

PROJECT MUSE

Digital Melville and Computational Methods in Literary Studies

Dennis Mischke, Christopher Ohge

Leviathan, Volume 25, Number 2, June 2023, pp. 35-60 (Article)

Published by Johns Hopkins University Press *DOI: https://doi.org/10.1353/lvn.2023.a904374*



➡ For additional information about this article https://muse.jhu.edu/article/904374

For content related to this article https://muse.jhu.edu/related_content?type=article&id=904374

Digital Melville and Computational Methods in Literary Studies

DENNIS MISCHKE

Freie Universität Berlin

CHRISTOPHER OHGE

University of London

The use of computational tools for the study of literature can facilitate new perspectives and avenues for critical work. Reflecting on the recent emergence of (and increasing hype around) large languages models such as GPT, this essay argues that the creation of "smart" data sets and corpora as new forms of literary objects requires and enables the development of computational methods and tools that can create "data stories". Smart data sets continue a humanistic tradition of textual scholarship and bibliography while also preparing text data to be "read" by machines. In telling "data stories" about Herman Melville, we bridge the gap from "numbers to meaning" in a variety of examples from *Billy Budd.* The essay closes with a broader reflection on reading Melville in the age of machine learning and artificial intelligence.

Digital Approaches to Melville Studies?

Herman Melville's writing has a rich and storied history, with a long-standing place in the literary canon and a lasting impact on American literature and cultural studies. However, in recent years, the field of Digital Humanities has sought to analyze and interpret Melville's texts through the lens of technology and digital tools. While these approaches may offer some new insights, it is important to critically examine the role of technology in our interpretation of literature. In this essay, we will explore the intersection of Melville and Digital Humanities, considering the potential limitations and pitfalls of using digital technologies to analyze and interpret his works. We will also consider how these approaches may risk reducing the complexity and nuance of Melville's writing and obscuring the deeper cultural and historical contexts in which it was produced.

n introduction to an academic essay on the topic of Melville and Digital Humanities might begin with these words above. But, mindful of Tommo's cry in Melville's *Typee* that "Appearances all the world over

Vol. 25.2 (2023): 35-60 © 2023 The Melville Society and Johns Hopkins University Press

are deceptive," (206) we must add, ceci n'est pas un abstract; the text above was generated by OpenAI's ChatGPT. ChatGPT is already unprecedented both for its widespread use and its feverish treatment in the press. It is a tool for generative artificial intelligence, based on a large language model (LLM) trained on the entire internet and all publicly available human produced text until roughly the year 2021.¹ It appears that ChatGPT "writes" new text, but it is actually a predictive software (like an enhanced autocomplete): it is good at predicting what word should come next given a series of prompts; each prompt then instructs the tool to generate new text. When advances in Artificial Intelligence (AI) are made to write a research paper on Herman Melville (or any other topic for that matter), the "stochastic parrots"² of large language models such as ChatGPT will produce seemingly convincing models of academic writing and analysis and yet remain neither artificial nor intelligent. What may result is a "textpocalypse," a "tsunami of text" which, according to Matthew Kirschenbaum, will redefine our relationship to writing. Yet many human writers are themselves stochastic parrots who-consciously or not-engage in formulaic thinking, text re-use, and various forms of copying. As tempting as it may seem to have an omniscient oracle and automaton at one's side, the production of research at the push of a button has never been a primary objective of digital humanities researchers. A large language model only reflects what humans have already said, but it is not discriminating nor does it understand human motives, context, or intentions. Such computational systems also do not understand ambiguities and abstractions. Despite their seeming understanding of two topics such as Herman Melville and Digital Humanities, these systems have no semantic understanding or hierarchical processing models that are essential to other kinds of language tools. Why, then, should AI tools and LLMs like GPT matter for literary studies (other than the obvious issue of automating the creation of text), and how do they stand in relation to other Digital Humanities tools? Do infrastructures of computation and machine learning actually intersect with the literary text as such? If so, how do the infrastructures of textuality reflect back on the questions we ask of literary texts and figures like Melville? Finally, as many other have already asked, how do scholars and teachers guard against the dangers of ChatGPT while also highlighting its gains?

These questions turn upon the important issue of methodology in digital literary studies and Digital Humanities more generally. The last twenty-odd years of labor-intensive digital projects such as Melville's Marginalia Online (MMO), the Melville Electronic Library (MEL), and Melville's Print Collection Online (MPCO) have turned the object of investigation in literary and cultural studies into a different object with different affordances. Digital scholarly editions and archives consist of what Christof Schöch calls "smart data"—that is, data collections of literary texts such as MEL's Billy Budd that can be processed with a variety of digital tools. These tools are representations of a smart data set; they can range from a print book (or e-book) that can be read and studied by human beings as well as machine-actionable data that can be processed and displayed on the computer (see Chapters 1 and 5 of Ohge 2021). As Ryan Cordell emphasizes, digitization "does not remove a historical artifact from material culture, but adds another stratum of computational materiality to its social text" (20). In a world increasingly shaped and driven by digital tools, digital infrastructure, algorithms, and data, digital literary studies must reckon with data as a new source for study. Such a new method entails a variety of questions of knowledge structure, provenance, and deployment through tools, interfaces, and what we call "data stories." Considering the literary dataset as a corpus, and as a distinct genre, as suggested by Michael Gavin's recent work, Literary Mathematics, we will demonstrate the various ways that Digital Humanities methods can furnish "data stories." We will illustrate these data stories by progressing through various DH approaches in a more or less chronological order: structured data from the Melville Electronic Library, text analysis tools that analyze unstructured text data from the same source, and a reflection on AI tools that make use of computational dictionaries and LLMs. Before we can discuss possible ways of connecting data and narration, numbers and interpretation, let us consider the human labor and achievements upon which state-of-the-art language models and Digital Humanities methodologies rest.

Textual Scholarship, Text Analysis, and Narratives of Smart Data

E arly digital approaches to Melville were indebted to a rich scholarly tradition, inherited from bibliography and textual scholarship, which focused on the manual curation of digital texts and metadata for the purposes of various tools—scholarly editions, digital archives, and other resources on the Internet. An important method of these approaches is to formalize traditional techniques of philological and bibliographical study into digital text markup. For example, Melville's Marginalia Online, under the direction of Steven Olsen-Smith, rendered Merton Sealts, Jr.'s bibliographic checklist of books Melville read, borrowed, and consulted into a database. The project has since organized Melville's marginalia into structured XML data and released impressive data visualization tools. The Melville Electronic Library also expanded traditional methods of textual studies for digital media. As John Bryant argues in "Access and Affiliation: A Biography of Digital Melville" (9–34), MEL uses a

"fluid text" approach to textual editing, and uses digital technology to publish editions that enable users to navigate a work's textual fluidity (Bryant 2002). The goal of both enterprises is to make texts and images of primary materials more accessible, but they are also creating valuable datasets that are grounded in scholarly traditions. Such enhanced accessibility means not only more available texts that can be read and searched but also more reliable data sets on which we can apply analyses of various kinds.

On the other hand, a new tradition of semi-automated and automated analyses and generation of texts is burgeoning. For a couple of decades, this approach has been represented by corpus linguistic approaches, dictionarybased topic clustering, Natural Language Processing (NLP) tools (in which a programming language automatically organizes and / or tags texts and produces statistics on them), stylometric tools, and (recently) machine learning and artificial intelligence tools. Both manual and automated approaches have gains and losses, and their ultimate success depends, as always, on the critical payoffs of the results. These approaches also raise questions about the crucial elements of judgment and discourse with regards to these tools. Are these computational tools the generators of facts or knowledge? Or, are they tools for producing more useful metaphors, or more reliable narratives about the significance of our literary heritage and literary practices themselves? In an age of machine learning, neural networks, and artificial intelligence, what is a literary text made of? Who made it? Where is the text in this computer? Grappling with these questions highlights the age-old critical questions of aboutness-the intentionality of writers and readers, the elements of style, the material facts of textual fluidity. These digital tools will never supplant critical thinking, bibliographic rigor, and the material histories of texts, as Katherine Bode has persuasively argued (see Bode 2017 and 2020). By iterating between data and discourses of knowledge, digital methods can sharpen our critical facilities.

Given the ubiquity of digital tools now, we might even think of computation and the datafication of text in ever larger digital text repositories as the infrastructures of everyday life—including many aspects of academic work. In viewing computational tools as forms of external cognition and "computational offloading" (Rogers and Brignull; Miller)—think of the cell phone in your pocket that yields more computational power than the largest supercomputer 40 years ago—we can begin to see that computation yields much more than merely calculating numbers. In fact, the recent revolution of LLMs and other AI systems based on neural networks are slowly but surely developing into frameworks for thinking, platforms for reading, and infrastructures of writing. Intellectual work has always been reliant on infrastructures (libraries, printing presses, publishing houses, labor, and the like); however, the infrastructural aspects of computation in the present moment increasingly act as an omnipresent "sub-sub librarian" who is "picking up whatever random allusions" as a "painstaking burrower and grub-worm," to borrow from Moby-*Dick*. The product of LLMs like ChatGPT, whose neural architecture is in large part based on the ability to utilize and detect what Vaswani has called "long range dependencies" between individual tokens and words of texts, can be seen as an almost Melvillean aggregation of quotes schematically like "Extracts" at the beginning of Moby-Dick. The machine-based aggregation of texts that lies at the heart of generative systems such as ChatGPT would be unthinkable without the long tradition of text encoding and annotation. The very idea of machine learning is fundamentally predicated on the human intellectual labor of encoding (annotating) data sets to train machine learning systems. Thus, as many commentators in DH have suggested, text encoding and annotation not only facilitate the creation of smart data and new forms of publication; they also constitute a form of close reading that complements distant readings and machine-generated textual output. As a result, it is not uncommon to find digital textual scholars offering interpretations of their manual encoding (see Cummings, Posner, Singer, and Eve).

To demonstrate the importance of text encoding as a kind of close textual analysis, we will focus on *Billy Budd*, which has so far received the most granular digital treatment of all Melville's works. The digital edition of *Billy Budd* was first released in 2019, after roughly 10 years of collaborative planning, data entry, and annotation.³ The workflow was painstaking: following the 1962 Hayford-Sealts genetic transcription of the manuscript, MEL editors created new diplomatic transcriptions of the *Billy Budd* manuscript leaves, matched revisions to the digital facsimiles of the manuscript, and encoded each revision with semantic XML vocabulary using the Text Encoding Initiative's P5 guidelines, which describe features such as the purpose, composition stage, and medium of each revision (see Fig. 1).⁴

Much of this work demonstrated what Hershel Parker, Ohge, and John Wenke have suggested—namely, that Melville complicated the mental state of Captain Vere as he revised, but the manuscript's incompleteness makes it difficult to ascertain Melville's intentions in revising and refining the story (Parker 1990; Ohge 2017; Wenke 2022). In a basic sense, the availability of edited transcriptions of the manuscript makes it possible for users to examine these ideas for themselves. But the digital work also foregrounds Melville's process rather than the old method of establishing the text that adheres to editorial theories of

into	
tion	
crip	D
rans	odin
tic ti	enco
ene	ML
C	X

Billy Budd, Sailor, ed. Harrison Hayford and Merton Sealts, Jr. (U of Chicago P, 1962) In the time before steamships, and [p Ba < and > or then] here traquently than now [p Ba add comma] a strother elong the doels of any considerable sea-port would occasionally have his attention arrested by a group of bronzed mariners, man-of-war's men or merchant-sailors in holiday attire ashore on liberty. In certain instances [p <In certain instances >Sometimes xx (of the set of the set

Melville Electronic Library edition

Ш

<ab>

<add place="inline" rend="no-caret _ink2"
hand="#HM" change="StEa" facs="#img_5-0019"> "dMH hand="#HM" change="StEa" facs="#img_5-0019" hand="#HM" change="StEa" facs="#img 5-0020" <add place="above" rend="caret " hand="#HM"</pre> <metamark place="inline" function="caret"</pre> <lb/>sea-port would occasionally have his change="StEa" facs="#img_5-0019" ><subst
facs="#img_5-0019" ><de1 rend="erasure_1"</pre> <lb/>slong the docks of any considerable <lb/><lb/></lb/> <lb/>slb/>In the time before steamships, rend="no-caret">, </add> a stroller facs="#img_5-0024" place="inline" rend="caret _ink1" change="StEa" facs="#img_5-0030" >^</metamark> more frequently than now<add or then</add></subst></add> >or then >and

Fig. 1. Comparison of the Hayford-Sealts genetic transcription of the Billy Budd manuscript (left) with the MEL encoding of the manuscript (right).

</ab>



Next Leaf **O**

Diplomatic

Base



Fig. 2. Diplomatic manuscript transcription of the opening of Billy Budd, Melville Electronic Library (https://app.textlab.org/ transcriptions/16900). Melville's "final" or "latest" intentions. The same image in the previous figure can be examined in MEL's "diplomatic" view to give readers direct engagement with a snapshot of Melville's process.

When the user hovers over a revision (shown in Fig. 2 with Melville's substitution of "or then" for "and" in the opening line of Chapter 1), the TEI-XML encoding becomes in the interface a mouse-over note with information about the revision (the place of the revision, the hand, the image, the type of revision, how it was rendered, and at which change/composition stage it occurred).

Doing this editorial work can also provide critical insights that may have never occurred to even the most experienced readers of Melville's texts. One of the most basic reasons for undertaking the text encoding is to generate statistics of the encoded text. This method can be accomplished in a variety of ways, but the two most common—and useful—ones are to use XPath to query the XML data and then to use text analysis methods to tokenize⁵ and calculate word frequencies of the data. Such a workflow enables us to count the number of revision types (Fig. 3) in the manuscript, and then sequence the revisions per composition stage (Fig. 4).

Two facts immediately appear from this analysis: Melville deleted more than he added overall, and he revised more in later stages of composition (stages F, G, and p). It would have been nearly impossible to generate these facts without having carefully encoded each revision in the manuscript with the compositional information from the Hayford-Sealts printed transcription.

We can then reveal other important analyses. On a basic level, text analysis starts with calculating word frequencies in a text; it is basic counting. Given that we are interested in revision, we can quickly generate relevant information such as word lists of revision-words (Fig. 5).

This first list shows the top results—mostly function words (which, according to linguists, can still be worthy of attention)—but notice that even though in revising Melville deleted more often than he added, he added more words (which is not surprising because we know he was adding new material in late stages). Here we could also pursue a line of thought about Melville's tendency to delete and whether he used a high frequency of negation words while revising the text (Fig. 6). This list generates some proof for the claim that Melville has a higher frequency of negation-words ("not," "never," "nothing," for example) in added text than in deleted text.

Related to this idea of negation is sentiment—to what extent did Melville's increased use of negation words in *Billy Budd* change the shape of the narrative?⁶

To address this question, we processed the TEI-XML data of the *Billy Budd* manuscript into unrevised and revised texts, which along with composition



Fig. 3. Bar graph showing the number of revision types in the Billy Budd manuscript.



Number of revisions per composition stage of Billy Budd

Fig. 4. Bar graph showing the number of revisions per composition stage in the *Billy Budd* manuscript. The abbreviations for composition stages follow the Hayford and Sealts designations for Melville's manuscript that use letters (A, B, C, and so on) to represent inscription sequences.

	del_word_v	Freq	add_word_v	Freq
1		11289		31920
2	the	349	the	784
3	of	191	of	448
4	а	141	to	286
5	and	129	a	284
6	to	127	in	252
7	in	105	and	246
8	it	84	his	162
9	his	82	was	157
10	that	81	as	151
11	was	66	it	147
12	as	59	that	141

Deleted and Added Words in the Billy Budd MS

Fig. 5. Table of the top twelve most frequent words deleted and added in the *Billy Budd* manuscript.

Negation words in added text

15	not	51	he	107
16	is	49	at	103
17	at	43	him	98
18	him	40	not	88
76	tho	10	are	21
77	what	10	never	21
		1	1	
89	fleet	8	nothing	18

Fig. 6. Table of word frequencies of deleted words (left) and added words (right) in the *Billy Budd* manuscript, with selected negation words bolded.

stage data, further enables us to approximate earlier and later versions of the text. Sentiment analysis algorithms were then applied to the early and later drafts to gauge whether there was a noticeable difference in sentiment between Melville's substantive versions (Fig. 7; in color insert).

Applying sentiment analysis to literary texts can be fraught with difficulties, but Matthew Jockers's Syuzhet package offers compelling results.⁷ Rule-based approaches to natural language processing and digital literary studies such as sentiment analysis have been based on basic look-up algorithms that compare lists of words against other lists of words such as dictionaries. Sentiment analysis is a process of chunking textual units (into word roots, single words, n-grams, and whole sentences) and joining them to sentiment dictionaries which classify those textual units into sentiment scores or designations such as positive, negative, or neutral. Other dictionaries, such as NRC, can also group sentiment words into categories such as fear, joy, and anticipation.⁸ The joining operation renders a list of textual units with sentiment scores or designations in a data frame that can be ordered and counted for analysis (see fig. 8). Digital scholars such as Matthew Jockers and Katherine Elkins have also shown that sentiment analysis can also be a gauge of narrative development.

Sentiment analysis nevertheless comes with several caveats. Sentiment dictionaries are based on contemporary language pulled from contemporary Internet sources, which sometimes distort the results of historical texts. To give one brief example, running a sentiment analysis on Shakespeare reveals a high frequency of the word "drunk" as a negative sentiment word, but most readers of Shakespeare would probably agree that several instances of that word in Shakespeare have a positive valence. Perhaps it is needless to say, but tagging any word or series of words with a "sentiment score" is problematic because the sentiment of all words depends not only on the context of surrounding words but also on myriad external factors about the text that would affect the sentiment of the language. The scoring may also seem rather arbitrary. However, to evoke Gavin, these quantitative approaches are not meant to be singular explanations but rather textual data for further measurement and interpretation; the quantitative results form "one small piece of a large and diffuse interdisciplinary project devoted to learning how to use textual modeling to describe and explain society" (Gavin 5).

The Jockers sentiment method has the virtues of analyzing the whole sentence (as opposed to one or two words), using four sentiment dictionaries, and of being created by scholars who are trained in literary studies. The analysis in Fig. 7 shows that Melville's revisions to the manuscript increased Early draft (blue) and edited text (red) of Billy Budd, with Rolling Means



Mischke and Ohge Fig. 7: A graph comparing the sentiment analysis of an early draft of Billy Budd compared to the edited text.

Sentence index	Emotional valence	Sentence
668	-6.5	Which appeal caused but a strange dumb gesturing and gurgling in Billy; amazement at such an accusation so suddenly sprung on inexperienced nonage; this, and, it may be, horror at the accuser's eyes, serving to bring out his lurking defect and in this instance for the time intensifying it into a convulsed tongue-tie; while the intent head and entire form straining forward in an agony of ineffectual eagerness to obey the injunction to speak and defend himself, gave an expression to the face like that of a conderned Vestal priestess in the moment of being buried alive, and in the first struggle against suffocation.
178	-5.5	They may add, too, that at Trafalgar it was in effect nothing less than a challenge to death; and death came; and that but for his bravado the victorious Admiral might possibly have survived the battle, and so, instead of having his sagacious dying injunctions overruled by his immediate successor in command he himself when the contest was decided might have brought his shattered fleet to anchor, a proceeding which might have averted the deplorable loss of life by shipwreck in the elemental tempest that followed the martial one.
161	-5.45	Though after parleyings between Government and the ring-leaders, and concessions by the former as to some glaring abuses, the first uprising–that at Spithead–with difficulty was put down, or matters for the time pacified; yet at the Nore the unforeseen renewal of insurrection on a yet larger scale, and emphasized in the conferences that ensued by demands deemed by the authorities not only inadmissible but aggressively insolent, indicated—if the Red Flag did not sufficiently do so,–what was the spirit animating the men.
1125	ς- -	Ignorant though they were of the secret facts of the tragedy, and not thinking but that the penalty was somehow unavoidably inflicted from the naval point of view, for all that, they instinctively felt that Billy was a sort of man as incapable of mutiny as of wilful murder.
556	-4.85	It was as if his precocity of crookedness (and every vulgar villain is precocious) had for once deceived him, and the man he had sought to entrap as a simpleton, had through his very simplicity ignominiously baffled him.
394	-4.5	Now something such an one was Claggart in whom was the mania of an evil nature, not engendered by vicious training or corrupting books or licentious living, but born with him and innate, in short "a depravity according to nature."

Fig. 8. Snippet of a data frame of sentence-level sentiment values in Billy Budd, sorted by descending values.

the negative sentiment. The graph also shows some noticeable deviations in sentiment during revisions (take note of the differences in the Narrative Time after 0.8 in the figure above).

The sentence to which Syuzhet assigned the most negative sentiment in the revised version of *Billy Budd* (and the most negative score overall at -6.5) is the passage in Chapter 19 when Vere demands a response from Billy about Claggart's accusation, "Speak! defend yourself":

Which appeal caused but a strange dumb gesturing and gurgling in Billy; amazement at such an accusation so suddenly sprung on inexperienced nonage; this, and, it may be, horror at the accuser's eyes, serving to bring out his lurking defect and in this instance for the time intensifying it into a convulsed tongue-tie; while the intent head and entire form straining forward in an agony of ineffectual eagerness to obey the injunction to speak and defend himself, gave an expression to the face like that of a condemned Vestal priestess in the moment of being buried alive, and in the first struggle against suffocation. (https://melville.electroniclibrary.org/editions/versions-of-billy-budd/ chapter-19)

It should be immediately evident why this lengthy sentence returns such a negative score: it has "dumb gesturing" and "gurgling", then "horror," "defect," and "agony," and concludes with the comparison of Billy to a "Vestal priestess" being buried alive. Crucially, though, this moment in the text is a product of late revision—it is Melville's late attempt to work out a trope that he had attempted in previous works. Billy has a problem with stuttering and violent outbursts, but so do earlier characters in Melville's work.

In Chapter II of *Redburn*, for example, the eponymous character also struggles with stuttering, feelings of suffocation (anxiety), and anger, in the scene when he is confronted by the captain's clerk who asks for him to pay for the ticket:

My whole soul was soured within me, . . . I buttoned up my coat to the throat, clutched my gun, put on my leather cap, and pulling it well down, stood up like a sentry before him. He held out his hand, deeming any remark superfluous, as his object in pausing before me must be obvious. But I stood motionless and silent, and in a moment he saw how it was with me. I ought to have spoken and told him the case, in plain, civil terms, and offered my dollar, and then waited the event. But I felt too wicked for that. He did not wait a great while, but spoke first himself; and in a gruff voice, very unlike his urbane accents when accosting the wine and cigar party, demanded my ticket. I replied that I had none. He then demanded the money; and upon my answering that I had not enough, in a loud angry voice that attracted all eyes, he ordered me out of the cabin into the storm. The devil in me then mounted up from my soul, and spread over my frame, till it tingled at my finger ends;

and I muttered out my resolution to stay where I was, in such a manner, that the ticket man faltered back. (12-13)

Redburn engages in a tense exchange with the captain's clerk on the Hudson steamship and ends with his admission that behavior like this is typical of boyhood.⁹ Later, in Chapter VII, he recalls his father's downfall and notices that "something rises up in my throat and almost strangles me" (36). The sense of words being strangled or suffocated—the inability to render experience into verbal representations, and the latent anger accompanying that inability ("I almost hated them," he says about his fellow passengers)—was a longstanding concern of Melville's.

The word "suffocation" in the *Billy Budd* passage is also interesting because it only appears nine times in all of Melville's fiction—three of which appear in *Pierre*. Consider this passage from the concluding paragraph to Book IX of *Pierre*:

Impossible would it be now to tell all the confusion and confoundings in the soul of Pierre, so soon as the above absurdities in his mind presented themselves first to his combining consciousness. He would fain have disowned the very memory and the mind which produced to him such an immense scandal upon his common sanity. Now indeed did all the fiery floods in the Inferno, and all the rolling gloom in Hamlet suffocate him at once in flame and smoke. The cheeks of his soul collapsed in him: he dashed himself in blind fury and swift madness against the wall, and fell dabbling in the vomit of his loathed identity. (171)

Here the "combining consciousness"—an inheritance from Hamlet—results in the collapse of sanity, of madness and anger. In both *Redburn* and *Pierre*, Melville is examining the combining consciousness of the boy consumed with uncertainties, anger, and the inability to use words to express their cognitive dissonance. Yet that sentence in *Billy Budd* still stands out, not only because we have evidence of revision that we do not have in the other works, but also because he is disrupting a previously used trope of violent male youth.

In *Billy Budd*, the simile of a "Vestal priestess" is shocking in its implications and adds to Melville's previous attempts to analyze the rage and violence and sense of suffocation. In the revisions to *Billy Budd* Melville compares Billy to a virgin woman, something he could not quite do in previous fictions. Billy is not perfect—he is a repressed and violent character, and this is a new dimension, again, that Melville clarified in revision. Now, many readers may have taken note of the simile, or parts of this lengthy sentence, or even pondered how it reflects Billy's tragic flaw. Nevertheless, it is unlikely that anyone would be able to situate this passage within the context of sentiment values, and sentiment values between two rough versions of the text that only textual editing can facilitate. There is a factual basis to this negative sentence that spurs further critical considerations. This digital analysis enables a new form of attention that lends itself to new critical analyses. Why is this sentence so "negative"? Melville's sentence concerns the inner psychology of stunned silence and the physical manifestations of the horror—all wrapped in a terrifying simile about a vestal priestess being buried alive.

In response to this critical analysis, we could imagine a skeptical reader or scholar asking "how is this different from the way we read books—maybe I have already noted the passage and connected it to other passages with traditional methods?" Or, perhaps, "how can I trust these calculations"? One important way in which this method is different harkens back to Gavin's idea of the corpus. The text to which we are attending is a data frame rather than a book or document. The quantitative results are also meant to be analyzed with our familiar critical tools of reading. So in the previous example of *Redburn*, we do not have the book open on our table but the corpus open on our computer, which can itself be the site of the kind of close reading we aim to achieve, not of a single text but of the writer's creative process over time (see Fig. 9 and Fig. 10).

The locus of a reader's attention is the textual unit rendered in digital code that is then processed in a dataframe that can order those textual units in a variety of quantitative and logical permutations. The linearity of reading is disrupted to allow for non-sequential connections between texts. In essence, this activity recalls the original conception of hypertext, as formulated by Ted Nelson (Fig. 11), which holds that hypertext is a non-linear "unified text object" which cannot be usefully printed and should connect one or more *editorially* related things (Nelson 6).

The factual computation leads to a focused form of data-driven storytelling through editorially connected things. Echoing R. B. Braithwaite's articulation of deductive conceptualization, this kind of computational methodology allows hypotheses to become "theoretical concepts" which are connected to the facts of the text(s) arrived at through complicated logical and numerical relationships (Braithwaite ix). However, as this is not strictly a science, we must also acknowledge the contingencies involved in our articulation of such relationships. To borrow from Poincaré's formulation of models of belief systems, we cannot start with all the facts of the text(s); rather we start with intuitively selected facts that form a network consisting of established and new facts.

In one sense, then, in the previous example we are modeling a selective chain of associative processes of Melville's thought—misanthropic boyhood, anger, violence, the inability to express the angst of youth and injustice which was arrived at through a data analysis (the facts in this case would be

AntConc				
KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud				
Γot	al Hits: 26 Pa	age Size 100 hits 😋 🔅 1 to 26 of 26 hits	0	
	File	Left Context	Hit	Right Context
1	k_confiden	this man of straw is a whit less wise than	anger	against a man of flesh, Madness, to be mad
2	k_confiden	hristian light," he added pensively, "I don't know that	anger	against this man of straw is a whit less
3	m_clarel.txt	there is a memory. Within a garden walking see The	angered	God. And where the vine And olive in the
4	d_redburn	round me, and if I had not felt so terribly	angry,	I should certainly have felt very much like a
5	a_typee.txt	ill impression he might have received. But the ireful,	angry	chief was not so easily mollified. He rejected my
6	a_typee.txt	t last Kolory speaks a little louder, and soon growing	angry,	comes boldly out with what he has to say
7	l_battle	ustere, he crowns the swaying perch, Flapped by the	angry	flag; The hurricane from the battery sings, But his
8	g_moby	quickest perils; only when within the eddyings of his	angry	flukes; only on the profound unbounded sea, can the
9	a_typee.txt	he most violent disapprobation, manifesting itself in	angry	glances and gestures, and a perfect torrent of passion
10	g_moby	ou reddenest and palest; my heat has melted thee to	anger-	glow. But look ye, Starbuck, what is said in
11	c_mardi.txt	to be sure! At times, brilliantly changeful as opal; in	anger,	glowing like steel at white heat. Belisarius, be it
12	i_piazza	on rippled him. Had there been the least uneasiness,	anger,	impatience or impertinence in his manner; in other we
13	p_billy	mptorily broke in the superior, his face altering with	anger,	instinctively divining the ship that the other was abou
14	e_white	ds swung round; a little confusion ensued; and, with	anger	on his brow, Captain Claret came forward to see
15	d_redburn	felt very much like a fool. But my being so	angry	prevented me from feeling foolish, which is very lucky
16	i_piazza	Men have committed murder for jealousy's sake, and	anger'	s sake, and hatred's sake, and selfishness' sake,
17	c_mardi.txt	shings, all that day was overcast. We sailed upon an	angry	sea, beneath an angry sky. Deep scowled on deep;
18	a_typee.txt	in our teeth, and it was one of those chopping	angry	seas in which it is so difficult to row.
Search Query V Words Case CRegex Results Set All hits Context Size 10 token(s) C				
\banger \bangry Start Adv Search				

Fig. 9: A corpus search for "anger" and "angry" in the AntConc software (note that the search query "\banger\bangry" is a regular expression search that explicitly finds words that begin with "anger" or "angry").

		AntCor	nc	
		KWIC Plot File Cluster	N-Gram Collocate	Word Keyword Wordcloud
ot	al Hits: 9 Pa	age Size 100 hits 😧 🔇 1 to 9 of 9 hits	Θ	
	File	Left Context	Hit	Right Context
L	o_billy	f being buried alive, and in the first struggle agains	st suffocation.	Though at the time Captain Vere was quite ignorant
2	d_redburn	aves were stowed, heel and point, like logs, and th	e suffocated	and dead were unmanacled, and weeded out from th
3	g_moby	in earthquake beneath us. The whole crew were ha	If suffocated	as they were tossed helter-skelter into the white
1	e_white	urgeon. Cuticle, almost dead, was dragged from th	e suffocating	atmosphere, and it was several days ere he complete
5	n_pierre.txt	in the Inferno, and all the rolling gloom in Hamle	t suffocate	him at once in flame and smoke. The cheeks
5	n_pierre.txt	on reasonably and say—Tell me, Pierre, does not th	e suffocating	in thee of poor Bell's holy claims, does
1	n_pierre.txt	t immortal sadness from it. She seemed as dead; a	s suffocated,—	the death that leaves most unimpaired the latent trar
3	d_redburn	¹ unshaven men, smoking tea-leaves, and creating	a suffocating	vapor. But this vapor was better than the native
, ,	e_white	rberating din, and my eyes and nostrils were almost	t suffocated	with the smoke, and when I saw this grim
iea suf	r ch Query 😡 focat*	V Words □ Case ✔ Regex Results Set All hits	Conte	xt Size 10 token(s) 0
or	Options	Sort to right 😌 Sort 1 1R 😌 Sort 2	2R Sort 3	3R C Order by freq C

Fig. 10. A corpus search for various forms of "suffocate" in the AntConc software.



Fig. 11. Ted Nelson's schematic for hypertext in "New Media and Creativity Systems (1965)," Ted Nelson Archive, Internet Archive. https://archive.org/details/nmcs_20210811/page/n5/mode/2up. "anger" and "angry" appearing 26 times, and "suffocate," "suffocating," and "suffocation" appearing 9 times, in Melville's corpus). Data analysis becomes a generative process of new associative thinking about Melville's process. Each locus of editorial attention in *Billy Budd*'s encoded text can be the basis of data analysis across Melville's corpus.

Data stories; or, how to get from numbers to meaning?

• he encoding and analysis of *Billy Budd* raises an important question: What stories can be told with data that are rich with symbolic representations? As the cosmopolitan jests in The Confidence-Man, "The data which life furnishes, towards forming a true estimate of any being, are as insufficient to that end as in geometry one side given would be to determine the triangle" (194). The question of significance and meaning, of interpretation and understanding data, has never been a trivial one, as the cosmopolitan here rightfully seems to argue. In our increasingly data-driven world, our ability to make sense of data is a new skill and even a new form of literacy to be taught at institutions of higher education around the world. However, how can we be critical toward that data? Can computational analyses of Melville teach us anything that we did not know previously? Ryan Heuser and Long Le-Khag have posed the question, "how do we get from Numbers to meaning?" They add, "When all text encoding work is done, all calculations made and all computation executed, how do we move from this kind of evidence and object to qualitative arguments and insights about humanistic subjects"? (46). They suggest that we distinguish between signal and concept. A signal can be any textual feature that can be analyzed computationally-anything from term frequencies, word or sentence lengths, stylometric features, or sentiment scores (as shown above). Similar to Braithwaite's theory, a concept is what we as humanistic scholars "take a signal to stand for, or the phenomenon we take the signal to reveal" (47). To be sure, arguments are the building-blocks of the narratives we tell ourselves about the world and about other narratives, concepts, and arguments that lie at the core of humanistic inquiry. But less has been said about the role of signals in humanistic scholarship (except perhaps in new media studies via information theory), and particularly the ways that they can furnish new kinds of arguments that would not have been possible without computation. "Few," as Heuser and Le-Khag rightly posit, "would be interested in a key, overlooked difference in the term frequencies of the 50 most frequent words between two authors, but if, instead, we found a key, overlooked difference in authorial style, more ears would probably perk up" (47). Heuser and

Le-Khag propose that scholars get even more data in order to harvest more signals to bridge the gap from signal to concept. In moving toward ever greater distance from a close reading, they suggest "what may be needed in fact is more numbers" (48).

Now, while the recent breakthroughs in LLMs seem to confirm that more data produces more powerful systems of text prediction and generation, these advances in "natural language processing" do not necessarily translate into "natural language understanding." Even more so, the move towards ever larger data sets raises the risk of "incurring documentation debts"¹⁰ as well as increased resources and energy consumption.¹¹ Size and quantity alone will not suffice to make digital tools and their output more explainable, whether that output would be text, numbers, or in any other modality. What then remains in bringing the power of computation down to the layers of human understanding, communication, and understanding? Our data-driven approach to reading Melville with computers is neither predicated on big data nor on distant reading procedures, but rather on an iterative, pragmatic approach that combines close and corpus-based forms of reading.

As MMO, MEL, and MPCO demonstrate, digital scholarship has come a long way to provide us with textual data and digital facsimiles of Melville's work and collection of books and prints. But, as Gavin has asked, what exactly are these new textual things (corpora) and what kinds of forms of criticism will they facilitate? As was shown in the example of Melville's revisions in *Billy Budd*, using a combination of digital tools and resources to investigate the revision process can facilitate a more detailed understanding of fluid texts. However, as indicated earlier, the Syuzhet Package was inspired by what Kurt Vonnegut proposed to call the "shape of stories." Following Vonnegut, Jockers implemented the Syuzhet Package based on the finding that changes in the emotional valence (sentiment) of a text does seem to correspond with plot movement. It was Vonnegut himself who held that there is "no reason why the simple shapes of stories cannot be fed into computers."¹²

Plotting the shape of stories alongside the editorial fluidity of the text already demonstrated the many potentials of computationally enhanced reading of literary texts. In what is to follow, we want to return to sentiment analysis to draw yet another connection between data and narrative, by relating another data story about Melville. As indicated in the previous section, the Syuzhet package performs a vectorized look-up in a variety of standard dictionaries. Among those is the "NRC Word-Emotion Association Lexicon" (in short, EmoLex). EmoLex works by assigning collectively attributed sentiment values to eight emotional categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust (see also Plutchik). Applied to *Billy Budd*, the NRC-lexicon, as implemented in Syuzhet, allows us to examine the occurrence of words that can be associated with the eight basic emotions mentioned above (Fig. 12).

This analysis shows a significant and continuous rise in words associated with fear in the novel and a smaller but noticeable increase in words of sadness from roughly the second half of the novel. The initial spike in trust-associated words in the plotted values deserves closer scrutiny. The Syuzhet package tokenizes the novel into individual sentences (1166 in total) and matches the sentiment score of every word relative to eight emotional categories. The 215th sentence in the novel here scores the highest number in trust-words (7 points) and registers fairly high on the anticipation scale (5 points).

How such a designation happened to fall upon one who whatever his sterling qualities was without any brilliant ones was in this wise: A favorite kinsman, Lord Denton, a free-hearted fellow, had been the first to meet and congratulate him upon his return to England from his West Indian cruise; and but the day previous turning over a copy of Andrew Marvell's poems had lighted, not for the first time however, upon the lines entitled Appleton House, the name of one of the seats of their common ancestor, a hero in the German wars of the seventeenth century, in which poem occur the lines, This 'tis to have been from the first In a domestic heaven nursed, Under the discipline severe Of Fairfax and the starry Vere.¹³

This sentence from chapter 6 of the novel, in which Captain Vere is introduced, displays an interesting crescendo of building Vere's image as an intelligent, well-read, disciplined, and cultured man with a strong sense of morality, duty, and responsibility—in short, a man of trust and integrity. It is also noteworthy that this sentence also has an explicit allusion to another author important to Melville, Marvell (allusion as trust?). What stands out from the chart of emotions above is the continuous decline in the emotion of joy as we follow the tragedy of the Handsome Sailor. With fear steadily increasing as the narrative proceeds through the novella, the most fearsome sentence according to the NRC-lexicon coincides with the key scene in which Billy is first confronted with the charge of mutiny in Chapter 19, and, as already mentioned above, it is the sentence with the most negative sentiment value in the novel.

Two stories stand out from these experiments. First, Melville did more than complicate the character of Vere as he revised Billy Budd; he changed the overall sentiment and thereby "the shape of the story" significantly to a darker (more negative) end. Secondly, *Billy Budd* reveals a much more nuanced





'Emotional Prevalence in 'Billy-Budd' read with NRC-Dictionary (EmoLex)'

philosophical and ethical-judicial picture of trust and judgment. The prevalence of words associated with anticipation remains undecided until the very end of the novel, when fear and anger reach their peak and trust plummets to its lowest level. It is especially the steep decline in trust-related words toward the end of the novel, associated with a rise of sadness and fear, that Billy Budd leaves us bewildered with Vere's regret of Billy's execution, despite his moral intuitions. Having followed the naval-codes upon which much of captain Vere's trust and trustworthiness seem to have rested, the emotional "shape of the story" of Billy Budd ends in a dissolution of hope and optimism as well as an almost ontological distrust in the ordering principles from which trusting relationships emerge. It is Melville's own fundamental undecidedness about the philosophical underpinnings of justice and injustice in Billy's death sentence that dictates the emotional shape of Billy Budd. The irresolvable conflict between the strictness of the law and the need for exceptions seems to leave little room for trust towards the end of Billy Budd. Seen from this perspective, Billy Budd continues and concludes Melville's intensive studies of the fabric of trust and confidence undertaken by The Confidence-Man and earlier novels-a perspective that is certainly worthy of further analysis. And we really want to stress further analysis, for the nature of this new methodology of computational and corpus-based research is to experiment with the quantitative results and pursue other creative and critical modes of interpreting them.

Dictionary-Based Analysis and Rule-Based Approaches to AI

espite the famous—or perhaps infamous¹⁴—history of the term "distant reading," as coined by Franco Moretti, using computers to read literary texts is not inextricably connected to a stance of epistemological detachment. As Martin Eve and others have shown, close reading with computers yields many promising extensions of humanities methods. Digital Humanities may often start with counting words, but it rarely ends there. In current developments of computational tools for Natural Language Processing (NLP), Artificial Intelligence (AI), but also Psychology, we find an interesting disjuncture that relates to the difference in distant and close ways of reading in a rarely acknowledged way. In light of recent breakthroughs and media hype of generative AI tools such as ChatGPT, the cognitive scientist Gary Marcus has repeatedly made the point that research in the field of AI has been focusing extensively on methods of Machine Learning¹⁵ in which enormous computational power can train mathematical models of neural networks on vast and "unfathomable" amounts of data, as Emily Bender has also emphasized.¹⁶ This truly "distant" form of reading the texts of the world has, owing to recent advances in computational design and the availability of ever greater supercomputers, created a strong and problematic bias toward solving more and more problems in AI and related fields with deep learning to the disadvantage of "rule-based" approaches.¹⁷ These rule-based approaches are—despite many problems of their own-much more akin to models of close reading developed and cultivated in the humanities. An interesting example of a rule-based system is the Syuzhet Package. As described above, the package divides text into distinct chunks and compares, following strict and well-defined rules, each chunk (words or sentences) with existing sentiment dictionaries. Those dictionaries, such as the one used above, are previously created lists of words rated by a host of different human readers according to a quantified sentiment value. An ethical caveat has to be raised here. The NRC-Emotion Lexicon was put together in a collaborative effort brought about by a hired workforce on the crowd-sourcing platform Amazon Mechanical Turk. A host of ethical problems arise from this. First, the modes of production of precarious labor shall not be glossed over here. Outsourcing the important work of data curation of any digital projects is ultimately damaging the profession. Despite the fact that the NRC lexicon did not explicitly start as a Digital Humanities project we want to acknowledge the time and labor of those many invisible people (see Mohammad and Boyles for more general discussions of Digital Humanities and precarious labor).

The reason why we bring up the dictionary or rule-based approach is two-fold. First of all, rule-based approaches have advantages and disadvantages, but the fact that they currently seem to lose the race with Machine Learning methodologies may have to do with the fact that most rule-based systems struggle to deal with a whole array of linguistic phenomena that are germane to literary writing—ambiguity, irony, double negations, contextual references (long-range dependencies in texts), and denotation, for example. However, taking into account the amount of labor, time, and resources it requires to train LLMs such as GPT or BERT, they turn out to be remarkably underwhelming compared to rule-based approaches. As Elkins remarks:

When Jon Chun and I began working with the latest transformer models our natural assumption was the newer AI models would work best. This was probably partially due to our work with GPT-2 and GPT-3 (Elkins and Chun, 2020), which revealed that, for language generation, transformer models far outperform simpler language models. Surprisingly, state-of-the-art models can often struggle with common text as can be seen when different transformer models like BERT, RoBERTa, and T5 produce more disagreements than simpler lexical models like Syuzhet, Bing, and NRC (Chun 2021). Like a human foreign language learner, AI transformers rely on a more sophisticated level of interpretation that can sometimes be woefully wrong. (86)

Using resource-intensives models such as GPT or BERT, she holds (and it is worthwhile to note here that her experiments with sentiment analysis and plot-detection were carried out long before the release of ChatGPT) that despite their at times uncanny performance in text creation, LLMs are not particularly useful in the contexts of digital literary studies.

Melville in the Age of Artificial Intelligence

et's again consider the opening epigraph generated by ChatGPT. It illustrates a few instructive problems: for one, LLMs are better at producing ■ what Harry Frankfurt called "bullshit," which is to say plausible and apparently valid text that is hollow, dubious, and, more often than not, not truthful. The fact that the opening it produced sounds a lot like an academic essay should ask us to consider whether our own mode of writing should be less predictable. Perhaps a silver lining in this tool is that it could push our students and colleagues to sound less programmatic and more creative in their approaches to essay writing. Given the fact that LLMs like GPT are in a substantial way built and "trained" on gigantic swaths of human-created text-texts that range from Shakespeare to instruction manuals of washing machineshow can we be original in the face of not only an incomprehensible collection of human textuality via the Internet, but also its machine-driven synthesis available at any time? While the discussions of what LLMs will eventually mean for higher education in general have just begun,¹⁸ we will discover that a world of artificial intelligence will be a world in which truthfulness and authenticity are even more difficult to proclaim. One possible remedy might be to use such technology as a means of analysis and reading that ties texts back to reliable and repeatable findings instead of deracinated discourse alone. Such modes of analysis enable scholars to ask different and novel research questions, and to find hitherto unexamined primary sources or to see material in a different light. But a more crucial point is that quantification should be brought back into the fold of literary studies. Quantification and statistical approaches to text make arguments harder, but in a fruitful way. Our purpose here is to examine the affordances and limitations that computational methods bring to the critical discourses on Melville and literary studies. A tool like ChatGPT will certainly impact the way we organize and conduct teaching and research. What Bender et al. call the fundamental "unfathomability" of the deep neural architectures upon which most current AI systems and models such as GPT rely will continue

to be problematic, but they are humanistic problems, as they concern the nature of discourse and the cultural biases and negotiations that shape our linguistic reality. Given the unbelievable amount of text and engineering required to train GPT, and given the fact that its sheer size and complexity confounds even its own creators, we might ask ourselves, "How can that be trustworthy that teaches distrust?" (*Confidence-Man* 243, 251). While certainly "something further will follow from this masquerade!" these technological developments only highlight the importance of the creative and critical potentials of "smart" data that are grounded in established humanistic methods.

Notes

¹The text was the result of two prompts we gave to the Chat bot, asking it to write the opening of an essay on Herman Melville and Digital Humanities, in an academic style, with a critical stance toward technology.

²Emily Bender et al. have critically compared the technology behind Large Language Models (LLM) systems such as ChatGPT with "stochastic parrots" that, devoid of anything such as a human understanding of text, can nonetheless generate a sequence of words based on predictions of statistical word occurrences. See https://doi.org/10.1145/3442188.3445922.

³See John Bryant's "Access and Affiliation" in this issue for more on the history of MEL's technological development.

⁺For more information on the Text Encoding Initiative, the most widely accepted standard for encoding humanities texts, see https://tei-c.org/.

⁵ Tokenization is one of the fundamental steps of undertaking any text analysis operation. The process of tokenizing is to segment unstructured text data into chunks of information that can be considered as discrete elements (characters, words, ngrams, sentences, and so on). Those tokens can then be vectorized (put into numbered lists) in such a way to facilitate numerical representations and statistical analyses of the text data.

⁶ To clarify, adding negation words does not necessarily entail a more "negative" story (litotes, for example, do mean doubly negative, but they are still a form of understatement that change the sense of what would otherwise be a direct phrase). As we explain, all words come with emotional valence values that affect the overall shape of the narrative.

⁷For the documentation of the Syuzhet package, see https://cran.r-project.org/web/packages/ syuzhet/vignettes/syuzhet-vignette.html.

⁸ For more on the NRC dictionary, see http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.

⁹ See Bryant 2021, Ch. 16, "A scar that the air of paradise cannot erase" (179–80), and Ch. 40, "An Apology for Adolescence" (475–76), for a discussion of "fictional strangulation" and adolescent anger in *Redburn*.

¹⁰ See Bender et al. who, long before the first release of ChatGPT, made the point that the trend towards larger and larger data sets to train language models opens a host of severe ethical, environmental, and political problems.

¹¹ Digital Humanists have begun to advocate for awareness of the environmental impact of digitization and other digital tools. See, for example, the Digital Humanities Climate Coalition's Green Digital Humanities Toolkit: https://sas-dhrh.github.io/dhcc-toolkit/.

¹² Vonnegut first expressed this idea in a Master's thesis submitted to the University of Chicago. The thesis was rejected, but he revisited the idea in a lecture that is now available on YouTube. https://www.youtube.com/watch?v=oP3c1h8v2ZQ.

¹³ See the MEL text of Chapter 6: https://melville.electroniclibrary.org/editions/versions-ofbilly-budd/chapter-6.

¹⁴On the contested story of the concept and Franco Moretti, see Klein.

¹⁵See Marcus. https://www.nytimes.com/2023/01/06/opinion/ezra-klein-podcast-gary-marcus.

html.

¹⁶ See Bender et. al.

¹⁷ See Marcus. ¹⁸ See Herman.

See Herman

Works Cited

- Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 2021. 610–623. https://doi.org/10.1145/3442188.3445922
- Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History," *Modern Language Quarterly*, vol. 78. 1 (2017): 77–106.

. "Why You Can't Model Away Bias," Modern Language Quarterly, 81.1 (2020): 95–124.

Boyles, Christina, et al. "Precarious Labor and the Digital Humanities." *American Quarterly*. 70.3 (2018):693–700. Web.

Braithwaite, R. B. Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science. Cambridge UP, 1968.

Bryant, John. "Access and Affiliation." Leviathan 25.2 (2023): 9-34.

- ———. The Fluid Text: A Theory of Revision and Editing for Book and Screen. U of Michigan P, 2002.
 ———. Herman Melville: A Half Known Life. 2 Vols. Wiley-Blackwell, 2021.
- Cordell, Ryan. "Q i-jtb the Raven': Taking Dirty OCR Seriously." *Book History*. 20 (2017): 188–225. doi:10.1353/bh.2017.0006.
- Cummings, James. "The Text Encoding Initiative and the Study of Literature." In A Companion to Digital Literary Studies. Ed. Ray Siemens and Susan Schreibman. New York: Wiley-Blackwell, 2013. 451–76.
- Elkins, Katherine. The Shapes of Stories: Sentiment Analysis for Narrative. Cambridge UP, 2022.
- Eve, Martin Paul. Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas. Stanford UP, 2019.
- Frankfurt, Harry G. On Bullshit. Princeton UP, 2005.
- Gavin, Michael. Literary Mathematics : Quantitative Theory for Textual Studies. Stanford UP, 2023.
- Heuser, Ryan, and Le-Khag, Long. "A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method." Stanford Literary Lab Pamphlet 4. https:// litlab.stanford.edu/LiteraryLabPamphlet4.pdf
- Herman, Daniel. "The End of High-School English." *The Atlantic*, Dec. 9, 2022. https://www.the atlantic.com/technology/archive/2022/12/openai-chatgpt-writing-high-school-english-essay/ 672412/

Kirschenbaum, Matthew. "Prepare for the Textpocalypse." *The Atlantic*, March 8, 2023. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/.

- Klein, Lauren. "Distant Reading After Moretti." Lauren F. Klein. Web. May 9, 2018. https://lklein. com/digital-humanities/distant-reading-after-moretti/
- Marcus, Gary. Interview with Ezra Klein. "A Skeptical Take on the A.I. Revolution—What if ChatGPT isn't as intelligent as it seems?" *New York Times*, Jan. 6, 2023. https://www.nytimes.com//01/06/opinion/ezra-klein-podcast-gary-marcus.html
- Melville, Herman. *Redburn: His First Voyage.* Ed. Harrison Hayford, Hershel Parker and G. Thomas Tanselle. Evanston: Northwestern UP and the Newberry Library, 1969.
 - . Typee : a Peep at Polynesian Life. New York: Library of America, 1982.
- ———. Pierre: or The Ambiguities. Ed. Harrison Hayford, Hershel Parker and G. Thomas Tanselle. Evanston: Northwestern UP and the Newberry Library, 1971.
- . The Confidence-Man: His Masquerade. New York: W.W. Norton, 2006.
- ———. Versions of Billy-Budd, Sailor (An Inside Narrative), ed. John Bryant, Wyn Kelley, and Christopher Ohge. Melville Electronic Library. https://melville.electroniclibrary.org/versions-ofbilly-budd.

Miller, Chris. Chip War: The Fight for the World's Most Critical Technology. New York: Scribner, 2022.

Mohammad, Saif M. "Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis." arXiv.org, 2022.

Moretti, Franco. "Conjectures on World Literature." New Left Review 1. 1 (2000): 54-68.

- Nelson, Theodor Holm. "New Media and Creativity Systems (1965)," Ted Nelson Archive, Internet Archive. https://archive.org/details/nmcs_20210811/page/n5/mode/2up.
- Ohge, Christopher. "Melville's Late Reading and the Revisions in the *Billy Budd* Manuscript." In *Critical Insights*: Billy Budd, Sailor. Ed. Brian Yothers. Ipswich, MA: Salem, 2017. 93–111.

—. Publishing Scholarly Editions: Archives, Computing, and Experience. Cambridge UP, 2021.

Ohge, Christopher, and Wyn Kelley, eds. A New Companion to Herman Melville. Wiley-Blackwell, 2022.

Parker, Hershel. Reading Billy Budd. Chicago and Evanston: Northwestern UP, 1990.

- Plutchik, Robert. *Emotion: Theory, Research, and Experience*. Theories of Emotion, vol. 1, New York: Academic, 1980.
- Poincaré, Henri. *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method.* New York: UP of America, 1913.

Posner, Miriam. "See No Evil." Logic, April 1, 2018. https://logicmag.io/scale/see-no-evil/

- Rogers, Yvonne, and Brignull, Yvonne. "Computational Offloading: Supporting Distributed Team Working Through Visually Augmenting Verbal Communication." *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 25, no. 25, 2003. https://escholarship.org/uc/ item/0945s427.
- Schöch, Christof. "Big? Smart? Clean? Messy? Data in the Humanities." Journal of the Digital Humanities, vol. 2, no. 3, 2013, pp. 2–13.
- Singer, Kate. "Digital Close Reading: TEI for Teaching Poetic Vocabularies." The Journal of Interactive Technology and Pedagogy 3 (2013). https://jitp.commons.gc.cuny.edu/digital-close-reading-teifor-teaching-poetic-vocabularies/

Vaswani, Ashish et al. "Attention Is All You Need." (2017): n.p. https://arxiv.org/abs/1706.03762.

- Vonnegut, Kurt, and Daniel Simon A Man without a Country. New York: Bloomsbury, 2014.
- Wenke, John. "Melville's 'Ragged Edges': Billy Budd, Sailor and the Arts of Incompletion." In A New Companion to Herman Melville, ed. Christopher Ohge and Wyn Kelley. New York: Wiley-Blackwell, 2022, pp. 184–96.

Yothers, Brian, ed. Critical Insights: Billy Budd, Sailor. Ipswich, MA: Salem, 2017.