rough estimate reported in this article, back in 2013; we can reasonably expect that this number is an underestimate for the years since, up to 2020. Nick Morgan, [“Publishing Your Book in 2013? Here’s What You Need to Know,”](#) *Forbes*, January 8, 2013.

⁶ As reported in this article: Lucas Shaw, [“Hollywood Made 532 TV Shows in 2019, and It’s Going to Make More,”](#) *Bloomberg News*, January 9, 2020.

⁷ As currently reported on Wikipedia’s entry for the website: Wikipedia, s.v. [“Archive of Our Own,”](#) last modified January 10, 2021, 9:06.

⁸ As reported by Salman Aslam, [“Instagram by the Numbers: Stats Demographics & Fun Facts,”](#) *Omnicores*, January 6, 2021.

⁹ Avidit Acharya, Matthew Blackwell, and Maya Sen, *Deep Roots: How Slavery Still Shapes Southern Politics* (Princeton: Princeton University Press, 2018).

¹⁰ A growing number of research articles placed in major journals, like *PMLA* and *Critical Inquiry*, central to the discipline, and a series of monographs published by major academic presses, such as the University of Chicago Press and Columbia University Press, indicate that this is changing. Here I will flag a few recent books: Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago: University of Chicago Press, 2019); Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary Study* (Ann Arbor: University of Michigan Press, 2018); Andrew Piper, *Enumerations: Data and Literary Study* (Chicago: University of Chicago Press, 2018); Richard Jean So, *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction* (New York: Columbia University Press, 2020), and Eric Bulson, *Ulysses by Numbers* (New York: Columbia University Press, 2020). But overall, the percentage of literary and cultural studies scholars working in this area is relatively miniscule.

¹¹ Timothy Brennan, “The Digital Humanities Bust,” *Chronicle of Higher Education*, October 15, 2017.

¹² For example, see Wendy Chun, *Programmed Visions: Software and Memory* (Cambridge: MIT Press, 2011); Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018), and Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Books, 2016).

¹³ See for example, David Allington, Sarah Brouillete, and David Golumbia, “Neoliberal Tools and Archives: A Political History of Digital Humanities,” *Los Angeles Review of Books*, May 1, 2016.

¹⁴ I have already cited work by these scholars; here I would just add Catherine D’Ignazio and Lauren Klein, *Data Feminism* (Cambridge: MIT Press, 2020) and a piece I co-wrote with Edwin Roland, “Race and Distant Reading,” *PMLA* 135, no. 1 (January 2020): 59-73.

¹⁵ For the first type of claim, see Nan Z. Da, “The Computational Case Against Computational Literary Studies,” *Critical Inquiry* 45 (Spring 2019): 601-639.

¹⁶ Rachel Burma and Laura Heffernan, *The Teaching Archive: A New History for Literary Study* (Chicago: University of Chicago Press, 2020), 11.

¹⁷ Ted Underwood’s “A Genealogy of Distant Reading,” *DHQ* 11, no. 2 (2017) offers a useful history of quantitative methods in literary criticism; see also Michael Gavin’s helpful “Vector Semantics, William Empson, and the Study of Ambiguity,” *Critical Inquiry* 44 (Summer 2018): 641-673.

¹⁸ Matthew Handelman, *The Mathematical Imagination: On the Origins and Promise of Critical Theory* (New York: Fordham University Press, 2019), 2.

¹⁹ *Ibid.*, 2.

²⁰ In an earlier footnote I cite the work of Chun, Noble, and O’Neil; here I would add Frank Pasquale, *Black-Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015)

and the more recent, Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (New York: Polity Press, 2019).

²¹ Ted Underwood, "Machine Learning and Human Perspective," *PMLA* 135, no. 1 (January 2020): 92-109.