Imposters and Transphobes: The Function of Fear in Grad School

Thank you for having me here today. I am honored to have been asked to come and share my perspective on theme of graduate work, and also on the question of how our work as students in English might speak to each other.

I think this question is a crucial one at this moment in time, when the public discourse seems to be beyond the reach of (or even desire for) deep, critical thinking. Right now, we're seeing incredible amounts of mis-information about what's happening on college campuses across the country, and we're witnessing a kind of fascism that twists the definitions of tolerance and free expression to serve the interests of imperialist violence.

I think students of English have unique tools for thinking through contemporary issues like mis-information and propaganda. But I also think that we have a lot holding us back as a discipline, a lot of constraints in our discipline that keep certain people out and discourage certain kinds of work.

For this talk, I'm going to organize my comments around the theme of fear. I'm going to think through the emotion of fear, which is traditionally seen as a negative emotion, to see how it can become a productive force that guides reading practices.

I'm going to begin by sharing some of my experience with fear, from the perspective of a student doing what is considered to be "non-traditional" work in English departments. Then I'm going to talk about critical methods that transform negative feelings like fear into something that is generative and productive. Finally, I'm going to close by talking about my most recent scholarship, which considers another kind of fear, that of transphobia, that is currently sweeping this country in the form of an anti-trans movement.

fear

I'll start with a kind of fear that many of many of you can probably relate to—that is, impostor

syndrome.

Throughout my time as a student here, I felt a strange contradiction—that I didn't really belong in an English department, but also, that the English department is the only place where I can do my kind of work. That's because my dissertation is not really about literature, it's about software, specifically, software that handles and transforms language data. My main research question is how digital tools engage with the semantic and figurative qualities of language, in how these tools reduce the complexities, ambiguities, and nuances of language forms into computable text. My goal was to examine the mechanism of digital tools, to find where they make this reduction, and then to find ways of re-working the tools to capture some of the expressivity of language that has been lost. To study this effect, I used queer literature, a body of literature that is rich in narrative and figurative forms, as a test case. I experimented with various digital tools, like text analysis for example, exploring how these tools manipulate and transform literary language into data.

In addition to my written dissertation, then, I ended up coding software applications, using programming skills that I taught myself deliberately as part of my graduate training. For my work on text analysis, for example, I used the Python programming language to create a simple text analysis application that analyses gender terms in novels, like Virginia Woolf's *Orlando*.

[SLIDE 2 - SIMILARITY FOR WOMAN]

Here's a screenshot of one component of that text analysis application. It shows a list of words that are computationally "similar" to the word "woman," in the novel *Orlando*. The application determines word similarity through context: it examines a target word, like "woman," and the words that surround "woman," what's called a context window, and then it compares those contextual words to other words and their contexts throughout the novel. It works like almost every other text analysis task: by counting words and compiling massive lists of word counts. From these word counts, it then determines words that tend to have similar contexts. Here, the words "flight" and "abase" have similar contexts to the word "woman," and those words appear next to their

probabilities, representing how closely they relate to "woman".

Besides just showing word similarities, my text analysis application also juxtaposes this so-called "distant reading" view of the words with "close reading" views of each word in context. The idea is that a reader might flip between these two views of text, close and distant, as they explore words in the novel. My goal with the project was to facilitate a new reading practice where distant reading might seep into and inform close-reading analysis.

[SLIDE CUNY ACADEMIC WORKS & GITHUB]

While my written dissertation is available on CUNY Academic Works, my coding work is hosted on a website called Github, where a lot of software (especially DH projects!) is shared and published in a format that's open and collaborative. You can see on the right my homepage on Github, which features my projects or "repositories" that are digital components to my dissertation (the project title is the text in blue, with the description right below it).

I know this kind work might seem impressive now, after I've completed it, but for many years, I really struggled to share it with folks in English. When I did share it, the most common responses were blank facial expressions, usually accompanied by the phrase, "Oh, that's really interesting. But I don't know anything about that." This was also the response I got, a couple of times, when I asked professors to be on my committee. I don't blame them (especially because technology has a tendency to make people feel flustered or frustrated), but it was really discouraging, especially early on in the project when I didn't have a good sense of where I was going. And I know this is probably a relatable feeling for many grad students, regardless of the topic of their research. There were many moments where I had to kind of power through and convince myself that the work was worthwhile.

For me, the impostor syndrome manifested most strongly in one particular form: as an intense fear of public speaking. It began my first semester of graduate school, when, in the middle of speaking to my seminar class, my mind went completely blank. I don't know exactly what caused it,

but I know it has something to do with the intensity of my passion about what I was saying. Something about that intensity threw me off. For many years afterward, I was afraid of it happening again, so I avoided going to conferences and from speaking too much in my seminars. The times that I did speak, I would write down everything, and I mean everything, even if it was just a two-sentence comment, that I planned to say before I opened my mouth. Sometimes, I would also shake while I spoke, in a way that recalls how Helene Cixous describes the shaking woman in her beautiful essay, "The Laugh of the Medusa."

[CIXOUS QUOTE]

> "Listen to a woman speak at a public gathering (if she hasn't painfully lost her wind). She doesn't 'speak,' she throws her trembling body forward; she lets go of herself, she flies; all of her passes into her voice, and it's with her body that she vitally supports the 'logic' of her speech."

In the essay, Cixous makes the vital point that expression (both written and spoken) is an embodied activity—this is what the shaking signifies precisely. She explains that women can either choose to stay trapped in their own bodies by a language that does not allow them to express themselves, or they can use the body as a way to communicate. (and if you haven't read the essay, I highly recommend it)

what does knowing do?

My experience with public speaking, unfortunately, was not as glorious as Cixous's rendering of the shaking woman. But it does make me think about fear, and to wonder how fear might be re-deployed as a tool for drawing something productive from the repressive structures that create it in the first place. In re-thinking the function of fear, I'm inspired by the scholarship of Eve Kosofsky Sedgwick, who as you may know, is a major figure in Queer Studies, and who also taught in this department until 2009, when she passed.

[SEDGWICK'S BOOKS]

Here you can see some of her books, which are among the most famous books in Queer Studies, and helped to shape the field, like *Between Men*, *the Epistemology of the Closet*, and *Touching Feeling*.

[SLIDE 5 – SEDGWICK IN ENGLISH LOUNGE]

And here, you'll see an image of Sedgwick with Robert Reid Pharr, who also taught in this department, and Samuel Delany, the Science Fiction writer. The background of the photo should be familiar to all of you.

I'm inspired by the way that Segdwick has framed her relationship to reading practices over her career. Throughout her work, she shifts from a kind of reading that she calls "paranoid reading" into a new mode called "reparative reading." By paranoid reading, Sedgwick means a critical practice that seek to unveil truth by searching for hidden meaning in text. In her famous essay on this topic, Sedgwick frames the difference between paranoid and reparative reading with a question about truth, and the role of unveiling or exposing truth in critical analysis:

> "Moving from the rather fixed question Is a particular piece of knowledge true, and
> how can we know? to the further questions: what does knowledge *do*–the pursuit of it,
> the having and exposing of it" (124, *Touching Feeling*)

In this essay, Sedgwick alleges that truth, on its own, does not motivate action. Merely knowing that something is true, revealing the presence of systematic oppression, injustice, discrimination, for example (and here, Sedgwick is theorizing within the context of the AIDs crisis), is not enough to "enjoin that person to any specific train of epistemological or narrative consequences" (123).

Truth does not motivate action, because, Sedgwick explains, action is not the end goal of paranoia. Paranoia is by nature suspicious: it suspects that there must be something hiding under the surface. Because of this suspicion, paranoia can never be surprised, only satisfied, and only

temporarily. So when it finally "discovers" the secret, that discovery only affirms what it has known all along. One is never "moved," to to speak.

Paranoid reading practices abound in literary criticism, especially, as Sedgwick admits, in her own criticism. In her earlier books, in *The Epistemology of the Closet*, published in 1990 (and pictured here as the middle book), for example, she uses paranoid methods to expose what she calls the unstable binaries between heterosexual and homosexual categories. Through close-readings of fiction, she exposes the inherent instability of these binaries — where one term is not symmetrical or simply subordinated to another, but rather, depends the other for its meaning through "simultaneous subsumption and exclusion" (10). Sedgwick explains that such binaries are "sites that are peculiarly densely charged with lasting potentials for powerful manipulation" (10).

In her later work, such as that in *Touching Feeling*, which is the book on the right, Sedgwick explores other critical reading methods that go beyond what she calls "the logic of repression" (a logic which she traces to Foucault) that guides much of paranoid reading, including her own earlier work. She proposes a mode of "reparative reading," which focuses on connection rather than exposure. Reparative reading "surrender[s] the knowing, anxious paranoid determination that no horror, however apparently unthinkable, shall ever come to the reader as new" (146). In a reparative reading practice, a reader allows herself to be taken by surprise.

She demonstrates reparative reading by analyzing the affect of shame. What if, Sedgwick asks, we take something that is typically seen as a negative, structuring force in queer identity, and examine how it unlocks creativity and productivity? Sedgwick describes shame as a contagious affect, which may be read as a mobilizing and creative force in text: she explains that,

[SLIDE 6: SHAME QUOTES]

> "Shame—living, as it does, on and in the muscles and capillaries of the face—seems to be uniquely contagious from one person to another." (63 *Touching Feeling*).

She also describes shame as:

> "not a discrete intrapsychic structure, but a kind of free radical that (in different people and different cultures) attaches to and permanently intensifies or alters the meaning of—of almost anything: a zone of the body, a sensory system, a prohibited or indeed a permitted behavior, another affect such as anger or arousal, a named identity, a script for interpreting other people's behavior toward oneself" (62)

She demonstrates this reading practice by analyzing metaphors that are made possible through shame. For example, in the fiction of Henry James, she connects moments of "blushing" and "flushing" to a fantasy of the skin being entered. Shame, in this reading, is a way of pulling other affects and images into relation. This is opposed to paranoid reading, which might plumb shame for what it reveals about a hidden or repressed sexuality. She explains that, "When we tune into James's language on these frequencies, it is not as superior, privileged eavesdroppers on a sexual narrative hidden from himself; rather, it is as an audience offered the privilege of sharing his exhibitionistic enjoyment and performance of a sexuality organized around shame" (54).

my work

I'm interested in this move that Sedgwick makes, of taking what is typically seen as a negative, repressive affect, like shame, and seeing how it opens up possibilities for reading new connections in text.

In my current work, I am exploring how I might enact something similar with the feeling of fear. In the work I've been doing since the dissertation, I have been interested in a particular kind of fear, of transphobia, and how it emerges in our current political moment.

[ROGD PAPER]

For example, one of the things I'm studying is the (now disproved) clinical phenomenon of so-called "Rapid Onset Gender Dysphoria" (ROGD), which stipulates that trans-ness can be

contagious among adolescents. The phrase was coined by Dr. Lisa Littman in her study of parents of transgender youth, published in 2018. In her study, Littman interviews over 250 parents of transgender children and concludes that what she calls "peer contagion" of gender dysphoria may be a contributing factor for adolescents who decide to transition. Almost immediately following its publication, Littman's study was criticized, including by the publisher, and its methodology and findings have been disavowed by every major medical association since then.

Although ROGD is not recognized as a valid diagnosis, it has been used and is still used as fodder for anti-trans propaganda and discrimination. It has made its way into the public lexicon, appearing in books, shows, and most importantly for my project, legislative bills that are being written, debated, and passed across the United States. For those who are unfamiliar, this legislation limits trans peoples' access to basic healthcare, public facilities, legal recognition, and more.

[ANTI-TRANS LEGISLATION MAP]

Here, you can see a map of where these bills are most concentrated across the country, and a chart of how many bills are proposed and passed over the last four years. Notice that this year, though we are hardly in the fifth month, we've almost caught up to the total bills for last year.

I'm interested in this threat of gender transgression, and specifically, in the language outlawing gender transgression, of transitioning from one gender to another, or of opting out of binary systems of gender. Why is this particular kind of transgression so controversial among a large part of our population? Why is the fear of this transgression itself so *contagious*? I am interested here in two kinds of fear: in the fear of transness being contagious (in the false phenomenon of ROGD), and in how this fear of transness *itself* is contagious, in the form of transphobia.

To study these phenomena of fear, I apply the same analytical spirit from my dissertation, where I use close-reading strategies to analyze aspects of digital tools, to look into their inner mechanisms, toward tracing their technological processes. For this project, I've decided to study AI tools, known technically as a "large language models" or LLMs, which include chatbots like ChatGPT. My focus

on this project, for the past several months, has been to understand how these models are created, how they are "trained," so to speak, so I can trace how they *perpetuate* biases, like transphobia, from their training data into the text that they generate.

I'm going to give a quick overview of how this training process works, just so we are all on the same page. The training process begins with the model taking and processing massive amounts of text from the internet. From processing this training data, word by word, it starts to compile numerical probabilities for which words tend to surround other words. It represents these probabilities with numbers, with actually a very large list of numbers. Here's an example of what one of these lists might look like, for the word "protest" (and this is from a model taken from twitter data)

[PROTEST SLIDE]

Like the first slide that I showed you from the text analysis application, probability here is important. Words are defined by their relationship to other words in the database. Here, we see words that are most closely associated to "protest" in the twitter database.

You can think of these numbers, or scores, functioning like definitions, which represent the word's *meaning* for the computer. To us, these scores look just like a long list of numbers, but to a computer, the scores represent a given word's meaning through its *relationship to every other word in the entire dataset of words*. That's why, by the way, these models are so large, and why they take so long to train. It's because every single word is represented by a massive list of probabilities, probabilities for how that word relates to every other word in the language. A language model will *generate* content by doing math with the scores attached to each word in its database. And the math that they use to make generate text is actually math that many of us have heard of before in math class: things like matrix multiplication and cosine similarity.

Here is a famous formula that introduced this technology of scoring words, which is technically called "word vectors" to the world.

[KING - MAN + WOMAN = QUEEN]

The idea is that by taking all the numbers that represent king, then subtracting the ones that represent man, and adding the ones that represent woman, you will get queen. I won't get into the sexism of this formula (what exactly is being subtracted, for example?), but I want to point out that it has great currency as it is THE formula that introduced this technology to the world.

[WORD VEC PAPER]

And here is an image of the original paper itself, "Efficient Estimation of Word Representations in Vector Space," which was published by Google researchers in 2013. Just for some historical context, in the history of Machine Learning: this paper contributed to the development of a model architecture, called the "transformer architecture" in 2017, also by a research team at Google. And *that* development, which allowed AI models to take more complex levels of word contexts into the text generation process, is what led to the explosion in AI models that we are currently seeing today.

So, to sum up: these tools collect and examine data from the internet in order to glean patterns of language from which it can generate new content. Put simply: it generalizes how language works by studying examples of language forms. Given how much the training data, and the specific configurations of words in the training data, affect the model's text output, I am very interested in using AI tools to study anti-trans bias, and particularly, the fear of contagion, of ROGD.

Right now, I am currently the middle of a data gathering and model training process, to study how models perpetuate this fear from their training data onto their outputs. I am compiling a dataset of gender and related terms from these bills. I'm interested specifically in how these bills, which limit trans peoples' rights, are framing what it means to be trans and to receive gender affirming healthcare.

[SLIDES OF DEFS FROM BILLS]

Here's some examples of the language that I extracted so far from the legislation. The dataset is

10

focused on definitions of gender identity and related terms, like "biological sex" and "gender transition procedure," which are underlined here. To get these definitions, I had to scrape the bill text from congress.gov's servers. And when I had the bill text, I wrote a pattern matcher to extract textual patterns that resemble definitions. (If you're interested in this process, all my code is on github).

[SOME WORDS IN BLUE]

Here, the text in blue indicates words and phrases that seem to suggest underlying assumptions about binary gender and gender transgression. Right now, I am interested how these assumptions are being constructed in subtle ways, in seemingly harmless formulations. For example, I am interested in the word "regardless," which appears in many of definitions, and collapses an essentialized notion of binary gender with biological sex, and contrasts that with gender as expression.

As I continue to build and clean my dataset (which is a long, technical process that involves data gathering, processing, and cleaning), I've also been dabbling with using them to train AI models.

Leaving aside all the hype about AI, and whether or not it is "intelligent," or moving toward what the industry calls "general intelligence," AI tools like large language models are really good at one thing: at making *predictions*. At generating content that is *plausible*. This is a fascinating phenomenon, because it makes them very good at guessing or improvising, but not at all good at being creative, at innovating. A language model can only generate what it has already seen before. Even a phenomenon like "hallucination," that a language model spews text that has no bearing in reality, is based on the tendency of models to repeat what they've already seen. They hallucinate not because they are creative or random, but because they are designed from statistical processes to generate what is most plausible rather than accurate.

This tendency toward plausibility creates an interesting perspective for me to think through how Trans Studies scholars have characterized trans affects. Typically, these scholars describe trans

affective modes by distinguishing them from "queer" modes. In a roundtable called "Thinking with Trans Now" published in *Social Text*, trans studies scholar Eliza Steinbock explains,

> "trans analytics have (historically, though not universally) a different set of
> primary affects than queer theory. Both typically take pain as a reference point,
> but then their affective interest zags. Queer relishes the joy of subversion.
> Trans trades in quotidian boredom. Queer has a celebratory tone. Trans speaks
> in sober detail. Perhaps the style of trans studies has been for the most part
> realist, but this should not be mistaken for base materialism. Even speculative
> thinking requires enough detail to launch into new realms."

Other trans scholars like Marquis Bey and Andrea Long Chu have made similar points; with Bey making the point that queer's intervention can be described as "anti" or militant, while trans is "non" or based in refusal ("Thinking with Trans Now"); and Chu has remarked that trans studies, rather than resisting norms, "requires that we understand–as we never have before–what it means to be attached to a norm, by desire, by habit, by survival" ("After Trans Studies" 108).

This makes me wonder, could AI-generated text, as a kind of approximation, a normalization, of its training data, be used to study the attachments to norms and the quotidian that characterizes trans affective modes? Could the same processes also be used to study the attachment to norms that characterizes the opposite movement, in transphobia, like perspectives driven by the fear of ROGD? What might outputs from AI text generation suggest about the allure, the threat, the "seduction," as Trans Studies scholar Cassius Adair puts it, of gender transgression?

While this project might sound very ambitious, I'll admit that, so far, my results are not very encouraging. I need to continue to add more training data and to tweak my model configuration, probably numerous times, before I find something really interesting.

Nonetheless, here are some excerpts of my language generated by my model, which I trained by

feeding it some examples of anti-trans legislation that I have already prepared.

[SLIDE SHOWING THE GENERATED TEXT]

Here, you can see the prompt text (so, text that I entered as a prompt to the model, in italics) and the AI model's responses (where it provides a continuation of my prompt) in normal, unitalicized text.

As you can see from skimming the results, the models are showcasing the tendency toward plausibility, specifically in the tendency to repeat itself, which is a fascinating concept in machine learning.

I will close now by coming back to this idea of fear, and particularly the fear of contagion, which drives some strains of transphobia. Cassius Adair offers a useful perspective for thinking through the fear of contagion. In his study of trans erotics, and specifically "trans for trans" or "t4t erotics," Adair asks, "Why shouldn't transness be transmissible or contagious? Why can't the erotic be a site of producing trans identity or practices?" He points out that, after all, cis people do it all the time: they use sexuality and sexual encounters as sites of identity formation.

Here, I see Adair doing for contagion what Sedgwick does for shame: turning something that is traditionally seen as a negative into something that may be generative and productive.

It is the same kind of thing I hope to accomplish with this project, and something that I think is possible by using the tools that we gain in English departments—that is—by close reading, or what Sedgwick calls, "imaginative close reading."

This is a kind of reading that allows one to take what has been a tool of oppression and turn it into a creative resource. Sedgwick explains that this kind of reading exposes "the ways selves and communities succeed in extracting sustenance from the objects of a culture—even of a culture whose avowed desire has often been not to sustain them" (*Touching Feeling* 151). Thank you.

13