CaladoF

Table of Contents

[1. thank you](#orgc1ce09e)

[2. 1. overview of diss - critical look at technology - 10](#orgeb62c1e)

[2.1. main idea](#org1bda688)

[2.2. orlando close reading](#org9d549ae)

[2.3. reduction](#org6ad1c2f)

[3. detail of my work - text encoding - 20 min.](#org7b7f3ca)

[3.1. queer studies trajectory](#orgb8b03a4)

[3.2. text encoding chapter overview](#org58f37c7)

[3.3. hierarchy](#org7ba4df8)

[3.4. Dominance structures in the archive](#orgd89a171)

[3.5. disidentification to queer form](#org78115c0)

[4. 3. where my work is now going - algorithmic bias - 10](#orge704b12)

[4.1. anti trans & text generation](#org6998c4f)

# thank you

Thank you for having me. Today I'll be talking about my dissertation, which I defended in November in the English department at the Graduate Center at CUNY. For that project, I applied a methodological approach based in Queer Studies to analyze Digital Humanities tools that work with language and language forms. I will show how these digital tools flatten or reduce some expressive qualities of language, particularly language that relates to sex, gender and sexuality, and how Queer Studies can be deployed to surface these qualities. Toward the end of the talk, I will discuss how this work has led me to a follow-up project, which applies the same analytical spirit to examine what AI tools and gender bias, specifically by using language models to analyze legislation that limits transgender peoples’ rights, which is currently proliferating across the United States.

# 1. overview of diss - critical look at technology - 10

## main idea

First, my dissertation, entitled “Since No Expressions Do: Queer Tools for Reading Literature”, uses literature, and queer literature in particular, as a test case for exploring what happens to language when it is transformed into electronic formats and processed with digital tools. Each of my chapters takes up a different aspect of computation–such as programming logics, data structures, and interface effects–to explore how that aspect handles stylistic and formal aspects of sex, sexuality and gender. I find that digital processes, which by design flatten semantic expressiveness of language into text data that can be categorized and counted by the machine—that these same processes can be reframed to index the semantic expressiveness that has been lost.

My methodology is to do “close readings” of digital tools, to examine how each one handles the complexity and nuances of language forms. I look into the mechanism of the tool, tracing the way that it transforms language into text data and manipulates that data for electronic processing and display. I closely examine the logical structures of computer code, across programming and markup languages such as Python, JavaScript, and XML. From the code’s internal logics, I isolate a specific constraint, an aspect of the tool that collapses or reduces expressive qualities of language into computable data. Then, I draw from theorizing in Queer Studies and its critiques to find ways of working within this constraint. Queer Studies offers this project methodologies for locating and resisting totalizing and oppressive systems, such as that of heteronormativity. I look to these systems as models for reading the strict and disambiguating logics of computer codes.

## orlando close reading

Before elaborating on my theoretical intervention, I'll share a small example of what this close reading method looks like, to concretize the kinds of reductions that are inherent to computational processes, and how these reductions specifically affect language data. Drawing from Literary Studies, close reading is a mode of analysis that pays attention to textual detail. It focuses on specific words, phrases, and punctuation for interpreting the tone, mood, and meaning of passages. It is a way of grounding the reading of a text in the materiality of language and language forms.

SLIDE 2: Tilda Swinton image from *Orlando* film.

I’ll begin with a close reading from the first sentence of Orlando, a novel by Virginia Woolf. For those who are unfamiliar with this text, it tells the story of Orlando, 16th century English nobleman who has a sex change halfway through the book, and lives the rest of her life as a woman. Here’s an image of a famous film adaptation, by Sally Potter.

SLIDE 3: first sentence of Orlando.

The story opens with a scene of racialized violence, which embeds important assumptions about gender:

“He–for there could be no doubt of his sex, though the fashion of the time did something to disguise it—was in the act of slicing at the head of a Moor which swung from the rafters” (11).

Here, a teenage Orlando swipes at a severed human head that his ancestors brought back to England from the Crusades. Within this backdrop of imperial conquest, the language makes some qualifications about gender: first, the reassurance that there could be “no doubt of [Orlando's] sex,” belies itself: it doesn't serve to reassure, but rather, is conspicuous in its certitude. That this reassurance follows an em dash interrupting the first word, a bold gender declaration–HE–draws even more attention to gender as something that is in contention. The associations here deepen when we connect this certainty, and even defensiveness, about gender with the imagery of racialized violence, the “slicing at the head of a moor.” Masculinity here is co-constructed with nationality and race, which heralds a central tension of the story.

If one were to pass this text into a computer program for text analysis, that processing would evacuate the subtle ways that gender is unsettled in the sentence. That is because all text analysis is based on counting: counting the number of times a certain word appears, for example, and how many times its surrounding words appear, and what words tend to surround those words, within their own contexts, and so on, with the ultimate goal of deriving patterns from these word frequencies. In order to count the words, however, the text must be prepared: Semantic features which are deemed minor and frequently occurring (like articles, prepositions, punctuation) must be removed. This process of “text cleaning” ensures that minor details won't skew the counting, the word frequencies, of the text, which is the basic foundation of all text analysis tasks. For this sentence in particular, that means the pronoun “HE” and the em dash following that “HE” would be removed, so that the sentence would appear as this list of words:

SLIDE 4: list of tokens of the first sentence after cleaning

‘could’, ‘doubt’, ‘sex’, ‘though’, ‘fashion’, ‘time’, ‘something’, ‘disguise’, ‘act’, ‘slicing’, ‘head’, ‘moor’, ‘swung’, ‘rafter’.

While this process makes the text amenable to quantitative analysis, creating a computable dataset of words, it also loses the subtle ways that pronouns and punctuation affect the understanding of gender in this sentence.

## reduction

Quantitative methods for analyzing language, also known as “Distant Reading” methods, can lead to interesting and provocative results. By stripping minor details, connections between more substantial words and language patterns can emerge. My work is not intervening in those kinds of analysis. Rather, I’m interested in reading the reductions required by computational processes as opportunities to unleash textual meaning. In my work, reduction is a starting point for theorizing.

Here, I turn to Queer Studies, which arose in the early 1990s in the wake of the AIDs crisis, and offers modes for analyzing (and revealing the contradictions of) sex, gender, and sexuality as social constructions. Because Queer Studies seeks strategies for not only recognizing oppressive structures, like compulsory heterosexuality, but of working within and against them simultaneously, I find them to be useful models for re-thinking the ways that machine processes create a kind of compulsory disambiguation of data.

For example, with text analysis methods, I make a connection between the processes of cleaning a text with the way that gender is defined according to Judith Butler's famous theory of Gender Performativity, a theory that helped to inaugurate the field of Queer Studies. This connection centers on the concept of iteration, which is central to both text cleaning and gender performativity.

For text cleaning, I examine how the Python programming language (a popular language for Natural Language Processing tasks) works with text data through a mechanism known as the “loop,” which is a concept that exists across programming languages and is a core building block of algorithms. The loop, technically called the “for-loop” enables a program to do things en masse to data, by repeatedly applying the same series of actions to each piece of data, one by one. The technical language describes loops as a process of “iterating over” data.

SLIDE 5: definition of iteration from python.org

Here is a definition of the loop from the Python documentation.

“... loops are used when you have a block of code which you want to repeat a fixed number of times. The “for-loop” … iterates over the members of a sequence in order, executing the block each time” (https://wiki.python.org)

I then turn to Judith Butler’s concept of gender performativity, which defines gender as a series of repeated acts (of behaviors) that conform to an authorizing norm. Her central intervention is that behavior determines identity, or at least the illusion of identity. According to Butler, it is by adopting gendered modes of behavior that subjects come to exist at all.

SLIDE 6: definition of performativity by Butler.

Here is an excerpt of Butler’s definition of Gender Performativity:

“Performativity must be understood not as a singular or deliberate “act,” but, rather, as the reiterative and citational practice by which discourse produces the effects that it names” (Butler 2)

From this definition of Gender Performativity, I am interested in the characterization of gender as a “re-iterative… practice”. For Butler, this “re-iterative” practice is the key to resistance from withina co: by repeatedly performing gender constraints in ways that deviate from social expectations, subjects can subvert gender norms.

Taking this shared quality of iteration between Python and gender, I propose a text analysis practice that combines different tasks for analyzing text. I demonstrate this process in a custom text analysis application, which I designed as part of a software package, “The Queer Text Toolkit,” that is supplementary to my written dissertation. I coded this application using Python libraries for text analysis, like NLTK and Word2Vec, for those who are familiar with Natural Language Processing. Resisting traditional “reproducible” social science methods for text analysis, my custom text analysis application, which is available online, offers an iterative method that flips between close, textual detail and distant views of language data.

# detail of my work - text encoding - 20 min.

## queer studies trajectory

For this text analysis project, Butler's theory of Gender Performativity offers a model for re-deploying the programmatic logics of Python programming toward an innovative text analysis practice. Similarly, throughout my dissertation, I look to Queer Studies for strategies for resisting oppressive structures, particularly as they pertain to gender, sex, and sexuality, but also increasingly, to race.

With prominent theorists like Judith Butler and Eve Kosofsky Sedgwick, Queer Studies began by disentangling the contradictions within gender, sex, and sexuality, and the potential for (and evidence of) transgression within hetero-patriarchy. But the trajectory of Queer Studies as a field changed with the Queer subject’s increasing incorporation into mainstream politics and progressive agendas, such as the prioritization of same-sex marriage and inclusion in the military. As a result, numerous critiques emerged to resist this movement toward normativity (sometimes called “homonormativity”). On the one hand, there was an embrace of affective modes such as shame, failure, despair, and loss as constitutive of Queer subjectivity. On the other hand, there was a doubling down on transgression at all costs, which manifested in theorizing around a total rejection of mainstream acceptance, known as the “the anti-relational” or “anti-social” turn.

One of these moves, called “Queer of Color Critique,” opens the purview of Queer Studies in a way that influences my approach toward analyzing technology throughout my project.

SLIDE 7: Cruising Utopia cover

Here, I turn to Jose Esteban Munoz's critique of Queer Studies, who brings intersectional thinking from Women of Color Feminism in the 80s and 90s to re-energize Queer's political potential. As a corrective to Queer Studies' implicit (and sometimes not so implicit) whiteness in its critical frameworks, where gains in the Queer community are trapped within what Munoz describes as “the limiting, normative time of the present,” he offers a vision for queerness as futurity.

SLIDE : Munoz’s quote

He explains that,

“Queerness is a structuring and educated mode of desiring that allows us to see and feel beyond the quagmire of the present… Queerness is a longing that propels us onward, beyond romances of the negative and toiling in the present. Queerness is that thing that lets us feel that this world is not enough, that indeed something is missing.” (Cruising Utopia)

Munoz’s thinking opens possibilities of theorizing from political and social structures that allow some subjects visibility and acceptance while attempting to erase the existence of others. Queerness theorized in this way, as a desire for what is not present, drives Queer Studies forward, pushing it beyond the boundaries of sex, sexuality, and gender and into its imbrications with other oppressive structures, like race. In my dissertation, this move in Queer of Color Critique encourage my continual re-thinking of ways that computational structures constrain language into computable elements, into components that function within larger, totalizing systems.

## text encoding chapter overview

I’ll provide an example from another chapter of my dissertation, where I draw from critical methods in the history of the Black Atlantic to explore the constraints of data formats.

SLIDE 9: XML

In this chapter, I consider a data structure called XML (short for eXtensible Markup Language), which is widely used for encoding, or “marking up” written content within a hierarchical structure, similar to HTML.

SLIDE 10: XML hierarchy

Like HTML, this data structure uses “tags” to wrap written content. Here, the text is written in gray, and the tags appear in blue. The tags have descriptive labels like <page> for a page, <quote> for a quotation, and <del> for a piece of deleted text.

All written content must be enclosed with these descriptive tags, which must themselves be nested neatly within one another, so that so-called “child” elements are contained within their “parent” elements. Due to this nested format, XML code resembles a tree structure, with one “root” element and nested within it, a number of “branches,” which might contain more branches.

Like my critique of text analysis's programming logics, I was interested in this tool’s constraint—how it reduces expressive aspects of gender, sex, and sexuality in language. To seek out this constraint, I used it in an editorial project, for transcribing the written text of a literary manuscript that was edited multiple time to remove evidence of homosexuality.

SLIDE 11: Dorian Gray MS & Wilde

I chose the manuscript of Oscar Wilde's novel, The Picture of Dorian Gray, which Wilde revised heavily before its first publication in 1890. For those who don't know, the story, it takes place in late 19th century London, where a beautiful youth named Dorian Gray makes a deal so he can stay forever young. While Dorian keeps his youth, his portrait, painted by his artist friend named Basil Hallward, would grow old and ugly in his place. The story then follows Dorian's descent into moral abandon, through which he eludes the marks of age and corruption, remaining young and apparently untouched. The novel was variously criticized for its apparently ambiguous stance toward Dorian's hedonistic lifestyle, as well as undertones of homoeroticism, which emerge in its various exhortations of masculine youth and beauty that recall Greek idealisations of the male form.

The critical scholarship on the manuscript explores how Wilde’s revisions stifle some of the more overt suggestions of homoeroticism. His strategy for doing so, critics generally agree, is to transform eroticism and passion into an appreciation for aesthetics and a focus on art. The changes that Wilde makes are often small ones, such as in word choice, that have significant effects on tone, mood, and connotation.

SLIDE 12: image from MS with diplo rendering of page 9

In one example, Wilde makes subtle changes to mute the intensity in the dialogue, replacing it with a sense of levity. In this scene, the painter, Basil Hallward, is about to explain to his friend, Lord Henry, why he cannot exhibit his portrait of Dorian Gray. I will read the text, prior to revision:

Lord Henry hesitated for a moment. “And what is that?” he asked in a low voice.

“I will tell you,” said Hallward, and a look of pain came over his face.

“Don't if you would rather not,” murmured his companion. (9)

In the revision, Lord Henry’s hesitation becomes a laugh, and Basil’s look of pain becomes an expression of perplexity. These changes trade affliction for confusion, and diminish the fatalistic connotations of Basil's passion. They also suggest an underlying sense of shame or restraint, especially from Lord Henry, who no longer hesitates or speaks in a low voice, but instead encourages Basil to share his story. Together, these changes work to lighten the mood of the scene.

***SLIDE 13: XML of page 9***

To encode these revisions from the manuscript into the XML data format, I customized a variant of the XML language. Using the XML data structure as a base, I created a schema with custom attributes for registering Wilde's revisions according to thematic categories, like “passion,” “intimacy,” “fatality,” and “beauty.” I chose these labels for the way they work together to aestheticize Dorian’s character, turning him from an erotic object into an aesthetic subject. In addition to theme, I also wanted to markup the text in a physical dimension, which would help indicate each instance of revision. So I created an attribute to mark the pen strokes on each revision.

This image of XML code here shows how I encoded the same passage from the manuscript, using my custom attributes to indicate theme and pen strokes. In the code, we see the tree-like structure of the XML, along with the tags (in dark yellow) that are particular to editorial projects, like tags for indicating deleted text in <del>, added text in <add>, and their parent tag, <mod>, which stands for modification. Additionally, we can see attributes in green, like “type” to indicate the type of modification (in this case, a substitution of one text for another), “rend” to indicate how a bit of deleted text should be rendered, such as with a strikethrough; and “place” to indicate where the addition was made, which is above the line. Finally, we also see my custom attributes: @theme and @strokes, to indicate the category for revision and the number of pen strokes that Wilde used on the section.

## hierarchy

My work on the minute, detailed labor of encoding the manuscript reveals a constraining quality about the XML data structure—that of hierarchical dominance. The XML requires that all data be contained as discrete components within this bounded structure, and they cannot overlap unless the inner element is fully nested within an outer element. When child elements do not fit neatly within the parent, the file will display error.

The problem is that language structures can be overlapping: for example, syntactic and phrasal structures may overlap with semantic structures, as happens often with in poetry, when the syntactic (or sentence-based) structure can overlap with a prosodic (or rhythm based) one. A classic example of this in poetry is called “enjambment,” when a sentence continues across a line break, running over one line into the next in a poem. In XML, we could not encode the sentence as a unit without creating an overlap between the line element and the sentence element, thus creating an error in the code.

Similarly, in Wilde’s manuscript, it was difficult to indicate the different rounds of revision with accuracy. For example, one page of the manuscript shows what are apparently multiple rounds of revision, that together have the effect of stifling suggestions of intimacy and passion from the text.

SLIDE 14: image of MS page 20 with transcription:

In this scene, the painter explains the effect that Dorian Gray has on his art:

“You remember that landscape of mine… It is one of the best things I have ever done. And why is it so? Because, while I was painting it, Dorian Gray sat beside me, and as he leaned across to look at it, his cheek just brushed my cheek. The world becomes young to me when I hold his hand, as when I see him, the centuries yield up all their secrets!”

His friend’s response, as you can see, is mostly illegible, but I could make out some words and phases, like “you must not talk”, “his power,” and “slave”, as well as the final line, which reads: “It is worse than wicked, it is silly. I hate Dorian Gray.”

In revising the passage, Wilde replaces “cheek just brushed my cheek,” with “hair just touched my hand,” before striking out the entire line, as well as the following sentence. He also strikes out much of Lord Henry’s response, replacing Lord Henry’s intensity with a relatively subdued interest when he says, “this is quite wonderful. I must see Dorian Gray.”

SLIDE 15: arrows showing areas of overlap in MS

A span of text, indicated by the blue text and arrow, creates an obstacle for the encoding. Here, Wilde layers the revision of “cheek just brushed my cheek” with “hair just touched my hand”, within a larger revision that deletes the rest of the paragraph. It is difficult to disentangle these layers while keeping the integrity of each revision as its own entity. For example, here is an encoding that shows the two layers of revision:

SLIDE 16: encoding of flat revision

We can see one substitution that is enclosed within a larger span of deleted text. However, the strict, disambiguated format of the XML requires more precision, which leads to a problem of overlap. Because Wilde, in this case, struck three words, “cheek,” “brushed,” and “cheek” over a longer span of deleted text, each of those words has to be encoded as separate deletions. The problem is that this encoding breaks up the integrity of the revision as a larger phrase. It separates what was in fact one revision into multiple instances, as indicated in this image on the right.

SLIDE 17: encoding of deep revision

Here, there are three modifications, or <mod> tags, within the larger deletion tag.

The XML’s data structure, and its need for strict disambiguation, makes it impossible to mark up the two rounds of revision. Its hierarchical structure, which requires that elements be completely nested within one another, cannot register each round as its own discrete entity. The only option is to create more granularity, which makes it difficult to draw meaningful connections between the rounds of revision. As a labelling tool, the XML surfaced moments where queerness, in this case, the suppression of queerness, transgresses the bounds of the data structure.

## Dominance structures in the archive

The more that I worked with the XML, the more I was interested in the hierarchical data model, and the kinds of dominance structures that are contained within them. Within its tree-like architecture, information is not only encapsulated or bound, it is delineated by the standards of each governing XML tag. There is an underlying lesson in power here, about who has it and who is subject to it.

In Wilde’s text, for example, the dominant power is that of the writer himself, a writer with nearly every privilege—racial, gender, financial, educational, cultural—who censors his own writing. What about texts whose writers or subjects are subscribed to larger systems of domination, to historical forces of discrimination and oppression?

SLIDE 18: Wicked Flesh cover

To better understand the role of dominance in delimiting data formats, I turned to a historical project that has been totally dominated by hierarchical power structures. This is a history of Black women in the 17th and 18th century across the Atlantic, by Black Studies scholar Jessica Marie Johnson. Johnson’s book, *Wicked Flesh*, works from archival records like birth certificates, marriage licenses, censuses, as and travel accounts to trace the existence and lives of Black women in the early Atlantic world. These archival records, which are all written by slave-owning men, traders, and colonial officials, contain minimal information about the lives of Black women, only what these officials deemed necessary to record. From this sparse dataset, she finds methods of surfacing a history despite the incomplete information. In the same vein as other scholars of the Black Atlantic, like Saidiya Hartman, Johnson uses narrative strategies to weave a story from the fragments. Her readings of these present a complicated and nuanced picture of black womens’ lives and how they negotiated their own freedom practices within the slave-owning, male-dominated world.

From a very constraining data and structures that, on their own, tell a story of bondage and subjection to power, John constructs vivid scenes that foreground marginalized figures from the archive.

SLIDE 19: catti quote

For example, she presents a dinner party hosted by Seignora Catti, “a wealthy merchant” [who] had leveraged her status as the wife of a European official “against her commercial savvy… “for her own benefit” (Johnson 16). Catti used her knowledge of the region and her role as a hostess to strategically gain standing.

As impressive as she sounds, there appears to be no documents in the archive or in historical writings that feature her name.

SLIDE 20: notes on barbot

A glance at the notes reveals that the source material for Catti’s biography comes indirectly, from biographical writings about a commercial agent, named Jean Barbot, who was based in Senegal. Here are a list of historical and critical writings featuring Barbot, who was an apparently important figure in the French/Senegambian slave trade. In Johnson’s narrative, however, Barbot’s role is delimited to a supporting character, to a guest at Catti’s dinner party whose presence serves to bring Catti into the foreground. Johnson works within these constraints of indirect sources to weave a story from the margins of the archive, to fill its gaps of missing data.

Thinking back to the XML, I imagine this story, and others from her book, functioning like the overlap between XML elements, as entities that go beyond the bounds of what is admissible within the data format, by the precision that such a format requires. Spanning numerous fragments, I doubt that the facts of Catti’s story can be traced to its source material in direct lines, in a one-to-one correspondence.

SLIDE 21: johnson on historians

And at the end of her book, Johnson suggests as much, when she asserts that:

Historians, bound by archives, may scrape dusty folios for sources, may question whether women and girls will appear or worry that when they do appear, they emerge as legends, myths, and motifs representing more than themselves. That is not the intellectual tradition this book was written in.

Rather, Johnson’s historical project requires more than assembling the fragments that survive; it requires narrating from the overlapping forms in between them.

## disidentification to queer form

My work in the dissertation as a whole seeks out such forms, ones that figure the contour, the boundary around the elusive identities, repressed desires, and other coded elements of queerness in text.

I find that the constraints in technological processes and codes are opportunities for reading the workings of power and other oppressive structures, and in that reading, revealing not only what has been flattened or reduced by the machine, but also an incommensurable element, which cannot be measured or captured. Whether these elements are the ways that punctuation and pronouns affect meaning from a “distant reading” vantage point, or whether they operate in overlapping and permeable ways on the page.

Posing these incommensurable qualities of queerness against the necessary disambiguation of technological processes, I try to access what Queer Studies theorists Kadji Amin, Amber Musser, and Roy Perez describe as “Queer Form,” that is, “an aesthetics that moves persistently around the visual,” “mak[ing] difference a little less knowable, visible, digestible” (235). According to these theorists, Queer Form “resist[s] the dictates of transparency normally required of non-normative subjects by illuminating the unseen” (233).

SLIDE 22: image of Queer Text Encoding

I tried to demonstrate Queer Form in practice with two digital applications that I created (one for text analysis, and one for visualizing XML encoding, pictured here), which I call “the Queer Text Toolkit”. The goal with these digital projects was not to build reproducible schemas and models for analyzing queerness. Rather, it was to harness the opacity of queer forms. I see this experimental work as an attempt to, as Judith Butler puts it, “begin, without ending, without mastering, to own—and yet never fully to own—the exclusions by which we proceed” (25).

# 3. where my work is now going - algorithmic bias - 10

Now I turn to my follow-up project, which I’ve begun researching as part of my current work at the Princeton University Library.

This project applies a similar analytical spirit of deonstructing technical mechanisms toward a new object: Large Language Models. Now, I am in the middle of a data gathering and cleaning process with the goal of fine-tuning a LLM to study how it perpetuates gender bias in its training data. So far, I am have gathered definitions of gender identity and related terms from Federal legislation that limits transgender peoples’ rights, known as the “anti-trans” legislation, which is being currently debated and passed throughout the US. Next, I will use this dataset to train a Text Classification model that can score gender bias in language.

Primarily, I am interested how the training processes for these tools perpetuate and amplify the sexist, racist, and other hegemonic perspectives from their training data.

## anti trans & text generation

The anti-trans legislation movement is a fascinating one. It's being used a galvanizing tool, as a way of conscripting folks who may hesitate to condemn homosexuality into what is ultimately a hetero-hegemony, a hetero-normativizing project. It seems to me that the threat to gender has to do with the binary gender paradigm that is inalienable, that cannot be transgressed, which is used as their rallying cry. (Of course, as trans activists and scholars have been saying for years, this is only the beginning. Because once they get the trans folk, they will move on to the rest of the non-normatives.)

I think it's important that we understand, today, the threat of gender transgression. And this threat is what I'm interested in. I'm interested in examining the language outlawing gender transgression, and this particular 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, so contentious? why is it so seductive to much of the public?

So I'm going to look at the definitions of words like gender identity, using LLMs to study them. I am also interested in what we can learn about the mechanism of AI tools, and claims about it's so-called “intelligence,” and “world modelling,” as AI researchers refer to it. There's this really strange phenomenon where the LLMs are very good at generating synthetic, plausible text but not so good at improvising from details. It cannot say something original, such as explain why a particular metaphor is powerful, like a human reader can.

It is a text generator, a synthetic text generator, in the words of Emily Bender, the author of the famous (in the field of AI Ethics) paper, “Stochastic Parrots”. It can only generate what it has already seen before. Even a phenomenon like “hallucination,” this idea that a language model can “create” text that is not true, that no bearing in reality, is based on the tendency of LLMs to reproduce what they've already seen before. They hallucinate not because they are “creative”, but because they are designed from statistical processes meant to generate what is most plausible.

Ted Chiang, science fiction writer (author of the story that inspired the “Arrival” film) describes it like a blurry photograph, where the compression program guesses what pixel to use through estimation and calculation. He explains that:

common technique used by lossy compression algorithms is interpolation—-that is, estimating what’s missing by looking at what’s on either side of the gap. When an image program is displaying a photo and has to reconstruct a pixel that was lost during the compression process, it looks at the nearby pixels and calculates the average.

The compression means that the tool is guessing, estimating. What we would call an educated guess. Is prediction the same as creativity? What about improvisation? These are larger questions looming in the background.

But being in this gap, in the blur between prediction and creativity, is to study the construction of gender. What can we distill, compress, or create about the way we understand gender and sex in the juridical and carceral realms?

I hope to find something in the language, something that tells us something about the allure, the threat, the seduction, as trans scholar Cassius Adair puts it. For example, there is a widespread fear of Rapid Onset Gender Dysphoria, this idea that transness is something that transmits from trans person to cis person, like a contagion. Adair (and other trans scholars) caution that we need to get ahead of this “trans panic”: “to anticipate the emergent rhetorics of anti-trans sexual panic and draw a line in the sand around our right to find each other”.

Thank you.

FEEDBACK:

* Slow down
* Explain close reading
* Add context for violence at the start
* More signposting
* Siraj: Emphasize where the problem with the “digital” is
  + What is my intervention? What is the effect of XML? Did it work?
  + What do I mean layering syntax and meaning?
* When answering questions, be practical. Walk them through answers with examples. It’s information studies.
* Visual dimension: add aides for Orlando, Wilde, Bender, Erin Reed.
  + Use colors for definitions
  + Don’t leave Wilde up on the screen when I’ve moved on.
* Matt: why is exploring data from a humanities angle important to students of information?
  + What does the literary/sex/gender allow me to bring into Info studies?
* Add more work in Queer DH. Gesture toward it at least, for the Q&A
* What about irony?