

Facets of Friction: Investigating epistemological friction between computing and the humanities to support Digital Humanities computing education

Anna Sollazzo <anna_dot_sollazzo_at_glasgow_dot_ac_dot_uk>, University of Glasgow  <https://orcid.org/0009-0008-1085-1999>

DOI: <https://doi.org/10.63744/xrxmu33g3t6t>

Abstract

Learning computing can often be challenging to Digital Humanities (DH) students because its knowledge-production practices and values do not necessarily align with those of humanistic inquiry. Further, the breadth of the technologies used in the field can make it difficult to identify which computational ideas and skills are most essential to teach in DH. This paper argues that, in order to teach students to enact a paradigm of 'humanistic computing' which is reflective of humanistic interests and ways of knowing, it is necessary to centre the epistemological friction that exists between computing and the humanities. It proposes a four-faceted framework to aid in the identification and exploration of that friction, with the intention that it be used by DH educators to structure their computing curriculum – regardless of the technologies or level of abstraction it engages with – and by students or even practitioners to scaffold a critical approach to computational DH work.

Introduction

Computing can be understood as the mapping of human conceptions of structure and process to machine interpretable ones. Computing education pioneer Alan Perlis asserted that the aim of any computing course should be to “reveal [...] how analysis of some intuitively performed human tasks leads to mechanical algorithms accomplishable by a machine” [Greenberger and others 1962]. Learning computing is a question of interrogating this process of translation.

1

When translating between natural languages, the best translation is not one which preserves exact vocabulary, but rather one which preserves tone and intention. Rendering such a translation depends on understanding the structures that govern the languages involved, which are a reflection of the way in which the cultures from which they emerged view and makes sense of the world [Kramsch 2014]. This remains true with respect to the translation of structure and process.

2

As C.P. Snow argued, academic disciplines *are* cultures [Snow 1959]. Karin Knorr Cetina (1999) adds a level of precision to Snow’s definition in describing academic fields specifically as *epistemic cultures*, naming epistemology as the defining factor [Cetina 1999]. Epistemology is concerned with how we know what we know and why we believe it. Domain cultures are characterised by what types of knowledge and ways of knowing a domain considers important and legitimate. A domain’s *epistemic machinery* is the specific set of methodologies and tools used for knowledge creation and validation; they are a concrete manifestation of its epistemology.

3

Natural language translation becomes more complex as a function of the cultural distance between languages. Translating between French and Italian, for instance, is fairly straightforward. They emerged from cultures with similar values and ways of understanding the world. This is reflected in the similar structure of their languages, making it

4

relatively simple to find equivalencies. Now consider a translation between French and languages such as Diné (Navajo), in which the level of animacy of agents — understood according to a continuum which reflects indigenous beliefs — impacts many structural elements [Young, Morgan, and Young 1987] ; or Yolmo, a Peruvian language with evidentiality, meaning that the way a speaker knows things must be encoded in the grammar [Gawne 2014]: in these cases, ideas and information are much more likely to be lost or altered in translation. Similarly, computing and humanistic disciplines are more culturally — epistemologically — different than, say, computing and mathematics. This distance is a central challenge of computing in the Digital Humanities (DH).

DH comprises the “the thoughtful use of computing in humanistic inquiry and the thinking through of computing from the perspective of the traditions of the humanities”^[1] Some might describe it as trying to use computers to tackle “non-computational” problems. However, there is nothing inherent to the problems that makes them “non-computational”; rather, it is that the priorities and interests of humanities inquiry are different from those of the scientific culture from which computing emerged and which its associated tools and procedures correspondingly embody [Drucker 2012]. Consequently, doing computing in a way that preserves humanistic outlooks and values requires attention to be explicitly paid to points of epistemological friction.

If the broad aim of computing education is to gain an understanding of the translation process, then DH-specific computing education should centre the interrogation of epistemological friction. In support of this aim, this paper proposes a framework for describing and modelling friction between computing and the humanities. This work is subject to two simplifying assumptions. First, though it is evidently not the case, the humanities are treated as a homogeneous entity, and one which includes disciplines such as history which could also be classified as social sciences. Second, *computing*, in the abstract, is associated with *computer science* and its corresponding knowledge culture, as this is how it is situated both in the majority of educational institutions as well as, to draw a broad generalisation, in the consciousness of the lay public.

The framework comprises four facets of epistemological friction:

- **Negative Knowledge:** This facet looks at what disciplines choose to know, can not know, and the errors they make in knowing, exploring how these elements are viewed and incorporated into knowledge production practices.
- **Positionality and Context:** A domain’s epistemology is inextricably linked to its conception of the agency of knowledge creator(s) and the influence of perspective. This second facet grapples with the dissonance between detached scientific knowledge and situated humanistic knowledge, and interrogates how computational techniques both resist and can be made compatible with uniqueness, plurality, ambiguity, and nuance.
- **Relationship to Tools:** Technology can be understood as impartial and a means of circumventing human bias or as a product and reflection of human culture and values. The third facet explores these different conceptions and how they translate to the use of technology in knowledge production, with particular attention to notions of objectivity and independence, and a tool’s capacity to (intentionally or unintentionally) influence and alter work.
- **Hermeneutic Attention:** The final facet considers the tension between a scholarly tradition whose procedures aim to eliminate the need for interpretation and one to which it is central.

The framework is principally intended for use by DH computing instructors and students — to guide the development of curriculum and scaffold critical practice. An instructor looking to introduce databases might choose to structure their lesson around *Positionality and Context* and *Negative Knowledge*, having their students create schemata grounded in specific perspectives and analyse what is highlighted and lost in each. A student applying a text analysis tool might be encouraged, through use of the facets as framework for analysis, to question the extent to which their intentions match up with the structural and procedural logics imposed by the computational medium (*Relationship to Tools*); the use of the facets as a reflection tool when creating a visualisation might also lead a student to more explicitly incorporate a representation of their process of interpretation into the digital product (*Hermeneutic Attention*).

The facets are not expected to be comprehensive but highlight common sites of friction which may play into a variety of DH projects. More importantly, they illustrate the concept of epistemological friction in the abstract, such that individuals might extend and adapt the framework to their own contexts.

9

Negative Knowledge

Artist Mimi Onuoha has a filing cabinet. Several, in fact — an oddity in the age of increasingly digital paperwork. But the increasing digital-ness of society is precisely why they exist. Onuoha's cabinets are full of folders with labels such as 'People excluded from public housing because of criminal records,' 'White children adopted by POC' and 'Quantifiable effect of corruption in lean economies.' Every single one is empty. Onuoha calls her collection the *Library of Missing Datasets*; the folders represent collections of data that do not exist — because those with the resources to create them lack the incentive, because the subjects do not lend themselves to favoured modes of data collection, because the work collection would require outweighs the perceived benefit, or because there are advantages to non-existence.^[2] Ohunua's artistic practice turns on what is forgotten when "globalised, quantified societies require the fluid messiness of people to be made into legible forms of data"; she asserts that "What is missing is still there. This is where my work begins."^[3]

10

Karin Knorr Cetina coined the term *Negative Knowledge* to designate not "nonknowledge" but rather "knowledge of the limits of knowing, of the mistakes we make in trying to know, of the things that interfere with our knowing, of what we are not interested in and do not want to know" [Cetina 1999, 64] We might consider the core question of epistemology to be "How and why do we know what we know?" but consideration of the inverse — the photo negative — equally stands to outline the characteristics of a given epistemic approach. Different domains opt, to varying degrees, to disregard specific knowledge so as to enable the creation of strong knowledge claims and ontologies of what is knowable, or to leverage missingness and ambiguity to generate avenues of exploration. Onuoha's work specifically points to the negative as a notable friction point when it comes to the intersection of computers and humanity.

11

What we choose not to know

A folder labelled "Nuanced racial descriptions of American residents — 1890" would have fit right in in the *Library of Missing Datasets*. The 1890 American census is noteworthy in that it saw the introduction of Herman Hollerith's tabulating machine — considered a direct ancestor of the modern computer. After being collected by surveyors, data was transferred to punch cards to be tabulated automatically. This new scheme required that all answers be discrete, or made to be so. Race^[4] was one of the characteristics encoded on the punch cards, and the 1890 census introduced several new options. The questionnaire was the first to include "Japanese," (previously there had been no distinction between East-Asian countries) and "Indian" (first nations). Where, since 1850, only the categories "Black" and "Mulatto" (any person of mixed race) had been included, the rise of the eugenics movement drove congress to mandate the introduction of additional categories based on blood quantum. The instructions given to the enumerators explained that an individual having 3/4 or more "black blood" should be labelled as "Black," 3/8-5/8 as "Mulatto," 1/4 as "Quadroon" and 1/8 or less as "Octoroon" [Lee 1993]. It is not clear the extent to which the precision of the categories was government dictated or a product of the need for clean delineations, but we can be more sure of two things: First, that the transfer of data to punchcards served to erase the presence of any uncertainty or ambiguity that might have been noted by surveyors; and, second, that the tabulating machine facilitated the imposition and propagation of seemingly strict racial categories — highly contextual political constructions reified by the machine — creating data that would be the foundation of government policies that would continually re-inscribe them.

12

13

"What we choose not to know" in computing is principally a question of encoding. The power of computing lies in the ability to manipulate information at a scale and speed beyond human capability. Prerequisite to this is an underlying model which operates at some level of abstraction — necessary to accommodate volume. The process of creating a reduced-complexity rendering of a real-world phenomenon is inherently lossy, a noted limitation to the "correctness" of computers [Smith 1985]. The challenge this presents is exacerbated in a humanities context as "the humanities

14

historically have adopted methodologies that are less amenable to abstract generalization than those used by the sciences”; they belong to a tradition of “telling detail” as opposed to “seeking generalizations” [Quamen 2012]. Further, the computational form “accepts only that which can be told explicitly and precisely” [McCarty 2005], a requirement which drives the choice of what to know. Correspondingly, the knowledge prioritised by the disciplines can be antipodal, the very specificity or variation which would make something an interesting object of study in the humanistic tradition [Ginzburg 1979] incompatible with normalisation. The strict encoding required by Hollerith’s machine — and, indeed, by modern computers — allowed for faster and larger-scale processing; the 1890 census received 25% more responses than the one in 1880 but was tabulated three and a half times faster. At the same time, it made negative context and ambiguity, and, correspondingly, not all the changes it encouraged, if not imposed, proved productive. The fine grained African-American racial classifications introduced in 1890 were removed before the 1900 census, “after census officials judged the data ‘of little value and misleading’” [Parker et al. 2016, 25]. Echoes of similar encodings, however, continue to persist — independently of the knowledge they made negative. In their investigation of contemporary datasets used for image recognition, for instance, Scheuerman et al. found that the racial categories used in the Microsoft Celeb database were “Caucasian,” “Mongoloid,” and “Negroid” [Scheuerman et al. 2020]. They wonder at the contrast of the use of these terms, which are evidently sociohistorical, and the fact that, in modern practice, “classifications of race [...] are portrayed as insignificant, indisputable, and apolitical.”

Faults identified in the use of such large Machine Learning (ML) datasets are often also a product of Negative Knowledge. ML models rely on the ample, working under the assumption that with enough volume comes generalisability. This is amusing when a model ends up relying on human fingers to identify fish because they are principally represented in the training data being held up by triumphant fishermen [Brendel and Bethge 2019], but distinctly more troubling when the widespread absence of racially diverse facial data [Buolamwini and Gebru 2018] means that proctoring software exhibits significant bias against students with darker skin tones [Yoder-Himes et al. 2022]. danah boyd and Michael Golebiewski’s notion of “data voids” further illustrates, on a structural level, how computational infrastructure is not designed to account for absence — as demonstrated by the fact that its negative spaces are its weakness. Data voids, as they define them, are places where search queries yielding few results leave space for exploitation by “media manipulators with ideological, economic, or political agenda” [Golebiewski and boyd 2019, 1]. These vulnerabilities “exist because of an assumption baked into the design of search engines: that for any given query, there exists some relevant content” [Golebiewski and boyd 2019, 16]. With ample data but no true understanding of content; these systems can only ever return what has been said, glossing over the nuance of novel queries. It is a question of volume, never of interpretation. This foundation in generalisation and macro-level aggregation characteristic of computing is not in and of itself problematic or even undesirable, but it can nonetheless create vulnerabilities whose severity is exacerbated by the fact that they may be less expected and accounted for than attacks such as DOS, which are also grounded in the computational facility for — and prioritisation of — scale. Somewhat ironically, the very principles which created the data voids are those that allow them to do damage. As attackers fill in identified gaps with targeted material, the increased volume of content causes it to proliferate widely and quickly. boyd and Golebiewski highlight that there is no fix for data voids; rather, it is a question of awareness — of their existence, of the circumstances of their creation, and of the potential consequences and how they might be mitigated.

A Humanistic data practice, as described by Ryan Cordell (2019), should be one that “seeks primarily to point our attention to the gaps in our digitization and critical practices” [Cordell 2019]. Cordell comes from the world of literary history; in this context, in addition to a more pronounced materiality, which stands to reify absence, there is a methodological tradition of archiving which foregrounds questions of provenance and politics of preservation undergirding phenomena such as canon formation. He insists that humanists must not lose sight of these perspectives when it comes to computational work, asserting that a humanistic approach “must begin with discussions of data construction rather than data analysis.” Attention to Negative Knowledge in DH relates not only to the critical interrogation of knowledge made negative by the use of computers, but also the ways in which the use of computers can bring to light knowledge made negative by the same human factors Onuoha describes with respect to digitisation — which, for all it presents unique challenges, is also simply another way of creating and preserving information that traces a specific narrative.

Cultural critic Walter Benjamin's conception of "aura" could be considered a type of Negative Knowledge. He argues that the aura of a piece of art, "its presence in time and space, its unique existence at the place where it happens to be," could never be captured by mechanical reproduction [Benjamin 2006, 114] — mechanical reproduction notably requiring similar decontextualisation and normalisation to computational transformation. Bruno Latour and Adam Lowe, to the contrary, argue that the aura of a work is not eroded but rather extended and enriched by its facsimiles [Latour and Lowe 2011]. The myriad reproductions together form an *assemblage*: a collection of sources that, though each analysable in isolation, when considered in conjunction with one another "have irreducible properties [...] that emerge from the interactions between parts" ([DeLanda 2019, 10] Each component part of the assemblage has its own trajectory — story of how it moved through the world — and, though together they "do not form a seamless whole," their conjunction and interaction is generative, continually adding to the aura. What distinguishes digital texts and, more broadly, digital collections, from their analog equivalents is their *massive addressability*. They are flexibly interrogable from myriad angles and levels of granularity, which facilitates linking [Witmore 2010]. Further, there is a flattening of hierarchies born of the equal accessibility of higher- and lower- level components; [Manovich 2013] argues, for instance, that a defining characteristics of databases as a New Media form is the way in which they invert the relationship between syntagm and the paradigm. They "materialize" the myriad possibilities (paradigms) represented by an individual row, diminishing the dominance of an isolated linear syntagm. A digital collection, with its emphasis on individual trajectories and linking, could correspondingly be understood as an assemblage; for all that computational tools can make negative some knowledge — just as mechanical reproduction did, especially with respect to uniqueness — the kinds of multi-level and combinatoric interrogations they permit can help expose and explore different sites of missingness.

In a name-based search of the *Papers of Thomas Jefferson Digital Edition* — a digital collection of a significant portion of the man's correspondence — Lauren Klein found no results for James Hemings, despite the central role the enslaved man played in Jefferson's diplomatic dealings in his capacity as household chef [Klein 2013]. A global search turns up a single first-name mention, identifiable as Hemings only through a researcher's metadata note. This archive and its silences are a product of historical power dynamics; the documents preserved may in rare cases be written about marginalised populations, but are even more seldom written by them. Any effort to recover these silences using its documents or others subject to the same politics of creation and conservation (e.g. ledger books, sales receipts) risks "reinforcing the damaging notion that African American voices from before emancipation — not just in the archival record, but the voices themselves — are silent, and irretrievably lost" [Klein 2013, 665]. Modern humanities scholars have moved away from trying to fill archival silences to instead explicitly highlighting them in a way that invites critical speculation about the past; the shift is to a conceptual model that "works by illumination rather than demystification" and, "through explication rather than appropriation or empowerment" [Klein 2013, 668]. Klein demonstrates how digital techniques can be turned to this task. Using named entity recognition and co-reference extraction to derive relationships from letter *contents*, Klein created a "deformation" of the archive that disrupts the structures of power it inscribes: "Rather than privilege the relationships between letter writers, I sought to dismantle the letter as the unit of the archive, examining each word of content on equal plane" [Klein 2013, 672]. Unlike the links drawn in a visualisation of the correspondence network of the starting archive, "the arcs that link Jefferson to the men and women he enslaved" in an interaction network visualisation Klein created from her new version of the data "are much more prominent than those that link him to his family members and friends." This, she asserts, "conjures a sense of the dependence, on the part of Jefferson, on the men and women he enslaved, even as it cannot recreate what these people said in their conversations, where they went in order to conduct their transactions, and how they truly lived their everyday lives" [Klein 2013, 674].

Closing the circle, the massive addressability of digital objects can be used in service of inquiry which explicitly focuses on absence. For Deb Verhoeven, the absence in question is that of women in creative roles in the film industry. Verhoeven and her team use speculative network analysis to re-imagine film industry DEI policies, the premise being that network analysis puts "emphasis on relationships instead of individuals" [Verhoeven et al. 2020, 1] and is correspondingly suited to targeting systemic inequalities. They do not look purely at what *is* there, but at what *is not* but *could be*, adding nodes and vertices representing hypothetical diverse creatives and their relationships to existing collaboration networks and analysing how these changes impact the "health" of the network by comparing its overall

metrics to those of notably more equitable regional networks.

Foundationally humanistic work like Klein and Verhoeven's is not possible without the high-level manipulation and rapid calculation made possible by computation. Its use, however, requires a specific type of problem modelling — one whose focus on abstraction and generalisation can butt up against the humanities interest in nuance and specificity. This is a tension which must be recognised and navigated in DH work.

20

What can not be known

Every area of knowledge production deals with uncertainty. Potter et al. divide it into two high-level types. The first is *Aleatoric* uncertainty, which they define as “random uncertainty inherent to the problem.” Negative Knowledge is more closely tied to the second type, *Epistemic* uncertainty, which designates things “which could, in principle, be known, but in practice are not”; this type of uncertainty is “introduced through deficient measurements, poor models, or missing data” [Potter, Rosen, and Johnson 2012]. Disciplines develop their own distinct ways of representing and accounting for uncertainty, which, being derived from domain scenarios, are a product of their own epistemic standpoints. Given the differences in how knowledge creation and validation functions between disciplines, these techniques and standards can not necessarily be ported from one context to another.

21

The scholars of *Project Cornelia* encountered this challenge with respect to their data modelling relationship networks of seventeenth century Flemish painters [Brosens et al. 2019]. Art history has a record of using “best-guess” imputation to fill data gaps, with individual scholars proposing possible values based on available related sources. Founded as they are in “insights gained from literature, expertise, gut feeling, and ‘common sense,’” such imputations are “often imprecise and divergent” [Brosens et al. 2019, 117]; they do not possess the necessary conditions for replication, which is frequently associated with knowledge validation in computing. The *Cornelia* team assert that “while art history’s traditional ad hoc strategies and methods to tackle missing data can certainly be used to a certain extent to address qualitative and micro-historical questions, they fall short of meeting our data-driven research agenda” [Brosens et al. 2019, 118]; they instead turned to statistical imputation. Single imputation, they discovered, even when run on processed data subsets intended to be favourable to the technique through constructed homogeneity, “increased rather than reduced the level of uncertainty in the data set” [Brosens et al. 2019, 119]. The difficulty of imputation in cases like *Project Cornelia*’s is that even small variances can drastically alter the meaning of the data. Brosens et al. give the example of the missing birthdate of artist Antoon Salert; depending on whether he entered the network in 1579 or in 1590 — both values generated by variations of single imputation underwritten by different mathematical functions — he might have been considered a contemporary of renown painter Peter Paul Rubens or as belonging to the first generation of painters following in Ruben’s footsteps. The researchers’ next recourse was to attempt the more complex technique of multiple imputation, teaming up with biostatisticians in whose work the technique is commonly used. Multiple imputation, as the name suggests, utilises single imputation across multiple datasets to produce a mean value that is statistically valid. They found that “the idiosyncratic nature of the Cornelia data and the high levels of ‘missingness’” made even the results of this more robust technique varyingly useful.

22

Rebecca Sutton Koeser and Zoe LeBlanc describe this complicating factor of prevalent “missingness” as a “phenomenon that is endemic to humanistic scholarship” [Koeser and LeBlanc 2024]. In light of this inescapable characteristic of humanistic contexts, and in the knowledge that, as Enjoli alludes to, “archival silences are magnified in digitization projects,” they propose that DH could adopt a practice of speculative reading that moves “the margins of our analyses to the center,” framing “missingness as something that we can consider as an object of study outright.” The analyses they undertook in this line of inquiry did not look dramatically different from those of the Cornelia team; Koeser and LeBlanc used statistical methods, emerged from scientific and mathematical disciplines to make guesses about lacunae in humanistic datasets — in their specific case, the membership and borrowing records of Paris bookshop and library Shakespeare and Company. They are conscious, in their work, to maintain methodological ties to the humanistic, always contextualising as a means of understanding the efficacy and limits of their work. In their attempt to use forecasting methods typically employed in finance to estimate missing borrowing activity, they identified that the predicted bottom line, which seems high relative to subscriptions, could potentially be explained by the bookstore owner’s propensity — as ascertained through additional documents — for granting but not documenting free

23

subscriptions. Koeser and LeBlanc also experimented with the application of techniques used by ecologists to correct for bias originating from the exclusion of species that are difficult to observe to the problem of missing books, but acknowledge that, for all there have been parallels drawn between cultural and ecological diversity, there are assumptions in the unseen species model that are not analogous to their problem — notably the rate of change in the system and the probable frequency difference between a book, which they point out could quite reasonably be borrowed only once, and a species for which multiple sightings are expected. Another of their experiments, a recommender-system inspired model aiming to predict possible borrowing history during a period of missing records, turned out to be quite accurate, the author's personal correspondence confirming a number of its suggestions for Ernest Hemingway.

Even when, as is the case with the Hemingway predictions, their results were proven correct, Koeser and LeBlanc emphasise that correctness, or even the results, are not the point. In the practice of speculative reading, the models are intended to “provide a means to experiment in knowledge creation.” They propose that missingness should not be considered merely a “technical category for measurement” but rather positioned as “a conceptual framework” for considering data. They align themselves with the Uncertain Archive Project's aim of “carv[ing] out an epistemic space for modes of inquiry that are motivated less by providing answers than by posing questions, exploring uncertainties, and offering material and speculative approaches” [Agostinho et al. 2019, 432]. Under this paradigm, missingness is not something bothersome to be accounted for, but something to be used generatively.

24

Errors in knowing

Scientific knowledge production and validation are strongly rooted in a logic of trial and error. There is an expectation of potential sources of error being reported and, through replication, discounted. Some argue that, by virtue of its disciplinary heritage, replication ought to be incorporated into the scholarship of DH [Tiefenbach Keller 2024]; however, as with uncertainty, the portability of epistemic practices merits consideration.

25

In his 2022 study intended as a replication of Langer et al.'s finding that “the frequency and diversity of biological taxa [in fiction] have been declining steadily since the first half of the nineteenth century,” DH scholar Andrew Piper's primary conclusion was that he disagreed with the way in which the authors of the initial paper defined fiction in the first place [Piper 2022]. One part of Piper's study comprised what we might call a “traditional” replication — testing the initial methods on the same data. In this particular case, however, though also drawn from Project Gutenberg, Piper's data differed somewhat from Langer et al.'s. Though he reports applying the filtering criteria specified in the original paper, Piper's search yielded only 13,075 texts, as compared with the original 15,238. In running his approximation of their dictionary-based biological term instance identification model on this corpus, Piper found routine discrepancies of ten to fifteen years with regard to inflection points and, when he ran the experiment on a subset of the data constructed using a different means of identifying “fiction,” he observed the opposite of Langer's reported trend. Piper also undertook a conceptual replication [Tiefenbach Keller 2024], introducing an additional ML-based model and two additional datasets to investigate the result in question. These produced further contradictory trends. For Piper, these results cannot be understood as disproving the initial result; he asserts that they are, rather, a starting point of alternative interpretations of the subject matter and, more abstractly, a case to drive discussion of methodological questions in cultural analytics.

26

In her historical examination of the notion of learning from errors in science, Jutta Schickore describes two distinct types of error-related practices: arguments from error and enquiries into error. The former consists of “those methodological arguments that the practitioners put forward to indicate that their research took the potential sources of error into account” [Schickore 2005, 554]; conventions such as error bars and the practice of direct replication fall into this category. Piper's exploration, in particular the conceptual replication, aligns more closely with the latter type, where “errors are not merely uncovered to be cast aside” but, rather, “become the object of systematic exploration and discourse” [Schickore 2005, 555]. This outlook is echoed by DH literary scholars Hoyt Long and Richard So with respect to their work using machine learning to explore the evolution of the Haiku as a distinct poetic form [Long and So 2016]. They explicitly leveraged errors, in the understanding that they occurred at points of class overlap, asserting that “what the machine learning literature treats as misclassifications [...] we treat as opportunities for interpretation” [Long and So 2016, 261].

27

Negative knowledge, when it comes to computing, is often a product of formalisation — whether it be the strict encoding of datafication or the normalisation of messy and lacunar data required by model building. An epistemologically-aware approach allows for this formalisation to move beyond being “simply a necessary and straightforward condition for computation” to become “a rich and productive concept to think about computation, the humanities, and their encounters” [Van Zundert et al. 2012, 280]. Negative knowledge, can be a tool to generate discourse [Cetina 1999] — an awareness of what is not there providing a new lens for investigating what is. In their theorization of Humanistic HCI, Shaowen and Jeffrey Bardzell, propose speculative “design futuring” as an emblematic practice; they specifically tie this exercise of “interpretively explor[ing] alternate worlds, to discover the possible and the preferable in them, and to construct both pathways and the collective will to pursue them” [Bardzell and Bardzell 2016, 26] to the notion of emancipation, which they assert is a fundamental aim of humanistic knowledge production. For all that computational tools and methodologies can be the instigators of oppression, they can — with an awareness of Negative Knowledge — equally be used to foster and facilitate imagining otherwise.

28

Positionality and Context

On July 11th 2021, Maxime Bernier, the leader of the People's Party of Canada tweeted^[5]:

29

There are no feminist math, indigenous math, black math, or any other kind of math.

30

There is only mathematics.

31

Only physics.

32

Only astronomy.

33

He was quickly met with a rebuttal from mathematician, educator, and researcher Michole Enjoli, in a thread^[6] beginning:

34

Whew. I've been waiting for this one.

35

cracks knuckles

36

WHERE and WHO we get mathematical knowledge from, shapes the content of the math being taught.

37

Enjoli asserts that “The belief system of a society motivates which math is taught,” pointing to the work of Education scholar Gloria Ladson-Billings who argues that the values being promoted in the U.S. system echo those of the dominant White culture which “demands efficiency, consensus, abstraction, and rationality” [Ladson-Billings 1997, 699]. She further highlights the work of mathematics education scholar [Martin 2013], who argues that the kinds of mathematics emphasised by the current U.S. system are a reflection of the value it places on economics: “...If we lived in a society with a different set of values,” concludes Enjoli, “then our mathematics and the education system along with it would again look different and be used differently.”

38

The reaction to Enjoli's response testified to the fact that it was a complete novelty to many who, though they disagree with Bernier's far-right politics, initially found themselves nodding along in recognition of his statement.

39

Knowledge validation in science rests on uniformity and replicability. This implies a certain separation of the methods and procedures from the people who perform them and the circumstances in which they are undertaken. Though there is interpretation of results that pushes knowledge forward, the knowledge is 'held' in the resulting detached axioms. This severing of knowledge from creator and context is what makes it easier for a specific cultural construction of Math to be viewed as 'just Math.'

40

Humanistic knowledge production, to the contrary, is not looking to achieve consensus in findings, but to create a dialogue between them. This entails a focus on individual nuance and specificity, which is tied up in context of creation. Further, critical interpretation is the core of its knowledge creation; correspondingly, validity is born of acknowledging and incorporating the influence of the creator and their positionality into its epistemic practices rather than designing

41

them to remove it. As DH scholar Harvey Quamen describes, “charges of anthropocentrism or ethnocentrism can often be read, especially by scientists, as a need for greater objectivity, but in humanistic circles they are more often assertions that the anthropos and the ethnos remain central to scholarly inquiry” [Quamen 2012, 10]. These contrasting approaches to positionality and context are an inevitable source of friction in DH work.

Nowhere and Somewhere

The lab is a central space of scientific knowledge production. Objects studied in the lab are necessarily removed from their spatial and temporal contexts, and it is possible to manipulate the usual restrictions that those impose. For example, scientists might simulate conditions to accelerate the growth cycles of bacteria or alter the circadian rhythms of mice. It is also permissible and expected to “substitute transformed and partial versions” [Cetina 1999, 27] of an object for the whole. Though computing could be said to lack the same flavour of lab culture, the computer itself stands to act as a lab and encourage a parallel detachment; data, as seen, is a partial and decontextualised representation, and processes such as simulations similarly perform artificial and normalised manipulations. Lab procedures are stringently codified in an effort to minimise space for individual influence. As historian of science Theodore Porter describes, [a] highly disciplined discourse helps to produce knowledge independent of the particular people who make it” [Porter 1995, ix]. Computing, again, emphasises these properties. It enforces a strict encoding and affords quantification, which is “a technology of distance” [Porter 1995, xxi]. Further, in facilitating high-scale operation, it minimises the impact of individual deviation. Uniformity is rigorous; human bodies and minds are not uniform. They become more so, however, when viewed from a distance, which is to say their specificity, as either the observer or the observed, is abstracted away.

42

These kinds of “realist” models of knowledge, asserts Johanna Drucker, “depend above all upon an idea that phenomena are observer-independent.” The aim is to take a “view from nowhere,” the implication being that the act of observation — its context and the particular standpoint of the observer — should not have an impact on the phenomenon. This detachment is constructed through standards and procedures. However, “[r]endering observation [...] as if it were the same as the phenomena observed collapses the critical distance between the phenomenal world and its interpretation,” which “und[oes] the basis of interpretation on which humanistic knowledge production is based” [Drucker 2011]. Feminist standpoint theory contends that all knowledge is inherently positional, dependent on individual identities and structural power dynamics. As a counterpoint to scientific objectivity, Donna Haraway offers Feminist objectivity, which “means quite simply situated knowledges” [Haraway 1988, 581]. This outlook is emphatically not an effort to account for every possible perspective — in the way that its counterpart looks to eliminate them all — but a recognition of the existence of the knowledge creator and the incorporation of their positionality and context into praxis. This notion of a view from somewhere is congruent with interpretive — and, correspondingly, humanistic — knowledge production practices.

43

Roopika Risam’s comparative analysis of two visualisations of global migration illustrates how these contrasting understandings can manifest in digital artifacts [Risam 2019]. The first, *The Flow Towards Europe (FTE)*, was produced by a Finnish startup using a dataset from the United Nations Refugee Agency. It depicts migration towards Europe from African, Middle Eastern, and Asian countries with waves of dots moving in straight lines from source to destination. It is built around a map projection centred on the global North, with a black and turquoise colour scheme that serves to clearly highlight country borders; its authority is drawn from quantification at scale and the use of established conventions. *Crossing the Mediterranean Sea by Boat (CMSB)*, her second example, conversely looks to convince through a focus on individual experience. CMSB is a collection of story maps built around interviews with over 250 migrants. Representations of migration are accompanied by motivations (e.g. “escaping sexual violence,” “escaping civil war”) and navigation between migrant stories is based on familial and relational bonds (e.g. “brother,” “young mother”), “offer[ing] humanizing representations of migrants by placing them within social relations” [Risam 2019, 573].

44

Each story map includes stops made along the way and integrates the voices of migrants through quotations describing their journeys. Risam highlights that the visualisation does not shy away from assuming a specific position; alongside the visualisations, it directly puts questions to the user such as “Unauthorised journeys through Europe are often frayed with dangers, and many have died trying to reach their destinations. Does this change your views on the need to open

45

legal ways for people to travel?” and “On arrival to their destination, many wait long periods during the asylum process while trying to learn about the rules and regulations in a new country. Can you imagine what it feels like to be waiting in this situation?” The crux of her argument, however, is that the choices underlying FTE relating to positionality and context — of the researchers and of the subjects — are no less of an assumed stance than those more plainly legible in CMSB. She argues that the “[u]se of a political map absent of topographical features emphasizes national borders and depicts cohesion of the nation state (however fictive)” [Risam 2019, 572]. The choice of the map also creates a focus on the impact of migration on the destinations — some origin points are even out of frame. Risam further points to the fact that the movement of the dots in “unimpeded waves” serves to “emphasiz[e] the magnitude of migration” but does not reflect how migration “take[s] many more forms than travel from point A to point B [and] does not occur uniformly over time in the flows depicted in the visualisation” [Risam 2019, 572–73]. In order to produce the uniformity needed to apply the chosen methods, the team behind FTE created departure dates not included in the original dataset and calculated using arrival dates and average human walking speed. They also chose to aggregate the migrants; each dot actually represents a group of 25 individuals who were not necessarily travelling together. This decision was notably “determined by the limitations of the visualization platform and the creators’ sense that if each migrant were represented by a dot, the visualization would be too crowded and create performance issues” [Risam 2019, 572].

The choices made by the FTE team are a reflection of an epistemic tradition which prioritises generalisability and associates validity with detachment. Computational mechanisms and conventions are instantiations of this position, and therefore in tension with the values of humanistic inquiry, for all that they can be used in service of it when the humanities, too, are interested in large numbers and high-level patterns. Historian of Slavery Jessica Marie Johnson, for instance, argues that while digital resources such as the Trans-Atlantic Slave Database “immediately reshaped debates about numbers of women and children exported from the continent,” they “could not function as a window into the everyday lives of Africans, who remained faceless, anonymous, disembodied” [Johnson 2018, 64]. The database reifies the logics of the documents used to construct it — slave ship manifests, colonial censuses, and other European bureaucratic documents — which were the very mechanisms that permitted the slave trade and, as such, reflect colonizer attitudes and values antithetical to those of the researchers looking to use it as a scholarly resource.

46

[Alexander 2020] encountered a similar mismatch of epistemic priorities in the context of her work on the digital archiving of the Federal Writers’ Project Slave Narratives. She highlights how choosing to tag the interview files exclusively by the state in which they were conducted and the surname of the interviewee serves to situate them on a macro-level within the Library of Congress collections but does not at all facilitate more micro-level explorations. These recitations, which give voice to former enslaved persons, have the potential to act as a counterpoint to the coloniser-generated data described by Johnson. However, the fact that even non-unique details of their lives (e.g. family separation, emancipation) and especially the violence committed against them (e.g. corporal punishment, sexual assault) remain untagged severely limits the scope and possibility of digital discoverability and exploitation along these axes. Alexander suggests that the archiving choices are a product of computational conventions, which favour normalisation for abstraction, and provides eight recommendations for the creation of a digital archiving system which “consciously choos[es] to center the humanity of [the narratives]” [Alexander 2020, 58]; these include in-depth content tagging anchored in prosopography and the deliberate inclusion of contextual information — additional documents and scholarship relating to cultural and historical context but also information relating to how the data was created (the conduct of the interviews and the project itself). They follow the ethos of “contextual digitisation” which aims to “provide an apparatus by which users can understand the provenance, history, curation and technologies that have served to produce the new intellectual output” [Gooding 2023, 597]. Alexander’s recommendations represent a kind of useful mid-way point for DH work.

47

The techniques employed by the Library of Congress and equally by the FTE team are, in part, a means of adapting to scale; they are standard convention in many areas of computing precisely because the power of computation lies in its facility with scale. While a thought-provoking use of affect, the deep context and relationality underpinning the CMSB approach would be infeasible and, arguably, less impactful, using comparable number of records. Alexander’s recommendations strike a balance; they emphasise the inclusion of context and specificity, while also accounting for the higher-level more abstract structures — such as tagging system — whose analytical affordances are the reason that

48

humanists were drawn to the use of computation in the first place.

Reshaping Convention

An essential part of the work of DH, as in the case of Alexander's speculative guidelines, is to (re)imagine and adapt conceptions and practices born of one of its constituent disciplines' epistemic traditions to align with the other. Drucker takes this back to the most base level with her reframing of 'data' to 'capta'. "Data," with its Latin root of 'given,' implies passivity; data is a "natural representation of pre-existing fact" "Capta," signifying 'taken,' conveys the "constructedness" of a representation, "acknowledg[ing] the situated, partial, and constitutive character of knowledge production" [Drucker 2011].

49

The notion of "capta" is at the heart of sociologist Maggie Walter's work on data-driven government policy and Australian indigenous communities. Walter describes the data about indigenous communities resulting from the prevailing paradigm as BADDR — Blaming, Aggregate, Decontextualised, Deficit, and Restrictive [Walter 2018]. A theoretically neutral and realistically dominant standpoint engenders "othering," measuring in terms of deviation from a norm. Aggregation creates and reifies an imaginary homogeneity. Removing context devalues lived experience. Metrics designed in external contexts which do not align with a population's worldviews, values, and priorities will inevitably find them wanting. "[S]tatistics that perpetually describe 'the problem' are themselves part of the problem, reproducing on a seemingly endless repeat the trope of Indigenous deficit," writes Walter; these patterns are elevated and perpetuated by the insistence that "data are just data [...] immutable and indisputable." As a counterpoint to typical quantitative models, Walter introduces the idea of *nayri kati* — "good numbers," in the palawa Tasmanian Aboriginal language [Walter and Andersen 2013]. Broadly, *nayri kati* refer to metrics emerging from standpoint-informed methodologies. As one example, she contrasts standard public health data with that collected by the Albuquerque Area Southwest Tribal Epidemiology Centre; the latter's metrics account for the way in which

50

for Indigenous peoples, health is not just about maintaining physical health, such as through exercise or taking medications to prevent and manage diseases, it is connected to their ways of being and doing that are unique to their identity and understanding of the world [...] for example, a way of life that supports wellness and spirituality and ceremonies, traditional medicine, heritage languages, family and community connectedness, agricultural way of life, and physical wellness

[Walter and Suina 2019, 238].

She notes that the introduction of new metrics is not the most important aspect of this kind of intervention, emphasising rather the importance of making visible the contrast and shift in methodology. Walter associates western quantitative methodologies, especially scientific ones, with a lack of "epistemological reflexivity" stemming from the choice not to engage with standpoint; "they have been the norm, and 'normal' gets conflated with 'natural.' These methodologies contain within them the basic tenets of societally dominant ways of knowing, but like a fish in water, most researchers don't subject these ways of knowing to scrutiny" [Walter and Andersen 2013, 43]. Computation does rely on the quantitative, which may appear less amenable to situatedness than the qualitative, but Walter demonstrates how quantitative work can be made compatible with positionality.

51

Drucker herself illustrates the data-capta shift — focusing less on collective values and more on questions of perspective and embodied experience — through a series of experimental 'capta' visualisations. In one example she augments a traditional bar chart representing the publication of nineteenth century novels with layered horizontal artifacts so as to communicate the context of their creation — the whole timeline of steps (e.g. writing, acquisition, editing, pre-press work) leading up to the event of publication that had previously been the sole focus of the chart. In another series of visualisations, she explores different ways of warping linear time scales to reflect how time is perceived by the subject whose data is being represented. The latter plays into some of the principles of the emergent movement of affective data visualisation — or visceralisation — which attempts to disrupt and extend the conventions of scientific computing by challenging the binary and hierarchy between reason and emotion, positioning the latter as a component of, rather than a risk to, valid knowledge production. Lauren Klein and Catherine D'Ignazio give the example of two visualisations of United States gun-related violence statistics. The more 'typical' example is a bar chart of the

52

aggregated number of active shooter events per year [D'Ignazio and Klein 2020]. The affective approach is an animated chart where arcs are drawn along a time-marked x-axis, each corresponding to a single gun-related fatality. Each arc is divided into two colour coded segments: An initial orange section encodes the age to which a victim of gun-violence lived, and the rest of the arc, drawn in white, stretches to the point of their projected life expectancy, according to demographic-specific health data. As each line is drawn, a name and age are shown, and a global counter displays the increasing total number of “stolen years.” Like FTE, this visualisation manufactures attributes not in the original dataset (projected life expectancies); however, it does so in a much more contextual way — using individualised demographic data — and also makes it clear that this data is not sure. Further, it deliberately embraces what FTE sought to avoid through the clustering of migrants into groups of 25: the illegible 'hairball' created by including a visual representation of each person. Visualisation is argument; rather than using detachment and aggregation in the aim of persuading through logos, this affective visualisation leverages pathos in the form of the emotional potential of overwhelming individuality.

Humanistically Data-Driven?

The modern data economy, asserts Kate Crawford, is premised on “the dual operations of abstraction and extraction” [Crawford 2021, 217]. Data itself is an extraction of the perceived most salient parts from a phenomenon, and the aim of this processing is to create representations which all conform to an abstract model. The labour of the creators is similarly extracted, while their presence is abstracted away. As an example, she points to data training sets which are at the foundation of our increasingly AI-driven world. Echoing aspects of the BADDR data issues Walter identified, Crawford highlights the fact that one of the most well established datasets for facial recognition (MEDS) is composed of prison mugshots collected over a subject’s multiple arrests, used without their consent and often featuring severe injuries, and that a similarly well-regarded dataset for emotion classification (CK+) is principally composed of images of feigned emotion. Yet, when ingested into a dataset, “the personal, social, and political meanings” of images representing individual people and the power exerted by societal structures “are all imagined to be neutralized” [Crawford 2021, 93]. When integrated into a dataset, argues Crawford, there is a shift “from image to infrastructure, where the meaning and care that might be given to an individual person, or the context behind a scene is presumed to be erased at the moment it becomes part of an aggregate mass that will drive a broader system” [Crawford 2021, 93]

53

Mechanisms such as Ghebru et al’s “Datasheets for Datasets” [Ghebru et al. 2021] make preliminary strides toward attending to the issues raised by Crawford. This exercise, which emerged out of the critical computing movement, asks developers to reflect on questions regarding a dataset’s:

54

1. Motivation: Who funded and undertook the creation of the dataset, and what was their intended purpose?
2. Composition and Processing: What does each instance represent, how is it labelled or broken down, what is the coverage of the dataset as a whole, and where might there be gaps?
3. Uses, Maintenance, and Distribution: How will the data be used and how could it be used poorly; how will data sharing and updating be controlled; and who will be responsible for these actions?

Mitchell et. al’s “Model Cards” look to fulfil a similar purpose with respect to the models built from and for these datasets. In recognition of the fact that these models are often used in “high-impact tasks in areas such as law enforcement, medicine, education, and employment” [Mitchell et al. 2019, 220], the authors propose that models be published with an accompanying model card, which details its intended use (and contexts where it should not be used); provide benchmarked evaluations across conditions involving different groups (cultural, demographic, phenotypic) of subjects, including intersectional groups; describe the agents and procedures involved in its creation; and provide ethical considerations.

55

Both datasheets and model cards are working toward transparency in the aim of democratising of Machine Learning (ML) and, specifically, Artificial Intelligence (AI) technologies. However, whatever checks and balances these provide, there remains the challenge in DH that these technologies embody epistemic principles which, in many cases, run counter to those of humanistic inquiry; they are, as Crawford describes, frequently fed decontextualised data and are operating under the premise that consensus is success, leaving no room for positionality.

56

For Lucy Havens, there needs to be “a shift away from removing social biases and towards acknowledging them,” so as

57

to “creat[e] data and models that surface the uncertainty and multiplicity characteristic of human societies” [Havens 2024, iv, emphasis mine]. As a step toward the creation of a “Bias-Aware” Methodology for the use of AI/ML in DH, she highlights four major priorities of AI/ML and describes corresponding humanistic recalibrations — all of which relate back to questions of context and positionality. The first priority Havens highlights is Quantity. AI/ML practice and, in particular, performance evaluation has a noted focus on the data set size and volume of model parameters, which “communicates an underlying assumption that bigger is better” [Havens 2024, 24]. The corresponding recalibration is Quality; Havens suggests using qualitative methods to critically analyse dataset and model composition. To counter the priority of Efficiency, which often manifests as the creation of “good-enough” metrics which are over-simplified, vaguely defined, or “reduce subjective tasks to a single interpretation” [Havens 2024, 25], she offers Accuracy — not in the sense of the success of a classification model on a dataset contrived to be amenable to its logics, but regarding the creation of metrics which are meaningful in real world contexts. These importantly can not remain fixed but must be continually adapted as society evolves. The third priority is Convenience, referring to the way in which AI/ML culture will reach for easily-obtained data, often created through online crowd work or — as in the case of MEDS — by exploiting marginalised populations. The corresponding recalibration is Representativeness — a focus on the curation of datasets which are representative of the population to which the models will be applied, so as to avoid the widespread existing issue of models which demonstrate obvious bias and prejudice. The lack of consideration of representativeness is in part a consequence of the final priority: Universal Thinking. This is a product of a valuation of the generalisable and neutral. Situated Thinking is the evident humanistic counter.

Maxime Bernier insisted that there is no such thing as “Indigenous math”; from this we might infer that he would similarly not believe in “Indigenous computing.” That, however, is precisely what nehiyaw-Métis artist and programmer Jon Corbett has worked to develop. Corbett describes how, in his journey learning computing, he continually fought to avoid “conforming to the design and language of the computer,” instead “breaking and rewriting as many computational rules as possible” (e.g. altering to be less efficient, but more reflective of the processes he was modelling), to instead have “the computer understand how I wanted to [process/create information]” [Corbett 2023, 91–92] — in a way that was reflective of an Indigenous worldview. This experience culminated in his creation of Cree#, a programming language that is Indigenous in its structure, symbology, syntax, and purpose. It is built around the metaphor of “code as story” and written entirely in Aboriginal Syllabics. The declaration of a function translates to this is a collected story called [function name], and the keywords for looping and conditionals are drawn for natural metaphors. The act of “smudging,” which is an indigenous cleansing ritual, has been computationally transcribed as a command which prepares the program for running by clearing the screen and any cached memory. An intended possible output of a Cree# program is a rendering of beaded artwork that is a traditional means of encoding and transmitting stories.

Posing a challenge of sorts to the DH community, Drucker questions whether it is possible to undertake meaningful humanistic scholarship using tools that “preclude humanistic methods from their operations because of the very assumptions on which they are designed: that objects of knowledge can be understood as self-identical, self-evident, ahistorical, and autonomous” [Drucker 2012, 86]. No matter where one stands on this question, the reality is that, despite these mismatches — which are specifically a reflection of differing epistemological stances relating to context and positionality — digital humanists are actively choosing to use computational tools and methods because they are interested in the analytic possibilities and perspectives these represent. Doing this “humanistically,” whether it is a question, as Corbett has done, of reimagining tools to align with humanistic intentions or, more simply (and accessibly), of accounting for and adapting to the epistemological underpinnings of existing ones, an awareness of this friction point is essential to the knowledge production work of DH.

Relationship to Tools

Born scarcely twelve years apart — in 1911 and 1923 respectively — Marshall McLuhan and Peter Elias both lived through the passage of computers from room-sized systems limited to small academic and governmental spheres toward much more ubiquitous and accessible technology. Being witness to this development seemingly gave them foresight to predict where computing was headed: toward independence and invisibility.

Elias, a computer scientist and information theory pioneer, believed that the ultimate goal for computing was to become

58

59

60

61

“frictionless.” He imagined a future where the machine would be able to carry out complex processes with minimal instruction and intervention, insisting that “if it continues to demand as much effort to learn how to speak to machines as it costs us to teach students a course for a couple of semesters, then we have failed” [Greenberger and others 1962, 203]. The interest, for Elias, lay not in dialoguing with the computer, but in the development of tools which could “understand” problems and solve them independently. Implicit to this proposal is the equivalence — at least perceived, if not actual — of human and machine logics.

McLuhan foresaw a similar convergence or at least confusion of logics, not as a technologist but as a media theorist. He placed the computer in a long history of technologies — media — that have altered the way in which we view and interact with the world. In a pattern that has occurred again and again, when it first emerges, a new technology is strange and new, and we are correspondingly highly aware of the way in which it is mediating our experience; as time passes, the technology and the processes surrounding its use become more formalised, normalised, then internalised such that we are less likely to notice its presence. Unlike Elias, McLuhan was wary of this eventuality. McLuhan describes media in terms of extensions and amputations of the body. A shovel, for instance, is an extension of the arm, enhancing its capability; similarly, a microscope is an extension of the eye [McLuhan 1994]. When pushed to an extreme, these extensions become replacements for specific capabilities and processes. Media amputate what they render obsolete, altering our behaviour; the telephone, for instance, extends the voice but amputates skills in penmanship and composition. He had a particular anxiety around digital media because he understood them as extensions of the nervous system and, more importantly, the brain. McLuhan worried about the “utter human docility” [McLuhan 1994, 57] that the extreme of amputating core cognition could represent.

62

In our modern context, as advancements in computing continually work to blur the boundary between user and interface intention [Cellard and Masure 2018] and McLuhan’s fears have come to bear in the extreme in the form of generative AI [Lee et al. 2025], the Digital Humanities are left to grapple with the place of the constantly elided human in computational work. What does interpretive and embodied knowledge production look like when it is, in part, the product of a machine? How do we go about analysing and critiquing its influence? The relationship between human and machine is central to the core methodological discussions of DH [Van Es, Wieringa, and Schäfer 2018].

63

Separation and Transplantation

The prevalent use of computers in scientific disciplines is consistent with the way in which they have historically used other tools: in service of mechanical objectivity. This notion of using machines to augment and automate procedures in order to circumvent the faults of the human senses which might obscure the objective truth predates computers. The introduction of photography is a seminal example, the mechanical image as amputated eye becoming “the emblem for all aspects of non-interventionist objectivity” [Cordell 2019]. The realities of the use of photography in its earliest days give the strongest insight into the motivation behind its inclusion: It “was not because the photograph was more obviously faithful to nature than handmade images — many paintings bore a closer resemblance to their subject matter than early photographs, if only because they used color — but because the camera apparently eliminated human agency” [Cordell 2019].

64

Since it extends the brain, it is the “faults” of cognition that are addressed by computation. In the early days of computing, “faults” essentially referred to procedural errors that might occur in mathematical calculation — of Bernoulli numbers, of missile trajectories, etc. What is being extended is not the reasoning behind the problem solving, but the speed, scale, and precision with which it can be done. In cases such as these, the delegation of cognition is arguably largely unproblematic; the problems are solved with strictly defined procedures that could be used by any person to arrive at an identical answer — the assumptions subsumed by the amputation are minimal and well-understood. Scientific instruments, explains Lisa Gitelman, are “matters of consensus” within a given community. When a new instrument is invented or introduced, there is a period where its proponents must “demonstrate persuasively that it does or means what they say.” After this point, as its use becomes more common, “its general acceptance soon helps make it a transparent fact of scientific practice” [Gitelman 2008]. The tool and its surrounding protocols are rendered self-evident such that scientists “look through the instrument the way one looks through a telescope, without getting caught up in battles already won over whether and how it does the job” [Gitelman 2008, 5].

65

In computing, these battles were won at a time where the capabilities of the technology were more limited and the user base more constrained, but the established transparency means that the impact of the mediation continues to remain unquestioned. However, using a limited set of machine instructions to automate the calculation of mathematical tables and using a high-level language to sketch a model to suggest insurance premiums or parole sentences, for instance, represent two very different models of working with machines. In the latter case, what is being extended and, eventually, amputated, is no longer well defined and delimited procedures, but reasoning and intuition. This usage is understood in the continuity of mechanical objectivity — circumventing cognitive bias as opposed to simple faults of the senses — and is attributed the same tacit trust despite not having been subjected to the same requirements of general agreement and understandability that underpinned the invisibilisation of the former.

66

Media are culturally defined. Our interactions with technology drive its evolution which in turn influences future interaction. This circular influence means that media come to reflect the beliefs and values of the culture from which they emerge [Latour 1990], which is what allows them to become invisible. In light of this, the invisibility of a medium should theoretically be disrupted when transposed to a different context. However, the mechanistic nature of early DH allowed norms to be ported without interrogation or adjustment. In the now infamous case of Father Busa's Index Thomisticus [Nyhan and Passarotti 2019], for instance, the problem of creating a large concordance resembles the problems of early computing in that the procedure is well defined and associated assumptions minimal, explainable, and generally well understood by the relevant community. This does not hold in the case of the interpretive and ambiguous problems of more modern DH, where the more complex extension of reasoning is at play [Presner 2010]. Here, scientific epistemic notions embedded in the technology stand to alter and shape humanistic inquiry in a way that might contradict the values of the latter tradition — an influence that often passes unnoticed because of the pre-established invisibility of the medium.

67

The choice of the fruit fly as an “instrument” has played a major role in defining genetics as a discipline, argues Gitelman. If a different organism were used “genetics itself would be a substantively different field” — insofar as, for instance, the pacing of research and framing of questions are concerned. She recognizes that “[t]he particular authority of science makes this an uncomfortable claim,” but one that in that in other contexts, however, would seem to be evident:

68

[J]ust as it makes no sense to appreciate an artwork without attending to its medium (painted in watercolors or oils? sculpted in granite or Styrofoam?) it makes no sense to think about “content” without attending to the medium that both communicates that content and represents or helps to set the limits of what that content can consist of. Even when the content in question is what has for the last century or so been termed “information,” it cannot be considered “free of” or apart from the media that help to define it. [Gitelman 2008, 7]

Many humanistic disciplines have a history of considering the impact of mediation — any scholar of early modern literature, for instance, would argue the necessity of considering the nature of the novel as a medium when analysing literary trends. In DH, however, there is a certain blind spot when it comes to computer mediation.

69

Media scholar Anthony Masure insists that DH digital tools rarely reflect the epistemological tensions between computing and the humanities because they “hide the primary characteristic of computers: their facility for processing symbols” [Masure 2018]; layers of abstraction obscure the base mechanisms to create the ‘natural’ interactions of Elias’ imagination. In order to be able to actually see and analyse the impact of the computational medium, it is necessary to understand its specificity. Media specificity is defined by essentially what Masure is pointing to: how the medium inscribes information. Computing is somewhat unique in that it inscribes both procedure and artefact. ‘Doing’ computation is mapping human understandings of structure and process to a specific format that is manipulable by a machine. The mediation is the adjustment made to the original conception to conform to a paradigm that is defined and bounded by the way in which information is represented. “We cannot continue to ignore how machine algorithms ‘read’ literary information,” insist literary DH scholars Long and So, “while blindly relying on them to enhance our own reading and interpretative practices.” [Long and So 2016, 236–37]. Understanding what machine ‘reading’ represents requires understanding its “underlying computability” — which is to say the structured reasoning and symbolic encoding that

70

A Complex Relationship Dynamic

The implications of underlying computability are precisely the subject of D. Sculley and Bradley Pasanek's treatise on the use of Machine Learning in DH. Their core argument is that computational learnability theory, which guarantees the learnability of arbitrary concepts, makes assumptions that are at odds with those of humanistic knowledge production. For instance, it first assumes that "data is produced by some process with constant probabilistic qualities" [Sculley and Pasanek 2008, 411] — which crucially means it does not change over time — and that examples can be drawn from it identically and independently. This assumption is what enables formal generalisation and creates the conditions for the proof that a hypothesis learned on early data can be used to make predictions about later data. It is, however, an assumption that rarely holds with respect to humanities data. Historians and scholars of culture, for instance, are frequently interested in change over time; they fully expect the dataset they are working to exhibit some non-constant behaviour. The required existence of a fixed distribution also implies the existence of some level of fixed truth which, as seen, is a troubling notion in a humanistic context. The further learnability assumptions of a fixed size hypothesis space and representative data similarly butt up against the notions of situated knowledge and data constructedness. Consequently, Sculley and Pasanek assert that ML tools and techniques, which are built on the foundation of this theory of learnability, can not be used to "prove" claims in the humanities in the same way they do in the natural sciences, where "the validity of the evidence lies inherent in the technology" [Sculley and Pasanek 2008, 421]. They argue that ML models in DH are only made relevant through critical interpretation — not only on the level of the content, but also of the experimental method. The latter demands engaging with the ways in which the specific demands of the computational tools influenced the procedure and results [Underwood 2014], something that is notably particularly difficult in the realm of ML/AI, where the explainability of deep learning models evolved through training remains an open challenge [Minh et al. 2022].

71

In his work on sentiment analysis of the poetry of the Black Arts Movement (BAM), Ethan Reed offers at once a critique and defence of the use of computational methods in humanities contexts — both of which rely on "look[ing] carefully at the nitty-gritty details [of] what these tools are, how they work, as well as the biases and assumptions they bring to any analyses they might perform" [Reed 2019, 111]. It is Reed's awareness of the demands and limitations of the computational medium which led him to ascertain from the start that his corpus was rife with what he terms "computational resistance." The "formal features" that create meaning in the poetic form (e.g. use of vernacular, irregular white space, deliberate word misspellings or transformations) would "defy, bewilder, or otherwise frustrate computational reading practices at its various stages" because of their connotations which do not align with those of dominant "canonical" forms; however, facilitating analysis by manipulating these texts such that they are "more amenable to computational forms of reading [...] often involves manipulating the voice of the poet by making changes to the original text" [Reed 2019, 99–100]. Reed insists that computational techniques can nonetheless be useful in the study of these poetic texts if the analysis places results in dialogue with these challenges. Using PatternAnalyzer, a pattern-based classifier, Reed was surprised to find that it classified what he perceived to be one of the most positive poems in the corpus — 'Come Sing a Song' — as the single most negative text. He discovered that this came down to the model routinely classifying the word "Black" as negative, despite the racial sense being designated as neutral in the model lexicon. Reed hypothesised — calculations appearing to confirm — that, failing to resolve the term's polysemy, the model simply averaged the valence values of the three possible senses of black', resulting in an overall negative. This course of action, though guessable because of the rule-based nature of the model, was completely black-boxed and not even featured in documentation. The classifier similarly struggled with terms whose sentiment was made opposite in the context of the poem (e.g. "bad" in the line "sing a 'bad' freedom song"). Studying the classifier's struggles, Reed insists, contributed to a deeper consideration of the question of audience and, correspondingly, tone, which can be conceptualised in terms of an author's perception of their audience. A positive reading of "Come Sing a Song" would have relied on a specific interpretation of terms that is born of belonging to the community for which the poem was intended. One of the most terminologically violent poems — 'The True Import of the Present Dialogue, Black vs. Negro' — was determined by PatternAnalyzer to be relatively neutral; this occurred through a combination of the fact that the classifier simply marked vernacular terms for which it had no lexicon entry as neutral and the way in which 'negative'

72

terms were often accompanied by 'positive' ones (e.g. "shoot straight"), the scores essentially cancelling out. The latter phenomenon, though born of misunderstanding (in human terms), is an interesting computational reflection of the concept of 'measured protest', which scholars had previously identified in BAM poems.

A different classifier, Vader, which was trained in a way that sees it perform better on vernacular text and demonstrate more sensitivity to sentiment intensity, while coming closer to the mark with "Come Sing a Song," performed a surface-level reading of "The True Import," marking it as extremely negative. It failed to capture the "positive negativity" — strategically employed anger — that Reed and critics see in the poem by virtue of knowledge of "vision of a broader social purpose" [Reed 2019, 126]. That said, Reed asserts that "the light shed by VADER's peculiar evaluative logic as to why this poem seems so negative allowed me to reflect more seriously on my own understanding of the poem, as well as that of the many other critics who have noted its particular affective charge" [Reed 2019, 131]. The value Reed finds in digital tools is in the way in which they "[do] *not* transform the text the way that a human reader does" [Reed 2019, 96]; the challenge lies in not conflating the two when the computational mechanics are often black-boxed away for "ease of use," which, as described by Stephen Ramsay and Geoffrey Rockwell "install[s] the user at a level of abstraction far above whatever theoretical claims might lie beneath." [Ramsay and Rockwell 2012, 80]

While Reed's explorations focused on the use of tools in DH, there is also the question of the construction of computational artefacts. The act of 'building' has been central to DH since the field's inception; however, there is complex debate over whether it (in isolation) 'counts' as scholarship under the epistemologies of both constituent disciplines. [Ramsay and Rockwell 2012] offer three interrelated framings of DH tools and three corresponding arguments for "building" as a valid form of humanistic, and therefore DH, scholarship. They first argue that DH tools can be considered theories (in the humanistic more so than scientific sense — something that explains rather than predicts) because they embody a specific discourse. Digital artifacts have a procedural rhetoric; the information they choose to display and the way they afford specific user interactions is an argument. Because of this inherent orientation, Ramsey and Rockwell assert that computational tools can also act as theoretical lenses, in that they help us see things differently; as "hermeneutical instruments" they are undoubtedly valid under a knowledge-culture built around interpretation. Finally, they propose the notion of the digital as a theoretical model. Computing is a modelling discipline because "any computing system is an explicit, delimited conception of the world or 'model' of it" [McCarty 2005, 21]. Ramsey and Rockwell draw a parallel between the coding used to create such a model and the writing used to express specific conceptions of the world in more traditional humanistic scholarship. They argue that the understanding that comes of "the manipulation of features, objects, and states of interest" in programming is equivalent to that born of having to put things into writing. They notably highlight that the words (or code) by themselves are not the scholarship, but the modelling process behind them, which is highly dependent on the medium of expression.

What unites the three framings Ramsey & Rockwell propose — and indeed, Sculley & Pasanek and Reed's work — is that they are all a question of thinking with computers. Each relationship described relies on an understanding of computational mediation, on seeing the ways in which the machine's inscription and processing of information transform phenomena and treating it as another voice in an ongoing discourse, to be accounted for in analysis and used as part of idea generation. There is a fine line between this thinking with and the alternative of thinking like computers — which is to say deliberately, if often unconsciously, trying to structure our conceptions to be amenable to computation [Doueiri 2011]. David Berry argues that we are witnessing an increasing operationalisation or mechanisation of thought; when we cede intention to computational suggestions we are confounding human reasoning with "algorithmic conceptualizations" of their ideas, the latter of which are increasingly constructed from information collected from millions of users and are therefore inescapably anchored in hegemonic frameworks such as capitalism and white supremacy. As Berry describes,

[...] Rather than acting as bicycles for the mind, ^[7] these technologies replace certain cognitive functions of the mind. Being owned and controlled by corporate organisations they tend to weaken explanatory and critical thinking and instead nudge and influence human behaviour in directions that are profitable

[Berry 2023]. The sort of replacement Berry outlines could more moderately be understood as the normalisation and

73

74

75

invisibilisation that, as seen, is part of the life cycle of any representational medium. In this light, we might see the core challenge of DH's relationship to machines as walking the line between like and with — analysing the nature and impact of mediation so as to be able to decide when and to what extent delegation is acceptable, and when and how to undertake generative co-creation.

Edsger Dijkstra is famously known to have argued that “computer science is no more about computers than astronomy is about telescopes.”^[8] In this, he implies that there can exist a complete separation between logical process and the machine being used to express it. Dijkstra was strongly against what he saw as the de-mathematisation of computing, which opened it up to broader use. He spoke out, for instance, against the development of early high-level programming language, COBOL, insisting that it “sought to replace intellectual discipline by management discipline” [Dijkstra 1984]. Perhaps in Dijkstra's idealised framing, it might have been feasible to separate human logic from machine expression; if the applications remain limited and carried out by experts, there can be a clear and well-understood mapping between human and machine problem representations. The impact of mediation is minimal because the problem domain and the context from which the medium emerged are one and the same. As — contrary to Dijkstra's wishes — the reach of computing has broadened, the mapping between problem and representation has become more complex and less evident, but the assumptions baked into the acceptance and invisibilisation of the medium remain those established in that limited context and expert community. The introduction of the digital to humanistic scholarship has provoked a major change in how information can be viewed and processed, and birthed new methodologies which, for all they may mimic the old ones, inescapably operate differently. “Computation forces us to rethink our current disciplinary practices in the humanities from the ground up,” asserts Andrew Piper — “What counts as evidence? What is the relationship between theory and practice? How do we account for the technological mediations of our critique?” [Piper 2016, 2]. These questions all come back to the influence of the way the machine “thinks” on humanistic work. The machine, of course, does not “think”; the way it structures and processes information is a product of a specific epistemic perspective. Now that we are pointing it at so much more than the stars, it is essential to consider the telescope.

76

Hermeneutic Attention

When working with computer scientists on digital visualisation tools, art historians on Project Cornelia were left feeling that “the tech kids are running too fast” [Lamqaddam et al. 2018]. Houda Lamqaddam — one of the “tech kids” — explains that the art historians found speed of development and release was at odds with the “contemplative nature of their research.” Artworks, through the methodological lens of art historians, are “elements that trigger reflection rather than contain answers,” an outlook that contrasts digital interfaces which “present the user with synthesized views of data and answers to research questions.” Contributions in art history, due to the inherent subjectivity of the field, “often consist of making a claim denoting a unique perspective” — one that is based in “in-depth artwork analysis” that “put[s] into evidence previously unexplored nuances.” Though, as seen, it is arguably embodied in the tools, “the use of aggregates, statistics and data-crunching algorithms by computer scientists is viewed as an approach where the reflection and the nuance disappear.” For art historians, Lamqaddam asserts, the intervention of the digital “threatens their operational paradigm.” In investigating their reticence to engage with the digital, she focuses on the perceived speed difference, but this mismatch could be viewed as an element of a broader question of divergent approaches to interpretation.

77

Interpretation is how information becomes knowledge; this sense-making is inherently a part of any discipline. There is a common process of pattern matching and formalisation — a collection of examples will be interpreted to propose a more or less abstract theory demonstrating or explaining a phenomenon [Dixon 2012]. Where domains differ is in the emphasis they place on aspects of this process. The heart of the scholarship for the art historians is the process by which they arrive at their conclusions. They were discomforted by the work of the developers, as the interpretation that went into the choice of encoding entities and relationships to illustrate specific theories of connection was, if not hidden, certainly not easily legible to the historians presented with the finished interface — an object which, for the developers, was understood to be the primary contribution. Underwood proposes that “humanists tend to think of computer science as an instrumental rather than philosophical discourse” [Underwood 2014]. This is untrue, he argues, highlighting how, for instance, the Bayesian statistics driving many machine learning algorithms encode a specific way of reasoning in

78

their acknowledgement that we “approach every question with some previous assumptions (called ‘prior probabilities’), as well as particular kinds of uncertainty” [Underwood 2014, 66]. This interpretive cycle of existing assumptions iteratively changing through the introduction of new evidence, Underwood insists, should be familiar to humanists, and recognisably valid under their paradigm of knowledge creation; however, it is seldom visible or legible. Digital humanists are caught between — and left to reconcile — implicit and explicit hermeneutic traditions.

Divergent Traditions

“There is no self-critical hermeneutic tradition in the sciences comparable to that taken for granted by other scholars,” argues [Markus 1987], as described by [Franklin 1995, 165] — the “other” gesturing to the humanities and social sciences. Markus is not making a critique, but an observation; he posits that it is a case of scientific domains constructing themselves such that they do not need a strong hermeneutic tradition. Markus points to scientific knowledge diffusion as a prime example. In interviews with biologists, [Cetina 1999] noted that many described an aspect of intuition as essential to correctly carrying out experimental procedures, but equally that this was not in any way integrated into methodology or reporting. Since the late 19th century, scientific publications have been characterised by their passivity; the figure of the author — the knowledge interpreter — is effaced. This is not intended to imply that no interpretation has taken place but rather that the reader’s interpretation, specifically of the procedures and logic from which a conclusion was derived, will never differ from that of the author. “The ‘ideology’ [...] of the natural sciences,” Markus argues “regards any acceptable scientific text as totally self-sufficient as to its meaning (and therefore as unambiguously clear to any reader with adequate competence)” [Markus 1987, 9]. The “adequate competence” to which he is referring is knowledge of rigidly defined and strictly encoded axioms and procedures. Curtailing alternative interpretations is one of the functions of the “language” of the natural sciences^[9], which is defined by these “sign-systems.” By virtue of their use, it is understood that the writer and the reader have the same conception of what is being described. When the writer and the reader are interchangeable, there is nothing to be interpreted.

79

Elements of the sign-system are viewed as beyond interrogation — a framing enabled by the decontextualized and ahistorical way in which they are taught. Markus notes that, whereas humanities students are expected to interrogate Aristotle’s works in light of the context of their creation and with respect to different readings over time, the version of Newton’s laws written in textbooks is “something Newton should have written had he used modern mathematical notations, contemporary physical concepts, etc.” For all that, as Markus implies, the meaning of these and other theories and expressions changes as science progresses, they are “posited as designating well-identifiable, singular historical phenomena and events” [Markus 1987, 32–33]. The interpretation behind building blocks of scientific knowledge production is subsumed by definitive “signs.” Mathematics education researcher Stephen Brown relates the anecdote of a physics professor who insisted that “the only things that count in proving anything are axioms, definitions, rules of logic and previously established theorems” [Brown 1996, 1292]; having forgotten one of the axioms of vector space, the professor, mumbling under his breath, turned his back to the class to derive it on the blackboard, taking care to hide his work and hastily erasing it before returning to his lecture in attempt to remain consistent with his earlier assertion. Brown argues that this handing down of detached axioms in “an authoritarian and non-controversial way” leads us to “believe that they were created in a way that involved no labor pains and in particular no negotiation between what was considered worth proving and what was taken for granted” [Brown 1996, 1292]. Interpretation is considered a key step of theoretical mathematical modelling, however, the process of modelling is rarely taught. Even in the rare cases that it is, student assessment has a tendency to focus on the overall outcome rather than considering the learner’s progress through modelling steps or sub-processes [Berry 2002]. Hermeneutics in the sciences, per Markus, is not synonymous with learning — as might be said in the case of the humanities — but prerequisite to it. Interpretation is a “tacit dimension” [Markus 1987, 7].

80

Interpretation is anything but tacit in the humanities. As previously discussed, the humanities lack a conception of absolute truth. In the absence of a definitive answer, there is nothing left but interpretation. In his book on literary criticism and computational text analysis, *Reading Machines*, Ramsay writes:

81

However far ranging a scientific debate might be, however varied the interpretations being offered, the assumption remains that there is a singular answer (or a singular set of answers) to the question at hand.

Literary criticism has no such assumption. In the humanities the fecundity of any particular discussion is often judged precisely by the degree to which it offers ramified solutions to the problem at hand. We are not trying to solve Woolf. We are trying to ensure that discussion of *The Waves* continues. [Ramsay 2011, 15]

Ramsay's assertion highlights a key distinction between the two hermeneutic traditions. The humanities tend toward the dialogic, looking for the interplay of multiple interpretations, whereas as the sciences are aiming for consensus and, correspondingly, are more monologic in their practices. Computer science is grounded in a problem solving paradigm and the humanities in a problematising one [Van Zundert 2015].

Because of this difference in aim, technology use as it has been implemented in one discipline may not be identically useful in the other. This mismatch was at the forefront of the minds behind the Humanities Networked Infrastructure (HuNI) project [Burrows 2013]. HuNI, which looks to make Australian cultural data from myriad institutions centrally accessible and interrogable, was funded as part of a larger network of "digital laboratories" to support data-centred research work, the rest of which were anchored in scientific contexts. Correspondingly, the proposed parameters very much followed in line with a "big science" model of data processes — aiming to simplify and streamline established data workflows. The team behind HuNI did not see these priorities as suitable for humanities research needs and set out to create a modified paradigm deliberately constructed to "enabl[e] researchers to express, share and discuss their differing interpretations of the data" [Burrows and Verhoeven 2016, 119]. HuNI is structured around graphs generated by user-created connections between cultural data entities (e.g. works, places, people) imported from different organisations. All connections are public by default, are not required to fit into any pre-defined ontology, and can be annotated with detailed meaning such that "different perspectives between (and within) disciplines are preserved and foregrounded, instead of being hidden behind a normalized, authoritative framework." [Burrows and Verhoeven 2016, 119]. Its creators note that HuNI is specifically designed to promote the kind of "serendipitous encounters" that are fruitful in the humanities [Ramsay 2010], rather than looking to scaffold any sort of "optimal" trajectory or set of operations. An environment that affords serendipity is one in which there is enough structure to maintain relevance but not so much as to "suppress the unexpected or unpredicted" and where there exists multiple perspectives and the "ability to meander among these different understandings and conceptualisations of the world and knowledge" [Burrows and Verhoeven 2023, 32].

Encountering the same friction as the HUNI team, [Hinrichs, Forlini, and Moynihan 2019] describe a pivotal scene for their work on visualisation in DH where one member of an interdisciplinary project committee posed the pointed question, with respect to project outcomes: "Are we building tools or just sandcastles?" The source of the committee member's frustration, they argue, is a mismatch in how the term 'tool' is understood in science and the humanities. In the former context, it is related to 'prototype,' and understood as the stabilised and often generalised endpoint of iteration. The successive prototypes are a means to an end rather than "object[s] of inquiry" [Hinrichs, Forlini, and Moynihan 2019, i82] in their own right. Humanistic understandings tend instead to frame tools as narrative and generative — an element of a process rather than a product. DH visualisations are, correspondingly, often small scale and highly specific. Hinrichs et al. defend the metaphor of the sandcastle as it does not permit a focus product. A sandcastle is always transient, will always be taken by the tide at the end of the day. The "point" is to build and rebuild, learning and changing without the constraints of a defined endpoint or a need to reach it. The emphasis is therefore on the ludic act of "sandcastling," which they describe as a "dynamic process wherein speculation and re-interpretation advance knowledge" [Hinrichs, Forlini, and Moynihan 2019, i80]. The success of humanistic visualisations, argue Sinclair et al., should be evaluated as a function of "how well they support interpretive activity," asserting that "the humanities approach consists not of converging toward a single interpretation that cannot be challenged but rather of examining the objects of study from as many reasonable and original perspectives as possible to develop convincing interpretations" [Sinclair et al. 2013].

Developing Digital Hermeneutics

"For as long as humanities scholars have made use of digital tools," asserts Eef Masson, "they have met with critique from fellow practitioners [...] stem[ming] from the perception that the projects conducted do not do justice to the critical-

interpretive legacy of much humanities research” [Masson 2017, 31]. With its departure from traditional close reading and its use of computational tools and methods in ways which do not necessarily align with the conventions of hypothesis-driven scientific research, detractors of DH accuse it of being both methodologically bereft and interpretation-poor, and correspondingly lacking in validity under the either disciplinary epistemology [Fish 2012] [Allington, Brouillette and Golumbia 2016]. Out of this critique has emerged a nascent “digital hermeneutics,” with a two-pronged goal of making explicit the interpretation implicit to digital methods and tools, so as to situate their use in the humanistic hermeneutic tradition, and of articulating the ways in which their design and use can be adapted to make space for interpretation.

The coordinators of the University of Luxembourg’s “Digital History and Hermeneutics” Doctoral Training Unit (DTU) insist that digital hermeneutics must concern itself with how “digital tools and infrastructures are transforming historical research practices in all stages of the iterative research process” and should operate as “a critical framework for making the methodological and epistemological tensions in current history practices explicit” [Fickers, Tatarinov and Van Der Heijden 2022, 7]. They correspondingly chose to structure the DTU’s training and projects around the notion of hands-on “Thinkering”: a “heuristic mode of doing” that “combines playful tinkering with critical thinking” [Fickers, Tatarinov and Van Der Heijden 2022, 7]. Emblematic of this approach, kiara and Lumy are a paired set of the DTU’s research projects that together form a virtual research environment which scaffolds “Thinkering” [Crouy-Chanel 2022]. kiara — which acts as a backend to Lumy but can also be used independently via command line — is described by its creators as a “data orchestration engine.” It supports the creation and analysis of network graphs by breaking workflow steps that would typically be subsumed into a single script down into functional modules. These can be rearranged and chained to create any number of bespoke pipelines. A user can also create custom modules; each is specified by an expected input and output, process description, and associated process python code. Though things are compartmentalised, nothing is blackboxed. What makes kiara different from similar workflow management tools emerged from a software development context is its focus on metadata management and iterative user interaction. It keeps a tree-structured log of every user interaction, allowing the user to reflect on their process and easily trace the “lineage” of any data result. Lumy is similarly intended to “encourage users to adopt recursive, self-reflexive and reproducible work practices.” It sits on top of kiara, giving GUI access to the network manipulation functionality in one part of the interface, while also providing space in another panel for users to take notes on their practice. The aim of associating metadata snapshots with researcher reflections is to “cultivat[e] the holistic practice of interweaving methodology, data, code, and metadata with a narrative of researcher’s choices and actions” [Crouy-Chanel 2022, 37]. This deliberate pairing of the “the ability to record what the software is doing with the ability to record what the scholar is doing” [Crouy-Chanel 2022, 37] brings the interpretation underlying the use of computational tools to the forefront.

Emergent digital hermeneutics must honour and encompass established disciplinary hermeneutic approaches. A significant proportion of DH work is connected to literary studies; however, the kind of abstract data that is generated from texts using “statistical and computational methods” is not “the kind of data that literary scholars practicing an [sic] hermeneutic approach are concerned with” [Kleymann and Stange 2021]. Literary hermeneutics is an “approach to produce meaning through an iterative, non-deterministic, and subjective procedure,” which relies on three premises: a distinction between the intention of the author and the text, in the acknowledgement of meaning as something that is not inherent but constructed; a circular process of building understanding, which holistically considers both the whole and its parts; and a dependency between text and recipient, which implies a contextual and subjective reasoning process. Though they acknowledge that some digital tools successfully replicate the conditions of analogue hermeneutic processes, Rabea Kleymann and Jan-Erik Stange argue that these essential premises are largely disregarded by tools more anchored in digital methods, especially in the realm of data visualisation. In response, drawing heavily on the work of [Meister, Drucker and Rockwell 2017] and [Drucker 2018], they propose four “postulates” for hermeneutic data visualisation which attend to the premises. The first, Two Way Screen, refers to the quality of being manipulable rather than limited to rendering. A hermeneutic visualisation should not only facilitate interpretive activities, such as annotation and information structuring, but display them. This bidirectional interaction encourages the circular and holistic approach. The second postulate, Quality, asks that data be shown as constructed and subjective, as understood from the text-recipient relationship premise. This does not necessarily require novel conventions, but might simply mean using standard visual variables to represent “interpretative dimensions like salience or relatedness.” Parallax, the third

postulate, emphasises the importance of polyvalence. Access to multiple views of an object of inquiry highlights its ambiguities, and, correspondingly, interpretive possibilities. The aim of Parallax is to deliberately engender uncertainty and doubt, which stands in stark contrast to the Tuftean conception of the purpose of visualisation as being to reveal the truth of data [Tufte and Graves-Morris 1983] — the not a — that dominates CS visualisation [Munzner 2014]. The final postulate, Discourse, is focused on the back and forth between visualisation and interpreter in the construction of argument and is particularly interested in forefronting the influence of the digital representation on the way in which the argument is constructed. In practice, this may look like the kind of visualisation and written interpretation pairing facilitated by Luxembourg DTU's Lumy interface.

While Kleymann and Stange focus on bidirectionality and polyvalence, others insist that transparency is central to supporting interpretation [Berry 2023] [Dobson 2021]; inputs, procedures, and results must evidently be interpretable in order to be interpreted. Transparency, in this case, is not the transparency of eyeglasses, where the aim is for transparency to be synonymous with invisibility — the kind of transparency which, as mentioned, is prevalent in computational interfaces [Cellard and Masure 2018] — but the transparency of a skeleton watch, where the point is to expose the inner workings. Exposing the “guts of the process” is the central premise of CLAIRON, a natural language database interface (NLIDB) for the theatrical data of the Comédie-Française Registers Project [Sollazzo 2022]. Like other tools in this class, CLAIRON does natural language question-to-query (SQL) translation, displaying the result of the latter. However, unlike other NLIDBs, it exposes each step of the translation process to the user in an effort to get them to “engage with the human to machine translation whose complexities and necessary compromises are at the heart of enacting a more humanistic style of computing” [Sollazzo 2022, 20]. Question-query highlight correspondence shows users how each of their questions is reformulated for execution, which may not match their initial conception, encouraging critical reflection on their intention. Users can directly modify queries that are produced, which are accessible through an interactive log, or even write queries from scratch, which facilitates the circularity evoked by Kleymann and Stange and reflects the dialogic nature of humanistic inquiry. The process of entity matching and disambiguation (establishing a correspondence between question terms and database entities and selecting between entities of the same name) is explicated through a series of visualisations, including notation for when foreign keys or calculated fields are used, and a series of successive query and expression trees show how linguistic dependencies and keywords are used to produce the query (e.g. implied negation, aggregation, ordering etc.). Notably, the design of CLAIRON makes no attempt to take on implicit interpretation. If asked for the best of Molière's plays, the interface returns an error message saying that it does not understand the meaning of “best”; however, questions about the most profitable or most frequently performed plays are translated and executed without issue. This feature forces users to confront their assumptions, which draws attention to the contextual and subjective nature of interpretive knowledge construction.

88

There exists a prevalent assumption in DH that “if the computer is to be useful to the humanist, its efficacy must necessarily lie in the aptness of the scientific metaphor for humanistic study” [Ramsay 2011]. However, regardless of how closely a humanistic object of study can be made to resemble a scientific problem, the ways in which disciplines seek to interact with it remain different. The aim of digital hermeneutics, as illustrated by tools such as *kiara/Lumy* and CLAIRON, is to establish new conventions of design and use which leverage the kind of interrogations made possible by digital transformation but which also make explicit sites of interpretation born of this transformation and include mechanisms for manipulation and reflection to support interpretive activity.

89

During her time as Artist in Residence at CERN, Su Wen-Chei observed that, when faced with a complex problem, the scientists would split into two groups, each of which would individually derive an answer; if the two solutions matched, the solution was considered to be correct. When she later related her experience in a workshop, one artist-participant pointed out that, were the same procedure followed with two groups of artists, if they arrived at the same answer, the conclusion would be not that the solution was correct, but that the question was wrong [Escobar Varela 2021]. These interactions exemplify the different positions interpretation occupies in knowledge production across disciplines and the assumptions underpinning them. In the sciences, there is a notion of a single correct solution and, correspondingly, and implicit aim of consensus. The primary interest lies in arriving at a solution, at which point the process of interpretation leading to it is considered tacit. In the arts, on the other hand — which are epistemologically closer, in this respect, to

90

the humanities — the interest lies in the development and comparison of different interpretations. DH exists at the crossroads of a strong hermeneutic tradition and one which is doing its best to fade away. The explicit interpretation central to humanistic scholarship is challenged by the implicit interpretation of computational artefacts which presume to embody it. At the same time, interpretive methodologies, based as they are in intuition and perspective rather than formal shared language, are difficult to precisely specify, making their use in concert with computation challenging. In order to reconcile these traditions and undertake knowledge production that is valid under both, it is essential to attend to their respective understandings of the role of interpretation and the assumptions which permit these framings.

Conclusion

The first wave of DH introduced humanities scholars to the potential of high-level quantitative computational techniques. As they transform their objects of inquiry and methodologies to be amenable to distant rather than close reading, there are aspects which are de-emphasised — made negative — as others come to the forefront.

91

Second wave DH saw a renewed focus on the qualitative, with a particular emphasis on the experiential and affective, both of which are rooted in the notion of positionality. This reorientation toward “the Humanities’ core methodological strengths” [Presner 2010], most notably its attention to complexity, depth, and context, begs the question of how these might be preserved, integrated, and re-imagined for computational work, given its often divergent priorities and assumptions.

92

The still speculative third wave [Berry 2011] pushes for increased attention to “the organizing logic that underlies our work” [Posner 2015] and the ways in which it may stand in opposition to the values and interests of the field. This demands interrogating our interactions with machines on multiple levels of abstraction — notably considering the extent to which they are used in addition to and instead of human intervention as well as how our ideas are transformed through their use, as a function of the affordances and limitations of the computational medium.

93

Throughout it all, the validity of DH scholarship has been challenged from both sides over questions of interpretation and its role in knowledge creation.

94

Successive developments in the field have introduced new challenges and questions brought about by the integration of humanistic and computational theories and methodologies. What is clear is that DH is doing computing in a way that is different to how it is understood in CS. It is not merely a question of content but of epistemology; computing is used for theory generation and creativity more so than confirmation and generalisation, for enhancing complexity rather than reducing it, for exploring the interplay of — rather than sharply delineating between — the mechanical and the embodied and the rational and emotional.

95

DH, however, lacks a dedicated computing curriculum that attends to these differences. That said, the development of one is not necessarily of interest to the community. Given the number of topics and methodologies involved, the difficulty of defining itself has been a constant of the field; this ever-evolving nature coupled with the rapid pace of technological development “should make us wary of efforts to encapsulate Humanities Computing into a list of acceptable perspectives and methods that are supposed to have any universality or permanence” [Sinclair and Gouglas 2002, 168]. The breadth and variation of DH is its strength, and the success of “shared teaching endeavours” in DH rests upon “accommodating th[e] wonderful variation and richness and situatedness” [Gambell et al. 2021, 9] inherent to the discipline and its instructors.

96

The facets are envisaged as a curricular middle ground. They do not prescribe any specific set of technologies, topics, or competencies, but can serve as a theoretical anchor for the development of a humanistic approach to computational work — for instance as a scaffold to construction and analysis activities or a framework for curriculum design.

97

Finally, Computer Science, especially in higher education, is epistemologically closed [Turtle and Papert 1990], which has been linked to its lack of diversity and argued to be a weakness of the discipline with respect to preparing its students for the complex ways in which computing is embedded in society [Ko et al. 2020]. Though designed with DH in mind, the facets framework could potentially also be of use in CS — in particular to those working in the areas of

98

CS4All, computing for non-majors, or critically conscious computing — in helping to introduce and explore the potential of alternative ways of viewing and understanding the act of doing computing and its role in society [Connolly 2020].

Acknowledgements

Thank you to Professor Quintin Cutts for his thoughtful supervision and eagle eyes!

99

Notes

- [1] Definition from Geoffrey Rockwell, selected from a list of crowdsourced definitions collected as part of annual “Day of DH” events: https://github.com/hepplerj/whatisdigitalhumanities/blob/master/dayofquotes_full.csv
- [2] <https://github.com/MimiOnuoha/missing-datasets>
- [3] <https://mimionuoha.com/about>
- [4] ‘Ethnicity’ would be the term used today to describe many of the options, but ‘race’ is the term that was used in the census.
- [5] <https://x.com/MaximeBernier/status/1414302864672006146>
- [6] <https://x.com/mathchole/status/1414653170232532999>
- [7] A phrase initially coined by Steve Jobs, meant to describe, according to Berry, the way in which computers stood to “**augmen[t]** the cognitive capacities of the user, making them faster and more capable.”
- [8] Attribution is contested
- [9] For all computing is arguably not one of the natural sciences, it nonetheless largely mirrors their operation.

Works Cited

- Agostinho et al. 2019** Agostinho, D., D'Ignazio, C., Ring, A., Thylstrup, N.B., and Veel, K. (2019) “Uncertain archives: Approaching the unknowns, errors, and vulnerabilities of big data through cultural theories of the archive”, *Surveillance & Society* 17(3/4), 422–41. <https://doi.org/10.24908/ss.v17i3/4.12330>
- Alexander 2020** Alexander, C. (2020) “Enabling access to the Federal Writers' Project slave narratives: A case study in digital archive design”. Master's thesis, University of Alberta.
- Allington, Brouillette and Golumbia 2016** Allington, D., Brouillette, S., and Golumbia, D. (2016) “Neoliberal tools (and archives): A political history of digital humanities”, *Los Angeles Review of Books*. Available at: <https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/>
- Bardzell and Bardzell 2016** Bardzell, J., and Bardzell, S. (2016) “Humanistic HCI”, *Interactions* 23 (2), pp. 20–29. <https://doi.org/10.1145/2888576>
- Benjamin 2006** Benjamin, W. (2006) “The work of art in the age of mechanical reproduction”, *Visual Culture: Experiences in Visual Culture* 4, pp. 114–37.
- Berry 2011** Berry, D.M. (2011) “The computational turn: Thinking about the digital humanities”, *Culture Machine* 12. Available at: <https://culturemachine.net/wp-content/uploads/2019/01/10-Computational-Turn-440-893-1-PB.pdf>
- Berry 2023** Berry, D.M. (2023) “The explainability turn”, *DHQ: Digital Humanities Quarterly* 17(2). Available at: <https://dhq.digitalhumanities.org/vol/17/2/000685/000685.html>
- Berry 2002** Berry, J. (2002) “Developing mathematical modelling skills: The role of CAS”, *ZDM—Mathematics Education* 34(5), pp. 212–20. <https://doi.org/10.1007/BF02655824>
- Brendel and Bethge 2019** Brendel, W., and Bethge, M. (2019) “Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet”, *arXiv*. <https://doi.org/10.48550/arXiv.1904.00760>
- Brosens et al. 2019** Brosens, K., Aerts, J., Alen, K., Beerens, R.J., Cardoso, B., De Prekel, I., Ivanova, A., et al. (2019) “Slow digital art history in action: Project Cornelia's computational approach to seventeenth-century Flemish creative communities”, *Visual Resources* 35(1-2), pp. 105–24. <https://doi.org/10.1080/01973762.2019.1553444>
- Brown 1996** Brown, S.I. (1996) “Towards humanistic mathematics education”. In *International Handbook of Mathematics Education: Part 1*, pp. 1289–1321. Springer.
- Buolamwini and Gebru 2018** Buolamwini, J., and Gebru, T. (2018) “Gender shades: Intersectional accuracy disparities in

Burrows 2013 Burrows, T. (2013) "A data-centred 'virtual laboratory' for the humanities: Designing the Australian Humanities Networked Infrastructure (HuNI) service", *Literary and Linguistic Computing* 28(4), pp. 576–81. <https://doi.org/10.1093/lc/fqt064>

Burrows and Verhoeven 2016 Burrows, T., and Verhoeven, D. (2016) "Aggregating data for social linking in the humanities and creative arts: The 'Humanities Networked Infrastructure (HuNI)'". *Signa*, 25, pp. 109-119

Burrows and Verhoeven 2023 Burrows, T., and Verhoeven, D. (2023) "Serendipity and knowledge organisation". In *Serendipity Science: An Emerging Field and Its Methods*, 31–48. Springer.

Cellard and Masure 2018 Cellard, L., and Masure, A. (2018) "Le design de la transparence: Une rhétorique au cœur des interfaces numériques", *Multitudes* 73(4), pp. 100–111.

Cetina 1999 Cetina, K.K. (1999) *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Connolly 2020 Connolly, R. (2020) "Why computing belongs within the social sciences", *Communications of the ACM* 63(8), pp. 54–59. <https://doi.org/10.1145/3383444>

Corbett 2023 Corbett, J. (2023) "ᐱᓄᓂᓕᓴᓂᓗ ᓂᓯᓲᓪᓃᑦᐱᓐᓇᓐ ᓂᓯᓲᓪᓃᓂᓗ/Mâmitoneyihcikanihkân Nîmîkisistahikêwin Ekwa Nîmitonêyihcikan: My reflections of beading with a computer", *Transmotion* 9(1&2), pp. 86–109.

Cordell 2019 Cordell, R. (2019) "Teaching humanistic data analysis", *Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching: Proceeding of the Gale Digital Humanities Day at the British Library*.

Crawford 2021 Crawford, K. (2021) *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crouy-Chanel 2022 Crouy-Chanel, M. de (2022) "Introducing the DHARPA project: An interdisciplinary lab to enable critical DH practice", *DH Benelux Journal*, 29–41. Available at: <https://journal.dhbenelux.org/journal/issues/004/article-2-Cunningham.html>

D'Ignazio and Klein 2020 D'Ignazio, C., and Klein, L. (2020) "3. On rational, scientific, objective viewpoints from mythical, imaginary, impossible standpoints". In C. D'Ignazio, C. and L. Klein (eds.) *Data Feminism*.

DeLanda 2019 DeLanda, M. (2019) *A New Philosophy of Society*. Bloomsbury Publishing.

Dijkstra 1984 Dijkstra, E.W. (1984) "The threats to computing science".

Dixon 2012 Dixon, D. (2012) "Analysis tool or research methodology: Is there an epistemology for patterns?" In D.M. Berry (ed.) *Understanding Digital Humanities*, pp. 191–209. Springer.

Dobson 2021 Dobson, J. (2021) "Interpretable outputs: Criteria for machine learning in the humanities", *DHQ: Digital Humanities Quarterly* 15(2). Available at: <https://dhq.digitalhumanities.org/vol/15/2/000555/000555.html>

Doueïhi 2011 Doueïhi, M. (2011) "Un humanisme numérique", *Communication & Langages* 167(1), pp. 3–15. <https://doi.org/10.4074/S033615001101101X>

Drucker 2011 Drucker, J. (2011) "Humanities approaches to graphical display", *Digital Humanities Quarterly* 5(1), pp. 1–21. Available at: <https://dhq.digitalhumanities.org/vol/5/1/000091/000091.html>

Drucker 2012 Drucker, J. (2012) "Humanistic theory and digital scholarship", in M.K. Gold (ed.) *Debates in the Digital Humanities* pp. 85–95.

Drucker 2018 Drucker, J. (2018) "Non-representational approaches to modeling interpretation in a graphical environment", *Digital Scholarship in the Humanities* 33(2), pp. 248–63. <https://doi.org/10.1093/lc/fqx034>

Escobar Varela 2021 Escobar Varela, M. (2021) *Theater as data: Computational journeys into theater research*. University of Michigan Press.

Fickers, Tatarinov and Van Der Heijden 2022 Fickers, A., Tatarinov, J., and Van Der Heijden, T. (2022) "Digital history and hermeneutics—between theory and practice: An introduction", *Digital History and Hermeneutics Between Theory and Practice*. Berlin, Boston: De Gruyter, pp. 1–22.

- Fish 2012** Fish, S. (2012) "Mind your P's and B's: The digital humanities and interpretation", *New York Times* 23(1). Available at: <https://archive.nytimes.com/opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation/>
- Franklin 1995** Franklin, S. (1995) "Science as culture, cultures of science", *Annual Review of Anthropology* 24(1), pp. 163–84. Available at: <https://www.jstor.org/stable/2155934>
- Gambell et al. 2021** Gambell, S., Gooding, P., Hughes, L., Doran, M., Murphy, O., Tupman, C., Winters, J., et al. (2021) "Communicating the value and impact of digital humanities in teaching, research, and infrastructure development", *Digital Humanities Research Network*. <https://doi.org/10.31219/osf.io/z8v9c>
- Gawne 2014** Gawne, L. (2014) "Evidentiality in Lamjung Yolmo", *Journal of the Southeast Asian Linguistics Society* 7, pp. 76–96.
- Gebru et al. 2021** Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé, H., III, and Crawford, K. (2021) "Datasheets for datasets", *Communications of the ACM* 64(12), pp. 86–92. <https://doi.org/10.1145/3458723>
- Ginzburg 1979** Ginzburg, C. (1979) "Clues: Roots of a scientific paradigm", *Theory and Society* 7(3), pp. 273–88. Available at: <https://www.jstor.org/stable/656747>
- Gitelman 2008** Gitelman, L. (2008) *Always already new: Media, history, and the data of culture*. MIT Press.
- Golebiewski and boyd 2019** Golebiewski, M., and boyd, d. (2019) "Data voids: Where missing data can easily be exploited", *Data & Society Research Institute*. Available at: <https://datasociety.net/library/data-voids-where-missing-data-can-easily-be-exploited/>
- Gooding 2023** Gooding, P. (2023) "Informational abundance and material absence in the digitised early modern press: The case for contextual digitisation", *The Edinburgh History of the British and Irish Press*, Edinburgh University Press, Edinburgh, Beginnings and Consolidation, 1640–1800.
- Greenberger and others 1962** Greenberger, M., and others (1962) *Management and the computer of the future*. M.I.T Press & Wiley, New York.
- Haraway 1988** Haraway, D. (1988) "Situated knowledges: The science question in feminism and the privilege of partial perspective", *Feminist Studies* 14 (3): 575–99. <http://www.jstor.org/stable/3178066>.
- Havens 2024** Havens, L.J. (2024) "Recalibrating machine learning for social biases: Demonstrating a new methodology through a case study classifying gender biases in archival documentation". PhD thesis, University of Edinburgh.
- Hinrichs, Forlini, and Moynihan 2019** Hinrichs, U., Forlini, S., and Moynihan, B. (2019) "In defense of sandcastles: Research thinking through visualization in digital humanities", *Digital Scholarship in the Humanities* 34(Supplement_1): pp. i80–i99. <https://doi.org/10.1093/llc/fqy051>
- Johnson 2018** Johnson, J.M. (2018) "Markup bodies: Black [life] studies and slavery [death] studies at the digital crossroads", *Social Text* 36(4), pp. 57–79. <https://doi.org/10.1215/01642472-7145658>
- Klein 2013** Klein, L.F. (2013) "The image of absence: Archival silence, data visualization, and James Hemings", *American Literature* 85(4), pp. 661–88. <https://doi.org/10.1215/00029831-2367310>
- Kleymann and Stange 2021** Kleymann, R., and Stange, J.-E. (2021) "Towards hermeneutic visualization in digital literary studies", *Digital Humanities Quarterly* 15(2). Available at: <https://dhq.digitalhumanities.org/vol/15/2/000547/000547.html>
- Ko et al. 2020** Ko, A.J., Oleson, A., Ryan, N., Register, Y., Xie, B., Tari, M., Davidson, M., Druga, S., and Loksa, D. (2020) "It is time for more critical CS education", *Communications of the ACM* 63(11), pp. 31–33. <https://doi.org/10.1145/342400>
- Koeser and LeBlanc 2024** Koeser, R.S., and LeBlanc, Z. (2024) "Missing data, speculative reading", *Journal of Cultural Analytics* 9(2). <https://doi.org/10.22148/001c.116926>
- Kramsch 2014** Kramsch, C. (2014) "Language and culture", *AILA Review* 27(1), pp. 30–55. <https://doi.org/10.1075/aila.27.02kra>
- Ladson-Billings 1997** Ladson-Billings, G. (1997) "It doesn't add up: African American students' mathematics achievement", *Journal for Research in Mathematics Education* 28(6), pp. 697–708. <https://doi.org/10.2307/749638>
- Lamqaddam et al. 2018** Lamqaddam, H., Brosens, K., Truyen, F., Beerens, R.J., De Prekel, I., and Verbert, K. (2018) "When the tech kids are running too fast: Data visualisation through the lens of art history research", *IEEE Transactions on Visualization and Computer Graphics*.

- Latour 1990** Latour, B. (1990) "Technology is society made durable", *The Sociological Review* 38(S1), pp. 103–31.
- Latour and Lowe 2011** Latour, B., and Lowe, A. (2011) "The migration of the aura exploring the original through its fac similes", *Switching Codes Thinking Through Digital Technology in the Humanities and the Arts*, pp 275–97.
- Lee et al. 2025** Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., and Wilson, N. (2025) "The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers". In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI '25. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713778>.
- Lee 1993** Lee, S.M. (1993) "Racial classifications in the US census: 1890–1990", *Ethnic and Racial Studies* 16(1), pp. 75–94. <https://doi.org/10.1080/01419870.1993.9993773>
- Long and So 2016** Long, H., and So, R.J. (2016) "Literary pattern recognition: Modernism between close reading and machine learning", *Critical Inquiry* 42(2), pp. 235–67. <https://doi.org/10.1086/684353>
- Manovich 2013** Manovich, L. (2013) "Database as symbolic form", in R. Perry (ed.) *Museums in a Digital Age*, pp. 64–71. Routledge.
- Markus 1987** Markus, G. (1987) "Why is there no hermeneutics of natural sciences? Some preliminary theses", *Science in Context* 1(1), pp. 5–51.
- Martin 2013** Martin, D.B. (2013) "Race, racial projects, and mathematics education", *Journal for Research in Mathematics Education* 44(1), pp. 316–33. <https://doi.org/10.5951/jresmetheduc.44.1.0316>
- Masson 2017** Masson, E. (2017) "Humanistic data research: An encounter between epistemic traditions". In *The Datafied Society: Studying Culture Through Data*, 25–38. Amsterdam University Press. <http://www.jstor.org/stable/j.ctt1v2xsqn.6>.
- Masure 2018** Masure, A. (2018) "Vers des humanités numériques critiques". Repéré à Dlis. Hypotheses. Org/2088.
- McCarty 2005** McCarty, W. (2005) *Humanities computing*. Springer.
- McLuhan 1994** McLuhan, M. (1994) *Understanding media: The extensions of man*. MIT Press.
- Meister, Drucker and Rockwell 2017** Meister, J.C., Drucker, J., and Rockwell, G. (2017) "Modelling interpretation in 3DH: New dimensions of visualization". In Digital Humanities Conference 2017. Available at: <https://dh2017.adho.org/abstracts/058/058.pdf>
- Minh et al. 2022** Minh, D., Wang, H.X., Li, Y.F., and Nguyen, T.N. (2022) "Explainable artificial intelligence: A comprehensive review", *Artificial Intelligence Review*, pp. 1–66. <https://doi.org/10.1007/s10462-021-10088-y>
- Mitchell et al. 2019** Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., and Gebru, T. (2019) "Model cards for model reporting". In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220–29. <https://doi.org/10.48550/arXiv.1810.03993>
- Munzner 2014** Munzner, T. (2014) *Visualization analysis and design*. CRC Press.
- Nyhan and Passarotti 2019** Nyhan, J., and Passarotti, M. (2019) *One origin of Digital Humanities: Fr Roberto Busa in his own words*. Springer.
- Parker et al. 2016** Parker, K., Horowitz, J.M., Morin, R., and Lopez, M.H. (2015) "Race and multiracial Americans in the US census", *Pew Research Center*. Available at: <https://www.pewresearch.org/social-trends/2015/06/11/chapter-1-race-and-multiracial-americans-in-the-u-s-census/>
- Piper 2016** Piper, A. (2016) "There will be numbers", *Journal of Cultural Analytics* 1 (1).
- Piper 2022** Piper, A. (2022) "Biodiversity is not declining in fiction", *Journal of Cultural Analytics* 7(3). <https://doi.org/10.22148/001c.38739>
- Porter 1995** Porter, T.M. (1995) *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press. Available at: <http://www.jstor.org/stable/j.ctt7sp8x>
- Posner 2015** Posner, M. (2015) "What's next: The radical, unrealized potential of digital humanities", *Miriam Posner's Blog*. Available at: <https://miriamposner.com/blog/whats-next-the-radical-unrealized-potential-of-digital-humanities/>
- Potter, Rosen, and Johnson 2012** Potter, K., Rosen, P., and Johnson, C.R. (2012) "From quantification to visualization: A taxonomy of uncertainty visualization approaches". In *Uncertainty Quantification in Scientific Computing: 10th IFIP WG 2.5 Working Conference, WoCoUQ 2011, Boulder, CO, USA, August 1-4, 2011, Revised Selected Papers*, pp. 226–49.

- Presner 2010** Presner, T. (2010) "Digital humanities 2.0: A report on knowledge". Citeseer.
- Quamen 2012** Quamen, H. (2012) "The limits of modelling: Data culture and the humanities", *Scholarly and Research Communication* 3(4). <https://doi.org/10.22230/src.2012v3n4a69>
- Ramsay 2010** Ramsay, S. (2010) "The hermeneutics of screwing around; or what you do with a million books", in K. Khee (ed.) *Pastplay: Teaching and Learning History with Technology*, pp. 111–20.
- Ramsay 2011** Ramsay, S. (2011) *Reading machines: Toward and algorithmic criticism*. University of Illinois Press.
- Ramsay and Rockwell 2012** Ramsay, S., and Rockwell, G. (2012) "Developing things: Notes toward an epistemology of building in the digital humanities", in M.K. Gold (ed) *Debates in the Digital Humanities*. Available at: <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfd1e/section/c733786e-5787-454e-8f12-e1b7a85cac72>
- Reed 2019** Reed, E. (2019) "Forms of frustration: Unrest and unfulfillment in American literature after 1934". PhD thesis, University of Virginia.
- Risam 2019** Risam, R. (2019) "Beyond the migrant 'problem': Visualizing global migration", *Television & New Media* 20(6), pp. 566–80. <https://doi.org/10.1177/1527476419857679>
- Scheuerman et al. 2020** Scheuerman, M.K., Wade, K., Lustig, C., and Brubaker, J.R. (2020) "How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis", *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW1). pp. 1–35. <https://doi.org/10.1145/3392866>
- Schickore 2005** Schickore, J. (2005) "'Through thousands of errors we reach the truth' — but how? On the epistemic roles of error in scientific practice", *Studies in History and Philosophy of Science Part A* 36 (3), pp. 539–56. <https://doi.org/10.1016/j.shpsa.2005.06.011>
- Sculley and Pasanek 2008** Sculley, D., and Pasanek, B.M. (2008) "Meaning and mining: The impact of implicit assumptions in data mining for the humanities", *Literary and Linguistic Computing* 23(4), pp. 409–24. <https://doi.org/10.1093/lilc/fqn019>
- Sinclair and Gouglas 2002** Sinclair, S., and Gouglas, S.W. (2002) "Theory into practice: A case study of the humanities computing master of arts programme at the University of Alberta", *Arts and Humanities in Higher Education* 1(2), pp. 167–83. <https://doi.org/10.1177/147402220200100200>
- Sinclair et al. 2013** Sinclair, S., Ruecker, S., Radzikowska, M., and Inke, I. (2013) "Information visualization for humanities scholars", *Literary Studies in the Digital Age-an Evolving Anthology*.
- Smith 1985** Smith, B.C. (1985) "The limits of correctness", *ACM SIGCAS Computers and Society* 14(1, 2, 3, 4), pp. 18–26. <https://doi.org/10.1145/379486.37951>
- Snow 1959** Snow, C.P. (1959) "Two cultures", *Science* 130(3373), pp. 419–19. <https://doi.org/10.1126/science.130.3373.419>
- Sollazzo 2022** Sollazzo, A. (2022) "Embracing the friction: Towards a computationally aware approach to humanistic data interfaces". Master's thesis, University of Alberta.
- Tiefenbach Keller 2024** Tiefenbach Keller, S. (2024) "Towards a culture of replication in the digital humanities: An understanding of transparency and openness practices in published DH papers". PhD thesis, McGill University.
- Tufte and Graves-Morris 1983** Tufte, E.R., and Graves-Morris, P.R. (1983) *The visual display of quantitative information*. Graphics Press Cheshire, CT.
- Turkle and Papert 1990** Turkle, S., and Papert, S. (1990) "Epistemological pluralism: Styles and voices within the computer culture", *Signs: Journal of Women in Culture and Society* 16(1), pp. 128–57. Available at: <https://www.jstor.org/stable/3174610>
- Underwood 2014** Underwood, T. (2014) "Theorizing research practices we forgot to theorize twenty years ago", *Representations* 127(1), pp. 64–72. <https://doi.org/10.1525/rep.2014.127.1.64>
- Van Es, Wieringa, and Schäfer 2018** Van Es, K., Wieringa, M., and Schäfer, M.T. (2018) "Tool criticism: From digital methods to digital methodology". In *Proceedings of the 2nd International Conference on Web Studies*, pp. 24–27. <https://doi.org/10.1145/3240431.3240436>

- Van Zundert et al. 2012** Van Zundert, J., Antonijevic, S., Beaulieu, A., van Dalen-Oskam, K., Zeldenrust, D., and Andrews, T.L. (2012) "Cultures of formalisation: Towards an encounter between humanities and computing", in D.M. Berry (ed.) *Understanding Digital Humanities*, pp. 279–94. Springer. https://doi.org/10.1057/9780230371934_15
- Van Zundert 2015** Van Zundert, J.J. (2015) "Screwmenetics and hermenumericals: The computability of hermeneutics", in S. Schreibman, R. Siemens, and J. Unsworth (eds.) *A New Companion to Digital Humanities*, pp. 331–47. <https://doi.org/10.1002/9781118680605>
- Verhoeven et al. 2020** Verhoeven, D., Musial, K., Palmer, S., Taylor, S., Abidi, S., Zemaityte, V., and Simpson, L. (2020) "Controlling for openness in the male-dominated collaborative networks of the global film industry", *PloS One* 15(6): e0234460. <https://doi.org/10.1371/journal.pone.0234460>
- Walter 2018** Walter, M. (2018) "The voice of indigenous data: Beyond the markers of disadvantage", *Griffith Review*, 60, 256–63. Available at: <https://www.griffithreview.com/articles/voice-indigenous-data-beyond-disadvantage/>
- Walter and Andersen 2013** Walter, M., and Andersen, C. (2013) *Indigenous Statistics: A Quantitative Research Methodology*. Taylor & Francis.
- Walter and Suina 2019** Walter, M., and Suina, M. (2019) "Indigenous data, indigenous methodologies and indigenous data sovereignty", *International Journal of Social Research Methodology* 22(3), pp. 233–43. <https://doi.org/10.1080/13645579.2018.1531228>
- Witmore 2010** Witmore, M. (2010) "Text: A massively addressable object", *Wine Dark Sea* 31. Available at: <https://winedarksea.org/?p=926>
- Yoder-Himes et al. 2022** Yoder-Himes, D.R., Asif, A., Kinney, K., Brandt, T.J., Cecil, R.E., Himes, P.R., Cashion, C., Hopp, R.M.P., and Ross, E. (2022) "Racial, skin tone, and sex disparities in automated proctoring software". In *Frontiers in Education*, 7:881449. <https://doi.org/10.3389/educ.2022.881449>
- Young, Morgan, and Young 1987** Young, R.W., Morgan, W., and Young, R.W. (1987) *The Navajo Language: A Grammar and Colloquial Dictionary*. University of New Mexico Press Albuquerque.

Recommendations

DHQ is testing out three new article recommendation methods! Please explore the links below to find articles that are related in different ways to the one you just read. We are interested in how these methods work for readers—if you would like to share feedback with us, please complete our short evaluation survey. You can also visit our documentation for these recommendation methods to learn more.

SPECTER Recommendations

Below are article recommendations generated by the SPECTER model:

1. Humanities Computing as Digital Humanities, 2009, Patrik Svensson, Umeå University
2. The Sound of Many Hands Clapping: Teaching the Digital Humanities through Virtual Research Environment (VREs), 2012, Craig Bellamy, VerSI, University of Melbourne, Australia
3. Tractable Tensions: A Review of Digital Humanities: Knowledge and Critique in a Digital Age by David M. Berry and Anders Fagerjord, 2024, Onyekachi Henry Ibekwe, University of Nigeria
4. Towards a Conceptual Framework for the Digital Humanities, 2012, Paul S. Rosenbloom, Department of Computer Science and Institute for Creative Technologies, University of Southern California
5. User Experience (UX) Heuristics for the Digital Humanities, 2025, Elizabeth Rodwell, University of Houston; Kristina Neumann, University of Houston; Peggy Lindner, Missouri University of Science and Technology

DHQ Keyword Recommendations

Below are article recommendations generated by DHQ Keywords:

1. Digital Humanities: On Finding the Proper Balance between Qualitative and Quantitative Ways of Doing Research in the Humanities, 2013, Helle Porsdam, University of Copenhagen
2. Introduction, 2017, Emily Christina Murphy, Queen's University; Shannon R. Smith, Bader International Study Centre, Herstmonceux Castle, Queen's University
3. The Digital Future is Now: A Call to Action for the Humanities, 2010, Christine L. Borgman, Professor & Presidential Chair in Information Studies, UCLA
4. Video-gaming, Paradise Lost and TCP/IP: an Oral History Conversation between Ray Siemens and Anne Welsh, 2012, Ray Siemens, University of Victoria; Anne Welsh, University College London; Julianne Nyhan, University College London; Jessica Salmon, University of Trier

5. Student Labour and Training in Digital Humanities, 2016, Katrina Anderson, Simon Fraser University; Lindsey Bannister, Simon Fraser University; Janey Dodd, University of British Columbia; Deanna Fong, Simon Fraser University; Michelle Levy, Simon Fraser University; Lindsey Seatter, University of Victoria

TF-IDF Recommendations

Below are article recommendations generated by the TF-IDF Model:

1. The Landscape of Digital Humanities, 2010, Patrik Svensson, HUMlab, Umeå University
2. Voluptuous Disintegration: A Future History of Black Computational Thought, 2022, Romi Ron Morrison, University of Southern California (USC)
3. Scholarly Infrastructure as Critical Argument: Nine principles in a preliminary survey of the bibliographic and critical values expressed by scholarly web-portals for visualizing data, 2020, Jo Guldi, Southern Methodist University
4. Unpacking tool criticism as practice, in practice, 2023, Karin van Es, Utrecht University
5. Sensitivity and Access: Unlocking the Colonial Visual Archive with Machine Learning, 2024, Jonathan Dentler, Catholic University of Paris; German Historical Institute, Washington D.C.; Lise Jaillant, Loughborough University, UK; Daniel Foliard, Université Paris Cité, LARCA (UMR 8225); Julien Schuh, Université Paris Nanterre; Maison des Sciences de l'Homme Mondes



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.