INFO 290 ▸ CULTURAL ANALYTICS

Final project requirements

KENT K. CHANG DATE: August 19, 2025

TL;DR

- **Final project** (essay/technical paper/hybrid), **40%** of the grade; aim for your *dream venue*.
- Must combine *quantitative evidence* with *qualitative/interpretive* analysis of a broadly defined cultural artifact (from novels to YouTube shorts and LLMs).
- Quizzes, HW 1, HW 2, and Labs scaffold the project (but no requirement to tie HWs directly).

This course culminates in a final project, which is worth 40% of your semester grade. Detailed specifications for the associated assignments (proposal, midterm report, presentation, and final paper) will be posted on bCourses in due course. This document provides additional context and motivation for the project.

1. FINAL PROJECT AND THIS CLASS

Other graded components are designed to help you prepare for your final project:

- *Quizzes* bring awareness to topics that are potentially useful for research, and help you spot gaps in your knowledge, as you read more widely in search of methods and inspirations for your project.
- *Close reading paper* (Homework 1) helps you identify a cultural phenomenon of interest and articulate why it matters.
- *Workshop notebook* (Homework 2) helps you run preliminary method experiments on your data and gain practical coding experience with a research question in mind.
- *Labs* give you space to workshop ideas, encounter new methods, and solicit peer feedback.

It appears as a separate grading component on the syllabus because the trajectory—from the first homework to your final paper—is more ideal than real. You may hit walls along the way and want to switch gears completely. On the flip side, your preliminary exploration might go *too well* and leave you with nothing more to say by HW 2; in that case, you can and should pursue something else. Accordingly, there is no *explicit* requirement that your other homework connect directly to the final project—and that is fine.

2. SCOPE OF THE FINAL PROJECT

As stated on the syllabus, the goal is a project conceived with your *dream* venue in mind. It may take the form of an argumentative essay or a technical paper presenting methods and experiments—or a hybrid of both. In all cases, your project **must synthesize quantitative evidence with qualitative observation** (informed by interpretive practices and situated in the appropriate critical context) about a **cultural artifact** of your choosing.

We adopt a broad definition of *cultural artifact*: any symbolic form—that is, any medium such as a text, image, video, or cultural technology[1]—that conveys and reshapes meaning within social, economic, and institutional contexts. In this sense, a novel, a poem, a YouTube short, or even a large language model can all be studied as artifacts of cultural production. You will use computational methods to shed some light on such artifacts, if not also the contexts from which they emerge. Here are some examples of what you might do:

- propose new approaches for studying cultural artifacts—focused on **methods and exprirments**, e.g., Soni et al., "Grounding Characters and Places in Narrative Text" (2023)[2]—or focused on **data and analysis**, e.g., Lucy et al., "Racial and Ethnic Representation in Literature Taught in US High Schools" (2025)[3]

- advance **original arguments** related to culture (socialties, histories, cognition, the literary) supported by quantitative evidence—e.g., Shechtman, "The Medium Concept" (2020)[4]; Zhou and Bamman, "Once More, with Feeling" (2024)[5]

- offer **informed critiques** of computational methods and their application to cultural artifacts—e.g., Bode, "Why You Can't Model Away Bias" (2020)[6]

- conduct **systematic examination** of models and their behaviors when applied to cultural artifacts—e.g., Walsh, Preus, and Antoniak, "Sonnet or Not, Bot?" (2024)[7]

- design **new tools** to process, retrieve, or organize cultural artifacts—e.g., Arnold and Tilton, "Automated Image Color Mapping for a Historic Photographic Collection" (2024)[8]

The examples above, while certainly not exhaustive, demonstrate productive ways of combining computational methods with the study of culture: some projects may emphasize *original arguments*, some may focus on *methodological innovation*, while others may center on critical reflection and interrogate what is lost or gained when cultural questions are translated into computational terms. You might also choose to examine models themselves as cultural objects, asking what their behavior reveals about language, representation, or bias. Or, you may design new tools and infrastructures that open up novel ways of organizing or accessing cultural data.

1. Henry Farrell et al. (Mar. 2025). "Large AI Models are Cultural and Social Technologies". In: *Science (New York, N.Y.)* 387.6739, pp. 1153–1156.
2. Sandeep Soni et al. (July 2023). "Grounding Characters and Places in Narrative Text". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, pp. 11723–11736.
3. Li Lucy et al. (Mar. 2025). "Racial and Ethnic Representation in Literature Taught in US High Schools". In: *Journal of Cultural Analytics* 10.1.
4. Anna Shechtman (May 2020). "The Medium Concept". In: *Representations* 150.1, pp. 61–90.
5. Naitian Zhou and David Bamman (Nov. 2024). "Once More, with Feeling". In: *arXiv [cs.CL]*.
6. Katherine Bode (2020). "Why You Can't Model Away Bias". In: *Modern Language Quarterly* 81.1, pp. 95–124.
7. Melanie Walsh, Anna Preus, and Maria Antoniak (Nov. 2024). "Sonnet or Not, Bot?" In: *Findings of the Association for Computational Linguistics: EMNLP 2024*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, pp. 15568–15603.
8. Taylor Arnold and Lauren Tilton (Nov. 2024). "Automated Image Color Mapping for a Historic Photographic Collection". In: *arXiv [cs.CV]*.

3. WHAT MAKES A PROJECT SUCCESSFUL

Successful projects will effectively bring computational methods and the study of culture together: they use machine learning models as *measuring devices* to characterize a phenomenon, or to support an argument with rigorous quantitative evidence. Such evidence may take the form of tables or figures, along with appropriate confidence intervals or error analysis. As we will see over the semester, this often involves the following steps:

1. **Flesh out a research question**: What is interesting, why is it interesting, and how has it been studied (qualitatively and quantitatively)? What can I add to this discourse?
2. **Devise a preliminary research plan**: Will qualitative analysis be sufficient? Is quantitative analysis feasible (do I have the data and computing resources)? Why or why not? Which computational tools that I know might speak to this question? What's more pressing and can be adequately justified—develop a new tool or present my arguments? Reframe the research question accordingly.
3. **Figure out how best to apply those tools** to serve your research question: How do I translate (or, *operationalize*) the concept? What's the simplest thing (or, *exploratory analysis*) I can do to see if this is worth pursuing? If I need to do some experiments, what are some reasonable baselines?
4. **Evaluate those tools for this use case**: How do you convince your readers that the evidence you present is robust? Even if it's just counting words, what do I mean by a "word," and what's the story of the data in which that word appears? If you are using LLMs for classification, how do you know your readers can trust the models (i.e., they aren't simply guessing, and there's some form of interpretability study).
5. **Write up the project**, following the formatting conventions specified by the journal or conference you plan to submit to.

To support you through these steps, the *project proposal* (one page maximum), asks you to articulate the intellectual stakes of your project and outline an initial plan for early feedback. The *midterm report* (four pages or a detailed paragraph outline) will emphasize literature review, as well as preliminary results or main arguments. The *presentation* is your opportunity to solicit concrete feedback for revision before the project report is due.

A strong final project will demonstrate:

- originality and significance of the research question or argument
- appropriateness and rigor of the chosen methods
- clarity, coherence, and persuasiveness of the writing and overall presentation

Bringing the qualitative and quantitative together is at the heart of this assignment; even so, the balance of evaluation will depend on the type of project you choose: a *technical* paper will be evaluated primarily on the rigor and transparency of its methods, the care with which results are validated, and the clarity with which those results are communicated. An *argumentative* essay, on the other hand, will be evaluated on the originality and significance of its central claims, as well as the coherence and persuasiveness of the argument as it unfolds across the paper. Hybrid projects should aim to integrate both dimensions in a way that makes each stronger than it would be on its own.

4. NOTES ON THE USE OF GENERATIVE AI

A key learning outcome of this class is to develop an intuitive understanding of how generative AI systems—including large language models—work, and to use them sensibly and responsibly. In general, you are welcome to use generative AI for low-stakes tasks such as brainstorming or surface-level editing (e.g., spelling or phrasing). You do not need to declare this kind of usage. However, you may *not* delegate the intellectual substance of your work—especially the design of methods, the interpretation of results, or the core argument of your project—entirely to LLMs, and this would constitute a misrepresentation of your own original work.[9]

Violations that may lead to a redo or a grade of 0 include, but are not limited to, the following:

- failing to submit an assignment, or submitting work that is missing required components
- any form of plagiarism[10]
- any form of fabrication, such as citing papers that do not exist
- inability to demonstrate or discuss the core ideas, methods, or arguments in your project

Generative AI can be a powerful ally when used thoughtfully, and part of this class is learning how to make it work *for* you rather than *in place of* you. When in doubt, please reach out. Ultimately, we want you to feel confident that the project you submit reflects your own skills, ideas, and growth.

Above all, this class is about learning how to ask meaningful questions, bring computational methods into dialogue with cultural interpretation, and produce work that is both rigorous and original. The final project is your opportunity to push yourself toward the kind of research you'd be proud to present at your dream venue (even if that venue has yet to take institutional shape), and to step a little outside your comfort zone in a supportive, low-stakes environment. Approach it not only as a chance to demonstrate what you've learned, but also to discover how your own ideas can shape the study of literature and culture in the age of data and generative AI.

Key dates of this course:

Quiz 1	20 pts[11]	September 11
Close reading paper	10 pts	September 26
Quiz 2	20 pts	October 2
Project proposal	5 pts	October 3
Workshop notebook	10 pts	October 31
Project midterm report	10 pts	November 7
Project presentation	5 pts	December 4
Project final report	20 pts	December 12

9. For any aspects not fully addressed above, please refer to the NeurIPS 2025 policy: https://neurips.cc/Conferences/2025/LLM.
10. See https://writing.berkeley.edu/academic-honesty.
11. Only one quiz (better of the two) counts toward the semester grade.