CHAPTER

The digitalization of the humanities: Recent trends and future prospects

14

James E. Dobson

Department of English and Creative Writing, Dartmouth College, Hanover, NH, United States

Introduction

Depending upon how we think about computational and digital methods, the digital turn in the humanities may have occurred contemporaneously with other academic fields undergoing a similar transformation. From another perspective, the digital turn was slow to arrive and much belated. What seems less up for debate, however, is that the computational transformation of the humanities has been uneven and contested, and remains incomplete. While those primarily quantitative or experimental fields were relatively quick to adopt computers in the processing and analysis of their data, researchers in the humanities have, on the whole, approached computation with some hesitation. This hesitation, to be sure, is partly the result of working with unruly and multidimensional objects such as images, texts, sounds, and material objects. Those researchers already accustomed to rendering their objects of inquiry into rows and columns had a far easier task in the transition of these items into digital formats. The historical dimension, too, of humanities research, in both its attention to objects created in the past and the ways in which noncontemporaneous methods influence historical interpretation, complicates the task of modeling human culture.

Over the past several decades there have been sustained critiques of the digitization of the humanities and resistance to modeling and what many see as the reduction of objects to that which can be quantized and quantified. There have also been important philosophical critiques of the digital as such (Fazi, 2018; Galloway, 2014) and its overdetermining binary logic (Evens, 2024). Despite these critiques, transformations of methods, shifts in research paradigms, and waning institutional support for some large-scale projects, there has been a marked increase in scholarship making use of computation and computational models in the humanities. In surveying recent and contemporary humanities scholarship employing computation, clear trends and prospects for the future emerge. While once almost obligatory, the rhetorical move to demonstrate the commensurability of computational methods with nondigital methods has become less important and attention has shifted to what might be learned from new ways of extracting evidence from objects of interest. Researchers have emphasized and will continue to underline the importance of asking to what

Digital Transformation in Artificial Systems. https://doi.org/10.1016/B978-0-443-32964-7.00016-0

degree digital tools shape the sort of questions that one might ask and inquiring into the ways in which these tools transform the objects to which they have been applied. It is precisely this move that renders these research activities, even when they are identical in form to those used by social scientists and others, compatible with the humanities.

Because humanities fields, in general, consider themselves to be interpretive, we find the lack of consensus to be a major impediment to digitalization of the humanities, even within highly specialized and narrowly defined humanistic fields, regarding the proper ways of interpreting objects. Those objects themselves, as John Guillory argues, might also be multiple, which is to say composite in character, entangled, or imbricated with other objects. For Guillory, following midtwentieth century art historian Erwin Panofsky, what makes the humanities distinctive from the social and natural sciences is attention to both *monuments*, "those artifacts, events, or ideas that have the most urgent meaning for us at any present moment and that most demand our recognition or study," and *documents*, "all of those artifacts or traces of human making, action, or thought surviving into the present, the total accumulation of human artifacts, events, and ideas" (Guillory, 2016, p. 12). In splitting the object of humanities research, in recognizing their diversity and multiplicity, Guillory gives us a vision of the humanities with a complex object-method relation, insofar as each object may necessitate the researcher to deploy different methods. Such complications frustrate attempts on the part of the humanities researcher to determine what and how to make their objects digital, what to model, how to analyze these digital objects, and finally, how to interpret the entire workflow and the many decisions made along the way.

In part because of assumed easier formatting of textual data and the accessibility of digital texts, the field of literary studies has played an important and indeed central role in the digitalization of the humanities. Computation and textual structure, as Dennis Tenen argues, might not be the strange bedfellows they initially appear to be: "We may liken books, paragraphs, and sentences to nesting dolls: data structures that contain within them further smaller arrangements of information. A word fits inside a sentence, the sentence within a paragraph, the paragraph within a chapter, the chapter within a book, the book within an archive, and so on" (Tenen, 2017, p. 95). Such analogs have made it easier for those humanists eager to use computation in the analysis of their objects of interest to conceptualize the (re)formatting of text into textual data and processes that enable comparisons of different modeled units produced from these objects. The explosive growth, especially in the first 2 decades of the 21st century, resulting from the combination of methods derived primarily from computational linguistics and data science, including the now mature natural language processing techniques, and the ready availability on the Internet of high-quality digitized texts has also been instrumental in drawing attention to the insights that could be gained from text mining and modeling. Humanists using these paradigms from what we might term already transformed (in terms of the application of computational methods) fields have found ready partners, collaborators, and models for the understanding of a variety of objects of humanistic interest.

As the range of computational methods and the scope of digital culture continue to expand and develop, some distinctions are increasingly becoming apparent. Divisions have emerged between scholars studying digitized historical or archival material, especially in fields such as digital literary and art history, and humanists examining contemporary digital culture, so-called born digital new media objects (Hayles, 2004). Lev Manovich's coining of "cultural analytics" in 2007 was an explicit response to what he saw as the limitations of the then dominant discourse of digital humanities. Manovich wanted to focus attention on contemporary and invented cultural analytics to give a name to practices that included the "computational analysis of patterns and trends in contemporary digital culture (as opposed to only historical culture) and can cover analysis of all kinds of media (as opposed to only texts)" (Manovich, 2020, p. 7). While Manovich's intervention was to reconceptualize the study of culture with computers as a new kind of media studies, even within familiar digital approaches to historical literary studies, there exist a plurality of methods capable of addressing much more than text or the text as conceived. As I have written elsewhere on the topic of digital literary studies, while deploying computational approaches in the study of literature is an "orientation or mode of research ... compatible with the goals of many humanists ... it does not come close to exhausting the range of inquiry found within literary studies" (Dobson, 2019, p. viii). The humanities, as primarily interpretive disciplines, continually confront what we might want to call obstinate objects that make claims on the present (Gadamer, 2004). These claims generate a seemingly interminable set of questions and approaches, and as long as there are new readers, viewers, and listeners, these objects remain meaningful in many new ways. As new computational technologies emerge and are applied in the production of cultural objects and in the scholarly analysis of culture of both the past and present, these crucially important questions about meaning and meaning making will only multiply.

From humanities computing to the digital humanities

Computational work in the humanities, research assisted by the use of computation, long preceded the label of the digital humanities. Known prior to the coining of the term "digital humanities" as humanities computing, researchers working in this area deployed a wide range of applications and methods. Many point to Father Roberto A. Busa's assembly of concordances drawn from the works of theologian Thomas Aquinas with an IBM computer in the 1940s as an important precursor for humanities computing. Busa would later describe this work as participating in what he termed "textual and hermeneutics informatics" (Busa, 2007, p. xxi). Digitization of documents and images, the supplementing of these objects with descriptive metadata or hierarchical tagging of subobjects found within these objects, the creation of databases, the querying of these databases, and the presentation of these materials were all common humanities computing and early digital humanities projects. The digital environment and the low barrier of entry for producing websites and

digital editions enabled the recovery of previously ignored or niche objects such as texts and images that were deemed unprofitable for print editions. Commenting on some of these early digitizing projects, Susan Schreibman notes that "from the late 1980s to the mid-1990s, many digital resources in the field of literature were conceived more as digital libraries than research projects that utilized the medium as a site of interpretation" (Schreibman, 2002, p. 284). This earlier moment was marked by field-transformative projects that were also continuous with the political stakes of recovery, such as the Women Writers Project that was initially created in 1986 to recover, digitize, and encode texts in English authored by women (Wernimont & Flanders, 2010). For scholars curating archives and conducting research in these gathered materials, the digital environment offered a high-quality distribution platform with relatively low cost, although frequently labor and expertise for the maintenance and upkeep of these archives and sites were underestimated. This work continues with a number of digital archives of primary sources, editorial commentary, and scholarly research such as the Colored Conventions Project (CCP) dedicated to collecting 19th-century material connected with the Black convention movement in North America (Foreman et al., 2021). Contemporary projects such as the CCP have expanded beyond collection to enable interpretation, often with the assistance of hosted or archive-specific computational tools.

Capturing (Drucker, 2011) or representing humanities objects in digital form can be complicated, not least in part because often researchers, in addressing their objects, are also interested in something that exceeds the object itself. Take, for example, a work of visual art, a painting. While one may imagine that the only object of interest would be the canvas itself, the painting cannot be separated from its frame. And the frame itself is much more than just the physical object that contains and adorns the canvas—the frame opens out into the world, into documents, its geography, its provenance. In the case of a painting, the frame might be said to function as what Jacques Derrida calls a *parergon*, "neither work (*ergon*) nor outside the work [*hors d'oeuvre*], neither inside nor outside, neither above nor below" (Derrida, 1987, p. 9). The interpretation of this work thus includes a whole range of different kinds of objects that can expand the schema by which one might represent the object. The same applies to texts. While the digitization of printed books might appear to involve only the encoding of the text and potential complications involving what Gérard Genette termed "paratexts" (Genette, 1987), those text objects such as title pages, prefaces, acknowledgments, and end notes included before (peritexts) and after (epitext) the main text, expand well beyond the covers for scholars using book history methods, taking up similar sites of inquiry as the previously mentioned painting. An increasing number of bibliographic studies deploy computation to model collections of texts, literary systems (Bode, 2018), and research connections, differences, and commonalities among these collective objects (Cordell, 2020).

The phrase "digital humanities" (DH) appeared in 2004 and was first used for a volume titled *Companion to Digital Humanities*, edited by Susan Schriebman, Ray Siemens, and John Unsworth and published by Blackwell (Kirschenbaum, 2010). The book featured a foreword by Roberto A. Busa along with an editors'

introduction and sections dedicated to history, principles, applications, production, dissemination, and archiving. These key thematic areas looked backward at the early years of humanities computing and gave a sense of possible future directions for this nascent field through chapters addressing, among other topics, classification, stylistics, modeling, and textual analysis. Since its incarnation, there have been numerous attempts to define digital humanities. There has been much written about defining the methods and objects of digital humanities (Svensson, 2014, 2016) and reimagining the field (Terras et al., 2013), and there have also been several DH projects dedicated to studying the discourse of the digital humanities (Liu et al., 2022; Luhmann & Burghardt, 2022).

Despite the different and competing visions offered, there have been some common threads within DH. The trajectory from the early humanities computing work identified by Schreibman as concerned with archive and library building to the present digital humanities, for example, runs through work that theorizes the field as essentially concerned with the creation and preservation of the digital cultural record. Roopika Risam gives a clear articulation of the stakes of this work: "The great promise of digital humanities—yet to be fully realized—lies in how it can be used to both uncover and complicate our understanding of the relationship between 'humanity' and 'technology'—that is, of course, if the *human* in *humanity* is inclusive" (Risam, 2021). Risam's corrective to earlier accounts of digital humanities identifies the shortcomings, which is to say the exclusions and biases, that have shaped the existing cultural record and the promises of the digital to recover, recenter, and refocus a more expansive human record. There has also been a longstanding interest in taking seriously the affordances of digitization; "quantification," Jacqueline Wernimont argues, "is not inert or 'merely' descriptive but rather always already engaged in the processes by which bodies and people have become and are becoming visible to themselves, to others, and to nation-state" (Wernimont, 2021, p. 430). The absence of data, too, as Jessica Marie Johnson has pointed out, in her reading of the "shadowed spaces of empty cells, null values" has important consequences for what is and what is not counted (Johnson, 2018, p. 70).

In her influential "Making a Case for the Black Digital Humanities," Kim Gallon makes an even stronger claim about the limitations to existing accounts of the human and the troubling reproductions of those conceptions in the digital humanities: "I would argue that any connection between humanity and the digital … requires an investigation into how computational processes might reinforce the notion of a humanity developed out of racializing systems, even as they foster efforts to assemble or otherwise build alternative human modalities" (Gallon, 2016, p. 44). These critiques about the relation between accounts of the human and digitality have become even more important as the discourse surrounding artificial intelligence (AI) and contemporary machine learning so frequently invokes an overdetermining natural/artificial binary. Others have examined the material relations among researchers within DH and the university and argued that the digital humanities have less to do with methods and data than the political and economic demands of the 21st-century university (Allington et al., 2016).

The computational turn

When digital archives of digitized texts—ranging from the no-frills, plain-text editions long available from Project Gutenberg to the professionally managed HathiTrust Digital Library with Pandas-ready columnal data objects from its Extracted Features Dataset—became readily accessible, the computational turn began in earnest. Literary scholar Stephen Ramsay understood the availability of digitized texts as instantiating something he called the "Screwmeneutical Imperative":

> *There are so many books. There is so little time. Your ethical obligation is neither to read them all nor to pretend that you have read them all, but to understand each path through the vast archive as an important moment in the world's duration—as an invitation to community, relationship, and play. (Ramsay, 2014, p. 119)*

Such calls to experiment were initially free from disciplinary protocols for quantitative analysis, for these were essentially nonexistent in the humanities. Ramsay's invocation of play draws a line between exploratory work in the humanities and the use of potentially similar methods in the social sciences; absent imperatives to present statistical rigor, such as tests of significance, this opening to quantitative work in the humanities presented a relatively low barrier. This was also enabled by the explicit analogy drawn between two different proposed reading practices that were set against each by scale and also, as Ramsay suggests, time.

While not necessarily computational, the paradigms of distant reading or distant viewing (Arnold & Tilton, 2019), in the case of visual objects, represent important shifts in thinking about scale that assisted in making the computational turn in the humanities possible. Franco Moretti provided the stakes of his distant reading intervention in literary studies by way of a critique of close reading in the understanding of literary history. Moretti writes of close reading:

> *At bottom, it's a theological exercise—very solemn treatment of very few texts taken very seriously—whereas what we really need is a little pact with the devil: we know how to read texts, now let's learn how* not *to read them. Distant reading: where distance, let me repeat it,* is a condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something. (Moretti, 2013, pp. 48–49)*

Moretti's justification of distant reading was that only by focusing on larger scales, objects, and crucially, as he writes, of both larger and smaller objects, could knowledge of the construction and operation of systems, literary systems, be produced. These simultaneously larger and small objects are on the display in an important series of pamphlets, the Stanford Literary Lab, directed at the time by Franco Moretti. Collected together in a book published in 2016, these pamphlets demonstrated research and results drawn from a wide variety of methods from word

frequency and distribution, through which they studied the semantic field of texts by tracking keywords and collocates across collections, to tracking lexical density in canonical and prestige authors (Moretti, 2017). The computational turn in literary studies expanded to include studies of genre (Underwood, 2019), fictionality (Piper, 2018), and influence (Jockers, 2013) in which scholars modeled collections using a variety of natural language processing and machine learning techniques. There has been considerable interest in studying narrative structure and the use of sentiment analysis as either a proxy for the shape of the narrative (Archer & Jockers, 2016) or the sentiment arc, as a measure of the relative intensity of emotion registered in text over time (Elkins, 2022). The demand for disciplinary protocols relating to the presentation of tests of significance to validate these models, however, has been slow in coming, and the shift of attention from quantitative textual analysis to the use of and concerns raised by generative AI has only deferred much of this work.

The development of popular and easy-to-use algorithms used for topic modeling—the unsupervised organization of key terms or phrases from a collection of texts into a predefined number of topically or thematic buckets, such as latent Dirichlet allocation and nonnegative matrix factorization (Blei, 2012), had uptake in the humanities. Following successful application to the problems of knowledge discovery and information retrieval from scientific publications, researchers in the humanities applied these models archives of literary texts (Jockers, 2013) and metadata from art historical archives (García-Zorita & Pacios, 2017). One of the reasons that made topic models successful was the use of a simple bag-of-words data model, the document-term matrix that maps word frequencies to documents or document segments. This same data model enabled different higher-level transformations and modeling methods. Rendered as a highly structured vector space, researchers applied unsupervised k-means clustering and supervised classification to these same sources. Hoyt Long and Richard Jean So's use of a supervised Naïve Bayes classifier enabled them to identify common patterns and compare short modernist poems to English-language haiku (Long & So, 2016). Ted Underwood's *Distant Horizons: Digital Evidence and Literary Change* uses supervised classification in the form of logistic regression to examine the historical transformation of genres such as science and detective fiction (Underwood, 2019).

While the document-term matrix data model was the enabling technology underlying Underwood's use of machine learning in his investigation of literary history, what made this research possible was the abovementioned HathiTrust Digital Library. Created from agreements with Google that led to the creation of Google Books, HathiTrust provides access to the collective archives of a number of research university libraries. Initially the archive was available for metadata and keyword searching of the body of the text; the HathiTrust Research Center (HTRC) then developed the Extracted Features Dataset and a software library (htrc-feature-reader) with Pandas-ready models of text with book, volume, and page-level part-of-speech tagged word frequency data. Underwood's use of HathiTrust gave him access to high-quality data models of public domain texts, and these same datasets

were available to anyone who wanted to reproduce his research. The rendering of books in these literary archives as document-term matrices made it trivial to apply the earlier mentioned techniques to model subsets of the archive further. The Extracted Features Dataset had another major advantage: distributed as frequency data rather than linear text, these objects were considered nonconsumptive in the sense that a human reader could not easily recover the original word order. As a nonconsumptive representation of a text, these data were able to be shared with researchers with fair use justification. HathiTrust also made available full-text data, human readable and linear text files, to researchers with signed agreements and access via a capsule running a virtual machine that limited the export of the original files. This capsule model was also used by ITHAKA, the parent organization of the scholarly library JStor. ITHAKA's Constellate platform provided limited capsule access to full-text objects in the JStor library. These digital libraries and platforms appeared at the right time in the development of computational methods for the modeling of texts and at a moment of heightened interest in the digital humanities. Unfortunately, the capsule-based approach seems to have been short-lived (most likely because these were difficult to use, required constructing complex workflows and multiple platforms, and had high staffing and technology demands). The demise of the HTRC and Constellate terminated one possible route for text and data mining the humanities. Other archives of public domain texts and the HTRC Extracted Features Dataset, however, remain available and viable sources for accessing bag-of-word based document-term matrices of texts of interest to humanists.

In the field of history, which has perhaps less suspicion of quantification and has long used quantitative data in research (Crymble, 2021), text mining in particular has been an important technology for the development of a digital history. In explaining her preference for text mining as an organizing method, historian Jo Guldi writes that "Unlike the phrase digital history, text mining for historical analysis carries no suggestion that computer-driven history will somehow substitute for history by traditional means ... Text mining for historical analysis rather sits across the many domains that use text, AI and statistics to produce certain kinds of knowledge, useful to certain purposes" (Guldi, 2023, p. 13). Guldi makes a convincing argument for something she calls a "smarter data science," a hybrid approach for researchers in history to take up in which "concerns about the bias of data and algorithms from the humanities and questions from historical theory meet mathematical modeling" (Guldi, 2023, p. 419).

Another thread from the humanities computing moment was the methodological application of computational transformations to the objects themselves. In the case of text-based objects, this work quickly moved from the production of summary statistics about an object, such as frequently used terms and phrases, and the extraction of keywords and those words in context (as in a concordance), to the modeling of text as a vocabulary distribution. Building on work in adjacent social scientific fields and the widespread availability of machine learning methods in programming environments such as R and Python, humanities researchers began to build models or new representations of their objects and collections. Responding to accounts, such

the one offered by Franco Moretti, that positioned distant reading and computational methods against the familiar close reading practices of literary studies, some scholars have explicitly redescribed their methods in formalist terms and applied computation to the small units favored by close reading. Comparing his set of algorithmic manipulations of a single text with the instruments of the sciences connected with investigation at smaller scales, in particular the microscope, Martin Paul Eve finds common ground between his approaches and the close reading practices of many literary scholars:

> *I have also attempted throughout this book to weave my data-driven approaches back into a more traditional set of literary hermeneutics—that is, to use the computational methods as data gathering for literary interpretation. If, in one sense, this is what close reading has always done—looked closely for features that could serve to promote or discredit an interpretative standpoint—then it is actually somewhat the case that there is little difference between close and computational distant reading.* (Eve, 2019, p. 156)

Using "computational formalism" as part of the subtitle for his *Close Reading with Computers*, Eve gestures toward the Stanford Literary Lab's coining of "Quantitative Formalism" in their first pamphlet publication and to "the repetitious task-based nature of the work here conducted using computers, as opposed to the purely numerical approach signaled in the Stanford Lab's publication" (Eve, 2019, pp. 156–157). This repositioning of a computational literary studies draws out commonalities in attention to the objects of study and the perhaps especially apt ways in which formal analysis is the appropriate analog for computational approaches to esthetic objects.

Advances in computer vision and image analysis have also transformed approaches to art history and introduced new methods. The Reassessing Vincent van Gogh project applied image analysis methods to produce pigment concentration maps and reconstruct the color palette originally used by Vincent Van Gogh as well as the changes to these colors over time (Kirchner et al., 2018). As part of his broad and developing cultural analytics investigations, Lev Manovich and his colleagues applied sampling, created montages, and visualized a number of image sets, archives, and sources including films, video games, *Time Magazine* covers, and publicly accessible geotagged Instagram images (Manovich, 2020). Humanists have also theorized the digital image, its relation to earlier and other forms of media (Denson, 2020), and its compressed rearticulations and copies (Steyerl, 2009).

While digital art has existed for decades, computer-aided interpretation of both digitized and born-digital artworks has added increased scale and capacity for the searching and cataloging of these works. With computer vision models trained for object recognition, this cataloging far exceeds the searching of descriptive information about a work by enabling the creation of additional metadata through the cataloging of the parts, the subobjects of a primary object. Adapting the formalisms in humanities fields, several humanities scholars with computational expertise have produced compelling accounts of a computational formalism with continuities

between these nondigital methods of analysis and recent methods, especially those using machine learning to model more complex visual and textual features. For art historian Amanda Wasielewski, computational methods promise a return to formalist concerns. In deploying computer vision techniques to study art historical images, Wasielewski describes her work: "opening up an interdisciplinary dialog between computer science and art history, the aim is to move beyond superficial or one-sided collaboration to something approaching understanding, both for theory and for practice" (Wasielewski, 2023, p. 37). Like borrowing methods from computational and corpus linguistics for literary studies, a successful digital art history, Wasielewski argues, should remain in conversation with researchers in other disciplines who are also creating and applying computational transformations to visual objects.

Unequivocally answering yes to the question of the existence of a digital art history, Leonardo Impett and Fabian Offert apply the pretrained CLIP vision transformer model to image archives for retrieval tasks and argue that this transformed field "has emerged as a crucial route to understanding, exposing, and critiquing the visual ideology of contemporary AI models" (Impett & Offert, 2022, pp. 186). Impett and Offert implicitly respond to earlier critiques launched at work in the digital humanities. Sharing arguments for a self-reflexive critical approach that interprets computational workflows alongside the results of this analysis emerging from adjacent fields, Impett and Offert direct their attention to the tools, on the ways vision transformers, as key parts of contemporary visual culture, see image data. Likewise, Wasielewski's call to foreground "interpretation and criticality" (Wasielewski, 2023, p. 19) provides an important corrective for earlier computational work in the humanities that was not as concerned with questions about how these tools function and how they transform research. David Berry describes some of the aspects of what we might call criticality in his account of a *critical digital humanities* that "draws attention to the politics and norms that are embedded in digital technology, algorithms, and software. It foregrounds these questions by affirming human emancipation as a goal making a distinction between what society is and what it hopes to be" (Berry, 2022, pp. 130–131).

In addition to these crucially important questions about the relation between method and ideology, there are ongoing questions that one might ask about the match or fit between an object and a selected method as well as the reliability of the methods themselves. The components of contemporary computational workflows, such as particular algorithms and data objects, have been subject to media histories and archaeologies. A media archaeology, as articulated by Vivian Sobchack (Sobchack (2011) and Jussi Parikka, concerns the oft-complex relation between the material archive of the past and the present affordances and functions of media. Asking "What are the conditions of existence of this thing, of that statement, of these discourses and the multiple media(ted) practices with which we live?" (Parikka, 2012, p. 18), media archaeology has become an important critical practice for examining continuities, recurrences, and short circuits that bridge the predigital past with the computationally saturated present (Dobson, 2023a). Critics have also

demonstrated that digital technology, even the most advanced systems and methods of the present, remains incredibly material, from the labor conditions involved in the training of machine learning to the energy and environmental costs of powering thousands of CPUs and GPUs (Bender et al., 2021). As individual humanities fields continue to undergo their own computational turn, familiar problems from other fields regarding the development of research and data standards, and especially issues connected with reproducibility and explainability, have been pressing (Ries et al., 2023). There is an urgent need for academic organizations and publications to respond with policies and encouragements for the public sharing of data, code repositories, and workflows.

The era of generative artificial intelligence

The advent of high-quality and high-performance generative AI has helped usher in an epochal shift in the digitalization of the humanities. This transformation has taken place on three different registers. The first and highest-level register concerns the wider understanding and perception of neural networks and advances in language models, within both the humanities and the larger public. The second register of the generative AI transformation affects the ways in which these language models alter the pipelines and methods used by computational researchers in the humanities. The final register links both technical advancements in terms of the capabilities and capacities of generative AI with their cultural and social impact through the development of new paradigms for theoretical and philosophical understandings of language, modeling, and world making.

The widespread availability of high-quality generative AI applications such as OpenAI's ChatGPT suddenly brought attention to the ongoing advances in machine learning. OpenAI's tremendous technical advantages, their processing facilities used to train and run inference on large language models, and their use of very large datasets, combined with sophisticated fine-tuning for instruction and conversation produced a platform for generative AI that was generic enough to enable users to find widespread uses for this technology. These uses ranged from casual entertainment, interactive conversation that was generally harmless and free from the problems of offensive and deeply biased outputs that plagued prior machine learning–based chatbots, to productive tasks such as composing documents, writing computer code, and helping to generate or refine ideas. Although ChatGPT, in this initial moment, was less useful for information extraction and knowledge discovery from its training data, that did not prevent many users from querying the language model for facts and other information.

ChatGPT's largest impact in relation to the humanities, and perhaps academia as a whole, may have been in relation to its capacities to produce extended documents that could mimic well-known genres such as the academic essay. The academic essay, a mainstay assessment strategy used by instructors to teach thinking and measure and evaluate learning, was put under unprecedented pressure by OpenAI's

release of ChatGPT. Developers have also integrated generative AI into the basic functions of a growing number of more general-purpose applications, including productivity tools such as Google Docs, Microsoft Word, and Adobe Acrobat. Writing, especially in the humanities, does much more than report on insights gained and communicate information. In writing, scholars think through difficult problems and lay out an exploratory answer. The act of writing, both in the process of writing and in the act of reading the essay, develops, refines, and complicates the problem. It is a mode of learning (Emig, 1977). What looks simple becomes complex; what seems unsurmountable and opaque becomes clarified and comprehensible. In cutting short the writing process, in limiting the often-laborious revisions, rethinkings, and reorganizations that have long characterized academic writing, the advent of ChatGPT and other generative AI technologies presented what felt to many like an existential threat to the humanities and writing instruction (Baron, 2023).

Digital studies, media studies, and the digital humanities, of course, have long had an interest in exploring the theoretical dimensions of technology and culture, but the key questions raised by interactions with contemporary machine learning technologies have returned scholars to foundational theoretical questions and frameworks. Alexander Galloway's ideological analysis of distributed systems and protocols (Galloway, 2006) and Wendy Hui Kyong Chun's critique of metaphors in computation (Chun, 2011) or the function of social capital in network science are foundational contributions to these fields (Chun, 2021). Historical and theoretical work from Bernard Dionysius Geoghegan (2023) examining continuities between cybernetics and literary theory and from Orit Halpern (2022) on the social and political epistemologies of early machine learning and neural network research have been crucial to efforts to bridge theoretical and computational research. Early computational work, too, found resources in theoretical descriptions of language as a system for understanding the segmenting and processing of patterns in digital texts (Smith, 1989). While the computational turn and methods using natural language processing and machine learning had prospects for returning and reopening theoretical questions, it was the development of large language models and generative AI that has been the primary impetus for bridging computational and theoretical work. While in 2016 Hoyt Long and Richard Jean So could frame debates over the use of computational methods in terms of the reduction of humanities objects to data, generative AI's production of linear text as output (through decoding) obscures the underlying data model and minimizes this earlier anxiety and renders some of these prior arguments moot. Critiques are less likely in the generative AI moment to be framed in terms of the risks of "reducing the complexity of literary texts to mere 'data' or for being incommensurable with the goals of critical theory" (Long & So, 2016). With the development and widespread availability of generative AI applications, the barrier to entry has been lowered and new life has been given to the many key theoretical concepts and questions that informed earlier work in the humanities.

Structural linguistics and deconstructive theories of language and semiotic theories of image meaning making, in particular, have become compelling explanatory

frames for some researchers. Foundational theoretical claims about the meaning and function of authorship, such as Roland Barthes's "death of the author" or Michel Foucault's "author function" have been seen by some as valuable resources for conceptualizing the generated text and for reading the relationship between the prompt and the output (Coeckelbergh & Gunkel, 2024). The generated text has also produced exciting complications and new possibilities for rethinking what might have previously been considered a staid textual and literary hermeneutics (Henrickson & Merono-Peñuela, 2023). For others, theoretical claims about language and literature have become testable in ways that they were not in prior moments. As Leif Weatherby argues, "the claims of literary theory are suddenly subject to experimental hypothesis testing, because generative AI establishes a kind of statistical-textual condition that allows for experimentation on the literary features of language" (Weatherby, 2024). The advent of large multimodal models that contain networks mapping between learned features from images and text also promise to provide humanists with resources for the analysis of visual culture and rich objects to examine similarities across these representational spaces (Bajohr, 2024). The cross-disciplinary area of research known as model interpretability has brought together theorists of language and text, neuroscientists, and computer scientists to probe large language models, making progress in mapping the learned feature space of contemporary transformer-based neural networks and understanding the influence of training data, fine tuning, parameterization, and prompting on generated outputs. Methods from this developing work in model interpretability and explainability (frequently called explainable AI), such the use of model-agnostic perturbed inputs (Dobson, 2023b) or Shapley values for identifying feature attributions, have been used to probe models and understand "algorithmic bias not in terms of a single origin ('cause'), but with respect to a broader set of social and technical conditions at play that (re)produce these disparities" (Ziosi et al., 2024).

An emergent field known as critical AI promises to foreground theoretical and critical questions about the new generation of AI technologies in humanistic terms (Raley & Rhee, 2024). Joanna Zylinska makes the case for exploring outputs from image-based generative models by pointing to the way in which these models defamiliarize and expose the modeled world: "By returning our values and meanings back to us for further reflection, machines can help us see ourselves as embodied beings who exist in a sociocultural context" (Zylinska, 2023, p. 140). While some of this work focuses on familiar topics such as dataset construction and use (Bode & Goodlad, 2023), there are also entirely new methodological questions focused on the kinds of representations found in neural networks (Weatherby & Justie, 2022). In reviewing a computer science paper on convolutional neural networks, Fabian Offert argues that "machine learning models are cultural artifacts. They are trained on (limited) real-world data and often designed to make decisions with real-world impacts" (Offert, 2023). Many scholars are coming to realize that the deep learning transformer networks of today and other recent technologies are subject to the same critique as those from earlier generations (Dhaliwal et al., 2024). Theories and practices of photography informing the underlying data and perceptual

models (e.g., ImageNet), as Nicolas Malevé and Katrina Sluis argue in their analysis of computer vision pipelines, persist even in contemporary larger semisupervised models (Malevé & Sluis, 2023). Philosophers have also found renewed interest in interrogating the discourse of AI by problematizing offered ontologies, critiques of the methods and meaning of knowledge extraction, and its ideologies (Bunz, 2019; Parisi, 2019).

Theoretical work addressing conceptual problems arising from the modeling and generation of language with these networks connected to philosophical frameworks—especially those pertaining to agency, language, and authorship—are likely to continue to be of critical importance to any future for computational work in the humanities. That theoretical and philosophical intersection joined with insights derived from the operational mechanisms of the deep learning neural networks used in large language models and generative AI applications, as well as interpretive analyses of these networks and their outputs, will continue to shape the trajectory of computational work in the humanities.

References

Allington, D., Golumbia, D., & Brouillette, S. (2016). Neoliberal tools (and archives): A political history of digital humanities. *Los Angeles Review of Books*.

Archer, J., & Jockers, M. (2016). *The bestseller code: Anatomy of the blockbuster novel*. St. Martin's Press.

Arnold, T., & Tilton, L. (2019). Distant viewing: Analyzing large visual corpora. *Digital Scholarship in the Humanities, 34*(Supplement_1), i3–i16.

Bajohr, H. (2024). Operative ekphrasis: The collapse of the text/image distinction in multimodal AI. *Word & Image, 40*(2), 77–90.

Baron, N. S. (2023). *Who wrote this? How AI and the lure of efficiency threaten human writing*. Stanford University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623).

Berry, D. (2022). Critical digital humanities. In J. O'Sullivan (Ed.), *The bloomsbury handbook to the digital humanities* (pp. 125–135). Bloomsbury Academic.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.

Bode, K. (2018). *A world of fiction: Digital collections and the future of literary history*. University of Michigan Press.

Bode, K., & Goodlad, L. M. E. (2023). Data worlds: an introduction. *Critical AI, 1*(1-2).

Bunz, M. (2019). The calculation of meaning: On the misunderstanding of new artificial intelligence as culture. *Culture, Theory and Critique, 60*(3–4), 264–278.

Busa, R. A. (2007). Foreword: Perspectives on the digital humanities. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A companion to digital humanities* (pp. xvi–xxi). Blackwell.

Chun, W. (2011). *Programmed visions: Software and memory*. MIT Press.

Chun, W. (2021). *Discriminating data: Correlation, neighborhoods, and the new politics of recognition*. MIT Press.

Coeckelbergh, M., & Gunkel, D. J. (2024). ChatGPT: deconstructing the debate and moving it forward. *AI & Society, 39*, 2221–2231. https://doi.org/10.1007/s00146-023-01710-4

Cordell, R. (2020). Speculative bibliography. *Anglia, 138*(3), 519–531.

Crymble, A. (2021). *Technology and the historian: Transformations in the digital age*. University of Illinois Press.

Denson, S. (2020). *Discorrelated images*. Duke University Press.

Derrida, J. (1987). In G. Bennington, & I. McLeod (Eds.), *The truth in painting*. University of Chicago Press.

Dhaliwal, R. S., Lepage-Richer, T., & Suchman, L. (2024). *Neural networks*. University of Minnesota Press.

Dobson, J. E. (2019). *Critical digital humanities: The search for a methodology*. University of Illinois Press.

Dobson, J. E. (2023a). *The birth of computer vision*. University of Minnesota Press.

Dobson, J. E. (2023b). Objective vision: Confusing the subject of computer vision. *Social Text, 41*(3), 35–55.

Drucker, J. (2011). Humanities approaches to graphical display. *DHQ: Digital Humanities Quarterly, 5*(1).

Elkins, K. (2022). *The shapes of stories*. Cambridge University Press.

Emig, J. (1977). Writing as a mode of learning. *College Composition & Communication, 28* (2), 122–128.

Eve, M. P. (2019). *Close reading with computers: Textual scholarship, computational formalism, and David Mitchell's Cloud Atlas*. Stanford University Press.

Evens, A. (2024). *The digital and its discontents*. University of Minnesota Press.

Fazi, B. A. (2018). *Contingent computation: Abstraction, experience, and indeterminacy in computational aesthetics*. Rowan and Littlefield.

Foreman, P. G., Casey, J., & Patterson, S. L. (2021). *The colored conventions movement: black organizing in the nineteenth century*. Chapel Hill: University of North Carolina Press.

Gadamer, H.-G. (2004). In J. Weinsheimer, & D. G. Marshall (Eds.), *Truth and method*. Bloomsbury Academic.

Gallon, K. (2016). Making a case for the black digital humanities. In M. K. Gold, & L. F. Klein (Eds.), *Debates in the digital humanities 2016* (pp. 42–49). University of Minnesota Press.

Galloway, A. (2006). *Protocol: How control exists after decentralization*. MIT Press.

Galloway, A. (2014). *Laruelle: Against the digital*. University of Minnesota Press.

Garcia-Zorita, C., & Pacios, A. R. (2017). Topic modelling characterization of Mudejar art based on document titles. In *Digital scholarship in the humanities*.

Genette, G. (1987). *Seuils*. Editions du Seuil.

Geoghegan, B. D. (2023). *Code: From information theory to French theory*. Duke University Press.

Guillory, J. (2016). Monuments and documents: Panofsky on the object of study in the humanities. *History of Humanities, 1*(1), 9–30.

Guldi, J. (2023). *The dangerous art of text mining: A methodology for digital history*. Cambridge University Press.

Halpern, O. (2022). The future will not be calculated: Neural nets, neoliberalism, and reactionary politics. *Critical Inquiry, 48*(2), 334–359.

Hayles, N. K. (2004). Print is flat, code is deep: The importance of media-specific analysis. *Poetics Today, 25*(1), 67–90.

Henrickson, L., & Meroño-Peñuela, A. (2023). *Prompting meaning: A hermeneutic approach to optimising prompt engineering with ChatGPT.* AI & SOCIETY.

Impett, L., & Offert, F. (2022). There is a digital art history. *Visual Resources, 38*(2), 189–209. https://doi.org/10.1080/01973762.2024.2362466

Jockers, M. (2013). *Macroanalysis: Digital methods and literary history.* University of Illinois Press.

Johnson, J. M. (2018). Markup bodies: black [life] studies and slavery [death] studies at the digital crossroads. *Social Text, 23.*

Kirchner, E., Van Der Lans, I., Ligterink, F., Geldof, M., Megens, L., Meedendorp, T., Pilz, K., & Hendriks, E. (2018). Digitally reconstructing Van Gogh's *field with Irises near Arles* part 3: Determining the original colors. *Color Research & Application, 43*(3), 311–327.

Kirschenbaum, M. G. (2010). What is digital humanities and what's it doing in English departments? *ADE Bulletin, 150,* 1–7.

Liu, A., Droge, A., Kleinman, S., Thomas, L., Baciu, D. C., & Douglass, J. (2022). What everyone says: Public perceptions of the humanities in the media. *Dædalus, 151*(3), 19–39. https://doi.org/10.1162/daed_a_01926

Long, H., & So, R. J. (2016). Literary pattern recognition: Modernism between close reading and machine learning. *Critical Inquiry, 42*(2), 234–267.

Luhmann, J., & Burghardt, M. (2022). Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology, 73*(2), 148–171.

Malevé, N., & Sluis, K. (2023). The photographic pipeline of machine vision; or, machine vision's latent photographic theory. *Critical AI, 1*(1–2).

Manovich, L. (2020). *Cultural analytics.* MIT Press.

Moretti, F. (2013). *Distant reading.* Verso.

Moretti, F. (2017). *Canon/Archive: Studies in quantiative formalism from the Stanford Literary Lab.* New York: N+1 Books.

Offert, F. (2023). Can we read neural networks? Epistemic implications of two historical computer science papers. *American Literature, 95*(2), 423–428.

Parikka, J. (2012). *What is media archaeology?* Polity Press.

Parisi, L. (2019). Critical computation: Digital automata and general artificial thinking. *Theory, Culture & Society, 36*(2), 89–121.

Piper, A. (2018). *Enumerations: Data and literary study.* University of Chicago Press.

Raley, R., & Rhee, J. (2024). Critical AI: A field in formation. *American Literature, 95*(2), 185–204.

Ramsay, S. (2014). The hermeneutics of screwing around; or what you do with a million books. In K. Kee (Ed.), *Pastplay: Teaching and learning history with technology* (pp. 111–120). University of Michigan Press.

Ries, T., Van Dalen-Oskam, K., & Offert, F. (2023). Reproducibility and explainability in digital humanities. *International Journal of Dental Hygiene, 5*(2–3), 247–251.

Risam, R. (2021). Digital humanities. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (Eds.), *Uncertain archives: Critical keywords for big data.* MIT Press.

Schreibman, S. (2002). Computer-mediated texts and textuality: Theory and practice. *Computers and the Humanities, 36*(3), 283–293.

Smith, J. (1989). Computer criticism. In R. Potter (Ed.), *Literary computing and literary criticism: Theoretical and practical essays on theme and rhetoric* (pp. 13–44). University of Pennsylvania Press.

Sobchack, V. (2011). Afterword: Media archaeology and Re-presencing the past. In E. Huhtamo, & J. Parikka (Eds.), *Media archaeology: Approaches, applications, and implications*. University of California Press.

Steyerl, H. (2009). In defense of the poor image. *e-flux Journal, 10*. https://www.e-flux.com/journal/10/61362/in-defense-of-the-poor-image.

Svensson, P. (2014). Humanities computing as digital humanities. In M. Terras, J. Nyhan, & E. Vanhoutte (Eds.), *Defining digital humanities: A reader* (pp. 159–186). Farnham and Burlington: Ashgate.

Svensson, P. (2016). Digital humanities as a field. In *Big digital humanities: Imagining a meeting place for the humanities and the digital*. University of Michigan Press.

Tenen, D. (2017). *Plain text: The poetics of computation*. Stanford University Press.

Terras, M., Nyhan, J., & Vanhoutte, E. (Eds.). (2013). *Defining digital humanities: A reader*. Routledge.

Underwood, T. (2019). *Distant Horizons: Digital evidence and literary change*. University of Chicago Press.

Wasielewski, A. (2023). *Computational formalism: Art history and machine learning*. MIT Press.

Weatherby, L. (February 26, 2024). *Artificial intelligence and the significance crisis*. The Chronicle of Higher Education.

Weatherby, L., & Justie, B. (2022). Indexical AI. *Critical Inquiry, 48*(2), 381–415.

Wernimont, J. (2021). Quantification. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (Eds.), *Uncertain archives: Critical keywords for big data* (pp. 427–431). MIT Press.

Wernimont, J., & Flanders, J. (2010). Feminism in the age of digital archives: The women writers project. *Tulsa Studies in Womens Literature, 29*(2), 425–435.

Ziosi, M., Watson, D., & Floridi, L. (2024). A Genealogical approach to algorithmic bias. *Minds and Machines, 34*(9). https://doi.org/10.1007/s11023-024-09672-2

Zylinska, J. (2023). Art in the age of artificial intelligence. *Science, 381*(6654), 139–140.