# What's the Story?

## Using Large Language Models for Policy Narrative Content Analysis at Scale

Master's Thesis for the Managment of Technology Program

Darsh Modi

Delft University of Technology

TU Delft

# What's the Story?

## Using Large Language Models for Policy Narrative Content Analysis at Scale

by

Darsh Modi

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Wednesday August 28, 2024 at 15:00.

| | | |
|---|---|---|
| Project duration: | February 28, 2024 – August 28, 2024 | |
| Thesis committee: | Dr. Nihit Goyal, | TU Delft, first supervisor |
| | Prof. dr. ir. Ibo van de Poel, | TU Delft, chair and second supervisor |

Cover: Generated by DALL-E (OpenAI)

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

TU Delft

# Preface

I would like to express my gratitude to my supervisors, always present during my thesis, and my academic journey. To Dr. Nihit Goyal, for his time and patience during these last 6 months while guiding me. To Dr. Ibo van der Poel, for his extensive feedback. To my friends and family who helped me keep going. To my parents, who paved the way for me to be able to get higher education and always supporting me, no questions asked.

As I write this, I realise that this will be the last few weeks of my formal education, something I will dearly miss, despite its numerous ups and downs. Although I won't be enrolled in school any more, I hope to always stay a student.

*Darsh Modi*
*Delft, August 2024*

# Executive Summary

Narratives are central to human communication, serving as the primary means by which we convey information. They permeate every aspect of human life and influence our lifestyles. Narratives also play a crucial role in the policy formation process. Discussions on complex topics often lead to commonly accepted practices, that eventually evolve into policies through debate and iterative refinement. Therefore, studying these narratives is essential for a deeper understanding of policy issues and for developing effective strategies to guide the process.

With the exponential population growth following the Industrial Age and the subsequent Information Age, there has been a proliferation of diverse narratives and platforms for their dissemination. Traditional manual methods of analyzing policy narratives frequently employed in qualitative research, such as content analysis and qualitative coding, are no longer adequate to consume the amount of available data. The size of data on the internet has grown at a rate of approximately 42% per year between 2010 and 2020, and the rate of its growth is only increasing. Thus, automated methods are necessary to analyze this 'big' data effectively.

As a potential solution for this task, this research examines the applicability of using natural language processing (NLP) techniques, specifically large language models (LLMs), for narrative policy analysis. Large language models are mathematical models of natural language trained on a large amount of language data to predict the next most probable word for a given input. Although it is a solution to a common NLP problem, emergent abilities of reasoning (confirmed by tests of math and common sense) make large language models a viable candidate for this task. The study involves qualitatively coding a dataset of policy narratives, manually and automatically, using a codebook derived from the Narrative Policy Framework (NPF)[1], and comparing the results. The codebook contains definitions of categories for elements such as plots, narrative strategies, belief systems, etc. and characters like heroes, villains, victims, etc. that are identified in the narrative data. The central research question of the study is:

**To what extent do Large Language Models (LLMs) accurately automate qualitative coding of policy narratives when compared to a manually coded dataset?**

The research question necessitates a research design that involves a comparative analysis between the outputs generated by a Large Language Model (LLM) on a specified classification and entity recognition task, and those produced by a human performing the same task. In this context, the human-generated output is regarded as the 'gold standard,' which is a widely accepted benchmark in qualitative research. The primary objective of this study is to evaluate the performance of LLMs on a complex task related to policy narrative analysis. Consequently, a case study approach, combined with systematic content analysis, has been selected as the most appropriate methodology.

The case study centers on the issue of air pollution, a critical concern in developing nations, with the setting being India. Despite the established scientific consensus on the adverse health impacts of air pollution, India has exhibited a lack of adequate policy responses. Notably, Indian cities consistently

[1]The NPF is a comprehensive framework in policy studies for analyzing how policy narratives influence policy processes, designs, and outcomes at multiple levels of analysis.

rank among the most polluted globally. By examining policy narratives surrounding air pollution, this study aims to uncover the complexities of the issue, assess the current state of air pollution policy in India, and evaluate proposed actions to inform future strategies for addressing this challenge.

The policy narratives under investigation are extracted from textual content in news articles sourced from an online newspaper. The search was conducted using the keywords 'air pollution' and 'India.' These narratives, when analyzed through the lens of the Narrative Policy Framework, will provide valuable insights into the policy problem, with the expectation of revealing its causes, mitigation efforts, affected stakeholders, and the positions of supporting and opposing entities, as well as the narrative structures employed.

The research design encompasses the following components:

1. **Dataset Retrieval and Codebook Generation** - The dataset of newspaper articles from the newspaper "The Times of India", the fourth largest in India by circulation (Audit Bureau, 2023) and the most circulated daily English-language newspaper in the world (Guiness World Records, 2009), is compiled by downloading and processing articles from online sources. A codebook based on the Narrative Policy Framework (NPF) is initially created deductively and subsequently refined inductively by manually coding a small sample of articles.
2. **Manual and Automated Coding** - Manual coding is performed by one human coder, and automated coding is performed twice by the same LLM. Coding involves reading (in case of the human)/ processing (LLM) a news article, categorising it into one of the many codes available in the codebook where appropriate, and saving the output in a structured data format.
3. **Model Evaluation** - The LLM is evaluated against the manually coded dataset using metrics such as Accuracy[2], Precision[3], Recall[4], F1 Scores[5], and Krippendorff's $\alpha$[6].

Twelve narrative components were identified for the codebook, each with their definitions, categories, and examples (section A.3).

Manual coding was performed at the document level on a representative sample of 297 articles from a total dataset of 14,578 articles, requiring 26 days to complete. Each article was reviewed multiple times to identify all possible components, with components not present in the text marked as zero.

The manual coding of air pollution narratives in India highlighted key figures: medical doctors emerged as the primary heroes spreading awareness, while farmers and stubble burning were seen as villains. The environment and vulnerable city dwellers were portrayed as victims and beneficiaries of the heroes' actions. Government, research universities, and law enforcement were identified as allies; industry and corrupt officials as opponents; and government interventions as sometimes ineffective. The narratives spanned across India, notably in the National Capital Region (Delhi, Gurugram, Noida), where Delhi is infamous for extreme pollution. The study found a "Regulatory Enforcement" and "Story of Decline" plot, with explicit policy solutions appearing as the moral in over half the texts. Egalitarian and Hierarchist belief systems were prevalent, with "Mobilization of Support" being the common strategy. Historically, the general trend of air pollution policy development in developed countries starts with high air pollution leading to increased public awareness, which puts pressure on the government

[2]The proportion of correct predictions (true positives and true negatives) out of all predictions.

[3]The ratio of true positives to the sum of true positives and false positives. It measures the accuracy of positive predictions.

[4]The ratio of true positives to the sum of true positives and false negatives. It measures the ability to find all positive instances.

[5]The harmonic mean of precision and recall. It balances the two metrics and is useful when you need a single measure of a models performance.

[6]Krippendorffs alpha is a statistical measure of the reliability of agreement between observers or raters. It generalizes several known reliability indices and is applicable to various data types (nominal, ordinal, interval, and ratio).

to act. The crucial finding is the growing awareness from 2010 to 2024 about air pollution's health impact, indicating the start of the journey of India's progressing efforts towards cleaner air. However, the timeline of effective policy action remains uncertain.

Manual coding revealed the complexity of this analysis, necessitating a nuanced understanding of the text and contextual knowledge of the country's culture due to the presence of 'Hinglish' (a blend of Hindi and English) and regional languages.

For the automated analysis, a single prompt for the LLM was developed, incorporating a persona (an LLM expert in policy narratives), context (the codebook), the task (coding articles using the NPF), and an output example (in JSON format).

While the LLM accurately identified key characters, it often misrecognized the villain as the policy problem itself. Medical doctors emerged as the primary heroes, raising awareness about air pollution. Farmers, particularly due to stubble burning, were frequently portrayed as villains. The environment and vulnerable urban populations were depicted as victims and beneficiaries of the heroes' actions. Key allies included government entities, research universities, and law enforcement, while industry and potentially corrupt officials were often seen as opponents. Government interventions and policies were sometimes described as ineffective. Geographically, the narratives spanned across India, with the highest concentration in the National Capital Region (NCR), particularly in Delhi, Gurugram, and Noida.

The resulting dataset, coded by the LLM, was compared to the manually coded dataset. The LLM faced challenges with classification, likely due to the numerous categories for each narrative component and difficulty in distinguishing between them. It performed poorly for 'Plot', achieving an accuracy of 24.9% compared to the manual coding. 'Moral' had the best performance among all categories, reaching an accuracy of 59.9%, while 'Belief System' and 'Narrative Strategy' trailed at 36% and 45.5% respectively. 'Moral' also had the lowest number of classes (3). Recall and Precision scores were low, indicating trouble identifying true positives, and Krippendorff's $\alpha$ scores were also low, ranging from -0.056 for 'Plot' to 0.279 for 'Moral'.

The Large Language Model (LLM), used as-is out-of-the-box, demonstrated notable consistency in its coding performance. Accuracy scores across different runs for the same model ranged from 78.1% to 87.9%, with precision between 43.8% and 76.3%, recall from 32.6% to 82.6%, and F1 scores from 32.5% to 75.8%. Krippendorffs alpha ranged from 0.508 to 0.818, with "Belief System" showing the best performance, likely due to clearer definitions in the codebook for this component.

However, named entity recognition presented challenges. Although the LLM consistently recognized major narrative roles, low cosine similarity scores (0.19 for opponents to 0.34 for victims) indicated poor performance in character identification. However, from N-grams of the most frequent words and group of words created to identify recurring characters, similar characters were present in manual and automated coding at the top positions, showing that it largely identifies the important characters in the categories, and low similarity could be arising from differences throughout the document. Enhanced prompt definitions with few-shot examples (multiple examples), hyperparameter tuning, and fine tuning through supervised learning could potentially improve output quality, as has been seen in similar LLM annotation studies.

The LLM required between 25-30 minutes per run (depending on internet speed) for a total of 2 runs to code the 297 articles, highlighting the potential for rapid scaling, constrained only by internet speed, computational power, and cost of API access. As previously mentioned, recommendations in similar LLM annotation studies for enhancing results include fine-tuning the LLM and improving data quality, which could be taken up in future research. LLMs' capability for transfer learning can be leveraged, and this methodology can be generalized for any research conducting a case study using qualitative content analysis for text. With the advent of multimodal large language models, image, audio, and video data can also be incorporated into the dataset.

In conclusion, this research demonstrates the limited ability of automating qualitative content analysis at scale, with a simpler codebook, fewer categories for the LLM to classify articles into, fine tuning, and few-shot examples possibly leading to higher performance. While the current accuracy of the output remains sub-par compared to manual analysis, there is potential for improvement as research in LLM technology progresses. LLMs, in their current state (August 2024) cannot be entirely relied on for narrative policy analysis, but they can be used as a starting point to understand the background of a complex topic such as air pollution policy at a high-level.

# Contents

# List of Figures

# List of Tables

# Nomenclature

## Abbreviations

| Abbreviation | Definition |
|---|---|
| LLM | Large Language Model |
| NPF | Narrative Policy Framework |
| NLP | Natural Language Processing |

# 1

# Introduction

## 1.1. Narratives

Humans are a narrative species, operating through stories rather than facts. As it is difficult to understand facts in isolation, narratives allow us to make sense out of those facts (Gabriel et al., 2004) by acting as a medium to digest those facts. Narratives are defined in literature as "*a story with a temporal sequence of events unfolding in a plot that is populated by dramatic moments, symbols, and archetypal characters that culminates in a moral to the story*" (Jones & McBeth, 2010). This power of delivering facts in an understandable format that narratives yield gives them the ability to shape opinions and individual beliefs (Shanahan, McBeth, & Hathaway, 2011).

In the information age, where every bit of public knowledge is a few clicks or taps away, the power of narratives to shape opinions is further bolstered. Even though strong communication technology has allowed for the rapid spread of information, messages have travelled far and wide without the internet as well. Throughout history, major religions have spread their teachings through stories as a medium. Christianity spread its teachings through the Holy Bible, as did Islam with the Quran, and Hinduism with the Bhagavad Gita. Religious texts have helped shaped human behaviour through their teachings, and in nations based on religion[1], they have even contributed to forming public policy[2].

Narratives today carry on this function in public and private policy processes, with individuals and coalitions using narrative tools like news media and social networks to portray a biased view of events to influence outcomes (Jones & McBeth, 2010), giving rise to the study of policy narratives. A policy narrative is a narrative with a reference to the policy problem that is being investigated (Shanahan, Jones, & McBeth, 2018b). However the policy problem is mentioned in the text, including, but not limited to being in the form of a policy solution, or a process leading to a solution, the reference to the policy problem is what makes it a policy narrative. Policy narratives can be constructed to paint a carefully thought-out picture to achieve a policy goal. For instance, during the healthcare reform debate in the United States, proponents of the Affordable Care Act (ACA) framed the narrative around expanding healthcare access and protecting vulnerable populations. In contrast, opponents focused

[1]The Saudi Arabian legal system is based on the Islamic Shari'a (Van Eijk, 2010)

[2]Although the church is separate from the state in the United Stated of America, anti-abortion activists in the country use passages from the Bible to paint a picture of women who abort and physicians who perform abortion procedures as murderers (Castle, 2011) to sway opinions in the abortion policy debate.

on themes of government overreach and economic burden from the increased spending, presumably to put an end to the act. These competing narratives were disseminated through various media channels to sway public opinion and political support (Shanahan et al., 2011). In the realm of biotechnology policy, narratives have been used to influence public perception and regulatory decisions. Finucane and Satterfield discuss how narrative frameworks can assess risk and value in biotechnology, illustrating that narratives can serve as a means to obtain more representative public opinions. By framing biotechnology in terms of potential benefits or risks, stakeholders have significantly impacted public attitudes and policy directions (Jones & McBeth, 2010). In the climate change debate, environmental organizations often use narratives that emphasize the urgent need for action and the catastrophic consequences of inaction. Conversely, climate change skeptics frame the issue as exaggerated or economically detrimental, aiming to delay policy measures (Shanahan et al., 2011). Studying policy narratives yields information about the policy problem, but there are also other reasons to study them systematically.

### 1.1.1. Why study policy narratives?

Studying policy narratives has many social use-cases. It can help in resolving conflicts, shine light on the interests of various actors involved in the decision making process, and make policy implementations more effective:

1. **Conflict Resolution:** Narrative mediation is an approach to conflict resolution that emphasizes the role of stories and discourses in shaping our perceptions and reactions to conflicts. According to Winslade and Monk, this approach involves deconstructing the traditional problem-solving or interest-based mediation processes and instead focusing on how conflicts are produced within specific sociocultural contexts. By examining the language and stories used to describe conflicts, mediators can uncover underlying biases and assumptions, gaining a deeper understanding of the origins of conflicts, helping disputants to reframe their understanding and promote the creation of new, collaborative narratives that can help parties move beyond difficult or intractable situations (Winslade & Monk, 2000), and find alternative solutions. For example, imagine a scenario where two neighborhoods are in conflict over the city's decision to build a new public park in one area but not the other.

   In this case, policy narrative analysis could be employed by mediators to examine the competing stories that each neighborhood tells about the issue. For example, one neighborhood might frame the conflict in terms of historical neglect and systemic inequality, arguing that the city's decision perpetuates existing disparities. In contrast, the other neighborhood might focus on their need for green space due to higher population density, framing the park as a necessary improvement to their quality of life.

   By analyzing these narratives, mediators could identify the underlying biases and assumptions driving the conflict, such as a perceived lack of fairness or historical grievances. Through narrative mediation, they could then help the parties to deconstruct these entrenched positions and collaborate on creating a new, shared narrative that acknowledges both neighborhoods' concerns. This might involve reframing the discussion from a zero-sum game to a broader dialogue about equitable resource distribution across the city or finding alternative solutions, such as developing a joint community initiative that benefits both areas.

   Ultimately, policy narrative analysis helps to uncover the deeper cultural and social contexts that fuel conflicts, enabling mediators to facilitate the creation of more inclusive and constructive narratives that promote resolution and cooperation.

2. **Understanding Interests of Different Actors:** In complex policy environments, policymakers often use cognitive and emotional shortcuts in their decision-making processes, and narratives can effectively harness these biases. By presenting compelling stories, policymakers can better grasp the underlying issues, values, and consequences of various policy options. This approach is particularly useful in health policy-making, where personal stories can highlight important issues, demonstrate the effectiveness of programs, and influence legislative changes (Fadlallah et al., 2019). Additionally, narratives help in identifying the social and cultural contexts that shape the interests and positions of different actors, providing a more nuanced understanding of their motivations (Winslade & Monk, 2000). Consider a situation in which a public health policy is being debated regarding the funding of mental health services in a rural community. Policymakers are faced with several competing priorities, and the allocation of limited resources is a challenging decision.

   Here, a powerful narrative might be presented by a mental health advocate who shares the personal story of a young individual from the community who struggled with severe depression but had no access to mental health services. This story would illustrate the dire consequences of inadequate mental health care, highlighting not only the individual suffering but also the broader impact on the community, such as increased unemployment and social isolation.

   By leveraging this compelling narrative, policymakers can better understand the human impact behind the abstract data and statistics. The story taps into cognitive and emotional shortcuts like the availability heuristic[3], allowing policymakers to quickly grasp the urgency and significance of the issue. It also sheds light on the social and cultural contextssuch as the stigma around mental health in rural areas and the lack of infrastructurethat shape the community's needs and the positions of various stakeholders.

   As a result, this narrative can influence policymakers to prioritize mental health funding, demonstrating how personal stories can effectively convey the values, consequences, and urgency of policy decisions in a way that traditional data might not. As Joseph Stalin has famously put "A single death is a tragedy, a million deaths are a statistic."

3. **Making Policies More Effective:** Narratives are also instrumental in making policies more effective by enhancing communication and engagement with policymakers and the public. Storytelling can make complex information more accessible, memorable, and persuasive, which is essential for influencing policy decisions. For instance, the use of narratives in policy advocacy can help scientists and researchers communicate their findings more effectively to policymakers, ensuring that evidence is taken up into policy. This approach leverages the psychological and social dynamics of policymaking, making it easier to convey the relevance and urgency of specific issues (Davidson, 2017). Moreover, narratives can shape public opinion and political attitudes, which in turn can drive policy changes. By constructing narratives that resonate with the audience's pre-existing beliefs and values, policymakers can garner broader support for their initiatives (Dennison, 2021). For example, scientists and public health officials have extensive data showing the benefits of vaccination, but simply presenting this data might not be persuasive enough to influence policy decisions or public behavior.

   By crafting a narrative that centers around a personal storysuch as a parent whose child suffered from a preventable disease due to lack of vaccinationthe campaign can make the scientific information more relatable and emotionally impactful. This narrative can help policymakers and the

[3] A cognitive shortcut where people assess the probability of an event based on how easily examples come to mind

public understand the real-world consequences of low vaccination rates, making the need for strong vaccination policies more urgent and compelling.

The use of storytelling in this context not only makes the complex scientific data more accessible and memorable but also aligns with the audience's values, such as protecting children and the community. As a result, this narrative approach can shape public opinion, leading to broader support for vaccination policies and potentially influencing lawmakers to adopt stricter regulations or funding for public health initiatives.

In day-to-day life, policy narrative analysis can assist with strategizing the precise narratives to create in a given scenario for a favourable outcome, providing powerful tools for both, policy lobbyists and policy makers. For instance, in the case of Cape Wind's proposal to install wind turbines to generate electricity in Nantucket, there were clear policy winners and losers identified in the study, with the winning coalition using the angel-shift[4] strategy and the losing coalition falling into a devil-shift[5] trap (Shanahan, Jones, Mcbeth, & Lane, 2013). This study was carried out based on the Narrative Policy Framework, but it is not the only way to study policy narratives.

### 1.1.2. Ways of Studying Policy Narratives

Although the systematic study of policy narratives is a relatively recent field, the existence and importance of stories in human understanding and communication are well known (Polkinghorne, 1988). Several frameworks have been developed to study policy narratives. They are -

1. **Argumentative Turn in Policy Analysis:** This postpositivist approach, pioneered by scholars like Frank Fischer and John Forester in the 1990s, recognizes that public policy is constructed through language and argumentation (Fischer & Forester, 1993), (Veselková, 2017). It examines the role of rhetoric, storytelling, and narratives in shaping policy debates and decisions.
2. **Causal Stories Approach:** Developed by Deborah Stone, this approach analyzes the causal stories or narratives that policymakers and stakeholders use to define policy problems, assign blame, and propose solutions. It focuses on the persuasive power of these narratives in framing policy issues (Stone, 1989).
3. **Discourse Analysis:** Discourse analysis methods, drawing from linguistics and social theory, can be used to study the narratives, frames, and discourses employed by policy actors. Approaches like critical discourse analysis examine how language and narratives are used to construct meaning and wield power in policy debates (Jones, McBeth, Shanahan, Smith-Walter, & Song, 2022).
4. **Frame Analysis:** Frameworks like Goffman's frame analysis can be applied to analyze the frames or narratives used by policy actors to define problems, diagnose causes, make moral judgments, and suggest remedies. This overlaps with the NPF's focus on narrative elements like characters and plots (Jones et al., 2022).
5. **Narrative Analysis:** Qualitative narrative analysis methods from fields like sociology, anthropology, and literary studies can be adapted to systematically analyze the structure, content, and functions of policy narratives (Jones et al., 2022).
6. **Narrative Policy Framework:** The NPF provides a comprehensive framework for analyzing how narratives influence policy processes, designs, and outcomes at multiple levels. Its core assumptions, structured elements, and emphasis on empirical testing make it a valuable tool for policy

[4] Angel-Shift Strategy - From the Narrative Policy Framework, it refers to the casting the winners in the narrative as the 'heroes'.
[5] Devil-Shift Strategy - From the Narrative Policy Framework, it refers to the casting the 'villains' in the narrative as the victors over the 'heroes'.

research. By integrating narrative strategies and belief systems, the NPF enriches the analysis, offering insights into the strategic use of narratives in public policy debates. Research is focusing on refining operational definitions, replicating studies, and testing new hypotheses to build a robust body of knowledge about the role of narratives in the policy process (Jones et al., 2022).

With all these different methods, it is important to choose the method that fits the purpose of a study. Each method has its advantages and disadvantages.

### 1.1.3. Strengths and Weaknesses of Various Methods

Methods to study policy narratives have the following strong and weak points (Volkens, 2007) -

| Method | Strengths | Weaknesses |
|---|---|---|
| **Argumentative Turn in Policy Analysis** | Emphasizes the role of rhetoric and argumentation in policymaking processes. Recognizes that policy debates involve competing claims and justifications. | May overlook the broader social and cultural contexts that shape policy narratives. Focuses primarily on the argumentative aspects, potentially neglecting other narrative elements. |
| **Causal Stories Approach** | Highlights the importance of causal explanations in policy narratives. Examines how narratives attribute causes and assign responsibility. | May oversimplify complex policy issues by reducing them to linear causal stories. Lacks a comprehensive framework for analyzing other narrative elements. |
| **Discourse Analysis** | Provides tools for analyzing the linguistic and rhetorical features of policy narratives. Considers the broader social and political contexts that shape discourse. | May prioritize textual analysis over the broader narrative structure and content. Lacks a specific focus on the unique characteristics of policy narratives. |
| **Frame Analysis** | Examines how policy issues are framed and presented in narratives. Highlights the role of framing in shaping public perception and policymaking. | Primarily focuses on issue framing, potentially overlooking other narrative elements. May lack a comprehensive framework for analyzing the overall narrative structure. |
| **Narrative Analysis** | Offers a systematic approach to analyzing the structure and content of narratives. Considers elements such as plot, characters, and themes. | May lack a specific focus on the unique characteristics of policy narratives. Lacks a comprehensive framework for analyzing the policy process and its contexts. |

**Table 1.1:** Strengths and Weaknesses of Various Methods

**Table 1.1:** (Continued)

| Method | Strengths | Weaknesses |
|---|---|---|
| **Narrative Policy Framework (NPF)** | Provides a comprehensive framework for analyzing policy narratives in their broader contexts. Considers narrative elements, policy beliefs, and the role of narratives in the policy process. Examines how narratives influence public opinion, policymaking, and policy implementation. | A relatively new approach, still evolving and being refined. May require a significant amount of data and analysis to fully apply the framework. |

All methods of narrative analysis consist the analysis of some form of media, usually text, and annotating that data in a defined and systematic way.

## 1.2. Role of Text in Policy Narratives

As illustrated in subsection 1.1.2, there are numerous methodologies for studying policy narratives, yet the unifying element across these approaches is the centrality of natural human language. Natural language is pivotal in the analysis of policy narratives as it constitutes the primary medium through which these narratives are constructed, communicated, and interpreted. While narratives can be disseminated through various formats, including audio, video, and print, the underlying messages are invariably conveyed through natural language. In the contemporary context, print remains one of the oldest and most readily accessible mediums for the distribution of information, highlighting the significant role of text in the study of policy narratives.

Policy narrative texts are available in a wide array of sources such as newspaper articles, policy documents, legislative records, speeches, interviews, social media posts, and more. These texts contain the stories that policymakers, stakeholders, and the public use to communicate and contest policy issues. Sources are different at different levels of analysis (Shanahan et al., 2018b). They are -

**Micro-Level Data:** This includes interviews, providing detailed insights into individual perspectives and experiences.

**Meso-Level Data:** Consists of public records, media content, speeches, and legislative records. This level captures the interactions among groups and organizations.

**Macro-Level Data:** Encompasses cultural narratives and institutional communications. These sources reflect broader societal values, norms, and ideologies.

The majority of the data discussed herein is qualitative and unstructured, originating from sources such as newspaper articles, transcripts of speeches or interviews, and various forms of media including audio, video, or images, as well as policy documents released by institutions. Aside from images, audio, and video (all of which can be transcribed or described), this narrative data is presented in natural language text.

Advancements in communication and storage technologies have led to an exponential increase in data generation. In 2009, the "Digital Universe"[6] was estimated to encompass approximately 0.8 Zettabytes[7], with projections indicating an expansion to around 35 Zettabytes by 2020 (Gantz & Reinsel, 2010). However, according to sources such as Raconteur[8] and Statista[9], the actual size of the digital universe in 2020 reached approximately 64 Zettabytes, nearly double the projected figure, with data generation continuing to accelerate exponentially each day.

As the volume of data increases, traditional manual methods of qualitative content analysis struggle to cope with the amount of available data, now falling entirely behind. Human coders have limited capacity for analysis, and the need for a higher number of coders becomes a significant bottleneck, hindering comprehensive analysis. This limitation increases the necessity for the automation of qualitative content analysis to manage and analyze the large amount of data effectively.

## 1.3. Addressing Challenges with Automation using Natural Language Processing Techniques

As manual text analysis has natural limits, the solution would be to automate this task with the help of computers. Previous research has employed NLP techniques to perform qualitative content analysis, but there are certain limitations present in traditional techniques.

### 1.3.1. Use of NLP Techniques in Previous Research and Their Limitations

Techniques such as topic modelling, keyword extraction, named-entity recognition, etc. (Olofsson, Weible, Heikkila, & Martel, 2018), (Wolton, Crow, & Heikkila, 2021) have been employed in policy analysis studies deploying the Narrative Policy Framework. While they have been proven valuable, they have several limitations -

- Lack of Context and Nuance
  - Traditional NLP techniques often struggle to capture the full context and nuance present in policy narratives and texts. They may overlook subtle implications, sarcasm, or figurative language that a human reader would easily understand (Jin & Mihalcea, 2022), (Zhou, Dai, Ren, Chen, & Chen, 2022).
  - Topic models and keyword extraction can identify prevalent themes and terms but may fail to grasp the deeper meaning, motivations, or underlying assumptions behind the policy narratives (Jin & Mihalcea, 2022), (Zhou et al., 2022).
- Oversimplification
  - By reducing complex policy narratives to topics, keywords, or named entities, traditional NLP techniques can oversimplify the richness and complexity of the original text. Important details, caveats, or counterarguments may be lost in this process (Jin & Mihalcea, 2022), (Zhou et al., 2022).
  - Named entity recognition can accurately[10] identify and classify entities like people, organizations, and locations, but it may struggle with more abstract or domain-specific concepts

[6] The Digital Universe is defined as all the "digital information created and replicated in the world" (Gantz & Reinsel, 2010).
[7] 1 Zettabyte = 1 Trillion Gigabytes
[8] Raconteur Source
[9] Statista Source
[10] Although Named Entity Recognition is largely considered a solved problem delivering highly accurate results, it suffers from problems of ambiguity in the very definition of what a named entity is, discrepancies in annotation guidelines across various evaluation forums, a lack of said forums, a requirement of large annotated corpora for training which limits their practical applicability for end users, and reporting on overfit results which questions the accuracy (Marrero, Urbano, Sánchez-Cuadrado, Morato, & Gómez-Berbís, 2013).

common in policy documents (Marrero et al., 2013).

- Bias and Subjectivity
  - The performance of traditional NLP models can be influenced by the data used for training, potentially introducing biases or reflecting the subjective perspectives of the training data (Jin & Mihalcea, 2022), (Zhou et al., 2022).
  - The interpretation of topics, keywords, or named entities extracted by traditional NLP models can be subjective and influenced by the researcher's own biases and preconceptions (Jin & Mihalcea, 2022), (Zhou et al., 2022).
- Data Quality and Preprocessing
  - The quality and preprocessing of the input text data can significantly impact the performance of traditional NLP techniques. Errors, inconsistencies, or noise in the data can lead to inaccurate or misleading results (Jin & Mihalcea, 2022), (Zhou et al., 2022).
  - Policy documents often contain specialized terminology, acronyms, or domain-specific language that may not be well-handled by general-purpose traditional NLP models trained on different types of text (Jin & Mihalcea, 2022), (Zhou et al., 2022).

Techniques like Recurrent Neural Networks, the state-of-the-art in natural language processing until 2018, had limitations such as limited context retention, scalability and computational limits arising from a sequential architecture, and a lack of transfer learning (requiring more task-specific training to achieve good performance) (Schmidt, 2019).

A recent natural language processing solution, next word generation using large language models, solves many of these challenges using a new software architecture.

### 1.3.2. Why are Large Language Models Any Better?

Like every technology with a sizeable user base, natural language processing has been getting better. Although no computer can understand "meaning" in the human sense of the word, there are techniques that can approximate and model "context" mathematically very well through the attention mechanism.

In 2017, Google researchers released a paper titled "Attention is All You Need" (Vaswani et al., 2017), introducing[11] the attention mechanism for language translation tasks.

This is how the architecture works -

1. Input Representation: Each word in the input sequence is first converted into a vector representation using embeddings.
2. Query, Key, and Value Vectors: For each word, three vectors are derived: the query vector (Q), the key vector (K), and the value vector (V). These vectors are obtained by multiplying the input embedding with learned weight matrices.
3. Attention Scores: The attention score for a pair of words is calculated by taking the dot product of their query and key vectors. This score is then scaled (by the square root of the dimension of the key vectors) and passed through a softmax function (normalization function) to obtain normalized weights.

[11] Although the transformer architecture, the backbone of large language models, was introduced by Bahdanau, Cho, and Bengio first in 2014, it gained popularity in 2017, when Google researchers used the transformer architecture for language translation tasks and released their landmark paper (Vaswani et al., 2017).

4. Weighted Sum: The final representation of each word is obtained by taking a weighted sum of the value vectors, where the weights are the attention scores. The value vectors are converted back into text and returned to the user as output.

In 2018, OpenAI, a research and development company applied this technique to create a generative predictive application which was performing favourably with next word prediction[12]. The OpenAI team scaled up the parameters in the model, the size of the training data, and released "ChatGPT". Large Language Models (LLMs) have since then become the state-of-the-art in natural language tasks and have shown performance in the top percentiles in standardized tests measuring language and math skills (OpenAI, 2024). These models have shown that abilities like reasoning in natural language emerge from training on a sufficiently large amount of data, making these models the only viable candidate to attempt this human level task previously not possible.

LLMs address the problems of context, nuance, and oversimplification in text using the self-attention[13] mechanism, which enables transformers to consider all positions in the input sequence, effectively capturing dependencies between distant words, and the multi-head attention[14] mechanism which dynamically weighs the importance of surrounding words, helping to disambiguate meaning based on context (for example, "Paris" in "Paris is beautiful" vs. "Paris Hilton is famous") in the transformer architecture.

Bias and subjectivity are a problem that exist in every human, and by extention, the data that we produce. Large language models are trained on large amounts of this very data, causing bias to creep in. Development of large language models is assisted with reinforcement learning with human feedback (RLHF) (Christiano et al., 2023), where labellers provide desired responses to avoid the LLM from providing toxic answers peppered with harmful sentiments (OpenAI, 2022), (Google Gemini Support, 2024) to mitigate this problem.

LLMs also overcome the problem of sequential processing in Recurrent Neural Networks, again, through the transformer architecture, which is highly parallelizable. It allows for the use of parallel computing for higher speeds (Vaswani et al., 2017). LLMs have much larger context retention (upto 1 million tokens[15] for Google Gemini 1.5 Advanced (Gemini Token Size, 2024)), compared to the Long Short Term Memory RNN, with context windows in practice ranging from about 50 to 200 words, with performance degrading with more (Graves, 2012).

Due to these features of LLMs, they are good candidates for qualitative content analysis using standard frameworks to analyse policy narratives, allowing text to be analysed at scale.

## 1.4. Thesis Focus

To examine the possibility of LLMs as a viable candidate for doing policy analysis, this thesis aims to examine the performance of large language models on qualitative content analysis of policy narratives using a policy analysis framework, using a case study. The approach is twofold: a manual analysis of the dataset, and analysis using large language models (LLMs). The Narrative Policy Framework has been chosen for this study to analyze policy narratives. The NPF's structured elements and emphasis

[12]Next word prediction is a common task in natural language processing (NLP) where the goal is to predict the next word in a given sequence of words. This task is foundational in many applications such as text completion, language modeling, and machine translation.

[13]In self-attention, each word attends to all other words (including itself) in the sequence, allowing it to incorporate contextual information.

[14]Multi-head attention involves using multiple sets of query, key, and value matrices, allowing the model to capture different types of relationships and nuances from different subspaces of the representation.

[15]100 tokens are equal to approximately 60-80 words in English (OpenAI, 2024), (Google, 2024)

on empirical testing provide a robust tool for policy research. It offers a rich set of vocabulary and examines different dimensions of narratives, making it suitable for this case study because of 2 reasons:

1. The Narrative Policy Framework allows for the development of a codebook with definitions of components that can be customized according to the case study. This makes possible the capture of complex elements that cannot be determined solely based on keywords in the text, making context important. This will serve to be the task that LLMs are specifically chosen for in this study.
2. The NPF is designed for repeatability in its application and to perform an empirical analysis of qualitative data, which is useful in checking the consistency of LLM coding across multiple runs.

By integrating narrative strategies and belief systems, the NPF enriches the analysis, offering insights into the strategic use of narratives in public policy debates. Although the approach is new, there are studies[16] that have used and developed upon the framework, and made available codebooks that can be customized according to need.

While each approach has its merits, the Narrative Policy Framework (NPF) also stands out as the most comprehensive and tailored approach for studying policy narratives. It offers a holistic framework that considers narrative elements, policy beliefs, and the broader policy process contexts. Its biggest advantage is its portability across contexts. The NPF has shown potential for portability across different geographies, political systems, and policy fields. This adaptability is due to the universal importance of narratives in human cognition and communication, making the NPF relevant in a wide range of settings (Schlaufer, Kuenzler, Jones, & Shanahan, 2022). The NPF is used to analyze policy narratives available in all kinds of media (audio, video, and text). In this study, it will be used to code textual information in the form of newspaper articles addressing the chosen policy problem. The chosen case for this study is air pollution in India, a critical and persistent policy issue.

Air quality, a paramount environmental and public health concern, has been improving worldwide (as measured through the concentration of pollutants) due to increasingly strong restrictions by governments (Baldasano, Valera, & Jiménez, 2003). However, in low and average-income countries in the Global South such as India, which unfortunately boasts some of the world's most polluted air (Kumari, Lakhani, Kumari, et al., 2020), air pollutant concentration is high and its trend is only upward at the ground level (Baldasano et al., 2003). This declining air quality poses severe health risks (Kampa & Castanas, 2008), contributing significantly to the burden of disease. The key causes of air pollution include man-made and natural sources such as factories, power generation plants, cars, planes, crop burning practices, volcanic activity, wildfires, etc. (Pratap Choudhary & Garg, 2013). Policy can play a key role in addressing the situation. Throughout history, governments have acted on air pollution through anti-pollution policy when awareness about its ill effects has spread (Zeng, Cao, Qiao, Seyler, & Tang, 2019).

Despite increasing evidence of these adverse effects, there remains a disproportionate response in policy action in India, especially in regions most affected by pollution, exacerbated by corruption and under-utilization of allocated funds (PTI, 2023), (Tiwari, 2023), (Singh, 2015). Studying about the problem of air pollution in India using policy narratives in the form of newspaper articles about the policy problem could provide a deeper understanding of the issue, the actors and stakeholders in the process, and the reasons for the lack of commensurate measures to tackle the problem.

With a suitable case study to conduct, the research moves forward with the following research question.

[16] Olofsson et al., Shanahan et al., Gupta, Ripberger, and Collins, Veselková, Crow and Berggren to name a few.

## 1.5. Research Question

**To what extent do Large Language Models (LLMs) accurately automate qualitative coding of policy narratives when compared to a manually coded dataset?**

As previously mentioned, to address the research question, a case study on air pollution in India will be employed, facilitating a comparative analysis of the coding results. The investigation begins with an exploration of the theoretical underpinnings pertinent to the research. Specifically, it necessitates a comprehensive review of the Narrative Policy Framework (NPF) to effectively apply it to this analysis. This theoretical review will elucidate the intricacies and uses of the NPF, to help make its application relevant to the current study.

Subsequently, a tailored codebook will be developed and refined to align with the case study. Initial data exploration will aid in identifying prevalent themes, contributing to the codebook's development. The codebook will be iteratively refined through its application manually and evaluation of the fit. A systematic research design will be established to ensure rigorous analysis. This will involve performing a manual coding process based on the refined codebook, with results presented in a standardized format.

Following the manual analysis, an automated analysis using large language models (LLMs) will be conducted. The large language model deployed will be used as-is out-of-the-box. This is because there are thousands of variants that can be created from a base model[17] for specialized applications (Hugging Face, 2024). Some specialized models may have better performance than base models for this particular task as they might have been fine-tuned on classification datasets and entity recognition, and may have their hyperparameters tweaked for optimal performance. This kind of model selection, however, is outside of this study's scope.

The findings from the automated analysis will be compared with those from the manual analysis. This comparison will enable a thorough evaluation of the performance of LLMs in qualitative coding, identifying areas where they are effective and where they fall short using predefined metrics. Finally, conclusions will be drawn regarding the implications of LLM performance for future research.

The sub-questions designed to address the main research question are as follows:

### 1.5.1. Sub-Questions

1. What are the key theories, concepts, and relevant studies in policy narrative analysis using the Narrative Policy Framework?

   To effectively deploy the Narrative Policy Framework, a theoretical background requires establishment. The underlying assumptions of the NPF, the constituents of the framework, the definition of the codebook, and the levels of analysis will be studied, and previous research relevant to this study will be reviewed.

2. What narrative components can be used to manually code news articles about the chosen case study to create a robust reference dataset against which LLM agreement will be analyzed?

   The policy problem of air pollution in India serves as a case study for assessing the capability of LLMs in annotating texts for the chosen narrative components. In assessing capabilities, it is crucial to compare labels presented by an LLM with a human-annotated dataset. To create this dataset, a comprehensive codebook based on the elements of the Narrative Policy Framework

[17] By base model or base state of the model, the version of the LLM as shipped by the developer, with no modifications whatsoever is implied

(NPF) will be developed and iteratively refined. A stratified sample of articles will be manually annotated, ensuring the coding framework comprehensively encompasses the content of the news articles.

3. What insights are derived from the manual coding of the case study on air pollution in India?

   Insights about the policy narrative of the chosen case study will be drawn in terms of the narrative components chosen to code the data. The recurring characters, narrative plots (story of decline, conspiracy, etc.), the belief systems present in the narrative (egalitarian, individualist, hierarchist, etc.), and the narrative strategies employed to tell the story (devil-angel shift, cost benefit distributions, etc.) will be identified[18] and the findings will be discussed. These findings will give context for comparing the results of the manual analysis to that of the LLM to aid the judgement of LLM performance.

4. What is the consistency of LLM-generated qualitative coding results for the same prompts across multiple runs?

   Repeatability is a must in qualitative studies, and the robustness of the coding is determined by inter-coder reliability. If LLMs are to be used for this work, it is important that they produce the same results, or results with high levels of similarity across multiple executions. Consistency will be evaluated by running the same prompt 2 times and comparing the results. Metrics[19] like accuracy, precision, recall, F1 scores, and cosine similarities will be used to measure the consistency of narrative elements identified by LLMs across different runs.

5. How does the LLM-generated qualitative coding agree with the manually coded dataset?

   Using a similar method as sub-question 4, the assessment of how well LLMs capture context, nuance, and complexity in policy narratives compared to human coders will be analysed here. Performance will be evaluated by comparing the accuracy, precision, recall, F1 scores, and cosine similarities between LLM-based qualitative coding and manual coding performed by a human.

   The advantages and challenges of using LLMs will be analyzed, including scalability and reliability. Specific strengths and weaknesses identified through comparative analysis with manual coding will be discussed. The implications of LLM-based qualitative coding for large-scale policy narrative research will also be explored, including how LLMs can enhance the speed with which large volumes of policy-related text data can be analyzed. Recommendations will be provided for the use of LLMs for policy narrative analysis.

Together, the sub-questions aim to collectively answer the main research question.

## 1.6. Approach

This research employs a qualitative case study approach: a detailed study of a complex phenomena in a contemporary setting, performed through manual analysis and automated policy narrative analysis using LLMs. The manual analysis involves traditional qualitative methods to examine policy narratives for the case study (air pollution in India) by coding the dataset of newspaper articles using a predefined codebook, while the automated analysis leverages large language models to perform the same task. The results of manual coding will be compared to the coding done by the LLM, and the consistency of the LLM coding will be checked.

[18] See section A.3 for the complete codebook and definitions.
[19] See subsection 3.6.2.

### 1.6.1. Scientific Relevance

This study attempts to address the limitations of current NLP techniques to do policy narrative analysis type of work by exploring the ability of LLMs in their base state to perform policy narrative analysis, and measures its agreement with a human coder on the same dataset to benchmark its performance. It contributes to the body of knowledge by demonstrating how advanced NLP tools can complement traditional qualitative methods and extend them by doing tasks currently not possible with such methods. On successful implementation, the expectation is that using LLMs for policy narrative analysis will enhance its robustness, speed, and scalability. This study will also make available datasets of articles coded by a human coder and an LLM, and a larger dataset of all the downloaded articles for researchers to use for their work in the future. The code written to pre-process the articles, make the graphs, run the LLMs, and post-process their output will also be made publicly available.

### 1.6.2. Practical Relevance

Understanding air pollution narratives is vital for addressing this significant policy issue. Effective narrative analysis can inform the development and diffusion of new technologies and policies aimed at mitigating air pollution. This research has practical implications for policymakers, researchers, and public health advocates by providing insights into the narratives that shape public opinion and policy decisions.

The study will also make available a fully crafted software pipeline to conduct such studies in the future, helping shift the focus on the subject matter of the analysis rather than the method.

## 1.7. Management of Technology Relevance

This thesis was a mandatory part of the MSc Management of Technology Program at TU Delft. Management of Technology aims to teach students of a purely technical background about the deployment of technology as a corporate resource. The goal of this research based study is to equip students to answer important questions that companies must answer, such as -

1. What technology is required by the company?
2. When is it needed?
3. Should it be developed in-house, collaboratively, or acquired externally?
4. How can existing and emerging technologies affect the mission, vision, strategy, objectives, and opportunities of the company?

Narrative analysis at scale possesses significant potential as a powerful tool for comprehending the intricacies of public sentiment. This method can effectively address the challenges brought up by the Management of Technology (MOT) program in the following ways -

1. **Determining the necessary technologies and their appropriate timing -** By analyzing market trends through prevalent narratives in public and policy discussions, companies can identify which technologies are gaining traction and recognize critical emerging issues. Moreover, narrative analysis offers insights into customer needs by revealing what customers are discussing, their pain points, and their expectations and a deep understanding of consumer attitudes and societal trends. This enables companies to develop technologies that address current and genuine needs, and allow for tailored products and marketing strategies. Using the tool developed in this study, coupled with a customized NPF codebook would enable this kind of market research at scale.

2. **Technology Procurement -** Through narrative analysis of technological capabilities and advancements, companies can assess their in-house skills and determine if external expertise is required. This analysis also uncovers potential collaborative opportunities by highlighting the strengths and weaknesses of potential partners, aiding in the formulation of effective collaboration strategies. Additionally, understanding market perceptions of technologies can inform decisions on whether to develop technology internally or pursue acquisition or licensing, especially if a technology is perceived as highly specialized and beyond current capabilities. Using the methodology of this study, corporations will be able to create maps of strengths, weaknesses, opportunities, and threats from a sea of data, aiding their long term strategy.
3. **Leveraging Technological Opportunities -** By examining narratives around existing technologies and their shortcomings, companies can identify unmet needs and innovation opportunities. Understanding societal and policy narratives ensures that technological development aligns with broader societal goals and company missions, promoting strategies that are both innovative and socially responsible. Narrative analysis at scale will make available customer demand information that corporations do not obtain from traditional sources of customer reviews and feedback, allowing them to innovate and create better products.
4. **Technology Impact Analysis -** To analyze technologies and their commercial impact involves investigating and comprehending the technological, economic, and social environments both within organizations and in relation to business partners. Narrative analysis provides the necessary tools for deep understanding of these environments. It also aids in anticipating wider societal trends, thus shaping strategic decisions about technological development and market entry. Analyzing the commercial impact of technologies through narrative insights enables predictions about consumer behavior and policy changes.
5. **Technology Management -** Strategic narrative management in business involves crafting compelling stories to effectively position products and technologies in the market, influencing stakeholder perceptions, and driving adoption and loyalty. In navigating regulatory and policy environments, narrative analysis helps anticipate and adapt to regulatory changes by understanding policy narratives, and advocates for favorable regulations through informed narrative engagement. Risk management and ethical considerations are addressed by understanding public and ethical concerns through narrative analysis, proactively mitigating risks, and ensuring responsible technology deployment.
6. **Technological Diffusion and Adoption -** Narrative analysis aids in devising strategies to enhance societal acceptance of new technologies and aligning or shifting prevailing narratives for smoother integration.
7. **Corporate Social Responsibility (CSR) and Puclic Relations (PR) -** Narrative analysis supports CSR and PR efforts by engaging with public narratives to maintain a positive corporate image and building public trust through understanding and influencing public perceptions.

## 1.8. Structure of the Document

In CHAPTER 2, the relevant background for the study and the current state of research in the Narrative Policy Framework will be discussed. CHAPTER 3 details the research design and the motivation for the choices made. CHAPTER 4 contains the results of the manual coding, an analysis of the results from an air pollution perspective, the insights in an Indian context. CHAPTER 5 contains the results of the coding done by the LLM, key observations, and comparison between multiple LLM runs. CHAPTER 6 contains the comparison of the results from the LLM to the results from the manual coding, and highlights the areas where the LLM performs well and those where it does not. CHAPTER 7 discusses the results obtained, and what they mean for qualitative content analysis of policy narratives by LLMs, and what they show about the policy landscape surrounding air pollution in India. CHAPTER 8 talks about takeaways from the study, limitations, scientific contribution, future research, and a conclusion.

# 2

# Theoretical Background

The research focuses on a systematic qualitative content analysis of policy narratives in an empirical manner, manually and through automation. The analysis is done using the Narrative Policy Framework (NPF) to code news articles about air pollution in India.

The Narrative Policy Framework offers a structured analytical approach to understanding the role and mechanics of narratives within public policy. Central to the NPF is the recognition of the inherent narrative nature of human cognition and communication, particularly in the context of policy formation and advocacy. The NPF is based on five core assumptions (Shanahan et al., 2018b), (Jones, Mcbeth, & Shanahan, 2014), (Jones et al., 2022):

**Social Construction of Policy Realities:** Policy reality is socially constructed, meaning that the significance of policy-related objects and processes varies based on human perceptions, best understood through collective and individual social constructions rather than objective truths. For example, consider climate change policy. In some communities, it might be seen as a critical issue due to perceived direct impacts, like rising sea levels, while in others, it might be seen as less urgent or even contested, depending on local experiences and beliefs.

**Bounded Relativity:** While social constructions create different policy realities, this variation is bounded by belief systems, ideologies, and norms within certain boundaries influenced by identity or culture, making it stable over time. For instance, the debate on gun control in the U.S. shows bounded relativity. While views on gun ownership vary widely, the debate is generally framed within the boundaries of the Second Amendment rights and cultural values surrounding personal freedom and safety.

**Generalizable Structural Elements:** Narratives possess specific, identifiable structures such as plots and characters that are consistent across contexts that can be counted and analyzed statistically. Consider discussions about healthcare reform in the U.S., where narratives often have a clear structure: healthcare providers as heroes, insurance companies as villains, and patients as victims. This structure can be seen across various such debates, regardless of specific policy details.

**Three Interacting Levels of Analysis:** Narratives operate simultaneously at micro (individual), meso (group/coalitional), and macro (cultural/institutional) levels. For example, in the immigration policy debate, an individual (micro level) might share a personal story about their immigrant journey, which resonates with a particular community or advocacy group (meso level), and these stories can influence broader national discourse and policy-making (macro level).

**Homo Narrans Model:** Humans primarily understand and communicate about the world through narratives or in narrative form. Emotions precede reason in narratives, and affective stories drive cognition, communication, and decision-making. For example, a politician might share a touching story about a struggling family to evoke an emotional response and support for social welfare policies, appealing to emotions rather than presenting only statistical evidence.

## 2.1. The Narrative Policy Framework

Proposed by Jones and McBeth (2010), the NPF contrasts with traditional poststructuralist[1] approaches by offering a systematic, empirical methodology for analyzing narratives in public policy. It emphasizes empirical and falsifiable research, integrating narrative analysis into the broader empirical study of policy processes with clear hypotheses and methodological guidelines.

It identifies narratives as having certain core components, which upon clearly defining, are reproducible across context.

### 2.1.1. Core Components of a Policy Narrative

There are 2 types of core components, Narrative Form and Narrative Content.

#### Narrative Form

Narrative Form concerns itself with the elements present in the narrative, which are distinctly identifiable across contexts (Jones et al., 2014). They are -

1. **Setting or Context:** The setting of a policy narrative refers to the space and time where the action unfolds, focusing the audience's attention on a specific context. It can range from a physical location to a broad socio-economic-political environment (Jones & McBeth, 2010) including legal, cultural, political, social, and economic factors (Jones et al., 2014), grounding the narrative in a particular reality that resonates with the audience.
2. **Characters:** Characters are the entities within the narrative who act or are acted upon, categorized into heroes, villains, or victims based on their roles in the policy issue. Heroes are seen as potential fixers of the issue, villains as the cause of the problem, and victims as those harmed. Additional character types may include beneficiaries of the proposed policy solution and allies or opponents to the hero's cause.
3. **Plot:** The plot organizes the actions within the narrative, linking characters to each other and to the setting. It highlights the causal mechanisms at play, often structured around common public policy storylines such as power struggles, change, or decline. The plot is crucial for constructing a compelling narrative that guides the audience through the unfolding of events and their implications.
4. **Moral of the Story:** Typically represented as the policy solution, the moral of the story is the takeaway message intended to prompt action. It may advocate for a specific policy change or suggest

[1] Post-structuralism is a theoretical framework that challenges the idea of fixed structures and meanings in language, culture, and society. It emerged as a reaction against structuralism, which posited that underlying structures govern human behavior and cultural phenomena. Post-structuralists argue that meanings are not stable but are instead constructed through language and discourse, constantly shifting and influenced by power relations (Williams, 2014).

intermediate steps towards a larger goal. The moral reflects the narrative's underlying values and the protagonist's (hero's) efforts to achieve a beneficial outcome for the victim or society at large.

### Narrative Content

Although a common claim by postpositivists[2] is that narratives are relative, the NPF proposes that generalizable elements are present in narratives that confine the variability in content (Jones & McBeth, 2010). 2 possible ways to put this claim to practice are -

1. **Belief Systems:** These are the values and ideologies underlying the narrative, shaping the interpretation of the narrative elements and influencing the persuasiveness of the narrative.
2. **Narrative Strategies:** These are techniques used to craft and deliver the narrative to maximize its impact. They may include framing the narrative to resonate with the audience's pre-existing beliefs and emotions and employing persuasive elements to shift opinions and behaviors.

In the Narrative Policy Framework (NPF), the fundamental components are narrative form and content. These components are further divided into categories and characters, some of which are universally present across all narratives. Additionally, there are customizable elements that can be adapted to align with the specific objectives of the study being conducted. The next section discusses them in the context of this study.

### Components of the Narrative Policy Framework

The categories present in narrative form and content used in extant NPF studies and literature (Shanahan, Jones, and McBeth (2018a), Jones and McBeth (2010), Olofsson et al. (2018), etc.) combined with categories that have been a part of existing narratives in contemporary and historical stories, such as the Biblical plot of David and Goliath, plots of movies, etc. provide a rich set of labels to assign to entities in the narratives to be studied, and to assign to the structure of the narrative itself. Characters such as heroes, villains, and victims are a part of most, if not all, narratives we see and hear in daily life. Many other such characters, plotlines, strategies of narration are described in the following figure and tables:

[2]Post-positivism is a philosophical approach that evolved as a critique of positivism. It acknowledges the limitations of positivism, particularly the idea that knowledge is entirely objective and can be fully captured through empirical observation. Post-positivists recognize that all observation is fallible and that research is influenced by the researchers perspectives, values, and biases (Phillips & Burbules, 2000).

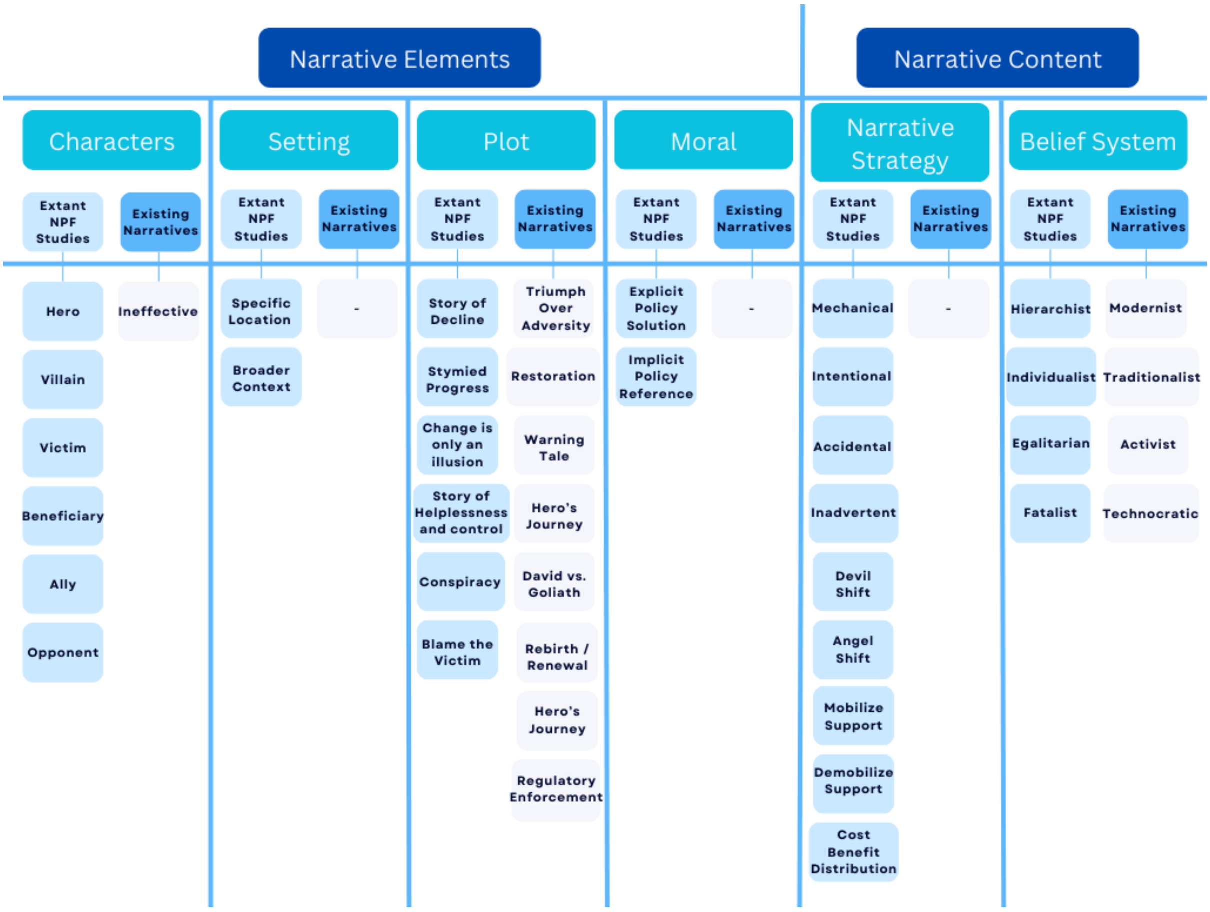



**Figure 2.1:** Narrative Elements and Content

Following is how the codebook is structured and how each component is defined -

Characters

| Type | Definition |
|---|---|
| Hero | The potential fixer of the policy issue, taking action with purpose to achieve or oppose a policy solution. Any actor depicted taking positive steps towards air pollution mitigation, advocating for clean air policies, or raising awareness about air pollutions consequences is considered a hero (Shanahan et al., 2018a). |
| Villain | The entity causing the policy problem, creating harm or opposition to the hero's aims (Shanahan et al., 2018a). |
| Victim | The one harmed by the villain, affected negatively by an action or inaction (Shanahan et al., 2018a). |
| Beneficiary | Those who benefit from the proposed policy solution; could be an animate character who is explicitly stated, directly linked to a hero, and the receiver of an action of a hero (Shanahan et al., 2018a). |
| Allies | Those aligned with the hero, supporting their efforts towards the policy solution (Shanahan et al., 2018a). |
| Opponents | An entity opposing a policy but distinct from a villain, often presenting alternative views or objections to the proposed solutions (Shanahan et al., 2018a). |
| Ineffective | An entity that performs an action that has no effect on the policy problem (Shanahan et al., 2018a). |

Setting

| Type | Definition |
|---|---|
| Specific Location | Settings can be specific locations, like a fracking site (Shanahan et al., 2018a). |
| Broader Context | Settings can be broader contexts, like the American West (Shanahan et al., 2018a). |

Plot

| Type | Definition |
|---|---|
| Story of Decline | This plot describes an initial state of well-being that deteriorates over time, highlighting the urgent need for action (Shanahan et al., 2018a). It may start with a good situation that worsens or begin at a point where things are already dire. |
| Stymied Progress | This plot outlines a trajectory of improvement that is halted or reversed by external interference, emphasizing thwarted efforts towards betterment (Shanahan et al., 2018a). |
| Change-Is-Only an Illusion | This narrative reveals that perceived changes in a situation are misconceptions, with the real situation being stable or moving in the opposite direction (Shanahan et al., 2018a). |
| Story of Helplessness and Control | This plot describes a dire situation initially seen as unchangeable but later shown to be amendable through specific actions or revelations. |
| Conspiracy | This plot involves a progression from an apparently predetermined state to one where control is exerted by a select few who have been manipulating circumstances for their own benefit (Shanahan et al., 2018a). |
| Blame the Victim | This plot centres on issues where those suffering from a problem are inaccurately held responsible for causing it (Shanahan et al., 2018a). |
| Triumph Over Adversity | This plot revolves around overcoming significant obstacles through resilience and ingenuity (Kimberly Fiock, 2023), focusing on successful mitigation of air pollution. |
| Restoration | This plot focuses on restoring the environment or social system to its original, pristine condition after suffering degradation, highlighting restorative efforts (Blignaut & Aronson, 2020). |
| Warning Tale | This plot serves as a cautionary story about the dire consequences of inaction or improper actions, often projecting a bleak future to motivate current action (Soriano & Frey, 1969). |
| Hero's Journey | This plot follows a protagonist or group undergoing a transformative journey to resolve a crisis, featuring various trials and eventual success. |
| David vs. Goliath | This plot highlights the struggle of a seemingly powerless individual or group against a far more powerful adversary, emphasizing justice and equity (Holy Spirit, Original text written circa 6th century BCE). |
| Rebirth or Renewal | This plot focuses on transformation and new beginnings after a period of decline or catastrophe, promoting an optimistic and forward-looking perspective (Duffy, 2017). |
| Pendulum Swing | This plot captures the cyclical nature of policy and public sentiment, where attitudes and conditions swing from one extreme to another over time (Rieder, 2018). |
| Regulatory Enforcement | This plot typically involves a structured progression where regulations are introduced, enforced, and the consequences of these actions are observed and analyzed (Kuhlmann & Blum, 2021). |

Narrative Strategies - Causal Mechanisms

| **Type** | **Definition** |
| --- | --- |
| Mechanical Cause | The excerpt associates intended consequences by unguided action with a policy problem (Shanahan et al., 2018a). |
| Intentional Cause | The excerpt associates intended consequences by purposeful action with a policy problem (Shanahan et al., 2018a). |
| Accidental Cause | The excerpt associates unintended consequences by unguided action with a policy problem (Shanahan et al., 2018a). |
| Inadvertent Cause | The excerpt associates unintended consequences by purposeful action with a policy solution (Shanahan et al., 2018a). |
| Devil Shift | Casting villains as the victors over the heroes (Shanahan et al., 2018a). |
| Angel Shift | Casting the heroes as the winners (Shanahan et al., 2018a). |
| Mobilization of Support | Rally support for a particular policy position (Shanahan et al., 2018a). |
| Demobilization of Support | Diminish support for opposing views (Shanahan et al., 2018a). |
| Diffusing Costs and Concentrating Benefits | This strategy involves how the costs and benefits of a proposed policy are distributed among the characters in the narrative. The elite few get the advantage, while the common people pay for it (Shanahan et al., 2018a). |
| Concentrating Costs and Diffusing Benefits | This strategy involves how the costs and benefits of a proposed policy are distributed among the characters in the narrative. Costs are concentrated and benefits are intended for a larger audience, generally the public (Shanahan et al., 2018a). |

## Moral/Policy Solution

| **Type** | **Definition** |
| --- | --- |
| Explicit Policy Solution | The policy solution in the policy narrative, frequently culminating in a call to action (Shanahan et al., 2018a). |
| Implicit Policy Reference | Intermediary steps leading to a larger policy solution (Shanahan et al., 2018a). |

Belief Systems

| Type | Definition |
|---|---|
| Hierarchist | Hierarchist belief systems focus on the need for structured regulations and state-led initiatives (Thompson, 2018) to tackle air pollution. They advocate for stringent enforcement of emission standards, urban zoning laws, and industrial compliance with environmental regulations. |
| Individualist | Individualist belief systems emphasize innovation, economic growth, and the role of market mechanisms (Thompson, 2018) in solving air pollution. Narratives might promote the development and adoption of new technologies like electric vehicles and market-based solutions such as carbon trading. |
| Egalitarian | Egalitarian belief systems stress community involvement and the impacts of air pollution on public health, particularly for the most vulnerable populations. They advocate for policies that ensure equal distribution of clean air as a shared resource and demand radical changes (Thompson, 2018) to reduce emissions, such as banning certain pollutants outright. |
| Fatalist | Fatalist belief systems are characterized by skepticism about the efficacy of any interventions (Thompson, 2018). In the context of air pollution, this translates into narratives that depict efforts to improve air quality as doomed to fail due to overwhelming systemic challenges or corruption. |
| Modernist | Modernist belief systems emphasize the power of scientific progress and technological innovation to solve problems (Burns, 1913), including environmental ones. This belief supports large-scale technological solutions to air pollution, such as the installation of state-of-the-art air purification systems or the development of advanced low-emission public transportation. |
| Traditionalist | Traditionalist belief systems emphasize the importance of cultural heritage, continuity, and adherence to historical lifestyles and practices (Legenhausen, 2002). This belief advocates for the preservation of traditional practices that have a smaller ecological footprint or criticizes modern industrial methods for disrupting natural and social orders. |
| Activist | Activist belief systems focus on direct action and social change, particularly in the face of perceived government inaction or corporate malfeasance. Narratives driven by this belief system mobilize public demonstrations or campaigns to pressure policymakers into taking action against air pollution. |
| Technocratic | Trusts in experts and technical solutions over political or public opinion, emphasizing the role of educated elites and technologists in crafting policy solutions (Bertsou & Caramani, 2020). This could lead to advocacy for solutions based on scientific research and data-driven approaches. |

The aforementioned elements constitute the codebook utilized in this study. Research employing the Narrative Policy Framework (NPF) can be conducted at various levels of analysis.

### 2.1.2. Levels of Analysis

The NPF distinguishes between three primary levels of narrative analysis:

**Micro Level:** Focuses on public opinion and the individual-level effects of narrative persuasion. This level examines how specific narrative elements influence personal beliefs and attitudes towards policy issues.

**Meso Level:** Concerns the strategic use of policy narratives by groups or coalitions to advance their policy agendas. At this level, the analysis explores how narratives are constructed and deployed to shape policy discourse and influence policy outcomes.

**Macro Level:** Examines the cultural and institutional narratives that operate at a broader societal level. This includes the long-term and widespread narratives that influence public policy across different contexts and time periods.

The research design depends on the level of analysis. This study focuses on the meso level because it is particularly well-suited for examining specific policy subsystems, which consist of groups of actors such as government agencies, non-governmental organizations, and industry stakeholders engaged in a particular policy domain. In the context of air pollution in India, a meso-level analysis allows for an in-depth exploration of the narratives and interactions within a particular policy subsystem, such as environmental regulation, public health, or urban planning. This level of analysis is preferable to the micro level, which would focus on individual actors, or the macro level, which would consider the nation as a whole.

Furthermore, the issue of air pollution in India is characterized by significant complexity, involving a diverse array of stakeholders, regional variations, and intersecting policy areas, including energy, transportation, and public health. A meso-level analysis provides a framework for capturing this complexity without being overwhelmed by the granular details of individual narratives, as would occur at the micro level, or the overly broad generalizations typical of macro-level analysis.

At the meso level, patterns of coalition formation, the alignment of narratives, and the influence of various actors within the policy subsystem can be effectively identified and analyzed. This level of analysis is particularly advantageous for understanding how specific narratives surrounding air pollution gain traction, shape policy decisions, and facilitate interaction and negotiation among stakeholders within the subsystem.

Lastly, the meso level offers an intermediate scope that balances depth and breadth, making it well-suited for case studies that require a detailed exploration of narratives and their impacts within a manageable and contextually relevant framework. This makes it particularly appropriate for a case study on a complex issue like air pollution in a diverse and populous country such as India.

### 2.1.3. Research Design and Methods

NPF studies can employ both experimental and non-experimental designs:

**Experimental Designs:** Typically involve control groups and various narrative treatments to study their effects on dependent variables.

**Non-Experimental Designs:** Include qualitative approaches, content analysis, and descriptive statistics to explore the relationships and impacts of policy narratives.

This study employs a non-experimental design due to its exploratory nature, with the objective of gathering insights into the policy problem under investigation. A more detailed discussion of the research design is provided in chapter 3. Data collection is structured according to the level of analysis, with various types of data being available at different levels.

### 2.1.4. Data Collection and Analysis

**Micro-Level Data:** Collected through surveys, interviews, and experimental treatments.

**Meso-Level Data:** Derived from public records, media content, speeches, and legislative documents.

**Macro-Level Data:** Comprised of cultural narratives and institutional communications.

The Narrative Policy Framework is the underlying theoretical basis for this study. Based on the framework, the analysis of narrative data will involve coding for specific narrative elements through manual and automated methods. The methodological approaches and the rationale behind them are elaborated in chapter 3.

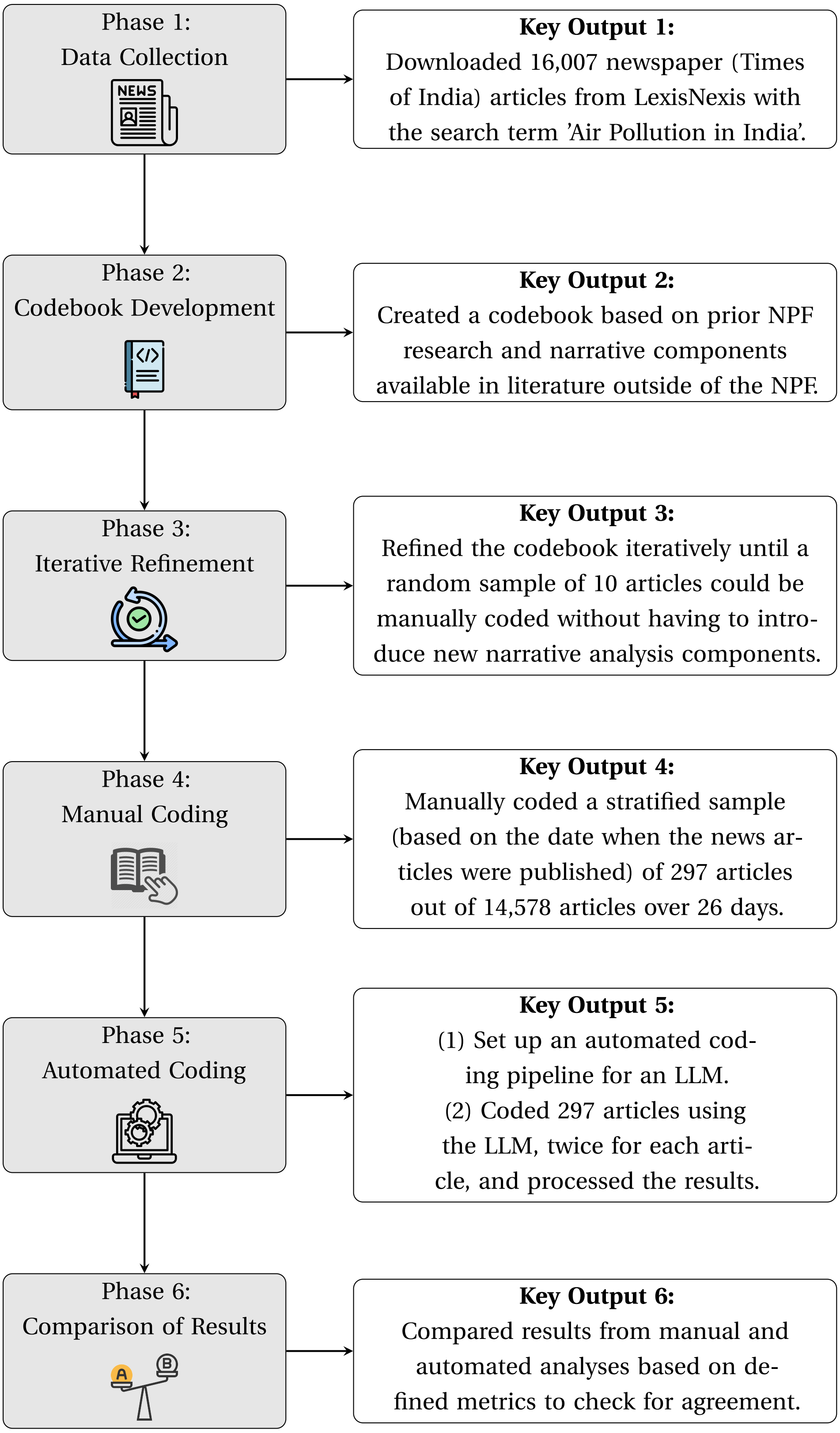
# 3

# Research Design

Th research design is based on a case study approach, with the topic of the case study being 'Air Pollution in India'. The research is designed to examine how diverse groups construct policy narratives by elucidating insights derived from both manual coding and automated coding, and facilitating a comparative analysis of these methods. Such an inquiry is characteristic of a meso-level research approach, which typically investigates the interplay between policy narratives and policy outcomes within specific contexts at the coalition level for actors and stakeholders (Jones et al., 2022). The central goal of the research is to do an exploratory study of the policy problem and comparing the outcomes of the manual analysis with the automated analysis. By adopting this methodology, the study seeks to enhance our understanding of the potential advantages and limitations associated with advanced automated coding techniques, particularly those enabled by large language models (LLMs), in the realm of qualitative policy research.

This study employs qualitative content analysis as a method of doing text analysis within the case study, a method extensively utilized within the Narrative Policy Framework (NPF) research domain (Jones et al., 2022). In addition, qualitative content analysis, with its systematic, theory-guided procedures, is well-suited for analyzing complex social phenomena in case studies (Kohlbacher, 2006) and can significantly enhance the rigor, validity, and reliability of case study research (Kohlbacher, 2006). Specifically, it adopts a meso-level, non-experimental, time-series design. The non-experimental nature of the study is characterized by the absence of control and treatment groups, while the time-series dimension is defined by the examination of narrative policy data pertaining to the chosen case study over the period from 2010 to 2024, based on available data at the time of this study. The study utilizes newspaper articles as the meso-level data source to explore various groups and coalitions.

## 3.1. Steps in Conducting the Research




## 3.2. Data - Census and Sample

**Census:** The data for this study consists of newspaper articles on the topic of "Air Pollution in India" sourced from 'The Times of India'. This newspaper was selected due to its extensive readership in India, the availability of English-language articles, and the ease of accessing these articles online.

Articles were retrieved using the Lexis Nexis, Nexis Uni database. The search term "Air Pollution in India" was used, with a filter applied to include only articles from 'The Times of India'. This search strategy ensured that each article contained the terms 'air', 'pollution', and 'India', making them relevant to the policy issue under study (Lexis Nexis Search Guide). Articles published between January 2010 and April 2024 were considered (all available articles in existence at the time of download), resulting in an initial total of 16,007 articles. Due to download limits, articles were downloaded in batches of 100, with a cap of 1000 articles per session (Lexis Nexis Download Limits).

The downloaded articles were processed using Python (Jupyter Notebook). This involved extracting the articles into a structured tabular format and removing duplicates, resulting in a final dataset of 14,578 news articles. The duplication in articles were likely due to manual downloading errors. Although Lexis Nexis offers ways to download more than the aforementioned limits through special request, this was outside the time and financial scope of the study.

The dataset had an average of 405.7 words per article, with a median of 377 words. Initial data analysis included examining the number of articles published each year, extracting keywords, and mapping article locations (Python Notebook). Here are some insights from the raw data -

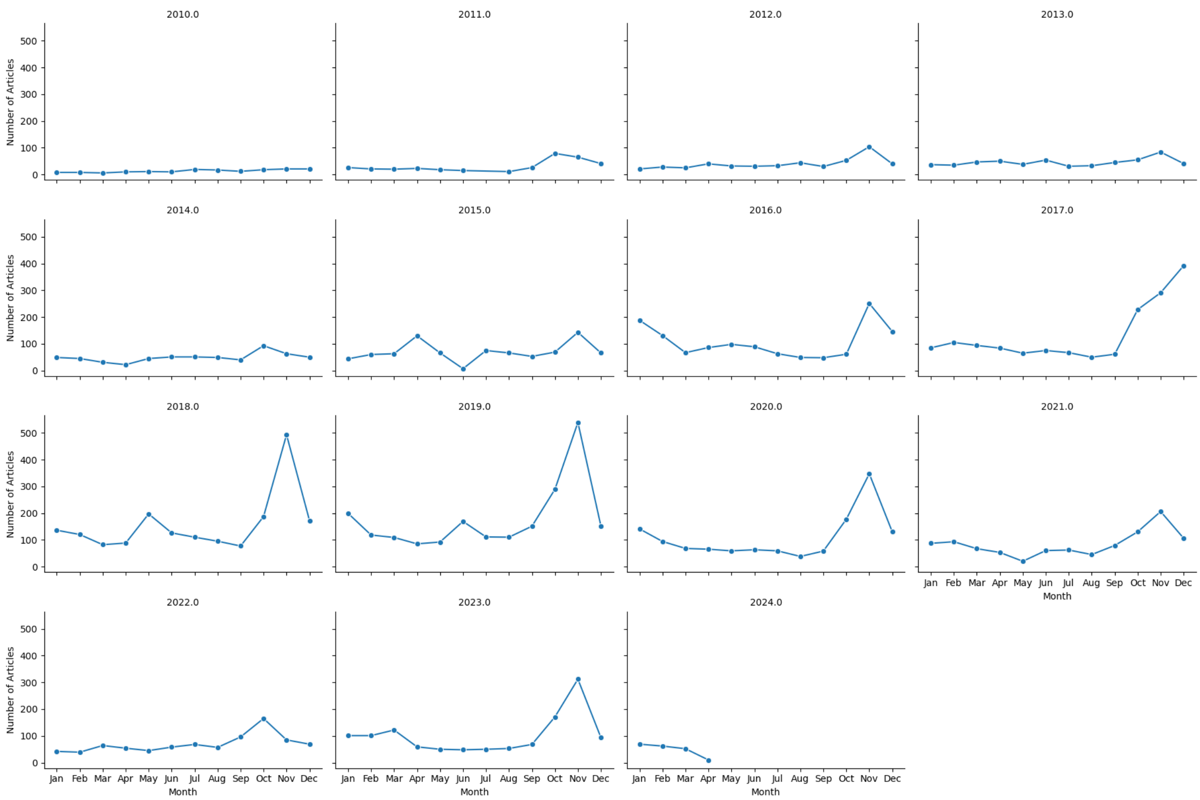



**Figure 3.1:** Number of Articles over the Years

**Sample:** A stratified sample of 297 articles was drawn out of the total data. The stratification was done basis the share of articles in a particular time period. The share of articles by year is the same (upto 2 decimal points of the percentage) in the census and the sample.

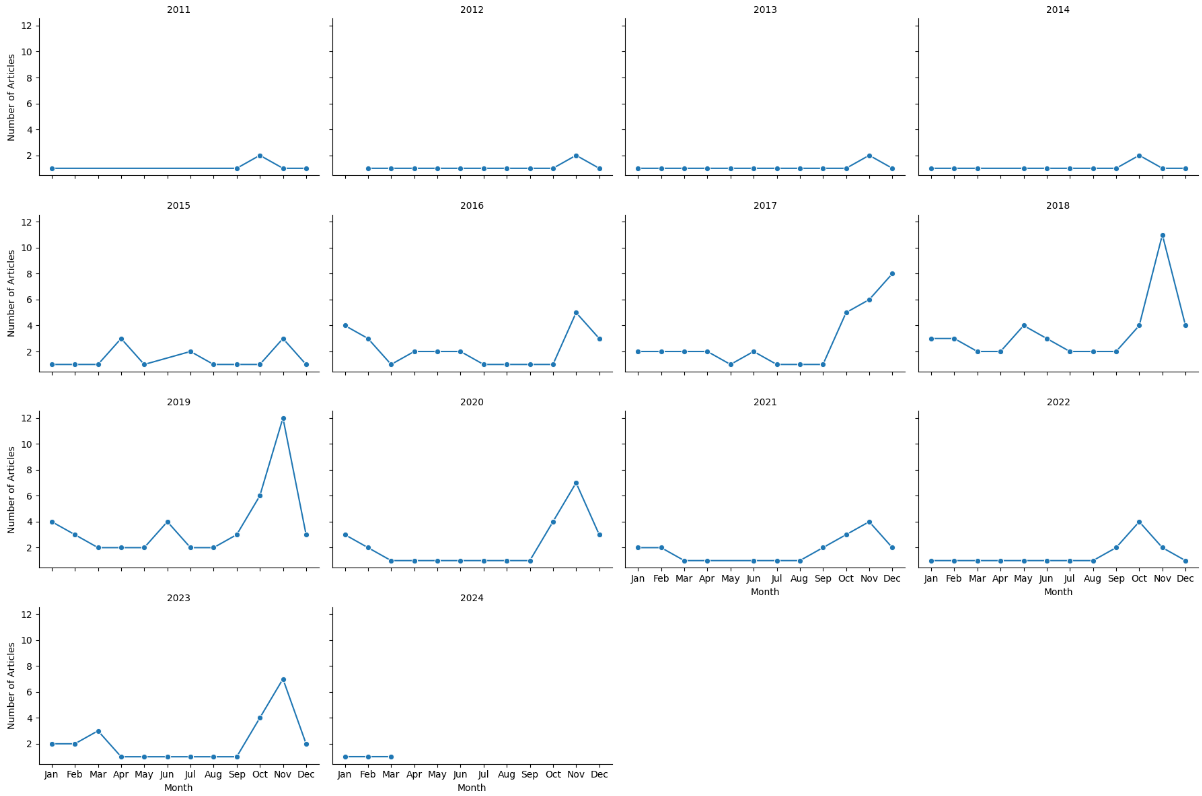
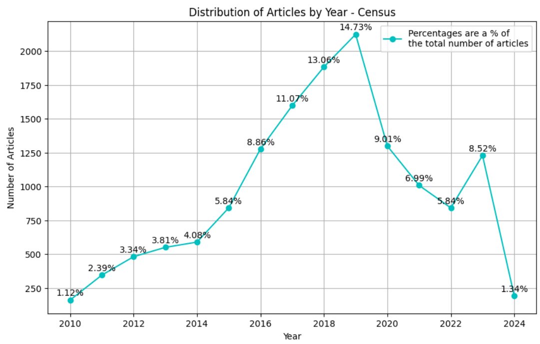


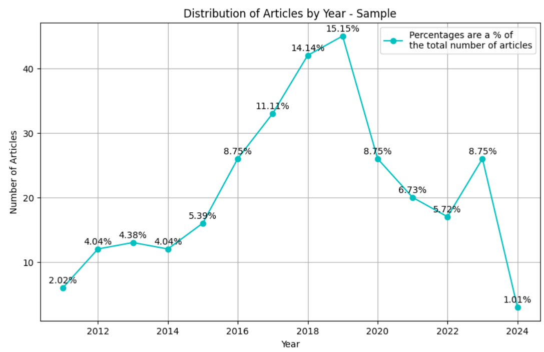

**Figure 3.2:** Number of Articles over the Years

## 3.2.1. Representativeness of the Sample:

The sample was drawn to be representative of the census. To validate its representativeness, the following visualizations were created -



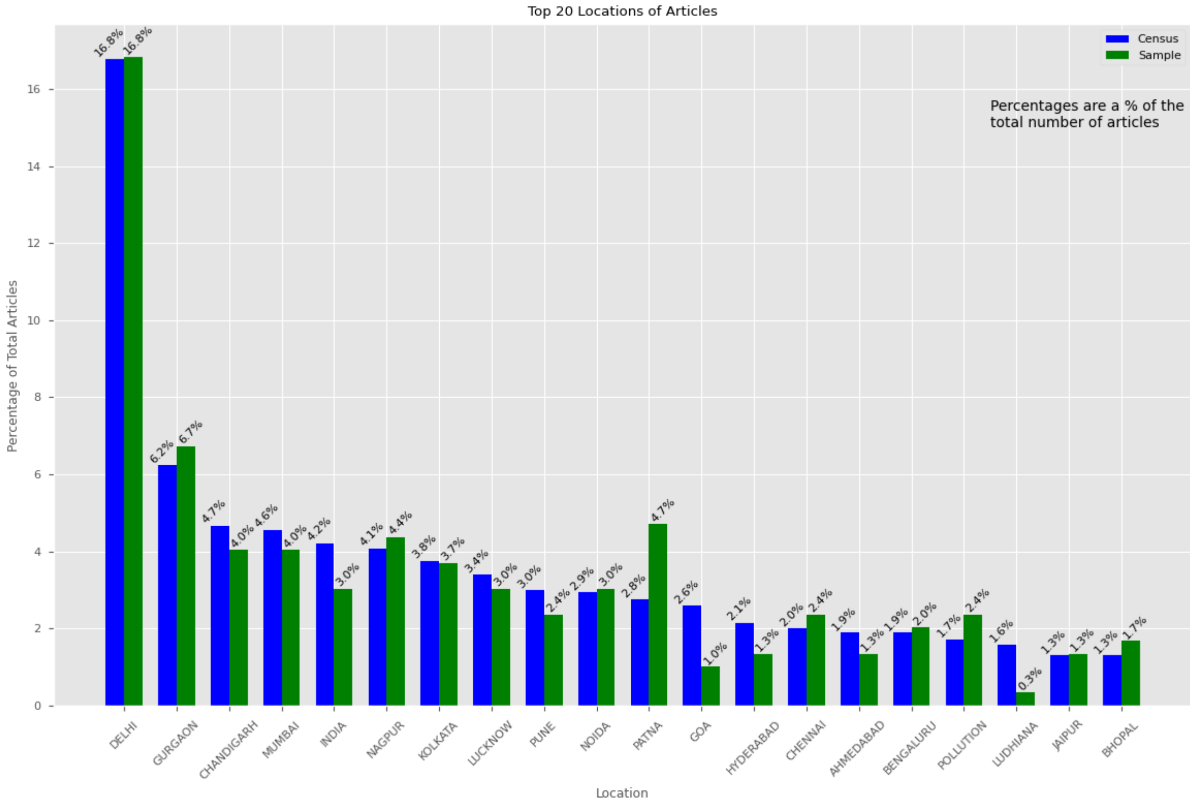

**(a)** Number of Articles by Year - Census



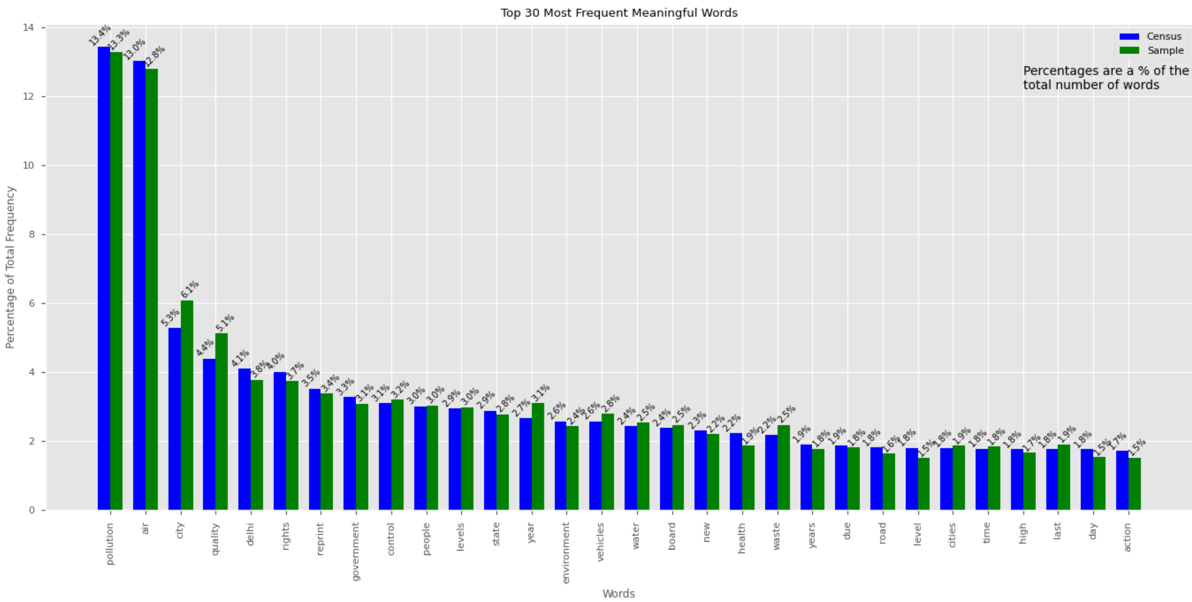

**(b)** Number of Articles by Year - Sample

**Figure 3.3:** Census-Sample Distributions

## Comparison of Census and Sample - Data Description




**(a)** Comparison of Number of Articles by Publishing 'Section' Across All Years




**(b)** Most Frequent Words (noise words removed) Across All Years

**Figure 3.4:** Census-Sample Comparison


**(a)** Latent Dirichlet Allocation (LDA) Method Word Cloud Across All Years - Census

**(b)** LDA Method Word Cloud Across All Years - Sample

**(c)** Latent Semantic Analysis (LSA) Method Word Cloud Across All Years - Census

**(d)** LSA Method Word Cloud Across All Years - Sample

**Figure 3.5:** Census-Sample Comparison - Word Clouds

Although there is no metric for comparison here, glancing at the word clouds shows that the word cloud of the sample is a reduced version of the word cloud of the census. These figures are merely meant to be representative to obtain a rough idea of the content of the dataset.

## 3.3. Defining the Codebook

One self-trained trained human coder created the codebook by taking elements from the standard NPF codebook and customizing it further by analyzing the sample narrative data. Ideally, two or three independent coders should read and code a small sample of the data independently. They then meet to reconcile their codes, discussing discrepancies and establishing decision rules. Due to budget limitations, this was not possible in this study. There were two parts to creating the codebook for this study:

**Components from the Standard Codebook:** The NPF employs a standard codebook that includes definitions and operationalizations of narrative components such as setting, characters, plot, and moral of the story. The codebook for this study was defined by definitions of narrative form and content used in contemporary studies in the analysis of air pollution policy narratives in India (Olofsson et al., 2018), (Costie & Olofsson, 2022), and other NPF studies (Shanahan et al., 2018b).

**Customization of the Codebook:** The codebook was used to code 10 random news articles, and iteratively changed until it contained the required elements needed to comprehensively code the breadth of articles present in the dataset. The codebook can be viewed in section A.3.

### 3.3.1. Unit of Coding Analysis

**Levels of Analysis:** The unit of analysis can vary from headlines, sentences, and paragraphs to entire documents or collections of documents. To check for which level of analysis was suitable for this study, 2 articles were coded at a paragraph level and at the document level each, and the results were slightly different at different levels of analyses. To maintain research rigor and quality, and to ensure that the

LLM and manual coding happen at the same level, the unit of analysis was chosen to be at the document level.

## 3.4. Manual Coding

**Human Coding:** The process of manual coding was conducted by a single, self-trained coder, who is a graduate STEM student. As previously discussed, the codebook was iteratively developed by this coder. Prior to commencing the final coding process, the coder thoroughly familiarized themselves with the codebook, repeatedly reviewing each definition. Documented decision rules were consistently applied throughout the coding process. The coder successfully completed the coding of 297 articles over a span of 26 days, with an average of approximately 12 articles coded per day. Each news article was read multiple times before being coded. To ensure consistency and prevent burnout, regular breaks were incorporated into the coding process. The coding was performed in Excel, with characters and settings written out as they appear in the text while being as consistent between articles for repeating entities as far as possible (example: ’police’, ’director’, ’Delhi’, ’New Delhi’, etc.), separated by commas wherever appropriate. For plot, moral, belief system, and narrative strategy, each column had a data validation check, where only pre-defined categories could be selected from a drop down list.

## 3.5. Automated Coding

For the automated coding process, a large language model was selected due to its demonstrated effectiveness in processing extensive amounts of natural language data with high accuracy and top percentile scores on natural language tasks (Gemini Performance, 2024). The newspaper articles and the codebook were inputted into the chosen large language model, which was tasked with identifying and outputting the narrative elements in a JSON format to facilitate post-processing. The primary advantage of automated coding lies in its ability to efficiently handle large datasets, 14,578 newspaper articles for instance.

### 3.5.1. Selection of Large Language Model

The Google Gemini large language out-of-the-box model with no fine-tuning and default hyperparameters was selected for this study, as there are many ways to customize this family of models for high performance on specific tasks. Although customizing may provide better results, this was outside the scope of this study due to time constraints. Google Gemini demonstrates state-of-the-art performance, comparable to GPT-4 on various reasoning tests (Google Blog) as seen in the table below. Additionally, Gemini offers a limited number of free API requests through their Google AI Studio environment, along with $300 in sign-up credits for making API calls outside of Google Studio. This allocation was sufficient for the purposes of this study. The choice of Gemini was influenced by its cost-effectiveness, free usage, and ease of use in conjunction with Google Colab, the online Jupyter notebook platform utilized for all coding in this research. The specific variant of Gemini employed was the 1.5 Flash version, which Google describes as their "fastest multimodal model with excellent performance for diverse, repetitive tasks and a 1 million context window" (Google AI Studio). The performance of Gemini 1.5 (also known as Ultra) on various benchmarks is detailed below.

**Table 3.1:** Comparison of Gemini Ultra and GPT-4 Capabilities

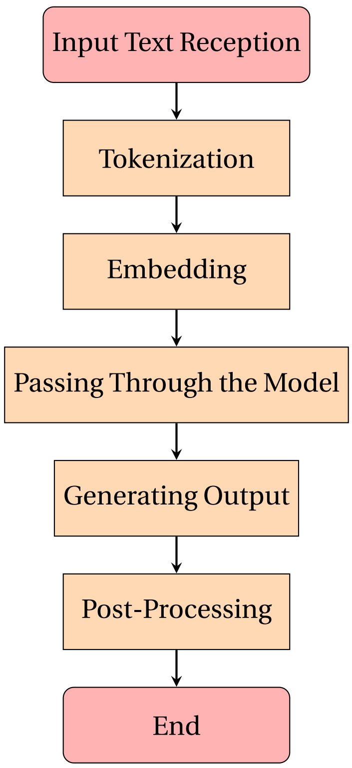
| Capability | Benchmark | Description | Gemini Ultra | GPT-4 |
|---|---|---|---|---|
| **General** | MMLU | Representation of questions in 57 subjects (incl. STEM, humanities, and others) | 90.0% CoT@32* | 86.4% 5-shot** (reported) |
| **Reasoning** | Big-Bench Hard | Diverse set of challenging tasks requiring multi-step reasoning | 83.6% 3-shot | 83.1% 3-shot (API) |
| | DROP | Reading comprehension (F1 Score) | 82.4 Variable shots | 80.9 3-shot (reported) |
| | HellaSwag | Commonsense reasoning for everyday tasks | 87.8% 10-shot* | 95.3% 10-shot* (reported) |
| **Math** | GSM8K | Basic arithmetic manipulations (incl. Grade School math problems) | 94.4% maj1@32 | 92.0% 5-shot CoT (reported) |
| | MATH | Challenging math problems (incl. algebra, geometry, pre-calculus, and others) | 53.2% 4-shot | 52.9% 4-shot (API) |
| **Code** | HumanEval | Python code generation | 74.4% 0-shot (IT)* | 67.0% 0-shot* (reported) |
| | Natural2Code | Python code generation. New held out dataset HumanEval-like, not leaked on the web | 74.9% 0-shot | 73.9% 0-shot (API) |

* See the technical report for details on performance with other methodologies
** GPT-4 scores 87.29% with CoT@32 - see the technical report on Google AI Studio for full comparison

**Table 3.2:** Gemini Ultra vs GPT-4 (Google AI Studio)

### 3.5.2. LLM Process

Gemini 1.5 Flash follows a sequence of steps to process textual data, following a structured pipeline from the reception of user input to the generation of the output. The steps for this study were as follows:



1. **Input Text Reception:** First, the textual input was provided as described.
2. **Tokenization:** The input text was then tokenized, i.e. it was decomposed into smaller units known as tokens. These tokens could be words, subwords, or even individual characters, depending on the Google Gemini Tokenizer. Tokenization is a critical step because LLMs function on tokens and not entire words as read and understood by humans. The tokenization was taken care of by the Google Gemini API call and initializing an instance of the 'GenerativeModel' in the 'generativeai' class:

```
!pip install -q -U google-generativeai
import google.generativeai as genai

model = genai.GenerativeModel('gemini-1.5-flash')
```

3. **Embedding:** Subsequently, the tokens were transformed into embeddings. These embeddings are high-dimensional vectors representing the tokens within a continuous abstract vector space, encapsulating semantic information about the tokens.
4. **Passing Through the Model** The token embeddings were then passed through the pre-trained model, which typically comprises multiple layers of self-attention mechanisms and feed-forward neural networks (Vaswani et al., 2017). Each layer processes and refines the embeddings based on the context provided by the entire input sequence.
   (a) **Self-Attention Mechanism:** The self-attention mechanism allowed the model to assign weights to each token relative to others in the sequence, thereby capturing dependencies and relationships between tokens (Vaswani et al., 2017).
   (b) **Feed-Forward Neural Networks:** Following the self-attention mechanism, the embeddings were processed through feed-forward neural networks, which further transformed the information.
5. **Generating Output:** The final layer of the model generated output embeddings, which were then converted back into tokens. These tokens were subsequently decoded to produce the final text output. The following code was used to generate and decode the output:

```
from IPython.display import Markdown

def to_markdown(text):
  text = text.replace('', '  *')
  return Markdown(textwrap.indent(text, '> ', predicate=lambda _: True))

response = model.generate_content("The prompt goes here")
print(to_markdown(response.text))
```

This was the response for a single query for testing the output. The whole program was run through a loop with a rate limit on requests to avoid charges beyond the free tier of the Gemini API[1], and saved at regular intervals to prevent data loss in the case of interruptions. There was also a provision added to load from a checkpoint in case of the aforementioned network loss.

```
    # Determine the starting point
start_index = df.dropna(subset=['Hero']).shape[0]  # Rows with 'Hero' filled are processed
for index in range(start_index, len(df)):
```

[1]At the time of implementation> The current billing scheme may be different.

```
        start_time = time.time()

        # The extract function
        result = extract(df.loc[index, 'Body'])
        df.at[index, 'Gemini Output'] = result
            # Save the checkpoint every 15 rows
        if (index + 1) % 15 == 0:
            save_checkpoint(df)
            print(f"Checkpoint saved at row {index + 1}")

        elapsed_time = time.time() - start_time
        if elapsed_time < 4:
            time.sleep(4 - elapsed_time)
    # Save the final state of the dataframe
    df.to_csv(final_output_path, index=False)
    print("Analysis complete. Final data saved.")
```

6. **Post-Processing** The generated text underwent post-processing to ensure it conformed to the desired format and quality. The JSON parts were extracted as a python dictionary and appropriate columns in the python pandas dataframe were populated with the elements and saved as a comma separated values file which could be accessed in Excel. Errors in the JSON output were checked for and rectified. Sometimes, the commas and apostrophes were at the wrong place, but these instances were few (4/297), which highlights the reliability of Gemini in terms of output quality. The postprocessing was done using pandas (Wes McKinney, 2010), numpy (Harris et al., 2020), and sci-kit learn (Pedregosa et al., 2011).

## 3.6. Validation of Results

### 3.6.1. Testing

Following the coding phase, the performance of the large language model (LLM) was systematically evaluated. For the narrative components, predetermined metrics were employed to determine the concordance between manual and automated coding, as well as between two separate runs of the automated coding process. Specifically, for elements such as plot, moral, narrative strategy, and belief system, the metrics of accuracy, recall, precision, and F1 score were computed, commonly used metrics in the field of machine learning for classification task evaluation (Rainio, Teuho, & Klén, 2024). In the case of characters and settings, cosine similarities were utilized due to the nature of these elements as part of an entity recognition task, rather than classification from a predetermined set of output classes. The metrics of accuracy, precision, recall, and F1 score assess equivalence in the most rigorous terms and tend to perform suboptimally if characters are not consistently present with identical spellings in the same order across the two coded datasets, making them inappropriate for characters.

### 3.6.2. Accuracy, Precision, Recall, F1 Score, Krippendorff's $\alpha$

**Accuracy:** The proportion of correct predictions (both true positives and true negatives) out of all predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where,

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

**Precision:** The ratio of true positives to the sum of true positives and false positives. It measures the accuracy of positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

**Recall:** The ratio of true positives to the sum of true positives and false negatives. It measures the ability to find all positive instances.

$$\text{Recall} = \frac{TP}{TP + FN}$$

**F1 Score:** The harmonic mean of precision and recall. It balances the two metrics and is useful when you need a single measure of a model's performance.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

**Krippendorff's $\alpha$:** It is a statistical measure of the reliability of agreement between observers or raters. Social scientists typically utilize data with reliability coefficients ($\alpha$) of 0.800 or higher. Data with reliability coefficients between 0.667 and 0.800 are considered for drawing tentative conclusions, while data with reliability coefficients below 0.667 are generally disregarded due to insufficient agreement (Krippendorff, 2013).

$$\alpha = 1 - \frac{D_o}{D_e}$$

where:

$D_o$ is the observed disagreement, calculated as:

$$D_o = \frac{1}{N} \sum_c \sum_{i \neq j} o_{c,i} \cdot o_{c,j} \cdot d_{i,j}$$

$D_e$ is the expected disagreement by chance, calculated as:

$$D_e = \frac{1}{N} \sum_{i \neq j} \sum_c (o_{c,i} \cdot o_{c,j}) \cdot d_{i,j}$$

$N$ is the total number of observed pairs. $o_{c,i}$ and $o_{c,j}$ are the observed proportions for category $c$ by raters $i$ and $j$. $d_{i,j}$ is the distance function, which varies depending on the type of data (e.g., nominal, ordinal). The data in this study is nominal.

### 3.6.3. Cosine Similarities

As stated in subsection 3.6.1, cosine similarities were calculated for identified characters in the dataset. Traditional metrics, such as recall, precision, accuracy, and F1 scores, are typically employed in classification tasks. These metrics evaluate a model's performance by comparing correctly identified instances relative to the actual instances. However, they are less suitable for tasks involving semantic comparison of text, as equivalence between two datasets in this context extends beyond character-by-character comparison. Characters might be identified in a different order in the output, or they might have different spellings in cases of multiple instances of the same character within a data point. Consequently, equivalence cannot be accurately captured by the aforementioned metrics. There are several advantages to using cosine similarities for evaluating the task of character recognition and labeling:

- Semantic Nuance Cosine similarity captures the subtle semantic nuances between textual descriptions of narrative roles. This is crucial for comparing character roles, where two different runs may generate text that varies in wording but is semantically equivalent.
- Robustness to Variations in Text The use of embeddings and cosine similarity accounts for variations in wording and phrasing, focusing on the underlying meaning, where different phrasings can convey the same role or concept.
- Interpretable Results Average cosine similarity provides a straightforward, interpretable measure of consistency between the runs. Higher similarity indicates greater consistency in narrative role generation.
- Comparative Analysis By comparing the average similarities across different narrative roles, we can identify specific roles where the model exhibits higher or lower consistency, which is not easily captured by traditional metrics.

Methodology

1. **Model Initialization:** The `SentenceTransformer` model (Reimers & Gurevych, 2019), specifically `paraphrase-MiniLM-L6-v2`, was utilized to generate semantic embeddings of text. This model is pre-trained to capture paraphrased sentences' meanings in a high-dimensional vector space, enabling nuanced comparison of textual data.
2. **Selection of Columns:** The comparison focuses on specific narrative roles: Hero, Villain, Victim, Beneficiary, Ally, Opponent, Ineffective, and Setting. Coding for these elements was not a classification task like for the plot, narrative strategy, moral, and belief system.
3. **Embedding and Similarity Calculation:** For each narrative role, texts from two datasets (df1 and df2) were embedded using the `SentenceTransformer` model. Cosine similarity was calculated between the corresponding embeddings from the two datasets. Cosine similarity measures the cosine of the angle between two vectors in a multi-dimensional space, providing a metric for the textual similarity that ranges from -1 (completely dissimilar) to 1 (identical).
4. **Average Similarity Calculation:** The similarity scores for each role were averaged to provide an overall measure of consistency for each narrative role.
5. **Post-Processing:**To improve the accuracy and clarity of the N-grams, the following steps were taken:
   - Refine Coding Process: The coding methodology was adjusted to filter out methodological terms like 'implied', 'explicitly mentioned', and 'implicitly'. This was achieved by excluding these terms during the coding phase or through post-processing of the N-grams.
   - Manual Review: A manual review of the N-grams was conducted to identify and remove any methodological artifacts. This will help ensure that the N-grams more accurately reflect the substantive content of the articles.

- Automated Filtering: Automated filtering techniques were implemented to identify and exclude common methodological terms from the final N-grams. This was done using regular expressions and stopword lists.

6. **Visualization:** The average cosine similarities for each narrative role were visualized using a bar chart. All the visualizations were made using Matplotlib (Hunter, 2007), and Seaborn (Waskom, 2021).

Result visualizations were made wherever necessary to facilitate comparison in the form of bar graphs, maps, co-occurrence plots, and confusion matrices.

# 4

# Results - Manual Coding

## 4.1. Manual Coding Results - Insights into Air Pollution

The analysis of the coded articles revealed the following about the presence and absence of narrative components in the coded articles:

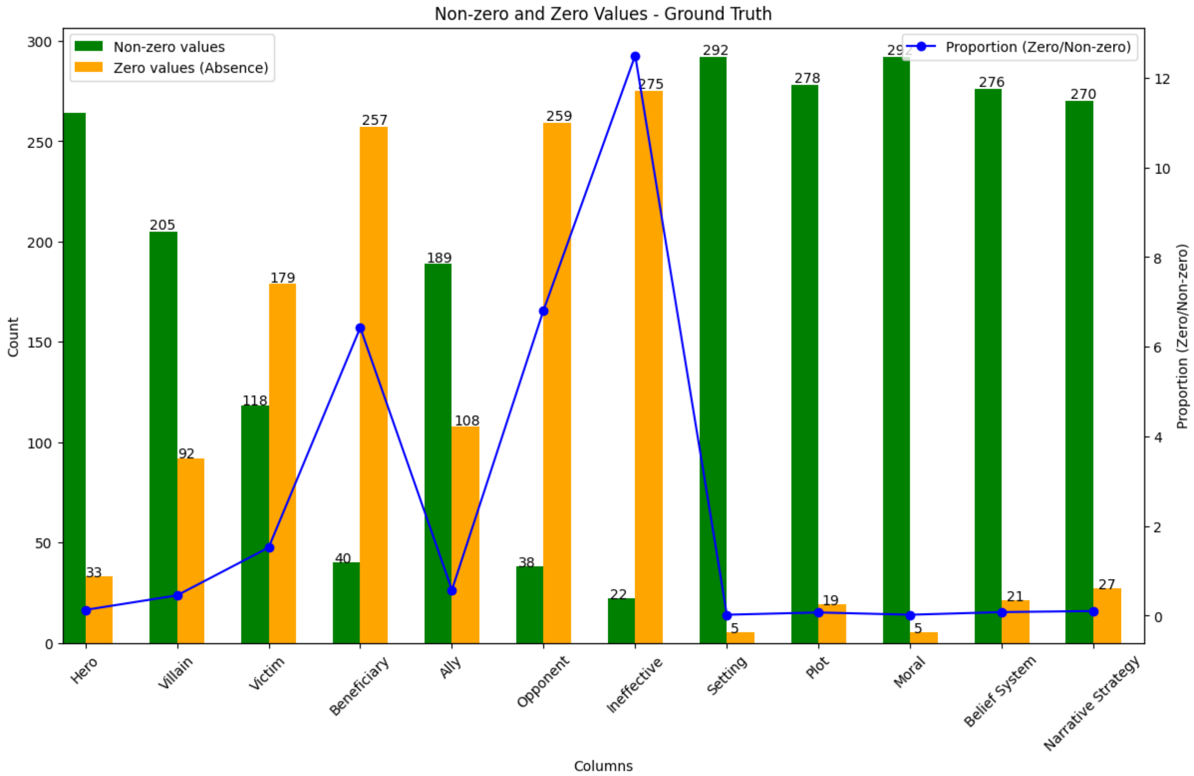



**Figure 4.1:** Presence and Absence of Narrative Components

As observed in the preceding analysis, policy narratives mention victims, beneficiaries, opponents, and ineffectives less often than they are omitted, whereas the incidence of heroes is notably the highest, occurring in 264 narratives among all sampled narratives. This predominance of heroes may be attributed to the narrative function they serve, indicating that measures are being undertaken to combat air pollution, or suggesting that the severity of the situation necessitates decisive action.

An in-depth analysis of all the components of the narrative provide more insight into the state of air pollution in India from 2010 to 2024.

### 4.1.1. Narrative Elements

Characters - Heroes

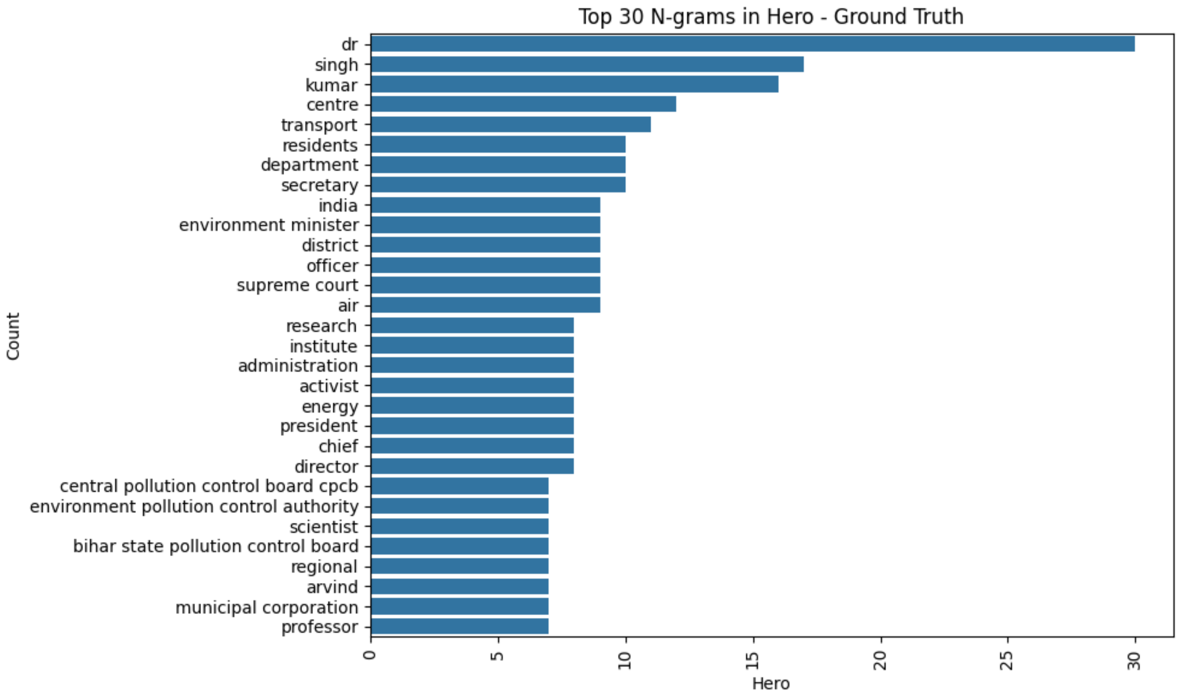



**Figure 4.2:** N-Gram of Heroes in Manual Coding

The hero is defined as the potential fixer of the policy issue, advocating for or taking positive steps towards air pollution mitigation. The analysis of N-grams (Figure 4.2) associated with heroes, reveals the following:

1. A significant prevalence of the term 'dr,' which, in this context, denotes Doctor of Medicine. The majority of the articles featuring doctors as heroes highlight their efforts in raising public awareness about the hazards of air pollution.
2. The surnames 'Singh'[1] and 'Kumar,' which are common in India, frequently appear in the data. These names often belong to individuals, either private citizens or those in influential positions, who actively voice their concerns about air pollution. They engage in various actions such as filing police reports against polluters, initiating legal proceedings, or formulating policies to address pollution issues within their capacity.
3. The term 'centre' refers to the Central Government of India, which is responsible for national gov-

[1]As present in the manual coding sheet - ['Jagbir Singh', 'N P Singh', 'Justice Dalip Singh and expert member P S Rao', 'pulmonary expert Dr B P Singh', 'housewife Smiriti Singh', 'additional solicitor general Maninder Singh', 'general secretary of Gurdwara Singh Sabha Jatinder Singh Sandhu', 'president of Gurdwara Dukhniwaran Sahib Pritpal Singh', 'Dr Virendra Singh', 'municipal commissioner Chandra Prakash Singh', 'Dr Virendra Singh', 'senior scientist and head of KVK Raj Karan Singh', 'HSPCB regional officer Kuldeep Singh', 'MLC Sanjeev Shyam Singh', 'Punjab environment minister Gurmeet Singh Meet Hayer', 'district agriculture officer KK Singh']

ernance and policy-making aimed at combating air pollution. Similarly, the state pollution control boards, central pollution control boards, environment pollution control authority, supreme court, ministers, and municipal corporations are identified as heroes. These entities, representing both state and central levels of government, play pivotal roles in regulating and mitigating air pollution.

4. Another category of heroes comprises secretaries, presidents, officers, ministers, chiefs, and directors of various civil and public institutions who contribute to the fight against air pollution.

5. Lastly, researchers, scientists, and professors are also recognized as heroes for their contributions. These individuals not only advocate against the dangers of air pollution but also engage in developing technologies to purify air, reduce emissions, and conduct research aimed at achieving a cleaner future.

### Characters - Villains

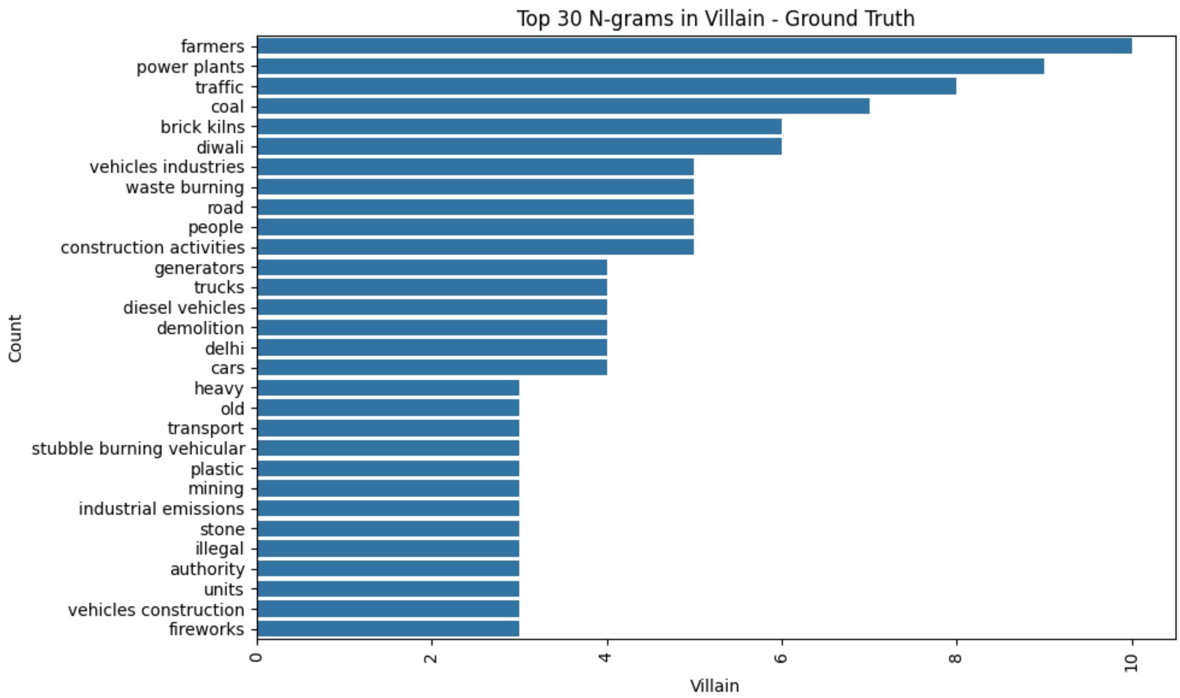



**Figure 4.3:** N-Gram of Villains in Manual Coding

The villain is defined as the entity causing the policy problem, creating harm or opposing the hero's aims. The N-grams of villains (Figure 4.3) identify the following:

1. The most frequently mentioned villain is 'farmers', often related to the practice of stubble burning, which significantly contributes to air pollution in the Northern plains of India, originating from the farming states of Punjab and Haryana, and spreading to the east towards the National Capital Region, Uttar Pradesh, Bihar, and so on.

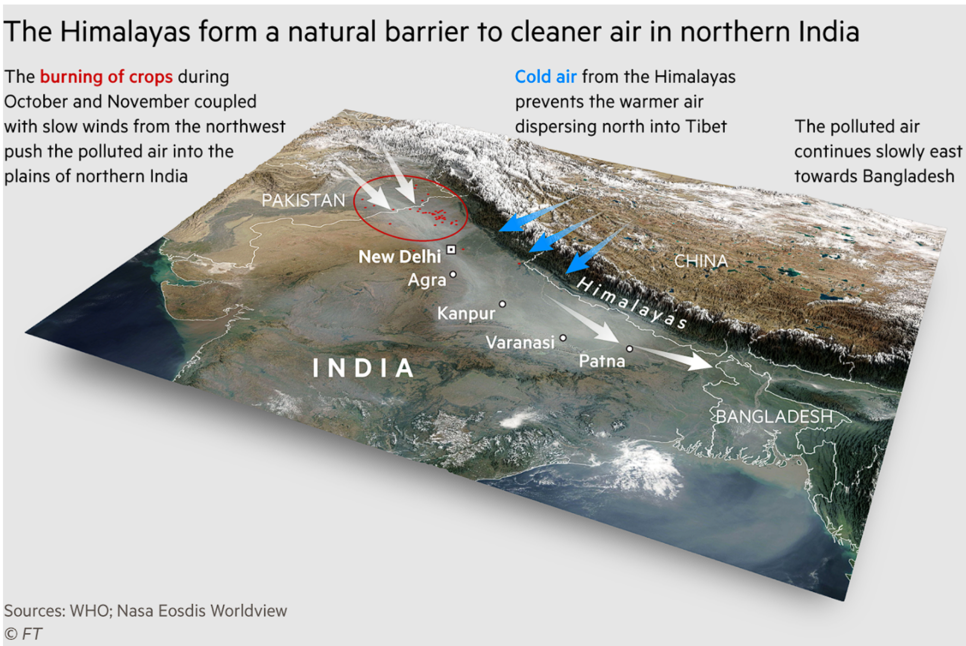



**Figure 4.4:** Direction of Air Flow in North India (Financial Times)

2. Next, we have ’power plants’ and ’traffic’, both of which are significant sources of emissions. ’Coal’ is another major villain, primarily due to its use in power generation and various industrial processes.
3. ’Brick kilns’ are another notable contributor, as they release substantial amounts of pollutants into the air. ’Diwali’ appears as a villain, due to the extensive use of fireworks during the festival, leading to a temporary but severe spike in air pollution. This can also be seen in Figure 3.1 with a spike in number of articles after September, as Diwali usually falls between mid-October to early November.
4. ’Vehicles industries’ and ’waste burning’ are also major sources of pollutants, contributing significantly to air quality degradation. ’Road’[2], ’people’[3], and ’construction activities’ are further listed, with construction activities generating a lot of dust and particulate matter.
5. ’Generators’, ’trucks’, ’diesel vehicles’, and ’demolition’ also appear as common villains, highlighting the diverse sources of air pollution. ’Delhi’ is frequently mentioned, indicating the high levels of air pollution in the city.
6. Other notable villains include ’heavy’[4], ’old’[5] (referring to old vehicles), ’transport’, ’plastic’, ’mining’, ’industrial emissions’, ’stone’[6] (stone crushing units), ’illegal’[7] (activities like waste burning), ’authority’[8], ’units’[9] (industrial), ’vehicles construction’, and ’fireworks’. Each of these contributes to air pollution through various means, from emissions to dust generation and beyond.

[2] ’digging roads’, ’road dust re-suspension’, ’construction work and road work’, ’road and C&D’

[3] ’people in the vicinity of the trash burning it’, ’people bursting firecrackers’, ’people lighting firecrackers’, ’people’, ’people burning tyres and plastic’

[4] As present in the manual coding sheet - ’heavy truck traffic’, ’heavy vehicles’, ’heavy traffic’

[5] ’old vehicles’, ’20-year old diesel vehicles’

[6] ’stone mining operations’, ’stone quarries’, ’stone crushers’

[7] ’Operator of the illegal dumpyard’, ’illegal crusher units’, ’operation of illegal industries’

[8] "Haryana’s State Environmental Impact Assessment Authority (SEIAA)", ’Noida Authority workers’, ’National Highway Authority of India (NHAI)’

[9] ’four highly polluting industrial units’, ’illegal crusher units’

### Characters - Victims

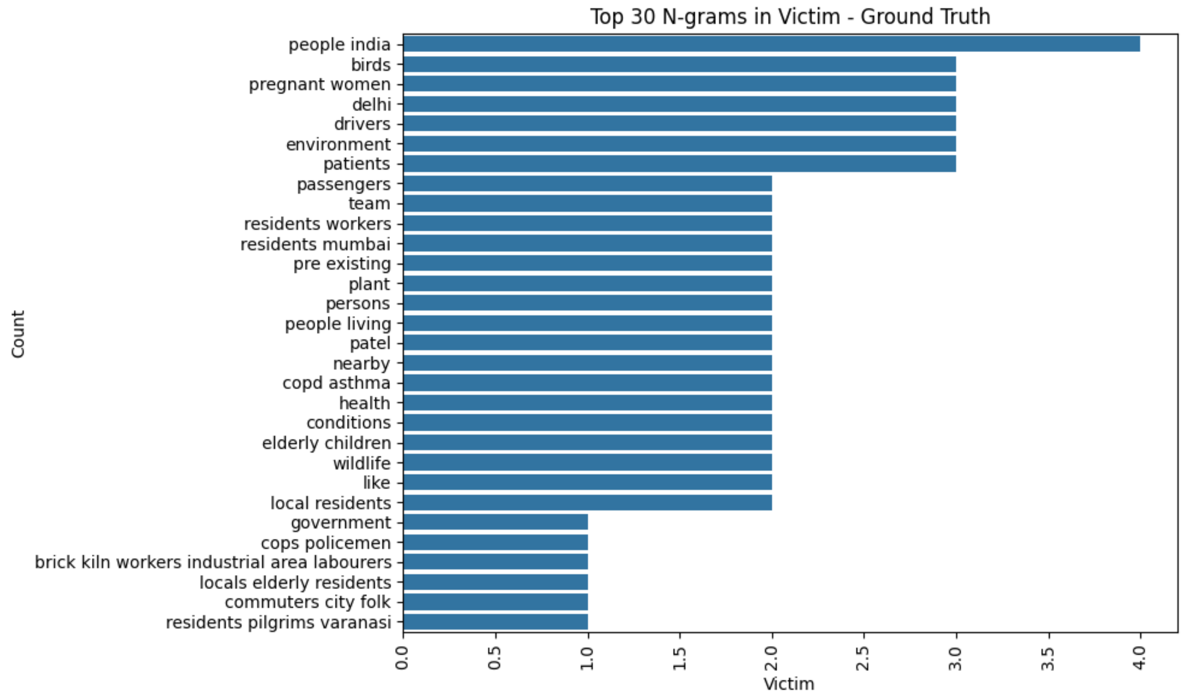



**Figure 4.5:** N-Gram of Victims in Manual Coding

The victim is the one harmed by the villain, or negatively affected by an action or inaction taking place in the determined setting. The N-grams of victims (Figure 4.5) show the following:

1. The most frequently mentioned victims are 'people india'[10], indicating the widespread impact of air pollution on the general population.
2. 'Birds' are also significantly affected, highlighting the ecological impact of air pollution. 'Pregnant women' and 'drivers' appear frequently, revealing the health risks to vulnerable populations and those frequently exposed to polluted environments.
3. 'Delhi'[11] is mentioned again, reflecting the city's severe air quality issues and its population being a victim. 'Environment' and 'patients'[12] are notable victims, with patients suffering from respiratory and other health issues exacerbated by poor air quality.
4. 'Passengers', 'team', 'residents workers', and 'residents mumbai' are also listed, indicating the widespread impact on various groups of people, from commuters to workers and residents of specific areas.
5. Other notable victims include 'pre existing' (referring to individuals with pre-existing 'health' 'conditions' like 'copd asthma'), 'plants', 'persons', 'people living' 'nearby', 'patel' (private individuals), 'elderly children', 'wildlife', 'local residents', 'government', 'cops policemen', 'brick kiln workers industrial area labourers', 'locals elderly residents', 'commuters city folk', and 'residents pilgrims varanasi'. Each of these groups experiences adverse effects from air pollution, highlighting the broad and varied impact on both humans and the environment.

[10] 'vulnerable people', 'people suffering from asthma', 'people living in the area', 'people with pre-existing breathing conditions', 'people of Kanpur'

[11] 'commuters on the Delhi-Gurgaon expressway', 'Delhiites', 'Delhi NCR', 'Delhi residents'

[12] 'patients of COPD and asthma'

### Characters - Beneficiaries

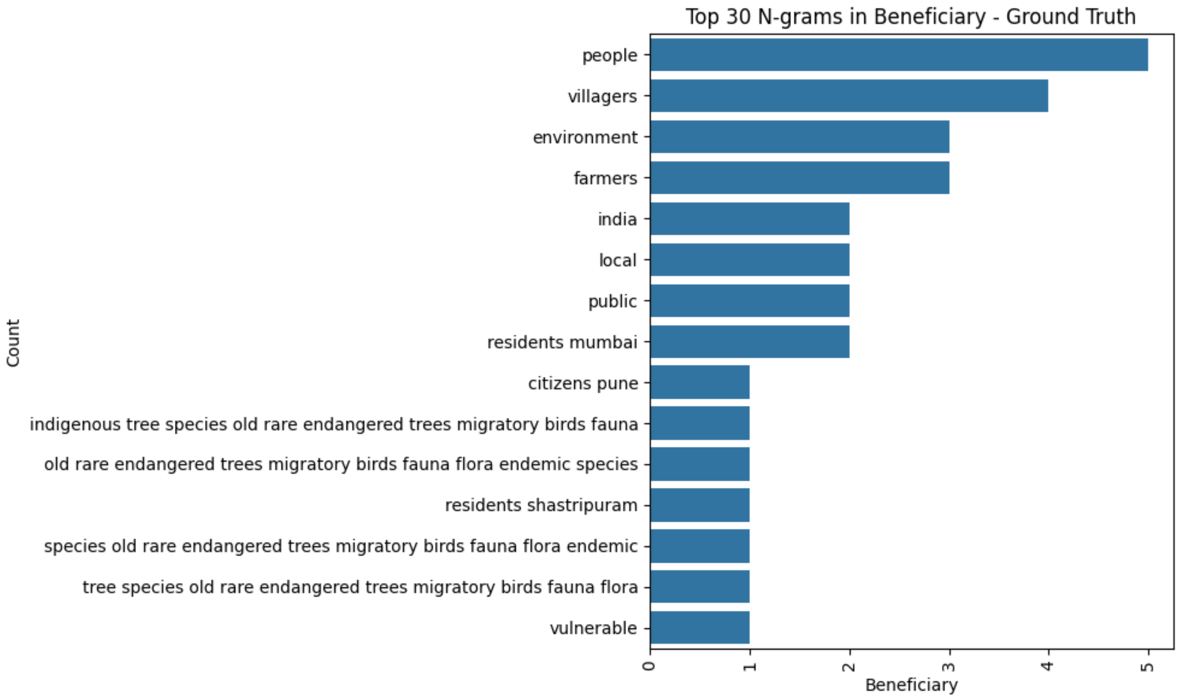



**Figure 4.6:** N-Gram of Beneficiaries in Manual Coding

Beneficiaries are those who benefit from the proposed policy solution, directly linked to a hero and receiving positive actions. The N-grams of beneficiaries (Figure 4.6) tell us the following:

1. The most frequently mentioned beneficiaries are 'people', indicating the broad impact on the general population.
2. 'Villagers' are also significant beneficiaries, likely due to the direct impact of clean air initiatives in rural areas. The 'environment' itself is a key beneficiary, highlighting the ecological benefits of reducing air pollution.
3. 'Farmers' benefit as well, possibly through reduced air pollution from practices like stubble burning. 'India' as a whole is a beneficiary, reflecting the nationwide impact of air pollution mitigation efforts.
4. Other notable beneficiaries include 'local', 'public', 'residents mumbai', 'citizens pune', and various mentions of 'indigenous tree species old rare endangered trees migratory birds fauna flora endemic species'. These groups and entities reflect the widespread and diverse impact of clean air initiatives, benefiting both human populations and the natural environment.

### Characters - Allies

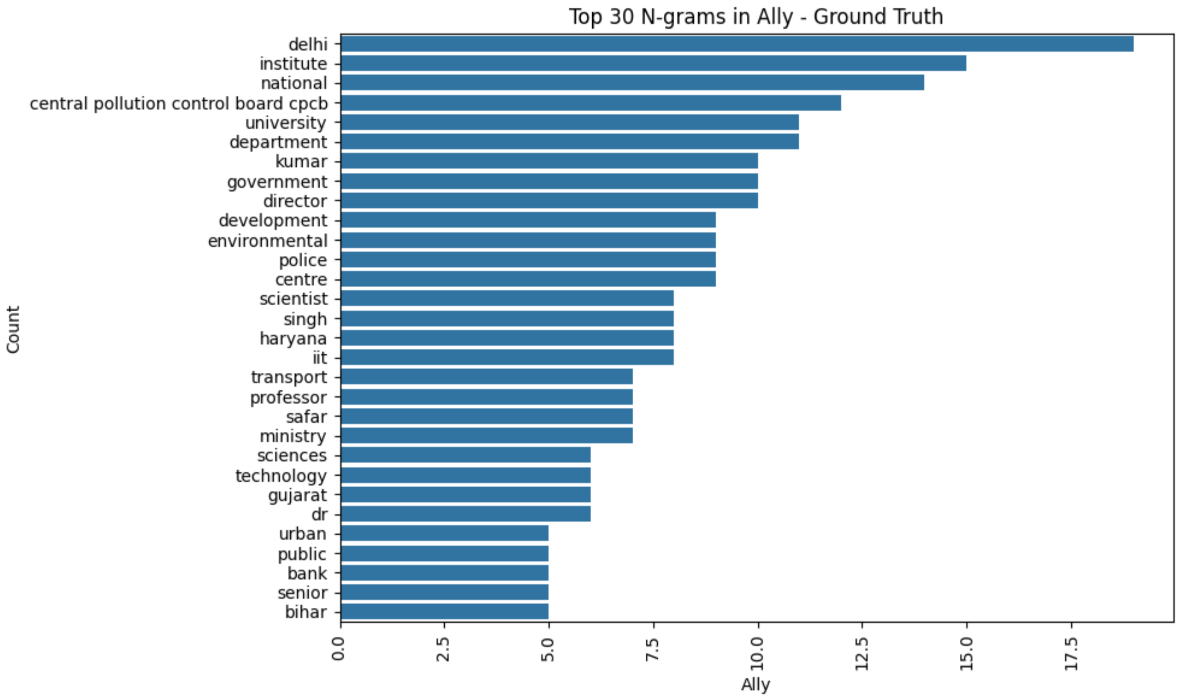



**Figure 4.7:** N-Gram of Allies in Manual Coding

Allies are those aligned with the hero, supporting their efforts towards the policy solution. The N-grams of allies (Figure 4.7) show the following -

1. 'Delhi'[13] is the most frequently mentioned, indicating various initiatives and efforts within the city by the government and people tied to Delhi institutions to combat air pollution.
2. 'institute', 'national', and 'central pollution control board cpcb' are notable allies, reflecting the involvement of research institutions and national regulatory bodies in addressing air pollution. 'University', 'department', and 'kumar' (a common surname in India) also appear frequently, suggesting the participation of educational institutions and individual advocates.
3. The 'government', 'director', and 'development' indicate the significant role of government officials and development initiatives. 'Environmental', 'police', and 'centre' (referring to the central government) are also key allies.
4. Other notable allies include 'scientist', 'singh' (common Indian surname), 'haryana' (a state in India), 'iit' (Indian Institutes of Technology), 'transport', 'professor', 'safar' (System of Air Quality Forecast and Research), 'ministry', 'sciences', 'technology', 'gujarat' (a West Indian state), 'dr' (Doctor), 'urban', 'public', 'bank', 'senior', and 'bihar' (a North Indian state). These allies represent a diverse group of entities, from government bodies and educational institutions to individual advocates and researchers, all working towards mitigating air pollution.

[13] 'New Delhi Municipal Council', 'municipal corporations of Delhi', 'Central Ground Water Authority and Delhi Development Authority', 'senior advocate Dushyant Dave for Delhi government', 'Delhi government', 'Delhi Police', 'Delhi government', 'Delhi University College of Medical Sciences', 'IIT Delhi', 'governments in Delhi', 'Delhi Pollution Control Board', "All-India Institutes of Medical Sciences in New Delhi", 'Delhi Traffic Police', "Delhi Congress' Ajay Maken", 'Delhi and Uttar Pradesh pollution department', 'IIT-Delhi professor Mukesh Khare', "Delhi government's urban department", "IIT-Delhi professor Sagnik Dey's SAANS (Satellite-Based Application For Air Quality Monitoring and Management at National Scale)", 'Delhi BJP president Virendra Sachdeva'

### Characters - Opponents

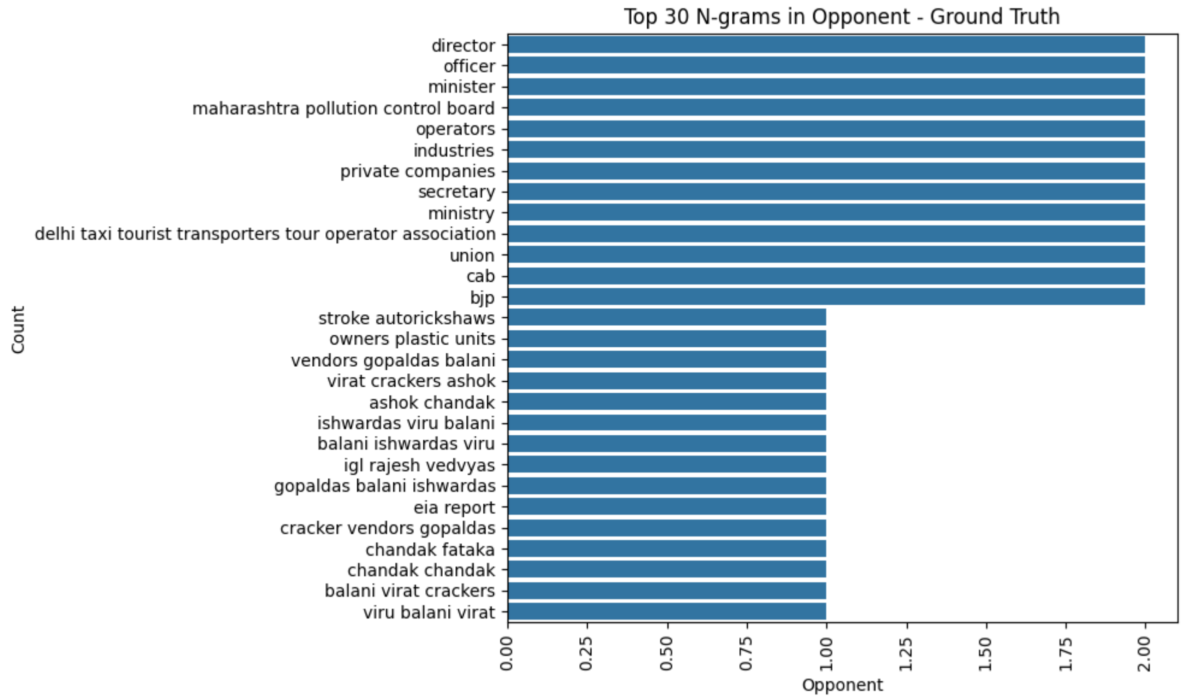



**Figure 4.8:** N-Gram of Opponents in Manual Coding

Opponents are entities opposing a policy but distinct from a villain, often presenting alternative views or objections. The N-grams of opponents (Figure 4.8) show that -

1. The most frequently (2 counts) mentioned opponents include 'director'[14], 'officer'[15], and 'minister"Former BJP minister S A Ramdas', 'chief minister Yogi Adityanath', reflecting individuals in positions of power who are resisting certain regulations or changes, possibly due to political reasons.
2. 'Maharashtra pollution control board' appears frequently (twice), suggesting potential resistance from this state-level regulatory body themselves. 'Operators', 'industries', and 'private companies' are notable opponents, indicating resistance from business interests potentially affected by stricter pollution controls.
3. 'Secretary', 'ministry', and various associations like 'delhi taxi tourist transporters tour operator association' also appear as opponents, reflecting institutional and organizational resistance.
4. Other notable opponents include 'union', 'cab', 'bjp' (right-wing political party with a majority in the government since 2014), 'stroke autorickshaws', 'owners plastic units', 'vendors gopaldas balani', 'virat crackers ashok', 'ashok chandak', 'ishwardas viru balani', 'balani ishwardas viru', 'igl rajesh vedvyas', 'eia report', 'cracker vendors gopaldas', 'chandak fataka', and other similar entities. These opponents represent a mix of political, business, and organizational resistance to air pollution mitigation efforts.

However, it is important to note that there are very few instances of opponents altogether, and the highest frequency corresponds to a count of 2.

[14]'deputy director of mines', 'horticulture director Om Bir Singh'
[15]'divisional forest officer', 'BMC executive health officer Dr Mangala Gomare', 'erring officers/contractors'

### Characters - Ineffectives

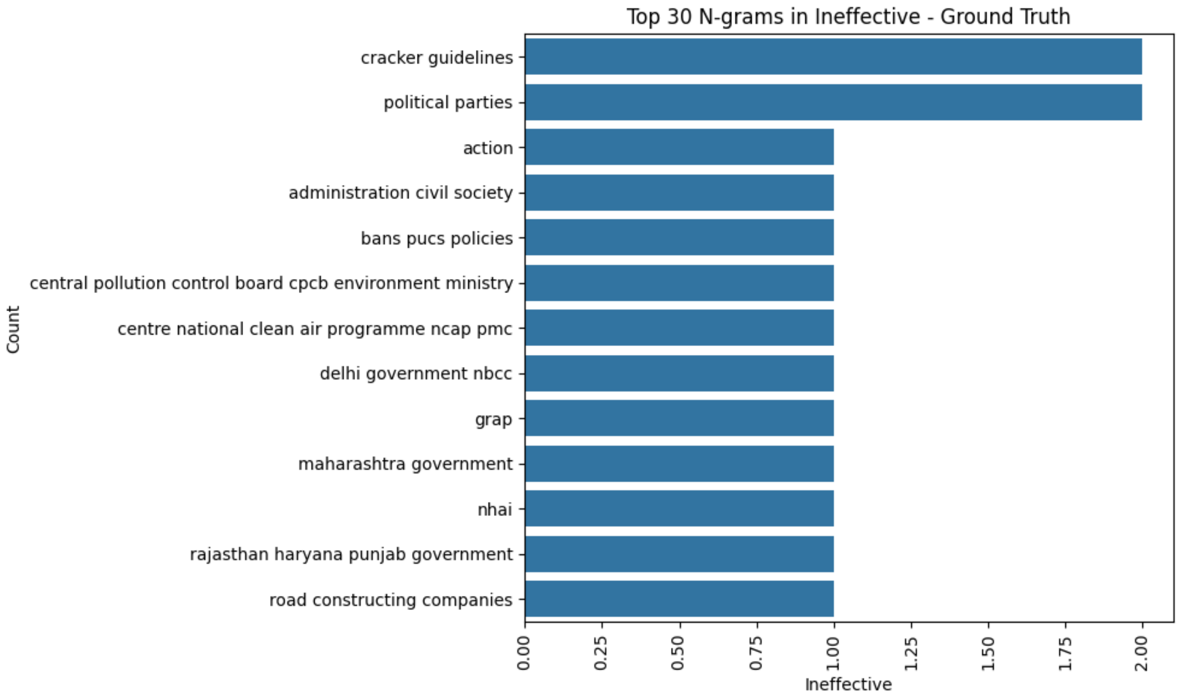



**Figure 4.9:** N-Gram of Ineffectives in Manual Coding

Innefectives are entities that perform an action that has no meaningful effect on the policy problem. The N-grams of ineffective efforts (Figure 4.9) highlight various entities and initiatives deemed ineffective in combating air pollution -

1. 'Cracker guidelines' are the most frequently mentioned, indicating ineffective regulations related to the use of fireworks.
2. 'Political parties' and 'action' are also significant, suggesting that political efforts and actions have not been effective in reducing air pollution. 'Administration civil society' and various 'bans puc policies' (PUC - pollution under control) appear frequently, reflecting the perceived ineffectiveness of administrative and policy measures.
3. Other ineffective efforts include 'central pollution control board cpcb environment ministry', 'centre national clean air programme ncap pmc', 'delhi government nbcc', 'grap' (Graded Response Action Plan), 'maharashtra government', 'nhai' (National Highways Authority of India), 'rajasthan haryana punjab government', and 'road constructing companies'. These entities and initiatives are perceived as failing to effectively address the problem of air pollution.

### Plots

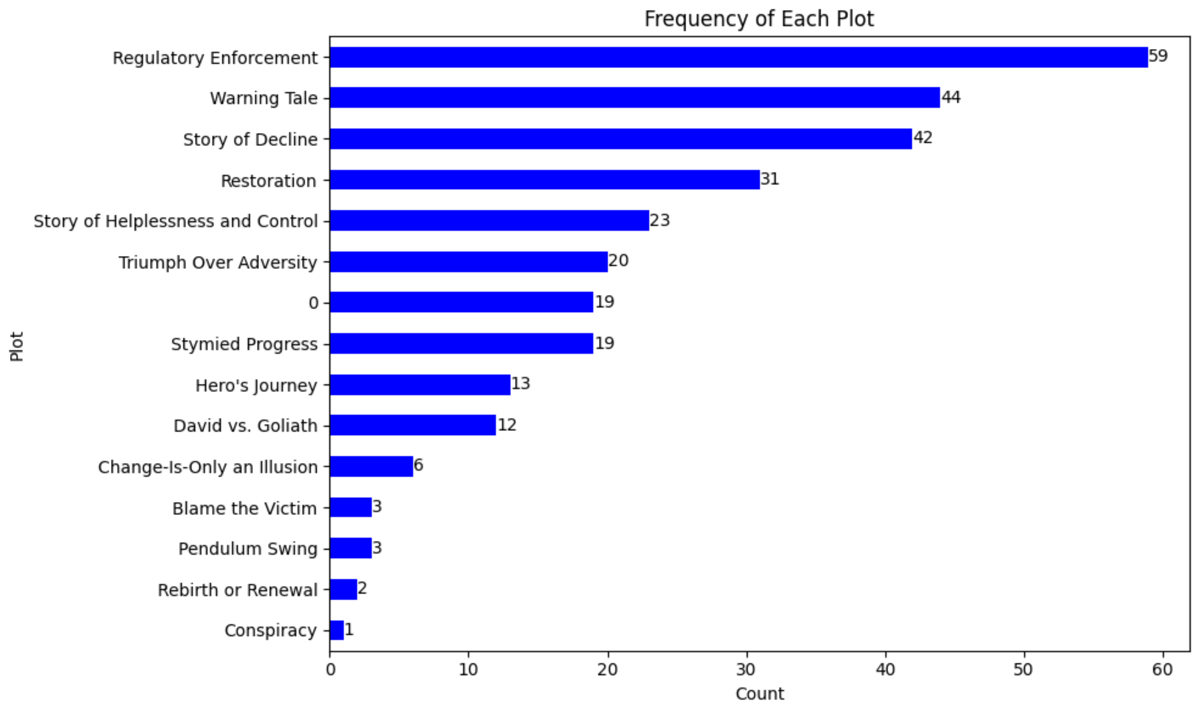



**Figure 4.10:** Plots in the Text

Figure 4.10 reveals the distribution of various narrative plots used in news articles about air pollution in India.

#### Dominant Plots

The most frequent plot is 'Regulatory Enforcement', with a count of 59. This suggests that a significant portion of the news articles focus on the implementation and impact of regulations designed to address air pollution. This plot emphasizes the role of government and regulatory bodies in controlling pollution through policy measures, enforcement actions, and legal frameworks.

An example from the data of what a 'Regulatory Enforcement' plot corresponds to:

> *"New Delhi: After the closure of the polluting Badarpur power plant due to environmental concerns, the Delhi government is looking to cut down on more coal-based electricity to combat air pollution this winter. While most of the blame for Delhi's poor air quality is put on stubble burning in nearby states, a lot of blame has also been placed on coal-based power plants with the Supreme Court-appointed EPCA ensuring the closure of the Badarpur station. "We have tried to cut back on coal-fired power for Delhi as much as possible this season. The city, however, still needs power from the Dadri II and Aravalli stations, which are critical for the islanding scheme," said Joshi."* (Verma, 2018).

This article talks about how authorities of RPCA, Delhi Government, and EPCA introduced and enforced regulation to close down coal power plants to control air pollution. The rest of the article talks about how Delhi will not need power from the main thermal power station that winter due to the another plant's increased capacity, and the city will rely on fewer coal-based plants and other gas-based stations.

The second most frequent plot is 'Warning Tale' (44 counts), followed closely by 'Story of Decline' (42 counts). The 'Warning Tale' plot serves as a cautionary narrative, highlighting the potential consequences of inaction or improper actions related to air pollution. This plot aims to raise awareness and prompt preventive measures. The 'Story of Decline' plot focuses on the deterioration of conditions, underscoring the urgency of addressing the air pollution crisis.

A news article coded as a 'Warning Tale':

> *"MANGALURU: The report noted that waste generation of Class I cities (with population between 0.1 million to five million) in India has been estimated to be around 80 per cent of country's total waste generation. Highlighting the concerns about per capita waste generation rate, the study said that presently it is about 300-400 gm/capita for medium cities and 400-600 gm/capita for large cities. " This is going to increase with the present trend of urbanisation and consumption patterns." On the need for proper waste treatment to generate environmental and monetary benefits, the study said that poorly managed waste has direct implications to urban environment, leading to air, water, and soil pollution, together with long-term health impacts, while it has indirect implications to our economy and growth prospects."* (Shenoy, 2017)

This story reports the effects of air pollution, and its consequences if left unchecked. The article talks about a joint study by ASSOCHAM and PwC highlights that by 2050, India will need 88 sq. km. of land for waste disposal, equivalent to the area of New Delhi Municipal Council, due to untreated waste. This land would be unusable for 50 years. The report urges a revamp of waste management practices, citing rapid urbanization and increasing waste production. Challenges include poor planning, complex systems, limited capacity and funding, and regulatory issues. The study recommends granting industry status to the waste management sector for better regulation and support.

#### Moderate Frequency Plots

Plots such as 'Restoration' (31 counts), 'Story of Helplessness and Control' (23 counts), 'Triumph Over Adversity' (20 counts), and 'Stymied Progress' (19 counts) have moderate frequencies. The 'Restoration' plot depicts efforts to revert the environment or social system to its original state after degradation. This plot offers a hopeful perspective, emphasizing rehabilitation and recovery.

The 'Story of Helplessness and Control' plot presents a narrative where dire situations can be improved through targeted actions or revelations, often portraying a sense of empowerment. 'Triumph Over Adversity' highlights overcoming obstacles through resilience and ingenuity, offering inspirational stories of success against the odds.

The 'Stymied Progress' plot portrays situations where improvement is halted or reversed due to external interference. This narrative can highlight the challenges and resistance faced in combating air pollution.

#### Less Frequent Plots

Several plots appear less frequently, including 'Hero's Journey' (13 counts), 'David vs. Goliath' (12 counts), 'Change-Is-Only an Illusion' (6 counts), 'Blame the Victim' (3 counts), 'Pendulum Swing' (3 counts), 'Rebirth or Renewal' (2 counts), and 'Conspiracy' (1 count). These plots, though less common, offer unique perspectives on the air pollution issue. Their absence from the available narratives may indicate that this type of narrative may not be getting the desired response from the public and might be less effective.

The 'Hero's Journey' plot focuses on a protagonist's transformative journey to resolve a crisis, providing a narrative of personal growth and achievement. 'David vs. Goliath' portrays the struggle of a powerless individual or group against a powerful adversary, highlighting themes of justice and resistance.

'Change-Is-Only an Illusion' suggests that perceived changes are misconceptions, with the real situation remaining stable or worsening. This plot can be used to critique superficial or ineffective measures. 'Blame the Victim' inaccurately holds victims responsible for the problem, potentially highlighting flawed narratives or unjust blame.

The 'Pendulum Swing' plot illustrates the cyclical nature of policy and public sentiment, swinging between extremes. 'Rebirth or Renewal' focuses on transformation and new beginnings after a period of decline or catastrophe, offering a narrative of hope and regeneration. The 'Conspiracy' plot, the least frequent, suggests manipulation of circumstances by a few for their benefit does not occur very often, with the plot itself adding a layer of intrigue and suspicion.

### Implications

The dominance of 'Regulatory Enforcement' and 'Warning Tale' plots indicates a strong focus on governmental actions and the potential consequences of inaction. This reflects the media's role in emphasizing the importance of regulation and raising public awareness about air pollution.

The prevalence of 'Story of Decline' highlights the urgency and severity of the air pollution issue, while 'Restoration' and 'Triumph Over Adversity' offer counter-narratives of hope and resilience. The moderate frequency of 'Story of Helplessness and Control' suggests an emphasis on empowerment and actionable solutions.

Less frequent plots like 'Hero's Journey' and 'David vs. Goliath' provide personalized and dramatic narratives, but their lower frequency indicates they are not the dominant framing in media coverage. Plots like 'Change-Is-Only an Illusion' and 'Blame the Victim' offer critical perspectives, potentially highlighting flaws in public perception or policy.

The 'Pendulum Swing' plot's low frequency suggests that cyclical narratives are less prominent, while 'Rebirth or Renewal' offers a rare but hopeful outlook. The 'Conspiracy' plot, being the least frequent, indicates that narratives of manipulation and hidden agendas are not common in the coverage.

### Moral

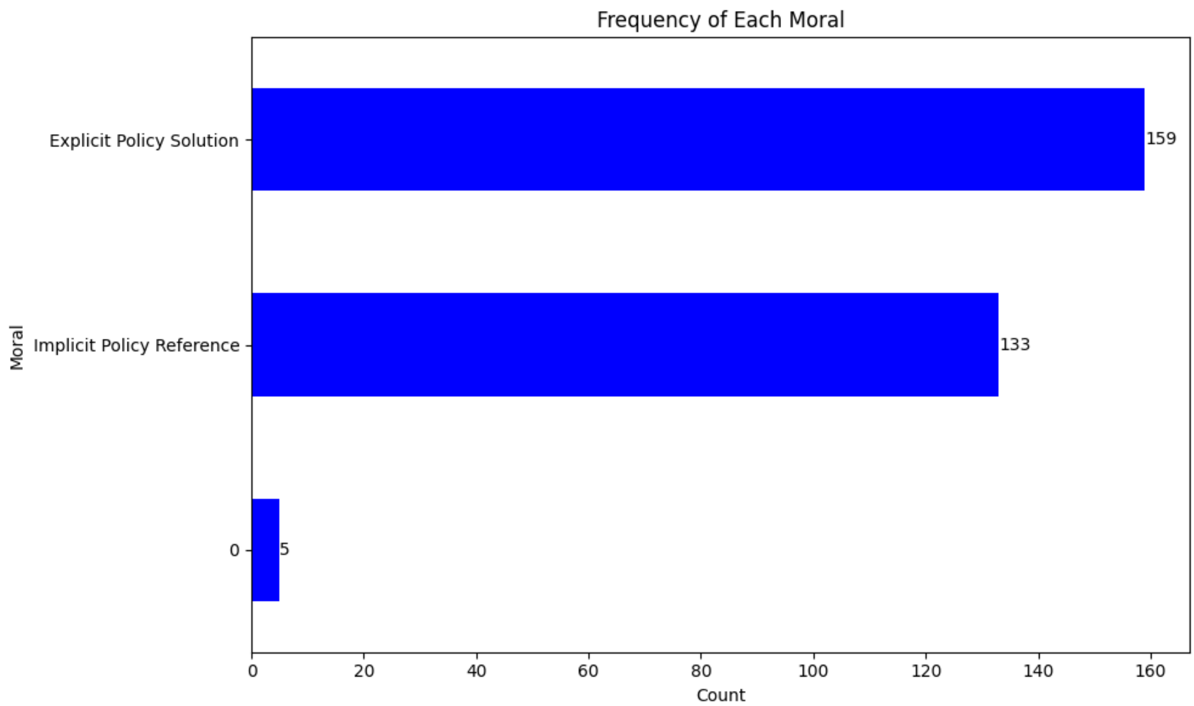



**Figure 4.11:** Moral in the Text

Figure 4.11 shows the distribution of explicit and implicit policy solutions in the narrative of news articles on air pollution in India.

### Morals

The two moral categories are ’Explicit Policy Solution’ and ’Implicit Policy Reference’.

’Explicit Policy Solution’ is the most frequently mentioned moral, with a count of 159. This suggests that a significant portion of the news articles directly state policy solutions, such as specific bans, regulations, or enforcement actions. This explicit approach indicates a clear and direct call to action, urging policymakers and the public to adopt or support specific measures to combat air pollution. An example of an article coded to have an ’Explicit Policy Solution’ is:

> *“NEW DELHI: Stressing that the current air pollution spike could have been averted if GRAP was implemented earlier and more stringently, she said GRAP had an ambitious set of measures under each category of alerts to be issued with action and implementation ranging from penalising the polluting car owners to capping of emissions from power plants across Haryana, Punjab and Uttar Pradesh. ”* (TOI, 2017b)

Here, the Graded Response Action Plan (GRAP) is an explicit policy solution introduced to control air pollution in the Indian National Capital Region. Independent researchers found that Delhi’s air quality dropped to emergency levels due to delayed and insufficient implementation of the graded response action plan (GRAP). Despite a notification in January, GRAP actions only began around October 18. Air pollution researcher Aishwarya Sudhir noted that air quality sharply declined from early October, reaching severe levels before GRAP’s implementation. The lack of coordinated efforts among NCR agencies and failure to address crop fires, which increased pollution by 25%, exacerbated the situation. The government’s delayed response, only intensifying efforts during the Under-17 FIFA World Cup, failed to prevent the crisis or adequately warn the public.

The second most frequent moral is ’Implicit Policy Reference’, with a count of 133. This indicates that many articles prefer to imply policy solutions indirectly, such as highlighting economic benefits, health improvements, or other advantages of addressing air pollution. This approach can be effective in persuading readers by illustrating the positive outcomes of policy actions without explicitly stating the policies. An example of an article coded to have an ’Implicit Policy Reference’ is:

> *“KANPUR: After having bumpy rides on the dug-up roads of the city in 2010, the commuters are praying for smoother roads to ride on in the new year. The ongoing work of laying of sewerlines under the JNNURM programme left the city folks gasping for breath as the digging of roads resulted in dust particles flowing all around in the air. It is now expected that in next six months, the work will be complete and only then will come relief for the city people. The long hours of power cuts forced people to depend on generators. The generators in turn not only created air pollution but also added sound pollution. The pollution in the Ganga river also remained a hot issue throughout the year. It was only after the Allahabad High Court issued special directives to the pollution board asking it to crack down on erring tanneries, that a few leather units were closed down and heavy penalties were imposed on some other units. Locals hope that year 2011 would see the pollution level going down in the holy river. It was in this direction only that professor Vinod Tare from IIT-Kanpur had demonstrated a ’zero toilet discharge system (ZTDS)’ developed by him in his lab. The ZTDS, as the name suggests, is a look alike of most of the conventional toilets with an only difference being that the water consumption in this system is based on ecological sanitation. The ZTDS consists of a separator device below the toilet that divides solids and liquids but the rest of the outside appearance of the ZTDS is completely similar to that of a conventional toilet. The solid waste passes through a different stream and liquid from a different stream. One part of the solid waste is mixed with two parts of cow dung and recycled for atleast 20 times before it is left to decay and form a quality compost. The liquid part is recycled and utilised as flush water. The water is collected, filtered, treated and then recycled to be used for flushing. Microbial agents are used for eliminating odour. This recycling technique removes the need for fresh water in flushing. This system can be installed as public toilets and hence the waste released from these toilets into the Ganga can be avoided. In 2011, not only the city but the entire nation would like to see the holy Ganga flowing clean.”* (TOI, 2010)

In the article, the focus is more on water pollution and a policy solution for that problem, but air pollution is mentioned as well. Focusing on the policy problem, the policy is only implicitly referenced.

### Absence of Morals

The category labeled '0' has a minimal count (5), indicating that very few articles fall into this category. This represents articles that do not present any clear moral or policy solution, instead are focusing on reporting facts or highlighting issues without suggesting specific actions or solutions. An example of an article is:

> *"NEW DELHI: The lieutenant governor on Monday reviewed the progress in providing land to the corporations to deal with solid waste mess as all the landfills Ghazipur, Okhla and Bhalswa are not just polluting and unscientifically built but are now saturated. DDA has identified eight sites. Another plot of about 60 acres has been identified which, according to DDA, can be shared by the corporations. One of the eight landfills is likely to be in Bawana. The LGs high-powered committee on air and water pollution has made recommendations to deal with water scarcity. It has told Delhi Jal Board to list and publicize the sewage treatment plants where treated waste water is available. It added that the water should be supplied to all government agencies like DJB, horticulture departments, DDA, PWD and NDMC. To curb dumping of waste in open drains, the panel has called for boundary walls and laying of a wire mesh along the drains. It also recommends more dhalaos, so that people dont dump waste in drains. The panel, however, doesnt mention anything about waste segregation or recycling. Experts have pointed out that Delhis model favours waste disposal but not recycling."*(TOI, 2014)

The policy problem of air pollution has been mentioned as the presence of a regulatory body on air and water pollution, but the policy problem in focus is not air pollution. This happens mainly in cases where the mention of air pollution is in a context not relevant to the policy problem, but shows up in the data anyway due to the way it was searched.

### Implications

The dominance of 'Explicit Policy Solution' as the primary moral suggests a strong emphasis on clear, actionable solutions in the media coverage of air pollution in India. This direct approach can be particularly effective in mobilizing public opinion and prompting policymakers to take concrete actions. By clearly stating policy solutions, these articles aim to drive immediate and specific responses to the air pollution crisis.

The significant presence of 'Implicit Policy Reference' indicates that while many articles favor direct calls to action, a substantial portion also employs a more subtle approach. By highlighting the benefits of addressing air pollution, these articles can appeal to a broader audience, including those who may be more persuaded by the positive implications of policy actions rather than direct mandates. This strategy can be useful in building consensus and support for long-term policy changes.

The minimal presence of articles without a clear moral or policy solution ('0') suggests that most media coverage of air pollution includes some form of call to action or solution-oriented narrative. This highlights the media's role not only in reporting on the issue but also in advocating for change and influencing public discourse.

### Setting

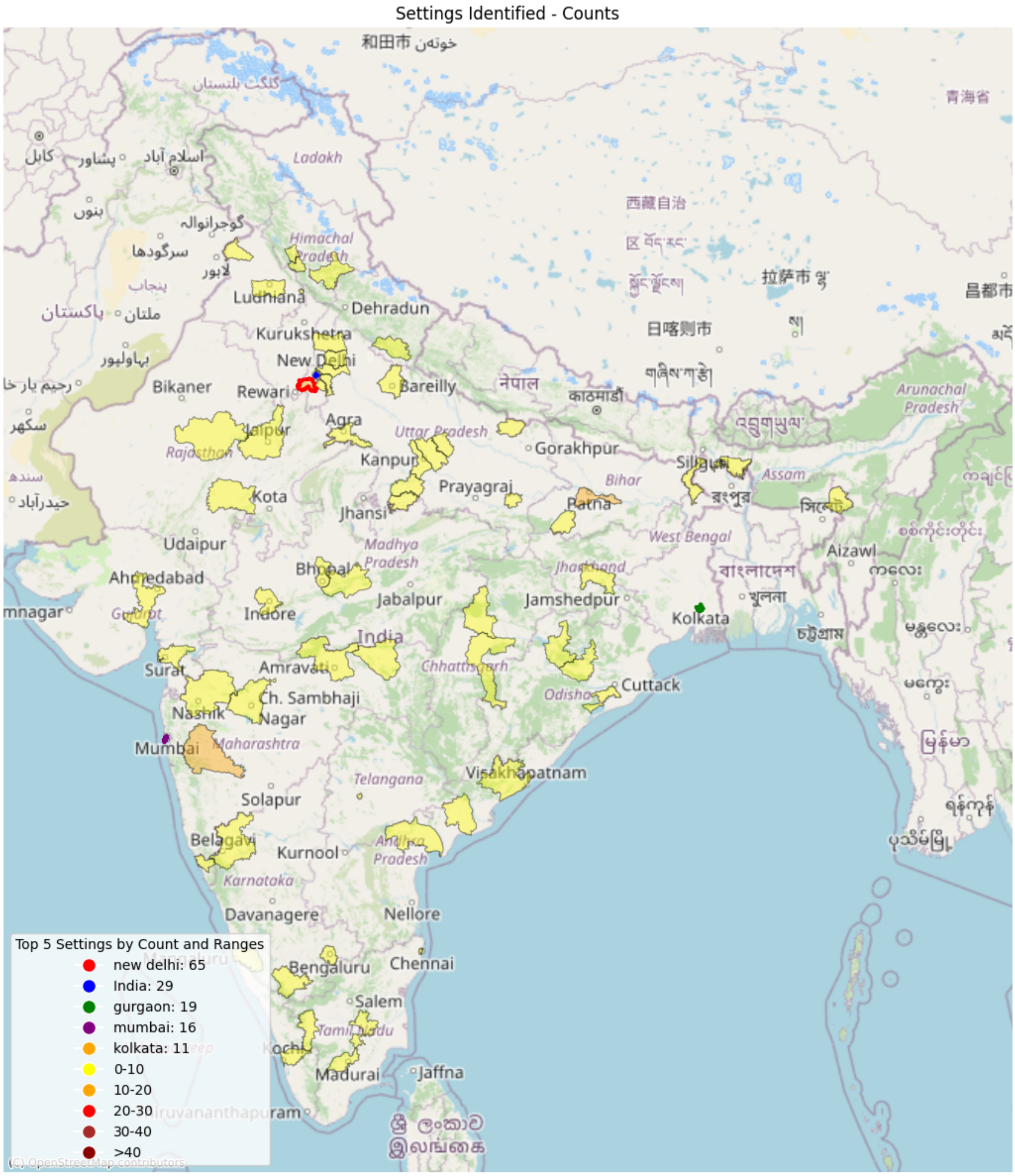



**Figure 4.12:** Settings in the Text

These are the settings identified in the data. Metro cities have the highest counts as settings, with maximum occurrence in Delhi (65), the city with the most polluted air in the world. Other metro cities also top the list, such as Mumbai and Kolkata, showing that the problem is exacerbated by Indian city life and suggesting that the problem in more pronounced in urban areas. There is no concentration of the setting in a particular area of India, which leads to the conclusion that it is a widespread problem, and not only a North Indian phenomenon attributable to stubble burning alone.

### 4.1.2. Narrative Content

#### Belief Systems

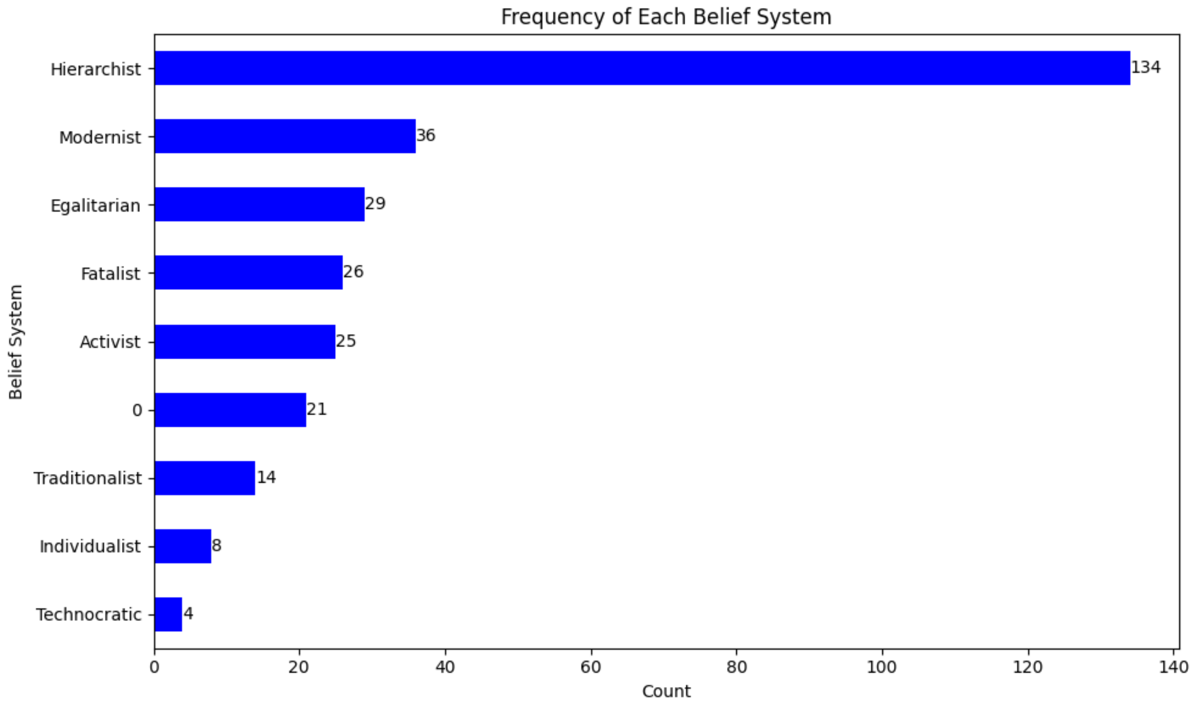



**Figure 4.13:** Belief Systems in the Text

#### Dominant Belief Systems

The most frequently mentioned belief system is 'Hierarchist', with a count of 134. This suggests that a significant portion of the news articles emphasize the need for structured regulations and state-led initiatives to tackle air pollution. Hierarchist narratives focus on stringent enforcement of emission standards, urban zoning laws, and industrial compliance with environmental regulations. This dominance indicates a strong advocacy for top-down regulatory approaches to managing air pollution. Here is an article of a 'Hierarchist' belief system:

> *"NEW DELHI: In the backdrop of severe air pollution, the odd-even scheme in the national capital will kick in from 8 am on Monday (November 4), with only even-numbered non-transport vehicles allowed on Delhi roads on the first day of the exercise. Chief minister Arvind Kejriwal appealed to people on Sunday to follow the rules for the sake of their children and the city. He has also asked the government machinery to ensure that no person faces inconvenience due to the restrictions. Hundreds of teams of Delhi Traffic Police and the transport and revenue departments have been deployed for a strict implementation of the scheme.".* (TOI, 2019a)

The article talks about the Delhi Government implementing the state-led initiative of the odd-even scheme in the National Capital Region to control emissions from vehicles.

#### Moderate Frequency Belief Systems

'Modernist' (36 counts), 'Egalitarian' (29 counts), 'Fatalist' (26 counts), and 'Activist' (25 counts) belief systems have moderate frequencies.

'Modernist' belief systems emphasize the power of scientific progress and technological innovation to solve air pollution problems. Articles within this framework advocate for large-scale technological solutions, such as renewable energy technologies and advanced low-emission public transportation. An article with the 'Modernist' belief system:

> *"New Delhi: Seeking to address the issue of stubble burning in a comprehensive and coordinated manner, the environment ministry has approved launching of a regional project to tackle the menace that adversely affects air quality and soil health. It said, "A slew of technological interventions will be undertaken for timely management of crop residue in addition to effective utilisation of existing machineries. Implementable and sustainable entrepreneurship models will be created in rural areas through upscaling successful initiatives and innovative ideas." The problem of crop residue burning has been intensifying over the years with Punjab, Haryana and Uttar Pradesh being the major stubble burning hotspots. These activities affects air quality in Delhi and national capital region (NCR) every post-harvest season. Increased mechanisation, declining number of livestock and absence of economically viable alternative to use crop residue are some of the reasons why farmers resort to stubble burning. It also asked the ministry to collaborate with the department of science & technology to ensure that independent data on stubble burning is available in real time for timely action."* (TOI, 2017c)

Here, new technology is being proposed to manage the leftover stubble after harvesting crops instead of burning it. The article talks about how the National Adaptation Fund for Climate Change (NAFCC) will phase in a Rs 100 crore project in Punjab, Haryana, Uttar Pradesh, and Rajasthan, focusing initially on awareness and capacity building to promote alternative farming practices and reduce stubble burning, with the aim to mitigate climate impacts, enhance adaptive capacity, and diversify livelihoods. The project's scope may expand based on performance. It complements a central task force's 12-point plan to reduce air pollution in Delhi and NCR, urging coordinated action from agriculture and rural development officials.

'Egalitarian' belief systems stress community involvement and the impacts of air pollution on public health, particularly for vulnerable populations. These narratives often demand radical changes and equitable distribution of clean air.

'Fatalist' belief systems depict efforts to improve air quality as doomed to fail due to overwhelming systemic challenges or corruption. This viewpoint is characterized by skepticism about the efficacy of interventions.

'Activist' belief systems focus on direct action and social change, often in response to perceived government inaction or corporate malfeasance. These articles highlight public demonstrations and grassroots campaigns demanding stricter air quality regulations. An example of the 'Activist' belief system:

> *GURUGRAM: Nearly 200 children from a government school marched for at least four kilometres from Sushant Lok-1 to Ardee City wearing masks and carrying plants on Saturday morning to underline the importance of a greener city. Students of Senior Secondary Government School, Chakkarpur, along with city-based doctors, also started an online petition, requesting Gurgaon authorities to take steps to combat air pollution. Over 1000 signatures were uploaded, urging the government to find a solution."* (TOI, 2017a)

This story is about school children mobilizing a public demonstration to pressure policymakers into taking action against air pollution. They also mention Bharat, a student in a march, who mentioned that his parents recalled Gurgaon as once green and unpolluted, now worse than Delhi, aiming to pressure authorities to combat air pollution. Principal Anjana Dhingra highlighted the voluntary participation of students, stressing the impact of pollution on health and education. Principal medical officer Dr. Pradeep Sharma emphasized collective responsibility in safeguarding the future.

### Less Frequent Belief Systems

Belief systems such as 'Traditionalist' (14 counts), 'Individualist' (8 counts), and 'Technocratic' (4 counts) appear less frequently.

'Traditionalist' belief systems emphasize cultural heritage and historical practices that have a smaller ecological footprint, advocating for the preservation of traditional methods.

'Individualist' belief systems highlight innovation, economic growth, and market mechanisms in solving air pollution, promoting new technologies and market-based solutions like carbon trading.

'Technocratic' belief systems trust in experts and technical solutions, emphasizing data-driven approaches and scientific research to craft precise policy interventions.

The category labeled '0' has a notable count (21), indicating articles that do not clearly align with any specific belief system or perhaps take a neutral stance without advocating for a particular framework, or are possibly not focusing on air pollution as the policy problem.

#### Implications

The dominance of the 'Hierarchist' belief system suggests a strong preference for structured and regulatory approaches to managing air pollution in the media coverage. This aligns with the widespread call for government-led initiatives and enforcement mechanisms to ensure compliance with environmental standards. The emphasis on hierarchist solutions reflects a belief in the effectiveness of top-down approaches in addressing complex environmental issues like air pollution.

The moderate presence of 'Modernist', 'Egalitarian', 'Fatalist', and 'Activist' belief systems indicates a diverse range of narratives beyond regulatory enforcement. 'Modernist' narratives highlight the role of technology and scientific advancements, appealing to those who believe in progress and innovation. 'Egalitarian' narratives bring attention to social justice and equity, advocating for policies that protect vulnerable communities. The presence of 'Fatalist' narratives points to a sense of despair and skepticism about the effectiveness of interventions, which can be a significant barrier to mobilizing public and political will. 'Activist' narratives emphasize the power of collective action and public pressure, showcasing the role of social movements in driving policy change.

The less frequent 'Traditionalist', 'Individualist', and 'Technocratic' belief systems highlight alternative viewpoints that focus on cultural heritage, market-based solutions, and expert-led approaches, respectively.

### Narrative Strategies

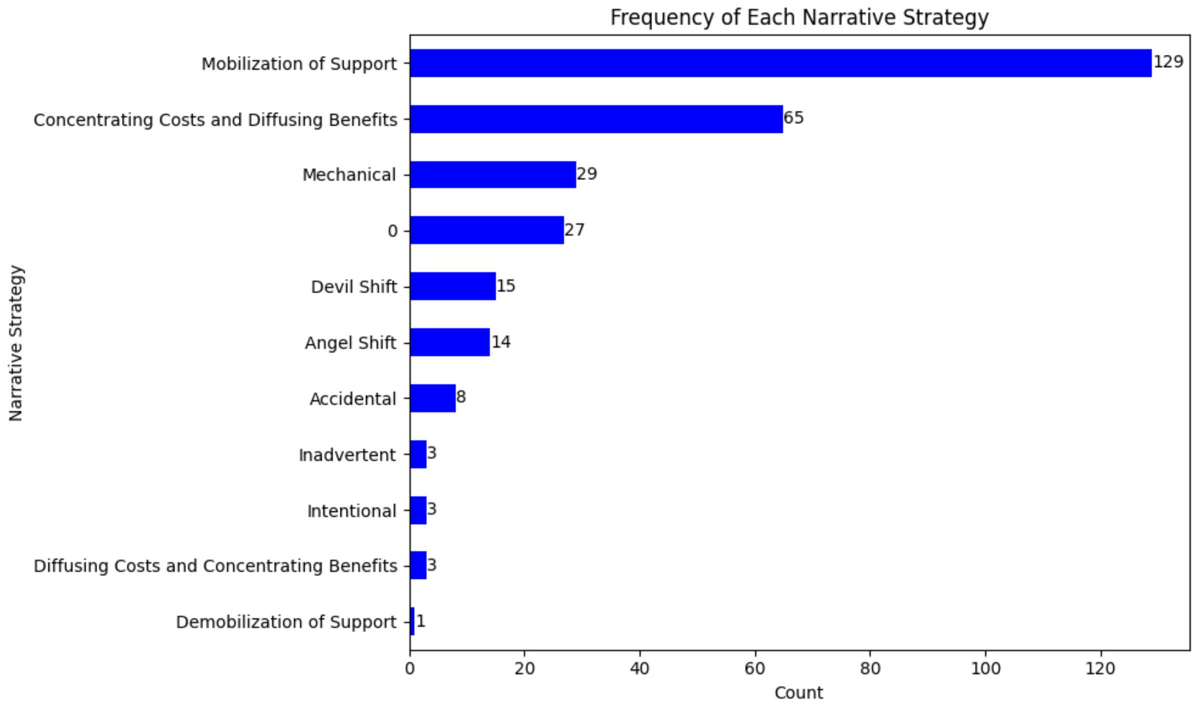



**Figure 4.14:** Narrative Strategies in the Text

### Dominant Narrative Strategies

The most frequently mentioned narrative strategy is 'Mobilization of Support', with a count of 129. This suggests that a significant portion of the news articles aim to rally support for particular policy positions. This strategy involves actively encouraging public engagement and support for clean air policies by highlighting the impacts of air pollution and calling for action. For example, articles might include stories of affected individuals, particularly children, to evoke emotional responses and mobilize citizens to demand policy changes. An example of an article coded with the 'Mobilization of Support' strategy is:

> *"HYDERABAD: Residents of Shastripuram, keen on securing a pollution-free neighbourhood, registered their protest against plastic units operating in their area with the local MLA and government officials who came to survey the residential colony on Saturday. In these columns TOI had earlier highlighted the plight of the residents of Shastripuram, which is Hyderabad's biggest residential colony. The residents have been up in arms against plastic units located in their neighborhood which have been steadily turning the streets into plastic dump-yards, and polluting air and water."* (TOI, 2012)

Here, the residents are mobilizing support for their cause to fight against plastic units, aiming to stir public emotions and encourage others to demand action from their representatives. They further mention that despite promises, only superficial cleaning was done. GHMC promised to request power cuts to these units by Monday. Residents have also approached the State Human Rights Commission (SHRC) for relief and may consider legal action if unresolved.

The second most frequent narrative strategy is 'Concentrating Costs and Diffusing Benefits', with a count of 65. This strategy involves depicting policies where the costs are concentrated on a specific group (e.g., polluting industries) while the benefits are diffused across a larger audience (e.g., the general public). This narrative often emphasizes the broader social and health benefits of reducing air pollution, despite the concentrated economic costs on specific sectors. An example of this strategy in the data:

> *"Surat: The standing committee of Surat Municipal Corporation (SMC) has approved a proposal to procure loan of Rs1,925 crore from World Bank. The money will be used on pollution-free development projects and also to conduct a study along with Gujarat Engineering Research Institute (GERI) and Gujarat Pollution Control Board (GPCB) on pollution in the city."* (TOI, 2019b)

Civic bodies are procuring funds by concentrating costs to spend on mitigating pollution, to diffuse benefits to the public. Projects include roads, bridges, gardens, and waste disposal. A proposal for air quality monitoring was approved, and SMC will collaborate with GPCB and GERI to measure pollution and recommend solutions. Additionally, 15 new air quality monitoring machines will be installed after the Lok Sabha elections, costing around Rs 50 lakh each.

#### Moderate Frequency Narrative Strategies

'Mechanical' (29 counts) and '0' (27 counts) are narrative strategies with moderate frequencies.

'Mechanical Cause' describes narratives where unintended consequences arise from unguided actions. In the context of air pollution, this might involve outdated industrial practices leading to high emissions due to bureaucratic inertia and failure to update regulations.

The category '0' indicates narratives that do not clearly align with a specific strategy or may take a neutral stance, only reporting facts and figures.

#### Less Frequent Narrative Strategies

Narrative strategies such as 'Devil-Shift' (15 counts), 'Angel-Shift' (14 counts), 'Accidental' (8 counts), 'Inadvertent' (3 counts), 'Intentional' (3 counts), 'Diffusing Costs and Concentrating Benefits' (3 counts), and 'Demobilization of Support' (1 count) appear less frequently.

'Devil-Shift' involves casting villains as victors over heroes. This might depict polluting companies as having defeated public interest groups by manipulating regulatory processes to prioritize profits over public health.

'Angel-Shift' involves casting heroes as winners, highlighting successful efforts by activists or communities to overcome challenges and achieve cleaner air.

'Accidental Cause' associates unintended consequences by unguided actions, suggesting air pollution results from unforeseen byproducts of urban development and increased vehicle use.

'Inadvertent Cause' involves unintended consequences by purposeful actions, where policies aiming to stimulate economic growth inadvertently lead to increased pollution.

'Intentional Cause' associates intended consequences by purposeful actions, depicting policymakers as deliberately protecting industrial interests at the expense of public health.

'Diffusing Costs and Concentrating Benefits' involves distributing policy costs broadly while concentrating benefits on a specific group, typically the elite few.

'Demobilization of Support' aims to diminish support for opposing views, often by spreading doubt or highlighting potential drawbacks of pollution control measures.

#### Implications

The dominance of the 'Mobilization of Support' strategy indicates a strong focus on rallying public and political support for air pollution control measures. This approach emphasizes the importance of public engagement and advocacy in driving policy changes. By mobilizing support, these narratives aim to create a sense of urgency and collective action to address the air pollution crisis.

The significant presence of 'Concentrating Costs and Diffusing Benefits' suggests that many articles frame air pollution control measures as beneficial to the broader public despite the concentrated costs on specific industries. This narrative strategy emphasizes the greater good and societal benefits of clean air policies, making a compelling case for their implementation.

The moderate frequency of 'Mechanical Cause' indicates recognition of systemic issues and bureaucratic inertia as significant contributors to air pollution. This narrative highlights the need for updating regulations and addressing institutional failures to effectively manage air quality.

The less frequent narrative strategies provide additional perspectives, such as highlighting the unintended consequences of policies, the role of deliberate actions by policymakers, and the importance of equitable distribution of policy costs and benefits.

### 4.1.3. Co-Occurrences of Codes

Plots and Morals

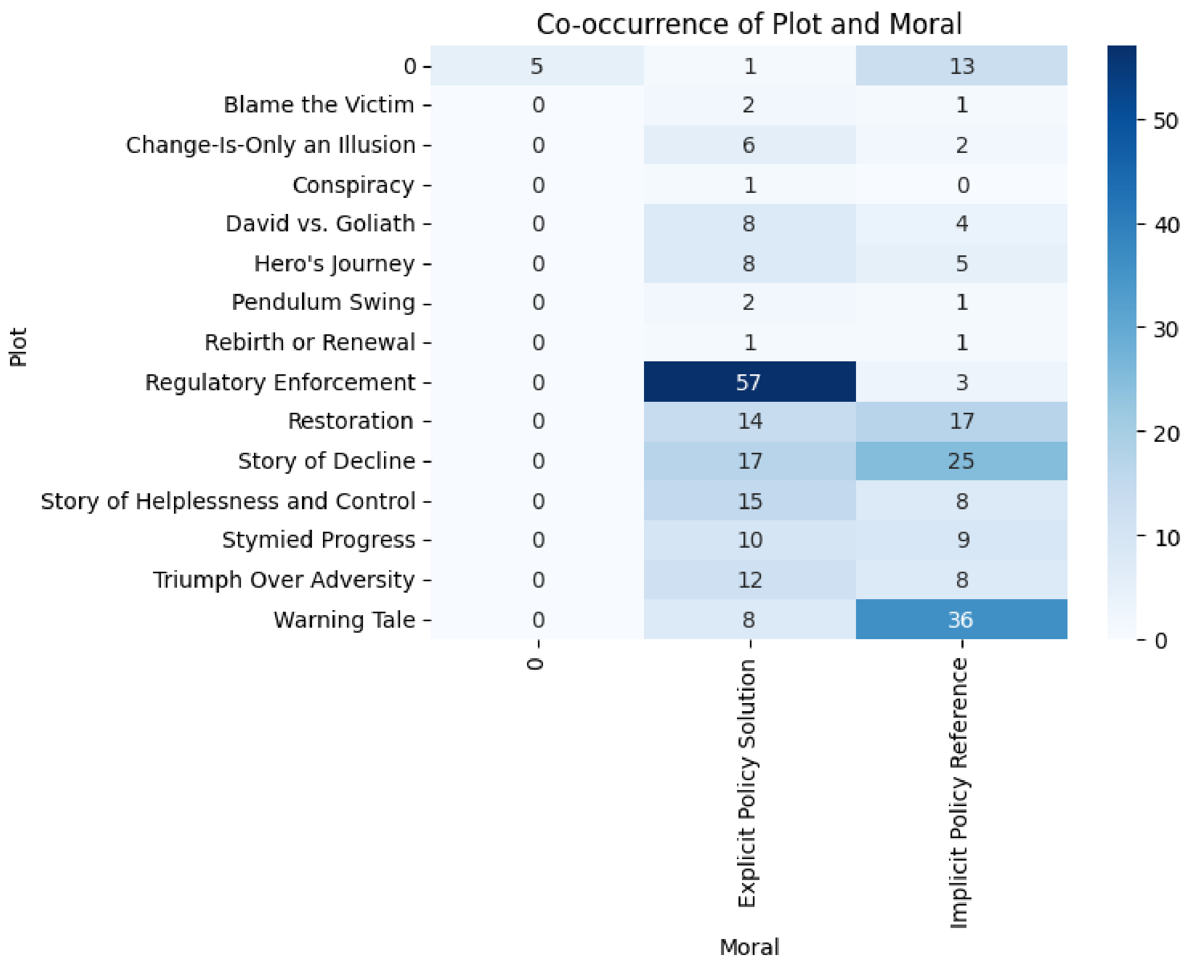



**Figure 4.15:** Co-Occurrences of Plots and Morals

Key Observations

1. **Regulatory Enforcement:**

   *Explicit Policy Solution:* This combination has the highest count (57), indicating that narratives involving regulatory enforcement heavily focus on explicit policy solutions. This suggests a strong call for government intervention and regulatory actions to combat air pollution.

2. **Story of Decline:**

   *Implicit Policy Reference:* This plot has a significant count (25) for implicit policy references. Narratives here emphasize the deterioration of conditions and suggest indirect solutions or benefits of addressing the decline. This indicates a narrative focused on the negative trajectory of air quality, pushing for a recognition of the problem before suggesting explicit actions.

3. **Restoration:**

   *Both Explicit and Implicit:* Restoration narratives are balanced between explicit (14) and implicit (17) policy solutions. This shows a dual approach where articles not only call for specific actions to restore previous conditions but also highlight the broader benefits of such actions.

4. **Story of Helplessness and Control:** *Both Explicit and Implicit:* Similar to restoration, this plot shows a significant count in both categories (15 explicit and 8 implicit). This plot presents a dire situation that can be amended, suggesting a mix of direct and indirect policy recommendations.

5. **Warning Tale:** *Implicit Policy Reference:* The second-highest implicit policy reference count (36) falls under this plot. Warning tales often highlight future risks of inaction, suggesting benefits of addressing air pollution without directly stating policies.

6. **Triumph Over Adversity, Stymied Progress, David vs. Goliath, Heros Journey:** *Balanced Approach:* These plots show moderate counts across both explicit and implicit solutions, indicating a balanced narrative approach. They suggest that overcoming obstacles (adversity) or struggles against powerful forces (David vs. Goliath) involve both direct actions and indirect benefits.

7. **Change-Is-Only an Illusion, Conspiracy, Pendulum Swing, Rebirth or Renewal, Blame the Victim:** *Lower Counts:* These plots have fewer occurrences and are less associated with explicit policy solutions. They might focus on explaining the status quo or shifting blame rather than proposing concrete solutions.

8. **0 Category:** *Explicit Policy Solution:* The significant count (5) suggests some articles might include methodological artifacts or unclear plot identification but still call for explicit policy solutions.

#### Insights

1. **Dominance of Regulatory Narratives:** The highest co-occurrence of explicit policy solutions with regulatory enforcement indicates a strong narrative push for regulatory measures. This underscores the media's role in advocating for government-led initiatives and enforcement as key solutions to air pollution.

2. **Importance of Decline Narratives:** The significant presence of implicit policy references in *'Story of Decline'* suggests that narratives focusing on worsening conditions are effective in highlighting the need for action without always specifying what that action should be. This can create a sense of urgency and awareness among the public and policymakers.

3. **Balanced Narrative Approaches:** Plots like *'Restoration'* and *'Story of Helplessness and Control'* showing balanced use of explicit and implicit solutions suggest these narratives are versatile. They can appeal to both policymakers (through explicit solutions) and the general public (through implicit benefits).

4. **Focus on Future Consequences:** The prominence of implicit policy references in *'Warning Tale'* indicates a narrative strategy focused on cautioning about future consequences. This approach

can be effective in mobilizing support by emphasizing the long-term benefits of addressing air pollution.

5. **Variety in Narrative Strategies:** The presence of various plots with moderate counts in both policy solution categories suggests that the media employs a diverse range of narrative strategies to discuss air pollution. This diversity helps cater to different audience segments and policy perspectives.

The dominance of regulatory enforcement with explicit solutions shows a clear advocacy for governmental action. Simultaneously, narratives of decline and warnings about future consequences underscore the urgency of the issue, leveraging both direct and indirect policy suggestions to mobilize support and action.

Plots and Belief Systems

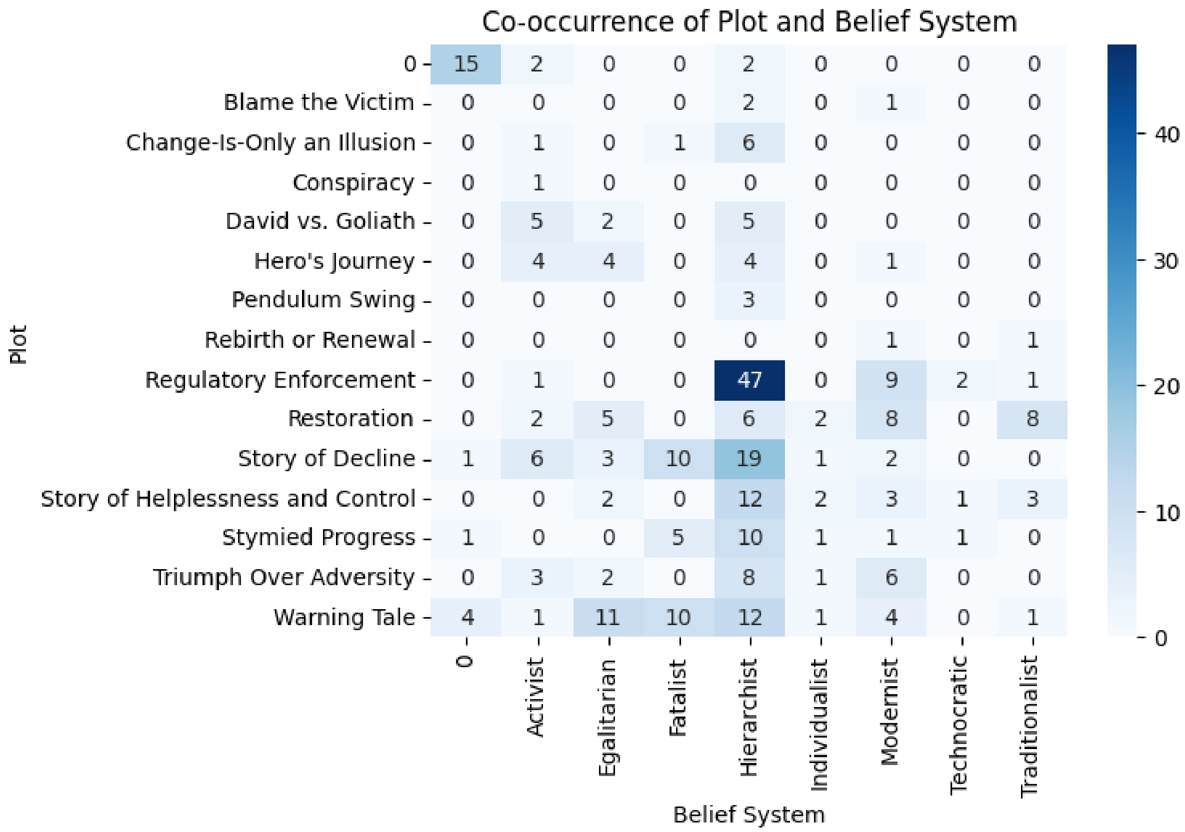



**Figure 4.16:** Co-Occurrences of Plots and Belief Systems

**Key Observations**

1. **Regulatory Enforcement**:

   - **Hierarchist**: The highest count (47) is observed in this combination, indicating a strong narrative emphasis on structured regulations and state-led initiatives to tackle air pollution. This belief system aligns with advocating for stringent enforcement of emission standards and compliance with environmental regulations.
   - **Modernist and Technocratic**: Lower count (9 and 2) suggests a minor but present narrative focus on technology-based or expert-led and data-driven policy solutions within regulatory enforcement stories.

2. **Story of Decline**:
    - **Hierarchist and Fatalist**: High counts (19 and 10, respectively) indicate narratives highlighting the deterioration of conditions often focus on structured regulatory responses and skepticism about the efficacy of interventions. This suggests a dual narrative where structured actions are deemed necessary despite perceived systemic challenges.
    - **Egalitarian**: Moderate count (6) shows a narrative focus on community involvement and public health impacts, advocating for equitable air quality improvements.
3. **Warning Tale**:
    - **Egalitarian and Hierarchist**: Significant counts (11 and 12, respectively) indicate cautionary narratives often stress the need for equitable policies and structured regulations to prevent future risks.
    - **Fatalist**: Moderate count (10) suggests some narratives convey skepticism about the effectiveness of interventions despite warnings.
4. **Restoration**:
    - **Individualist, Modernist, Technocratic, Traditionalist**: Balanced counts across these belief systems (6-8) indicate a diverse narrative approach. Stories of restoration often highlight technological innovations, scientific progress, traditional practices, and market mechanisms as solutions.
    - **Hierarchist**: Moderate count (6) aligns with narratives advocating for structured regulatory measures to restore conditions.
5. **Story of Helplessness and Control**:
    - **Hierarchist and Egalitarian**: High counts (12 each) suggest narratives focus on the potential for structured regulatory actions and community involvement to improve dire situations.
    - **Technocratic**: Moderate count (3) highlights the role of expert-driven solutions.
6. **Stymied Progress**:
    - **Hierarchist and Fatalist**: High counts (10 and 5, respectively) indicate narratives highlight structured regulatory challenges and skepticism about overcoming systemic barriers.
    - **Technocratic**: Lower count (1) suggests minimal focus on expert-led solutions in these narratives.
7. **Triumph Over Adversity**:
    - **Hierarchist and Egalitarian**: Significant counts (8 and 6, respectively) show narratives of overcoming obstacles often advocate for structured regulations and community-driven efforts.
    - **Modernist and Technocratic**: Moderate counts (6 each) highlight the role of technological and expert-driven solutions.
8. **Plots with Minimal Representation**:
    - **Conspiracy, Blame the Victim, Change-Is-Only an Illusion, Rebirth or Renewal, Pendulum Swing**: These plots have minimal co-occurrence with most belief systems, indicating less frequent narrative focus on these storylines.

Insights

1. **Dominance of Hierarchist Belief System**:
    - The 'Hierarchist' belief system shows the highest co-occurrence across multiple plots, particularly in 'Regulatory Enforcement', 'Story of Decline', and 'Warning Tale'. This indicates a strong narrative focus on structured regulations and government-led initiatives to combat air pollution.
2. **Skepticism and Challenges**:
    - The 'Fatalist' belief system appears frequently in 'Story of Decline', 'Warning Tale', and 'Stymied Progress'. This suggests narratives often highlight systemic challenges and skepticism about the effectiveness of interventions, emphasizing the need for robust regulatory actions despite these barriers.
3. **Community and Equity Focus**:
    - The 'Egalitarian' belief system shows significant co-occurrence in 'Warning Tale', 'Story of Decline', and 'Story of Helplessness and Control'. This reflects narratives advocating for equitable policies, community involvement, and protection of vulnerable populations from air pollution.
4. **Technological and Expert-Driven Solutions**:
    - The 'Modernist' and 'Technocratic' belief systems appear in 'Restoration', 'Triumph Over Adversity', and 'Regulatory Enforcement'. This indicates narratives often highlight the role of scientific progress, technological innovation, and expert-led solutions in addressing air pollution.
5. **Diverse Narrative Approaches**:
    - Plots like 'Restoration' and 'Story of Helplessness and Control' show balanced co-occurrence across multiple belief systems, suggesting diverse narrative strategies that incorporate regulatory, technological, community-driven, and traditional approaches.
6. **Limited Narrative Focus**:
    - Plots such as 'Conspiracy', 'Blame the Victim', and 'Pendulum Swing' have minimal representation across belief systems, indicating these storylines are less commonly employed in narratives about air pollution.

The dominance of the 'Hierarchist' belief system across multiple plots shows the strong advocacy for structured regulatory measures and state-led initiatives. Simultaneously, the presence of 'Fatalist', 'Egalitarian', and 'Technocratic' belief systems highlights the diverse narrative strategies emphasizing systemic challenges, equity, and expert-driven solutions.

Narrative Strategies and Belief Systems

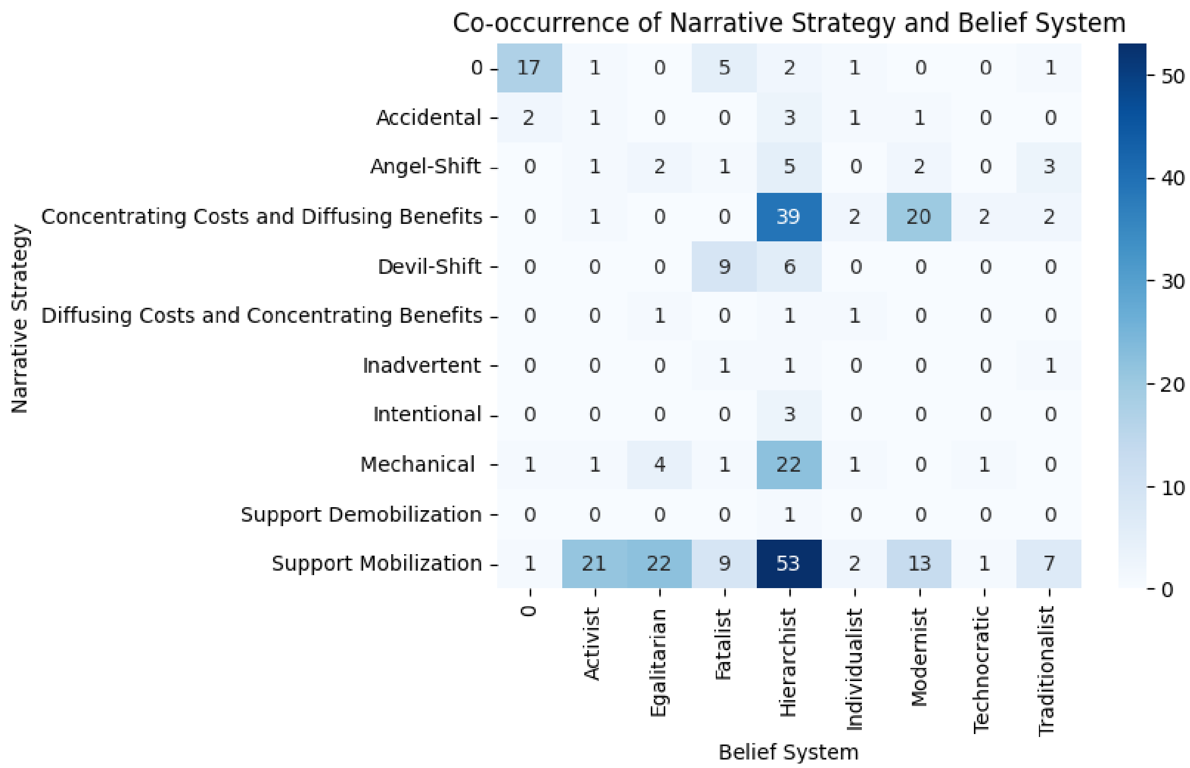



**Figure 4.17:** Co-Occurrences of Narrative Strategies and Belief Systems

Key Observations

1. **Support Mobilization**:
   - **Hierarchist**: The highest count (53) is observed in this combination, indicating a dominant narrative strategy of rallying support for structured regulatory actions to combat air pollution. This belief system aligns with advocating for stringent enforcement of emission standards and compliance with environmental regulations.
   - **Egalitarian**: Significant count (22) suggests that mobilizing support often involves advocating for community involvement and equitable policies to protect vulnerable populations.
   - **Activist**: Moderate count (21) reflects the narrative strategy of mobilizing public demonstrations or campaigns to pressure policymakers into action.
   - **Technocratic and Modernist**: Lower counts (13 and 7, respectively) indicate narratives focusing on expert-driven solutions and technological innovation.
2. **Mechanical**:
   - **Hierarchist**: High count (22) indicates narratives highlighting systemic or bureaucratic issues within structured regulatory frameworks.
   - **Fatalist**: Lower count (4) suggests some narratives emphasizing skepticism about overcoming bureaucratic inertia.
   - **Technocratic**: Moderate count (10) highlights expert-driven solutions to systemic issues.
3. **Devil-Shift**:

- **Hierarchist and Fatalist**: High counts (9 and 6, respectively) show narratives portraying regulatory bodies or systemic challenges as villains prioritizing other interests over public health.
- **Egalitarian**: Moderate count (3) indicates narratives framing certain groups or institutions as opposing equitable policies.

4. **Accidental**:
   - **Egalitarian and Hierarchist**: Moderate counts (2 and 3, respectively) suggest narratives highlighting unintended consequences of actions impacting public health and regulatory frameworks.
   - **Fatalist**: Lower count (1) reflects skepticism about the efficacy of addressing these accidental causes.
5. **Concentrating Costs and Diffusing Benefits**:
   - **Hierarchist and Fatalist**: High counts (39 and 20, respectively) indicate narratives focusing on how structured regulatory actions may impose costs on certain groups while diffusing benefits more broadly. This strategy often highlights systemic challenges in implementing equitable policies.
6. **Angel-Shift**:
   - **Hierarchist and Fatalist**: Moderate counts (5 and 2, respectively) show narratives portraying regulatory bodies or systemic challenges as heroes who successfully implement beneficial policies.
   - **Egalitarian**: Lower count (2) reflects narratives emphasizing the role of community-driven efforts in achieving positive outcomes.
7. **Intentional**:
   - **Hierarchist**: Moderate count (3) indicates narratives attributing purposeful actions by regulatory bodies to address air pollution effectively.
   - **Technocratic**: Lower count (1) highlights expert-driven intentional actions.
8. **Diffusing Costs and Concentrating Benefits**:
   - **Hierarchist and Egalitarian**: Moderate counts (1 each) show narratives emphasizing policies that distribute costs broadly while concentrating benefits on specific groups, often aligning with community-driven efforts.
   - **Technocratic**: Lower count (1) suggests expert-driven policies with similar distribution effects.
9. **Support Demobilization**:
   - **Hierarchist and Egalitarian**: Moderate counts (1 each) indicate narratives aiming to diminish support for opposing views within structured regulatory frameworks and community-driven efforts.
   - **Fatalist**: Lower count (1) reflects skepticism about the effectiveness of these demobilization efforts.

Insights

1. **Dominance of Hierarchist Belief System**:

- The ’Hierarchist’ belief system shows the highest co-occurrence across multiple narrative strategies, particularly in ’Support Mobilization’, ’Mechanical’, and ’Concentrating Costs and Diffusing Benefits’. This indicates a strong narrative focus on structured regulatory measures and state-led initiatives as primary solutions to air pollution.

2. **Community and Equity Focus**:

   - The ’Egalitarian’ belief system appears frequently in ’Support Mobilization’, ’Accidental’, and ’Devil-Shift’. This reflects narratives advocating for equitable policies, community involvement, and highlighting the unintended consequences of actions impacting public health.

3. **Skepticism and Challenges**:

   - The ’Fatalist’ belief system appears in ’Mechanical’, ’Devil-Shift’, and ’Concentrating Costs and Diffusing Benefits’. This suggests narratives often highlight systemic challenges, skepticism about overcoming bureaucratic inertia, and the uneven distribution of costs and benefits.

4. **Technological and Expert-Driven Solutions**:

   - The ’Technocratic’ and ’Modernist’ belief systems show significant co-occurrence in ’Support Mobilization’ and ’Mechanical’. This indicates narratives frequently highlight the role of scientific progress, technological innovation, and expert-driven solutions in addressing air pollution.

5. **Diverse Narrative Strategies**:

   - The heatmap reveals diverse narrative strategies across different belief systems, with a focus on mobilizing support, highlighting systemic challenges, and emphasizing the role of both community-driven and expert-led efforts.

6. **Limited Narrative Focus**:

   - Strategies such as ’Intentional’, ’Diffusing Costs and Concentrating Benefits’, and ’Support Demobilization’ have minimal representation across most belief systems, indicating these strategies are less commonly employed in narratives about air pollution.

The dominance of the ’Hierarchist’ belief system across multiple strategies underscores the strong advocacy for structured regulatory measures and state-led initiatives. Simultaneously, the presence of ’Egalitarian’, ’Fatalist’, and ’Technocratic’ belief systems highlights the diverse narrative strategies emphasizing community involvement, systemic challenges, and expert-driven solutions.

### Plots and Narrative Strategies

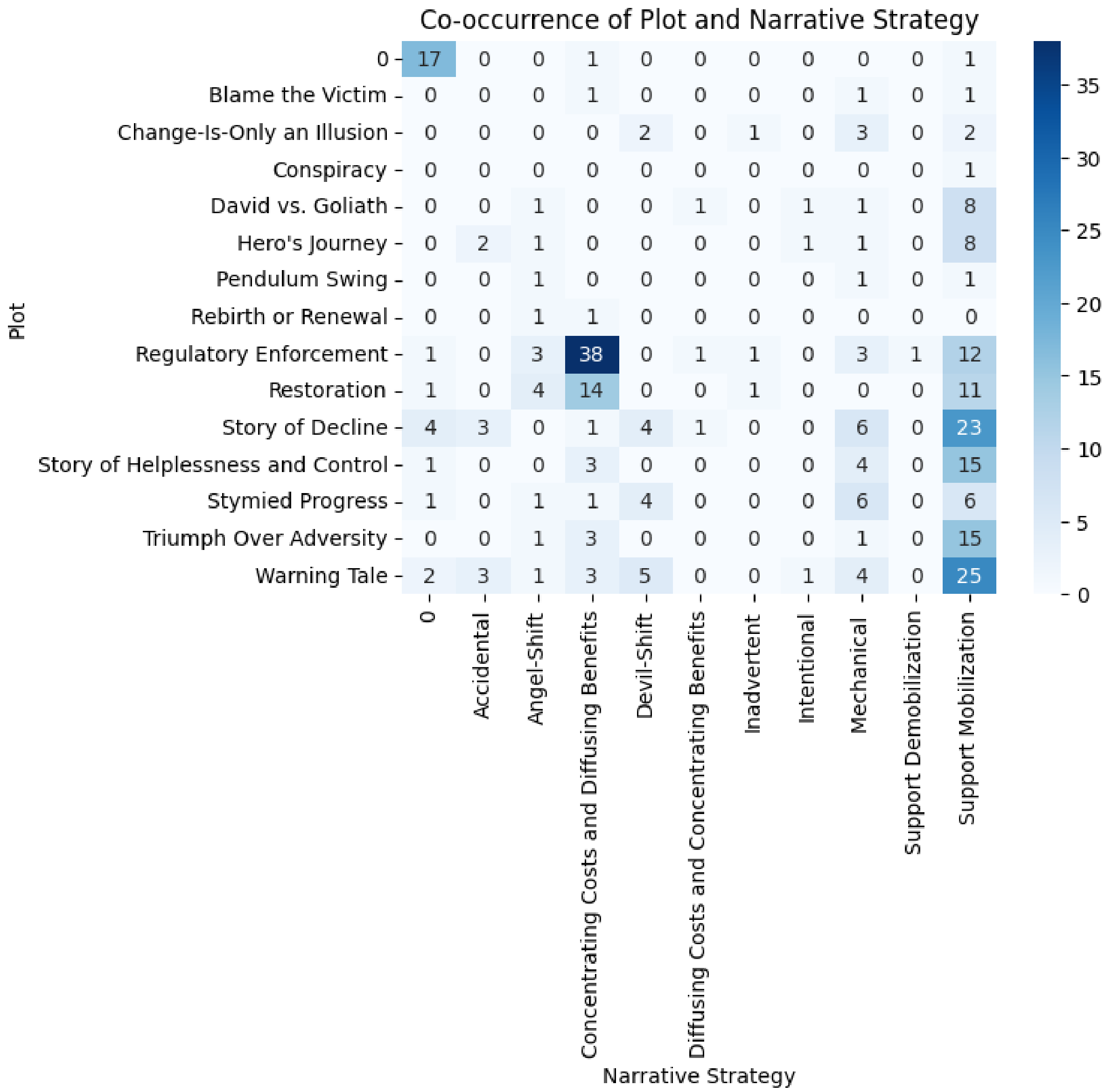



**Figure 4.18:** Co-Occurrences of Plots and Narrative Strategies

### Key Observations

1. **Regulatory Enforcement**
   - **Support Mobilization:** This combination has the highest count (38), indicating a strong narrative focus on rallying public and political support for regulatory measures to combat air pollution.
   - **Angel-Shift and Mechanical:** Lower but significant occurrences (3 each) show narratives casting regulatory bodies as heroes and highlighting systemic bureaucratic issues.
2. **Story of Decline**
   - **Support Mobilization:** A high count (23) suggests that narratives about the deteriorating condition of air quality are used to mobilize support for action.

- **Devil-Shift and Accidental:** Lower counts (4 each) indicate narratives portraying villains as victorious and highlighting unforeseen consequences of actions.

3. **Warning Tale**
   - **Support Mobilization:** High occurrence (25) emphasizes using cautionary stories to mobilize support by warning about future risks of inaction.
   - **Angel-Shift, Devil-Shift, Mechanical:** Moderate counts (3-4) suggest diverse narrative strategies within warning tales, from casting heroes and villains to highlighting systemic issues.
4. **Restoration**
   - **Support Mobilization:** Significant count (11) shows narratives focusing on the benefits of restoring conditions to their original state to rally support.
   - **Angel-Shift:** Moderate occurrence (4) indicates narratives portraying successful restoration efforts as heroic.
5. **Story of Helplessness and Control**
   - **Support Mobilization:** Moderate count (15) shows narratives emphasizing the potential for control and action to improve dire situations.
   - **Mechanical:** Some occurrences (4) highlight systemic issues needing rectification.
6. **Triumph Over Adversity**
   - **Support Mobilization:** High count (15) underscores narratives of overcoming obstacles to inspire and mobilize support.
   - **Angel-Shift:** Moderate occurrence (3) emphasizes narratives portraying individuals or groups who triumph as heroes.
7. **Stymied Progress**
   - **Mechanical:** Moderate count (6) suggests narratives focusing on how systemic or bureaucratic issues halt progress.
   - **Support Mobilization:** Significant occurrence (6) indicates a narrative push to overcome these obstacles through mobilization.
8. **Plots with Minimal Representation**
   - **Conspiracy, Blame the Victim, Change-Is-Only an Illusion, David vs. Goliath, Heros Journey, Rebirth or Renewal, Pendulum Swing:** These plots have minimal co-occurrence with most narrative strategies, indicating these narratives are less frequently linked to strategic mobilization or systemic issues.

Insights

1. **Dominant Use of Support Mobilization**
   - Across multiple plots, *'Support Mobilization'* is the most common narrative strategy, particularly in *'Regulatory Enforcement'*, *'Story of Decline'*, *'Warning Tale'*, *'Restoration'*, and *'Triumph Over Adversity'*. This indicates a strong focus on rallying support and advocating for action in the fight against air pollution.
2. **Systemic and Bureaucratic Issues**

- Plots like *'Stymied Progress'*, *'Story of Helplessness and Control'*, and *'Regulatory Enforcement'* frequently employ the *'Mechanical'* narrative strategy. This suggests that narratives often highlight systemic failures or bureaucratic inertia as significant obstacles to addressing air pollution.

3. **Casting Heroes and Villains**
   - The *'Angel-Shift'* and *'Devil-Shift'* strategies, though less frequent, are present in narratives involving *'Regulatory Enforcement'*, *'Warning Tale'*, and *'Story of Decline'*. This indicates that these stories sometimes frame regulatory bodies, activists, or industrial players as heroes or villains, adding a moral dimension to the narrative.
4. **Accidental and Inadvertent Causes**
   - The presence of *'Accidental'* and *'Inadvertent'* strategies in plots like *'Story of Decline'* and *'Warning Tale'* suggests that some narratives focus on unintended consequences of actions, portraying air pollution as a byproduct of other developments (e.g., urban sprawl, economic growth).
5. **Limited Narrative Diversity in Some Plots**
   - Plots such as *'Conspiracy'*, *'Blame the Victim'*, and *'Pendulum Swing'* have minimal representation across most strategies. This could indicate these narratives are either less common or not as effectively linked to broader strategic narratives in the context of air pollution.

The co-occurrence analysis reveals that narratives about air pollution in India are primarily geared towards mobilizing support for action. Regulatory enforcement and deterioration of air quality are the main storylines driving calls for intervention. Additionally, systemic and bureaucratic failures are frequently highlighted as significant barriers, reinforcing the need for mobilization to overcome these obstacles. Narratives also occasionally frame actors as heroes or villains, adding moral weight to the stories.

## 4.2. Summary

The landscape of air pollution narratives in India presents a complex tapestry of characters, plots, and belief systems, each contributing to the overarching story of a nation grappling with environmental challenges. At the heart of this narrative are the heroes doctors, government entities, and researchers who tirelessly advocate for and implement measures to mitigate air pollution. Their efforts are often portrayed in stark contrast to the villains, such as farmers engaged in stubble burning, power plants, and urban traffic, all of which contribute significantly to the air pollution crisis.

Victims of this environmental battle are widespread, ranging from the general population, especially in severely affected areas like Delhi, to vulnerable groups such as pregnant women and ecological entities like birds. These victims' stories highlight the human and environmental toll of unchecked pollution, screaming the urgency for effective intervention.

Beneficiaries of clean air initiatives, including villagers and the broader environment, reflect the positive outcomes of a few successful mitigation efforts. Allies in this fight come from various quarters, including government bodies, educational institutions, and individual advocates, all supporting the heroes' mission. However, the narrative is also peppered with opponents political figures, businesses, and associations who resist changes that threaten their interests. Alongside them are the ineffectives, entities whose efforts, such as cracker guidelines and certain political actions, fail to make a meaningful impact on reducing pollution.

The stories told through these characters are woven into various narrative elements. The most dominant plot is that of "Regulatory Enforcement" (59 out of 297 articles, 19.86%) which emphasizes the crucial role of government regulations in combating air pollution. This is followed by the "Warning Tale" (14.81%), a plot that cautions about the dire consequences of inaction, and the "Story of Decline" (14.15%), which paints a bleak picture of deteriorating air quality, urging immediate attention and action.

Moral narratives within these stories are often explicit (53.54%), with direct calls for policy solutions, while others subtly imply the benefits of addressing pollution through indirect references (44.78%). The belief systems underlying these narratives predominantly align with the "Hierarchist" perspective (45.11%), advocating for structured regulations and state-led initiatives. Modernist (12.13%) and egalitarian (9.76%) viewpoints also surface, highlighting the importance of technological innovation and community involvement, respectively.

Narrative strategies employed in these stories frequently aim to mobilize support (43.44%) for clean air policies, rallying public engagement by emphasizing the severe impacts of pollution and the collective benefits of mitigation efforts. Strategies like "Concentrating Costs and Diffusing Benefits" (21.88%) and "Mechanical" (9.76%) narratives further talk about government action and inaction.

# 5
# Results - Automated Coding

## 5.1. Results from LLM - Gemini

The same LLM was used twice to obtain codes based on the codebook. The following results and their comparisons between Run 1 and 2 were obtained -

### 5.1.1. Character N-Grams

Both runs of Gemini identified nearly the same important characters, showing a certain degree of repeatability as identified in the N-grams. However cosine similarity scores are low, with most characters having cosine similarity scores between 0.19 - 0.4 between run 1 and 2, indicating that it may not be capturing the actual similarity accurately, or possibly the highest counts are the ones identified consistently, with differences in less frequent characters identified. Setting has a higher similarity score than every character, suggesting a better performance in capturing the same geographical context repeatedly. The N-gram was created to capture the 30 more recurring meaningful characters, but it ended up capturing between 10-20 for each character, possibly due to the lack of repetition across each coded article.

#### Overview of Extra Words

Terms such as 'implied', 'explicitly mentioned', and other similar phrases appear in the N-grams, indicating that these words are likely artifacts of the coding process rather than substantive content from the articles. These terms do not provide meaningful information about the heroes, villains, victims, beneficiaries, allies, opponents, or ineffective efforts in the context of air pollution but instead reflect the methodological approach used in the analysis. Care has been taken to remove these from the analysis as far as possible.

#### Examples of Extra Words in Context

Implied: This term appears in several categories such as 'Villain', 'Victim', and 'Ineffective Efforts'. It suggests that certain actors or actions are inferred but not directly stated in the text. Its presence in the N-grams highlights the use of inferential coding techniques, possibly the LLM judging their implications from the context.

Explicitly Mentioned: This term is seen in categories like 'Opponent' and 'Beneficiary'. It indicates that the coding process involved identifying whether specific actors or actions were directly mentioned in the text. The term itself, however, is not useful in understanding the narrative content.

Implicitly: Similar to 'implied', this term appears in 'Villain' and 'Ineffective Efforts'. It suggests that certain details are indirectly referenced. Its inclusion in the N-grams indicates a methodological focus on inferred meanings rather than only coding if it was explicitly mentioned.

### Impact on Analysis

The presence of these extra words affects the clarity and usefulness of the N-grams. These terms do not contribute to understanding the narrative elements but instead reflect the coding process. They may obscure more meaningful content in the analysis if unchecked.

Refining the analysis process and implementing automated filtering techniques can help achieve this goal, ensuring that the N-grams more accurately reflect the substantive content of the articles.

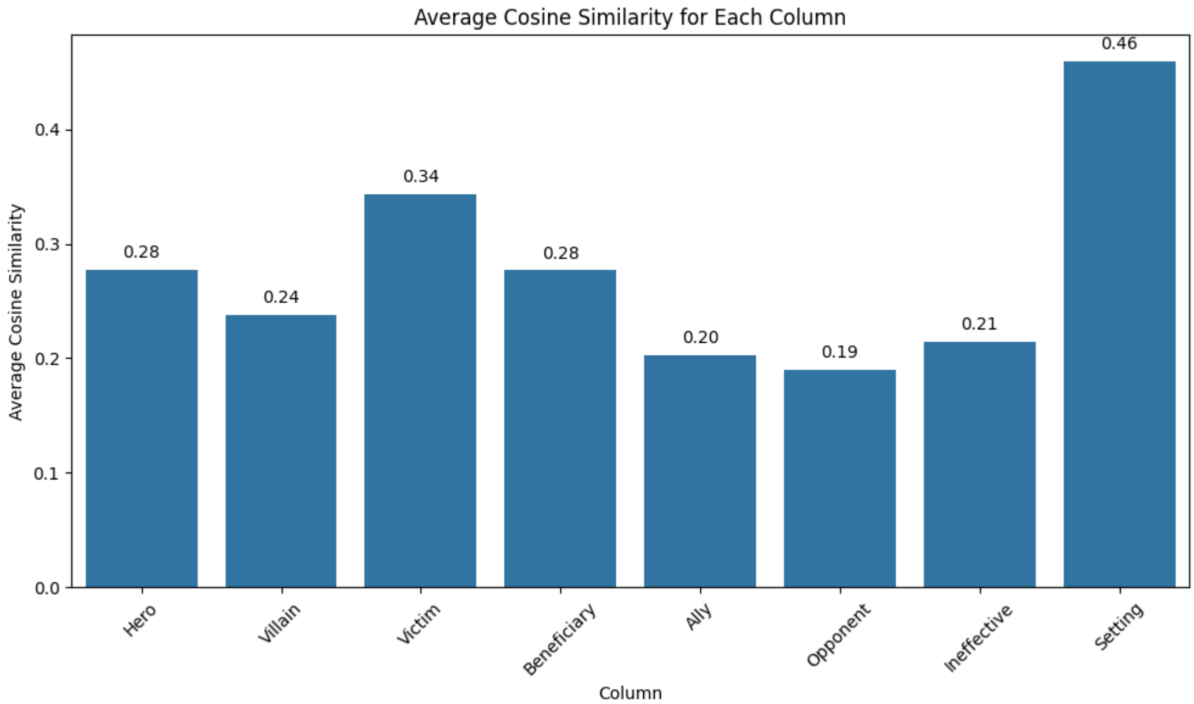



**Figure 5.1:** Between Gemini Run 1 and Gemini Run 2

The cosine similarities between the two runs is low, even though they recognize most of the same frequently occurring characters as coded by the LLM. The similarity for heroes is 0.28, Villains is 0.24, Victims is the highest at 0.34, Beneficiaries at 0.28, Allies at 0.20, Opponents are the lowest at 0.19, and Ineffectives at 0.21. Settings has the highest similarity score at 0.46, which is still low. This leads to the conclusion that characters are not consistently captured in the same category consistently according to the cosine similarity.

### Characters - Heroes

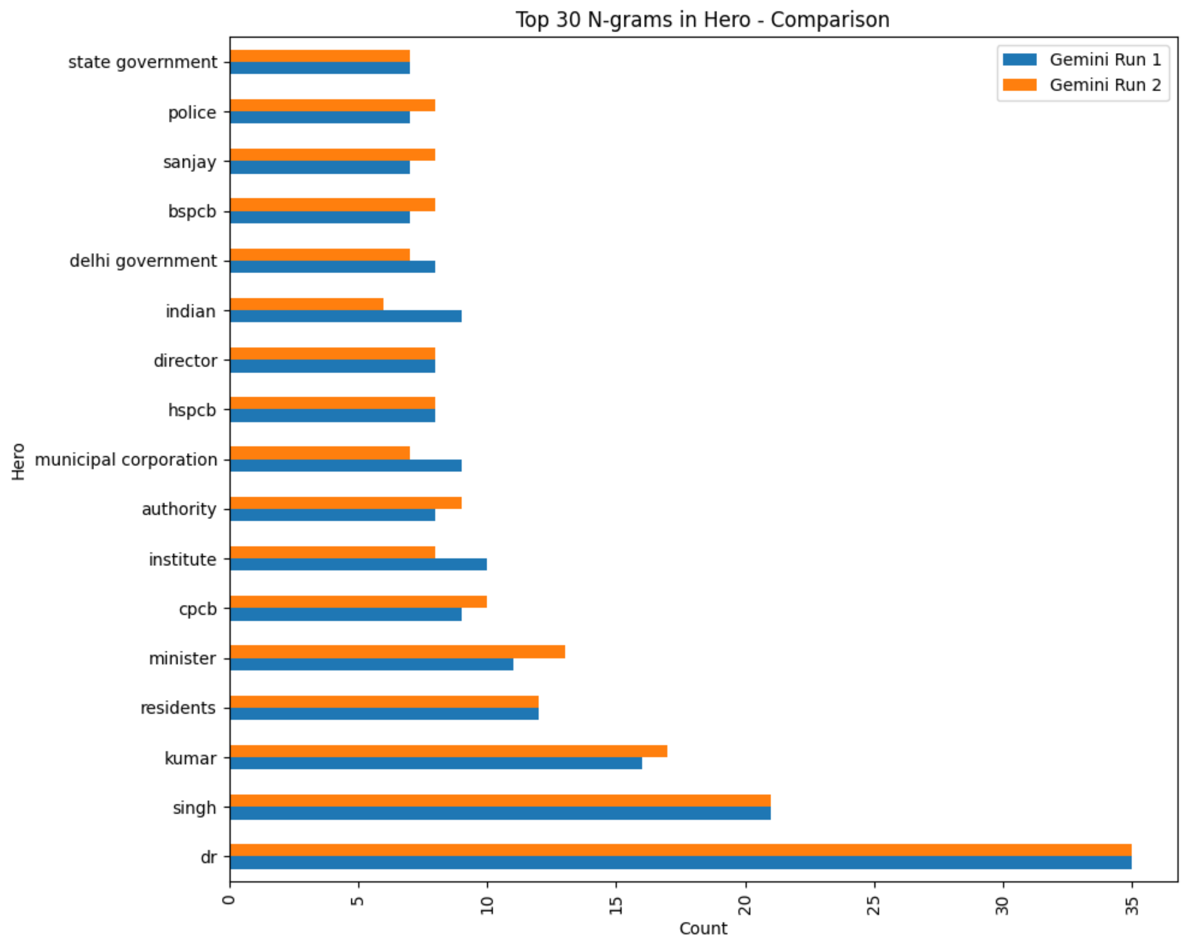



**Figure 5.2:** N-Gram of Heroes in Gemini

The heroes identified from the LLM are 'dr' (referring to Doctors of Medicine); 'singh'[1] and 'kumar'[2] (common surnames in India) who are private citizens, doctors, or people in positions of power taking actions against air pollution; residents of various areas; ministers as in ministers of the government (environmental, power, etc.); and so on. The example given in the prompt (section A.1) for heroes was - "**Example:** An environmental activist leading a successful campaign to ban single-use plastics in a major city, thereby reducing plastic waste and improving air quality." Even though the prompt is specific, the LLM manages to recognize other instances of heroes, highlighting its transfer learning capability.

Both LLM runs identified similar characters in approximately the same numbers, indicating agreement and repeatability. Heroes however, have low cosine similarity (0.28 on a scale of 0 to 1), indi-

[1]'Singh' occurrences in the raw data as identified by the LLM - Jagbir Singh (Manav Utthan Manch), SK Singh (ARTO), Dr B P Singh, Tikender Singh Panwar (Deputy Mayor), Maninder Singh (Additional Solicitor General), Jatinder Singh Sandhu, Pritpal Singh, Gurdwara Singh Sabha, Renu Singh (Principal, Amity International School), Dr Virendra Singh, Kamaldeep Singh Sangah, Prof Dhruv Sen Singh (Director, air quality monitoring station, Lucknow University), NTPC, R K Singh (Power Minister), Raj Karan Singh, Rupesh Singh, Dhruvsen Singh, Kahan Singh Pannu (Tandrust Punjab Mission Director), Punjab State Council of Science and Technology, Punjab Pollution Control Board, Hardeep Singh Puri, Rambali Singh, Gurmeet Singh Meet Hayer (Punjab Environment Minister), KK Singh

[2]'Kumar' occurrences in the raw data as identified by the LLM - 'Anjan Kumar Dutta (reputed structural engineer)', 'Nem Kumar Banthia', 'Siddhanta Kumar Dash', 'Girish Kumar', 'Inspector Anil Kumar', 'Sushil Kumar Modi', 'Dr Arunesh Kumar', 'Dr. Sanjay Kumar', 'Dr. Pratyush Kumar', 'Kumari Ria', 'Dr Arunesh Kumar', 'Ravindra Kumar (District Magistrate)', 'Shiv Kumar (farmer)', 'Dr Ashok Kumar Ghosh (BSPCB chairman)', 'Mahesh Kumar', 'Shri. Rajesh Kumar Pathak', 'Rajeev Kumar Mishra'

cating a diagreement in the details of heroes for each article, or that the method of measurement does not fully capture the equivalence of named entities, possibly due to their proper noun nature. There could be other methods to perform this task.

### Characters - Villains

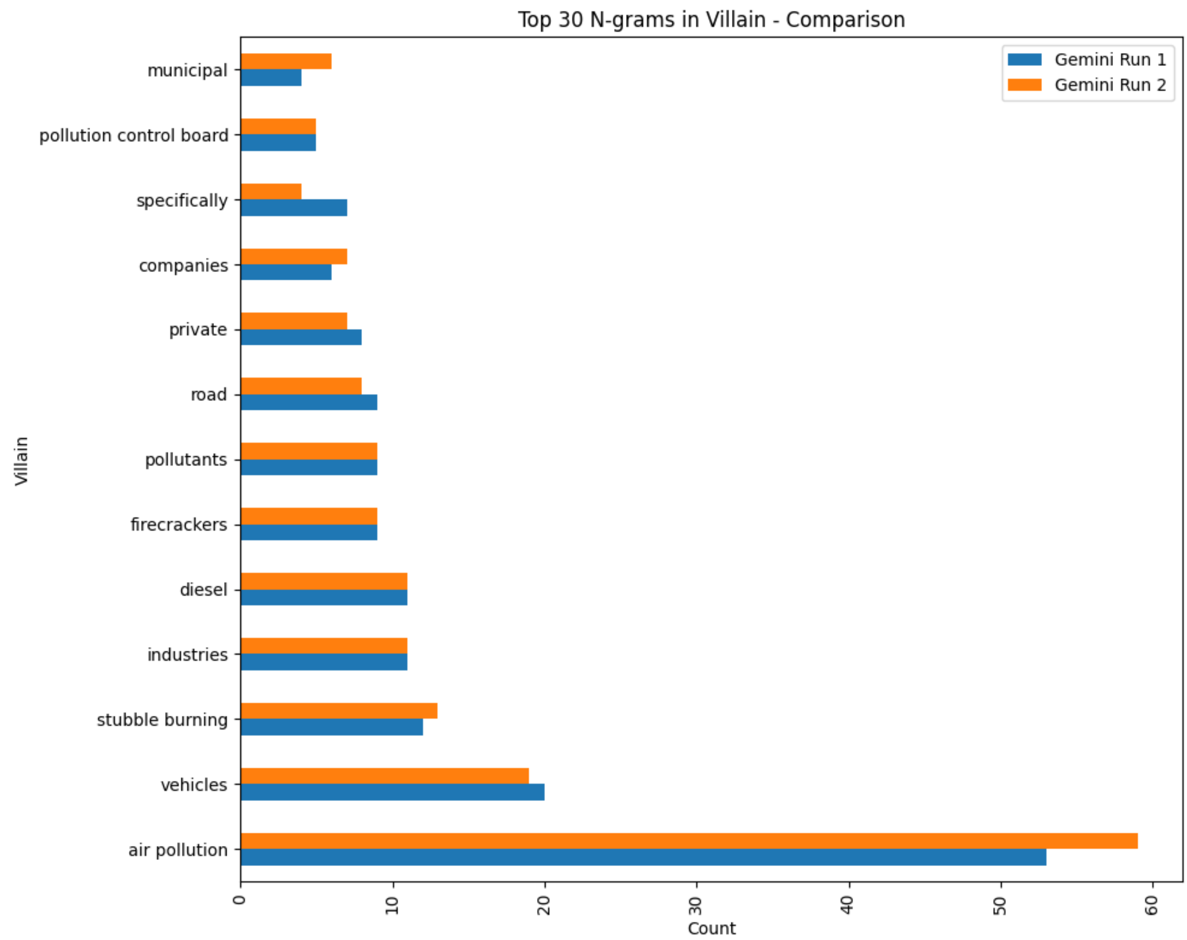



**Figure 5.3:** N-Gram of Villains in Gemini

Both runs have mistakenly identified the villain as the policy problem itself. Other than that, the results for villain also identify important sources of the policy problem, such as vehicles, stubble burning, industries[3], diesel (engines and trucks), firecrackers, etc. The cosine similarity scores for villains are low (0.24) as well, indicating the same limitations as identified in subsubsection 5.1.1.4. The example given in the prompt (section A.1) for identifying villains was - "**Example:** Large industrial corporations found guilty of illegally dumping toxic waste and emitting high levels of pollutants into the air, despite regulations."

[3] In the raw data as identified by the LLM - ['Polluting industries (specifically mentioned: petro-chemicals', 'industries emitting smoke', 'Union ministry of heavy industries and public enterprises', 'Chemical industries in Vapi', 'Coal-based industries', 'non-PNG industries', 'Polluting industries and coal mines in Chandrapur', 'Air Pollution Sources (specifically mentioned: road dust and industries)']

### Characters - Victims

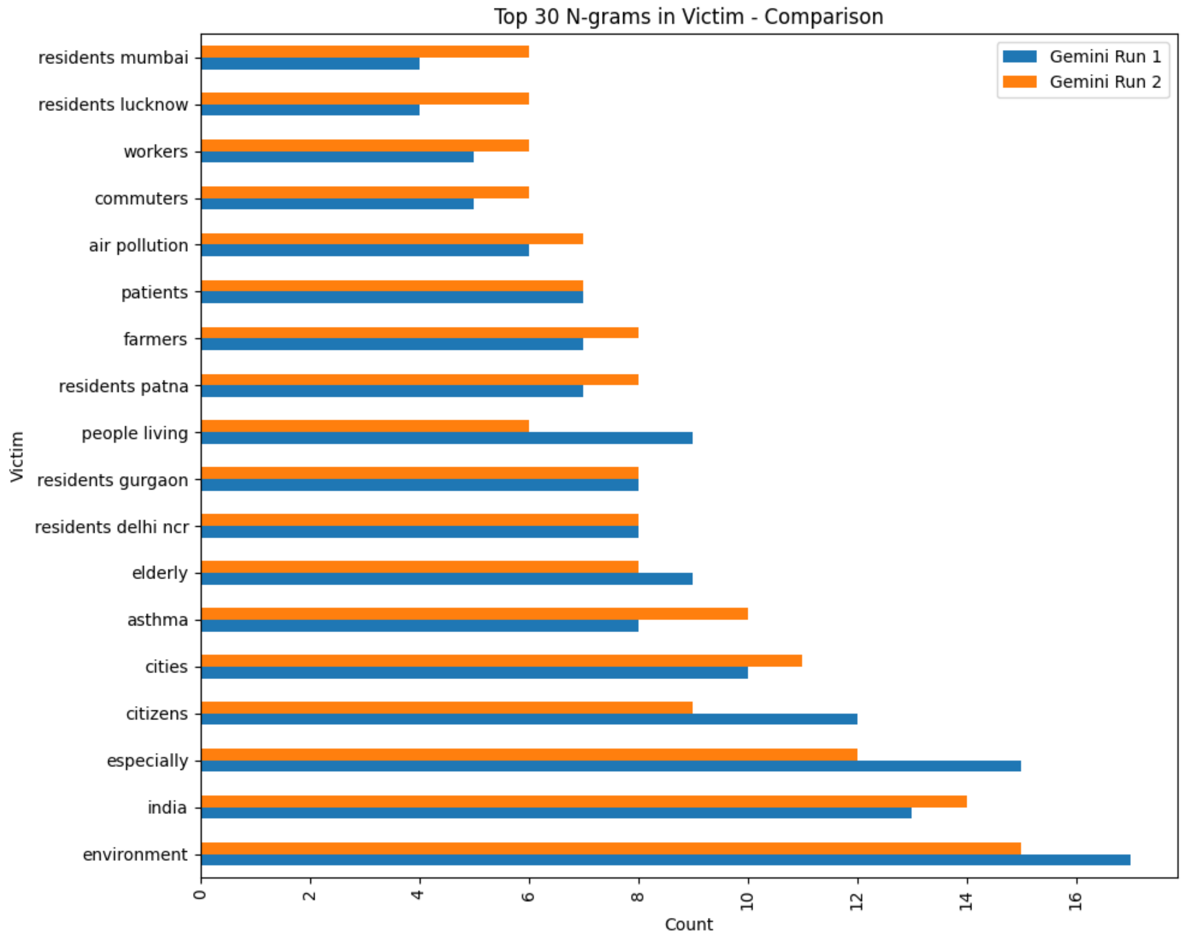



**Figure 5.4:** N-Gram of Victims in Gemini

It is a similar story with victims, where the LLM has managed to capture entities, such as the environment, successfully. Other than that, we see the expected victims such as residents and at-risk populations. Some examples such as 'asthma'[4], 'cities'[5], 'especially'[6] become clearer with the surrounding words for context (see footnotes). The example given in the prompt (section A.1) for identifying victims was - "**Example:** Residents of a community suffering from respiratory issues due to the nearby factory's unchecked emissions, highlighting the human cost of industrial pollution."

[4]'asthma patients', 'Students with medical ailments like asthma and bronchitis', 'people suffering from asthma in Delhi', 'including asthma', 'children with asthma'

[5]'Residents of Madhya Pradesh cities', 'Residents of other polluted cities like Gaya and Lucknow', 'and other polluted cities in Maharashtra', 'especially those in the 17 non-attainment cities', 'Residents of NCR cities', 'Residents of cities impacted by air pollution', 'Residents of major cities', 'Residents of Patna and other Bihar cities', 'and other cities in southern India', 'People in metro cities'

[6]'especially children and senior citizens', 'especially those living in areas with poor air quality', 'especially children (suffering from respiratory ailments)'

### Characters - Beneficiaries

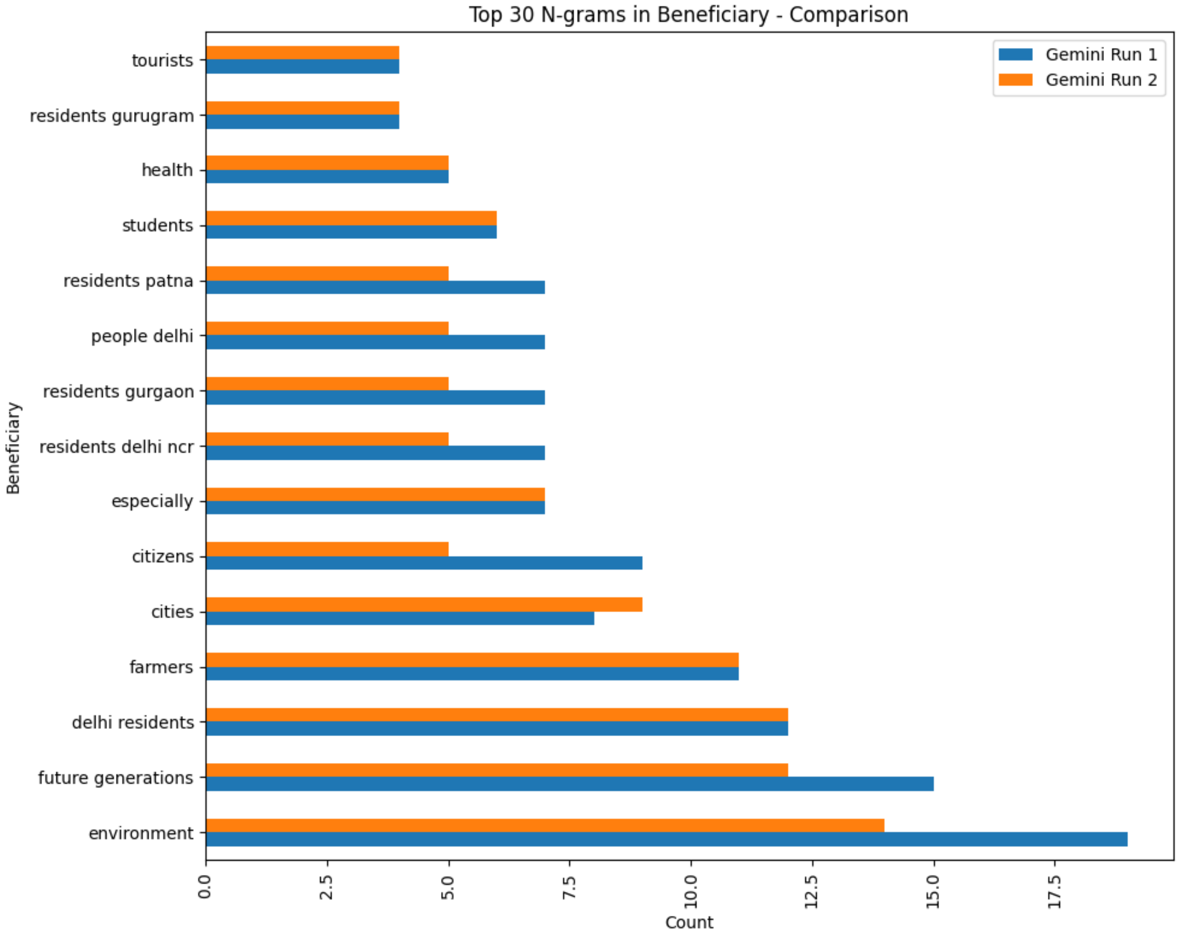



**Figure 5.5:** N-Gram of Beneficiaries in Gemini

Similar to victims, beneficiaries have been captured in the right context. Additional context is required for words such as 'cities'[7], 'especially'[8], and 'health'[9]. Discrepencies have started popping up here, where both runs have different frequencies for coding the same characters. The example given in the prompt (section A.1) for identifying beneficiaries was - "**Example:** School children in urban areas who experience improved health outcomes and reduced asthma rates after the implementation of stringent air quality standards."

[7]'Residents of NCR cities', 'Residents of cities impacted by air pollution', 'People of Mumbai and other cities', 'and other cities in Bihar', 'residents of metro cities'

[8]'especially those living in urban areas', 'Residents of Delhi (especially those with respiratory issues)', 'especially those with pre-existing respiratory conditions'

[9]'Those experiencing improved air quality and health', 'Public health', 'People who follow the health tips', 'but especially those with health concerns)', 'People who experience improved health outcomes'

### Characters - Allies

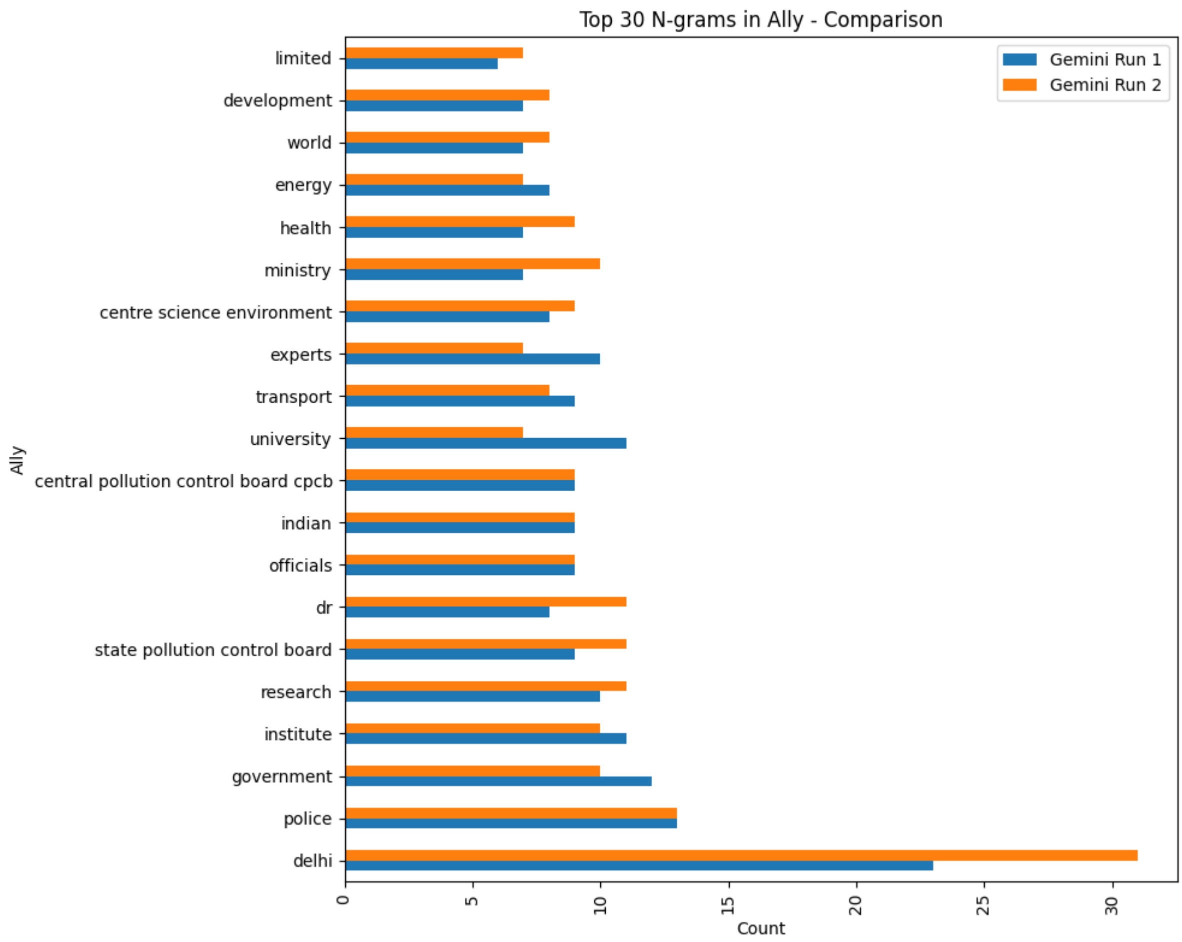



**Figure 5.6:** N-Gram of Allies in Gemini

With allies as well, 'delhi'[10] has been captured, but the surrounding context for the word appears later as 'police' and 'government'. Additional context for other words such as 'indian'[11], 'transport'[12], 'health'[13] also aids the analysis. The example given in the prompt (section A.1) for identifying allies was "**Example:** Non-governmental organizations (NGOs) and community groups collaborating with environmental activists to lobby for stricter air pollution controls and public awareness campaigns."

[10]'Delhi and Bombay IITs', 'New Delhi Municipal Council', 'Municipal Corporations of Delhi', 'Delhi Development Authority', 'Delhi Dialogues Commission', 'Delhi Pollution Control Committee (DPCC)', 'Delhi Taxi Tourist Transporters and Tour Operator Association', 'Delhi University College of Medical Sciences', 'IIT Delhi', 'Delhi Traffic Police', 'Delhi Government', 'Municipal Corporation of Delhi (MCD)', 'Delhi Transco Limited', 'Tata Power Delhi Distribution Limited', 'Delhi', 'Delhi Police', 'Delhi BJP President Virendra Sachdeva', 'IIT Delhi Alumni Association', 'Delhi Traffic Police'

[11]'Indian Railways', 'Environmental monitoring division of Indian Institute for Toxicology Research', 'Indian Institute of Toxicology Research', 'CEOs from Indian companies like ICICI', 'Indian Medical Association (IMA)', 'Indian Oil Limited', 'IOCL (Indian Oil Corporation Limited)', 'Indian Stroke Association (ISA)', 'Federation of Indian Chambers of Commerce and Industry (FICCI)'

[12]'Transport experts', 'Regional Transport Office (RTO)', 'Delhi Taxi Tourist Transporters and Tour Operator Association', 'Transport Department Official', 'Regional Transport Office (RTO)', 'Transport department', 'Transport and Revenue Departments', 'Transport Department', 'Transport Department Officials'

[13]'Researchers at Imperial College London and the National Institute for Public Health and the Environment in the Netherlands', 'Centre for Environment Occupational and Health)', 'World Health Organization (WHO)', 'World Health Organisation (WHO)', 'Health experts', 'Healthcare experts', 'Healthy Energy Initiative India', 'World Health Organisation (WHO)'

### Characters - Opponents

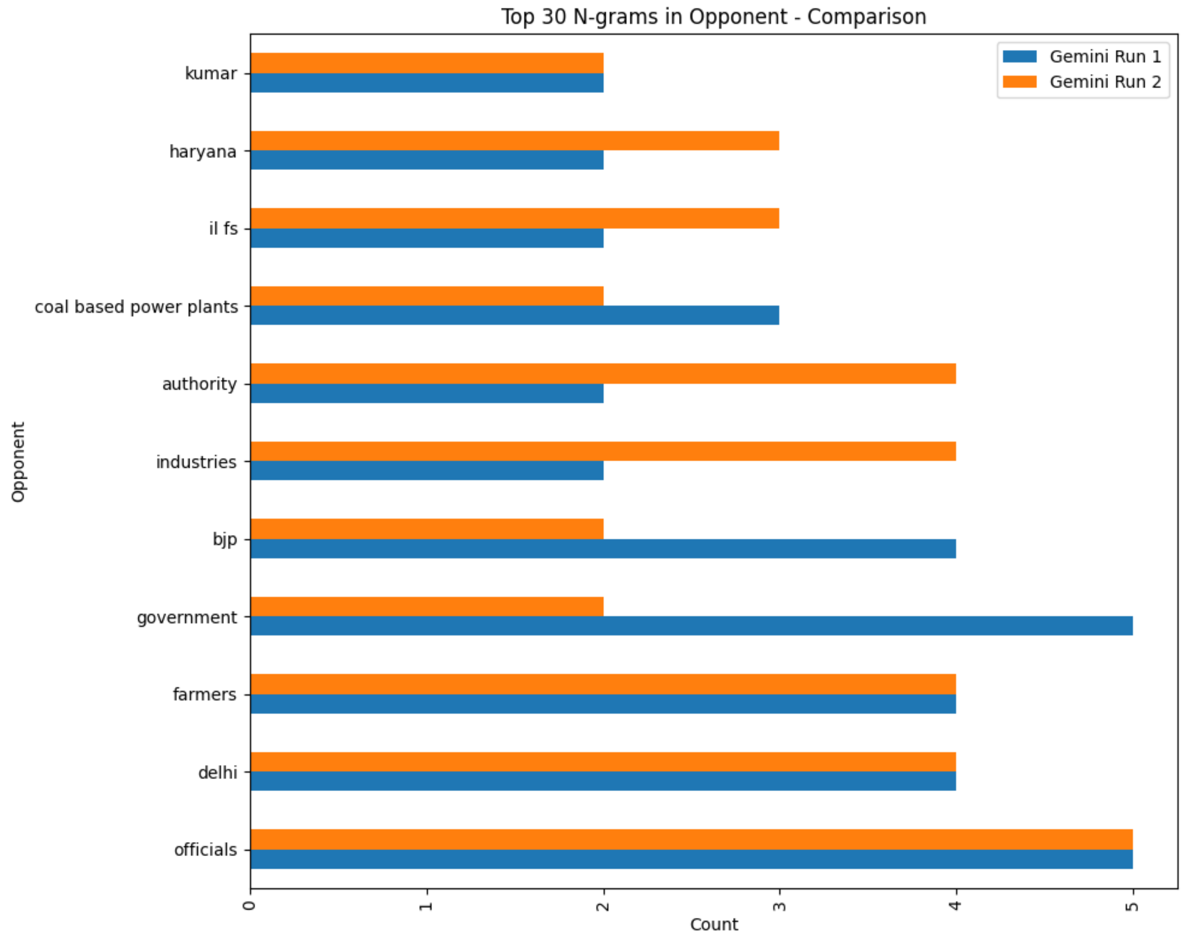



**Figure 5.7:** N-Gram of Opponents in Gemini

The government has been captured twice as many times in run 1 than in 2. However, given there are only 5 observations, this may change with a larger dataset. Context for 'il fs'[14], 'haryana'[15] (an Indian State), and 'delhi'[16] is in the footnotes. The example given in the prompt (section A.1) for identifying opponents was "**Example:** Local businesses that oppose new air quality regulations due to concerns over increased operational costs, arguing that the economic impact outweighs the benefits of cleaner air."

[14] 'Infrastructure Leasing and Financial Services Ltd. (IL&FS)', 'IL & FS'
[15] 'HUDA (Haryana Urban Development Authority)', 'BJP Governments of Uttar Pradesh and Haryana'
[16] 'DPCC (Delhi Pollution Control Committee) officials', 'Delhi Taxi Tourist Transporters and Tour Operator Association', 'Delhi Metro Rail Corporation (DMRC)', 'Delhi government'

### Characters - Ineffectives

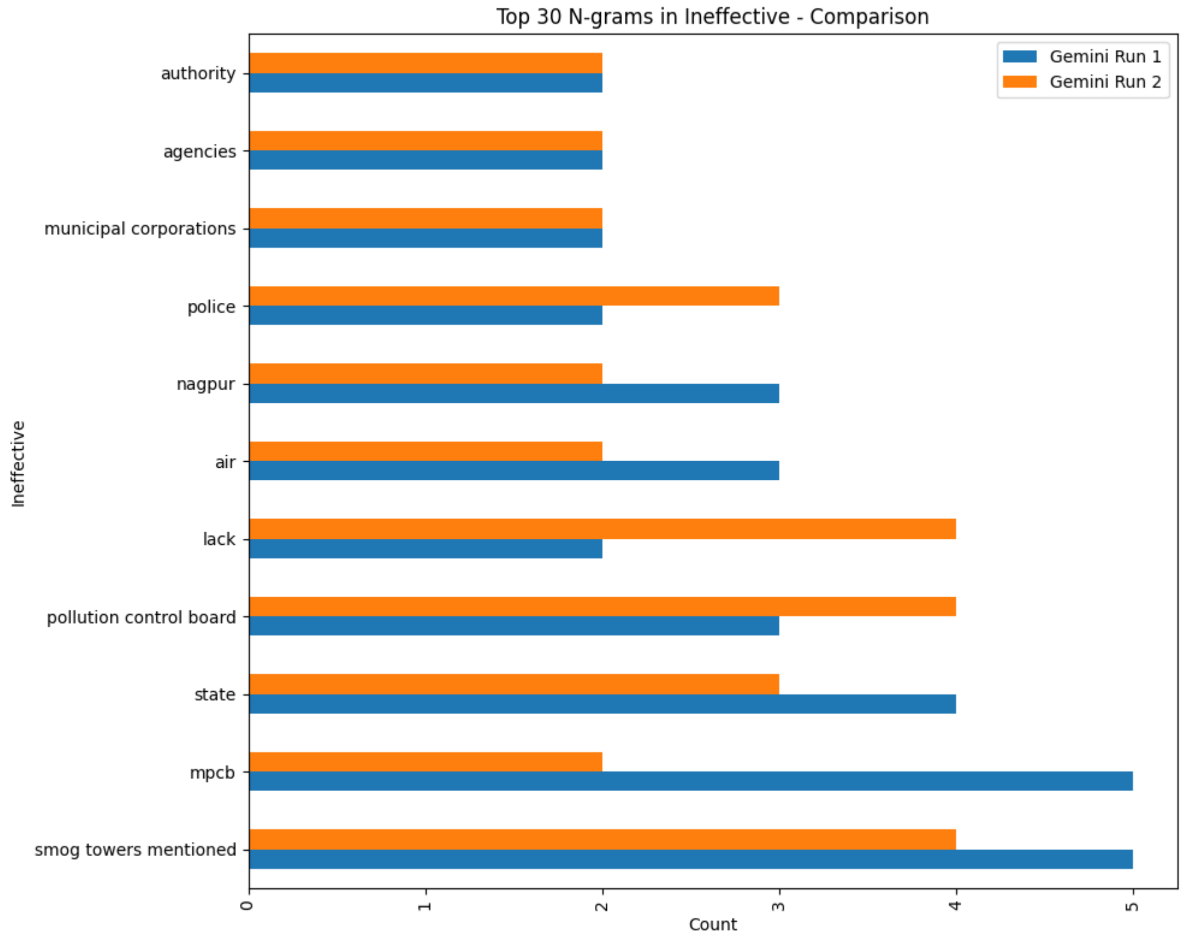



**Figure 5.8:** N-Gram of Ineffectives in Gemini

Here, the characters captured seem to be driven by the example given in the prompt (smog towers). The example given in the prompt (section A.1) for identifying ineffectives was "**Example:** The government installs smog towers in public areas to reduce air pollution, but the towers cannot work in that setting with that much volume, hence rendering their solution useless and their intervention ineffective." Other than that, the results seem appropriate for the definition. 'mpcb' stands for Maharashtra Pollution Control Board. Context for 'state'[17], 'air'[18], 'lack'[19], and 'nagpur'[20] is in the footnotes.

[17] 'State government (due to inaction)', 'RSPCB (Rajasthan State Pollution Control Board) officials', 'Different Agencies of the State', 'State Forest Department'

[18] 'Mysuru City Corporation (MCC) - spraying chemicals for air pollution', 'Air Purifiers (implied as potentially not effective for birds)', "Nagpur's Clean Air Action Plan", 'Commission for Air Quality Management (due to lack of support)'

[19] 'MPCB Officer (due to lack of enforcement power)', 'Commission for Air Quality Management (due to lack of support)

[20] 'Nagpur Municipal Corporation', "Nagpur's Clean Air Action Plan", 'NMC and NIT (Nagpur Improvement Trust)'

### 5.1.2. Plots

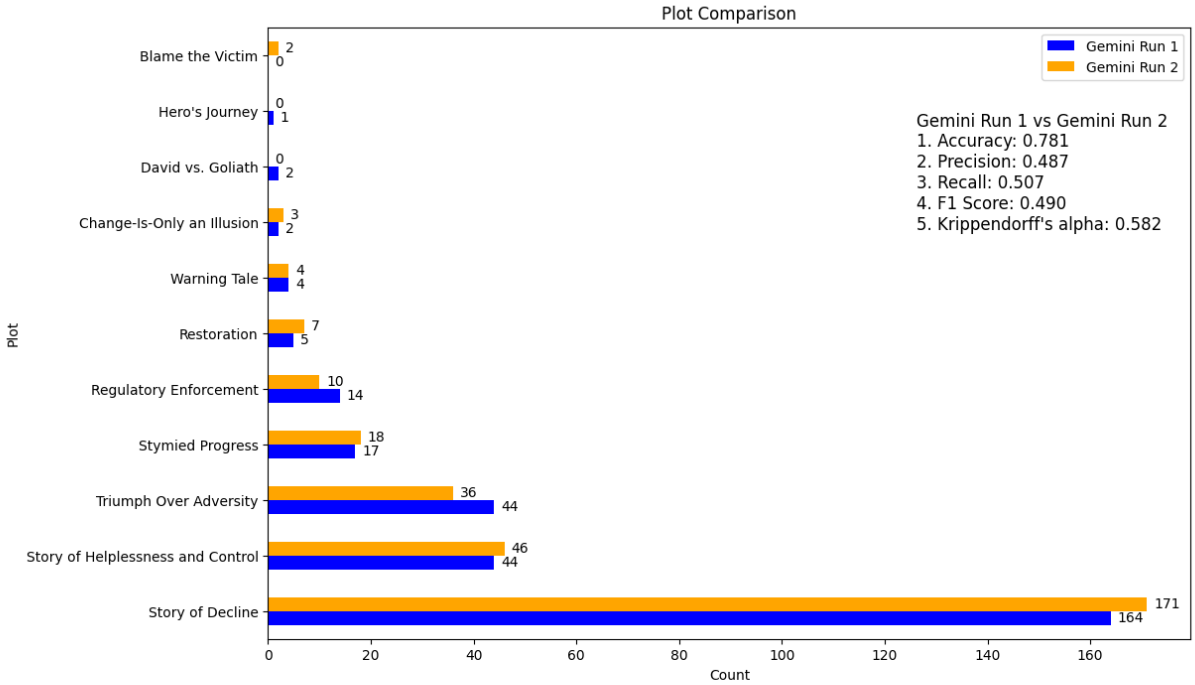



**Figure 5.9:** Comparison of Plot Frequencies

For plot elements, the results show moderate performance. The accuracy is 0.781, meaning that about 78.1% of the codes matched each other. The precision value is 0.487, indicating that 48.7% of the predicted codes were correct. The recall value is 0.507, reflecting that many true positives were missed, with only 50.7% correctly identified. The F1 score of 0.490 demonstrates a balance between precision and recall. Krippendorffs alpha value of 0.582 suggests moderate agreement, indicating a fair level of reliability in the systems coding.

Plot capturing is higher in similarity than characters between runs. The Krippendorff's alpha corroborates this agreement between runs, although it is still lower than the generally accepted standard of 0.8. From subsection 3.6.2, data with Krippendorff's $\alpha$ reliability coefficients below 0.667 are generally disregarded due to insufficient agreement. The LLM codes for 'Story of Decline' more (164-171) than any other class, the next closest is 'Story of Helplessness and Control' (44-46 times). Class imbalances[21] lead to a low score of $\alpha$ (Feng, 2014), which means that when the data is heavily skewed towards a particular category or if one category is rarely coded, it can result in a low alpha. This could be mitigated by using a larger dataset.

[21] Class imbalance occurs when the classification data has a skewed distribution with one or more classes making up the majority of the data.

### 5.1.3. Moral

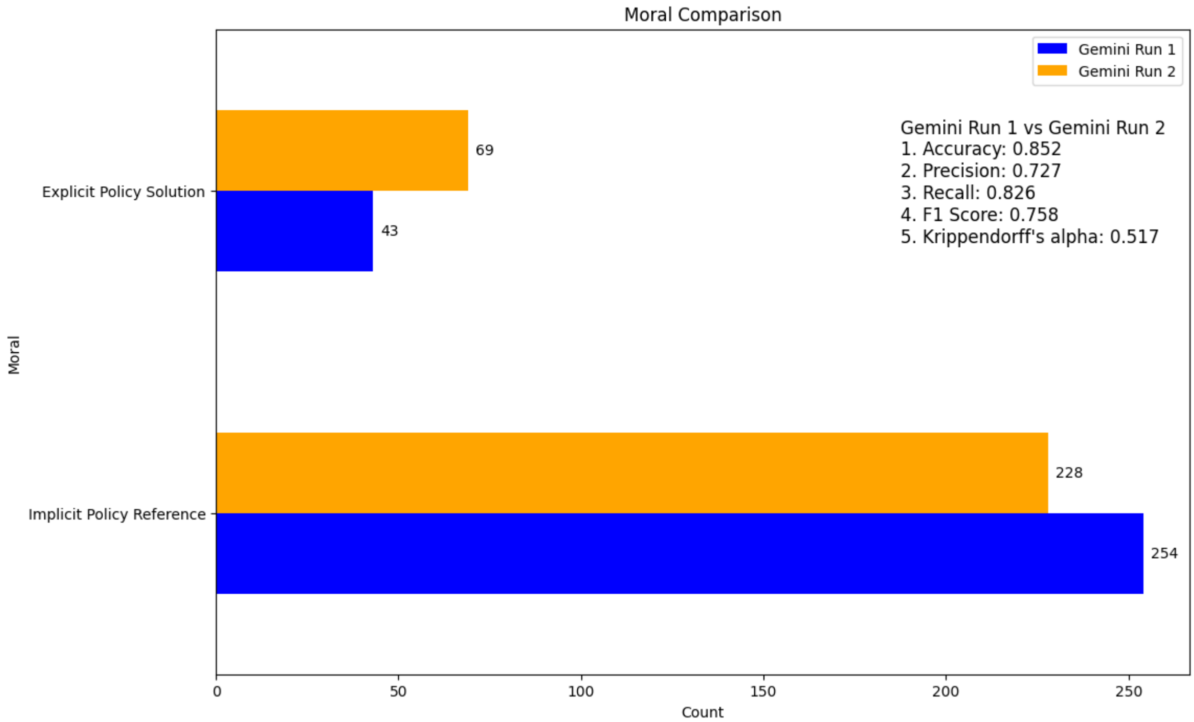



**Figure 5.10:** Comparison of Moral Frequencies

For moral elements, the results show moderate performance. The accuracy is 0.852, meaning that about 85.2% of the codes matched each other. The precision value is 0.727, indicating that 72.7% of the predicted codes were correct. The recall value is 0.826, reflecting that many true positives were identified, with 82.6% correctly identified. The F1 score of 0.758 indicates a relatively good balance between precision and recall. Krippendorffs alpha value of 0.517 suggests moderate agreement, indicating some reliability in the systems coding. Similar disagreement in moral is observed, with a score of 0.5167. Most articles (228-254) are coded to have an 'Implicit Policy Reference'.

### 5.1.4. Belief Systems

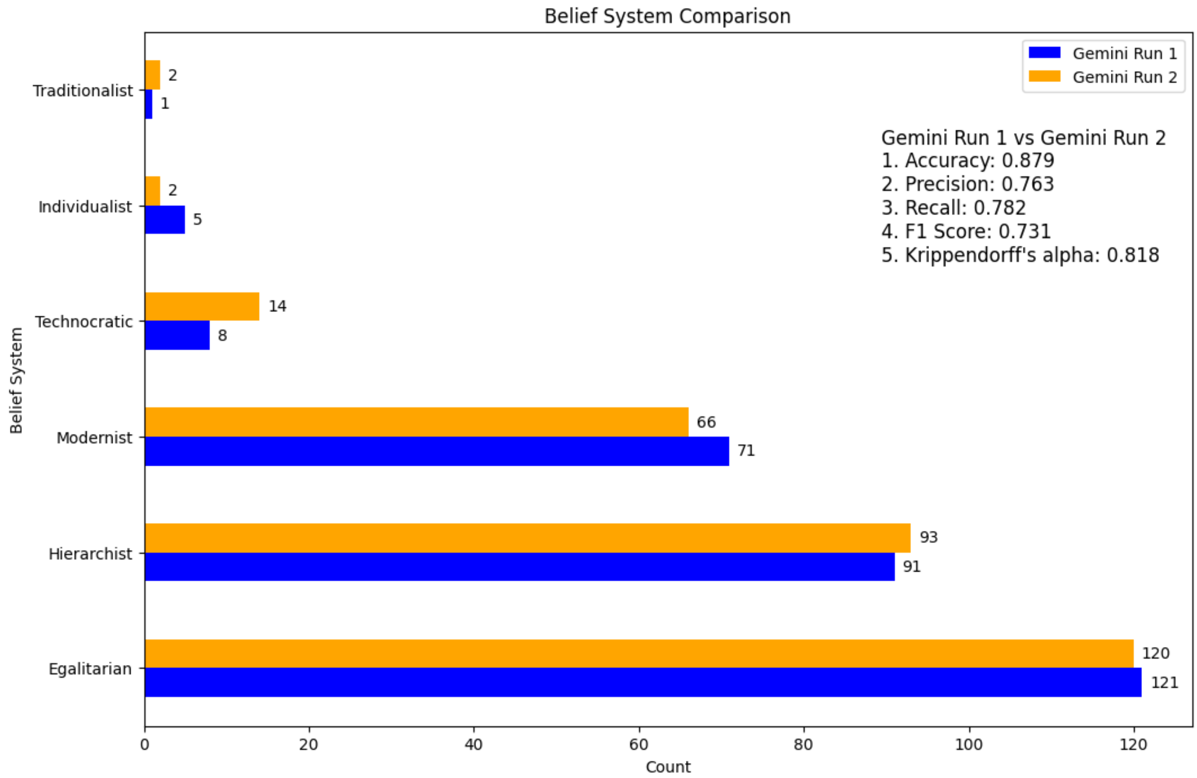



**Figure 5.11:** Comparison of Belief System Frequencies

For belief systems, the results demonstrate a strong performance. The accuracy is 0.879, meaning that approximately 87.9% of the codes matched each other. The precision value is 0.763, indicating that 76.3% of the predicted codes were correct. The recall value is 0.782, reflecting that a significant number of true positives were identified, with 78.2% correctly identified. The F1 score of 0.731 indicates a good balance between precision and recall. Krippendorffs alpha value of 0.818 suggests a high level of agreement, indicating the systems coding is reliable.

Belief system has a good agreement amongst subsequent runs, with the score of 0.818 which is higher than 0.8. This would be accepted as reliable by social science research standards, indicating the ability of LLMs to code this category consistently. The most common belief system is found to be 'Egalitarian' (120-121 instances out of 297), followed by 'Hierarchist' (91-93 instances out of 297) and 'Modernist' (66-71 instances out of 297), leading to the takeaway that most articles (according to the LLM) stress upon community involvement to advocate for policies that ensure equal distribution of clean air as a shared resource, followed by a focus on the need for structured regulations and state-led initiatives to tackle air pollution, and the role of technology in addressing the policy problem.

### 5.1.5. Narrative Strategies

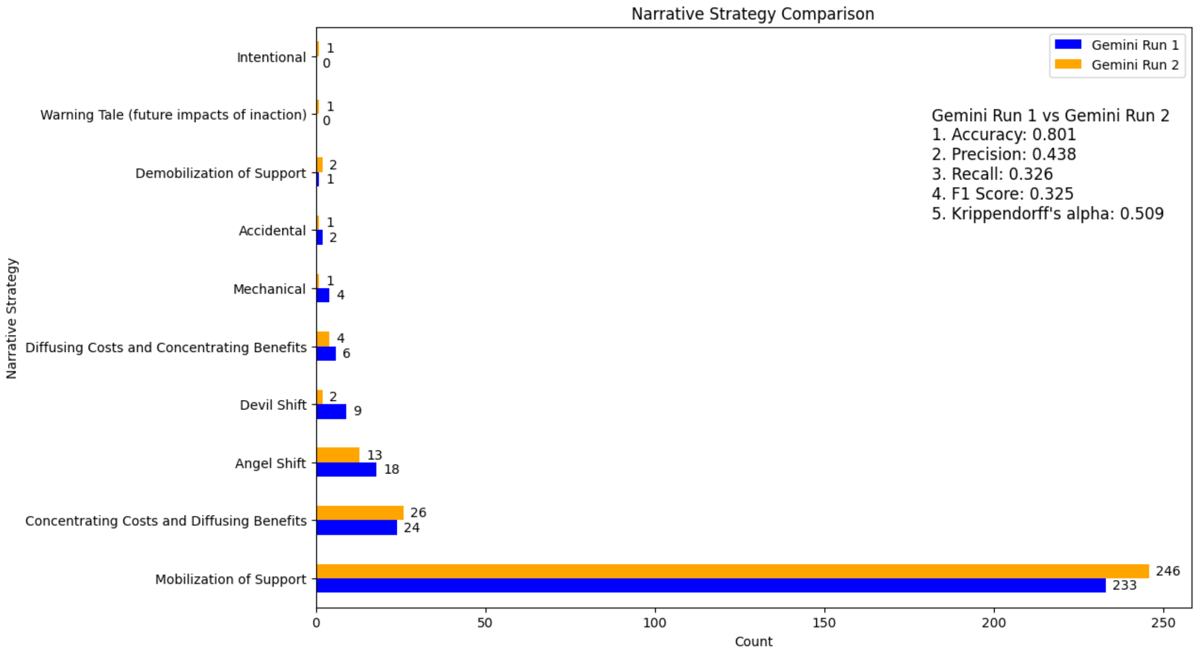



**Figure 5.12:** Comparison of Narrative Strategy Frequencies

For narrative strategies, the results indicate lower performance. The accuracy is 0.801, meaning that about 80.1% of the codes matched each other. The precision value is 0.438, indicating that only 43.8% of the predicted codes were correct. The recall value is 0.326, reflecting that many true positives were missed, with only 32.6% correctly identified. The F1 score of 0.325 demonstrates a poor balance between precision and recall. Krippendorffs alpha value of 0.509 suggests moderate agreement, indicating some reliability but also highlighting issues in the systems coding.

Krippendorff's $\alpha$ falls below the threshold for reliability. Run 2 wrongly identified warning tale (a plot) as a narrative strategy. Most articles (233-248 out of 297) are identified as having a 'Mobilization of Support' strategy, highlighting the class imbalance in the output.

## 5.2. Agreement between Consequent Runs

From Table 6.5, a promising result is the repeatability of results. In all the classification components, the lowest accuracy score is 0.78, showing a strong ability to identify the same instances with the same categories. However, with high accuracy, and lower scores otherwise indicate a few things -

1. High Accuracy with Moderate Precision and Recall in Plots: The method shows a significant improvement, suggesting better performance and consistency. The high Krippendorffs Alpha indicates strong agreement and reliability.
2. High Scores Across Metrics in Morals: The method is performing well in predicting moral elements, with strong precision and recall. The high Krippendorffs Alpha further confirms the reliability of the method.
3. Very High Scores in Belief Systems: The method excels in predicting belief system elements, showing high accuracy, precision, and recall. The very high Krippendorffs Alpha indicates excellent reliability.
4. High Accuracy but Lower Precision and Recall for Narrative Strategies: The method shows high

accuracy but struggles with precision and recall, suggesting some over-prediction of narrative strategies. The moderate Krippendorffs Alpha indicates good, but not perfect, reliability.

It is essential to acknowledge that these experiments were conducted using the default configuration of the Gemini 1.5 Flash model. Various inference hyperparameters can be fine-tuned to modify the output of a large language model (LLM) (Renze & Guven, 2024). Specifically, hyperparameters such as temperature, top-k, repetition penalty, and maximum token length have a substantial impact on both the quality and effectiveness of the LLMs generated output (Renze & Guven, 2024). To investigate the effects of such tuning, the temperature parameterintegral to sampling strategies that regulate the randomness of predictions generated by the language modelwas adjusted. Lower temperature values result in more deterministic, high-probability outputs, whereas higher temperature values introduce greater randomness, thereby producing more diverse outputs (X. Wang et al., 2023)(Gemini Documentation, 2024). In the Gemini 1.5 model, the temperature parameter ranges from 0 to 2, with the default setting being 1, where higher values correspond to increased output randomness (Gemini Documentation, 2024). For the purpose of this analysis, the temperature was set to 0, and the models were run twice. The resulting output was identical in all instances, demonstrating perfect repeatability (see Table A.2).

Moreover, while entirely repeatable coding is achievable, it may be more advantageous to obtain a range of outputs by varying the temperature parameter and synthesizing these outputs, especially in tasks that necessitate diverse reasoning approaches. A method known as "Self-Consistency," which involves this procedure, enhances the reasoning capabilities of LLMs (X. Wang et al., 2023). By altering the temperature, the model generates multiple reasoning paths for the same problem, potentially leading to different answers. The self-consistency method then aggregates these paths to identify the most consistent final answer, thereby improving the models accuracy (X. Wang et al., 2023). As previously noted, there are numerous approaches to optimizing the performance of an LLM for specific tasks, but these considerations are beyond the scope of this study.

The next step is to check how much automated coding agrees with the manual coding.

# 6

# Comparison of Manual and Automated Coding

## 6.1. Comparison between Manual Coding, Gemini Run 1, and Gemini Run 2

In this chapter, the results from chapter 4 and chapter 5 will be compared to see how much agreement the LLM coding has with the manual coding. The first point of comparison will be the characters, using a Cosine Similarity measure.

### 6.1.1. Cosine Similarities for Characters

For Characters, cosine similarities were calculated between Manual Coding, Gemini Run 1, and Gemini Run 2.

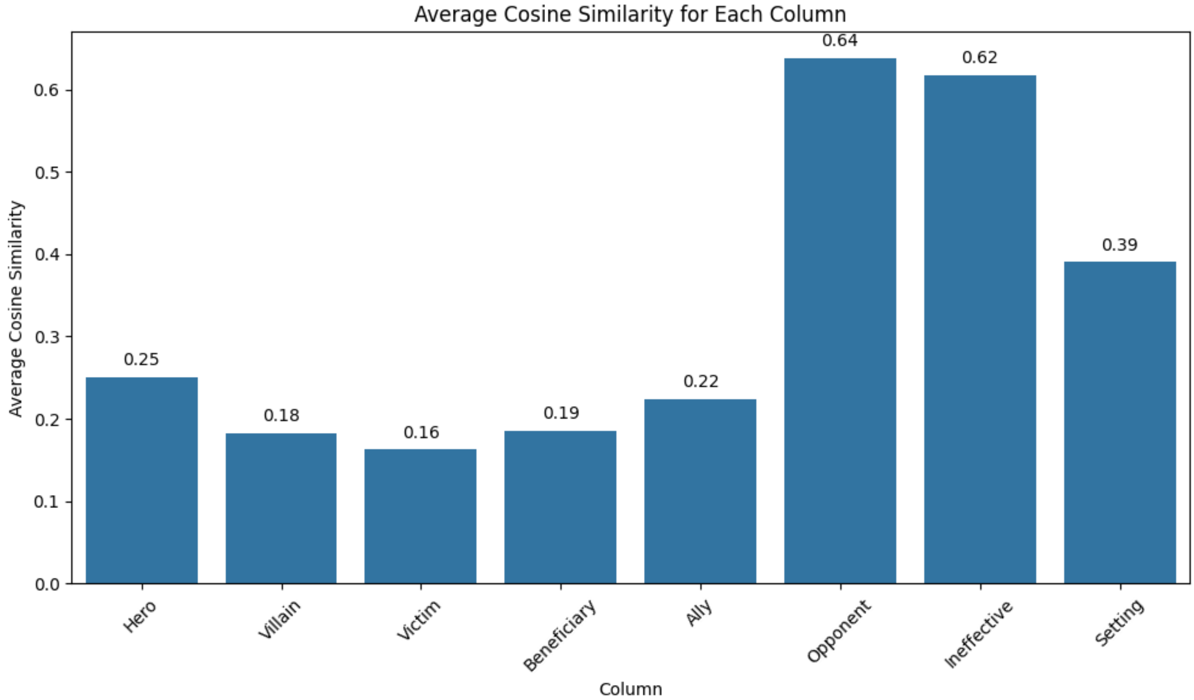



**Figure 6.1:** Between Manual Coding and Gemini Run 1

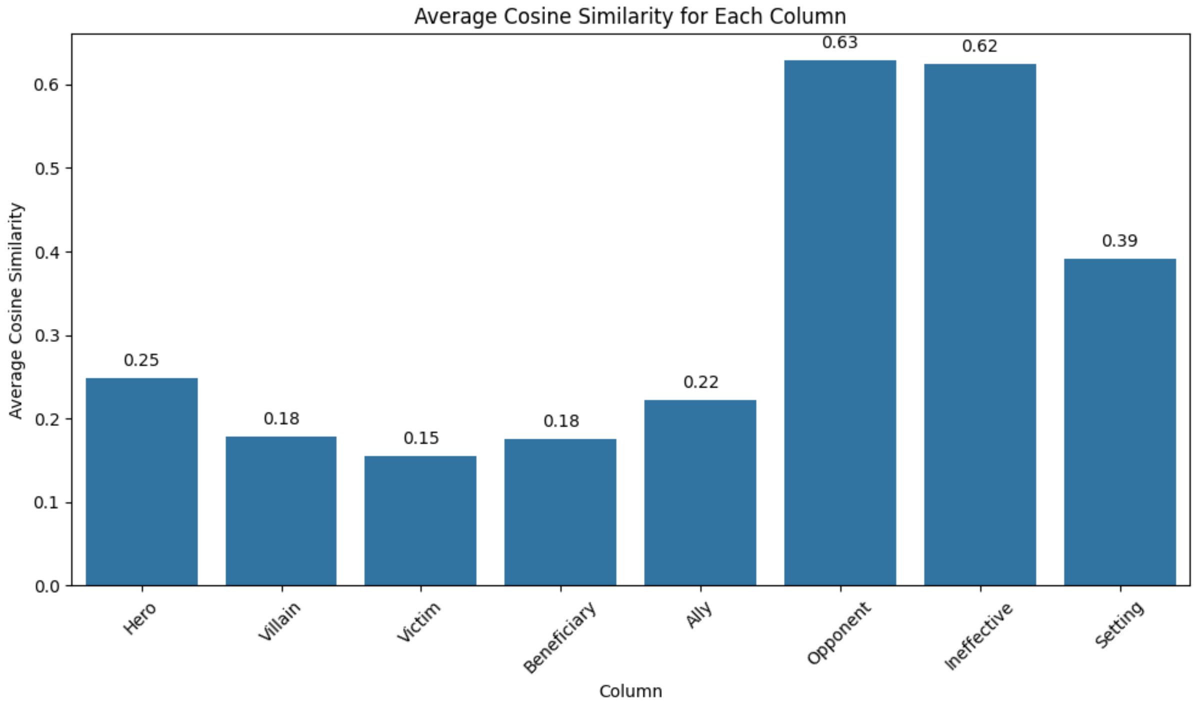



**Figure 6.2:** Between Manual Coding and Gemini Run 2

Cosine similarity is a measure used to determine how similar two vectors[1] are by calculating the cosine of the angle between them. It ranges from -1 to 1, where 1 indicates that the vectors are identical, 0 indicates that they are orthogonal (no similarity), and -1 indicates that they are diametrically opposed.

[1]The vectors being compared are the representations of the text in a particular category embedded in a high dimension space, done through the SentenceTransformers library (Reimers & Gurevych, 2019).

In the context of coding comparisons, a higher cosine similarity score indicates a greater alignment between the codes produced by the LLM (Large Language Model) and the manual coding.

For both runs, the cosine similarities between the LLM coding and manual coding are very consistent. The cosine similarity scores for each character category are as follows:

- **Hero**: 0.25 in both runs.
- **Villain**: 0.18 in both runs.
- **Victim**: 0.16 in Run 1 and 0.15 in Run 2.
- **Beneficiary**: 0.19 in Run 1 and 0.18 in Run 2.
- **Ally**: 0.22 in both runs.
- **Opponent**: 0.64 in Run 1 and 0.63 in Run 2.
- **Ineffective**: 0.62 in both runs.
- **Setting**: 0.39 in both runs.

The similarity in these scores between the two runs suggests that the LLM is consistently coding these categories in a manner that is comparable to manual coding. The high cosine similarity for categories like "Opponent" (0.64 and 0.63) and "Ineffective" (0.62 in both runs) indicates a strong alignment between LLM and manual coding in these categories. Conversely, categories like "Villain" (0.18 in both runs) and "Victim" (0.16 and 0.15) have lower cosine similarities, suggesting that there might be more variability or disagreement in how these categories are coded by the LLM versus manually.

The consistency of the cosine similarity scores across the two runs implies that the LLM is reliable in its coding practices. The higher scores for "Opponent" and "Ineffective" indicate that the LLMs coding is highly aligned with the manual coding for these categories, which may suggest that these categories have more distinct and identifiable features that the LLM can easily recognize and match. These are also the categories which have a higher proportion of '0' codes (component missing in the narrative), which could be the reason for this result.

On the other hand, the lower similarity scores for "Villain" and "Victim" suggest that these categories may be more challenging for the LLM to code accurately. This could be due to the more nuanced or subjective nature of these roles, as well as the fact that these categories also include many named entities with Indian names, which might not be as clearly defined or present in the training data used by the LLM.

### 6.1.2. Plots

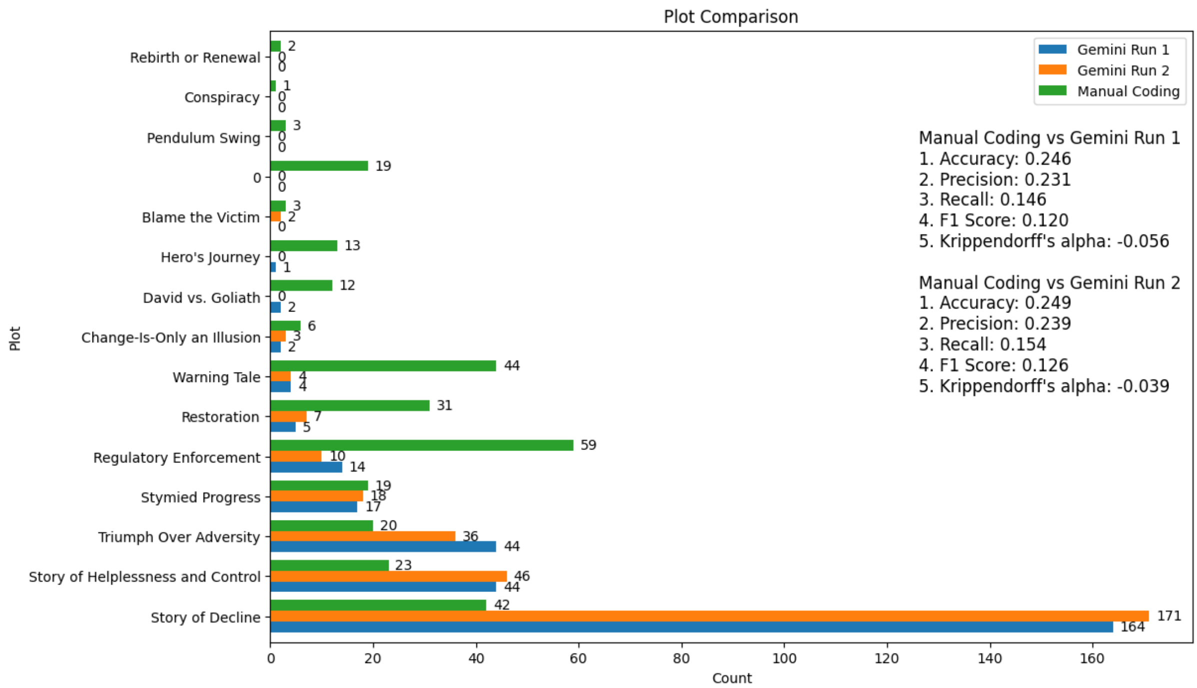



**Figure 6.3:** Comparison of Plot Frequencies

Key observations from the plot comparison:

- **Story of Decline**: The counts for this plot type are inconsistent between Groud Truth and LLM runs (which are closer to each other), with Manual Coding at 42, Gemini Run 1 at 164, and Gemini Run 2 at 171.
- **Story of Helplessness and Control**: There is a notable discrepancy between the Manual Coding (23) and the Gemini runs (44 and 46).
- **Triumph Over Adversity**: Significant variation is observed, with Manual Coding at 20, Gemini Run 1 at 36, and Gemini Run 2 at 44.
- **Regulatory Enforcement**: Manual Coding shows a high count of 59, while Gemini runs show significantly lower counts (14 and 10).
- **Restoration, Warning Tale, Change-Is-Only an Illusion, Hero's Journey, Blame the Victim, Pendulum Swing, Conspiracy, Rebirth or Renewal**: These plot types show considerable differences, with some plots having zero counts in one or both Gemini runs.

From Figure 6.3, we see that plots are not well captured by the LLM. With low scores on every metric in Table 6.5, the LLM does not appear to be very reliable when coding for this component. Many plots that were manually coded for, are entirely absent in the automated coding, such as Rebirth or Renewal, Pendulum Swing, Conspiracy, and the presence of no plot as well.

This is common across all components, showing the tendency of the LLM to provide an answer, even when there might not be one.

1. **Rebirth or Renewal, Pendulum Swing, Conspiracy, 0 (No identifiable plot):** These plots are present in Manual Coding but absent in both Gemini runs, indicating that coding has been performed everywhere, when the manual coder deemed the element to be missing.

2. **Hero's Journey, David vs. Goliath:** Significantly lower, or entirely absent in both Gemini runs compared to Manual Coding.
3. **Change-Is-Only an Illusion, Warning Tale, Restoration, Regulatory Enforcement, Triumph Over Adversity:** Manual Coding shows higher occurrences, indicating under-representation in Gemini runs.
4. **Story of Decline:** Although present in all, Manual Coding is substantially higher, showing that this dominant narrative plot is underrepresented by LLMs.

Low Scores Across All Metrics: For plot elements, Gemini Run 1 has an accuracy of 0.246 and Run 2 has 0.249, meaning that only about 24.6% and 24.9% of the codes matched the manual coding. Precision values are low at 0.231 for Run 1 and 0.239 for Run 2, indicating that only about 23% of the predicted codes were correct. Recall values are 0.146 for Run 1 and 0.154 for Run 2, reflecting that many true positives were missed, with only 14.6% and 15.4% correctly identified. F1 scores of 0.120 for Run 1 and 0.126 for Run 2 demonstrate poor balance between precision and recall. Negative Krippendorff's alpha values of -0.056 for Run 1 and -0.039 for Run 2 indicate that the agreement is worse than random chance, suggesting complete unreliability in the system's coding compared to the manual analysis.

Plot is also the component with the highest number of categories, some of which have subtle differences. This could be a reason that the LLM, even with its capability to mimic reasoning, fails to differentiate between nuances in text.

**Confusion Matrix:**

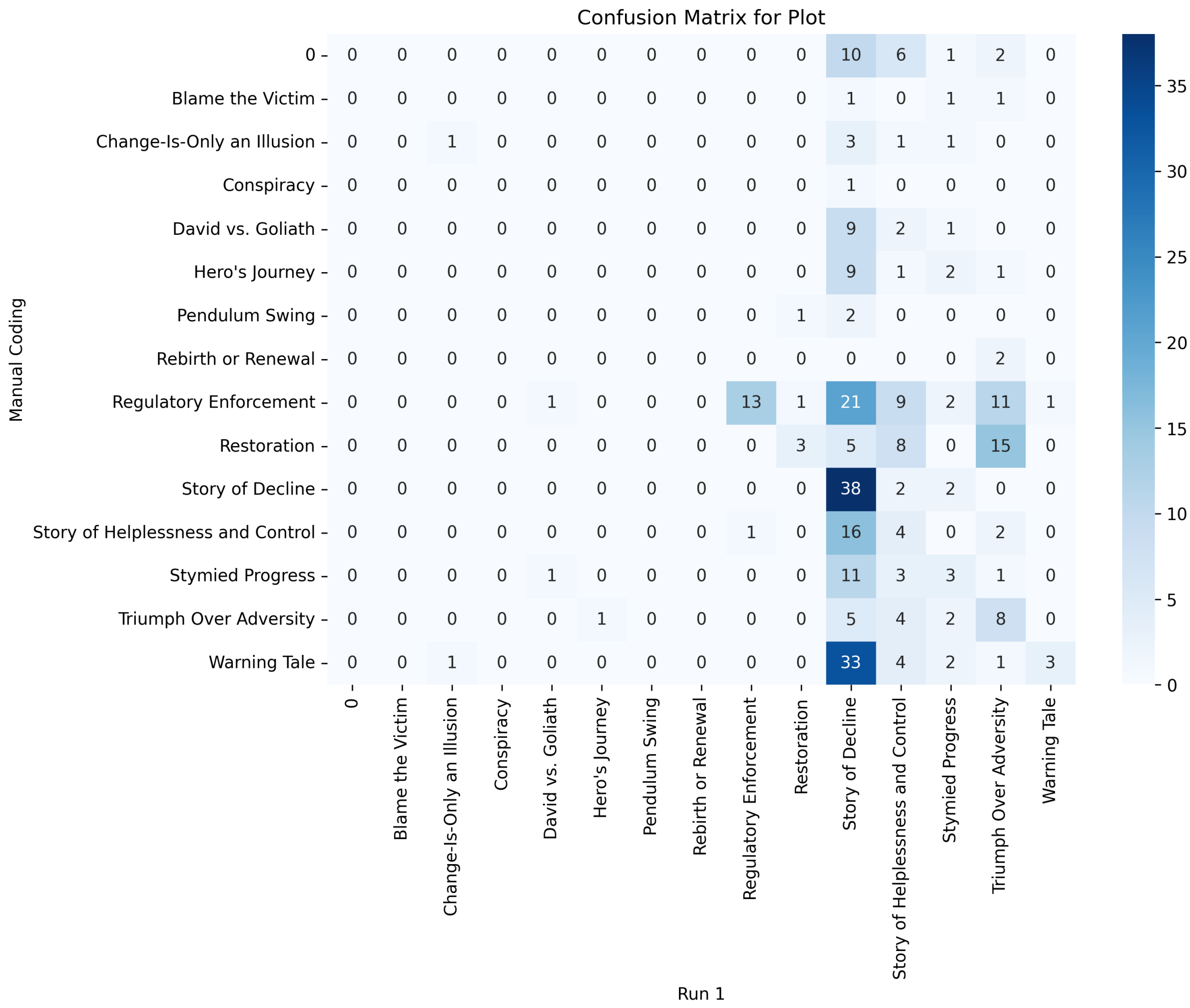



**Figure 6.4:** Instances of Coding: Manual vs Run 1

| Plot | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 19 | 278 |
| Blame the Victim | 0 | 0 | 3 | 294 |
| Change-Is-Only an Illusion | 1 | 1 | 5 | 290 |
| Conspiracy | 0 | 0 | 1 | 296 |
| David vs. Goliath | 0 | 2 | 12 | 283 |
| Hero's Journey | 0 | 1 | 13 | 283 |
| Pendulum Swing | 0 | 0 | 3 | 294 |
| Rebirth or Renewal | 0 | 0 | 2 | 295 |
| Regulatory Enforcement | 13 | 1 | 46 | 237 |
| Restoration | 3 | 2 | 28 | 264 |
| Story of Decline | 38 | 126 | 4 | 129 |
| Story of Helplessness and Control | 4 | 40 | 19 | 234 |
| Stymied Progress | 3 | 14 | 16 | 264 |
| Triumph Over Adversity | 8 | 36 | 12 | 241 |
| Warning Tale | 3 | 1 | 41 | 252 |
| Total | 73 | 224 | 224 | 3934 |

**Table 6.1:** Confusion Matrix - Manual Coding and Run 1

As illustrated in Figure 6.4, the LLM (Large Language Model) assigns the 'Story of Decline' plot significantly more frequently than any other plot category. Analysis of the confusion matrix for Run 1 (and Run 2 in subsection A.8.1) reveals that 'Story of Decline' exhibits a high incidence of false positives when compared to manual coding. Specifically, the LLM misclassifies instances identified manually as 'Regulatory Enforcement' 21 times, 'Warning Tale' 33 times, and 'Story of Helplessness and Control' 16 times as 'Story of Decline.'

Moreover, subsubsection 4.1.1.9 presents an example that was classified as a 'Story of Decline' by the LLM but is more accurately characterized as a 'Warning Tale,' according to manual coding. The LLM's rationale was: "Plot: The narrative follows a 'Story of Decline,' as it outlines the increasing problem of waste generation and its damaging consequences if left unaddressed." Nevertheless, further analysis of the data reveals instances where the LLM's coding may be considered more accurate than the manual coders. For instance, consider the following excerpt, which was manually coded as a 'Warning Tale':

> *"BAREILLY: The unprecedented level of air pollution enveloping National Capital Region (NCR) has now spread to western Uttar Pradesh reducing visibility to a low usually seen between December 25 and January 10, the peak winter days, due to low temperatures, humidity and absence of wind. On Sunday morning, poor visibility resulted in a pile-up on National Highway 24 involving eight vehicles. One person was killed and six persons were injured. The condition of one of the injured is stated to be serious.According to sources, owing to low visibility conditions, two trucks collided at about 7:30 am on Bilwa flyover in Bhojipura. This triggered a pile-up as vehicles going towards Lucknow began colliding into the first crash site, resulting in injuries to half-a-dozen commuters."We were on board a DCM truck carrying domestic goods to our village at Riccha, Bareilly when our vehicle suddenly collided with another vehicle on Bilwa flyover. There were several repeated impacts of collision, after which I fell unconscious," said Mohammed Yakub, 55. According to a witness, the collision between trucks was followed by a pile-up due to low visibility on the flyover. Some onlookers tried to make noise to deter incoming vehicles from crashing into the pile-up.The smog condition has made roads more vulnerable to accidents, with similar pile-ups being reported from other parts of NH 24. Around a dozen vehicles collided on NH 24 in Hapur, leaving another half a dozen injured. No fatalities have yet been reported from there.At least 12 people were injured in a pile-up of over 20 vehicles on Yamuna Expressway near Delhi on Thursday."* (Rai, 2016)

Upon further reflection, it appears that the 'Story of Decline' archetype is a more appropriate classification. There are instances in which the manual coder might reconsider their initial coding decision, acknowledging that manual coding is inherently susceptible to human error and subjectivity, which can vary over time. However, the LLM introduces a class imbalance by disproportionately coding for a single category, whereas manual coding exhibits a more evenly distributed categorization. This imbalance may reflect the LLM's suboptimal performance, potentially due to the inherent difficulty in distinguishing among multiple similar categories.

### 6.1.3. Moral

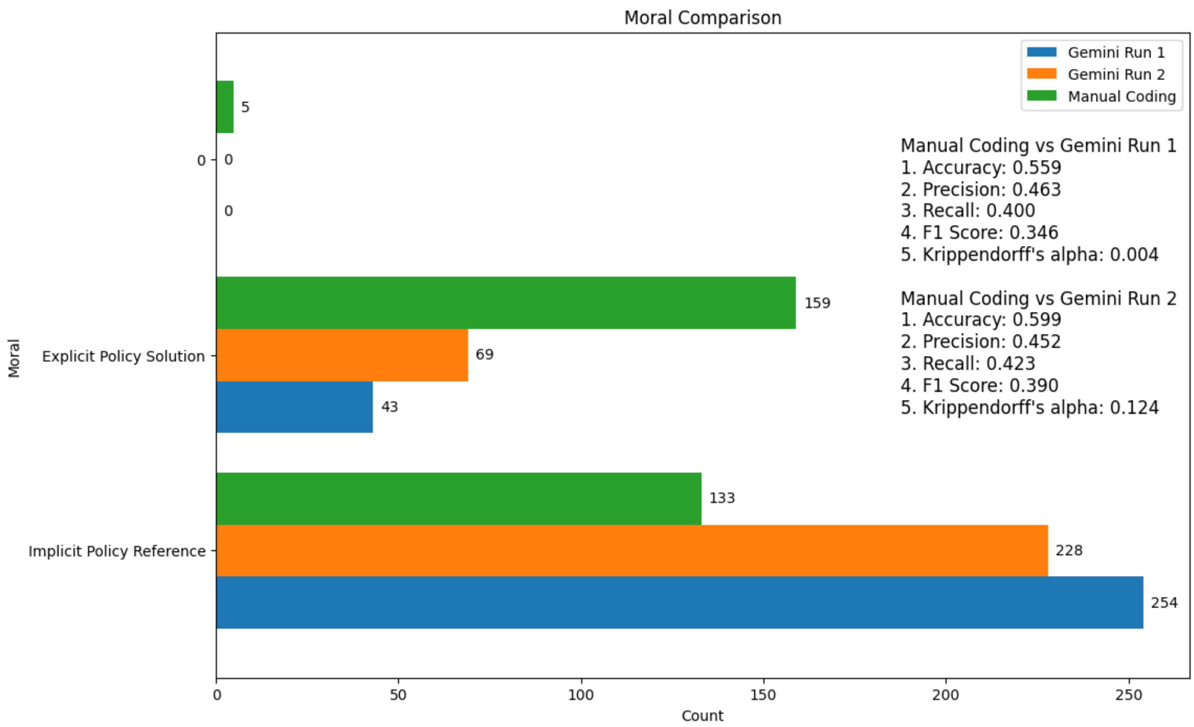



**Figure 6.5:** Comparison of Moral Frequencies

Key comparisons from the moral comparison:

- **Explicit Policy Solution**: The counts show a large discrepancy with Manual Coding at 159, Gemini Run 1 at 43, and Gemini Run 2 at 69.
- **Implicit Policy Reference**: There are notable differences, with Manual Coding at 133, Gemini Run 1 at 254, and Gemini Run 2 at 228.

For moral, Gemini Run 1 shows an accuracy of 0.559, and Run 2 shows 0.599, indicating that approximately 56% and 60% of the codes, respectively, matched manual coding, which is better than belief systems but still not high. Precision scores of 0.463 for Run 1 and 0.452 for Run 2 suggest that less than half of the predicted codes were correct. Recall values of 0.400 for Run 1 and 0.423 for Run 2 show that 40% and 42.3% of the actual positives were identified correctly, reflecting a moderate level of missed true positives. The F1 scores of 0.346 for Run 1 and 0.390 for Run 2 remain low. Krippendorff's alpha values of 0.004 for Run 1 and 0.124 for Run 2 imply almost no agreement beyond chance for Run 1 and slight agreement for Run 2. These results suggest that while Gemini runs are somewhat more effective in coding moral, their overall performance remains inadequate.

**Confusion Matrix:**

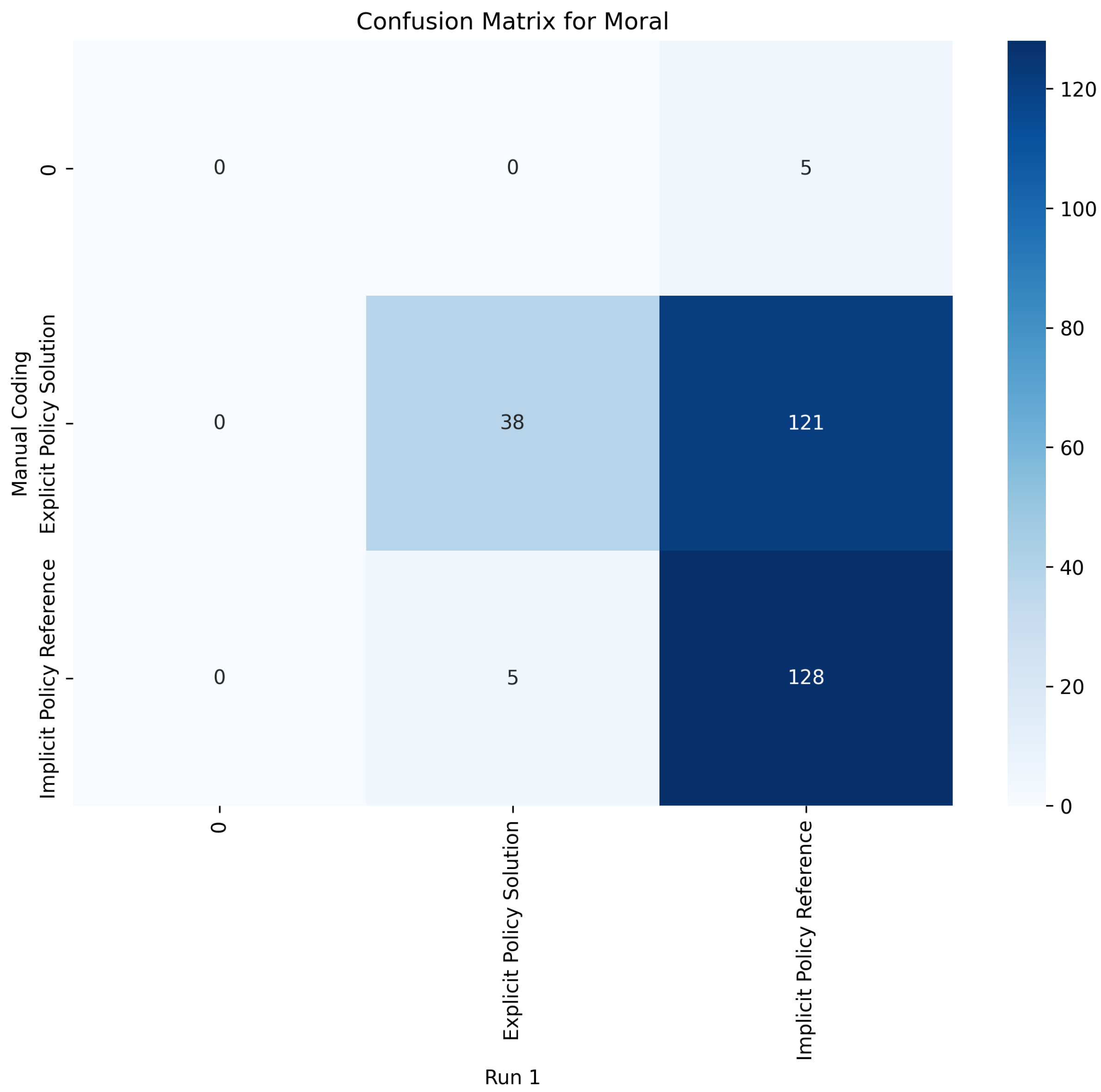



**Figure 6.6:** Instances of Coding: Manual vs Run 1

| Moral | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 5 | 292 |
| Explicit Policy Solution | 38 | 5 | 121 | 133 |
| Implicit Policy Reference | 128 | 126 | 5 | 38 |
| Total | 166 | 131 | 131 | 463 |

**Table 6.2:** Confusion Matrix Between Manual Coding and Run 1

The LLM identifies a significantly greater number of implicit policy references compared to explicit policy solutions, whereas manual coding presents a more balanced identification between the two categories. For example, in subsubsection 4.1.1.14 concerning the Graded Response Action Plan (GRAP), the LLM classifies the article as containing an implicit policy reference, justifying this with the rationale: “The implicit policy reference is the need for coordinated efforts, proactive action, and public awareness campaigns to effectively address air pollution”. This classification is made despite the article explicitly discussing the policy solution multiple times and questioning the delay in its implementation by the key stakeholder.

From the confusion matrix, it is evident that implicit policy references exhibit a large number of false positives, indicating that the LLM misclassified 107-126 articles (depending on the run) as 'Implicit Policy Reference' when they were manually coded otherwise.

In subsubsection 4.1.1.14, the article analyzed to have an implicit policy reference is identified by the LLM as containing an explicit policy solution with the reason: "The article implicitly suggests the need for waste segregation and recycling as a solution, as it highlights the lack of these practices in the current model". This interpretation addresses waste management as a policy problem, rather than air pollution, indicating a deviation from the issue at hand despite the prompt stressing the focus on air pollution as the policy problem. There are significantly fewer instances of the LLM coding for an 'Explicit Policy Solution' in disagreement with the manual analysis (38-57 cases); however, this also means that the LLM missed identifying an explicit policy solution in 102-121 cases that were recognized by the manual coder.

In instances where the moral is coded as '0' (not present), the articles typically center on topics unrelated to the primary policy issue. This is another area where the LLM tends to "lose focus". The LLM consistently assigns one of the non-zero codes to news articles, failing to recognize when the policy problem discussed in the narrative is unrelated to air pollution.

### 6.1.4. Belief Systems

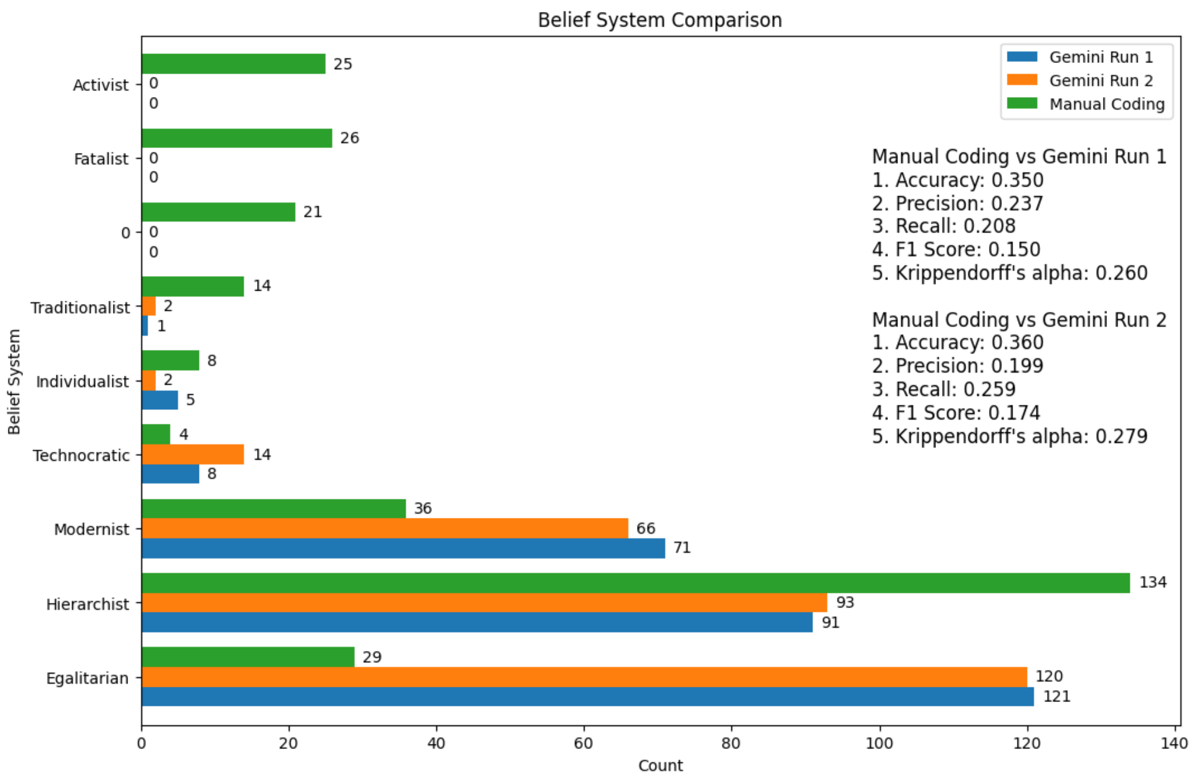



**Figure 6.7:** Comparison of Belief System Frequencies

Key comparisons from the belief system comparison include:

- **Activist**: The counts show a large discrepancy with Manual Coding at 25, Gemini Run 1 at 0, and

Gemini Run 2 at 0.

- **Fatalist**: Manual Coding shows a count of 26, while both Gemini runs have 0.
- **Traditionalist**: Manual Coding has 14, Gemini Run 1 has 1, and Gemini Run 2 has 2.
- **Individualist**: Manual Coding has 8, Gemini Run 1 has 5, and Gemini Run 2 has 2.
- **Technocratic**: Manual Coding shows 21, while Gemini Run 1 has 8 and Gemini Run 2 has 14.
- **Modernist**: Manual Coding has 36, Gemini Run 1 has 71, and Gemini Run 2 has 66.
- **Hierarchist**: Manual Coding shows 134, while Gemini Run 1 has 91 and Gemini Run 2 has 93.
- **Egalitarian**: Manual Coding shows 29, while both Gemini Run 1 and Gemini Run 2 have similar high counts of 121 and 120, respectively.

The accuracy for Gemini Run 1 is 0.350, and for Gemini Run 2, it is 0.360. This indicates that approximately 35 out of 100 codes from Gemini Run 1 and 36 out of 100 from Gemini Run 2 matched the manual coding, reflecting a modest level of correct identifications. Precision scores of 0.237 for Run 1 and 0.199 for Run 2 suggest that only about 23.7% and 19.9% of the predicted codes were correct, respectively, highlighting a high rate of false positives. Recall, at 0.208 for Run 1 and 0.259 for Run 2, shows that 20.8% and 25.9% of the actual positive codes were correctly identified, indicating that many true positives were missed. The F1 scores, which balance precision and recall, are 0.150 for Run 1 and 0.174 for Run 2, demonstrating a poor balance between the two. Krippendorff's alpha values of 0.260 for Run 1 and 0.279 for Run 2 indicate slight agreement beyond chance, which is minimal. Overall, these metrics show that the Gemini system struggles significantly to match manual coding for belief systems, with limited accuracy and reliability.

**Confusion Matrix:**

| Belief System | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 21 | 276 |
| Activist | 0 | 0 | 25 | 272 |
| Egalitarian | 23 | 98 | 6 | 170 |
| Fatalist | 0 | 0 | 26 | 271 |
| Hierarchist | 60 | 31 | 74 | 132 |
| Individualist | 0 | 5 | 8 | 284 |
| Modernist | 20 | 51 | 16 | 210 |
| Technocratic | 0 | 8 | 4 | 285 |
| Traditionalist | 1 | 0 | 13 | 283 |
| Total | 104 | 193 | 193 | 2183 |

**Table 6.3:** Confusion Matrix Between Manual Coding and Run 1

For the 'Hierarchist' belief system, there are 71 instances of false negatives, where the LLM failed to identify the articles as reflecting the 'Hierarchist' belief system, in contrast to manual coding. The LLM frequently misclassifies articles coded manually as 'Hierarchist' as 'Egalitarian' (46 instances in Run 1) and 'Modernist' (21 instances in Run 1). Similarly, the LLM often misclassifies articles as having an 'Egalitarian' belief system when they were manually coded as 'Fatalist' (16 instances in Run 1) and 'Activist' (23 instances in Run 1), as depicted in Figure 6.8.

In subsubsection 4.1.2.2, an article exemplifying the 'Hierarchist' belief system was identified by the LLM as having an 'Egalitarian' belief system, with the rationale: "The article aligns with an egalitarian

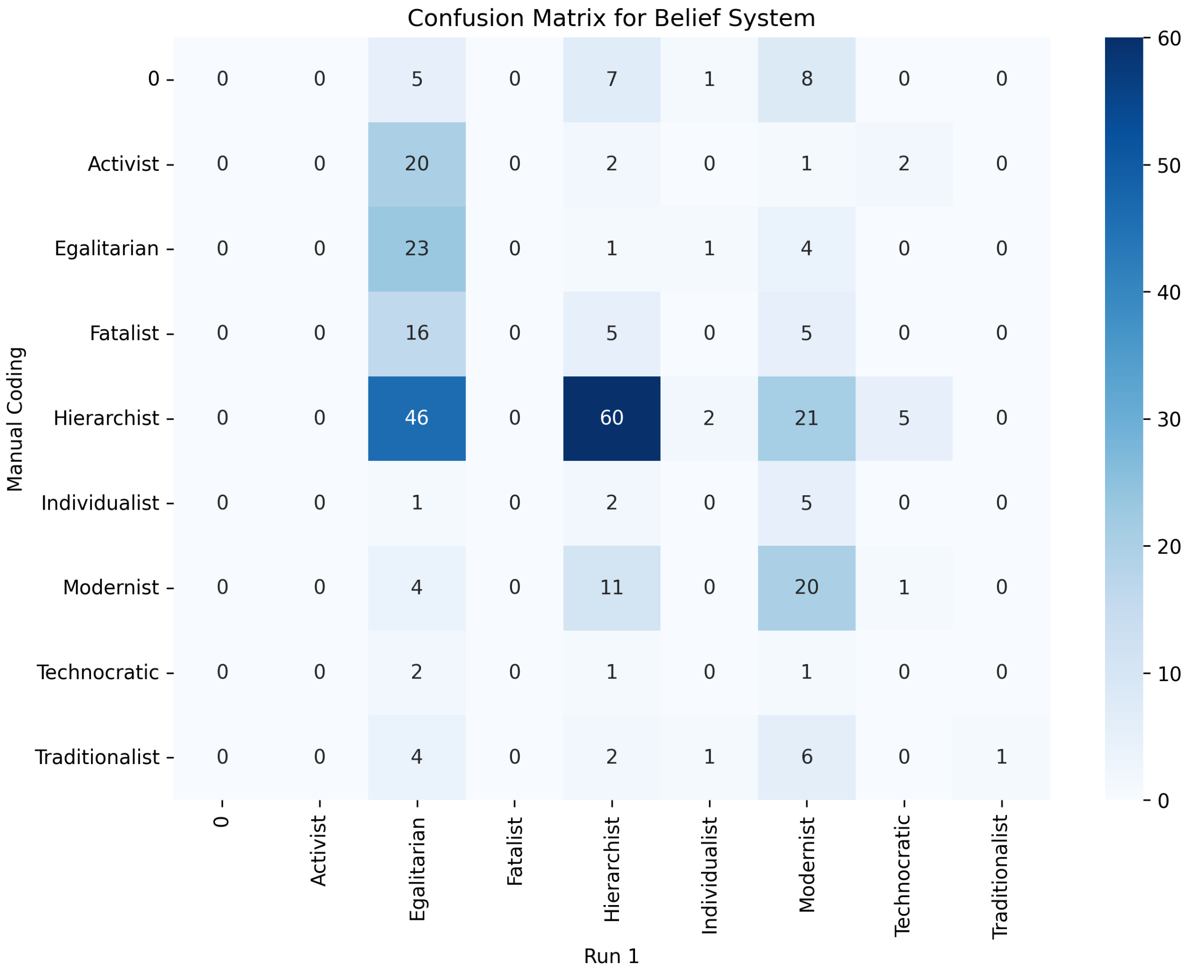



**Figure 6.8:** Instances of Coding: Manual vs Run 1

belief system, emphasizing the impacts of air pollution on public health, particularly for vulnerable groups like children".

> *"NEW DELHI: In the backdrop of severe air pollution, the odd-even scheme in the national capital will kick in from 8 am on Monday (November 4), with only even-numbered non-transport vehicles allowed on Delhi roads on the first day of the exercise. Chief minister Arvind Kejriwal appealed to people on Sunday to follow the rules for the sake of their children and the city. He has also asked the government machinery to ensure that no person faces inconvenience due to the restrictions. Hundreds of teams of Delhi Traffic Police and the transport and revenue departments have been deployed for a strict implementation of the scheme."*. (TOI, 2019a)

The narrative does not emphasize the impacts of air pollution; rather, it focuses on the government's implementation of state-led initiatives.

The article identified as 'Modernist' in subsubsection 4.1.2.2 was classified as 'Hierarchist' by the LLM, which justified this classification with the rationale: "The narrative emphasizes the role of the government and its agencies in tackling the issue, aligning with a Hierarchist belief system". This interpretation shifts the focus away from the stubble burning technology proposed. This raises the question of what the most important aspect of a narrative is when coding it. It is essential to consider that complex data such as this might have multiple plausible codes.

### 6.1.5. Narrative Strategies

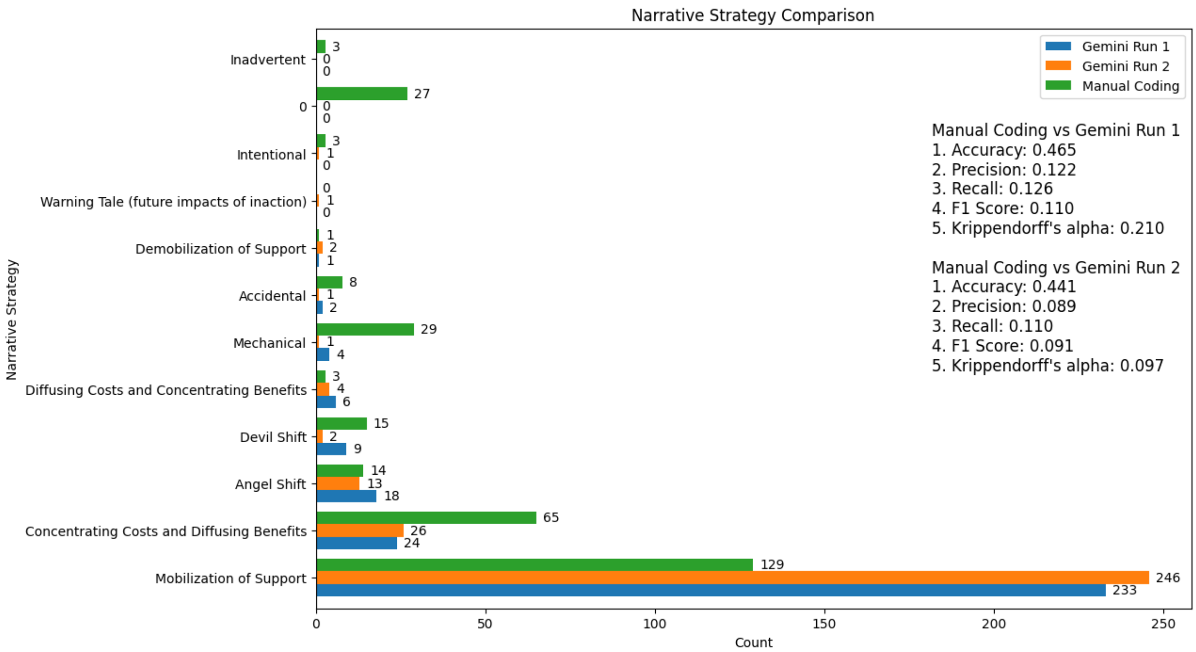



**Figure 6.9:** Comparison of Narrative Strategy Frequencies

Key comparisons from the narrative strategy comparison: There are notable differences throughout the graph, with Manual Coding far from both runs, and Gemini Run 1 and Gemini Run 2 closer to each other.

The accuracy for Gemini Run 1 is 0.465 and for Run 2, it is 0.441, meaning that about 46.5% and 44.1% of the codes matched the manual coding, which is slightly below half. Precision scores are very low at 0.122 for Run 1 and 0.089 for Run 2, indicating that only 12.2% and 8.9% of the predicted codes were correct, highlighting many false positives. Recall values of 0.126 for Run 1 and 0.110 for Run 2 show

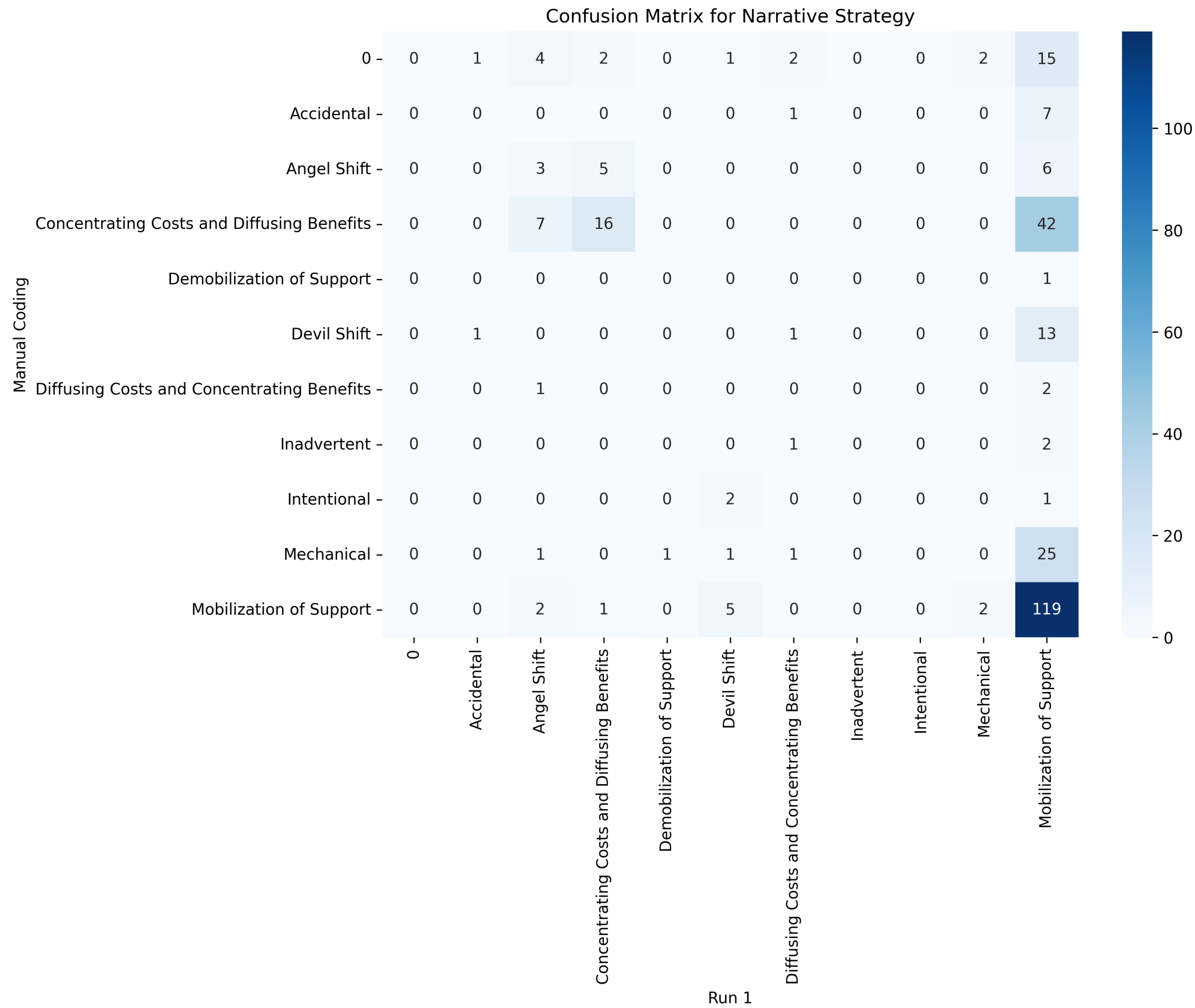



**Figure 6.10:** Instances of Coding: Manual vs Run 1

that a significant number of true positives were missed, with only 12.6% and 11% correctly identified. F1 scores of 0.110 for Run 1 and 0.091 for Run 2 reflect poor balance between precision and recall. Krippendorff's alpha values of 0.210 for Run 1 and 0.097 for Run 2 indicate slight to no agreement beyond chance. Collectively, these metrics reveal that the Gemini system faces significant challenges in accurately coding narrative strategies.

**Confusion Matrix:**

From the confusion matrices, it is evident that the strategy 'Mobilization of Support' has been predominantly coded for most of the articles. The LLM (Gemini) misclassifies 25 articles manually identified as 'Mechanical,' 13 articles as 'Devil Shift,' 42 articles as 'Concentrating Costs and Diffusing Benefits,' and 15 articles as '0' under the category 'Mobilization of Support' for Run 1. Similar misclassification numbers are observed for Run 2, resulting in a high number (114-126) of false positives for this category (subsection A.8.4, Figure 6.10, Figure A.7).

Although 'Mobilization of Support' is the most frequently coded strategy in manual coding as well, there is a substantial presence of other strategies in the distribution. In the example provided in subsubsection 4.1.2.7 for 'Mobilization of Support,' there is concordance between the manual coding and LLM coding. However, an example of disagreement is as follows:

*"PATIALA: The Punjab Pollution Control Board has decided to periodically measure the pollution lev-*

| Narrative Strategy | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 27 | 270 |
| Accidental | 0 | 2 | 8 | 287 |
| Angel Shift | 3 | 15 | 11 | 268 |
| Concentrating Costs and Diffusing Benefits | 16 | 8 | 49 | 224 |
| Demobilization of Support | 0 | 1 | 1 | 295 |
| Devil Shift | 0 | 9 | 15 | 273 |
| Diffusing Costs and Concentrating Benefits | 0 | 6 | 3 | 288 |
| Inadvertent | 0 | 0 | 3 | 294 |
| Intentional | 0 | 0 | 3 | 294 |
| Mechanical | 0 | 4 | 29 | 264 |
| Mobilization of Support | 119 | 114 | 10 | 54 |
| Total | 138 | 159 | 159 | 2811 |

**Table 6.4:** Confusion Matrix Between Manual Coding and Run 1

> *els at toll plazas during peak hours. Besides, asking project directors of privately managed National Highways and State Highways in the state to initiate measures as per their concession agreements to control dust pollution on the roads. KS Pannu, Chairman said the PPCB has asked the companies to ensure that proper plantation in the central verge of the roads and on both sides of the roads is done and maintained. He said it has observed that although the agreement of the road constructing companies includes developing a green cover on the central verge and the sides of the roads, the companies were found lacking in undertaking the job. The PPCB chairman said that the board officials have recently submitted a report in which dust being thrown in the air by vehicles using highways was found to be a major pollutant, especially during the winter season. He said the board has also asked the private companies to make certain that measures to collect dust from roads was put in place. The non-concreate paths between the over bridge and the service lanes was another major dust spots which were not being properly maintained by the road constructing companies resulting in air pollution due to dust. Pannu said PPCB would keep a constant watch on to ensure the directions given to the highway authorities were followed in letter and spirit."* (Sirhindi, 2018)

In the aforementioned article, the outcome of manual coding for narrative strategy was 'Concentrating Costs and Diffusing Benefits,' as the Punjab Pollution Control Board has required the private companies managing the highways to implement dust control measures. This approach concentrates the costs on these companies while spreading the benefits to the local population and highway users. However, the LLM coded this as 'Mobilization of Support,' justifying this with the rationale: "The article utilizes the "Mobilization of Support" strategy by highlighting the severity of the issue and presenting the PPCB as taking action, thereby encouraging public support for the agency's efforts". This classification, in the coder's judgment, is less appropriate. This issue recurs frequently, as many policy narratives are complex, involving multiple stakeholders, and may be relevant to more than one code, and in this case, multiple narrative strategies may be at work here.

## 6.2. Accuracy, Precision, Recall, F1 Score, and Krippendorff's Alpha

To reiterate, the following table outlines the scores on each of the metrics evaluated for. Any score above 0.7[2] for Accuracy, Precision, Recall, and F1 has been highlighted in Green, along with all scores above 0.8[3] for Krippendorff's $\alpha$.

| Category | Accuracy | Precision | Recall | F1 Score | Krippendorff's Alpha |
|---|---|---|---|---|---|
| **Manual Coding and Gemini Flash Run 1** | | | | | |
| Plot | 0.246 | 0.230 | 0.146 | 0.120 | -0.056 |
| Moral | 0.559 | 0.463 | 0.400 | 0.346 | 0.004 |
| Belief System | 0.350 | 0.237 | 0.208 | 0.150 | 0.260 |
| Narrative Strategy | 0.455 | 0.091 | 0.090 | 0.078 | 0.213 |
| **Manual Coding and Gemini Flash Run 2** | | | | | |
| Plot | 0.249 | 0.239 | 0.154 | 0.126 | -0.039 |
| Moral | 0.599 | 0.452 | 0.423 | 0.390 | 0.124 |
| Belief System | 0.360 | 0.199 | 0.259 | 0.174 | 0.279 |
| Narrative Strategy | 0.428 | 0.054 | 0.074 | 0.057 | 0.097 |
| **Gemini Flash Run 1 and Gemini Flash Run 2** | | | | | |
| Plot | 0.781 | 0.487 | 0.507 | 0.490 | 0.582 |
| Moral | 0.852 | 0.727 | 0.826 | 0.758 | 0.516 |
| Belief System | 0.879 | 0.763 | 0.782 | 0.731 | 0.818 |
| Narrative Strategy | 0.801 | 0.438 | 0.326 | 0.325 | 0.508 |

**Table 6.5:** Results for Gemini Run 1 and 2

| Column | Manual Coding and Run 1 | Manual Coding and Run 2 | Run 1 and Run 2 |
|---|---|---|---|
| Hero | 0.25 | 0.25 | 0.28 |
| Villain | 0.18 | 0.18 | 0.24 |
| Victim | 0.16 | 0.15 | 0.34 |
| Beneficiary | 0.19 | 0.18 | 0.28 |
| Ally | 0.22 | 0.22 | 0.20 |
| Opponent | 0.64 | 0.63 | 0.19 |
| Ineffective | 0.62 | 0.62 | 0.21 |
| Setting | 0.39 | 0.39 | 0.46 |

**Table 6.6:** Average Cosine Similarity Between Characters

### 6.2.1. Definitions

**Accuracy:** The proportion of correct predictions (both true positives and true negatives) out of all predictions.

**Precision:** The ratio of true positives to the sum of true positives and false positives. It measures the accuracy of positive predictions.

**Recall:** The ratio of true positives to the sum of true positives and false negatives. It measures the ability to find all positive instances.

**F1 Score:** The harmonic mean of precision and recall. It balances the two metrics and is useful when you need a single measure of a model's performance.

**Krippendorff's Alpha:** Krippendorff's alpha is a statistical measure of the reliability of agreement between observers or raters. It generalizes several known reliability indices and is applicable

[2]Threshold used in other similar studies concerning annotation using large language models (Pangakis et al., 2023), (Gilardi, Alizadeh, & Kubli, 2023), (Ding et al., 2023)

[3]As commonly used in social science research (Krippendorff, 2013).

to various data types (nominal, ordinal, interval, and ratio).

### 6.2.2. Result Review

Manual Coding and Gemini Flash Run 1 & 2

- **Plot:** The Gemini Flash model demonstrates significant difficulties in accurately identifying plot elements when compared to manual coding. In Run 1, the model achieved an Accuracy of 0.246, which marginally increased to 0.249 in Run 2. Precision improved slightly from 0.230 to 0.239, while Recall moved from 0.146 to 0.154. The F1 Score, an aggregate measure of Precision and Recall, showed a minor improvement from 0.120 in Run 1 to 0.126 in Run 2. Notably, Krippendorff's Alpha remained negative, shifting from -0.056 to -0.039, indicating a lack of agreement and reliability in plot identification.
- **Moral:** The model's performance in identifying moral elements displayed moderate accuracy and agreement. Accuracy remained the same with 0.559 in Run 1 and 0.599 in Run 2. However, Precision decreased slightly from 0.463 to 0.452, while Recall increased from 0.400 to 0.423. The F1 Score improved from 0.346 to 0.390. Krippendorff's Alpha showed a significant increase from 0.004 to 0.124, reflecting better agreement in Run 2.
- **Belief System:** For belief system identification, the Gemini Flash model exhibited moderate performance. Accuracy increased from 0.350 in Run 1 to 0.360 in Run 2. However, Precision decreased from 0.237 to 0.199, whereas Recall improved from 0.208 to 0.259. The F1 Score improved from 0.150 to 0.174. Krippendorff's Alpha, reflecting the reliability of the model, increased slightly from 0.260 to 0.279, indicating insignificant agreement.
- **Narrative Strategy:** The models identification of narrative strategies showed moderate accuracy but low precision and recall. Accuracy slightly decreased from 0.455 in Run 1 to 0.428 in Run 2. Precision dropped from 0.091 to 0.054, and Recall decreased from 0.090 to 0.074. Consequently, the F1 Score declined from 0.078 to 0.057. Krippendorff's Alpha also decreased from 0.213 to 0.097, indicating reduced reliability.

Gemini Flash Run 1 vs. Run 2

- **Plot:** The model demonstrated substantial consistency in plot identification between runs, with an Accuracy of 0.781. Precision was 0.487, and Recall was 0.507, resulting in an F1 Score of 0.490. Krippendorff's Alpha was 0.582, indicating better agreement between the two runs compared to the agreement between manual coding and the runs.
- **Moral:** The model achieved high accuracy and strong agreement for moral elements, with an Accuracy of 0.852. Precision was 0.727, and Recall was 0.826, resulting in an F1 Score of 0.758. Krippendorff's Alpha was 0.516, reflecting substantial agreement between runs, but lower than the threshold of 0.8 commonly used in social science studies (Krippendorff, 2013).
- **Belief System:** High accuracy and strong agreement were also noted for belief system identification, with an Accuracy of 0.879. Precision was 0.763, and Recall was 0.782, yielding an F1 Score of 0.731. Krippendorff's Alpha was 0.818, indicating high consistency between runs.
- **Narrative Strategy:** In narrative strategy identification, the model exhibited high accuracy with moderate precision and recall. Accuracy was 0.801, Precision was 0.438, and Recall was 0.326, resulting in an F1 Score of 0.325. Krippendorff's Alpha was 0.508, indicating reliable performance.

Average Cosine Similarity Between Characters

- **Hero, Villain, Victim, Beneficiary, Ally:** The similarity scores are relatively low (0.15-0.28), indicating moderate alignment between manual coding and model runs, and between the two model runs. This reflects some consistency in character identification but suggests room for improvement.

- **Opponent and Ineffective:** Both categories show high similarity between manual coding and the runs (0.62-0.64), but a substantial drop in similarity between Run 1 and Run 2 (0.19-0.21). This inconsistency between model runs suggests variability in model performance for these characters.
- **Setting:** The similarity is moderate (0.39-0.46), indicating reasonable consistency in identifying the setting across manual coding and model runs, and between the two model runs.

### 6.2.3. Summary

- **Low-Scoring Categories:** Hero, Villain, Victim, Beneficiary, Ally
  - Both Gemini runs exhibit struggles with these categories, indicated by low cosine similarity.
- **Moderate-Scoring Categories:** Moral, Setting, Plot, Narrative Strategy, Opponent, Ineffective.
  - These categories demonstrate moderate performance, with slightly higher Krippendorff's alpha values in combined runs.
- **Higher-Scoring Categories:** Belief System.
  - This category is identified with higher accuracy and reliability by both runs, though there are still disagreements with the Manual Coding. The agreement within runs is satisfactory, as shown by acceptable Krippendorff's alpha value.

# 7

# Discussion

## 7.1. Overview

The main objective of this study was to address the research question: "To what extent do Large Language Models (LLMs) accurately automate qualitative coding of policy narratives when compared to a manually coded dataset?" To answer it, a step-by-step analysis was performed using the methodology outlined in chapter 3. This chapter will highlight the findings of this study, discuss those findings, and acknowledge its limitations.

### 7.1.1. Findings

From the comparison of the chosen LLM and manual coding, it is evident that the LLM does not achieve significant agreement with manual coding in this multi-class classification task. The lowest accuracy was 0.246 for the 'Plot' element, indicating that out of 100 instances, only 24 matched with the manually assigned code when coded by the LLM. The highest accuracy was for 'Moral' at 59.9%. 'Narrative Strategy' and 'Belief System' have similar scores, 42.8% - 45.5%, and 35% - 36% respectively (run 1% - run 2%). Other metrics such as precision, recall, F1 score, and Krippendorff's $\alpha$ also reveal suboptimal performance, with scores as low as 5.4% for Precision, 7.4% for Recall, and 5.7% for F1 for the 'Narrative Strategy' component. Krippendorff's $\alpha$ ranges from -0.056 for 'Plot' in run 1 to 0.279 for 'Belief System' in run 2. Overall, 'Moral' (in run 2) has the best performance across all used metrics, possibly due to the lowest number of possible classes in its output. In other studies, LLMs have been shown to be particularly strong in tasks with fewer classes (Gilardi et al., 2023). 'Belief System' has the second best performance, with the highest $\alpha$ compared to other categories. The output demonstrated class imbalance, affecting the scores (Jeni, Cohn, & De La Torre, 2013), (Feng, 2014) by causing skewed agreement[1], reduced sensitivity to minority classes[2], bias in expected agreement[3], and an impact on

[1] In the presence of class imbalance, where one class is much more prevalent than others, raters may show high agreement simply by frequently choosing the majority class. This can inflate the observed agreement, potentially leading to a misleadingly high Krippendorff's alpha.

[2] If raters rarely encounter instances of the minority class, they may be less reliable in identifying and agreeing on these instances. This can lead to lower agreement on the minority class, which, although it might not drastically change the overall alpha due to the small number of such instances, can still reflect poor reliability in important, albeit less frequent, categories.

[3] Krippendorff's alpha compares observed agreement to expected agreement by chance. In imbalanced datasets, the expected agreement by chance will be skewed towards the majority class, potentially underestimating the difficulty of achieving agreement on minority classes.

interpretation[4].

Although research in LLM text annotation is limited, there is agreement between the findings in this study and previous studies. Pangakis et al. have shown that LLMs show promising results with a median accuracy of 0.850 and a median F1 score of 0.707 across 27 annotation tasks from 11 datasets, but there is significant variability in performance, with some tasks showing poor precision or recall. This variability underscores the need for task-specific validation and careful assessment before deploying LLMs for annotation tasks.

The LLM shows high repeatability in its coding, with best performance in the 'Belief System' category, with a Krippendorff's $\alpha$ greater than 0.8, the generally accepted metric. Gilardi et al. also show that the LLM in their study (ChatGPT) maintains high performance across various tasks, including relevance, stance, topic, and frame detection. Its performance is particularly strong in tasks with fewer classes (as mentioned before) and higher intercoder agreement among trained annotators on the same task (Gilardi et al., 2023). In another study, Nasution and Onan found that LLMs annotate content at a human level quality in tasks such as sentiment analysis, but human annotations by trained annotaters or experts outperform LLMs in more intricate and complex NLP tasks, especially with "understanding context and addressing ambiguity" (Nasution & Onan, 2024). LLM annotations were still found to be of higher quality than Amazon's crowdsourcing Mechanical Turk service (Nasution & Onan, 2024). It is also possible to increase the repeatability by varying the temperature parameter (Table A.2), but it does not give us better agreement with manual coding. In fact, Renze and Guven found that for multiple choice problem solving tasks, varying the LLM temperature from 0 to 1 does not show a significant impact on LLM performance, as is reproduced in Table A.2 as well. X. Wang et al. has shown that in complex tasks that benefit from diverse reasoning paths, the LLM performance can be made better by using outputs obtained from different temperature settings and using a vote-in mechanism to choose the most appropriate response, as if simulating an artificial "Wisdom of Crowds".

Several issues persist with NLP techniques used in qualitative content annotation, such as oversimplification, bias, and the effect of data quality on the output (Jin & Mihalcea, 2022; Zhou et al., 2022). These issues continue to affect the process when using LLMs. For instance, bias and subjectivity become apparent when the identified component is present in the example. For instance, smog towers were frequently identified as "ineffective" due to the specific example of the "Ineffective" character being about them (section A.3). Additionally, the coding from the LLM showed class imbalances, whereas the manual coding exhibited a more even distribution. This suggests that LLMs tend to favor one category over another, possibly due to difficulty in discerning nuances between categories, similar to limitations observed in traditional NLP techniques (Jin & Mihalcea, 2022; Zhou et al., 2022). In another study with GPT3.5 and Meta's LLAMA used for short answer scoring, it was found that for complex tasks, even well-framed questions may not suffice, indicating the limitations of current LLMs in handling intricate reasoning or domain-specific content (Chamieh, Zesch, & Giebermann, 2024). Plots such as 'Story of Decline'[5] and 'Warning Tale'[6] have definitions that would cause a coder to get confused about which category an article belongs to, suggesting a possibility of codebook redesign. Pangakis et al. found that refining the codebook can lead to modest improvements in accuracy and F1 scores. This iterative refinement[7] helps in addressing systematic misclassifications by the LLM.

[4]A high Krippendorff's alpha in an imbalanced dataset may not indicate true reliability across all classes. It may reflect the ease of agreement on the majority class while masking poor reliability on minority classes. Conversely, a low alpha might better indicate genuine issues with inter-rater reliability across all classes.

[5]Story of Decline: This plot describes an initial state of well-being that deteriorates over time, highlighting the urgent need for action. It may start with a good situation that worsens or begin at a point where things are already dire.

[6]Warning Tale: This plot serves as a cautionary story about the dire consequences of inaction or improper actions, often projecting a bleak future to motivate current action.

[7]Workflow proposed for iterative codebook refinement - section A.9.

The design of the codebook also influences the qualitative analysis especially if performed by an individual, potentially constraining insights into predefined categories subscribing to the individual's personal biases, which could be a reason for class imbalance.

Human perception of decision problems and the evaluation of probabilities and outcomes are governed by certain psychological principles, which can lead to predictable changes in preference depending on how the problem is framed (Tversky & Kahneman, 1981). When the same problem is framed in different ways, people's preferences can shift predictably; this phenomenon is known as the "framing effect" (Tversky & Kahneman, 1981). For example, a person might make a different choice if a situation is framed in terms of potential gains versus potential losses, even if the actual outcomes are equivalent (Tversky & Kahneman, 1981).

There may be labels appropriate for the content that do not exist in the codebook, thereby limiting the analysis. This could also influence the output to slide towards a particular outcome, due to loaded language[8] or leading examples[9], or from framing effects, leading to class imbalances, making codebook design all the more important. Codebook quality becomes better and the codebook becomes less biased as more people are involved in its design (Fonteyn, Vettese, Lancaster, & Bauer-Wu, 2008). Since the codebook was completely designed by one person in this study, there is scope for improvement here

Finally, the presence of non-standard English in the narratives, such as 'Hinglish' in India, poses challenges. It is an adaptation of English where Hindi words are written in the Latin Alphabet instead of the 'Devanagari' script. Words like *puja* (prayer), *taluk* (administrative district), *pradhan mantri yojana* (prime minister's scheme), and *tandrust* (strong, healthy) are terms present in the data that the LLM might not interpret correctly in context. Although a brief interaction with ChatGPT revealed that it understands these words[10], the context in complex tasks like qualitative coding could still pose difficulties.

As LLMs are general purpose tools, the study undertaken here can be performed in various other contexts. Performance on simpler tasks such as sentiment analysis or topic modelling using LLMs will likely deliver superior results, but for more complex tasks such as the one in this study, human validation would still be necessary.

### 7.1.2. Air Pollution in India

This analysis also revealed information about the policy problem itself. The heroes in these policy narratives have emerged as medical doctors, as identified through both manual and automated coding processes. Doctors have played a crucial role in raising awareness about the adverse effects of air pollution on public health and in educating individuals on managing these risks. Additionally, private citizens and residents of heavily affected areas have been recognized for their efforts. These individuals have actively engaged in protests, filed public interest litigations, and promoted awareness initiatives.

Moreover, various actors in positions of authority, including ministers, pollution control boards, and municipal corporations, oscillate between being perceived as heroes and villains, contingent upon

[8] Choice of words and sentence framing such that an emotional response is evoked in the reader or listener (Macagno & Walton, 2014).

[9] In law, a leading question is one that contains information that the questioner is looking to get confirmed, or is framing the question in a way that evokes a certain answer or 'leads' the other person there (Melilli, 2003). For example, “Have you stopped stealing from your brother?” is a question implying that the person has a brother, and has stolen from him at some point.

[10] User: What is the meaning of tandrust?
ChatGPT: The term "tandrust" is a word from Urdu, which means "healthy" or "in good health." It is often used to describe a person who is physically fit and free from illness. The word can be used in various contexts to refer to overall well-being and soundness of health.

media narratives. Entities such as farmers/stubble burning, coal-powered plants, brick kilns, vehicles, firecrackers, and construction activities are frequently portrayed as the primary contributors to the problem, often cast as villains in these narratives. In another study, the sources of air pollution in India have been found to include both anthropogenic and natural origins, such as industrial emissions, vehicular exhaust, agricultural burning practices, and natural events like dust storms and wildfires (Pratap Choudhary & Garg, 2013), corroborating this finding.

The detrimental impact of these so-called villains predominantly affects the victims, which include the environment, local residents (particularly vulnerable populations such as the elderly), children, individuals with cardiovascular and respiratory conditions, pregnant women, wildlife, as well as law enforcement officers and laborers. It has been shown in medical studies that that air pollution impacts the respiratory, cardiovascular, and neurological health of the population (Genc, Zadeoglulari, Fuss, & Genc, 2012; Bourdrel, Bind, Béjot, Morel, & Argacha, 2017; Glencross, Ho, Camina, Hawrylowicz, & Pfeffer, 2020)

When the heroes succeed, the primary beneficiaries are the general populace, especially in rural areas. Farmers also benefit, often arguing that they are depicted as villains due to the lack of viable alternatives to stubble burning. The environment, including numerous indigenous, endangered, rare, and ancient flora and fauna, also gains from the heroes' actions. Government bodies, pollution control boards, and institutions dedicated to research and technology frequently support the heroes by providing financial assistance, technological resources, or policy development.

However, opponents to the heroes' efforts include directors, officers, ministers, private companies, and worker unions, whose economic or operational interests are threatened by reduced air pollution, given the current lack of sustainable energy alternatives or methodological substitutes. Political agenda could also be driving this opposition. India is perceived to be the 93rd least corrupt country in the world (Transparency International, 2023), implying the plausibility of the people in power opposing air cleaning initiatives in exchange for goods and services. In this narrative, entities such as cracker guidelines (restrictions on the use of firecrackers during festivals), political parties (pledging to reduce air pollution during elections), bans and policies, and government interventions have been identified as ineffective, failing to make a measurable impact in this ongoing struggle.

The "Plot", as identified through manual analysis, predominantly focuses on themes of "Regulatory Enforcement", "Warning Tales", and a "Story of Decline". Automated coding corroborates the prevalence of the "Story of Decline" within the narrative, followed by a "Story of Helplessness and Control". This depiction underscores the dire state of air pollution in India, simultaneously indicating a movement towards heightened mass awareness regarding the issue, which is a crucial preliminary step towards developing effective policy responses (Zeng et al., 2019).

More than half (159 out of 297) of the sampled narratives explicitly propose policy solutions to the identified policy problem, as determined through manual coding. However, the language model (LLM) suggests that the majority of references are "Implicit Policy References". In this context, manual coding is deemed more accurate as the LLM seemingly lacks the capability to accurately distinguish "Explicit Policy Solutions" within the text. This discrepancy highlights that while policy solutions are being proposed, further efforts are needed to advance these proposals.

Gulia et al. discovered that air pollution studies in India have been largely limited to the Indo-Gangetic plain and metro cities, while a large part of South, Central, and Eastern India has been neglected, the settings for the narratives are spread throughout India, showing that it is a country-wide problem, and warrant a country-wide research effort.

A "Hierarchist" belief system predominantly characterizes the narratives, according to manual coding. This result also corroborates the hierarchical nature of Indian society (Meyer, 2014).

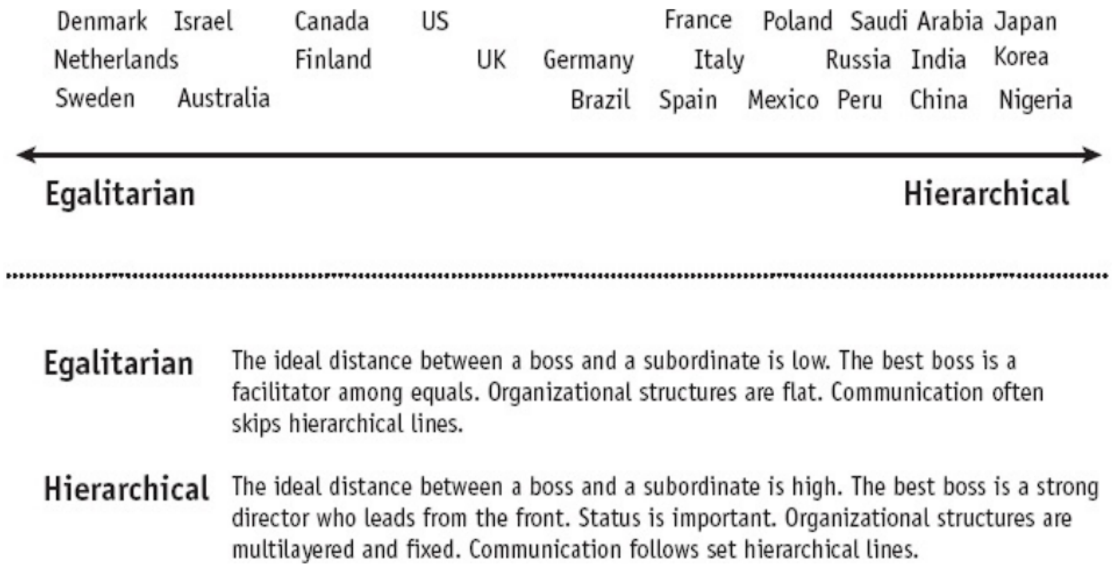



**Figure 7.1:** The Culture Map (Meyer, 2014)

Conversely, the LLM indicates that the narratives are largely "Egalitarian". This divergence suggests that multiple interpretative labels may be applicable. Given India's socialist foundation (India, 1950) and hierarchical societal structure (Meyer, 2014), this dual characterization is plausible. It reflects that while many narratives advocate for a systematic approach to the policy problem, grassroots efforts are also significant, leveraging community power and collective action. Additionally, "Modernist" belief systems are prominently featured, emphasizing the role of technology, whether domestically developed or imported, in addressing air pollution.

The narrative strategy shows consistency between manual and automated coding, both identifying the "Mobilization of Support" strategy as dominant. This alignment suggests that the media plays a critical role in this context by raising awareness about air pollution and mobilizing support, thereby appealing to authorities and policymakers to address the issue.

The policy problem of air pollution has taken form as India develops, uses more energy, develops infrastructure and technology, all of which are contributors to air pollution. The first step of raising awareness is already underway, with the time to form and implement commensurate policy fast approaching.

### 7.1.3. Interpretation of Findings

The results indicate that LLMs, in their current form, perform poorly in qualitative coding tasks compared to manual coding, particularly in identifying plot elements and narrative strategies. However, they show moderate performance in identifying moral elements and belief systems. The consistency between runs without hyperparameter tuning is a notable strength, suggesting potential for improved performance with further optimization.

Previous studies have highlighted limitations in NLP techniques, such as bias, subjectivity, and data quality issues (Jin & Mihalcea, 2022; Zhou et al., 2022). This study confirms that these limitations persist with LLMs. However, it also demonstrates that LLMs can achieve consistent results across multiple runs, which is an advantage over human coders whose coding behavior can vary over time (Belur, Tompson, Thornton, & Simon, 2021). In other automated text annotation studies, LLMs have been shown to be 30 times cheaper than Amazon's mechanical Turk with 25% greater accuracy on tasks such as judging the relevance of tweets or articles to specific topics, stance detection regarding Section 230 legislation, topic detection for tweets, frame detection for content moderation as a problem or solution, and policy frame detection for various policy-related aspects. (Gilardi et al., 2023).

The use of LLMs in this context is novel, and this study is among the first to conduct a comprehensive Narrative Policy Framework (NPF) analysis of news articles on air pollution in India using LLMs. Previous approaches have relied on traditional NLP techniques such as topic modelling and named entity recognition, which are less effective in handling nuanced and complex qualitative tasks such as multi-class classification (Peña et al., 2023) and recognition of named entities (Fritzler, Logacheva, & Kretov, 2019) by definition instead of keywords, especially when zero-shot performance becomes important due to lack of training examples.

### 7.1.4. Contribution

Although the primary contribution of this study is methodological, it also provides significant insights into the air pollution policy challenges in India. A comprehensive dataset comprising 14,578 newspaper articles on air pollution in India, spanning the years 2010 to 2014, has been made available in a structured format conducive to analysis. For transparency, the study has made available the coded datasets, and released open-source code for the complete analytical process, including pre-processing of all articles, analysis using a large language model (accessible via APIs or open-source with the potential for retrieval-augmented generation), post-processing for visualization, and calculation of relevant metrics. This would facilitate the replication of the study by future researchers in an efficient manner and enables enhancements to the methodology as deemed necessary.

The study tests a novel method (Large Language Models) to automate qualitative content analysis of policy narratives using the Narrative Policy Framework, to increase the speed with which studies can be carried out, and to increase the size of the data annotated for research. It details the parts where the method performs satisfactorily, where it fails to reach generally accepted standards, and what can be improved for future studies using this methodology. The entire pipeline has been made open-source for researchers to use and build on, and a preliminary codebook has been made available to refine and modify, to analyse narratives on air pollution.

### 7.1.5. Limitations of the Study

This study has some limitations which should be kept in mind while perusing the results:

- The manual coding was performed by a single coder who is a graduate student, with no previous qualitative coding experience, and without inter-coder reliability checks. The dataset was not annotated by experienced coders, which may have affected the quality of the manual coding.
- The study could have benefitted from further refinement of the codebook to improve its quality, to judge repeatability in coding with the same coder and make category distinctions clearer, which was not possible due to time constraints.
- The LLM used in the study was not fine-tuned or optimized for the specific task, potentially limiting its performance. It was used as is, out-of-the-box.
- Multiple LLMs could not be used to compare inter-LLM performance due to time constraints.

- The data for this study included all online available articles with the terms 'Air Pollution' and 'India'. The quality of the study could be made better by triangulating using multiple newspapers as sources, which was not possible due to download limits and time constraints.
- One of the largest limitations is the black-box, indeterminate nature of LLMs. Care was taken to use the same prompt, with the same parameters on every run to ensure result repeatability, but uncertainty still exists. This lack of transparency leads to many issues, such as hallucinations, toxic responses, and a complete misalignment with ethical values (Zhao, Yang, Shen, Lakkaraju, & Du, 2024). There is ongoing research into increasing the explanability of deep neural networks, but there is a long way to go (Samek, Montavon, Lapuschkin, Anders, & Müller, 2021).

# 8

# Conclusion and Future Work

## 8.1. Research Objectives and Questions Revisited

The primary objective of this study was to evaluate the capability of Large Language Models (LLMs) in automating the qualitative coding of policy narratives through a case study on air pollution in India. The research aimed to assess the performance of LLMs in terms of accuracy and consistency compared to a human-coded dataset.

The guiding research question was:

**To what extent do Large Language Models (LLMs) accurately automate qualitative coding of policy narratives when compared to a manually coded dataset?**

This research question was answered by addressing the following sub-questions:

1. What are the key theories, concepts, and relevant studies in policy narrative analysis using the Narrative Policy Framework?

   The Narrative Policy Framework (NPF) offers a systematic and empirical methodology for analyzing the role and impact of narratives in public policy. This framework, proposed by Jones and McBeth, contrasts with traditional poststructuralist approaches by emphasizing empirical and falsifiable research, integrating narrative analysis into the broader empirical study of policy processes with clear hypotheses and methodological guidelines.

   Central to the NPF is the notion of the social construction of policy realities, positing that policy realities are socially constructed, meaning the significance of policy-related objects and processes varies based on human perceptions. This perspective focuses on collective and individual social constructions rather than objective truths. This concept is complemented by the theory of bounded relativity, which acknowledges the variation in social constructions of policy realities. According to NPF, this variation is bounded by belief systems, ideologies, and norms, which are influenced by identity or culture, thereby ensuring relative stability over time. The NPF further identifies generalizable structural elements within narratives, such as plots and characters, which are consistent across contexts and can be statistically analyzed. This aspect of the framework underscores the empirical nature of the NPF, enabling the systematic study of narratives in a reproducible manner. The framework also operates on three interacting levels of analysis: micro (individual), meso (group/coalitional), and macro (cultural/institutional). This multi-level approach emphasizes the simultaneous operation of narratives across different scales, highlighting their pervasive influence in shaping policy discourse and outcomes. Additionally, the NPF is grounded in the Homo Narrans model, which suggests that humans primarily understand and communicate about the world through narratives. This model asserts that emotions often precede reason, driving cognition, communication, and decision-making. Within the NPF, policy narratives are dissected into two main components: narrative form and narrative content. Narrative form includes the elements present in the narrative, such as setting, characters (heroes, villains, victims, etc.), plot, and the moral of the story (policy solutions). Narrative content, on the other hand, involves the underlying belief systems and narrative strategies used to maximize impact. These strategies may include framing the narrative to resonate with the audiences preexisting beliefs and employing persuasive elements to shift opinions and behaviors.

To ensure systematic analysis, a codebook structure is created to define and categorize these narrative elements, facilitating reproducibility across different contexts. This structured approach enables researchers to systematically identify and analyze the components of policy narratives, contributing to a more nuanced understanding of their role in policy processes. The NPF emphasizes the importance of analyzing narratives at different levels. This study specifically focuses on the meso level, which examines the strategic use of policy narratives by groups or coalitions to advance their policy agendas. At this level, the analysis explores how narratives are constructed and deployed to shape policy discourse and influence policy outcomes. This meso-level focus is particularly relevant for understanding the dynamics of policy advocacy and the strategic use of narratives in shaping public perception and policy decisions.

In terms of research design, this study employs a non-experimental approach, consistent with many NPF studies. Specifically, a case study research design using qualitative content analysis is utilized to analyze text data. This method allows for an in-depth examination of policy narratives, capturing the complexities and nuances of narrative construction and deployment in the context of public policy. Data collection focuses on meso-level data, derived from relevant articles identified by the keyword "Air Pollution in India" from public records. This approach ensures the inclusion of diverse and relevant perspectives, facilitating a comprehensive analysis of policy narratives related to air pollution in India.

2. What narrative components can be used to manually code news articles about the chosen case study to create a robust reference dataset against which LLM agreement will be analyzed?

   A comprehensive codebook was created by taking a page out of the Narrative Policy Framework book (chapter 2) by deductively and inductively adding components. The components used were the narrative elements (character, plot, setting, moral) and narrative content (belief systems, narrative strategies). Characters were 'Heroes', 'Villains', 'Victims', 'Beneficiaries', 'Allies', and 'Opponents', as found in existing NPF literature, with the addition of 'Ineffectives', identified inductively from the data. 'Plot' underwent 8 such additions, 4 additions for 'Belief System', and no additions for 'Narrative Strategy'. This broke down the large amounts of text into components that could be empirically assessed while ensuring repeatability for future studies. The codebook was used for the manual coding, and then was used by the LLM for automated coding. The codebook has also been made available for transparency, and to be used and improved upon for similar future studies (section A.3).

3. What insights are derived from the manual coding of the case study on air pollution in India?

The study identified major characters in the air pollution narratives in India (chapter 4), with medical doctors coming up center stage as the most frequently identified heroes as they spread awareness, farmers/stubble burning coming on top as villains, the environment and vulnerable people of cities portrayed as victims of the villains, and the very same as the beneficiaries of the heroes actions. Government, research universities, and law enforcement crop up as allies, industry and (likely corrupt) officials as opponents, and government interventions and policies sometimes described as ineffective. The settings in the narratives were seen to be present throughout India, with the highest occurrence in the National Capital Region of India, containing the cities of Delhi (Infamous for being the city with the most polluted air in the world, with the AQI there exceeding 500 in extreme cases (Asia Pacific Foundation of Canada, 2023), (Hindustan Times, 2023), (The Guardian, 2023)), Gurugram, and Noida. The plot is largely that of regulatory enforcement and a "Story of Decline", with explicit policy solutions present in more than half the analyzed text corpora. "Egalitarian" and "Hierarchist" belief systems dominate the narrative, with the most commonly applied strategy being that of "Mobilization of Support". The most important finding of this case study is that there is an increasing amount of awareness (from 2010 to 2024) being spread about the detrimental effects of air pollution on health. This shows that India has already started its journey towards cleaner air, because as discussed previously, governments in history take action against air pollution when aware citizens raise their voice (Zeng et al., 2019). Commensurate policy action is scheduled for India, however, it is difficult to say when.

4. What is the consistency of LLM-generated qualitative coding results for the same prompts across multiple runs?

   The LLM was found to be quite consistent in its coding (chapter 5), with accuracy scores (when compared between runs) ranging from 78.1% to 87.9%, precision from 43.8% to 76.3%, recall from 32.6% to 82.6%, F1 from 32.5% to 75.8%, and Krippendorff's $\alpha$ from 0.508 to 0.818, with "Belief System" having the best performance, possibility due to higher perceived clarity in the codebook to distinguish between classes for that narrative component. However, named entity recognition results were challenging to interpret. Despite consistent recognition of major narrative roles, low cosine similarity scores, ranging from 0.19 for opponents to 0.34 for victims, indicated potential for using other metrics to judge character identification similarity (or doing it manually) and comprehensive prompt definitions with more examples to enhance output quality.

5. How does the LLM-generated qualitative coding agree with the manually coded dataset?

   The comparison between LLM coding and human coding revealed that the LLM did not perform well enough to have an agreement with the human (chapter 6), falling below the thresholds on almost all the metrics. Accuracy was low, ranging from 24.6% for 'Plot' to 59.9% for 'Moral'. The ability to correctly identify the true positives was also below par, with precision ranging from 5.4% for 'Narrative Strategy' to 46.3% for 'Moral', and recall ranging from 7.4% for 'Narrative Strategy' to 42.3% for 'Moral', indicating 'Moral' had the best performance amongst all the categories, possibly due to a lower number of output classes, as seen in similar LLM text annotation research[1]. F1 scores, the harmonic mean of precision and recall, were similarly low, ranging from 5.7% for 'Narrative Strategy' to 39.0% for 'Moral'. Krippendorff's $\alpha$ scores showed that the agreement was not enough even for drawing tentative conclusions, with worse than random results in 'Plot' (-0.056) to the highest in 'Moral' (0.279). The output showed a lot of class imbalance, affecting the scores likely by causing skewed agreement, reduced sensitivity to minority classes, bias in expected agreement, and an impact on interpretation.

[1] (Gilardi et al., 2023)

Characters suffered similarly, with low cosine similarity scores, ranging from a low of 0.16 for 'Victim' to a high of 0.62-0.64 for 'Ineffective' and 'Opponent'. 'Ineffective' and 'Opponent' were present only in some of the studied narratives(22/297 and 38/297 respectively), while being absent in most of them, possibly leading to high similarity due to this imbalance. The N-grams showed a match between the frequently appearing characters in the text, but the cosine similarity scores show low overall agreement, raising questions about LLM performance in character identification.

The response to the research question concerning the extent to which large language models (LLMs) can accurately automate qualitative coding of policy narratives is: not sufficiently, at present. LLMs demonstrate potential in handling complex tasks where context is critical; however, their agreement with manual coding, considered the gold standard for comparison in such studies, remains unsatisfactory (Krippendorff, 2013). Enhancements to LLM performance for task-specific requirements, such as fine-tuning through supervised training on labeled examples, prompt engineering, expert development of codebooks with high inter-coder reliability, implementing techniques like Retrieval Augmented Generation to mitigate hallucination, and utilizing AI agents for output quality control, could improve their effectiveness. Nevertheless, in their current state, LLMs require further development, and their application in text annotation necessitates additional research before they can be reliably utilized for policy narrative analysis.

### 8.1.1. Implications of the Findings

Keeping in mind the limitations, the findings have several implications:

#### Practical Implications

Large Language Models (LLMs) offer significant potential for reducing the time and resources required in qualitative coding tasks. However, their current level of performance is not yet adequate for reliable automation without considerable manual supervision. While LLMs can effectively and consistently identify specific terms, their utility for comprehensive coding remains limited. Enhancements such as model fine-tuning with a sample set of coded data (Chamieh et al., 2024), the development of a robust, expert-verified codebook (Pangakis et al., 2023), and the provision of few-shot examples for each category (Chamieh et al., 2024) can significantly improve LLM coding. With these improvements, LLMs could eventually achieve agreement with expert-level manual coding. At this stage, LLMs could function as text annotators, offering significant cost reductions and the ability to annotate data at speeds multiple times faster than human coders, while eliminating human error and fatigue (Gilardi et al., 2023). Utilizing specialized LLMs, rather than general-purpose models, can yield superior performance. For instance, GPT-NER, an LLM architecture specifically designed for named entity recognition, outperforms fully supervised models when data is limited (S. Wang et al., 2023), making it a suitable candidate for character identification tasks.

Another big implication is that since LLMs are general purpose tools for all types of language and math tasks (with limitations there as well), they can be used for the systematic analysis of large volumes of narrative data. The case study in this research is that of air pollution in India, but it could just as easily be that of water scarcity in South Africa, or the recent shift to right wing politics in the democratic countries, and just about any policy issue with narrative data available. The generalizability of this technology allows the reuse of the programming pipeline developed for this study, with the only change needed being the codebook and the data. LLMs are also multimodal now, i.e. they can take images, text, and video as input, and output all images, text, and video. This opens up many more sources of data that can be analyzed beyond text. Policy narratives may also be present in the form of political cartoons and parody videos that can become a part of the analysis.

Theoretical Implications

This study advances the understanding of the capabilities and limitations of LLMs in qualitative content analysis, delineating the contexts in which they can and cannot replace human judgment. The prevailing consensus in the literature, corroborated by this study, is that LLMs cannot yet be wholly relied upon for intricate natural language tasks, such as policy analysis, without human validation. Nonetheless, they provide a valuable starting point for deriving initial insights in complex policy narrative analysis. The use of the NPF helped the study, mainly due to its customizability and context-agnosticism, which allowed its application in the case study. This framework is also easy to use with Large Language Models, breaking the text corpora into classes and named entities through definitions, making it suitable for the task. This framework would likely become one of the top choices for a study such as this.

Policy Implications

Policymakers should exercise caution when considering the adoption of LLMs for qualitative analysis tasks until further advancements are realized. While the automated method presented in this study could serve as a useful initial approach for policy narrative analysis, the necessity for human oversight remains paramount to ensure accuracy and reliability.

### 8.1.2. Recommendations for Future Research

Future studies should address the findings of this study, and build upon the methodology before deploying it for large-scale policy narrative analysis. Following are recommendations that future researchers pursuing this type of study should keep in mind:

- Multiple experienced qualitative coders should create a codebook iteratively to achieve high inter-coder reliability, with as few categories and codes as possible to be mutually exclusive and collectively exhaustive on the sampled dataset. Krippendorff's $\alpha$ can be used as a metric for this.
- Datasets annotated by multiple experienced coders should be used, with inter-coder reliability checks to ensure the quality of manual coding to establish a reliable reference dataset.
- Use multiple sources of narratives for data triangulation. This could include newspaper from publishing houses with different political leanings, non-text data such as videos or images, etc.
- LLMs fine-tuned specifically for qualitative coding tasks in the required domain should be deployed, and techniques such hyperparameter tuning and Retrieval-Augmented Generation (RAG) should be used to improve performance.
- Task specific LLMs, such as GPT-NER, should be used in named entity recognition tasks.
- Larger and more advanced models than Gemini 1.5 Flash should be used to perform qualitative coding tasks.
- Principles of prompt engineering such as context, persona, format, and task should be a part of the deployed prompt, along with any other technique proven with statistical significance to increase LLM output quality.
- More examples for each narrative component should be given, allowing for a few-shot prompt for the LLM.
- Data Imbalance: For categories with high accuracy but low recall (e.g., Plot), addressing class imbalance through resampling may improve performance.

- AI agents can be deployed to control the output quality, with one agent being a generator of information, and another judging if it is right according to defined criteria.

# References


Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural machine translation by jointly learning to align and translate.* Retrieved from `https://arxiv.org/abs/1409.0473`

Baldasano, J., Valera, E., & Jiménez, P. (2003). Air quality data from large cities. *Science of The Total Environment*, *307*(1), 141-165. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0048969702005375` doi: https://doi.org/10.1016/S0048-9697(02)00537-5

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837-865. Retrieved from `https://doi.org/10.1177/0049124118799372` doi: 10.1177/0049124118799372

Bertsou, E., & Caramani, D. (2020). Measuring technocracy. In *The technocratic challenge to democracy* (pp. 91–109). Routledge.

Blignaut, J., & Aronson, J. (2020). Developing a restoration narrative: A pathway towards system-wide healing and a restorative culture. *Ecological Economics*, *168*, 106483. doi: 10.1016/j.ecolecon.2019.106483

Bourdrel, T., Bind, M.-A., Béjot, Y., Morel, O., & Argacha, J.-F. (2017). Cardiovascular effects of air pollution. *Archives of cardiovascular diseases*, *110*(11), 634–642.

Burns, C. D. (1913). The mysticism of a modernist. *Harvard Theological Review*, *6*(3), 316325. doi: 10.1017/s0017816000013298

Castle, M. A. (2011). Abortion in the united states bible belt: organizing for power and empowerment. *Reproductive Health*, *8*(1). doi: 10.1186/1742-4755-8-1

Chamieh, I., Zesch, T., & Giebermann, K. (2024, June). LLMs in short answer scoring: Limitations and promise of zero-shot and few-shot approaches. In E. Kochmar et al. (Eds.), *Proceedings of the 19th workshop on innovative use of nlp for building educational applications (bea 2024)* (pp. 309–315). Mexico City, Mexico: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2024.bea-1.25`

Chowdhary, K. (2020). *Fundamentals of artificial intelligence.* Springer India. doi: 10.1007/978-81-322-3972-7

Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2023). *Deep reinforcement learning from human preferences.* Retrieved from `https://arxiv.org/abs/1706.03741`

Costie, D., & Olofsson, K. (2022). Narrators and Narratives: A Study of Climate and Air Issues in Delhi, India. *International Review of Public Policy*, *4*(2), 171–190. doi: 10.4000/irpp.2698

Crow, D. A., & Berggren, J. (2014). Using the narrative policy framework to understand stakeholder strategy and effectiveness: A multi-case analysis. In *The science of stories: Applications of the narrative policy framework in public policy analysis* (pp. 131–156). Springer.

Davidson, B. (2017). Storytelling and evidence-based policy: lessons from the grey literature. *Palgrave Communications*, *3*(1). doi: 10.1057/palcomms.2017.93

Dennison, J. (2021). Narratives: a review of concepts, determinants, effects, and uses in migration research. *Comparative Migration Studies*, *9*(1). doi: 10.1186/s40878-021-00259-9

Ding, B., Qin, C., Liu, L., Chia, Y. K., Li, B., Joty, S., & Bing, L. (2023, July). Is GPT-3 a good data annotator? In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 11173–11195). Toronto, Canada: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2023.acl-long.626` doi: 10.18653/v1/2023.acl-long.626

Duffy, M. (2017). Rebirth and renewal in maya ritual from the precolonial period to the present. *The Thetean: A Student Journal for Scholarly Historical Writing*, *46*(1), Article 7. Retrieved from https://scholarsarchive.byu.edu/thetean/vol46/iss1/7

Fadlallah, R., El-Jardali, F., Nomier, M., Hemadi, N., Arif, K., Langlois, E. V., & Akl, E. A. (2019). Using narratives to impact health policy-making: a systematic review. *Health Research Policy and Systems*, *17*(1). doi: 10.1186/s12961-019-0423-4

Feng, C. (2014, 09). Mistakes and how to avoid mistakes in using intercoder reliability indices. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 1-10. doi: 10.1027/1614-2241/a000086

Finucane, M. L., & Satterfield, T. A. (2005). Risk as narrative value: a theoretical framework for facilitating the biotechnology debate. *International Journal of Biotechnology*, *7*(1-3), 128–146.

Fischer, & Forester. (1993). *The argumentative turn in policy analysis and planning*. Duke University Press. doi: 10.1215/9780822381815

Fonteyn, M. E., Vettese, M., Lancaster, D. R., & Bauer-Wu, S. (2008). Developing a codebook to guide content analysis of expressive writing transcripts. *Applied Nursing Research*, *21*(3), 165-168. Retrieved from https://www.sciencedirect.com/science/article/pii/S0897189706000991 doi: https://doi.org/10.1016/j.apnr.2006.08.005

Fritzler, A., Logacheva, V., & Kretov, M. (2019, April). Few-shot classification in named entity recognition task. In *Proceedings of the 34th acm/sigapp symposium on applied computing*. ACM. Retrieved from http://dx.doi.org/10.1145/3297280.3297378 doi: 10.1145/3297280.3297378

Gabriel, Y., et al. (2004). Narratives, stories and texts. *The Sage handbook of organizational discourse*(Part I), 61–78.

Gantz, J., & Reinsel, D. (2010). Idc–iview. *International Data Corporation Study*.

Genc, S., Zadeoglulari, Z., Fuss, S. H., & Genc, K. (2012). The adverse effects of air pollution on the nervous system. *Journal of toxicology*, *2012*.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023, July). Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, *120*(30). Retrieved from http://dx.doi.org/10.1073/pnas.2305016120 doi: 10.1073/pnas.2305016120

Glencross, D. A., Ho, T.-R., Camina, N., Hawrylowicz, C. M., & Pfeffer, P. E. (2020). Air pollution and its effects on the immune system. *Free Radical Biology and Medicine*, *151*, 56–68.

Google. (2024). *About google support*. https://support.google.com/a/users/answer/14200040?hl=en. (Accessed: June, 2024)

Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37–45). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-642-24797-2_4 doi: 10.1007/978-3-642-24797-2_4

Gulia, S., Shukla, N., Padhi, L., Bosu, P., Goyal, S., & Kumar, R. (2022). Evolution of air pollution management policies and related research in india. *Environmental Challenges*, *6*, 100431.

Gupta, K., Ripberger, J. T., & Collins, S. (2014). The strategic use of policy narratives: Jaitapur and the politics of siting a nuclear power plant in india. In *The science of stories: Applications of the narrative policy framework in public policy analysis* (pp. 89–106). Springer.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020, September). Array programming with NumPy. *Nature*, *585*(7825), 357–362. Retrieved from https://doi.org/10.1038/s41586-020-2649-2 doi: 10.1038/s41586-020-2649-2

Holy Spirit. (Original text written circa 6th century BCE). *The bible*. Various Publishers. (1 Samuel 17)

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

India. (1950). *The constitution of india*. Dehra Dun: Photolithographed at the Survey of India Offices. (Retrieved from the Library of Congress, https://www.loc.gov/item/57026883/)

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data: Recommendations for the use of performance metrics. In *International conference on affective computing and intelligent interaction and workshops (acii)* (pp. 245–251). doi: 10.1109/ACII.2013.47

Jin, Z., & Mihalcea, R. (2022). Natural language processing for policymaking. *Handbook of Computational Social Science for Policy*, 141-162. doi: 10.1007/978-3-031-16624-2_7

Jones, M. D., & McBeth, M. K. (2010, 5). A narrative policy framework: Clear enough to be wrong? *Policy Studies Journal*, *38*(2), 329–353. doi: 10.1111/j.1541-0072.2010.00364.x

Jones, M. D., Mcbeth, M. K., & Shanahan, E. A. (2014). *Introducing the Narrative Policy Framework* (Tech. Rep.).

Jones, M. D., McBeth, M. K., Shanahan, E. A., Smith-Walter, A., & Song, G. (2022, 1). Conducting Narrative Policy Framework Research: From Theory to Methods. In *Methods of the policy process* (pp. 137–180). Taylor and Francis. doi: 10.4324/9781003269083-6

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, *151*(2), 362–367.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 37133744. doi: 10.1007/s11042-022-13428-4

Kimberly Fiock, M. (2023). *A story of adversity and hardship and beauty and triumph.* https://awis.org/resource/story-adversity-hardship-beauty-triumph/. (Accessed: 2024-06-27)

Kohlbacher, F. (2006, Jan.). The use of qualitative content analysis in case study research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *7*(1). Retrieved from https://www.qualitative-research.net/index.php/fqs/article/view/75 doi: 10.17169/fqs-7.1.75

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology.* SAGE Publications. Retrieved from https://books.google.nl/books?id=s_yqFXnGgjQC

Kuhlmann, J., & Blum, S. (2021). Narrative plots for regulatory, distributive, and redistributive policies. *European Policy Analysis*, *7*(S2), 276302. doi: 10.1002/epa2.1127

Kumari, S., Lakhani, A., Kumari, K. M., et al. (2020). Covid-19 and air pollution in indian cities: Worlds most polluted cities. *Aerosol and Air Quality Research*, *20*(12), 2592–2603.

Legenhausen, M. (2002). Why i am not a traditionalist. *Religioscope. Available online: http://www.religioscope. com/pdf/esotrad/legenhausen. pdf (accessed on 10 June 2017).*

Macagno, F., & Walton, D. (2014). *Emotive language in argumentation.* Cambridge University Press. doi: 10.1017/cbo9781139565776

Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., & Gómez-Berbís, J. M. (2013). Named entity recognition: fallacies, challenges and opportunities. *Computer Standards & Interfaces*, *35*(5), 482–489.

Melilli, K. J. (2003). Leading questions on direct examination: A more pragmatic approach. *Am. J. Trial Advoc.*, *27*, 155.

Meyer, E. (2014). *The culture map: Breaking through the invisible boundaries of global business.* Public Affairs.

Murukutla, N., Negi, N. S., Puri, P., Mullin, S., & Onyon, L. (2017, September). Online media coverage of air pollution risks and current policies in india: A content analysis. *WHO South-East Asia Journal of Public Health*, *6*(2), 41–50. doi: 10.4103/2224-3151.213791

Nasution, A. H., & Onan, A. (2024). Chatgpt label: Comparing the quality of human-generated and llm-generated annotations in low-resource language nlp tasks. *IEEE Access*, *12*, 71876-71900. doi: 10.1109/ACCESS.2024.3402809

Olofsson, K. L., Weible, C. M., Heikkila, T., & Martel, J. C. (2018, 10). Using Nonprofit Narratives and News Media Framing to Depict Air Pollution in Delhi, India. *Environmental Communication*, *12*(7), 956–972. doi: 10.1080/17524032.2017.1309442

OpenAI. (2024). *GPT-4 Technical Report* (Tech. Rep.).

Pangakis, N., Wolken, S., & Fasching, N. (2023). *Automated annotation with generative ai requires validation.* Retrieved from https://arxiv.org/abs/2306.00176

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(Oct), 2825–2830.

Peña, A., Morales, A., Fierrez, J., Serna, I., Ortega-Garcia, J., Puente, Í., … Córdova, G. (2023). Leveraging large language models forătopic classification inătheădomain ofăpublic affairs. In M. Coustaty & A. Fornés (Eds.), *Document analysis and recognition – icdar 2023 workshops* (pp. 20–33). Cham: Springer Nature Switzerland.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research.* Rowman & Littlefield.

Polkinghorne, D. (1988). *Narrative knowing and the human sciences.* State University of New York Press.

Pratap Choudhary, M., & Garg, V. (2013). *Causes, Consequences and Control of Air Pollution* (Tech. Rep.). Retrieved from https://www.researchgate.net/publication/279202084

PTI. (2023, December 21). *Ncap funds largely unused in delhi-ncr cities, govt data shows.* Retrieved from https://indianexpress.com/article/cities/delhi/ncap-funds-largely-unused-delhi-ncr-cities-govt-data-shows-9078071/ (Published in The Indian Express on December 21, 2023.)

Rai, P. (2016, November 6). One killed, six hurt as smog engulfs city, triggers pile-up on roads. *Times of India.* Retrieved from https://timesofindia.indiatimes.com/city/bareilly/one-killed-six-hurt-as-smog-engulfs-city-triggers-pile-up-on-roads/articleshow/55278047.cms

Rainio, O., Teuho, J., & Klén, R. (2024, March 13). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 6086. Retrieved from https://doi.org/10.1038/s41598-024-56706-x doi: 10.1038/s41598-024-56706-x

Ravindra, K., Sidhu, M. K., Mor, S., John, S., & Pyne, S. (2016). Air pollution in india: bridging the gap between science and policy. *Journal of Hazardous, Toxic, and Radioactive Waste*, *20*. doi: 10.1061/(asce)hz.2153-5515.0000303

Reimers, N., & Gurevych, I. (2019, 11). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing.* Association for Computational Linguistics. Retrieved from https://arxiv.org/abs/1908.10084

Renze, M., & Guven, E. (2024). *The effect of sampling temperature on problem solving in large language models.* Retrieved from https://arxiv.org/abs/2402.05201

Rieder, T. N. (2018). Theres never just one side to the story: Why america must stop swinging the opioid pendulum. *Narrative Inquiry in Bioethics*, *8*(3), 225231. doi: 10.1353/nib.2018.0071

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, *109*(3), 247-278. doi: 10.1109/JPROC.2021.3060483

Schlaufer, C., Kuenzler, J., Jones, M. D., & Shanahan, E. A. (2022). The narrative policy framework: A travelers guide to policy stories. *Politische Vierteljahresschrift*, *63*(2), 249273. doi: 10.1007/s11615-022-00379-6

Schmidt, R. M. (2019). Recurrent neural networks (rnns): A gentle introduction and overview. *arXiv preprint arXiv:1912.05911.*

Shanahan, E. A., Jones, M. D., & McBeth, M. K. (2018a). How to conduct a narrative policy framework study. *The Social Science Journal*, *55*(3), 332–345.

Shanahan, E. A., Jones, M. D., & McBeth, M. K. (2018b, 9). How to conduct a Narrative Policy Framework study. *Social Science Journal*, *55*(3), 332–345. doi: 10.1016/j.soscij.2017.12.002

Shanahan, E. A., Jones, M. D., Mcbeth, M. K., & Lane, R. R. (2013, 8). An angel on the wind: How heroic policy narratives shape policy realities. *Policy Studies Journal*, *41*(3), 453–483. doi: 10.1111/psj.12025

Shanahan, E. A., McBeth, M. K., & Hathaway, P. L. (2011, May). Narrative policy framework: The influence of media policy narratives on public opinion. *Politics amp; Policy*, *39*(3), 373400. doi: 10.1111/j.1747-1346.2011.00295.x

Shenoy, J. (2017, June 25). India would need to bring 88 sq. km. land under waste disposal by 2050: Study. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/india/india-would-need-to-bring-88-sq-km-land-under-waste-disposal-by-2050-study/articleshow/59307503.cms`

Singh, D. (2015, June 29). *Delhi air pollution: Vehicles funds unused by aap government.* Retrieved from `https://www.indiatoday.in/mail-today/story/delhi-air-pollution-vehicles-funds-unused-aap-government-260104-2015-06-28` (Published in India Today on June 29, 2015.)

Sirhindi, M. (2018, February 28). Punjab pollution body to measure pollution levels at toll plazas during peak hours. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/chandigarh/punjab-pollution-body-to-measure-pollution-levels-at-toll-plazas-during-peak-hours/articleshow/63114069.cms`

Soriano, M., & Frey, J. B. (1969). From tales of warning to formulettes: The oral tradition in french children's literature. *Yale French Studies*(43), 24–43.

Stone, D. A. (1989, jan). Causal stories and the formation of policy agendas. *Political Science Quarterly*, *104*(2), 281300. doi: 10.2307/2151585

Thompson, M. (2018). *Cultural theory*. Routledge.

Tiwari, A. (2023, October 16). *Air pollution: How is india fighting air pollution? unused funds to tackle it may pose problems.* Retrieved from `https://www.indiatoday.in/diu/story/air-pollutiona-how-is-india-fighting-air-pollution-unused-funds-to-tackle-it-may-pose-problems-2449782-2023-10-16` (Published in India Today on October 16, 2023.)

TOI. (2010, December 31). Locals want clean ganga, smooth roads in 2011. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/kanpur/locals-want-clean-ganga-smooth-roads-in-2011/articleshow/7197688.cms`

TOI. (2012, June 25). Teacher's colony promised respite from plastic by mla. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/hyderabad/teachers-colony-promised-respite-from-plastic-by-mla/articleshow/14363801.cms`

TOI. (2014, June 14). 8 new landfill sites identified by dda. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/delhi/8-new-landfill-sites-identified-by-dda/articleshow/36510528.cms`

TOI. (2017a, December 17). 200 kids march with masks for cleaner air. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/gurgaon/200-kids-march-with-masks-for-cleaner-air/articleshow/62101576.cms`

TOI. (2017b, November 9). Agencies dragged feet on graded response: Experts. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/delhi/agencies-dragged-feet-on-graded-response-experts/articleshow/61568392.cms`

TOI. (2017c, December 29). Plan to combat stubble burning approved. *Times of India*. Retrieved from `https://timesofindia.indiatimes.com/city/delhi/plan-to-combat-stubble-burning-approved/articleshow/62287093.cms`

TOI. (2019a, November 4). Exemptions, rules, fines and more: All you need to know about delhi odd-even scheme. *Times of India*. Retrieved from https://timesofindia.indiatimes.com/city/delhi/exemptions-rules-fines-and-more-all-you-need-to-know-about-delhis-odd-even-scheme/articleshow/71880238.cms

TOI. (2019b, March 2). Smc plans to borrow rs1,925 crore from world bank. *Times of India*. Retrieved from https://timesofindia.indiatimes.com/city/surat/smc-plans-to-borrow-rs1925-crore-from-world-bank/articleshow/68225525.cms

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453458. doi: 10.1126/science.7455683

Van Eijk, E. (2010). Sharia and national law in saudi arabia. *Sharia incorporated: A comparative overview of the legal systems of twelve Muslim countries in past and present*, 139–180.

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., ... Polosukhin, I. (2017). *Attention Is All You Need* (Tech. Rep.).

Verma, R. (2018, October 26). Delhi will use less coal power this winter. *Times of India*. Retrieved from https://timesofindia.indiatimes.com/city/delhi/city-will-use-less-coal-power-this-winter/articleshow/66369595.cms

Veselková, M. (2017). Narrative policy framework: Narratives as heuristics in the policy process. *Human Affairs*, *27*(2), 178–191. Retrieved 2024-06-12, from https://doi.org/10.1515/humaff-2017-0016 doi: doi:10.1515/humaff-2017-0016

Volkens, A. (2007). Strengths and weaknesses of approaches to measuring policy positions of parties. *Electoral Studies*, *26*(1), 108120. doi: 10.1016/j.electstud.2006.04.003

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., ... Wang, G. (2023). *Gpt-ner: Named entity recognition via large language models.* Retrieved from https://arxiv.org/abs/2304.10428

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., ... Zhou, D. (2023). *Self-consistency improves chain of thought reasoning in language models.* Retrieved from https://arxiv.org/abs/2203.11171

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. Retrieved from https://doi.org/10.21105/joss.03021 doi: 10.21105/joss.03021

Wes McKinney. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (p. 56 - 61). doi: 10.25080/Majora-92bf1922-00a

Williams, J. (2014). *Understanding poststructuralism.* Routledge.

Winslade, J., & Monk, G. (2000, 01). Narrative mediation: A new approach to conflict resolution. , *36*.

Wolton, L. P., Crow, D. A., & Heikkila, T. (2021, 11). Stepping Forward: Towards a More Systematic NPF with Automation. In *Narratives and the policy process : Applications of the narrative policy framework.* Montana State University. doi: 10.15788/npf3

Zeng, Y., Cao, Y., Qiao, X., Seyler, B. C., & Tang, Y. (2019, 5). Air pollution reduction in China: Recent success but great challenge for the future. *Science of the Total Environment*, *663*, 329–337. doi: 10.1016/j.scitotenv.2019.01.262

Zhao, H., Yang, F., Shen, B., Lakkaraju, H., & Du, M. (2024). *Towards uncovering how large language model works: An explainability perspective.* Retrieved from https://arxiv.org/abs/2402.10688

Zhou, L., Dai, D., Ren, J., Chen, X., & Chen, S. (2022). What is policy content and how is the public's policy support? a policy cognition study based on natural language processing and social psychology. *Frontiers in Psychology*, *13*. Retrieved from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.941762 doi: 10.3389/fpsyg.2022.941762

# A
# Appendix

## A.1. Gemini Prompt

This is the prompting function written for getting an output from Google Gemini -

```python

#@title Gemini Prompt
def extract(article):
    response = model.generate_content(prompt)
    return response.text
```

where the prompt is -

You are an expert on narrative policy analysis and on using the narrative policy framework. In this policy narrative, (*insert article here*), using the narrative policy framework and the following codebook:

A: **Unit of Analysis** - code everything at a document level.

B: **Narrative Elements - Characters**

**Definition:** Characters are entities portrayed within the narrative, including individuals, groups, organizations, animals, natural phenomena, or abstract concepts. They are depicted as having agency or being acted upon, shaping the plot, themes, and conflicts of the story.

1. **Hero Definition:** The potential fixer of the policy issue, taking action with purpose to achieve or oppose a policy solution. Any actor depicted taking positive steps towards air pollution mitigation, advocating for clean air policies, or raising awareness about air pollutions consequences is considered a hero. **Example:** An environmental activist leading a successful campaign to ban single-use plastics in a major city, thereby reducing plastic waste and improving air quality.
2. **Villain Definition:** The entity causing the policy problem, creating harm or opposition to the hero's aims. **Example:** Large industrial corporations found guilty of illegally dumping toxic waste and emitting high levels of pollutants into the air, despite regulations.
3. **Victim Definition:** The one harmed by the villain, affected negatively by an action or inaction. **Example:** Residents of a community suffering from respiratory issues due to the

nearby factory's unchecked emissions, highlighting the human cost of industrial pollution.

4. **Beneficiary Definition:** Those who benefit from the proposed policy solution; could be an animate character who is explicitly stated, directly linked to a hero, and the receiver of an action of a hero. **Example:** School children in urban areas who experience improved health outcomes and reduced asthma rates after the implementation of stringent air quality standards.
5. **Allies Definition:** Those aligned with the hero, supporting their efforts towards the policy solution. **Example:** Non-governmental organizations (NGOs) and community groups collaborating with environmental activists to lobby for stricter air pollution controls and public awareness campaigns.
6. **Opponents Definition:** An entity opposing a policy but distinct from a villain, often presenting alternative views or objections to the proposed solutions. **Example:** Local businesses that oppose new air quality regulations due to concerns over increased operational costs, arguing that the economic impact outweighs the benefits of cleaner air.
7. **Ineffectives Definition:** An entity that performs an action that has no effect on the policy problem. **Example:** The government installs smog towers in public areas to reduce air pollution, but the towers cannot work in that setting with that much volume, hence rendering their solution useless and their intervention ineffective.

C: **Narrative Elements - Setting Definition:** Setting is the space where the action of the story takes place over time. It is used to focus the audience's attention on a particular space and time. Settings can be specific locations, like a fracking site, or broader contexts, like the American West. They can be one of the following:

1. Specific Location (e.g., fracking site)
2. Broader Context (e.g., American West)

D: **Narrative Elements - Plot Definition:** Plot is the narrative element that links characters to each other and to the setting, organizing actions, and often highlighting the moral of the story. Plots can be traditional story arcs or thematic frameworks used to convey specific messages. The plots used in this analysis are:

1. **Story of Decline Definition:** This plot describes an initial state of well-being that deteriorates over time, highlighting the urgent need for action. It may start with a good situation that worsens or begin at a point where things are already dire. **Example:** Articles detailing the gradual increase in air pollution in New Delhi due to unchecked urban expansion and vehicular emissions, emphasizing the resulting severe health impacts and the pressing need for stringent pollution control measures.
2. **Stymied Progress Definition:** This plot outlines a trajectory of improvement that is halted or reversed by external interference, emphasizing thwarted efforts towards betterment. **Example:** Coverage on how the introduction of the National Clean Air Programme (NCAP) initially led to improvements in air quality across Indian cities, but subsequent political changes and economic pressures, including industrial lobbying, led to a rollback of these gains.
3. **Change-Is-Only an Illusion Definition:** This narrative reveals that perceived changes in a situation are misconceptions, with the real situation being stable or moving in the opposite direction. **Example:** Reports on Mumbai's air quality improvement efforts, which were believed to be successful due to reduced smog levels. However, further studies revealed

that the air quality data was inaccurately reported, and actual pollution levels remained unchanged or worsened.

4. **Story of Helplessness and Control Definition:** This plot describes a dire situation initially seen as unchangeable but later shown to be amendable through specific actions or revelations. **Example:** Articles about the persistent smog in Kanpur, initially believed to be an unavoidable consequence of industrial activity, which was later significantly reduced through the implementation of stringent environmental regulations and the adoption of advanced pollution control technologies.
5. **Conspiracy Definition:** This plot involves a progression from an apparently predetermined state to one where control is exerted by a select few who have been manipulating circumstances for their own benefit. **Example:** Investigative pieces uncovering that despite public efforts to reduce emissions, major coal-based power plants had been covertly lobbying against environmental regulations and manipulating emission data to falsely indicate compliance with national standards.
6. **Blame the Victim Definition:** This plot centres on issues where those suffering from a problem are inaccurately held responsible for causing it. **Example:** In articles discussing poor air quality in rural areas, initial public health campaigns blamed local communities for using traditional biomass for heating and cooking. However, later interventions highlighted the need for systemic solutions, such as providing affordable access to cleaner cooking technologies.
7. **Triumph Over Adversity Definition:** This plot revolves around overcoming significant obstacles through resilience and ingenuity, focusing on successful mitigation of air pollution. **Example:** Stories of how the city of Pune drastically improved its air quality by adopting electric buses, implementing green building codes, and launching successful grassroots campaigns that mobilized the community towards environmental stewardship.
8. **Restoration Definition:** This plot focuses on restoring the environment or social system to its original, pristine condition after suffering degradation, highlighting renewal efforts. **Example:** Reports on the restoration of air quality in Hyderabad through large-scale reforestation projects and the adoption of stringent industrial emission standards that have helped revert the city to cleaner air levels seen decades ago.
9. **Warning Tale Definition:** This plot serves as a cautionary story about the dire consequences of inaction or improper actions, often projecting a bleak future to motivate current action. **Example:** Articles projecting the future health impacts and economic costs of air pollution in Indian megacities if current trends of industrial emissions and vehicular pollution continue without significant intervention, serving as a dire warning to policymakers and the public.
10. **Hero's Journey Definition:** This plot follows a protagonist or group undergoing a transformative journey to resolve a crisis, featuring various trials and eventual success. **Example:** Features on a policymaker or activist who champions groundbreaking legislation or initiatives, such as the introduction of the Bharat Stage VI emission standards, facing significant opposition from various stakeholders but ultimately succeeding in implementing transformative changes.
11. **David vs. Goliath Definition:** This plot highlights the struggle of a seemingly powerless individual or group against a far more powerful adversary, emphasizing justice and equity. **Ex-**

**ample:** Narratives about small communities or environmental groups, like the residents of Singrauli, taking on large industrial conglomerates and winning legal battles to demand reductions in emissions and accountability for environmental degradation.

12. **Rebirth or Renewal Definition:** This plot focuses on transformation and new beginnings after a period of decline or catastrophe, promoting an optimistic and forward-looking perspective. **Example:** Stories of how Ahmedabad transformed from a heavily polluted city to a model of urban green living by implementing a series of successful environmental policies, including extensive green spaces, pollution control measures, and sustainable urban planning.
13. **Pendulum Swing Definition:** This plot captures the cyclical nature of policy and public sentiment, where attitudes and conditions swing from one extreme to another over time. **Example:** Articles illustrating the shifts in air pollution policy in India, from strict regulation under one administration to deregulation under another, and back again, reflecting the changing political landscape and its impact on air quality management.
14. **Regulatory Enforcement Definition:** This plot typically involves a structured progression where regulations are introduced, enforced, and the consequences of these actions are observed and analyzed. **Example:** Coverage on the introduction of the Graded Response Action Plan (GRAP) in Delhi, detailing how the regulations were enforced during periods of severe air pollution, the immediate impact on air quality, and subsequent analyses of the long-term effectiveness and compliance by various stakeholders.

E: **Narrative Elements - Moral/Policy Solution Definition:** Moral of the Story typically presents the policy solution in the policy narrative, frequently culminating in a call to action. It may directly state policy solutions or offer intermediary steps leading to a larger policy solution. They can be one of the following:

1. Explicit Policy Solution (e.g., ban on fracking)
2. Implicit Policy Reference (e.g., economic benefits)

F: **Narrative Strategies - Causal Mechanisms Definition:** Causal mechanisms in narratives explain how various factors within the story are linked causally. They involve intentional, inadvertent, mechanical, and accidental aspects.

1. Mechanical Cause - does the excerpt associate intended consequences by unguided action with a policy problem? **Example:** A bad policy might be explained as resulting from an unthinking bureaucracy. **Example:** A report may describe air pollution as a consequence of industrial practices that are outdated and no longer regulated properly. It might argue that the emissions are an inevitable outcome of an unthinking adherence to old technology and methods, implying that the bureaucracy has failed to update standards and enforce new regulations, thus unintentionally perpetuating high levels of pollution.
2. Intentional Cause - does the excerpt associate intended consequences by purposeful action with a policy problem? **Example:** Policymakers might be accused of making policies to increase their personal wealth. **Example:** Activists might claim that policymakers deliberately resist updating air quality regulations because they receive campaign contributions from big industrial companies. In this view, the persistence of air pollution is portrayed as a direct result of policymakers intentionally acting to protect their own interests and the interests of wealthy donors, at the expense of public health.

3. Accidental Cause - does the excerpt associate unintended consequences by unguided action with a policy problem? **Example:** Climate change might be explained as a natural occurrence having nothing to do with human action. **Example:** An article might argue that air pollution has worsened as an unforeseen consequence of urban sprawl and increased vehicle use. The narrative suggests that as cities grew, no one intended to increase pollution; rather, it was an accidental byproduct of the population's desire for more housing and the convenience of personal transportation.
4. Inadvertent Cause - does the excerpt associate unintended consequences by purposeful action with a policy solution? **Example:** The American Recovery and Reinvestment Act of 2009 might be explained as having raised inflation. **Example:** A policy analysis might discuss how efforts to stimulate economic growth through tax incentives for new factories inadvertently led to increased air pollution. Although the policy aimed to reduce unemployment and boost local economies, the unintended side effect was a significant rise in emissions due to increased industrial activity.
5. Devil Shift: Casting villains as the victors over the heroes.
6. Angel Shift: Casting the heroes as the winners.
7. Mobilization of Support: Rally support for a particular policy position. **Example:** A campaign might mobilize support for clean air policies by distributing images and stories of people, especially children, affected by asthma due to air pollution, aiming to stir public emotions and encourage citizens to demand action from their representatives.
8. Demobilization of Support: Diminish support for opposing views. **Example:** Industry groups might attempt to demobilize support for strict pollution controls by spreading doubt about the scientific consensus on the health impacts of air pollution or by highlighting the potential economic drawbacks, like job losses in the industrial sector.
9. Diffusing Costs and Concentrating Benefits This strategy involves how the costs and benefits of a proposed policy are distributed among the characters in the narrative. The elite few get the advantage, while the common people pay for it.
10. Concentrating Costs and Diffusing Benefits This strategy involves how the costs and benefits of a proposed policy are distributed among the characters in the narrative. Costs are concentrated and benefits are intended for a larger audience, generally the public.

G: **Belief Systems Definition:** Belief systems are viewpoints or ideological frameworks from cultural theory represented within the narrative, shaping how characters, actions, and events are framed and interpreted. Types of belief systems are:

1. Hierarchist **Definition:** Hierarchist belief systems focus on the need for structured regulations and state-led initiatives to tackle air pollution. They advocate for stringent enforcement of emission standards, urban zoning laws, and industrial compliance with environmental regulations. **Example:** Articles supporting the implementation of strict government policies to control industrial emissions, emphasizing the importance of hierarchical enforcement mechanisms to ensure compliance.
2. Individualist **Definition:** Individualist belief systems emphasize innovation, economic growth, and the role of market mechanisms in solving air pollution. Narratives might promote the development and adoption of new technologies like electric vehicles and market-based solutions such as carbon trading. **Example:** Reports highlighting successful startups developing cutting-edge air purification technologies and advocating for carbon trading markets to

incentivize pollution reduction.

3. Egalitarian **Definition:** Egalitarian belief systems stress community involvement and the impacts of air pollution on public health, particularly for the most vulnerable populations. They advocate for policies that ensure equal distribution of clean air as a shared resource and demand radical changes to reduce emissions, such as banning certain pollutants outright. **Example:** Stories emphasizing grassroots movements demanding equitable air quality improvements and policies aimed at protecting low-income communities from disproportionate pollution exposure.
4. Fatalist **Definition:** Fatalist belief systems are characterized by skepticism about the efficacy of any interventions. In the context of air pollution, this translates into narratives that depict efforts to improve air quality as doomed to fail due to overwhelming systemic challenges or corruption. **Example:** Articles portraying the air pollution crisis as insurmountable due to pervasive governmental corruption and deeply ingrained industrial practices.
5. Modernist **Definition:** Modernist belief systems emphasize the power of scientific progress and technological innovation to solve problems, including environmental ones. This belief supports large-scale technological solutions to air pollution, such as the installation of state-of-the-art air purification systems or the development of advanced low-emission public transportation. **Example:** Advocating for government investment in renewable energy technologies to reduce reliance on coal-fired power plants, highlighting the role of cutting-edge science in mitigating air pollution.
6. Traditionalist **Definition:** Traditionalist belief systems emphasize the importance of cultural heritage, continuity, and adherence to historical lifestyles and practices. This belief advocates for the preservation of traditional practices that have a smaller ecological footprint or criticizes modern industrial methods for disrupting natural and social orders. **Example:** Promoting the use of traditional biomass stoves, which are seen as part of cultural heritage despite their contribution to indoor air pollution, and advocating for a return to pre-industrial agricultural practices.
7. Activist **Definition:** Activist belief systems focus on direct action and social change, particularly in the face of perceived government inaction or corporate malfeasance. Narratives driven by this belief system mobilize public demonstrations or campaigns to pressure policymakers into taking action against air pollution. **Example:** Organizing mass protests demanding stricter air quality regulations and immediate government action to reduce air pollution, featuring vivid accounts of public demonstrations and grassroots campaigns.
8. Technocratic **Definition:** Trusts in experts and technical solutions over political or public opinion, emphasizing the role of educated elites and technologists in crafting policy solutions. This could lead to advocacy for solutions based on scientific research and data-driven approaches. **Example:** Supporting policies that fund extensive air quality monitoring and research into long-term health impacts of air pollution, used to craft precise interventions based on empirical evidence.

Tell me who the explicitly mentioned heroes, villains, victims, beneficiaries, allies, opponents, and ineffectives are in the news article. Also, identify the setting, the plot, the presence of ONLY ONE type of each of the following - moral (implicit policy reference or explicit policy solution), belief system, and narrative strategy of the article. Give it to me in a JSON format that is easily parsable, for example:

```
{
```

```
"Hero":"(Name of hero 1, Name of hero 2, etc)",
"Villain":"(Name of Villain 1, Name of Villain 2, etc)",
"Victim":"(Name of Victim 1, Name of Victim 2, etc)",
"Beneficiary":"(Name of Beneficiary 1, Name of Beneficiary 2, etc)",
"Ally":"(Name of Ally 1, Name of Ally 2, etc)",
"Opponent":"(Name of Opponent 1, Name of Opponent 2, etc)",
"Ineffective":"(Name of Ineffective 1, Name of Ineffective 2, etc)",
"Setting":"Name of the place/area",
"Plot":"Name of Plot",
"Moral":"Explicit Policy Solution/Implicit Policy Reference",
"Belief System":"Name of System",
"Narrative Strategy":"Name of Strategy"
}
```

## A.2. Python Code

All the code is available in the form of Python notebooks here - Github Repository.

## A.3. Codebook

# Codebook for Qualitative Content Analysis of Policy Narratives Using the Narrative Policy Framework

## General Coding Guidelines -

Unit of Analysis: Define whether coding by sentence, paragraph, or document.

Manifest Content: Only explicit content is coded; avoid inferring or interpreting meanings.

Repetition: Characters and key elements are coded once per document, regardless of how often they appear in the text.

Completeness: Fill all cells; use “0” for absence.

Documentation: Maintain a log of coding decisions and note any ambiguities or unusual cases in a separate notes column.

## 1: Policy Narrative Demographics

Unit of Analysis (sentence, paragraph, document)

## 2: Narrative Elements - Characters

Narrative Elements - Characters

Definition: Characters are entities portrayed within the narrative, including individuals, groups, organizations, animals, natural phenomena, or abstract concepts. They are depicted as having agency or being acted upon, shaping the plot, themes, and conflicts of the story.

1. **Hero**

**Definition**: The potential fixer of the policy issue, taking action with purpose to achieve or oppose a policy solution. Any actor depicted taking positive steps towards air pollution mitigation, advocating for clean air policies, or raising awareness about air pollution’s consequences is considered a hero.

**Example**: An environmental activist leading a successful campaign to ban single-use plastics in a major city, thereby reducing plastic waste and improving air quality.

2. **Villain**

**Definition**: The entity causing the policy problem, creating harm or opposition to the hero's aims.

**Example**: Large industrial corporations found guilty of illegally dumping toxic waste and emitting high levels of pollutants into the air, despite regulations.

3. **Victim**

**Definition**: The one harmed by the villain, affected negatively by an action or inaction.

**Example**: Residents of a community suffering from respiratory issues due to the nearby factory's unchecked emissions, highlighting the human cost of industrial pollution.

4. **Beneficiary**

**Definition**: Those who benefit from the proposed policy solution; could be an animate character who is explicitly stated, directly linked to a hero, and the receiver of an action of a hero.

**Example**: School children in urban areas who experience improved health outcomes and reduced asthma rates after the implementation of stringent air quality standards.

5. **Allies**

**Definition**: Those aligned with the hero, supporting their efforts towards the policy solution.

Example: Non-governmental organizations (NGOs) and community groups collaborating with environmental activists to lobby for stricter air pollution controls and public awareness campaigns.

6. **Opponents**

**Definition**: An entity opposing a policy but distinct from a villain, often presenting alternative views or objections to the proposed solutions.

**Example**: Local businesses that oppose new air quality regulations due to concerns over increased operational costs, arguing that the economic impact outweighs the benefits of cleaner air.

7. **Ineffective**

**Definition:** An entity that performs an action that has no effect on the policy problem.

**Example:** The government installs smog towers in public areas to reduce air pollution, but the towers cannot work in that setting with that much volume, hence rendering their solution useless and their intervention ineffective.

## 3: Narrative Elements - Setting

Definition: Setting is the space where the action of the story takes place over time. It is used to focus the audience's attention to a particular space and time. Settings can be specific locations, like a fracking site, or broader contexts, like the American West.

Specific Location (e.g., fracking site)

Broader Context (e.g., American West)

## 4: Narrative Elements - Plot

Definition: Plot is the narrative element that links characters to each other and to the setting, organizing actions, and often highlighting the moral of the story. Plots can be traditional story arcs or thematic frameworks used to convey specific messages.

**1. Story of Decline**

**Definition:** This plot describes an initial state of well-being that deteriorates over time, highlighting the urgent need for action. It may start with a good situation that worsens or begin at a point where things are already dire.

**Example:** Articles detailing the gradual increase in air pollution in New Delhi due to unchecked urban expansion and vehicular emissions, emphasizing the resulting severe health impacts and the pressing need for stringent pollution control measures.

**2. Stymied Progress**

**Definition:** This plot outlines a trajectory of improvement that is halted or reversed by external interference, emphasizing thwarted efforts towards betterment.

**Example:** Coverage on how the introduction of the National Clean Air Programme (NCAP) initially led to improvements in air quality across Indian cities, but subsequent political changes and economic pressures, including industrial lobbying, led to a rollback of these gains.

**3. Change-Is-Only an Illusion**

**Definition:** This narrative reveals that perceived changes in a situation are misconceptions, with the real situation being stable or moving in the opposite direction.

**Example:** Reports on Mumbai's air quality improvement efforts, which were believed to be successful due to reduced smog levels. However, further studies revealed that the air quality data was inaccurately reported, and actual pollution levels remained unchanged or worsened.

**4. Story of Helplessness and Control**

**Definition:** This plot describes a dire situation initially seen as unchangeable but later shown to be amendable through specific actions or revelations.

**Example:** Articles about the persistent smog in Kanpur, initially believed to be an unavoidable consequence of industrial activity, which was later significantly reduced through the implementation of stringent environmental regulations and the adoption of advanced pollution control technologies.

**5. Conspiracy**

**Definition:** This plot involves a progression from an apparently predetermined state to one where control is exerted by a select few who have been manipulating circumstances for their own benefit.

**Example:** Investigative pieces uncovering that despite public efforts to reduce emissions, major coal-based power plants had been covertly lobbying against environmental regulations and manipulating emission data to falsely indicate compliance with national standards.

**6. Blame the Victim**

**Definition:** This plot centres on issues where those suffering from a problem are inaccurately held responsible for causing it.

**Example:** In articles discussing poor air quality in rural areas, initial public health campaigns blamed local communities for using traditional biomass for heating and cooking. However, later interventions highlighted the need for systemic solutions, such as providing affordable access to cleaner cooking technologies.

**7. Triumph Over Adversity**

**Definition:** This plot revolves around overcoming significant obstacles through resilience and ingenuity, focusing on successful mitigation of air pollution.

**Example:** Stories of how the city of Pune drastically improved its air quality by adopting electric buses, implementing green building codes, and launching successful grassroots campaigns that mobilized the community towards environmental stewardship.

**8. Restoration**

**Definition:** This plot focuses on restoring the environment or social system to its original, pristine condition after suffering degradation, highlighting restorative efforts.

**Example:** Reports on the restoration of air quality in Hyderabad through large-scale reforestation projects and the adoption of stringent industrial emission standards that have helped revert the city to cleaner air levels seen decades ago.

**9. Warning Tale**

**Definition:** This plot serves as a cautionary story about the dire consequences of inaction or improper actions, often projecting a bleak future to motivate current action.

**Example:** Articles projecting the future health impacts and economic costs of air pollution in Indian megacities if current trends of industrial emissions and vehicular pollution continue without significant intervention, serving as a dire warning to policymakers and the public.

**10. Hero's Journey**

**Definition:** This plot follows a protagonist or group undergoing a transformative journey to resolve a crisis, featuring various trials and eventual success.

**Example:** Features on a policy maker or activist who champions groundbreaking legislation or initiatives, such as the introduction of the Bharat Stage VI emission standards, facing significant opposition from various stakeholders but ultimately succeeding in implementing transformative changes.

**11. David vs. Goliath**

**Definition:** This plot highlights the struggle of a seemingly powerless individual or group against a far more powerful adversary, emphasizing justice and equity.

**Example:** Narratives about small communities or environmental groups, like the residents of Singrauli, taking on large industrial conglomerates and winning legal battles to demand reductions in emissions and accountability for environmental degradation.

**12. Rebirth or Renewal**

**Definition:** This plot focuses on transformation and new beginnings after a period of decline or catastrophe, promoting an optimistic and forward-looking perspective.

**Example:** Stories of how Ahmedabad transformed from a heavily polluted city to a model of urban green living by implementing a series of successful environmental policies, including extensive green spaces, pollution control measures, and sustainable urban planning.

**13. Pendulum Swing**

**Definition:** This plot captures the cyclical nature of policy and public sentiment, where attitudes and conditions swing from one extreme to another over time.

**Example:** Articles illustrating the shifts in air pollution policy in India, from strict regulation under one administration to deregulation under another, and back again, reflecting the changing political landscape and its impact on air quality management.

**14. Regulatory Enforcement**

**Definition**: This plot typically involves a structured progression where regulations are introduced, enforced, and the consequences of these actions are observed and analyzed.

**Example**: Coverage on the introduction of the Graded Response Action Plan (GRAP) in Delhi, detailing how the regulations were enforced during periods of severe air pollution, the immediate impact on air quality, and subsequent analyses of the long-term effectiveness and compliance by various stakeholders.

## 5: Narrative Elements - Moral/Policy Solution

Definition: Moral of the Story typically presents the policy solution in the policy narrative, frequently culminating in a call to action. It may directly state policy solutions or offer intermediary steps leading to a larger policy solution.

A: Explicit Policy Solution (e.g., ban on fracking)

B: Implicit Policy Reference (e.g., economic benefits)

## 6: Narrative Strategies - Causal Mechanisms

Definition: Causal mechanisms in narratives explain how various factors within the story are linked causally. They involve intentional, inadvertent, mechanical, and accidental aspects.

A: **Mechanical** Cause - does the excerpt associate intended consequences by unguided action with a policy problem? EX: a bad policy might be explained as resulting from an unthinking bureaucracy.

**Example**: A report may describe air pollution as a consequence of industrial practices that are outdated and no longer regulated properly. It might argue that the emissions are an inevitable outcome of an unthinking adherence to old technology and methods, implying that the bureaucracy has failed to update standards and enforce new regulations, thus unintentionally perpetuating high levels of pollution.

B: **Intentional** Cause - does the excerpt associate intended consequences by purposeful action with a policy problem? EX: policymakers might be accused of making policies to increase their personal wealth.

**Example**: Activists might claim that policymakers deliberately resist updating air quality regulations because they receive campaign contributions from big industrial companies. In this view, the persistence of air pollution is portrayed as a direct result of policymakers intentionally acting to protect their own interests and the interests of wealthy donors, at the expense of public health.

C: **Accidental** Cause - does the excerpt associate unintended consequences by unguided action with a policy problem? EX: climate change might be explained as a natural occurrence having nothing to do with human action.

**Example**: An article might argue that air pollution has worsened as an unforeseen consequence of urban sprawl and increased vehicle use. The narrative suggests that as cities grew, no one intended to increase pollution; rather, it was an accidental byproduct of the population's desire for more housing and the convenience of personal transportation.

D: **Inadvertent** Cause - does the excerpt associate unintended consequences by purposeful action with a policy solution? EX: the American Recovery and Reinvestment Act of 2009 might be explained as having raised inflation.

**Example**: A policy analysis might discuss how efforts to stimulate economic growth through tax incentives for new factories inadvertently led to increased air pollution. Although the policy aimed to reduce unemployment and boost local economies, the unintended side effect was a significant rise in emissions due to increased industrial activity.

E: **Devil Shift**: Casting villains as the victors over the heroes.

**Example**: In discussions about air pollution control measures, proponents of strict regulations might depict companies opposing the regulations as villains who prioritize profits over public health, suggesting they have managed to defeat public interest groups by manipulating regulatory processes.

F: **Angel Shift**: Casting the heroes as the winners.

G: **Mobilization of Support**: Rally support for a particular policy position.

**Example**: A campaign might mobilize support for clean air policies by distributing images and stories of people, especially children, affected by asthma due to air pollution, aiming to stir public emotions and encourage citizens to demand action from their representatives.

H: **Demobilization of Support**: Diminish support for opposing views.

**Example**: Industry groups might attempt to demobilize support for strict pollution controls by spreading doubt about the scientific consensus on the health impacts of air pollution or by highlighting the potential economic drawbacks, like job losses in the industrial sector.

I: **Diffusing Costs and Concentrating Benefits –** This strategy involves how the costs and benefits of a proposed policy are distributed among the characters in the narrative. The elite few get the advantage, while the common people pay for it

J: **Concentrating Costs and Diffusing Benefits –** This strategy involves how the costs and benefits of a proposed policy are distributed among the characters in the narrative. Costs are concentrated and benefits are intended for a larger audience, generally the public.

## 7: Belief Systems

Definition: Belief systems are viewpoints or ideological frameworks from cultural theory represented within the narrative, shaping how characters, actions, and events are framed and interpreted.

1. **Hierarchist**

**Definition**: Hierarchist belief systems focus on the need for structured regulations and state-led initiatives to tackle air pollution. They advocate for stringent enforcement of emission standards, urban zoning laws, and industrial compliance with environmental regulations.

**Example**: Articles supporting the implementation of strict government policies to control industrial emissions, emphasizing the importance of hierarchical enforcement mechanisms to ensure compliance.

2. **Individualist**

**Definition**: Individualist belief systems emphasize innovation, economic growth, and the role of market mechanisms in solving air pollution. Narratives might promote the development and adoption of new technologies like electric vehicles and market-based solutions such as carbon trading.

**Example**: Reports highlighting successful startups developing cutting-edge air purification technologies and advocating for carbon trading markets to incentivize pollution reduction.

3. **Egalitarian**

**Definition**: Egalitarian belief systems stress community involvement and the impacts of air pollution on public health, particularly for the most vulnerable populations. They advocate for policies that ensure equal distribution of clean air as a shared resource and demand radical changes to reduce emissions, such as banning certain pollutants outright.

**Example**: Stories emphasizing grassroots movements demanding equitable air quality improvements and policies aimed at protecting low-income communities from disproportionate pollution exposure.

4. **Fatalist**

**Definition**: Fatalist belief systems are characterized by skepticism about the efficacy of any interventions. In the context of air pollution, this translates into narratives that depict efforts to improve air quality as doomed to fail due to overwhelming systemic challenges or corruption.

**Example**: Articles portraying the air pollution crisis as insurmountable due to pervasive governmental corruption and deeply ingrained industrial practices.

5. **Modernist**

**Definition**: Modernist belief systems emphasize the power of scientific progress and technological innovation to solve problems, including environmental ones. This belief supports large-scale technological solutions to air pollution, such as the installation of state-of-the-art air purification systems or the development of advanced low-emission public transportation.

**Example**: Advocating for government investment in renewable energy technologies to reduce reliance on coal-fired power plants, highlighting the role of cutting-edge science in mitigating air pollution.

6. **Traditionalist**

**Definition**: Traditionalist belief systems emphasize the importance of cultural heritage, continuity, and adherence to historical lifestyles and practices. This belief advocates for the preservation of traditional practices that have a smaller ecological footprint or criticizes modern industrial methods for disrupting natural and social orders.

**Example**: Promoting the use of traditional biomass stoves, which are seen as part of cultural heritage despite their contribution to indoor air pollution, and advocating for a return to pre-industrial agricultural practices.

7. **Activist**

**Definition**: Activist belief systems focus on direct action and social change, particularly in the face of perceived government inaction or corporate malfeasance. Narratives driven by this belief system mobilize public demonstrations or campaigns to pressure policymakers into taking action against air pollution.

**Example**: Organizing mass protests demanding stricter air quality regulations and immediate government action to reduce air pollution, featuring vivid accounts of public demonstrations and grassroots campaigns.

8. **Technocratic**

**Definition**: Trusts in experts and technical solutions over political or public opinion, emphasizing the role of educated elites and technologists in crafting policy solutions. This could lead to advocacy for solutions based on scientific research and data-driven approaches.

**Example**: Supporting policies that fund extensive air quality monitoring and research into long-term health impacts of air pollution, used to craft precise interventions based on empirical evidence.

### A.3.1. Sources of Narrative Data

All the narrative data for this study was in the form of newspapers as described in section 3.2. Here are some other sources that may be used in future studies -

**Interest Group Data** **Public Consumption Documents:** These include newsletters, reports, and other documents produced by interest groups for public dissemination. They often provide rich narrative data on policy debates.

**Newspapers** **Media Coverage:** Newspapers offer a wealth of narrative data, including editorials and news stories. Researchers must provide a rationale for including or excluding specific types of media coverage and apply the NPFs definition of a policy narrative to these texts.

**Transcripts, Speeches, and Digital Media** **Television Transcripts:** Sources like Lexis-Nexis provide access to TV news program transcripts.

**Speeches:** Elected officials, candidates, and other leaders often have their speeches transcribed and available online.

**Digital Media:** Platforms like YouTube, Facebook, and Twitter are increasingly used for narrative analysis, despite presenting unique challenges in accessing and coding the data. Coding visual narratives (e.g., YouTube videos, images in tweets) has shown to be effective in understanding the impact of policy narratives on public attention and emotional reactions (Jones et al., 2022).

**Interviews and Focus Groups:** Semi-structured interviews allow for targeted data acquisition related to NPF elements, while focus groups can reveal narrative understandings among participants.

## A.4. All Metrics

## A.5. Automated Coding

Context within narratives is crucial, which has led to most Narrative Policy Framework studies being conducted manually, typically by trained human coders (Jones et al., 2022; Shanahan et al., 2018b). Efforts to automate this process have been made through automatic dictionary lookup, wherein words associated with specific themes are identified and classified within the text. This method dates back to 1962 and has been utilized in contemporary policy research, such as Olofsson et al.'s (2018) study on air pollution narratives in Delhi (Wolton et al., 2021). However, this research is primarily limited to thematic analysis and does not fully capture the nuances of narratives, largely due to the constraints of Natural Language Processing techniques (Shanahan et al., 2018b; Wolton et al., 2021).

### A.5.1. Natural Language Processing

> "Natural Language Processing is a collection of computational techniques for automatic analysis and representation of human languages, motivated by theory" (Chowdhary, 2020).

Natural Language Processing aims to enable computers to understand, interpret, and generate human language in a meaningful and useful way. Natural Language Processing encompasses a broad range of applications, including machine translation, sentiment analysis, question answering, and information extraction.

### A.5.2. Components of NLP

NLP can be broadly divided into two parts: Natural Language Understanding (NLU) and Natural Language Generation (NLG) (Khurana, Koli, Khatter, & Singh, 2022).

**Natural Language Understanding (NLU)** NLU involves the comprehension of human language by a machine and includes several linguistic levels (Khurana et al., 2022):

- **Phonology**: Systematic organization of sounds in languages.
- **Morphology**: Study of word structure and morpheme formation, which are the smallest units of meaning.
- **Lexical Analysis**: Understanding the meaning of words and their roles in sentences, often through part-of-speech tagging.
- **Syntax**: Arrangement of words to form sentences and their grammatical structure.
- **Semantics**: Meaning of words and sentences, including disambiguation of words with multiple meanings.
- **Discourse**: Analysis of text structure beyond individual sentences to understand flow and coherence.
- **Pragmatics**: Consideration of context and intended meaning behind the text.

**Natural Language Generation (NLG)** NLG is the process of producing coherent text from a machine's internal representation and involves (Khurana et al., 2022):

- **Content Selection**: Deciding what information to include.
- **Textual Organization**: Structuring the information logically.
- **Linguistic Resources**: Choosing appropriate words and phrases.
- **Realization**: Generating the final text.

**Processing vs. Understanding** Processing and understanding are distinct yet interrelated aspects of NLP. Processing refers to the technical manipulation of language data, such as tokenization, parsing, and stemming. Understanding involves comprehending the meaning and context of the language, which is crucial for tasks like sentiment analysis and machine translation.

### A.5.3. Standard Techniques in NLP

Several techniques and models are employed in NLP to process the given input and produce a coherent output (Khurana et al., 2022), (Chowdhary, 2020):

- **Tokenization**: Breaking down text into individual units (tokens) such as words or phrases. This is an important part of all NLP tasks as computers only understand numbers, and series of tokens can be represented numerically.
- **Part-of-Speech Tagging (PoS)**: Assigning parts of speech to each token based on its role in the sentence.
- **Named Entity Recognition (NER)**: Identifying and classifying entities in the text such as names, dates, and organizations.
- **Parsing**: Analyzing the grammatical structure of sentences to uncover relationships between words.

- **Stemming and Lemmatization**: Reducing words to their root forms to aid in analysis. This removes the 'ing' or the plural form of 'es' in words.
- **Word Embeddings**: Representing words in a high dimensional vector space to capture semantic similarities. This is used to either compare words and phrases, or operations can be done on these vectors to change their meaning, as similarly spelled words may have different meanings in different contexts. For example - "The person is lying on the bed" versus "The person is lying to me".
- **Neural Networks**: Using deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to process and understand language. Multi-layer perceptrons (Neural Networks) are best known for their capability is approximating non-linear functions with remarkable accuracy. Processing natural language in terms of mathematical equations is accomplished in this way.

### A.5.4. Prompting

With the advent of large language models, the field of 'Prompt Engineering' has emerged. There has been a lot of research in this area. Specifically for Google Gemini, the Google team has released these set of instructions (Google, 2024) for creating prompts -

1. Use clear, concise, and natural language - Give clear instructions as to what the task requires, and what your expectations are.
2. Provide all the necessary context - Give any background information with keywords specific and relevant to the task provided. The NPF codebook was given as an input.
3. Use a chain of prompts to break down complex tasks - Provide step by step instructions if the task requires critical analysis.
4. Provide a persona to the LLM - For example, in the case of policy analysis, the prompt should start with "You are an expert on narrative policy analysis using the narrative policy framework."
5. Format - Describe exactly how the output is needed. In this study, the LLM was asked to provide the narrative element and content in a JSON format, along with an explanation for each as a string.

The prompt can be seen in the section A.1.

## A.6. The History of Natural Language Processing

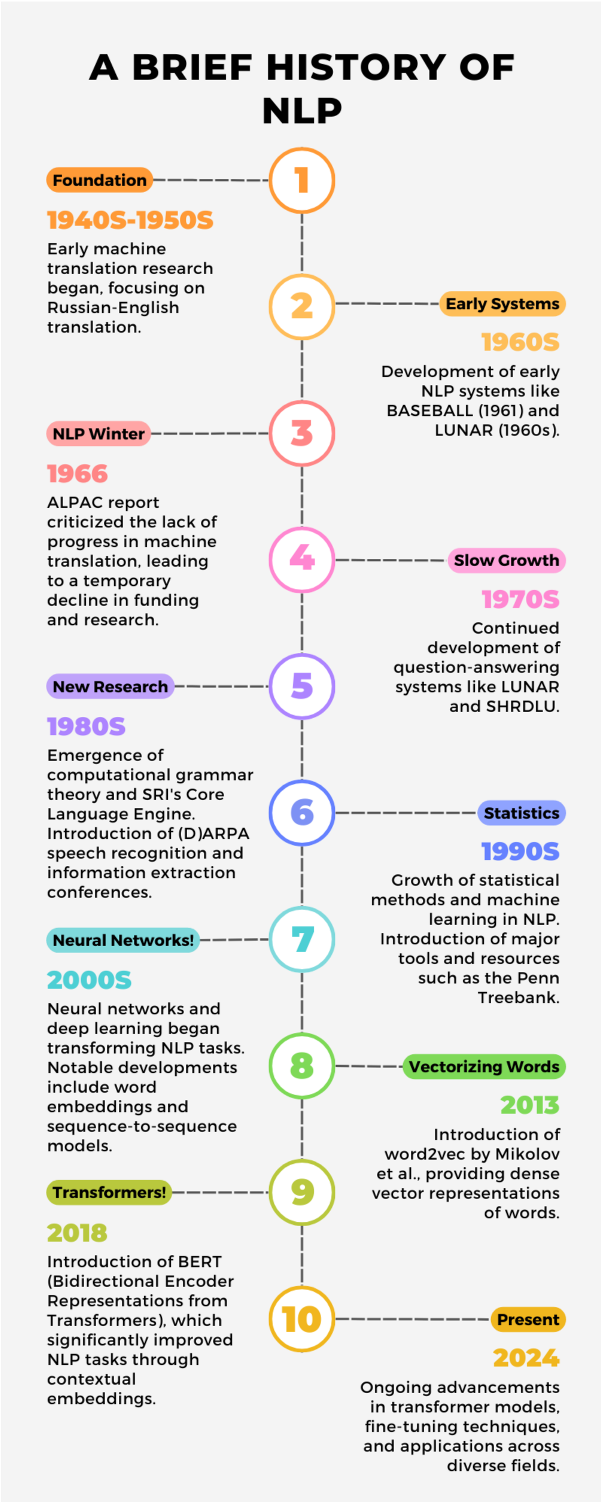



**Figure A.1:** Timeline (Khurana et al., 2023)

Natural Language Processing has come a long way from a rudimentary Russian-English translator to multi-billion parameter Generative Large Language Models.

## A.6.1. Applications of NLP

NLP has numerous applications across different fields:

- **Machine Translation**: Automatically translating text between languages, such as Google Translate.
- **Text Categorization**: Classifying text into predefined categories, used in spam detection and content moderation.
- **Sentiment Analysis**: Analyzing the sentiment expressed in text, widely used in social media monitoring and customer feedback analysis.
- **Question Answering**: Developing systems that can answer questions posed in natural language, such as virtual assistants and chatbots.
- **Information Extraction**: Extracting specific information from text, used in data mining and knowledge management.
- **Summarization**: Automatically generating concise summaries of longer texts, useful in news aggregation and document management.

## A.6.2. Challenges in NLP

- **Ambiguity**: Handling the multiple meanings of words and sentences.
- **Context Understanding**: Grasping the broader context beyond individual sentences.
- **Multilingual Processing**: Developing models that work across different languages and dialects.
- **Resource Limitations**: Lack of annotated data for training models in less-studied languages.
- **Bias and Fairness**: Ensuring that NLP models do not perpetuate or amplify biases present in training data.

| Coder | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| **Ground Truth and Gemini Flash Run 1** | | | | |
| Hero | 0.053872 | 0.026544 | 0.025483 | 0.024968 |
| Villain | 0.030303 | 0.018072 | 0.014992 | 0.015249 |
| Victim | 0.013468 | 0.011730 | 0.008814 | 0.008830 |
| Beneficiary | 0.013468 | 0.007326 | 0.003706 | 0.003748 |
| Ally | 0.030303 | 0.017886 | 0.019535 | 0.017930 |
| Opponent | 0.060606 | 0.006536 | 0.000454 | 0.000849 |
| Ineffective | 0.057239 | 0.017809 | 0.009708 | 0.010180 |
| Setting | 0.040404 | 0.027135 | 0.018841 | 0.020713 |
| Plot | 0.245791 | 0.230632 | 0.145902 | 0.120189 |
| Moral | 0.558923 | 0.462553 | 0.400467 | 0.345912 |
| Belief System | 0.350168 | 0.236790 | 0.207539 | 0.149685 |
| Narrative Strategy | 0.454545 | 0.090569 | 0.089895 | 0.078231 |
| **Ground Truth and Gemini Flash Run 2** | | | | |
| Hero | 0.060606 | 0.027896 | 0.024784 | 0.024642 |
| Villain | 0.037037 | 0.017195 | 0.015228 | 0.015377 |
| Victim | 0.013468 | 0.010324 | 0.008866 | 0.008882 |
| Beneficiary | 0.077441 | 0.006988 | 0.003877 | 0.004133 |
| Ally | 0.094276 | 0.022205 | 0.025930 | 0.023036 |
| Opponent | 0.545455 | 0.012932 | 0.010739 | 0.011605 |
| Ineffective | 0.501684 | 0.025042 | 0.021299 | 0.022620 |
| Setting | 0.047138 | 0.033403 | 0.027649 | 0.028649 |
| Plot | 0.249158 | 0.239119 | 0.153509 | 0.125842 |
| Moral | 0.599327 | 0.452263 | 0.422755 | 0.390120 |
| Belief System | 0.360269 | 0.199054 | 0.258664 | 0.174460 |
| Narrative Strategy | 0.427609 | 0.054074 | 0.074137 | 0.056703 |
| **Gemini Flash Run 1 and Gemini Flash Run 2** | | | | |
| Hero | 0.360269 | 0.211206 | 0.212348 | 0.210968 |
| Villain | 0.343434 | 0.172956 | 0.177658 | 0.174547 |
| Victim | 0.279461 | 0.135257 | 0.137946 | 0.135178 |
| Beneficiary | 0.164983 | 0.074676 | 0.077987 | 0.074013 |
| Ally | 0.218855 | 0.134353 | 0.137685 | 0.133566 |
| Opponent | 0.070707 | 0.038283 | 0.041115 | 0.038569 |
| Ineffective | 0.074074 | 0.030518 | 0.033874 | 0.030691 |
| Setting | 0.420875 | 0.220421 | 0.223487 | 0.217982 |
| Plot | 0.781145 | 0.487067 | 0.506579 | 0.489918 |
| Moral | 0.851852 | 0.726640 | 0.826451 | 0.757928 |
| Belief System | 0.878788 | 0.763222 | 0.781999 | 0.730984 |
| Narrative Strategy | 0.801347 | 0.437805 | 0.326037 | 0.324831 |

**Table A.1:** All Results for Gemini 1.5 Flash Runs 1 and 2 with Temperature 1

| Coder | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| **Manual Coding and Gemini Pro Run 1** | | | | |
| Hero | 0.040 | 0.020 | 0.014 | 0.015 |
| Villain | 0.064 | 0.019 | 0.018 | 0.019 |
| Victim | 0.027 | 0.013 | 0.007 | 0.008 |
| Beneficiary | 0.094 | 0.007 | 0.006 | 0.003 |
| Ally | 0.077 | 0.017 | 0.015 | 0.015 |
| Opponent | 0.232 | 0.022 | 0.014 | 0.016 |
| Ineffective | 0.205 | 0.011 | 0.002 | 0.004 |
| Setting | 0.205 | 0.069 | 0.034 | 0.041 |
| Plot | 0.296 | 0.207 | 0.153 | 0.128 |
| Moral | 0.727 | 0.485 | 0.495 | 0.489 |
| Belief System | 0.478 | 0.251 | 0.247 | 0.211 |
| Narrative Strategy | 0.391 | 0.113 | 0.090 | 0.076 |
| **Ground Truth and Gemini Flash Run 2** | | | | |
| Hero | 0.040 | 0.020 | 0.014 | 0.015 |
| Villain | 0.064 | 0.019 | 0.018 | 0.019 |
| Victim | 0.027 | 0.013 | 0.007 | 0.008 |
| Beneficiary | 0.094 | 0.007 | 0.006 | 0.003 |
| Ally | 0.077 | 0.017 | 0.015 | 0.015 |
| Opponent | 0.232 | 0.022 | 0.014 | 0.016 |
| Ineffective | 0.205 | 0.011 | 0.002 | 0.004 |
| Setting | 0.205 | 0.069 | 0.034 | 0.041 |
| Plot | 0.296 | 0.207 | 0.153 | 0.128 |
| Moral | 0.727 | 0.485 | 0.495 | 0.489 |
| Belief System | 0.478 | 0.251 | 0.247 | 0.211 |
| Narrative Strategy | 0.391 | 0.113 | 0.090 | 0.076 |
| **Manual Coding and Gemini Pro Run 2** | | | | |
| Hero | 1.000 | 1.000 | 1.000 | 1.000 |
| Villain | 1.000 | 1.000 | 1.000 | 1.000 |
| Victim | 1.000 | 1.000 | 1.000 | 1.000 |
| Beneficiary | 1.000 | 1.000 | 1.000 | 1.000 |
| Ally | 1.000 | 1.000 | 1.000 | 1.000 |
| Opponent | 1.000 | 1.000 | 1.000 | 1.000 |
| Ineffective | 1.000 | 1.000 | 1.000 | 1.000 |
| Setting | 1.000 | 1.000 | 1.000 | 1.000 |
| Plot | 1.000 | 1.000 | 1.000 | 1.000 |
| Moral | 1.000 | 1.000 | 1.000 | 1.000 |
| Belief System | 1.000 | 1.000 | 1.000 | 1.000 |
| Narrative Strategy | 1.000 | 1.000 | 1.000 | 1.000 |

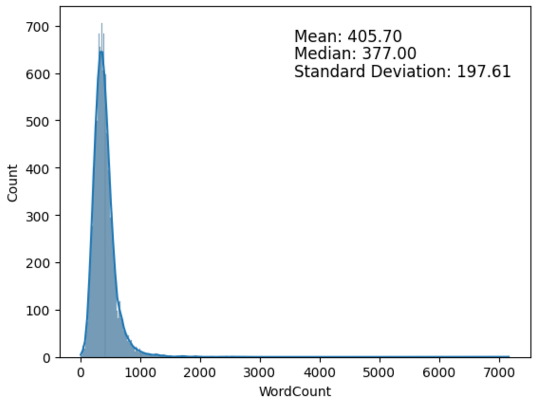
**Table A.2:** All Results for Gemini 1.5 Pro Runs 1 and 2 with Temperature 0

# A.7. Air Pollution Policy Research

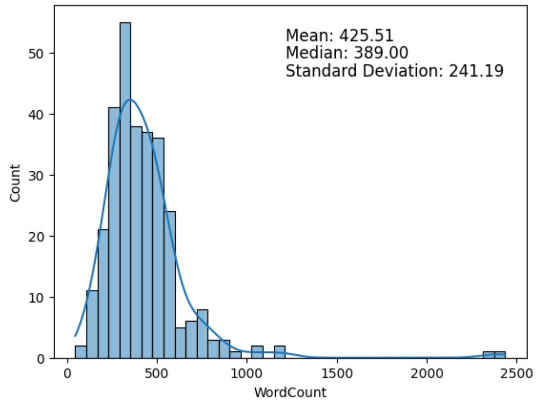
Research on air pollution in India has predominantly focused on the Indo-Gangetic Plain and metropolitan areas, often neglecting other regions (Gulia et al., 2022). While policy initiatives are evolving, there remains a significant gap between legislative actions and the required measures to meet air quality targets. Recommendations for bridging this gap include developing comprehensive air quality measurement networks, enhancing data analytics capabilities, and fostering robust science-policy interfaces (Ravindra, Sidhu, Mor, John, & Pyne, 2016).

Previous studies have systematically analyzed narratives within Indian news media, revealing topics such as the health risks of air pollution, urgent calls for action, and the frequent attribution of pollution to vehicular emissions (Murukutla, Negi, Puri, Mullin, & Onyon, 2017). Additionally, international events have influenced the prominence of air pollution in media coverage (Murukutla et al., 2017). Notably, a study using the NPF to analyze narratives in Delhi highlighted the role of government as a central actor, often depicted as both a hero and a villain (Olofsson et al., 2018).

**Aggregating Data:** Many NPF studies aggregate raw counts of narrative elements within individual documents and standardize these counts by the length of the narrative. This is how the distribution of the length of text looks like in the policy narrative documents in the census and the sample.



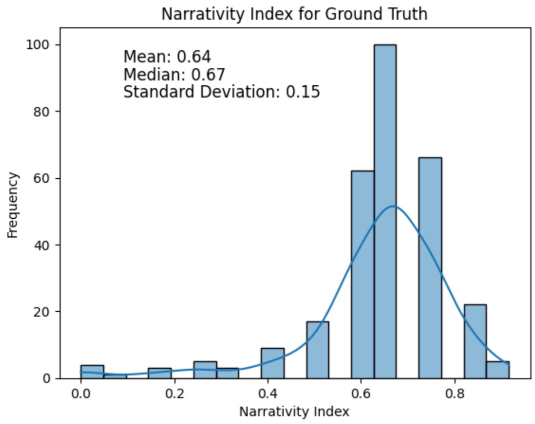

**(a)** Word Count Distribution in the Census



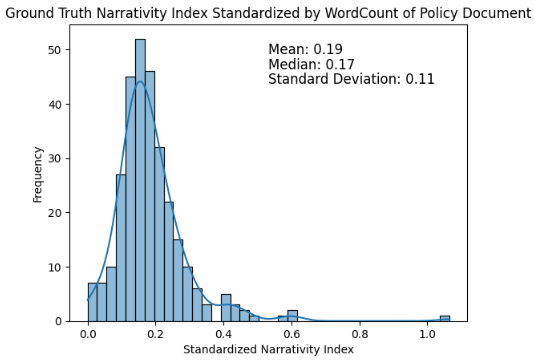

**(b)** Word Count Distribution in the Sample

**Figure A.2:** Word Counts in the Census and Sample

The narrativity index was created for each coded dataset. Narrativity index is a measure of the presence of narrative components in the policy narrative. It was calculated by dividing the number of narrative components present in the text by the total possible narrative components in the codebook. It is usually also normalized by word count of the policy narrative.




**(a)** Narrativity Index of Manual Coding




**(b)** Narrativity Index of Manual Coding - Standardized by Word Count

**Figure A.3:** Census-Sample Distributions

# A.8. Confusion Matrices

## A.8.1. Plot

| Plot | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 19 | 278 |
| Blame the Victim | 0 | 2 | 3 | 292 |
| Change-Is-Only an Illusion | 1 | 2 | 5 | 289 |
| Conspiracy | 0 | 0 | 1 | 296 |
| David vs. Goliath | 0 | 0 | 12 | 285 |
| Hero's Journey | 0 | 0 | 13 | 284 |
| Pendulum Swing | 0 | 0 | 3 | 294 |
| Rebirth or Renewal | 0 | 0 | 2 | 295 |
| Regulatory Enforcement | 10 | 0 | 49 | 238 |
| Restoration | 3 | 4 | 28 | 262 |
| Story of Decline | 38 | 133 | 4 | 122 |
| Story of Helplessness and Control | 6 | 40 | 17 | 234 |
| Stymied Progress | 5 | 13 | 14 | 265 |
| Triumph Over Adversity | 7 | 29 | 13 | 248 |
| Warning Tale | 4 | 0 | 40 | 253 |
| Total | 74 | 223 | 223 | 3935 |

**Table A.3:** Confusion Matrix - Manual Coding and Run 2

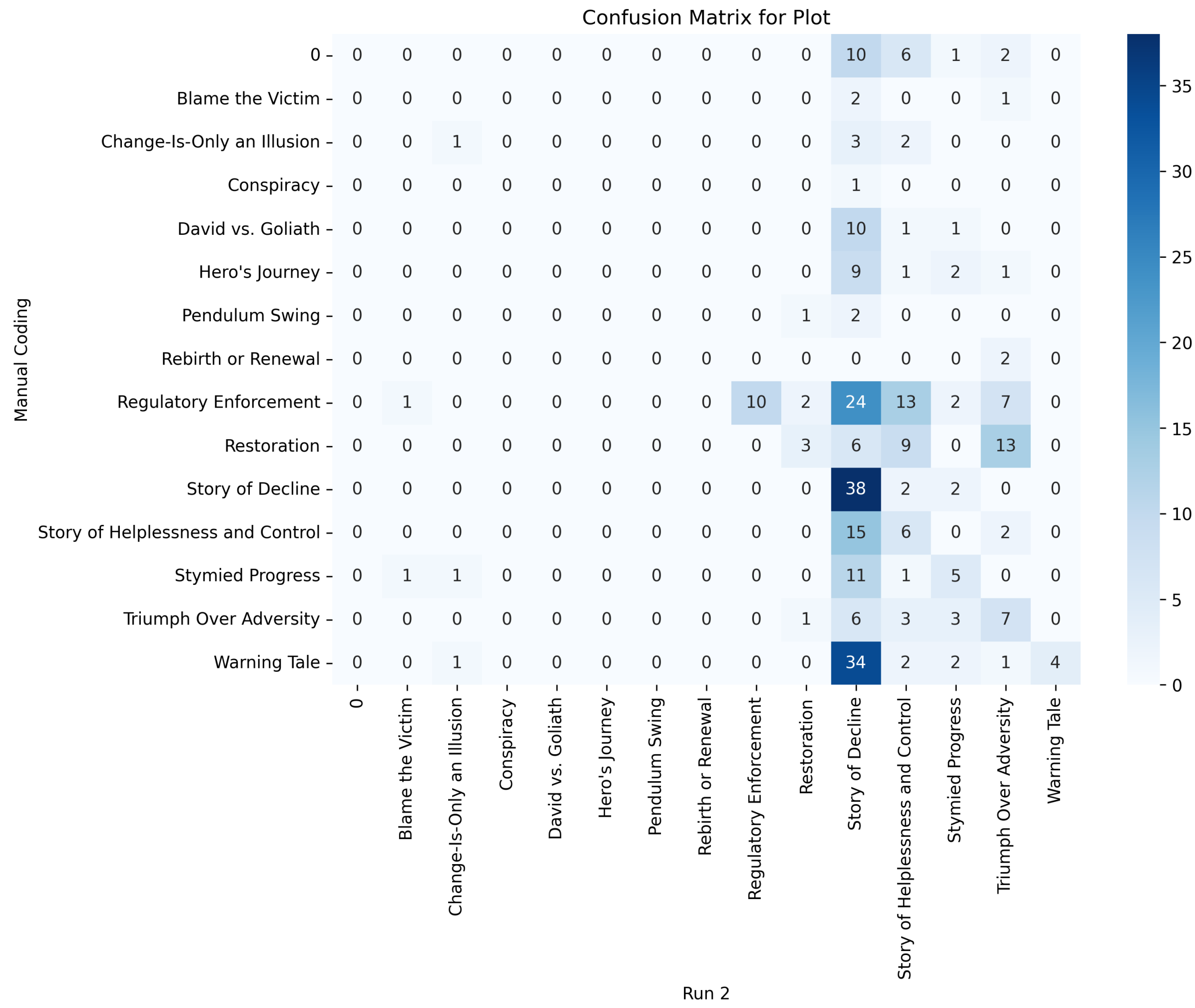



**Figure A.4:** Instances of Coding: Manual vs Run 2

## A.8.2. Moral

| Moral | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 5 | 292 |
| Explicit Policy Solution | 57 | 12 | 102 | 126 |
| Implicit Policy Reference | 121 | 107 | 12 | 57 |
| Total | 178 | 119 | 119 | 475 |

**Table A.4:** Confusion Matrix Between Manual Coding and Run 2

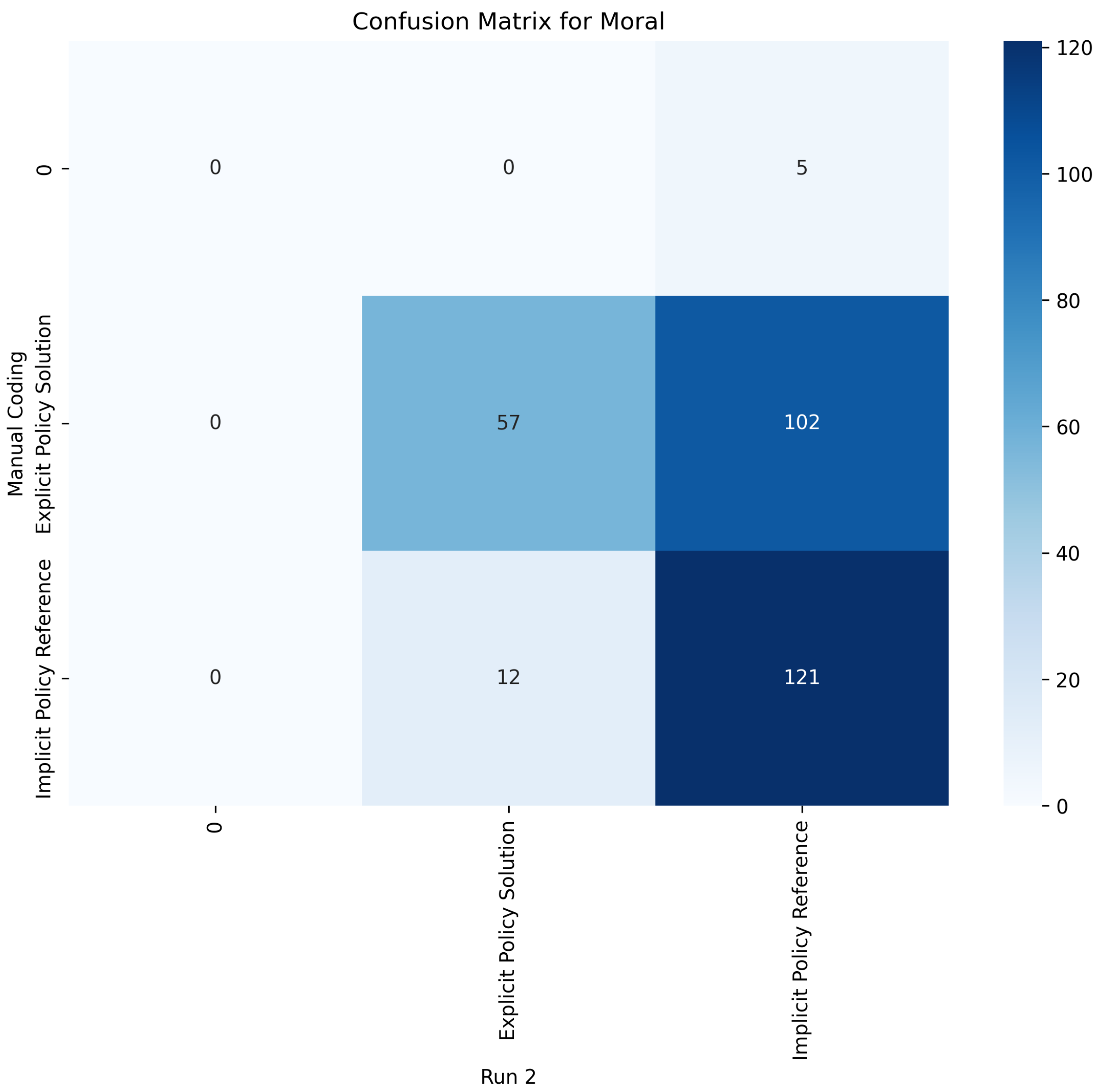



**Figure A.5:** Instances of Coding: Manual vs Run 2

### A.8.3. Belief System

| Belief System | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 21 | 276 |
| Activist | 0 | 0 | 25 | 272 |
| Egalitarian | 22 | 98 | 7 | 170 |
| Fatalist | 0 | 0 | 26 | 271 |
| Hierarchist | 63 | 30 | 71 | 133 |
| Individualist | 0 | 2 | 8 | 287 |
| Modernist | 19 | 47 | 17 | 214 |
| Technocratic | 2 | 12 | 2 | 281 |
| Traditionalist | 1 | 1 | 13 | 282 |
| Total | 107 | 190 | 190 | 2186 |

**Table A.5:** Confusion Matrix Between Manual Coding and Run 2

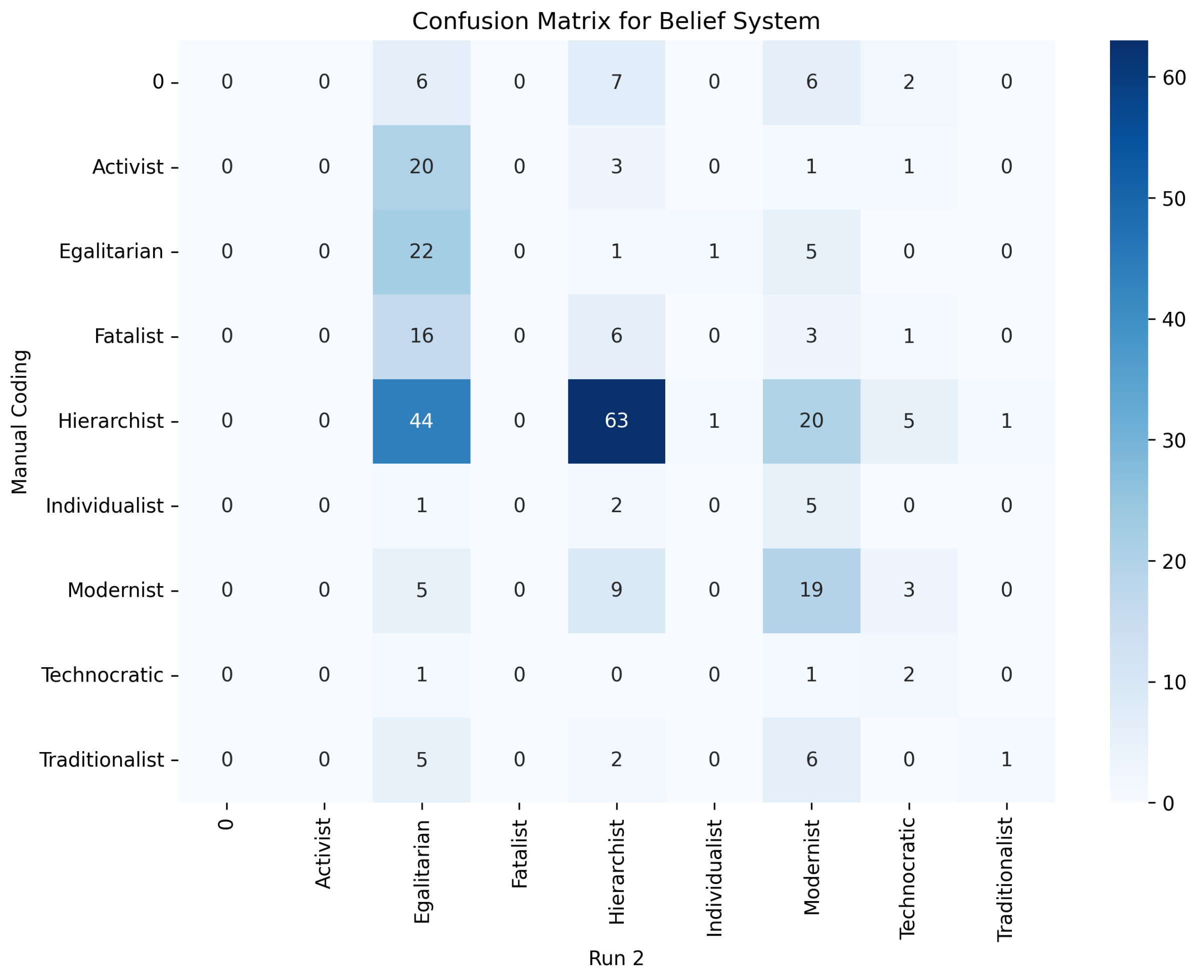



**Figure A.6:** Instances of Coding: Manual vs Run 2

## A.8.4. Narrative Strategy

| Narrative Strategy | True Positive (TP) | False Positive (FP) | False Negative (FN) | True Negative (TN) |
|---|---|---|---|---|
| 0 | 0 | 0 | 27 | 270 |
| Accidental | 0 | 1 | 8 | 288 |
| Angel Shift | 4 | 9 | 10 | 274 |
| Concentrating Costs and Diffusing Benefits | 7 | 19 | 58 | 213 |
| Demobilization of Support | 0 | 2 | 1 | 294 |
| Devil Shift | 0 | 2 | 15 | 280 |
| Diffusing Costs and Concentrating Benefits | 0 | 4 | 3 | 290 |
| Inadvertent | 0 | 0 | 3 | 294 |
| Intentional | 0 | 1 | 3 | 293 |
| Mechanical | 0 | 1 | 29 | 267 |
| Mobilization of Support | 120 | 126 | 9 | 42 |
| Warning Tale (future impacts of inaction) | 0 | 1 | 0 | 296 |
| Total | 131 | 166 | 166 | 3101 |

**Table A.6:** Confusion Matrix Between Manual Coding and Run 2


**Figure A.7:** Instances of Coding: Manual vs Run 2

## A.9. Workflow Proposed for Iterative Refinement of the Codebook by Pangakis et al.:

1. Creating task-specific instructions (a codebook).
2. Human annotation of a subset of text samples using the codebook.
3. LLM annotation of the same subset, followed by performance evaluation.
4. Refining the codebook based on initial results and repeating the validation if necessary.
5. Testing the final LLM performance on the remaining human-labeled samples.

## A.10. Use of Large Language Models in this Thesis

Since the study is about LLMs themselves, this thesis extensively used LLMs (Google Gemini and OpenAI ChatGPT) for formatting in Latex, code generation for Python, and image generation for the cover.