Why I'm Bringing AI Annotation into My American Studies Course

— Jens Kjeldgaard-Christiansen, Aarhus University (November 9, 2026)

AI has often felt like a negative and disruptive influence on my teaching, for all the familiar reasons: unreliable assessment, hallucinated references, lifeless prose, and the technology's temptations to certain kinds of intellectual laziness. Yet I am convinced that AI holds real potential for teaching, and that my students will need to learn to use it effectively and critically. I am going to act on this conviction by introducing AI as a central tool for student research in my department's American Studies course ("History, Society, and Culture 2: American History and Society"), in the spring semester of 2027. In this short essay, I want to make the case that this is a good idea.

The course syllabus will be nearly unaffected, and students will be able to approach the exam the same way they have in the past: by choosing a topic and writing a 9–12-page analytical essay, arguing, for example, that Frederick Douglass's 1869 "Composite Nation" speech offers an optimistic vision of American exceptionalism. But they will also be able to take on all of Douglass's published speeches. With **AI-assisted text annotation**, a student can ask whether the hope of 1869 was the rule or the exception across five decades of public oratory, amounting to hundreds of speeches, and then read the answer against the history Douglass lived through: the abolitionist struggle, war, Reconstruction, and its collapse. This kind of study—examining a large corpus of texts one by one, against the same complex criteria—used to be entirely unrealistic for a student's exam paper, but now it is very doable.

AI-assisted text annotation is a form of content analysis. "Annotation" here doesn't mean that the AI model writes notes in the margins of each text; it means that the model puts the same questions to every text—How hopeful is this speech? What does it say about the Constitution?—and records each answer as a label, a score, or a short description. At a very basic level, the method does what you can do right now by dropping a Word or PDF file into ChatGPT, Claude, or some other AI and asking it to analyze or score the text: its positive or negative tone, linguistic complexity, thematic composition, the attitudes it expresses, or something else. An annotation study scales and systematizes this procedure. You carefully control the model's settings and prompt-based instructions, and you ensure that it analyzes each text independently and returns its annotations in a consistent format. You then have it annotate hundreds or thousands of texts in this way so that you can draw conclusions about the corpus as a whole: how a theme rises and falls over time; how one author, genre, or period differs from another;

or whether a pattern you noticed in a handful of texts holds up across many. My students will use a website, the Text Annotation Lab (Kjeldgaard-Christiansen 2026), that I designed as a student-friendly introduction to the method and a workspace for using it.

Of course, annotating in a way that is *reliable*, *valid*, and *meaningful* is easier said than done. Annotations that are reliable come out the same, or nearly so, when the prompt is rerun or given to a different model. Annotations that are valid measure what they are designed to measure, usually gauged by comparison with informed human judgment. Annotations that are meaningful capture what matters: an annotation can be both reliable and valid and still tell you nothing worth knowing about your subject matter. Because their annotations have to be meaningful, students can't use the method convincingly without knowing a good deal about the topic of investigation. That knowledge is what allows them to ask the right questions and determine which instructions to give the model.

Figure 1 sketches a student workflow for AI-assisted text annotation, using a hypothetical study of Civil War soldiers' letters. Its question—what Union and Confederate soldiers believed they were fighting for, or at least what they told their families—is something historians have long debated, and the answer bears on how ordinary Americans understood the war. But one needs to be careful. Many Confederate soldiers wrote of defending home, and historians such as Chandra Manning (2007) have argued that "home" was bound up with slavery. A student who knows that debate will ask sharper questions and read the results more critically.

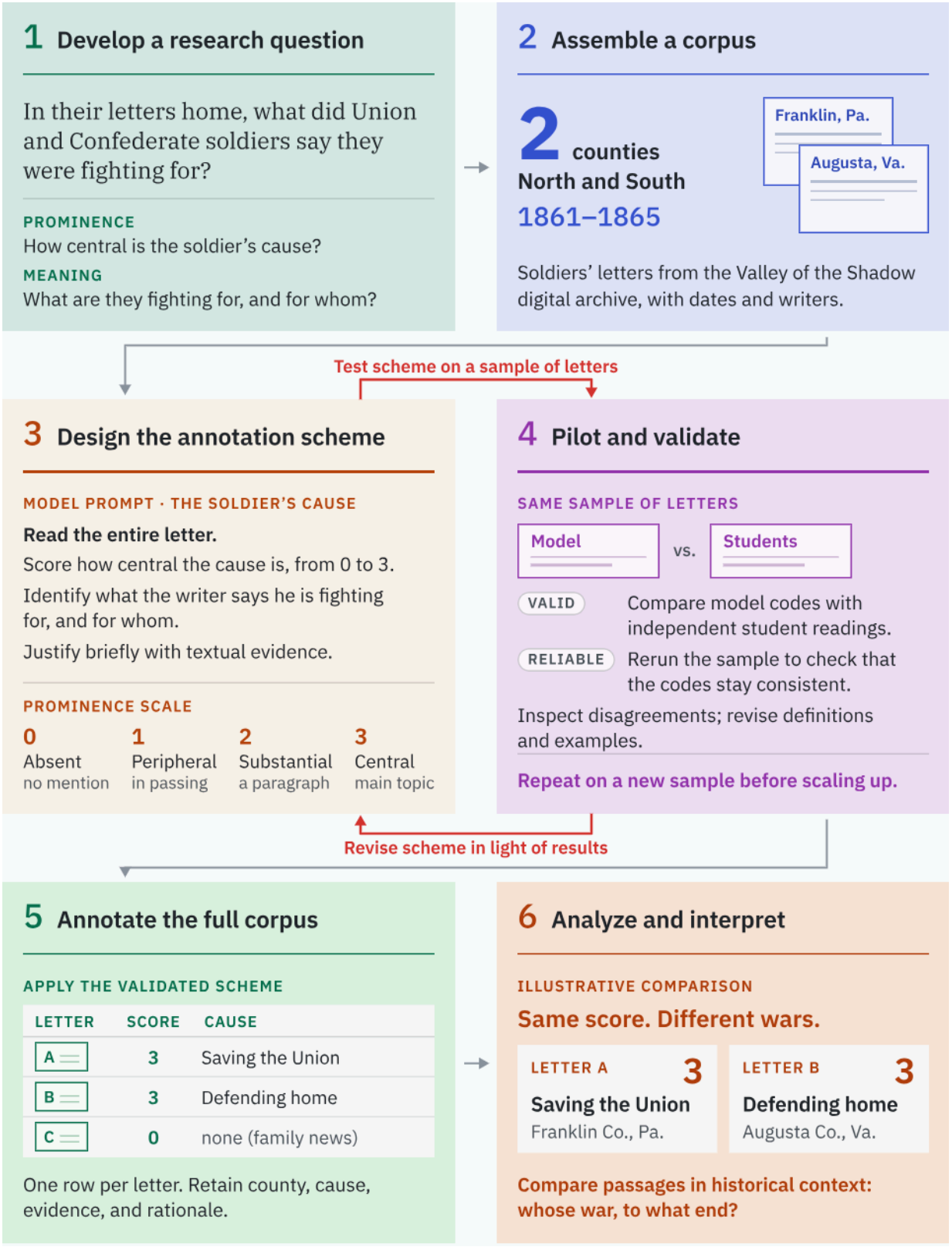
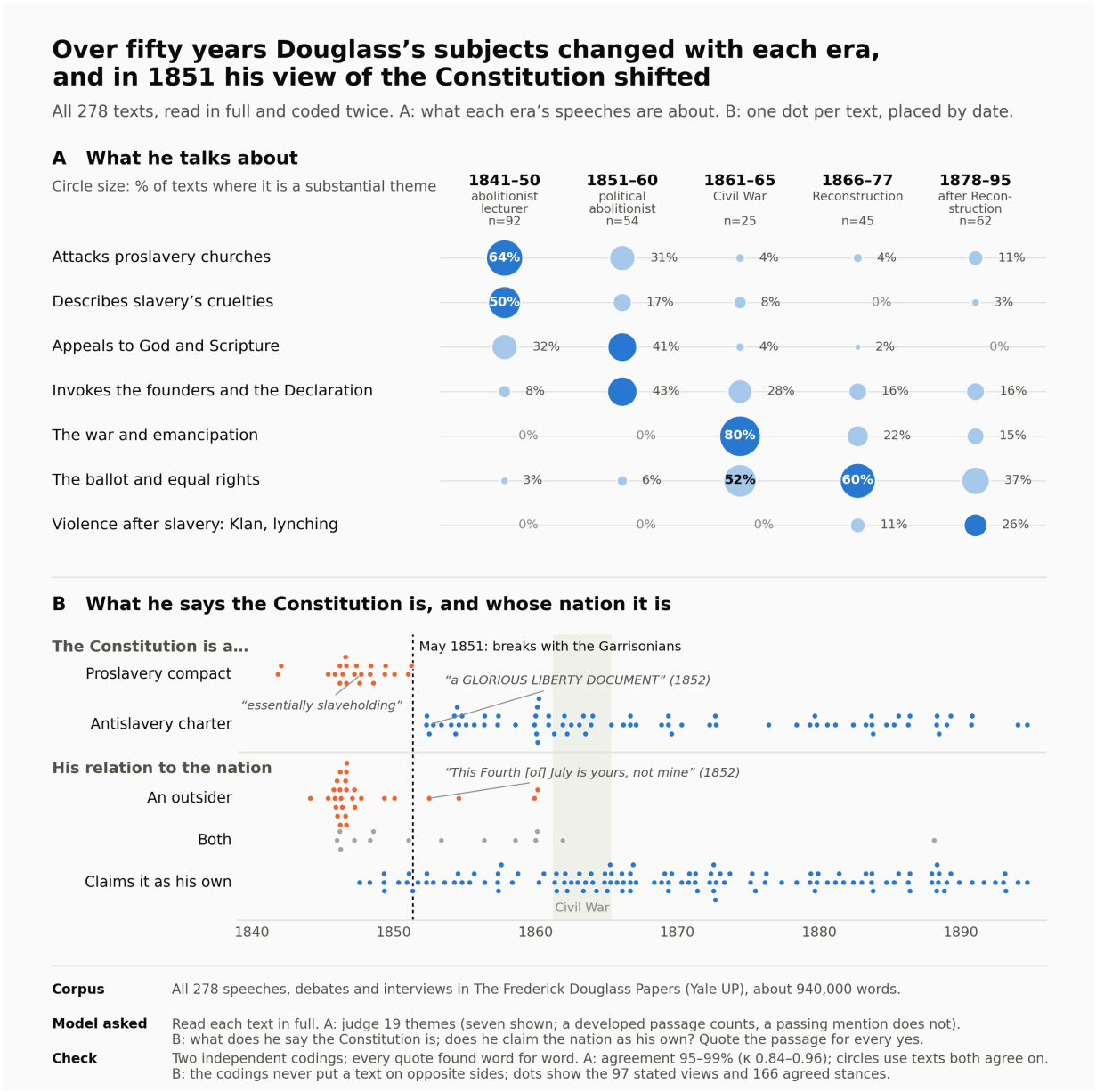
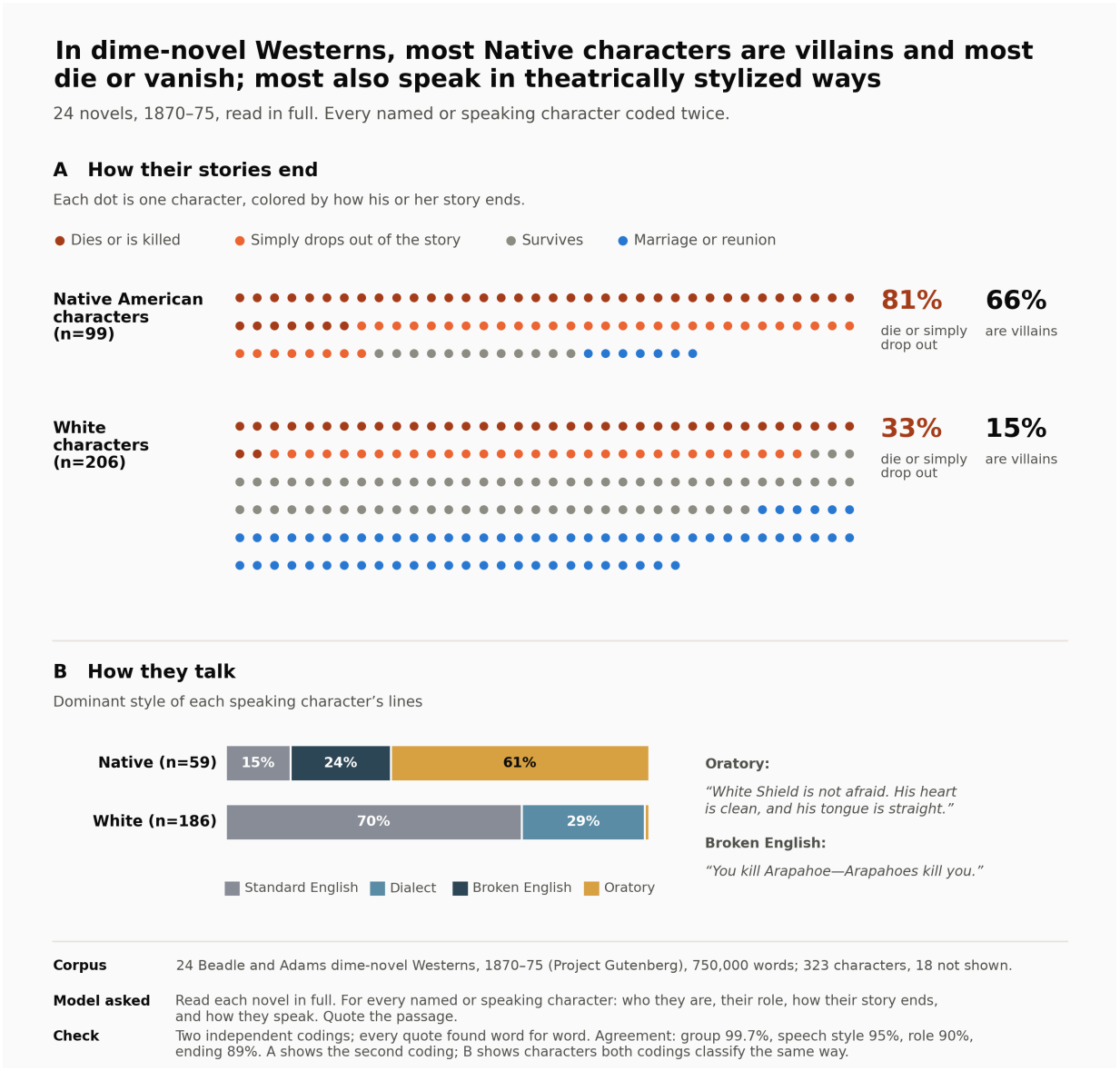


Figure 1. *Example annotation workflow*

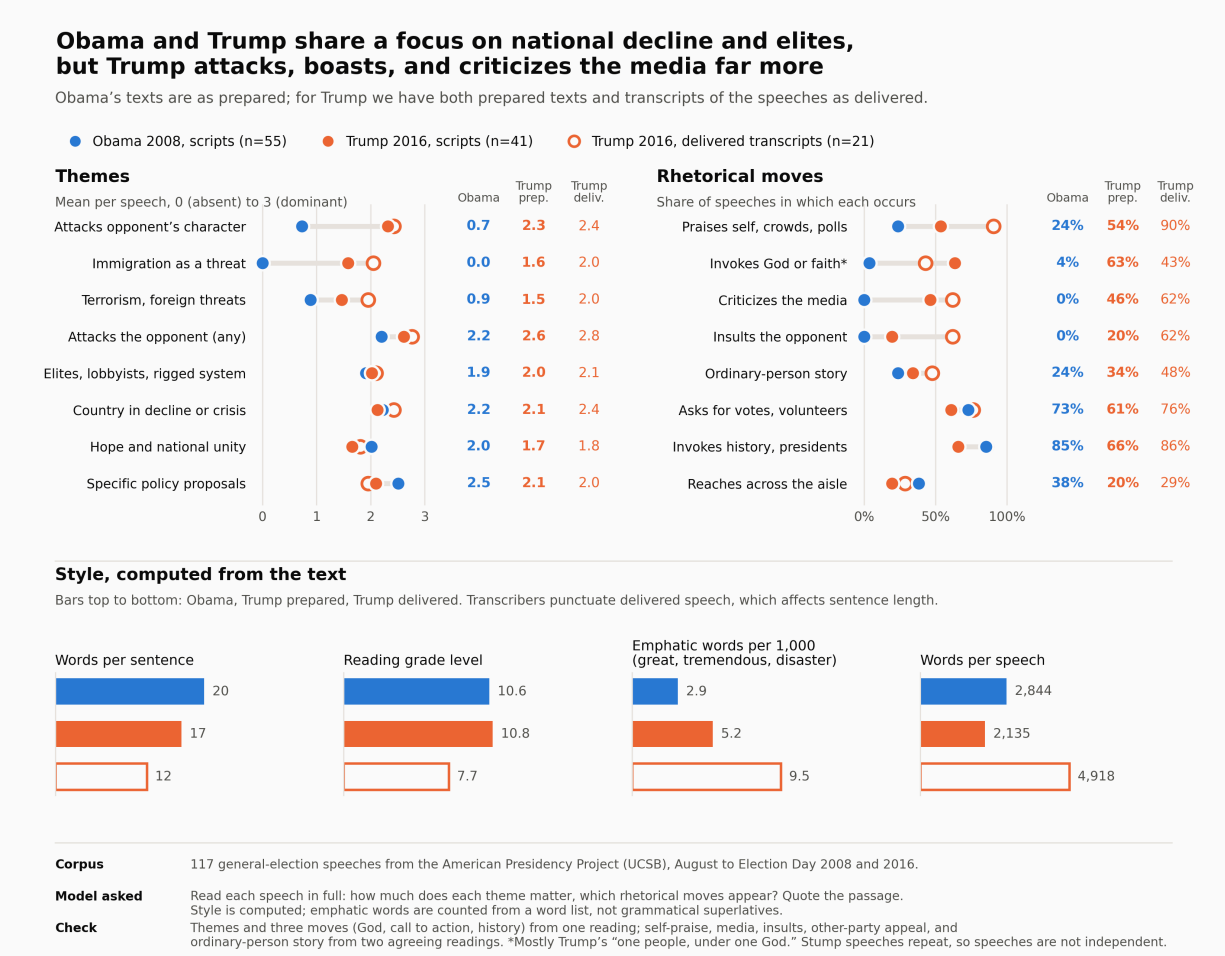
Below are a few examples of the kinds of results one can get from this method. All draw on freely available texts, curated and annotated by me using Anthropic's AI models (mostly Claude Opus 5.5). The annotations were not carefully controlled, and I present the following graphs only as illustrations.



Example 1. *Themes and attitudes toward the Constitution in the speeches of Frederick Douglass*



Example 2. *Depictions of white and Native American characters in dime-novel Westerns*



Example 3. *Themes and style in the campaign speeches of Barack Obama and Donald Trump*

So much for the method and its aims. More information—about prompt design, validation, reliability, and other topics—is available on the Text Annotation Lab and in the various resources listed on it. The website's worked example ("Study example") on the linguistic complexity of presidential inaugural addresses over time should give a good sense of how technically accessible the method can be. The example shows that the linguistic complexity of the speeches has fallen over time, as expected, but it also reveals exceptions to this rule that a student might want to drill down on by combining the method's breadth with close, interpretive reading.

Here are three main reasons why I think this method is worth introducing into my American Studies course.

1. Expanding opportunities for empirical research

I hope that the examples above help to make the first reason plain: you cannot have one student manually analyze and annotate hundreds or even thousands of texts, but you can have them design such a study and let the model take it to scale. It is also cheap. An efficient lower-tier model like OpenAI's GPT-6 Luna can annotate a couple of thousand texts for a dollar.

Importantly, cheap doesn't mean bad. AI annotations can match those of trained human coders, and often outperform crowd workers, in many different contexts (e.g., Heseltine and Clemm von Hohenberg 2024; Piper and Wu 2025). In an early study, for example, Gilardi, Alizadeh, and Kubli (2023) found that, with trained research assistants' coding as the benchmark, a GPT-3.5 model beat crowd workers by about 25 percentage points in accuracy, and that it was more consistent than the assistants themselves. To be sure, the models can be biased, and their output has to be checked. But a realistic comparison should also admit some sobering truths about human coders: we get tired, we lose focus, and we are not exactly free of bias and prejudice. And, of course, the models are getting much more powerful.

Computational "distant reading" methods (e.g., Moretti 2013; Underwood 2019) have long been used to analyze large text corpora based on semantic category distributions and other countable features of texts, with some interesting results. But AI-assisted text annotation is a different beast. The model reads and, in some important sense, understands the texts. Some of the method's proponents go so far as to say that it combines the interpretive insight of "close reading" with the scale of automated "distant reading," producing a kind of "automatic close reading" (for a critical discussion, see Beaufoy and Kjær 2026). Perhaps that's a stretch. But it seems clear to me that AI-assisted text annotation makes possible much richer, more holistic judgments about texts than traditional computational methods could readily provide. As one example of this, I recommend Mitran, Wu, and Piper's (2025) study of nearly 2,700 folktales from around the world, which uses AI-assisted annotation to investigate the moral values expressed through characters' actions. Looking at what characters do rather than counting moral vocabulary, they find that the "positivity bias" earlier research attributed to folktales is overstated: the tilt toward positive portrayal is modest, and many characters are morally ambivalent.

In my view, American Studies should at the very least be curious about this method. For a field that advances broad theses about American society and culture, and that has spent decades widening its archive beyond the canon, the method offers a way to test those theses against far more of that archive.

2. Developing critical AI literacy

As I write this, a two-panel meme circulating online depicts a student being scolded by a professor for using AI, with the second panel showing the same student, a few years older, being rejected at a job interview for having no experience with AI. As a criticism of AI-skeptical professors, that's a bit unfair: students should be discouraged from certain uses of AI that will hurt them academically *and* professionally—most obviously, outsourcing their thinking to it. At the same time, I think it's plain that much of the world that students will enter after their studies will run on AI, and they will need to learn to use it critically. And then there's the ultimate utilitarian yardstick: job prospects. Employers are noticing just how effective AI can be for doing all kinds of work. In the United States, a recent survey by the National Association of Colleges and Employers (2026) found that more than a third of entry-level jobs at surveyed employers already require AI skills. In Denmark, a survey by Dansk Industri

(2026) similarly found that more than a third of companies consider AI skills important when recruiting, while nearly nine in ten believe the education system should prepare students to understand and use AI. I agree: we ought to have much more to say about these technologies than "stay away."

I believe that having students actually use AI-based methods for their assignments cultivates far more critical understanding of the technology's workings and limitations than the "AI warnings" slides I earnestly talk them through before written exams. Critical evaluation of the AI's output is built into AI-assisted text annotation: students must assess how far the model's judgments can be trusted before drawing conclusions from them. And the prompts they use should be reported as a core component of their research, so that their use of AI can be evaluated as a method, not just checked against the rules.

3. Strengthening core humanities skills

I am now on version 35 of a codebook containing the prompts I will be using for a formal annotation study of historically shifting attitudes and emphases in American Studies scholarship. I am trying, again and again, to get my prompts' topical relevance, conceptual distinctions, handling of borderline cases, and scope *just right* so that I can answer my research questions precisely, convincingly, and without leading the silicon witness. On the other side of data collection, I will need to think just as hard about what the numbers, in the form of annotation distributions over time and publication venues, actually mean. The skills needed to meet these kinds of challenges are very much those of the humanist: the literary scholar, the historian, the linguist, the rhetorician, the cultural theorist. You will get nowhere and convince no one if you are not able to think critically about questions like:

- Which texts should I be looking at in the first place?
- What do I even mean by "patriotism," or "democratic optimism," or whatever I'm coding for? And do I mean what my authors meant?
- What should I do with borderline cases: passing mentions, quoted material, genre-specific understatement, ceremonial boilerplate, period vocabulary?
- Is the model reading the way I would? And where it reads differently, is it wrong, or is it showing me something I missed?

For me as a teacher in a department of English, there is a special affinity here to exploit. **The essential analytical tool of AI-assisted annotation is not a programming language or statistical software, but *English***. Writing a prompt that effectively distinguishes "patriotism" from "nationalism" without reading today's meanings into a nineteenth-century speech is an exercise in analytical writing, and doing it well depends on the same attention to nuance, ambiguity, and register that we spend years trying to teach.

There are practical costs to what I'm proposing. Most obviously, students who choose to work with AI-assisted text annotation will have to spend their own money. I plan to tell my students that they should not spend more than the equivalent of five dollars, and I suspect that most will spend far less

than that. The time for teaching students this method also has to come from somewhere, and I don't want the course's core objective—teaching students about American history and society—to suffer as a result. My tentative plan is to dedicate three one-hour workshops to introducing the method through cases drawn from the course's own periods and themes. By the end, students should have a foundation for asking method-appropriate research questions, curating a relevant corpus, writing well-designed prompts, validating the model's answers against their own best judgment, and interpreting the results.

Then there are the broader ethical questions that surround the use of AI, including its environmental footprint, the use of creative and scholarly work as training data, the concentration of power in the hands of a few large providers, and others (see Crawford 2021). These questions are not the subject of the course, and I won't pretend that what time I can dedicate to them will do them justice. But we will be discussing them, and no student will be required to use the method.

With all this in mind, I believe AI-assisted text annotation holds real promise for our students. I will report back on how they receive it.

REFERENCES

Beaufoy, Joanna, and Lars Kjær. 2026. "Can Automatic Reading Be Close? On the Implications of Using a Large Language Model to Identify Interactions between Lighting and Emotions in French Literature, 1838–1928." *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqag090.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press. https://doi.org/10.12987/9780300252392.

Dansk Industri. 2026. "Virksomheder: Uddannelsessystemet skal ruste fremtidens arbejdskraft til kunstig intelligens" [Companies: The education system must prepare tomorrow's workforce for artificial intelligence]. DI Analyse, September 2026. https://www.danskindustri.dk/arkiv/analyser/2026/09/virksomheder-uddannelsessystemet-skal-ruste-fremtidens-arbejdskraft-til-kunstig-intelligens/.

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences* 120 (30): e2305016120. https://doi.org/10.1073/pnas.2305016120.

Heseltine, Michael, and Bernhard Clemm von Hohenberg. 2024. "Large Language Models as a Substitute for Human Experts in Annotating Political Text." *Research & Politics* 11 (1): 20531680241236239. https://doi.org/10.1177/20531680241236239.

Kjeldgaard-Christiansen, Jens. 2026. *Text Annotation Lab*. https://asca.au.dk/annotation-lab.

Manning, Chandra. 2007. *What This Cruel War Was Over: Soldiers, Slavery, and the Civil War*. New York: Alfred A. Knopf.

Mitran, Luca, Sophie Wu, and Andrew Piper. 2025. "Probing Narrative Morals: A New Character-Focused MFT Framework for Use with Large Language Models." In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, edited by Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, 28514–29. Suzhou, China: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.emnlp-main.1449.

Moretti, Franco. 2013. *Distant Reading*. London: Verso. https://www.penguinrandomhouse.com/books/232676/distant-reading-by-franco-moretti/.

National Association of Colleges and Employers. 2026. *Job Outlook 2026 Spring Update*. Bethlehem, PA: National Association of Colleges and Employers. https://www.naceweb.org/docs/default-source/default-document-library/2026/publication/research-report/2026-job-outlook-report-spring-update.pdf.

Piper, Andrew, and Sophie Wu. 2025. “Evaluating Large Language Models for Narrative Topic Labeling.” In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*, edited by Mika Hämäläinen, Emily Öhman, Yuri Bizzoni, So Miyagawa, and Khalid Alnajjar, 281–91. Albuquerque, NM: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.nlp4dh-1.25.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226612973.001.0001.