

Emory University
Departments of English and Quantitative Theory & Methods

ENG 790 / QTM 490 – Spring 2023
Quantitative Literary Analysis: Theory and Practice
Tuesdays 1-3:45pm, Callaway C201

Instructors

Dr. Lauren Klein (lauren.klein@emory.edu)

Dr. Ben Miller (b.j.miller@emory.edu)

Land Acknowledgment

Emory University is located on Muscogee (Creek) land. Emory was founded in 1836, during a period of sustained oppression, land dispossession, and forced removal of Muscogee (Creek) and Ani'yunwi'ya (Cherokee) peoples from Georgia and the Southeast. Emory owes an immense debt to the Muscogee, Ani'yunwi'ya and other original peoples, and their descendants, who have cared for and inhabited these lands.

Read the full [Land Acknowledgment and History Statement](#) developed by Emory faculty.

Office Hours

Klein: M/W, 11:30am-1pm; Tu, 4-5pm (schedule [here](#))

Miller: T, 10-11:15 (and by appointment)

Course Description

One of the primary strands of digital humanities scholarship is quantitative literary analysis: the use of quantitative (and other digital) methods to study literary texts. These practices raise many major questions: What does large-scale analysis reveal about literature that cannot be discerned by reading alone? What happens when literary texts are converted into numbers, as they must be for any quantitative analysis to take place? What does it mean, both ethically and intellectually, to borrow methods developed in the sciences and/or from industry for literary and cultural studies scholarship? And what of the core concerns of literary and cultural studies? Can ideas about language, metaphor, style, labor, and power, among many others, be quantified, modeled, and/or otherwise explored at scale?

This course will take on these questions in both theory and practice, focusing our inquiry around a current state-of-the-art approach to quantitative literary analysis (and quantitative text analysis more broadly) involving what are known as large language models (LLMs). We will begin by contextualizing the advent of LLMs within the history of language modeling and predictive text generation, primarily in the area of electronic literature, before delving into the conceptual and technical details of LLMs as a class. This shared inquiry will enable us to think through possible projects in the area of quantitative literary studies that might engage the strengths—or, alternately, probe the limits—of LLMs as a method of analysis. The course will

culminate in final projects undertaken in small groups, consisting of students skilled in both literary analysis and quantitative methods, which demonstrate how quantitative analysis can contribute to current scholarly conversations in the field of literary studies, and vice versa.

Required Course Materials

All required readings will be posted on Canvas and/or are available online.

List of Graded Assignments

Your grade for the course will be calculated as follows:

- Attendance and participation - 20%
- “Explain an Explainer” – 10%
- “Explain a Research Paper” – 10%
- Final project proposal – 10%
- Final project presentation – 10%
- Final project – 40%

Description of Graded Assignments

Participation

Participation is often assumed to be a hazy concept, but it actually involves a careful assessment in five distinct areas. Here are short descriptions of each of these areas, adapted from grading criteria developed by Dr. Mark Sample of Davidson College:

- **Preparation:** Reading/reviewing any assigned material before class.
- **Presence:** Being verbally and nonverbally engaged during class.
- **Focus:** Avoiding distractions during class (both in person and online).
- **Asking questions** in class and in office hours, as well as via email when appropriate.
- **Specificity:** Referring to specific ideas from readings and prior class discussions when contributing to class discussion and/or in conversations during office hours.

“Explain a...” assignments

This description applies to both the “Explain an Explainer,” and the “Explain a Research Paper” assignments. With a partner, in 2-3 minutes and a 1/2 to 1-page summary, explain the argument, evidence, shortcomings, and value of an explainer and a research paper. The 2-3 minute talk can be by either person (or both). The summary should be by both people. The written summary can include 1-2 figures from the source.

Examples of explainers are videos or podcasts, such as Dale Merkowitz’ “Transformers, explained: Understand the model behind GPT, BERT, and T5,” (<https://youtu.be/SZorAJ4I-sA>) or posts on blogs, such as “Language Model Scaling Laws and GPT-3,” by Cameron Wolfe (<https://cameronrwolfe.substack.com/p/language-model-scaling-laws-and-gpt>).

Examples of research papers are what one might find in the proceedings of conferences like AAAI, ACL, EMNLP, CoLING, NACCL, FAACT, NeurIPS, CHR, and DH; in preprint venues like arXiv.org; in white papers by the various groups producing these language tools; or in

peer-reviewed journals. If you are having difficulty finding a paper that interests you, please contact us.

Final project proposal

Each project group will be required to submit a written proposal that documents their ideas and initial research for their final project. A detailed assignment sheet for this assignment will be distributed no later than two weeks before the scheduled due date.

Final project presentation

On the final day of class, you will be required to present your project-in-progress to the class. The presentation should be approx. 10 minutes long, and will be followed by questions and comments from the class.

Final Project

In addition to the assignments described above, you will be completing a final project which demonstrates how quantitative analysis can contribute to current scholarly conversations in the field of literary and cultural studies (or the humanities more broadly) or the other way around (i.e. how methods of analysis from literary and cultural studies can contribute to current scholarly conversations around quantitative analysis (or AI/computation more broadly)). You will be working in small groups, and your group will be required to submit a co-authored research paper of approx. 2000 words that documents your work. You will receive a letter grade (A-F) on the basis of your contribution, as well as written feedback.

Grading Process

All assignments will be graded on an A-F scale. At the end of the semester, assignments will be weighted as listed above.

To best utilize our skill-sets as faculty, students will have one consistent professor/mentor throughout the semester. We will determine the mentor assignments after the end of A/D/S. Note that this does not mean that you cannot meet with or solicit feedback from the other faculty member; only that a single professor will be responsible for your grading throughout the course.

Emory Grading Levels

Average Score College Grade LGS Grade

| | | |
|--------------|----|----|
| 93 and above | A | A |
| 90 to 92.99 | A- | A- |
| 87 to 89.99 | B+ | B+ |
| 83 to 86.99 | B | B |
| 80 to 82.99 | B- | B- |
| 77 to 79.99 | C+ | C+ |
| 73 to 76.99 | C | C |
| 70 to 72.99 | C- | C |

| | | |
|-------------|----|---|
| 67 to 69.99 | D+ | F |
| 60 to 66.99 | D | F |
| 0 to 59.99 | F | F |

Grading Rubric

To make sure that everyone is clear about the expectations for written work, below is the rubric for an “A” grade in any written assignment.

Contacting your Professors

Both of your professors can be reached via their Emory email addresses. We respond to email M-F 9am-5pm, and outside of those hours only if our schedules allow. Please allow 24 hours for a response, and 48 hours if your message is sent over the weekend.

Policy on Late/Skipped Assignments

All assignments are mandatory. Should you submit an assignment after the due date, your grade for that assignment will decrease by a 1/3rd letter grade for each day that it is late (e.g. B becomes B-). Should you fail to submit an assignment entirely, you will receive an F on that assignment. Should you need an extension, please contact us 24 hours *in advance* to negotiate the deadline.

Attendance and Punctuality

In ordinary years, we allow two excused absences, no questions asked, with your grade beginning to be lowered with the third absence. However, due to the ongoing pandemic, we do not want to pressure you to come to class if you might be sick. With that said, you are responsible for finding out what was discussed in the course on any days that you miss; we do not provide copies of our lecture notes. In addition, beginning with the third absence, you must email us to let me know that you will be missing class for health reasons. Finally, please be respectful to your colleagues and arrive on time. If you arrive more than 15 minutes late, you will be considered absent for that class.

Office of Accessibility Services

The Office of Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact OAS. It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed. Students registered with OAS who have a letter outlining their academic accommodations, are strongly encouraged to coordinate a meeting time with your professor that will be best for both to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible. Students must renew their accommodation letter every semester they attend classes. Contact the Office of Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at <http://equityandinclusion.emory.edu/access/students/index.html>.

Writing Center and ELL Program

Tutors in the Emory Writing Center and the ELL Program are available to support Emory College students as they work on any type of writing assignment, at any stage of the composing process. Tutors can assist with a range of projects, from traditional papers and presentations to websites and other multimedia projects. Writing Center and ELL tutors take a similar approach as they work with students on concerns including idea development, structure, use of sources, grammar, and word choice. They do not proofread for students. Instead, they discuss strategies and resources students can use as they write, revise, and edit their own work. Students who are non-native speakers of English are welcome to visit either the Writing Center tutors or the ELL tutors. All other students in the college should see Writing Center tutors. Learn more and make an appointment by visiting the websites of the ELL Program and the Writing Center. Please review tutoring policies before your visit. The Writing Center is located in Callaway N111. Their hours are Mon-Thurs 11am-8pm, Fri 11am-5pm, and Sun 12-8pm. Learn more about the Writing Center and make an appointment through the EWC website:

<http://www.writingcenter.emory.edu/>. Please review tutoring policies before your visit. A maximum of 2 appointments are allowed each week. Students need to bring hard copies of drafts to their appointments.

Honor Code

The Honor Code is in effect throughout the semester. By taking this course, you affirm that it is a violation of the code to cheat on exams, to plagiarize, to deviate from the teacher's instructions about collaboration on work that is submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You agree that the instructor is entitled to move you to another seat during examinations, without explanation. You also affirm that if you witness others violating the code you have a duty to report them to the honor council. Students who violate the Honor Code may be subject to a written mark on their record, failure of the course, suspension, permanent dismissal, or a combination of these and other sanctions. The Honor Code may be reviewed at:

<http://catalog.college.emory.edu/academic/policies-regulations/honor-code.html>.

A Closing Note on COVID

By the time this course begins, it will have been nearly three years since the pandemic began. As your professors, we are operating under the assumption that we have all experienced our share of hardships, and that we are all—in different ways and to different degrees--exhausted. Our goal is to make this course useful and interesting, and not to contribute to our collective (and ongoing) struggles. If you are having trouble related to the course in any way, *please let us know*. By the same token, we will all need to adapt as the Covid situation continues to evolve. We ask for your flexibility and understanding should the format of the course or the assignments need to change in response to new or unexpected developments.

Class-by-Class Schedule

Class schedule subject to change.

Please consult Canvas for the most current class schedule.

Current versions of all class labs can also be found on [GitHub](#).

Tues, January 10th – Course Introduction and Opening Examples

What is a language model and what do they look like when you encounter them online?

In class:

- Introductions and Syllabus Overview
- Explore “[Sea and Spar Between](#)”
- Mini-lecture / discussion on language models
- Explore [AI Dungeon](#)
- Mini-lecture on / discussion AI Dungeon as parable of the internet

Tues, January 17th – Large Language Models

What do language models look like right now and what should humanities scholars do with and/or about them?

Readings:

- Kevin Roose, “[The Brilliance and Weirdness of ChatGPT](#),” *The New York Times*, December 5, 2022.
- Emily Bender, Timnit Gebru, et al., “[On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?](#)” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (March 2021): 610-623.
- Marika Cifor, Patricia Garcia, et al., “[The Feminist Data Manifest-No](#)” (2019)
- Plato, from the *Phaedrus*, [247c-275b](#) (Canvas).
- Jacques Derrida, “Plato’s Pharmacy,” part 1, from *Dissemination*, trans. Barbara Johnson (Univ. of Chicago Press, 1981) (Canvas) ← *just get as far through / in as you can; we’ll work through this in class as well*

Explore:

- ChatGPT: <https://chat.openai.com/>

Optional:

- Abeba Birhane and Deborah Raji, “[ChatGPT, Galactica, and the Progress Trap](#),” *WIRED*, December 9, 2022.

Tues, January 24th – Literary Language Generation, week 1

What is the history of language generation in a literary context and what are some examples?

Readings:

- Pressman, Jessica. "Electronic literature as comparative literature." *Futures of Comparative Literature*. Routledge, 2017. 248-257.
<https://stateofthedisipline.acla.org/entry/electronic-literature-comparative-literature-o>.
- Nelson, Theodor H. "[Chapter 0: Hyperworld](#)." *Literary Machines*. (1981).
https://monoskop.org/images/b/be/Nelson_Ted_Literary_Machines_c1987_chs_0-1.pdf
- Carter, Shan, and Michael Nielsen. "Using artificial intelligence to augment human intelligence." *Distill* 2.12 (2017): e9. <https://distill.pub/2017/aia/>.
- Howe, Daniel C., and A. Braxton Soderman. "The aesthetics of generative literature: lessons from a digital writing workshop." *Hyperrhiz Journal of New Media Cultures* (2009).
<http://hyperrhiz.io/hyperrhiz06/essays/the-aesthetics-of-generative-literature-lessons-from-a-digital-writing-workshop.html>
- Johnston, David Jhave. "ReRites: Machine Learning Poetry Edited by a Human." *Glia*, <http://glia.ca/rerites> (2019).
- Fan, Lai-Tze. "Symbiotic Authorship: A Comparative Textual Criticism of AI-Generated and Human-Edited Poetry." *ReRites: Responses*, edited by Stephanie Strickland, Anteism, 2019, pp. 57-64. [*ILL Request in process*]

In-class lab: computationally-generated poetry (and other forms of literary text)

Tues, January 31st – Literary Language Generation, week 2

What does computer-generated literature look like right now, what are the models/tools used to make it, and how should literary scholars analyze it?

Readings:

- Mark Riedl et. al "An Introduction to AI Story Generation"
(<https://mark-riedl.medium.com/an-introduction-to-ai-story-generation-7f99a450f615>)
- Some GPT-generated text:
 - Pamela Mishkin. "Love and AI." <https://pudding.cool/2021/03/love-and-ai/>
 - James Yu. "Singular." <https://jamesyu.org/singular/>
 - Josh Dzieza. "The Great Fiction of AI"
<https://www.theverge.com/c/23194235/ai-fiction-writing-amazon-kindle-sudowrite-jasper>
 - Pick one: <https://wordcraft-writers-workshop.appspot.com/>
- Rettberg, Jill Walker. "Algorithmic failure as a humanities methodology: Machine learning's mispredictions identify rich cases for qualitative analysis." *Big Data & Society* 9.2 (2022): 20539517221131290.
<https://doi.org/10.1177/20539517221131290>

- Ippolito, Daphne, et al. "Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers." arXiv preprint arXiv:2211.05030 (2022). <https://arxiv.org/pdf/2211.05030.pdf>

Optional:

- Winters, Thomas, and Pieter Delobelle. "Survival of the Wittiest: Evolving Satire with Language Models." ICCV. 2021.
https://computationalcreativity.net/iccc21/wp-content/uploads/2021/09/ICCC_2021_paper_72.pdf
- Nichols, Eric, Leo Gao, and Randy Gomez. "Collaborative storytelling with large-scale neural language models." Motion, Interaction and Games. 2020. 1-10.
<https://dl.acm.org/doi/pdf/10.1145/3424636.3426903>

In-class lab: text generation with the HuggingFace transformers library and BERT

Tues, February 7 – Analyzing Language Models, week 1

What is the text used to train large language models and how has it been analyzed? How can literary scholars contribute to its analysis and/or critique?

Readings:

- Suchin Gururangan et al, "[Whose Language Counts as High Quality? Measuring Language Ideologies in Text Data Selection](https://arxiv.org/abs/2201.10474)," arXiv preprint arXiv:2201.10474 (2022) <https://arxiv.org/abs/2201.10474>
- Eun Seo Jo and Timnit Gebru, "[Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning](#)," *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020): 306-310
- Hannah Rose Kirk, Abeba Birhane, et al, "[Handling and Presenting Harmful Text in NLP Research](https://arxiv.org/abs/2204.14256)," arXiv preprint arXiv:2204.14256 (2022) <https://arxiv.org/abs/2204.14256>
- Jessica Marie Johnson, "[Markup Bodies: Black \[Life\] Studies and Slavery \[Death\] Studies at the Digital Crossroads](#)," *Social Text* 36.4 (December 2018).
- Sadiya Hartman, "[Venus in Two Acts](#)" *Small Axe* 26 (June 2008).

Tues, February 14 – Analyzing Language Models, week 2

What is the text that is generated by LLMs and can we use literature to analyze it? What are some other ways to make use of the text generated by LLMs and are they ethical? Can we look a little more under the hood to understand how LLMs make their predictions?

Readings:

- Li Lucy and David Bamman, "[Gender and Representation Bias in GPT-3 Generated Stories](#)," *Proceedings of the Third Workshop on Narrative Understanding* (2021): 48-55. Association for Computational Linguistics.

- Joon Sung Park et al., “[Social Simulacra: Creating Populated Prototypes for Social Computing Systems](#),” *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (October 2022): 1-18.
- Brandon Rohrer, “[Transformers from Scratch](#)” ← *spend as much or as little time on this as you’d like*

Explore:

- BERT for Humanists: <https://www.bertforhumanists.org/> (focus on textual apparatus; we will work through some of the tutorials in class this week and next)

Optional:

- Ashish Vaswani et al, “[Attention is All You Need](#),” *Proceedings of the 31st Annual Conference on Neural Information Processing* (December 2017): 6000-6010.
- Tom Brown, Benjamin Mann, et al., “[Language Models are Few-Shot Learners](#),” *Proceedings of the 34th Annual Conference on Neural Information Processing* (December 2020): 1877-1901
- Anna Rogers, Olga Kovaleva, and Anna Rumshisky, “[A Primer in BERTology: What We Know about How BERT Works](#),” *Transactions of the Association for Computational Linguistics* (2020) 8: 842–866.
- Enrique Manjavacas and Lauren Fonteyn, “[Adapting vs. Pre-training Language Models for Historical Languages](#),” *Journal of Data Mining and Digital Humanities* (2022)

In-class lab:

- More with the HuggingFace transformers library and BERT

Tues, February 21 – Using Language Models to Analyze Literature

What does literary scholarship that makes use of large language models look like right now?

What might it look like in the future?

Readings:

- Ted Underwood, “[Do Humanists Need BERT?](#)” *The Stone and the Shell*, July 15, 2019
- Rabea Kleymann, Andreas Niekler, and Manual Burghardt, “[Conceptual Forays: A Corpus-Based Study of ‘Theory’ in Digital Humanities Journals](#),” *Journal of Cultural Analytics* 7.4 (December 2022).
- M. Beshir Massri, Inna Novalija, et al., “[Harvesting Context and Mining Emotions Related to Olfactory Cultural Heritage](#),” *Multimodal Technologies and Interaction* 6.7 (2022)
- Hoyt Long, “Learning to Live with Machine Translation,” forthcoming in *American Literary History* (Canvas)

Optional:

- Jinbing Yang, Yann Ciaran Ryan, et al., "[Detecting Sequential Genre Change in Eighteenth-Century Texts](#)," *Computational Humanities* 2022
- Dallas Card, Serina Chang, et al., "[Computational analysis of 140 years of US political speeches reveals more positive but polarized framing of immigration](#)," *PNAS* 119.31 (2022)
- Li Lucy and David Bamman, "[Characterizing Language Variation Across Social Media Communities with BERT](#)," *Transactions of the Association for Computational Linguistics* (2021) 9: 538–556.
- Margherita Parigini and Mike Kestemont, "[The Roots of Doubt: Fine-Tuning a BERT Model to Explore a Stylistic Phenomenon](#)," *Computational Humanities* 2020

In-class lab: Using BERT to detect word similarity

Tues, February 28 – Guest lecture by [Jacob Eisenstein](#)

DUE: Final project proposal

Readings:

- Bernd Bohnet, Vinh Q. Tran, et al., "Attributed Question Answering: Evaluation and Modeling for Attributed Large Language Models," arXiv preprint arXiv:2212.08037 (2022). <https://arxiv.org/abs/2212.08037>

Optional:

- Jacob Eisenstein, Daniel Andor, et al. "Honest students from untrusted teachers: Learning an interpretable question-answering pipeline from a pretrained language model," arXiv preprint arXiv:2210.02498 (2022). <https://arxiv.org/abs/2210.02498>
- Sandeep Soni, Lauren Klein, and Jacob Eisenstein, "[Abolitionist Networks: Modeling Language Change in Nineteenth-Century Activist Newspapers](#)," *Journal of Cultural Analytics* (2021) (or just read [this shorter version](#) on *Public Books*)
- Jacob Eisenstein, David Bamman, and Tyler Schnoebelen, "[Gender Identity and Lexical Variation in Social Media](#)," *Journal of Sociolinguistics* 18.2 (2014)

Tues, March 7 – SPRING BREAK

Tues, March 14 – Open Questions, week 1

How do broader critiques of capitalism, colonialism, and "capture" inform our understanding of LLMs, their uses, and their limits? Are the goals of computer science research and (digital) humanities scholarship compatible at all?

Readings:

- Matthew Hannah, “Toward a Political Economy of Digital Humanities,” forthcoming in *Debates in the Digital Humanities 2023*, ed. Matthew K. Gold and Lauren Klein (Univ. of Minnesota Press, 2023): 3-26. (Canvas).
- Meredith Whittaker, “[The Steep Cost of Capture](#),” *ACM Interactions* 28.6 (November 2021): 50-55.
- Inioluwa Deborah Raji, Emily Bender, et al., “[AI and the Everything in the Whole Wide World Benchmark](#),” *of the 35th Annual Conference on Neural Information Processing* (2021).
- Toma Tasovac and Natalia Ermolaev, “[Parrots](#),” *Startwords* 3 (2022) and the three essays that issue includes:
 - Ted Underwood, “[Mapping the Latent Spaces of Culture](#)”
 - Gimena del Rio Riande, “[On Spanish-Speaking Parrots](#)”
 - Lauren Klein, “[Are Large Language Models Our Limit Case?](#)”
- Olúfẹ́ mi Táíwò, “Introduction,” in *Elite Capture: How the Powerful Took Over Identity Politics (and everything else)* (Haymarket Books, 2022) (Canvas)

Tues, March 21 – Guest Lecture (currently being scheduled)

Readings:

- TBD

Tues, March 28 – Open Questions, week 2

Continued from 3/14: How do broader critiques of capitalism, colonialism, and “capture” inform our understanding of LLMs, their uses, and their limits? Are the goals of computer science research and (digital) humanities scholarship compatible at all?

Readings:

- Nick Couldry and Ulises A. Meijas, “[Data Colonialism: Rethinking Big Data’s Relation to the Contemporary Subject](#),” *Television & New Media* 20.4 (2019).
- Karen Hao, “[Artificial Intelligence is Creating a New World Order](#),” *MIT Technology Review* (2022) and the four essays that issue includes:
 - Karen Hao and Heidi Swart, “[South Africa’s private surveillance machine is fueling a digital apartheid](#)”
 - Karen Hao and Andrea Paola Hernández, “[How the AI industry profits from catastrophe](#)”
 - Karen Hao and Nadine Freischlad, “[The gig workers fighting back against the algorithms](#)”
 - Karen Hao, “[A new vision of artificial intelligence for the people](#)”
- Jason Lewis, Noelani Arista, et al., “[Making Kin with the Machines](#),” *Journal of Design and Science* 4.3 (2018).
- Sabelo Mhlambi, “[From Rationality to Relationality: Ubuntu as an Ethical & Human Rights Framework for Artificial Intelligence Governance](#)” (2020)

Explore:

- [Indigenous Protocol and Artificial Intelligence Working Group](#)
- Malavika Jayaram, Aarathi Krishnan et al, "[AI Decolonial Manifesto](#)"

Optional:

- Paola Ricaurte, "[Data Epistemologies, The Coloniality of Power, and Resistance.](#)" *Television & New Media* 20.4 (2019).
- Abeba Birhane, "[Algorithmic Colonization of Africa](#)," *Scripted* 17.2 (August 2020)

Tues, April 4 - Open Questions, week 3

Topics/readings to come from class

Unit 4: Final Projects

Tues, April 11 – In-class project workshop (LK away)

Tues, April 18 – Final project presentations (BJM away)

Final projects due via Canvas by 2pm on Tuesday, May 2nd

