Small Models, Big Norms: Close-Reading Gender with Micro-Language Models

This presentation demonstrates how small, purpose-built language models—trained on tightly delimited corpora—can serve as tools for close reading and for analyzing the production of gender norms across divergent textual contexts. Rather than leveraging large-scale machine learning systems to identify generalized patterns, I construct a suite of micro-language models, each designed not for scale but for specificity. These models are trained on three corpora that represent distinct views on gender: (1) conservative-leaning essays and opinion pieces published by the Heritage Foundation;[1] (2) progressive-leaining essays and opinion pieces from the ACLU;[2] and (3) transcripts from the reality television series *Love Is Blind*.[3] Each of these datasets was gathered and curated individually for this project, and then used to fine-tune a small language model on the custom data. By building a separate model from each dataset, this project treats computational prediction not as a neutral or objective function but as an analytic mode capable of engaging how language patterns structure gender, biological sex, embodiment, desire, and social norms.

Despite its focus on technology in method, this approach is inspired by a theoretical shift from Literary Studies, specifically from Eve Kosofsky Sedgwick's reconceptualization of reading practices. In her famous essay on reading, "Paranoid Reading and Reparative Reading: Or, You're So Paranoid You Probably Think This Essay Is About You," Sedgwick urges scholars

---

[1] Calado, Filipa. *gpt2-hertiage_foundation-gender* model repository. Huggingface. https://huggingface.co/gofilipa/gpt2-hertiage_foundation-gender. 2025.
[2] Calado, Filipa. *gpt2-aclu-gender* model repository. Huggingface. https://huggingface.co/gofilipa/gpt2-aclu-gender. 2025.
[3] Calado, Filipa. *LoveIsBlind_Pods* model repository. *Gofilipa*, Huggingface. https://huggingface.co/gofilipa/LoveIsBlind_Pods. 2025; Calado, Filipa. *LoveIsBlind_Postpods* model repository. *Gofilipa*, Huggingface. https://huggingface.co/gofilipa/LoveIsBlind_Postpods. 2025.

to move beyond extraction, unveiling, and suspicion toward generative, reparative modes of interpretation.[4] I adapt this insight to machine learning: instead of treating ML as a tool for mining information from text or predicting real-world outcomes, I use it as a way to read *with* a corpus—to surface what is most patterned, habitual, and structurally insistent within language. Small models lend themselves to this work because of the underlying predictive process that determines their outputs, which amplifies the internal tendencies of their source texts. The result is an aggregated portrait of how a specific perspective organizes meaning.

Even within polarized discourses, the aggregated patterns reveal conceptual middle spaces—outputs that blend or intersect ideas from the training data. This new literary form, which I call an *aggregate form*, enacts a statistical condensation of how different texts might relate to one another. The outputs from the progressive model, for example, align with what one might expect from a perspective that affirms gender diversity and expression, characterizing gender as with empowering language like "liberation," "beauty", and "joy". Meanwhile, in a similarly predictable way, the conservative model associates gender with what are typically conservative ideals, affixing gender to notions of culture, tradition, and reproduction, which prioritize stability over empowerment. Within the results, however, something unexpected emerges around the specific term, "subjectivity." This term appears in ways that one wouldn't typically associate with the conservative viewpoint:

> Masculinity is a subjective self-perception, not a universal concept.
>
> Femininity is a subjective, internal sense of self.
>
> The gender binary is a subjective, malleable, and often incorrect idea.

[4] Sedgwick, Eve Kosofsky, ed. *Novel Gazing: Queer Readings in Fiction*. Duke University Press. 1997.

> The gender binary is a subjective, internal, and often transitory concept.
>
> The gender binary is a subjective, grammatically incorrect and illogical concept that conflates sex and gender identity.

These examples do not describe the far-right viewpoint on gender——that it is based on the "biological truth" of two sexes, for example, as stated in recent executive orders from the White House.[5] Unlike these documents, which double down on the primacy of sex, defining it as "binary and biological," the generated outputs above more closely resemble the progressive view of gender, which asserts that gender describes identity, based on social behaviors, roles, and expression, among other things.[6]

In fact, the particular phrase, "gender is subjective", does not reflect the conservative position from the Heritage Foundation data. Rather, it reflects a conservative frame for the progressive position. It represents what a conservative thinks a progressive person thinks gender is—as something insubstantial, as a feeling. The outputs, then, express not a single perspective of gender, but an *aggregation* of perspectives into a single statement. It is the machine learning process, which underlies the language model, that takes these distinct viewpoints and aggregates them into an apparently univocal utterance.

---

[5] The White House 2025, "Defending Women From Gender Ideology Extremism And Restoring Biological Truth To The Federal Government"; The White House 2025. "Keeping Men Out of Women's Sports."

[6] The American Psychiatric Association, for example, defines gender identity as "a person's inner sense of being a girl/woman, boy/man, some combination of both, or something else" ("What is Gender Dysphoria?"). Similarly, the World Health Organization defines gender identity as "a person's innate, deeply felt internal and individual experience of gender," and contrasts it to biological sex, adding that gender identity "may or may not correspond to the person's physiology or designated sex at birth" ("Gender and health" 2025).

Then, I move from exploring political discourse to popular discourse, using a reality TV show, *Love Is Blind*, as my textual source. The show takes heterosexual men and women, puts them into a "blind" dating experiment with the ultimate goal of marriage. First, the daters are placed into dating "pods" where they can hear but not see the other. Then, once they agree to get married, they meet and live together for one month until their wedding day, making their final decision, whether or not to remain together, at the altar. Similarly to the two models that I trained on conservative and progressive perspectives, I trained two models based on the transcripts of the show: the first model is trained on the transcripts from the episodes in the pods, when the participants fall in love and agree to get married without ever seeing the other; and the second model, I trained off the period following the pods, when the daters finally meet and live together, which is when the relationships tend to fall apart.

The inclusion of *Love Is Blind* introduces a different but related discursive domain—one in which gender and embodiment are not debated explicitly but enacted through the rituals of heterosexual romance. By building two additional micro-models—one trained exclusively on pod transcripts (in which participants converse without seeing each other) and one trained on post-reveal transcripts—I investigate how sensory conditions shape the way cisgender participants talk about their bodies, their desires, and each other. Here, I draw on Jay Prosser's theorization of the "transsexual trajectory," which conceptualizes the relationship between the visible body and the felt body-image as a site of narrative and phenomenological tension.[7] While Prosser addresses the experiences of trans subjects navigating dysphoria, medical transition, and the alignment of body and identity, *Love Is Blind* participants inadvertently encounter a milder

[7] Prosser, Jay. *Second Skins: The Body Narratives of Transsexuality*. Columbia University Press. 1998.

form of this tension when the show's structure temporarily divides their sensory experience: the disembodied voice in the pods versus the visible, tactile body after the reveal.

The pods model consistently produces language that reflects a heightened, inwardly directed awareness of bodily sensation. When prompted with terms like "my body," "touch," or "desire," the model generates statements filled with fantasy, abstraction, and metaphysical intimacy. These outputs highlight how the absence of visual information inflates the role of imagination. The pod environment encourages participants to construct a body-image of the other that is detached from physical reality and laden with desire. The model's generated language reflects this disembodied eroticism, which parallels (without equating) elements of Prosser's account of bodily split: the subject relates to an imagined, affectively charged body rather than the visible body that eventually arrives.

After the reveal, the postpods model reverses these patterns in ways that illuminate the reintroduction—and destabilization—of embodied presence. When prompted with "touch," the model generates sentences such as "When I touch you, I feel like I'm in my head," "When I touch you, you just feel like it's so weird," "When I touch you, it feels like a jab." The romantic idealism that saturated the pods gives way to ambivalence, discomfort, and embodied awkwardness. Visual confirmation collapses the distance between body-image and physical body, but it does not seamlessly reconcile the expectations built in the pods. Instead, it produces a new form of estrangement. Many real episodes of the show reflect this dynamic: participants report that hugs feel "weird," kisses feel "off," and physical closeness feels prematurely intimate. The models capture this instability and formalize it into aggregated linguistic patterns.

Bringing these corpora together—conservative and progressive political discourse, and the romantic dramaturgy of *Love Is Blind*—allows the project to demonstrate how gender norms are produced through the microdynamics of embodied interaction. The political models reveal how institutional rhetorics construct and contest the meanings of "biological sex," "womanhood," and "gender identity." The *Love Is Blind* models reveal how these concepts take on lived, affective dimensions through the ordinary scripts of cisheterosexual courtship. Together, the models show that gender normativity is not merely ideological or conceptual but materially and emotionally grounded.

By mobilizing small data and small models as interpretive tools rather than extraction engines, this project proposes a method of computational cultural analysis that is attentive to specificity and aligned with reparative reading. It offers a sustainable practice that can be adapted and re-appropriated across educational contexts (all the code for the project is openly licensed online[8]). The models themselves are adapted from an open source model, gpt-2,[9] and were trained on a single laptop, over a single afternoon, in less time than it takes to stream a movie.

Ultimately, this project demonstrates that small, purpose-built language models can serve as powerful tools for close reading, enabling scholars to analyze normative structures without reproducing them. By showing how discourses on gender organize themselves across political, legal, and popular cultural domains, the paper argues for an approach to machine learning that embraces close reading and small datasets.

---

[8] Calado, Filipa. *love_blind* code repository, *Gofilipa*, Github. https://github.com/gofilipa/love_blind. 2025.
[9] *Openai-community*. "Gpt2". Huggingface. https://huggingface.co/openai-community/gpt2. 2025.