



Unheard and Misunderstood
Tracing Hermeneutical Injustice in ADHD Narratives Generated
by Large Language Models

Miia Zhang
Supervisors: Jie Yang, Anne Arzberger
EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
26.06.2025

Name of the student: Miia Zhang
Final project course: CSE3000 Research Project
Thesis committee: Jie Yang, Anne Arzberger, Myrthe Tielman

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Abstract

This study investigates how large language models (LLMs) narrate ADHD-related experiences and whether their narrative forms give rise to hermeneutical injustice. Rather than comparing experience itself, this study analyzes how experiences are narrated. Using a hybrid coding strategy based on Reflexive Thematic Analysis, it compares LLM-generated outputs with first-person narratives from ADHD communities. The analysis identifies several recurring misnarration patterns, Truncated Subjectivity, One-Way Definition, Illocutionary Disablement, and Skewed Style Replacement. Each of these patterns constrains the interpretive space for expressing ADHD experience. Sub-themes are developed to further reveal injustice embedded in LLMs. These patterns are linked to both the training data and the optimization process. In addition, the underlying mechanism of LLMs lacks the différance structure that characterizes human narration.

1 Introduction

Imagine being called lazy, careless, or irresponsible and believing in it yourself. For many people with ADHD, these labels are not just external judgments but internalized interpretations. Why does this happen? Why are their own life experiences so often misunderstood, even by themselves? And can such misunderstanding be a form of injustice?

A commonly overlooked fact is that an event is not inherently meaningful the moment it occurs. It gains meaning only later, through interpretation and narration. Living a life is one thing; understanding it is another. When someone struggles to make sense of their own life, something has gone wrong in the process of knowing. *Epistemic injustice* refers to precisely this: the unfair conditions embedded in how knowledge is formed and distributed. A specific kind of epistemic injustice, *hermeneutical injustice*, occurs when people lack the conceptual resources needed to understand their own experiences. That is why a sentence appears again and again in online ADHD communities: "*My life finally makes sense!*" After decades of confusion and self-blaming, this is the bittersweet moment when a long-missing interpretive resource is finally acquired. This is the moment when hermeneutical injustice becomes visible, when a person can finally name their experience, and in doing so, bring light to years of confusion.

Hermeneutical injustices surrounding ADHD are not new. What is new is the presence of LLMs as powerful participants of interpretation. As these systems begin to shape how ADHD-related experiences are represented, a critical question arises and this study aims to answer the question: **What patterns of misnarration in LLM outputs contribute to hermeneutical injustice against ADHD communities?** It provides a necessary first step toward understanding and addressing such injustices. Before we can ask why hermeneutical injustices emerge in LLMs or how to mitigate them, we must first identify the forms they take. Through careful diagnosis, this research offers a conceptual foundation for future explanatory and solution-focused work.

This thesis is structured as follows: Chapter 2 provides the background and theoretical framework, introducing key concepts and reviewing related literature. Chapter 3 outlines the methodology. Chapter 4 discusses ethical considerations and reproducibility. Chapter 5 reports the result of the experiment. Chapter 6 interprets the results and explores their broader implications. Chapter 7 concludes the thesis by summarizing its contributions and limitations.

2 Background

This chapter situates the study within existing philosophical and technological discussions of epistemic injustice. It outlines four subsections: (1) the foundational concept of epistemic injustice and its relevance to AI, (2) the role of narrative in shaping and communicating experience, and (3) the specific research gap this study aims to address concerning narrative structures as mechanisms of epistemic harm. (4) the specific target group for this study.

2.1 The Concept of Epistemic Injustice

The idea of epistemic injustice was first developed by Miranda Fricker [1]. It refers to injustice related to knowledge. Specifically, when someone is unfairly treated in knowing, understanding, or communicating knowledge. Fricker identifies two key forms: testimonial injustice, where someone's words are not believed because of prejudice, and hermeneutical injustice, where someone's experience cannot be properly understood due to missing interpretive resources. The latter is the focus of this research.

Hermeneutical injustice happens when a group lacks the concepts or language to make sense of certain experiences, often because dominant groups shape which meanings are seen as legitimate. When this happens, people may not even fully understand their own experiences, or struggle to express them in ways others can hear or understand. Later researchers have extended this idea and applied it to new technologies. For example, Kay, Kasirzadeh, and Mohamed introduce the concept of *generative algorithmic epistemic injustice* to describe how generative AI systems can amplify testimonial injustice and reproduce hermeneutical ignorance[2]. Their work highlights how AI-generated content may reinforce misinformation, distort public discourse, and marginalize already underrepresented perspectives, especially in multilingual and cross-cultural settings. These developments show epistemic injustice is not only a philosophical concept, but it also happens in algorithmic systems that produce meaning and knowledge today.

2.2 Experience Through Narrative

In this study, the term "experience" is used a lot. Wilhelm Dilthey distinguishes two kinds of experiences, *Erfahrung* and *Erlebnis*—the former refers to empirical, structured, often scientific experiences; the latter to lived, immediate, subjective experience [3]. In this study, "experience" is used in the sense of *Erlebnis*, which concerns how individuals internally process and narrate events that shape their sense of self and world.

In this study, narrative serves as a starting point for examining lived experience. As Paul Ricoeur argues, narration gives form to experience, allowing individual experience to be organized and rendered meaningful over time [4]. Without such structuring, experiences may remain diffuse or unintelligible. Narratives make experiences speakable and shareable, enabling them to circulate socially and be taken up in broader processes of understanding and knowledge formation.

Having established the centrality of narrative to lived experience, narratives are now an important entry point for accessing the world of ADHD communities. Through narrating and reading narratives, individuals within these communities come to make sense of their own experiences and those of others. By examining how these stories are told, or fail to be told, we can trace the signs of hermeneutical injustice: the structural conditions that make experiences excluded from shared understanding.

2.3 Research Gap: Narration Patterns as Micro-level Diagnosis

Previous studies have proposed several frameworks for understanding hermeneutical injustice. For instance, Dotson identifies three types of epistemic exclusion[5], highlighting different levels at which knowledge systems fail to accommodate marginalized experiences. Medina offers four analytical parameters(source, dynamics, breadth, and depth)[6], that capture various aspects of interpretive injustice, ranging from the absence of naming to structural barriers in communication. These models offer important tools for analyzing how hermeneutical injustice arises and operates at a broad level.

However, these approaches often combine causes, conditions, and effects into general categories. What they do not examine in detail is how specific forms of narrative structure can themselves contribute to epistemic harm. While we know that some experiences are misrepresented or fail to gain recognition, the concrete ways in which such misnarration happens remain underexplored.

This study addresses that gap by focusing on the level of narrative form. It offers a finer-grained method for diagnosing how lived experiences are reshaped and flattened through narrating. In doing so, it complements existing frameworks by shifting attention to the micro-level mechanisms through which hermeneutical exclusion can take place.

2.4 ADHD and Hermeneutical Injustice

To examine these micro-level narrative distortions in a concrete context, this study focuses on ADHD-related personal accounts as its primary domain of analysis. ADHD is short for Attention deficit hyperactivity disorder. Adults with ADHD often face difficulties in articulating their experiences, partly due to the lack of socially available interpretive resources. Many do not receive a diagnosis until later in life, if at all, and spend years attributing their struggles to personal failure. Asherson et al. note that even high-functioning individuals frequently expend excessive energy to mask impairments, often experiencing chronic distress, mood instability, and low self-esteem[7]. These symptoms are commonly misunderstood or minimized, as cultural expectations and inadequate diagnostic frameworks obscure the reality of adult ADHD.

As many people with ADHD engage in long-term masking without access to adequate interpretive resources, they are especially vulnerable to hermeneutical injustice. Their experiences remain distorted not only by others, but also in their own attempts at self-understanding. This makes them a fitting target group for studying how LLMs handle experiences that are already epistemically marginalized.

3 Methodology

This chapter outlines the methodological approach used to investigate the research question. It is divided into four parts. Research Design introduces the qualitative strategy adopted by this study. Experimental Design describes how the experiment is designed. Data Analysis explains how the collected data is interpreted using thematic methods. Initial coding framework introduced the initial main themes developed from literature.

3.1 Research Design

This study aims to answer the question: What patterns of misnarration in LLM outputs contribute to hermeneutical injustice against ADHD communities? To do so, it adopts a qualitative interpretive design. While philosophical accounts of hermeneutical injustice identify conceptual conditions under which interpretive harm arises, they offer limited insight into how such harm is produced through specific narrative forms. This study addresses that gap by grounding its analysis in a close comparison between LLM-generated outputs and first-person narratives authored by individuals with ADHD.

The focus is not on content accuracy, but on narrative form: how meaning is structured. This contrasts with many existing LLM evaluations, which emphasize factual correctness or syntactic fluency but neglect how models handle narrative structure. The study situates its method within this underexplored dimension, examining whether LLMs constrain interpretive possibility through recurrent narrative patterns that flatten, moralize, or erase lived complexity.

3.2 Experimental Design

Thematic Comparison. To examine how LLMs narrate ADHD-related experiences, this study compares model-generated outputs with first-person narratives authored by individuals with ADHD. Prompts were designed to elicit outputs that mimic the tone and structure of community narratives, covering commonly discussed topics. For each topic, multiple completions were generated using GPT-4o to capture variation.

Community-authored texts were sourced from Reddit and selected based on narrative depth and relevance. For each topic, a corresponding set of model outputs was matched with Reddit narratives to enable close, source-aligned comparison.

These community-authored narratives serve as the interpretive baseline not because their factual claims are verified, but because they are genuine human narrations. This study does not aim to fact-check LLM outputs against external reality, but to assess whether their narrative forms align with how people actually tell their own stories. The baseline thus reflects lived, subjectively meaningful accounts, grounded in first-person perspective.

Data Collection from Online Communities To establish a reference point for comparison, first-person narratives authored by individuals with ADHD were collected from public online sources. These posts are collected from Reddit (e.g., r/ADHD, r/ADHDwomen). Texts were selected based on relevance, length, and narrative depth. The inclusion criteria required that the texts be written in the first-person, reflect lived experience, and address concrete aspects of life with ADHD.

To identify common experiential themes, a preliminary thematic scan was conducted on ADHD-related subreddits. Posts were manually reviewed, and frequently discussed topics were noted in Table 1. Each selected text was assigned a unique identifier and tagged by topic, allowing structured comparison with LLM outputs. All texts were compiled into a structured spreadsheet with identifiers.

ID	Topics
A	Relationships / social life
B	Daily life
C	Work / study
D	Diagnose

Table 1: Topic Labels (ID A–D)

Data Collection from LLMs This research used GPT-4o model, as it is widely used and integrated into public applications, making it a representative agent of how LLMs currently participate in shaping social narratives. To enable meaningful comparison, each prompt was written to closely mimic the structure and tone of ADHD community narratives. Prompts were phrased in the first-person and focused on specific topics previously identified through community posts. The aim was to elicit responses that could be directly compared to real-life narratives, both in voice and thematic content.

To capture a diverse range of ADHD-related voices, six distinct personas were designed. Each persona combined attributes such as gender, age, role and diagnostic status. A full table of persona specifications and an example prompt are provided in Appendix A. To capture the range of potential outputs, each of the prompt was submitted three times. And it was ensured that each prompt was processed independently, without memory carryover or contextual influence from previous inputs. Temperature was set at 0.7 and top-p at 1.0, balancing coherence with variation. All completions were preserved for analysis in order to keep the variation. Each response was labeled with a prompt ID , topic ID and sample number, then stored for later analysis. Outputs were not edited or post-processed.

3.3 Data Analysis

Reflexive Thematic Analysis. The data analysis approach used by this study is inspired by Reflexive Thematic Analysis (RTA) as developed by Braun and Clarke[8], due to its strong alignment with the epistemological and conceptual demands of the research.

RTA is grounded in interpretivism and constructionism, which understand meaning not as something objectively embedded in texts, but as something constructed in context through interactions among language, individuals, and researchers. This makes RTA especially well-suited for analyzing hermeneutical injustice, a phenomenon rooted in gaps within shared interpretive resources and the resulting barriers to understand lived experience.

In contrast, other common methods such as Codebook Thematic Analysis and Content Analysis are based on positivist assumptions. They tend to treat meaning as an observable object, rely on predefined coding frameworks. These methods prioritize reproducibility, but at the cost of reducing the human’s interpretive role and ignoring the contextual nature of meaning-making.

Procedures. This study uses a hybrid coding strategy that combines deductive and inductive approaches. The main themes were established deductively, based on existing theories of hermeneutical injustice. These themes provided the initial structure for organizing the data. Within each theme, sub-themes were identified inductively through close reading of the narratives. This allowed the analysis to remain open to recurring patterns that were not explicitly defined by prior theory. The process is as follows:

Topic grouping. All narratives from Reddit and LLM outputs were first grouped by topics identified earlier in Table 1. This allowed for aligned comparison across narrative types responding to similar experiential domains.

Familiarisation with data. Each narrative was read multiple times. Open-ended notes were taken on tone, emotional structure, framing devices, and moments of tension or contradiction.

Exploring narrative patterns. Within each topic, recurring features in LLM outputs were compared with those in community-authored narratives. Shared narrative structures, vocabulary patterns, and emotional tones were analyzed.

Application of theoretical codes. A set of several misnarration categories was developed based on existing theories of epistemic injustice. These were used as guiding lenses to examine the data. When a pattern in the LLM output aligned with one of these categories, it was marked.

Organising themes and sub-themes. The identified misnarration patterns were refined into main themes and related sub-themes. Each sub-theme was further defined based on recurring features observed across samples. While the themes were theory-driven, sub-themes were derived inductively from the narrative material.

The analysis was conducted at topic levels. Texts were grouped by topic, and for each topic, LLM and community-authored narratives were compared in terms of how experiences were framed, interpreted, or omitted. The goal was not to analyze isolated sentences, but to examine how meaning was constructed across full responses.

3.4 Initial Coding Framework

This experiment begins by establishing an initial analytical framework. Based on the literature study, six recurring patterns of how the experiences could be narrated improperly were identified as a basis for qualitative coding.

Lack of Naming. Fricker’s foundational account of hermeneutical injustice highlights the epistemic harm incurred when certain experiences lack the conceptual vocabulary needed for recognition and communication [1]. In this case, people have the experience but no words to name or explain it, so their stories remain vague or incoherent.

Truncated Subjectivity. Carel and Kidd emphasize how patients are often reduced to data points, denied epistemic agency in the interpretive process [9]. People’s subjectivity are truncated in this sense. The speaker’s first-person perspective gets ignored, they are seen as objects to be analyzed, not as narrators of their own experience.

One-Way Definition. Scrutton critiques how medical frameworks often monopolize interpretive authority, marginalizing lived experience in favor of predetermined diagnostic categories [10]. Here, experience is forced into authorized description, and whatever doesn’t fit gets excluded from the narrative.

Illocutionary Disablement. Relatedly, speech acts may be rendered entirely inert, receiving no uptake from the listener [11]. In such cases, the speaker is not merely misinterpreted but entirely silenced as a knowing subject. What they say does not count as meaningful at all, it’s ignored as if they never spoke.

Skewed Style Replacement. Potter describes how dominant diagnostic or clinical vocabularies often overwrite the speaker’s native expressive style [12]. Individuals may adapt their self-descriptions to conform to mainstream frameworks, resulting in distorted self-representation. To be understood, people change how they speak, and in doing so, they lose the original form of their experience.

These patterns focus on *how* experiences are misnarrated, not *what* experiences are missing or *why* they are missing. In the following chapter, the considerations of this study is discussed.

4 Responsible Research

This chapter reflects on the ethical dimensions, interpretive position, and reproducibility of the study. The first part outlines how ethical use of community data and model outputs was ensured. The second part is a positionality statement from the researcher, reflecting on the analysis’s relation to personal background. The third part discusses the limits and strategies of reproducibility in the qualitative experiment.

4.1 Ethical Considerations

All community-authored data used in this study was sourced from publicly accessible Reddit posts, in accordance with Reddit’s Public Content Policy ¹. No usernames, URLs, timestamps, or identifiable metadata were collected. Posts were anonymized to prevent re-identification and were used for non-commercial academic research. Deleted content, private communities, and personal communications were excluded.

Special attention was paid to the representation of individuals with ADHD. The study avoids deficit-based framings and does not aim to generalize the experiences of people with ADHD. Rather, the analysis focuses on identifying interpretive patterns in textual representations. All LLM-generated content was used for analytical comparison only and is not intended to substitute or simulate lived experience. No clinical or diagnostic claims are made.

4.2 Positionality Statement

I approach this project as someone outside the ADHD community, but several people closed to me do. I have had many conversations with them, listening to their reflections and frustrations. Their perspectives have shaped my interest in this topic and helped me to better understand their context.

While reading large amounts of community posts, I found myself often emotionally involved. I felt related, not because I shared the same symptoms, but because the struggle to make sense of oneself through language is something I deeply relate to. This emotional similarity has helped me read attentively, but I am also aware that it can bias how I interpret texts. I do not assume access to others’ experiences, and I try to stay cautious about reading my own reactions into their words. I also paid close attention to whether themes emerged across multiple texts, rather than from moments of personal resonance.

¹<https://support.reddithelp.com/hc/en-us/articles/26410290525844-Public-Content-Policy>

4.3 Reproducibility and Transparency

This study adopts a qualitative, interpretive framework, where reproducibility does not mean producing identical results, but rather enabling transparency and traceability. All prompts used to generate model outputs are documented, along with the fixed parameters and the number of completions per prompt. Outputs were preserved in full, versioned, and labeled to maintain a clear audit trail.

Themes were derived through a reflexive thematic analysis process, with emphasis on transparency rather than mechanical replication. Themes were grounded in multiple examples across different topics in both community-authored and LLM-generated texts. Although exact replication of model outputs is limited due to their stochastic nature, the analytic process, interpretive decisions, and underlying assumptions are made explicit to support conceptual reproducibility.

5 Result

This chapter presents the major forms of misnarration found in the LLM outputs, categorized under the main themes identified in the previous chapter. Each category includes several recurring sub-themes observed across multiple samples and examples from Reddit and LLM outputs.

5.1 Truncated Subjectivity

Template-Imposed Experience

Reddit narratives exhibit diversified emotional tones, narrative structures, and expressive styles, often preserving ambiguity, tension, or unresolved feelings. In contrast, LLM-generated texts tend to follow a fixed three-part structure: Problem – Insight – Resolution. Difficulties are first described, then framed through a diagnostic or psychological lens, and finally resolved through personal growth or emotional reframing. A typical example is:

*Having adhd has been a rollercoaster when it comes to friendships and social interactions... Moreover, my **emotional dysregulation** means I might react more intensely... **Despite these challenges**, I've learned to communicate openly... the support and patience from true friends have been invaluable. (LLM, 1A1 ²)*

This narrative follows a formula: a challenge is presented, then medicalized, then end with a positive outcome. The result is a constrained and overly resolved narrative, where struggle becomes a tool for insight and closure. Because all narratives follow the same structure, the differences in how people actually experience and process ADHD are lost. This is a case of Truncated Subjectivity: the speaker's emotional and narrative range is shortened, edited into a neat arc rather than an open or uncertain process.

Emotional Flattening

Reddit narratives often convey strong emotional tension. In contrast, LLM outputs tend to maintain a calm, neutral tone, even when describing experiences of long-term suffering or systemic failure. Consider these two narratives of late diagnosis:

²Each data point is labeled using a identifier: persona ID (1–6), topic ID (A–D), and for data from LLMs sample number (1–3).

It took 10 years for me to finally get an evaluation through Kaiser. I was asking for 10 years and they kept telling me they don't give adult evaluations for ADHD. But after 10 years, they finally gave me one! And I walked right into the office and three hours later I was diagnosed with having a severe case of ADHD. But I had to suffer my entire life (I'm 45 now). I can't believe how they can let me sit and suffer for 10 years but they did. (Reddit, D2)

After years of feeling out of sync, I finally have clarity. At 52, I was diagnosed with ADHD...This newfound awareness is bittersweet, but it brings relief. It's empowering to know there's a reason behind the chaos, and I'm eager to explore strategies to manage it better. This journey is just beginning, and I'm hopeful about the future. (LLM, 4D3)

The Reddit narrative is emotionally intense, emphasizing long-term neglect. The LLM version reframes the similar experience as a personal journey toward understanding, using soft language and future-oriented optimism. Here, the intensity and frustration of delayed diagnosis are replaced by calm, structured reflections that reduced anger and emotional urgency. The result is a loss of narrative realism. This sub-theme fits under Truncated Subjectivity because it limits the range of permitted emotions and excludes unresolved affect from the narrative space.

5.2 One-Way Definition

Symptom Template Overcoding

LLM outputs often describe interpersonal and emotional struggles using a limited set of diagnostic terms, regardless of the speaker's background or context. As a result, different personas, such as a 21-year-old student or a 50-year-old blue-collar worker, are made to appear as if they are dealing with identical problems in identical ways. In contrast, Reddit posts reflect greater variability in how people frame and interpret their difficulties. The following examples show how LLM narratives repeatedly fall into a narrow vocabulary of forgetfulness, impulsivity, and emotional dysregulation:

*Sometimes, I struggle with emotional regulation, which makes me react more **intensely** than others might expect... I often **forget plans** or find it hard to maintain consistent communication... (LLM, 1A2)*

*Maintaining friendships requires extra effort, as I can be **forgetful** or **impulsive**, leading to **canceled plans**... (LLM, 5A3)*

Despite demographic and contextual differences, the same diagnostic framing is applied. This results in different life stories being squeezed into the same clinical template, reducing the perceived individuality of each narrative. This sub-theme exemplifies One-Way Definition by imposing a single explanatory framework onto all forms of difficulty, excluding alternative interpretations or contexts.

Causal Reframing

Reddit narratives often situate individual struggles within broader social or structural contexts. Emotional and behavioral challenges are framed as the result of dynamic interactions between internal states and external environments. In contrast, LLM outputs tend to erase external causes, attributing all difficulties solely to internal deficits. This divergence is particularly clear in the following excerpts on workplace experience:

On the outside, I appear organized and efficient, but beneath this facade, there’s a constant whirlwind of thoughts and distractions... Meetings often feel overwhelming, as I struggle to maintain focus and resist the urge to interrupt or drift away. (LLM, 5C1)

Things start off in a new job well enough, and then colleagues—almost always people without ADHD—take a dislike to me... Eventually, a complaint is raised. It never comes with evidence, and it’s always either a lie or a deliberate distortion. (Reddit, C3)

In the LLM output, workplace difficulties are explained through internal traits, while the Reddit narrative points to social exclusion and systemic prejudice. The LLM version thus erases external causality, privatizing struggle. This sub-theme shows One-Way Definition by removing social or structural explanations, reinforcing a one-way model of explaining the experiences.

5.3 Illocutionary Disablement

Narrative Toolification

In many LLM outputs, personal experience is used as a tool for delivering some lessons or insights. The narrative is not about expressing lived experience in its complexity, but about serving a pre-defined arc of growth or resolution. The speaker becomes a tool for inspiration, rather than an agent of self-expression. Consider the following example:

*Navigating relationships with ADHD can feel like riding a rollercoaster with no seatbelt. My emotions surge like ocean waves, often leaving others struggling to understand the intensity. It’s not just about being overly enthusiastic or impulsive; it’s the constant whirlwind of thoughts and feelings that can be overwhelming for both me and those around me. I often find myself hyper-focusing on someone, showering them with attention and affection, only to get distracted by the next shiny thing that captures my interest. This can unintentionally hurt those who feel neglected when my focus shifts. Communication becomes a delicate dance, as I struggle to articulate the depth of my feelings without overwhelming others. **Yet, these challenges have taught me the importance of honesty and patience. I’ve learned to surround myself with those who understand my ebb and flow, and who appreciate the vibrant chaos I bring.** ADHD makes my relationships unique, intense, and sometimes challenging, but it also fuels my creativity and passion, which I channel into my art and connections with others. (LLM, 3A3)*

The speaker’s emotional reality is overwritten by a moral function: to turn pain into strength. The ending reframes the story so that its value lies not in what was felt, but in what was learned. This is a case of Illocutionary Disablement: the speaker is not allowed to just express contradiction, confusion, or emotional stuckness. Their narrative act is shaped into a motivational arc, disabling the possibility of simply expressing unresolved experience.

5.4 Skewed Style Replacement

Premature Positivization

Reddit narratives often reflect complex emotional responses, most of them negative, with many ending in expressions of anger, frustration, or bitter irony. Words like exhausted, ashamed, defeated, or furious appear frequently. In contrast, LLM outputs tend to end with a hopeful note, regardless of the severity of the struggle. Consider the following examples:

...Despite the challenges, I persevere, driven by the desire to excel and the hope that one

day I might feel truly at ease in my own mind. Sharing this struggle is daunting, but it's a step toward embracing my reality and finding strategies that align with who I am. (LLM, 5C3)

...On the flip side, ADHD also means I bring energy and creativity to my relationships, and I deeply appreciate the friends who accept me as I am. I'm learning to manage my symptoms better, but it's a work in progress, and having understanding friends makes all the difference. (LLM, 1A3)

These endings shift quickly from difficulty to growth, emphasizing acceptance, support, or self-discovery. In the Reddit data, such forced positivity is rare. Most posts end in ambivalence or emotional weight rather than resolution, making LLM narratives appear prematurely optimistic and emotionally unearned. This sub-theme illustrates Skewed Style Replacement because it replaces the chaotic or unresolved nature of real narratives with a calm tone that aligns with conventional ideas of how people should grow from difficulties.

This sub-theme is closely related to *Template-Imposed Experience*, as both reflect the structural tendency of LLM outputs to organize ADHD narratives into fixed arcs of difficulty, insight, and resolution. However, while *Template-Imposed Experience* concerns the overall narrative structure, *Premature Positivization* focuses specifically on tonal closure. It describes the pattern whereby narratives end on an emotionally positive or hopeful note, regardless of the preceding content.

Clichéd Metaphorical Flattening

LLM outputs frequently rely on clichéd metaphors such as rollercoaster, marathon, or storm to describe lived experiences. While such phrases appear expressive on the surface, their repetition across samples renders them formulaic and emotionally empty. In Reddit narratives, by contrast, metaphorical language is rare. Examples include:

Each day is an unpredictable dance... (LLM, 3B2)

Every day at work feels like running a marathon with invisible weights... (LLM, 5C1)

Navigating the professional world with ADHD has been a rollercoaster... (LLM, 5C2)

These metaphors substitute individual experiences with generalized imagery. They smooth over ambiguity, conflict, or contradiction, replacing specific emotional texture with templates. In this sense, they show patterns of Skewed Style Replacement: the replacement of complex, context-specific narrative voice with polished, stylistic conventions. The result is that emotions like confusion, bitterness, or resentment are removed from the narrative entirely.

5.5 Distribution Summary

Table 2 provides an overview of how each main theme and its sub-themes are distributed across the four narrative topics (A–D) identified in Chapter 3.

Main Theme	Sub-theme	A	B	C	D
Lack of Naming	—				
Truncated Subjectivity	Template-Imposed Experience	✓	✓	✓	✓
	Emotional Flattening	✓	✓	✓	✓
One-Way Definition	Symptom Template Overcoding	✓	✓	✓	✓
	Causal Reframing		✓	✓	✓
Illocutionary Disablement	Narrative Toolification	✓	✓	✓	✓
Skewed Style Replacement	Premature Positivization	✓	✓	✓	✓
	Clichéd Metaphorical Flattening	✓	✓	✓	

Table 2: Main and Sub-theme Presence across Topics A–D

Most of the themes appear across almost all topics, suggesting that the misnarration patterns identified are not isolated but structurally embedded in LLM outputs. Notably, Lack of Naming do not appear in any of the four topics. The absence of this pattern will be examined in detail in the following chapter.

A few exceptions stand out. For instance, Causal Reframing does not appear in Topic A (Relationships and Social Life), likely because this topic itself tends to invite internal attribution. Unlike workplace or diagnostic settings, interpersonal narratives more often emphasize personal traits, emotional patterns, or communication styles rather than structural causes.

Overall, apart from Lack of Naming, most misnarration types are widely distributed across topics

6 Discussion

This chapter reflects on the implications of the misnarration patterns identified in the previous chapter. First, it examines how the absence or presense of the misnarration forms can be traced to LLM training data and optimization processes. Then, it turns to human narration as a point of contrast to highlight the deeper limits of LLMs.

6.1 The Structural Absence of Lack of Naming

Among the six main types of misnarration identified in this study, Lack of Naming is the only one that did not appear in any of the LLM outputs. This absence might seem to suggest that the model succeeds in naming and recognizing relevant experiences. However, a closer look reveals that the pattern is missing not because the problem has been solved, but because it has been structurally avoided.

LLMs generate text based on linguistic patterns present in their training data. As a result, they can only describe experiences that have already been named and circulated within public discourse. If an experience is vague, emotionally ambiguous, or conceptually underdeveloped, it is unlikely to appear in model output. As Bender put it, a language model is best understood as a “stochastic parrot”: a system that “haphazardly stitching together sequences of linguistic forms it has observed in its vast training data... without any reference to meaning” [13]. In this sense, the model does not generate language for unfamiliar experiences—it omits them entirely.

This finding extends Fricker’s account of hermeneutical injustice [1], which occurs when individuals lack access to shared interpretive resources. Here, the injustice is compounded:

the model’s reliance on pre-existing discourse limits the range of narratable experience, and its outputs, which often integrated into broader knowledge system, become part of the interpretive environment themselves. Thus, LLMs both reflect and reinforce hermeneutical injustice.

6.2 Why Misnarration Emerges: Constraints from LLM Training Data and Optimization Process

While "Lack of Naming" reflects a structural omission, other forms of misnarration emerge as a result of constraints imposed by model training. These patterns appear consistently across different topics and personas, suggesting that they are not isolated design artifacts but rather systemic features of the model.

Evidence for these constraints is illustrated by the narrative patterns observed in the model’s outputs. Regardless of age, gender, or other context, LLM-generated narratives tend to share similar emotional trajectories and lexical frames. Experiences are often organized into structured sequences of difficulty, insight, and affirmation. Emotional intensity is muted, and language tends to converge around generalizable clinical terms. This narrative homogeneity contrasts with the diversity and ambiguity of Reddit posts, where users express hesitation, contradiction, and unresolved emotions.

One significant source of these constraints lies in how training data is curated. As Bender points out, common filtering practices systematically remove language associated with marginalized identities. Terms reclaimed by LGBTQ communities, along with other forms of non-normative or affectively charged expression, are often excluded under broad criteria of offensiveness or unintelligibility [13]. This reduces the model’s exposure to the voices of minority groups and contributes to the suppression of narrative styles and voices that depart from dominant cultural norms.

Further, LLMs are trained using reinforcement learning from human feedback (RLHF), which encourages the model to produce responses rated as helpful, safe, and emotionally appropriate [14]. In practice, this penalizes affective expressions like anger, despair, or uncertainty, states that appear frequently in authentic ADHD narratives. The model tends to favor resolution and closure, it defaults to endings that emphasize growth. This tendency, labeled in previous chapter as *Premature Positivization*, reflects a training regime that rewards emotional stability over emotional realism.

Together, these tendencies align with what Carel and Kidd describe as *Truncated Subjectivity*: when complex inner lives are reduced to simplified, externally legible categories [9]. As a result, lived experience of marginalized groups becomes harder to express and understand as their complexity is flattened by the model’s preferred narrative style.

6.3 Narration and Différance

The previous sections identified how the misnarration forms emerge from the data and optimization process. But to fully understand what is missing in LLM-generated narration, it is also necessary to consider how human narratives work. This section shifts focus from model mechanisms to the structure of human narration itself, in order to highlight the underlying capacities that LLMs fail to reproduce.

Human narration is often messy, recursive, and unfinished. Speakers hesitate, revise and discover meaning as they speak. This is clearly visible in Reddit narratives but almost entirely absent from LLM outputs. As one user said:

*Regarding one (now dear) friend, I didn't think of us as any more than acquaintances until one day, having known each other for over a year, mostly through social circles, she invited just me for a midday craft beer, an invitation that I accepted. We had a few and played chess (checkers maybe?), during which she told me she was glad we were friends and the kind who could play chess at a bar midday. I remember being taken aback and my thoughts on what I considered to be our acquaintanceship came out. She was maybe more flustered than offended, and luckily for me, she, being the forward and somewhat abrasive person she is, informed me of her disbelief and **asserted that we are in fact friends and had been for awhile**. I accepted this as fact, and it's been **a very good friendship ever since (and apparently before)**! (Reddit, A6)*

What is interesting about this post is that the speaker went for a drink with someone and was told they were already friends. She was genuinely surprised and only realized afterward that they had, in fact, been friends for quite some time. Here, understanding unfolds slowly with a delay. This open, unfinished nature of human narration is what Derrida calls *différance*, a concept that captures both "difference" and "deferral"[15]. According to Derrida, meaning is never fully present in a word or sentence. It is always delayed. In this view, narration is not a container for meaning, but a space where meaning happens slowly, often incompletely.

This structure is largely absent in LLM-generated outputs. LLMs are optimized to produce fluent and coherent responses based on surface-level patterns in the training data. As Bender and Koller note, LLMs often exploit surface-level patterns without internalizing the abstract relations they appear to express. This reliance on distributional cues helps explain why model-generated narratives may resemble human ones in form, yet diverge in interpretive depth [16]. This reflects a system built for statistical reproduction, not for exploratory narration.

Several misnarration patterns follow from this. *Template-Imposed Experience* reflects the model's reliance on established narrative forms. *Premature Positivization* mirrors the tendency to select emotionally resolved endings. *Narrative Toolification* arises when narration is used to fulfill expected communicative functions rather than to express open-ended experience. These are structural mismatches between the statistical logic of LLMs and the open structure of human narration (what Derrida calls *différance*).

7 Conclusions and Limitations

This study explored how LLMs narrate ADHD-related experiences and whether their narrative forms introduce hermeneutical injustice. By comparing LLM outputs with community-authored narratives, it identified several recurring misnarration patterns, including Truncated Subjectivity, One-Way Definition, Illocutionary Disablement, and Skewed Style Replacement. These patterns constrain on how experience can be expressed and interpreted.

LLMs produce narratives that follow templates, often emotionally flattened, prematurely resolved, and diagnostically overcoded. These outputs reflect a narrow standard of narrative legitimacy: they define which stories count as coherent, valuable, or complete. In doing so, they restrict the range of lived experience that can be meaningfully told.

Notably, the study did not find cases of Lack of Naming, not because this problem has been resolved, but because LLMs are unlikely to generate experiences that lack recognizable patterns in their training data.

In tracing the causes of these patterns, this study linked misnarration to training data

and optimization process, also to the model’s underlying narrative logic. Whereas human narration allows meaning to emerge slowly, LLMs often exploit surface-level patterns without internalizing the abstract relations they appear to express.

Taken together, these findings show that LLMs actively construct some mis-narrative forms. This study contributes a detailed account of how narrative conventions embedded in LLM outputs may limit the epistemic space available to marginalized communities. It offers a framework for understanding narrative distortion and provides a conceptual link between narrative structure and hermeneutical injustice.

This study also has several limitations. It focused on ADHD-related prompts and used Reddit posts as a comparison baseline, which may limit the generalizability of findings to other topics or communities. The six misnarration forms are now presented as separate categories, but their relations and conceptual boundaries remain underexplored and should be addressed in future work.

References

- [1] M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.
- [2] J. Kay, A. Kasirzadeh, and S. Mohamed, “Epistemic injustice in generative ai,” 2024.
- [3] W. Dilthey, *Poetry and Experience*. Princeton: Princeton University Press, 1996.
- [4] D. Pellauer and B. Dauenhauer, “Paul ricoeur,” 2025. Stanford Encyclopedia of Philosophy.
- [5] K. Dotson, “A cautionary tale: On limiting epistemic oppression,” *Frontiers: A Journal of Women Studies*, pp. 24–47, 2012.
- [6] J. Medina, “Varieties of hermeneutical injustice,” in *The Routledge Handbook of Epistemic Injustice*, Routledge, 1st ed., 2017.
- [7] P. Asherson, R. Akehurst, J. J. S. Kooij, M. Huss, K. Beusterien, R. Sasané, S. Gholizadeh, and P. Hodgkins, “Under diagnosis of adult adhd: Cultural influences and societal burden,” *Journal of Attention Disorders*, vol. 16, no. 5_suppl, pp. 20S–38S, 2012. PMID: 22377849.
- [8] V. Braun and V. Clarke, *Successful Qualitative Research: A Practical Guide for Beginners*. SAGE Publications, 2013.
- [9] H. Carel and I. J. Kidd, “Epistemic injustice in healthcare: A philosophical analysis,” *Medicine, Healthcare, and Philosophy*, pp. 529–540, 2014.
- [10] A. P. Scrutton, “Epistemic injustice and mental illness,” in *The Routledge Handbook of Epistemic Injustice*, p. 350, Routledge, 1st ed., 2017.
- [11] L. Anderson, *Epistemic Injustice and the Philosophy of Race*, pp. 139–148. Routledge Handbooks in Philosophy, United States: Routledge, Mar. 2017.
- [12] N. Potter, “Empathic foundations of clinical knowledge,” in *The Oxford handbook of philosophy and psychiatry* (K. W. M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, and T. Thornton, eds.), Oxford University Press, 2013.

- [13] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, “On the dangers of stochastic parrots: Can language models be too big?,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’21, (New York, NY, USA), p. 610–623, Association for Computing Machinery, 2021.
- [14] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, “Training language models to follow instructions with human feedback,” in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS ’22, (Red Hook, NY, USA), Curran Associates Inc., 2022.
- [15] J. Derrida, *Differance*, pp. 467–469. Montreal: McGill-Queen’s University Press, 1990.
- [16] E. M. Bender and A. Koller, “Climbing towards NLU: On meaning, form, and understanding in the age of data,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, eds.), pp. 5185–5198, Association for Computational Linguistics, July 2020.

Appendices

A Persona Profile

To generate LLM outputs that approximate the narrative diversity observed in ADHD communities, six persona profiles were constructed. These were used to condition the prompt design.

ID	1	2	3	4	5	6
Gender	Female	Male	Non-binary	Male	Female	Other
Job	Student	Blue-Collar Worker	Tattoo Artist	Retired	White-Collar	Student
Age	21	34	27	52	30	15
Status	Self-Aware	Minimal Formal Diagnosis	Emotionally Expressive	Late Diagnosis	Masked for Years	Undiagnosed, Seeking Validation

Table 3: Persona Profiles

Prompt Example: *You are a 15-year-old student who suspects you might have ADHD. You’re not sure what to think, and no one at home really talks about it. Write a post about your confusion or curiosity about yourself.*

B Use of Large Language Models

During the writing process, I discussed some of my ideas with LLM and used it to assist with refining grammar. I also used the model to generate some LaTeX commands for formatting purposes. All the chapters were authored by me. No verbatim output from LLM is included.