Cultural Analytics and Environmental News: A Novel LLM Approach to Cross-Cultural Understanding of Narrative Morals

Haiqi Zhou

A thesis submitted to McGill University
in partial fulfillment of the requirements of the degree of
Masters of Arts

Department of Languages, Literatures, and Cultures
McGill University, Montreal
August 2024

Table of Content

Abstract

This study investigates the use of Large Language Models (LLMs) like GPT-4 to uncover and analyze latent narrative messaging within environment-related news articles from North American and Chinese media. With the definition of "narrative messaging" as the intrinsic moral or lesson embedded in a story, LLM-aided moral extraction is applied to a dataset comprising approximately 15,000 news articles in English and Mandarin, categorized by climate-related themes and ideological alignments. The analysis of extracted results spans three comparative dimensions: North American versus Chinese-language sources, Liberal versus Conservative perspectives, and state (mainland China) versus offshore (Hong Kong and Taiwan) sources. The findings elucidate marked differences in how agents and narrative values are portrayed across diverse cultural and ideological contexts. For instance, North American media tend to highlight individualistic and crisis-driven themes, whereas Chinese media emphasize narratives centered on development and cooperation. This research underscores the capabilities of LLMs in dissecting and influencing climate and environmental communication, providing fresh perspectives on the collective belief systems that frame public discourse on environmental issues in varying cultural settings.

Cette étude examine l'utilisation de modèles de langage avancés (LLMs) tels que GPT-4 pour découvrir et analyser les messages narratifs latents dans les articles de presse relatifs à l'environnement issus des médias nord-américains et chinois. Avec la définition du "message narratif" comme la morale intrinsèque ou la leçon intégrée dans une histoire, l'extraction morale assistée par LLM est appliquée à un ensemble de données comprenant environ 15 000 articles de presse en anglais et en mandarin, catégorisés par thèmes liés au climat et

alignements idéologiques. L'analyse des résultats extraits s'étend sur trois dimensions comparatives : les sources nord-américaines versus chinoises, les perspectives libérales versus conservatrices, et les sources étatiques (Chine continentale) versus offshore (Hong Kong et Taïwan). Les résultats mettent en lumière des différences marquées dans la manière dont les agents et les valeurs narratives sont présentés à travers différents contextes culturels et idéologiques. Par exemple, les médias nord-américains ont tendance à mettre en avant des thèmes individualistes et axés sur la crise, tandis que les médias chinois soulignent des récits centrés sur le développement et la coopération. Cette recherche souligne les capacités des LLMs à disséquer et à influencer la communication sur le climat et l'environnement, offrant de nouvelles perspectives sur les systèmes de croyances collectifs qui encadrent le discours public sur les problèmes environnementaux dans différents contextes culturels.

Acknowledgments

I am profoundly grateful to my supervisor, Professor Andrew Piper, whose guidance and unwavering support have been pivotal throughout the course of this project. I extend my thanks to David Hobson for his meticulous efforts in the validation process of this thesis and to Professor Derek Ruths for his feedback on the research design. I also owe a debt of gratitude to the dedicated team of student annotators whose diligent work in validating the extractions from GPT and managing the dataset has been essential. Special thanks are due to Hye-jin Juhn, a librarian at McGill University, and the customer support team at Factiva Dow Jones, both of whom provided expert advice and assistance in accessing the necessary data. Lastly, my appreciation goes to Ashley Weng, whose insights into Taiwanese culture were immensely helpful. This thesis has truly been a collaborative endeavor, and I am thankful for the support and expertise of each individual involved.

Contribution of Authors

Haiqi Zhou, the author of this thesis, leads several key components: conducting a comprehensive literature review in Chapter One, researching available data, coordinating with Chinese-speaking annotators for Chinese-language news, and designing and validating the agent type prompting workflow in Chapter Two. Additionally, Zhou oversees the analysis of results throughout Chapter Three.

David Hobson, a PhD student in the Department of Computer Science at McGill University, is responsible for the validation process outlined in Section 2.4 of this thesis. He has developed the validation setup using Amazon Mechanical Turk, created the tables for this section, and analyzed the validation results.

List of Tables

List of Figures

List of Abbreviations

AMT	Amazon Mechanical Turk
ATU	Aarne-Thompson-Uther
GPT	Generative Pre-trained Transformer
LLMs	Large Language Models
MFT	Moral Foundations Theory
NLP	Natural Language Processing
NLU	Natural Language Understanding
TSMC	Taiwan Semiconductor Manufacturing Company

Introduction

One would rarely doubt the existence of common patterns guiding the way stories are told. Ever since ancient Greece, scholars like Aristotle have started to approach poetry in a scientific manner, identifying types and establishing overarching principles of literary experiences. Already, thousands of years ago, we have character archetypes that are perceived to embody the universality of human nature, like Aristotle's tragic hero. Over time, these archetypes have evolved, and today, our repertoire includes not just character-based archetypes but also those rooted in plot structures such as the hero's journey, and genres like romance and science fiction, commonly classified under genres. Our story schemas stem from the fact that humans tell collective stories that share very similar narrative characteristics, structures, or values that reflect our common perception of the world and our shared experience. However, to systematically capture these abstractions that we call narrative archetypes or story schemas, our approach has predominantly relied on predefined categories, and tends to focus on narrative facts ("what happened") and overlook overarching narrative values ("why it was told"). It is one thing to say that climate change is often talked about alongside carbon emissions, it is a whole different thing to say that these discussions are, for example, centered around the glorification of the government. And it is precisely with this element of "why" that we get to understand narratives as something embedded in social and cultural context, reflecting beliefs, norms, and overall sentiment.

This thesis aims to forge a comprehensive understanding of the narrative messaging behind environmental news through a novel approach that leverages Large Language Models

(LLMs) to extract narrative schemas, here more specifically referred to as "morals." Unlike traditional methods that rely on pre-existing, top-down categories, the methodology employed in this research *extracts* morals, which I define as *a general lesson that the narrator wishes to impart to the audience about the world*. Specifically, the term "moral," linked with an ancient narrative tradition, often serves as a guiding or instructive element within a story. This tradition spans various cultures, notably seen in Aesop's Fables in Western culture, as well as in Hindu (Panchatantra) and Buddhist (Jatakas) traditions dating back to around the fifth century BCE. For example, the moral of the story "The Little Red Riding Hood" can be summarized as the importance of not trusting strangers. While we might think that not every story comes with a moral, we'd likely agree that every story imparts some central "lessons" – a more profound message that the storyteller aims to share, whether consciously or not. Therefore, simply put, a moral is the higher-level lesson that a story conveys, which is more general and abstract than the story's contents. It focuses on the "why" questions, such as why the story was told, and answers what lesson or point the story aims to convey, relying on summarizing key elements without being a summary itself.

Historically, narrative schemas were manually summarized, involving scholars who selected and categorized key narrative elements, as exemplified by Christopher Booker's identification of the seven basic plots. However, the rise of computational narratology has prompted a shift from conventional top-down methodologies to a bottom-up approach in analyzing narrative structures, facilitated by the ability of computers to process large text corpora. This transition has shed light on the patterns and peculiarities of shared narratives, revealing the diverse angles from which narratives can be explored, including common plot arcs in literary works and collective narrative shifts on social media (Matthew Jockers; Zhao et al.).

Despite these advancements, the identification of narrative archetypes has yet to be automated with the use of large language models (LLMs). With their prowess in narrative understanding and generation, LLMs open up new avenues for annotating and analyzing vast collections of texts – tasks that previously required human annotators – and offer fresh perspectives on understanding narratives and, by extension, culture itself. As such, this thesis contributes to the field of computational narratology by proposing a multidimensional approach of moral extraction that acknowledges the complexity of collective narratives as carriers of cultural, ethical, and social significance, thus enriching the field and expanding its scope to include more holistic interpretations of texts.

Although the proposed methodology is applicable to a range of texts, including novels and folktales, this paper specifically explores morals within environmental news narratives. The rationale for choosing this specific object of analysis is twofold. Firstly, why the environment? With the increasing severity of anthropogenic impacts on the planet, the repercussions are found around the world in the form of catastrophic flooding, frequent and intense hurricanes, droughts, and resulting mass displacements and famines. It is thus imperative to understand the discourses surrounding environmental issues as they shape public and political decision-making, drive educational initiatives, inspire behavioral changes, and facilitate global collaboration on pressing issues like climate change and pollution.

This leads to the second question – why the news? The news media is selected for its critical role in shaping public perception, particularly through the effects of news framing, which will be discussed further in the first chapter. Although news is often perceived as objective, it is still undeniably a kind of narrative. Analyzing these narratives' implicit morals thus involves uncovering the subjectivity in reporting – what is reported, how it is written, and the conveyed

messages. Historically, environmental communication studies have focused on the concept of framing, as defined by Robert Entman in 1993, which involves emphasizing certain aspects of reality in texts to promote specific interpretations and solutions. However, this thesis positions morals as a construct both similar to and distinct from framing. Both morals and framings are a higher-level pattern that governs how narratives are understood and influence audience perceptions by shaping the interpretation of an event. However, while framing concerns itself with the presentation of issues, morals explore the ethical implications and lessons embedded within narratives, emphasizing the behavioral values each story promotes. As such, the scrutiny of environmental news morals provides deeper insight into the narrative messaging from various cultural and ideological perspectives, in the meantime also enhancing our understanding of news framing effects.

In this thesis, I analyze morals extracted from approximately 15,000 news articles from 2023, sourced from representative media outlets in mainland China, Hong Kong, and Taiwan (Chinese language news), as well as the United States and Canada (English news). These regions were selected because they are among the top contributors to global greenhouse gas emissions, which provides a pertinent backdrop for this study. In addition, Chinese sources provide an interesting dimension to understanding the variation of morals in and out of a censorial context and allow for potential comparison across cultures and languages. The articles, selected based on keywords such as "climate change," "pollution," "carbon emissions," and "sustainability," attempt to capture critical environmental issues that reflect current global challenges and priorities in environmental conservation and policy.[1] Consequently, the dataset of this study will allow for research questions like – What narrative surrounds climate change? How do morals differ between regions and ideologies? How are different agents represented? – which helps

[1] The rationale behind the choice of keywords will be explained in the second chapter.

explore how media discusses environmental issues, examining underlying sentiments and the ultimate objectives of these narratives.

This thesis is divided into three chapters to define, set up, and analyze the moral-based analysis for understanding the cultures of storytelling in environmental news. The first chapter starts by examining the significance of narrative classification for organizing knowledge and providing insights into the worldviews that guide the production and reception of stories. Here, I look at the history and inherent limitations of predefined textual systems through several well-known constructs, including genres, archetypes, and the Aarne-Thompson-Uther (ATU) index for folktales. I then present some existing literature in the field of computational narratology and Natural Language Processing (NLP) that serves as a catalyst for the conception and design of this research. As I will demonstrate, these previous studies unveil new perspectives on the dynamics of collective narratives but remain insufficient due to the lack of a generalizable methodology that takes into account underlying beliefs. This is why I then present some of the newest studies of text analysis using LLMs to justify the choice of tool and propose a study of narrative archetypes using morals, which can be extracted automatically only through the use of LLMs. After an extensive definition of morals and the role they play in this research, the chapter segues into an examination of media studies and the news framing effects, demonstrating how different framings of news significantly influence public perception. As mentioned above, morals and framing are conceptually similar in the sense that they both recognize the presence of underlying structures that influence how information is presented and interpreted. The effect of framing underscores the insufficiency of the aforementioned systems that solely rely on narrative facts and justifies the use of morals as the higher-ordering principle. The first chapter ends with an exploration of related literature in environmental media and communication studies, including

what scholars have garnered from the intersection of media, communication, and environmental topics as well as the limitations of these studies.

Chapter Two demonstrates in detail the processes of moral extraction from large corpora of texts with the use of LLMs. Specifically, the research setup is broken down into six sections. The first section presents the data's source, detailing its size and features while also characterizing the various news outlets. The second section focuses on the reasons for choosing GPT for moral extraction. The third section describes the prompting design – what questions we give to the model and why we ask these questions based on narratological frameworks. The fourth section demonstrates the human and automated validation processes to determine the feasibility of using LLMs as an agent in large-scale cultural analysis, which involves a comparison between the responses to our list of questions from human annotators and from GPT-4. In the fifth section, an enhanced prompting design is introduced and detailed to specifically address the main agents in the news stories. The final section addresses the limitations of each step of this research design and the possibility of future improvements.

In the concluding chapter, I discuss in length the results of moral extraction and undertake a comprehensive comparative analysis of narratives that surfaced from each news outlet on three levels, i.e. differences between North America and Chinese-speaking regions, between state- and non-state media in the context of Chinese-language news, and between Conservative and Liberal sources in North America. This analysis is built upon the examination of insights from media studies, scrutiny of each outlet's historical trajectory and political leanings, and a contextualization within recent significant environmental events. The analysis examines three aspects: the valence associated with different agent types, the roles of agents as archetypes (such as hero, villain, or victim), and the underlying morals presented. In order to

offer potential explanations for observed discrepancies in morals, I synthesize insights gleaned from prior research exploring distinctions in reporting styles between, for example, Liberal and Conservative media entities, as well as between mainland China and Chinese-speaking regions with differing degrees of press freedom, all while being aware of the degree to which the findings can be generalized. Furthermore, the insights gleaned from moral extractions are juxtaposed with traditional text analysis methods like topic modeling to highlight the benefits of using LLMs for moral extraction. It's important to note that this thesis adopts a comparative and exploratory approach, emphasizing a bottom-up discovery process rather than hypothesis testing. This is due to the innovative use of moral perspectives as a lens for narrative analysis. Consequently, this methodology lays the groundwork for future research to concentrate on formulating and testing additional hypotheses.

The findings of this thesis reveal a range of narrative emphases: North American media often highlight individualistic and crisis-driven themes, whereas Chinese sources tend to focus on developmental and cooperative narratives. Liberal media typically emphasize the impacts of climate issues on people, while Conservative outlets are more likely to discuss market dynamics alongside the environment. Additionally, state-controlled Chinese media generally adopts an international perspective, in contrast to offshore Chinese media, which prioritizes community-centered narratives and places a stronger emphasis on individual behaviors. Through analyses of this kind, this study offers a nuanced understanding of the intricate interplay between media representation, political context, and environmental discourse, contributing to broader conversations surrounding narrative messaging and its impact on public perception and policy formulation.

Overall, this research endeavors to illuminate the intricate tapestry of storytelling by leveraging the capabilities of LLMs to dissect and understand narratives on a more profound level. It challenges the traditional reliance on predefined categories by introducing a novel approach focused on extracting "morals" that capture the core lessons or messages intended by the storyteller. This method transcends mere narrative elements to explore the deeper, often implicit, values and sentiments that stories convey, situating narratives within their broader social and cultural contexts. By applying this methodology to the analysis of environmental themes in news narratives across diverse linguistic and geopolitical landscapes, the thesis contributes to the fields of computational narratology and cultural analytics as well as showcases the versatility and depth of LLMs in extracting nuanced insights from vast text corpora. It provides insights into how narratives, particularly those related to pressing global issues like climate change, are framed and disseminated by public media. Through its multidimensional analysis, this research aims to enrich our understanding of the intricate ties between narrative, culture, and ethics, and offers a comprehensive framework for examining the morals underpinning collective stories and their impact on our worldviews and actions.

Chapter One: Environment News Through a New Lens

In this thesis, I present a method of leveraging LLMs for the extraction of morals to understand the manifestations of cultural and ideological differences in environmental news. But first, what are morals? Here, I define morals as *a general lesson that the narrator wishes to impart to the audience about the world*. Linked with an ancient narrative tradition, the term "moral" often serves as a guiding or instructive element within a story. To establish a clearer definition of morals, I summarize its characteristics as the following:

1. A moral is the higher-level lesson that a story conveys.
2. A moral is more general and abstract than the contents of the story.
3. A moral is not a summary, but does rely on summarizing key elements.
4. A moral focuses on why questions (why was this told?) not what questions (what happened?)
5. A moral answers the question: "The lesson that this story wishes to convey is...." or "The point of this story is to convey…"

In essence, every story carries an underlying message, whether conveyed explicitly or subtly, which the storyteller aims to impart to influence or affirm the audience's beliefs or objectives. This element is an integral part of the narrative "schema," influencing both the storytelling approach and the facets of life that the narrator chooses to emphasize. In the first chapter, I aim to address three main questions surrounding the design of this research: Why morals? Why LLMs? Why understand environmental news through LLM-extracted morals? The first section lays the foundation for a moral-based analysis of storytelling cultures in

environmental news, starting with an examination of narrative classification systems such as genres, archetypes, and the Aarne-Thompson-Uther (ATU) index for folktales (Section 1.1). While existing narrative classifications offer a useful foundation, they often fall short of capturing the nuanced intricacies inherent in narratives. This is why, in Section 1.2 (The Age of Computational Analysis), I discuss relevant literature in computational narratology and Natural Language Processing (NLP) that informs the design of this research. It then explores the use of Large Language Models (LLMs) for analyzing complex narrative information, highlighting the unique capabilities of LLMs for narrative analysis. Section 1.3 (Morals, News Framings Effects, and Environmental Communication) delves into the core of this research by justifying the use of morals for the analysis of environmental communication. It posits morals as a concept related to media studies and news framing effects, illustrating how different framings impact public perception and emphasizing the conceptual similarities between morals and framing. The section concludes with a review of environmental media and communication studies, discussing their findings and limitations. As such, in the first chapter, I aim to establish the theoretical groundwork for my project by delving deeper into narrative messaging. Insights from this literature review and the proposal of the methodology are intended to benefit those concerned with climate change, offering innovative methods for studying environmental communication.

1.1 Predefined Textual Systems

We live in a world of textual systems – books are divided into genres like romance and science fiction, news articles are separated by categories such as politics and finance, and our education, governments, and society at large are all structured by different systems. These mechanisms of classification help us orient ourselves in a rather chaotic world. For instance, when we recommend a book to our friend, we inevitably bring up the genre of the book, so that

our friend can get a better grasp of its content. Narrative classification systems like genres organize narratives in a certain way, often through themes, forms, characters, or plots, and these structures, in turn, reflect what is inherent in our shared values and common worldviews. Simply put, by looking at the way we group knowledge together, we can learn something about the functioning of our society and what we, as a collective species, care about. However, we often take for granted the systems that have come to shape our everyday lives when they are anything but uncontested grounds.

To most readers today, genres have become frameworks entrenched in their heads. Like it or not, the genre system has entered our everyday habit of reading, from the organization of books at libraries and bookstores to the way we talk about books. However, as much as we are used to the system now, the concept of genre is constantly evolving and a long history of debates has yet to yield a consensus on the topic (Genette). To further complicate this, genre is not the only generic structure that has come to delineate groups of texts. It frequently intersects with, and is discussed in relation to, other concepts such as archetypes (Jung; Frye), scripts (Schank and Abelson), schemas (Chambers and Jurafsky), tale types (Uther), motifs (Thompson), and functions (Propp). These concepts, though varied in their applications and origins, share a common thread: they serve as implicit structures that guide both the creation and interpretation of stories. Specifically, they suggest that storytelling is not just a creative act but also one deeply embedded in psychological, cultural, and structural patterns that transcend individual tales. This interconnectedness aids in the creation of narratives by providing a shared set of expectations between storytellers and audiences and by offering a framework through which narratives can be understood and related to broader human experiences. Let’s take a closer look.

In the list of concepts above, one of the terms with the longest history is "archetype", which originates from ancient Greek and can be traced back to Plato's Theory of Forms or Ideas, in which he describes a form (*lexis*) as the archetype, or perfect model, of the attributes it embodies. The word itself is composed of "arche" (ἀρχή), meaning "origin" or "beginning," and "typos" (τύπος), meaning "type" or "pattern." Hence, archetype essentially refers to an original model from which other models or types derive. This concept underlies all the previously-mentioned systems and is also an idea that we shall pursue in this project. In a more modern context, archetypes have been widely used in both psychological and literary domains. Northrop Frye, in 1957, looked at myths as the very first step of studying archetypes. He regarded myths as "the most abstract and conventionalized of all literary modes" similar to the way religious Byzantine paintings show the peak in structural stylization (134). Specifically, Frye asserts that the structural principles of literature are tied to the natural cycles of the Earth, including the changing of seasons. As such, one of the central arguments in *Anatomy of Criticism* is that the foundational rules of literature should be understood through the lens of archetypal and anagogic criticism, which allows for the consideration of literature as a whole. Similarly, albeit in a psychological setting, Carl Jung ventures into the realm of "the universal images that have existed since the remotest times" in his 1959 book *The Archetypes and the Collective Unconscious* (5). Jung proposed several archetypes, including the persona, the shadow, the anima/animus, and the self, which are universal psychic predispositions that constantly emerge in dreams, myths, art, and literature, and play a critical role in shaping human behaviors and beliefs across different cultures and historical periods. Therefore, both Frye and Jung conceptualized archetypes as universal, symbolic patterns that transcend individual experiences, manifesting in

the collective unconscious and literature, respectively, to reflect fundamental aspects of the human condition.

There is one key distinction to make in the discourse around archetypes: unlike Plato's theory of the perfect form, which comes from the inside and cannot be verified, Jung based his works on empiricism, using experience and observation as the method of verification. According to Jung, since the arrival of empiricism, "the Idea (archetype) is no longer something a priori, but is secondary and derived" (76). In a way, both Northrop Frye and Carl Jung engage in the *derivation* of archetypes through methods that are analytical and interpretive in nature. The act of tracing patterns to extract archetypes is manifested, to varying extents, in the different narrative classification systems. In order to understand the organizing principles of folktales, the Aarne-Thompson-Uther (ATU) index was developed, categorizing folklore by its motifs and narrative structures. Similarly, to classify books according to themes and styles, the genre system evolved through time, facilitating the grouping of literature into distinct categories like science fiction and mystery. In addition, one of the most direct literary applications based on Jung's idea of archetypes is Christopher Booker's *The Seven Basic Plots: Why We Tell Stories*, which provides a compelling framework by proposing that all narratives essentially follow one of seven basic story archetypes. These plots – including "The Quest", "Overcoming the Monster", and "Rags to Riches", among others – suggest a universal structure underlying storytelling across cultures and time periods. The methodologies of narrative classification, as illustrated, leverage the identification of patterns through exhaustive literature review – ranging from hundreds to thousands of stories across diverse cultures – and meticulous classification, thereby unraveling and systematizing the intricate tapestry of storytelling.

The concept of a higher-order principle underlying our narratives is intriguing. Still, one might wonder, are there truly only seven plot archetypes? Is it accurate to categorize folklore into numerous divisions and subdivisions, such as "Tales of Magic" and "Supernatural Adversaries," using the Aarne-Thompson-Uther (ATU) system? Are we doing justice to books and literature with the use of genres? Indeed, behind the many doubts that people might have, there are two main concerns surrounding the validity and universality of narrative classification. First, there is an inherent problem in employing predefined classes that are not taxonomic units to classify stories. Despite the fact that scholars have come up with their categories after rigorous research, these categories are barely structural units that serve as the make-up of a typology (Dundes). In other words, attempting to apply a top-down approach by fitting a narrative into a finite number of predefined categories inevitably leads to problems. What if a book belongs to multiple genres? What if it belongs to none? One needs only to look at the discourse around the "death of genre" to understand the limitations of similar structures. According to Jonathan Culler, "What we value in literature is the singularity of a literary work, and to expect it to conform to the conventions of a genre or to approach it through the lens of the genre is to aim at something other than its distinctive literariness" (41). Although all generic structures restrict meaning in the process of making meaning, predefined categories that fail to uncover the actual inner workings within a text corpus can be dangerous to use as a part of typological systems. Such categories not only impose more significant constraints on interpretation but can also sometimes lead to inaccurate representations.

Admittedly, we can still appreciate the usefulness of these categories, as they provide insights into the narrative traditions and shared themes of cultures. Like what Dundes argues: "True, it is always dangerous to use ready-made patterns since there is the inevitable risk of

forcing material into the prefabricated Procrustean pattern. However, this technique is justified if it aids in solving a problem." (96). In our case, ready-made patterns like genres and Frye's archetypes are patterns that reveal underlying structures of storytelling without the need to get into the problems of typology. For example, genre, often regarded as "fluid and tenuous constructions," can reflect the movement of human interests in literary forms (Vint and Bould 48). Nevertheless, this leads to a second problem. Suppose we simply consider these predefined categories as a way to extract cultural information, how do we verify whether or not these systems accurately capture what they claim to capture? Going back to the questions surrounding Christopher Booker's seven plots, in which cultures and to what extent can these plots generalize? Systems that rely heavily on human readers inevitably encounter cultural biases, temporal constraints, and scalability issues. In light of these challenges, the emergence of computational narratology and computational literary studies in the late 20th century has revolutionized our approach to analyzing narratives. These advancements allow us to process vast corpora of texts and determine the scope of their representativeness, which provides new perspectives on literary and cultural analysis. In the next section, I will explore the text-related methodologies and technological advancements of the computer age that have significantly influenced this research.

1.2 The Age of Computational Analysis

Early-stage applications of computer-aided information retrieval from texts involve processes such as Natural Language Understanding (NLU). In 1977, in the quest to comprehend the architecture and expectations of knowledge embedded in narratives, Roger Schank introduced the concept of "scripts," which aids in identifying and understanding the predictable

sequences of events and roles that characters typically navigate in stories and real-life situations. Some typical scripts include the restaurant script, the shopping script, and the travel script, all of which represent common situations with a sequence of expected events. Echoing the mechanisms of previously discussed systems, script theory represents a search for archetypes of scenarios in narratives, an endeavor that has also advanced the ability of computers to comprehend the structuring of narratives.

Over the past two decades, the achievements of deep learning models have significantly advanced the field of literary analysis. We have seen the evolution of models like transformers that achieve near-human performance or even perform better than humans in many tasks, including sentiment analysis, topic modeling, question answering, and text classification. This rapid progression allows scholars to ask more sophisticated questions, such as understanding the intricate dynamics of character interactions, mapping narrative structures, and identifying key story features. Furthermore, the integration of visualization techniques also enhances our ability to interpret and explore complex literary datasets. Thanks to the advancements in Natural Language Processing (NLP) and other computational tools, we can now analyze previously unfathomable large text corpora, enabling us to rethink literary studies and verify existing narrative classification systems.

Specifically, the drastic increase in the scope of text that can be analyzed has given rise to the concept of distant reading, introduced by Franco Moretti in 2000 and later expanded upon in his 2013 book. Distant reading revolutionizes the exploration of textual information by focusing on large-scale patterns and trends across extensive datasets rather than traditional close reading methods. In his book, Moretti describes distant reading as follows:

> Distant reading: where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes-or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something. (48-9)

Of course, there has been criticism surrounding the concept of distant reading, where the disappearance of the text becomes problematic for some. Nevertheless, we have seen fruitful results regarding what distant reading can do in and beyond the realm of the text itself. From the gender representation of comic book characters (Amanda Shendruk) to the characteristics of best-selling books (Archer and Jockers), we have gained incredible insights into the way we tell stories through the lens of content, form, and styles, which, in turn, reflects what is inherent in our society – gender inequality, cultural differences, the stories that attract us, and so on. Expectedly, some scholars have also ventured to verify Booker's seven plots using digital methods. Researchers at the University of Vermont's Computational Story Lab, for instance, have applied sentiment analysis to thousands of English novels and found out that there are mainly six archetypal plots, namely, Rags to Riches (rise), Riches to Rags (fall), Icarus (fall-rise), Oedipus (rise-fall), Cinderella (rise-fall-rise), and Man in a Hole (fall-rise-fall) (Reagan et al.). Whereas Christopher Booker's seven plots are more focused on the thematic content and structure of the stories, such as Overcoming the Monster, Voyage and Return, and Rebirth, the computer-extracted emotional plot arcs show the sentiments associated with the experience of the protagonist and have even been used to reveal the common plot structures of bestsellers – best-selling books tend to be characterized by rhythmic rises and falls (Archer and

Jockers). Although these emotional plot archetypes resonate with and, in some respects, corroborate Booker's seven basic plots, we can still say that the evolution of computational analysis marks a significant leap in our understanding of narrative forms.

Here, we can return to the broader concept of narrative archetypes and examine how their study has evolved with the advent of computational analysis. Observing the vibrant landscape of our contemporary world, it's clear that the genre is far from obsolete. What has happened, according to Ted Underwood, is the resurgence of critical writings about genres in the past couple of decades, which is closely linked to popular culture and new critical theories or methods (3). In the last section, we mentioned the perception of genres as "fluid and tenuous constructions." Sherryl Vint and Mark Bould further elaborated on this concept in their chapter:

> Genres come into being "after the fact." Selected features of existing texts are eventually recognized, for a multitude of reasons, by discursive agents as collectively forming a genre. A key driving force in the creation of genre is the commercial impulse to reproduce textual characteristics which have sold well to the public and thus tend to guarantee a market for similar products. (48)

Specifically, it has become ever more convenient for cultural producers and platforms to target their audience with the collection of user data and the introduction of ever-finer categories, which, to an extent, explains the revival of genres highlighted by Underwood. Some notable examples given by Underwood that showcase the dynamics between computational cultural analytics and classification systems include the recognition of literary genres by computer-generated algorithms (Allison et al.), the reverse-engineering of Netflix's 77,000 microgenres (Madrigal), and the clustering and visualization of music microgenres on Spotify documented by Glenn McDonald's *Everywhere at Once*, all of which demonstrate the potential

of using computational methods to extract patterns and archetypes from metadata, texts, and all kinds of cultural productions.

At the end of the last section, we highlighted the limitations of predefined systems that rely heavily on human readers, such as cultural biases, temporal constraints, and scalability issues. We posed the question: how can we verify that these systems accurately capture what they claim to capture? The answer, we argue, lies in computational cultural analysis, which allows us to more precisely define the patterns we identify and determine the scope they can generalize to. Instead of requiring human annotators to read hundreds of books or stories, we can leverage computers to analyze thousands over extended periods. Furthermore, a key difference between traditional predefined classification systems and computationally derived genres is that the former employs a top-down approach, while the latter uses a bottom-up method. With a larger corpus of data, this bottom-up approach that applied automated extraction allows us to more effectively determine the extent to which we want to generalize our findings. In his paper "There Will Be Numbers," Andrew Piper discusses the problem of generalization: to what extent does the data represent the culture we claim they depict? By focusing on the "representativeness of our own evidence," Piper acknowledges that while the problem of generalization will always exist, computational tools give us the ability to address it (5):

> One of the key concepts operative in computational research that has so far been missing from traditional studies of culture is that of modeling. A model is a metonymical tool – a miniature that represents a larger whole. But it is also recursive in that it can be modified in relationship to its "fit," how well it represents this whole. There is a great deal of literature on the role of modeling in knowledge creation and this should become core reading for anyone undertaking cultural analytics. The more we think about our methods

> as models the further we will move from the confident claims of empiricism to the contingent ones of representation. (8)

This concept of modeling is particularly evident in the computer-aided genre studies we previously discussed, where computational modeling reveals how genres are not fixed taxonomies but are instead fluid and ever-changing. As Piper notes in his paper, "far from a space of universal knowledge, cultural analytics marks out more clearly the contingencies of knowledge," thus circumscribing the space of knowledge more accurately (8).

Although computational cultural analytics and literary studies have provided us with knowledge that was previously unobtainable, one can argue that there is only so much information that we gain from looking at sentiment plot arcs or topic modeling, however accurate they might be. Recently, however, the arrival of Large Language Models (LLMs) offers an opportunity to revolutionize text processing methods due to their advanced capabilities in analyzing textual data more thoroughly. Their text-generation capabilities have also reached unprecedented levels, enabling new forms of language understanding and interactive dialogue systems. For example, studies have shown that LLMs like ChatGPT outperform crowd-workers in text-annotation tasks, including relevance, stance, topics, and frame detection for tweets and news articles (Gilardi et al.). As well, ChatGPT has been found to demonstrate exceptional zero-shot sentiment analysis capabilities, often matching fine-tuned BERT and state-of-the-art models, effectively handling polarity shifts and performing well in open-domain scenarios (Wang et al.). More generally, researchers have examined the narrative understanding ability of LLMs, including their performance in narrative summarization and narrative question answering, emphasizing the crucial role of LLMs in advancing narrative understanding (Zhu et al.). Additionally, LLMs have undergone evaluations for theory of mind capabilities. The study

conducted by Strachan et al. reveals that GPT models excel in tasks involving beliefs, intentions, and non-literal expressions. Notably, GPT-4 has outperformed human benchmarks in tests focused on irony, hinting, and strange stories. This indicates that LLMs possess a profound ability to interpret narrative information and contexts at a sophisticated level. While there have been extensive debates about the extent to which machines can understand human languages, these studies have demonstrated the versatility of large language models and the diverse applications in which they can be employed.

Despite the advanced text analysis capabilities of large language models, few studies have thus far employed LLMs for cultural analytics or literary studies, with most focusing on the investigation of biases, regional opinions, and personality traits embedded within LLMs (Romero et al.; Miotto et al.; Lucy and Bamman; Durmus et al.). One research that applied LLMs to the study of cultural framing is Hamilton and Piper's "The COVID That Wasn't: Counterfactual Journalism Using GPT" published in 2022. The authors claim to be the first to use LLMs as a tool for "critically understanding the interpretation of events through media coverage or other forms of cultural framing." Specifically, they used GPT-2 (trained on data before 2020 and therefore without pre-existing knowledge of COVID-19) to generate alternative (counterfactual) news articles based on real headlines from the COVID-19 pandemic, comparing these with actual articles to assess differences in framing and attitude. This study thus demonstrates the potential of LLMs to analyze media coverage critically, offering a novel method for exploring "what-if" scenarios in news reporting and understanding media bias and editorial choices.

Inspired by experimentation of this kind, I decided to come up with the methodology of applying LLMs to the extraction of narrative archetypes. There remains a lot more questions that

we can ask about narrative archetypes, especially with the aid of computational tools. For example, what more can we learn about the patterns of storytelling and narrative messaging with the aid of LLMs? How can LLMs help us understand narrative archetypes differently? How can we automate the study of narratives using LLMs? In this study, we aim to propose a methodology that employs *distant close reading* – leveraging the narrative understanding capabilities of LLMs to analyze a large dataset and answer complex questions about narrative information that was previously coded only by human annotators. In a sense, unlike Moretti's definition of distant reading at the beginning of this section, *distant close reading* reappears the text and the deeper level information that was lost.

1.3 Morals, News Framing Effects, and Environmental Communication

Motivated by the concept of *distant close reading*, I conceived this research that centers around the extraction of morals using LLMs like GPT-4, which allows a deeper insight into the narrative value of large corpora of environmental discourse. As discussed in earlier sections, there are several compelling reasons to prioritize morals over other narrative schemas. Firstly, while subjective, story morals offer a valuable framework by encapsulating the narrative message intended by the author – a dimension not accessible through other schemas, such as genres. This focus on the overarching message allows us to transcend the literal words and content, exploring deeper story implications. Secondly, the advent of LLMs enables the automation of tasks previously performed manually, such as annotating story morals, which facilitates the extraction and analysis of morals on a large scale. Thirdly, analyzing narratives through morals adopts a bottom-up approach, enabling the distinct cultural characteristics, such as prevalent morals, to be identified and contrasted with those from different cultures or

ideologies. Together, these factors underscore the unique utility of morals as a tool to comprehend a variety of narratives.

It is useful here to clarify the concept of story morals a little more by examining the moral foundations theory (MFT), which was brought up by Jonathan Haidt and Craig Joseph in 2004. The MFT proposes that human moral reasoning is based on a set of innate psychological systems, or "moral foundations," which are shaped by both evolutionary processes and cultural influences. Examples of these moral foundations include "care/harm," "fairness/cheating," and "loyalty/betrayal." They have several characteristics such as nativism, meaning that there is a first draft of the "moral mind" and pluralism, which means that moral foundations are never singular. In this thesis, however, it is crucial to distinguish the concept of "story morals" from that of "moral foundations." Unlike moral foundations, which delve into the psychological underpinnings and origins of moral reasoning, story morals concentrate on the overarching messages of narratives – the central lessons or themes that the narrator aims to communicate. Like MFT, the strength of using story morals stems from its pluralistic approach, which welcomes a variety of lessons rather than relying on a predefined taxonomy that might not fully capture cultural behaviors. Nevertheless, the patterns or clusters we can identify through story morals are more nuanced and detailed than the five core concerns outlined by the MFT. Despite this distinction, the methodology of this research parallels the use of the MFT in comparative analysis, which has been employed to explore the moral orientations of different groups, such as Conservatives and Liberals. For instance, it has been shown that Liberals tend to rely more heavily on the Care and Fairness foundations, while Conservatives are more inclined towards the Loyalty, Authority, and Sanctity foundations (Graham et al.). This thesis adopts a similar comparative approach, utilizing this lens to examine narrative content across different contexts,

yet it does so without employing a predefined list of archetypes, which allows for a more fluid exploration of the narrative morals present in diverse contexts.

I have selected environmental news as the object of analysis mainly due to the pressing nature of issues like climate change and pollution, which have profound connections to media coverage and governmental response. The urgency is underscored by irreversible environmental impacts observed in recent events. For instance, in 2023, Canada experienced its worst wildfires on record, while the United States faced the highest number of billion-dollar disasters in a single year. Similarly, China witnessed unprecedented flooding, with record rainfall affecting at least 16 cities and provinces in the northeastern region during the summer of 2023. These environmental disasters directly impact populations and economies, leading to the loss of lives, large-scale population displacements, and significant economic downturns, including damage to infrastructure and agricultural losses that strain local and national economies. By examining morals within environmental news, this study aims to provide insights into how different regions and ideological frameworks are responding to environmental challenges and the types of actions they advocate. This, in turn, informs public understanding and can potentially influence policy and mobilize action to tackle challenges like climate change, ultimately affecting global sustainability.

While morals are inherent in all types of narratives, they are typically not associated with news. News is often perceived as an objective medium, ostensibly providing straightforward reports of reality. However, research has demonstrated that the presentation of news significantly influences how subjects are perceived, a phenomenon known as news framing effects. A well-known example of this is summarized by Sophie Lecheler and Claes H. Devreese in their book *News Framing Effects*:

> About two decades ago, Nelson, Oxley, and Clawson (1997) presented strong evidence for the impact of news framing. In a local news story about the Ku Klux Klan (KKK), one group of respondents was exposed to a news story that framed a planned KKK rally as a free-speech issue. The other group of respondents was exposed to a news story that framed the rally as a disruption of public order. Those reading the free-speech-issue frame exhibited more tolerance for the KKK than those reading the public-order-disruption frame (1).

Framing, sometimes also known as schemas, involves highlighting certain aspects of perceived reality in communicative texts to promote specific problem definitions, causal interpretations, moral evaluations, and recommended solutions (Entman). Matthew Nisbet defines frames as "interpretive storylines that set a specific train of thought in motion, communicating why an issue might be a problem, who or what might be responsible for it, and what should be done about it" (15). Even when different news outlets report on the same event or phenomenon, their distinct political ideologies and agendas result in varied framings. As such, utilizing morals allows us to examine the presence of frames in all types of narratives. While most existing classification systems focus on literary works, analyzing morals in news enables us to better understand the effects of news framing, i.e., how different presentations and emphases shape public perception and opinion. For example, both China and the United States are key players in climate change discourses, but we can hypothesize that the two countries would present very different morals when discussing climate change due to their distinct economic and political situations, which, in turn, would lead to varied priorities in climate discourse and action-taking within each country.

A prominent example of framing in the context of environmental communication is seen in the discourse surrounding climate issues in the United States, where conservative groups have effectively replaced the term "global warming" with "climate change." This strategic choice of language, shifting to a term with a less alarming and more neutral connotation, illustrates how framing can influence public perception and dialogue. Just like what George Lakoff points out in his paper "Why it Matters How We Frame the Environment," "The idea was that 'climate' had a nice connotation – more swaying palm trees and less flooded out coastal cities. 'Change' left out any human cause of the change. Climate just changed. No one to blame" (71). As evident from this example, framing can occur even at the level of word choices. Additionally, it has been found that news discourse framing also operates through syntactical structures, script structures, thematic structures, and rhetorical structures (Pan and Kosicki). In the following part of this section, I conduct a literature review to examine what scholars have discovered and how they have attempted to detect frames in the past. We will explore the kinds of frames they have found to be prevalent, the scope of their data, the regions and timespans of their focus, and identify what gaps remain in the field of environmental communication.

For researchers and climate advocates, climate change represents a "super wicked problem," in part due to its immense scale, which makes it challenging to motivate individual action. In addition, this "super wicked problem" is characterized by the urgent need for action, yet hindered by delayed impacts and insufficient institutional efforts (Levin et al.; Lazarus; Rodrigo-Alsina). Consequently, scholars have stressed that effective framing is essential to bridge the action gap, i.e. the gap between awareness and action in environmental protection. In their paper, Bushell et al. have delineated the two causes for the action gap, the first being the absence of a nation-level strategy for addressing climate change and the second being the failure

to "agree on and articulate the complex range of solutions, and the need to implement those solutions, in a compelling way" (39). The authors argue that the second cause is the reason why we need a "unifying strategic narrative" to coordinate and unify the various actors involved in addressing the action gap. In a similar vein, according to Matthew C. Nisbet:

> Reframing the relevance of climate change in ways that connect to a broader coalition of Americans – and repeatedly communicating these new meanings through a variety of trusted media sources and opinion leaders – can generate the level of public engagement required for policy action. (14)

In other words, Bushell et al.'s unifying strategic narrative can be understood as an effective frame or narrative schema that, when consistently repeated, triggers specific responses in the audience's brain and ultimately motivates action. Therefore, proper framing can be useful for addressing the action gap in the context of climate change – "Truth must be framed effectively to be seen at all. That is why an understanding of framing matters" (Lakoff, 80).

In the field of climate communication, a substantial body of literature has explored the various frames in climate discourse, albeit with different focuses. One of the first lists of frames was brought up by Mike Shanahan in an IIED (International Institute for Environment and Development) briefing paper, which Mike Hulme also summarizes in his 2009 book *Why We Disagree about Climate Change*. These frames include the scientific uncertainty frame, the national security frame, the polar bear frame, the money frame, the catastrophe frame, and the justice and equity frame, with each corresponding to a specific group of audience (e.g. the polar bear frame engages wildlife lovers while the money frame engages politicians and the private sector). Around the same period, Matthew Nisbet also compiled a similar typology of eight frames applicable to climate change, such as "social progress," "economic development and

competitiveness," and "morality and ethics" (see Table 1 for the full list). Nisbet describes the latent meanings of these frames in terms of how they define science-related issues. For example, the "economic development and competitiveness" frame defines science-related issues as an "economic investment; market benefit or risk; or a point of local, national, or global competitiveness," whereas the "morality and ethics" frame defines it as a "matter of right or wrong; or of respect or disrespect for limits, thresholds, or boundaries" (18). Nisbet developed these eight frames by compiling the frames that frequently recur across various science-related policy debates, including discussions on nuclear energy, food, and medical biotechnology. Already, one can observe a difference between Shanahan's and Nisbet's frames: Shanahan's are designed with the audience's reception in mind, whereas Nisbet's frames serve as distinct definitions of science-related issues. While no frame is inherently right or wrong, it is worth considering the arbitrariness of these methodologies, as both are selected by the author without clear criteria.

Scholars	**Frames**
Semetko and Valkenburg (2000) Dirikx and Gelders (2010) Jingrong Tong (2014)	Conflict frame Human interest frame Economic consequences frame Morality frame Responsibility frame
Matthew Nisbet (2009)	Social progress Economic development and competitiveness Morality and ethics Scientific and technical uncertainty Pandora's box/ Frankenstein's monster/ runaway science Public accountability and governance Middle way/ alternative path Conflict and strategy

Toby Bolsen and Matthew Shapiro (2018)	Scientific consensus/Uncertain science Economic consequences Environmental consequences Morality/Ethics Disaster Political conflict National security Public health Self-efficacy External efficacy Response efficacy
Mike Shanahan (2007) Mike Hulme (2009)	Scientific uncertainty frame National security frame Polar bear frame Money frame Catastrophe frame Justice and equity frame

Table 1. A non-exhaustive list of different sets of frames employed by different authors.

Over time, climate framing has continuously evolved, much like the development of literary genres, with various scholars proposing different sets of frames. For instance, Bolsen and Shapiro have expanded on the frames introduced by Nisbet. They analyzed U.S. news media coverage of climate change using eleven frames derived from multiple sources, including Nisbet's eight frames (Bolsen and Shapiro). Similarly, another lineage of frames can be traced back to Neuman et al.'s 1992 book *Common Knowledge: News and the Construction of Political Meaning*, and are later built on by Semetko and Valkenburg, with a focus on frames in news discourses and media studies in general. This resulted in a set of five frames: the conflict frame, the human interest frame, the economic consequences frame, the morality frame, and the responsibility frame (Semetko and Valkenburg). While these frames are not exclusive to environmental news, they are frequently employed in climate communication studies. For example, Astrid Dirikx and Dave Gelders applied these five frames to examine how Dutch and

French newspapers framed climate change during the annual United Nations Conferences of the Parties (2010), and Jingrong Tong used them to analyze investigative reports on environmental issues in major Chinese newspapers.

These examples (and more can be included) highlight the lack of consensus on frames within the field of climate communication and news framing studies in general. The cause of these diverging frameworks can be understood through Semetko and Valkenburg's discussion of the two possible approaches to content analyzing frames in the news: the inductive and deductive approaches. The inductive approach "involves analyzing a news story with an open view to attempt to reveal the array of possible frames, beginning with very loosely defined preconceptions of these frames," while the deductive approach "involves predefining certain frames as content analytic variables to verify the extent to which these frames occur in the news" (94). The authors note that while the inductive approach can identify numerous ways an issue might be framed, it is labor-intensive, typically relies on small samples, and is challenging to replicate. Conversely, the deductive approach is easily replicable, can handle large samples, and can efficiently detect framing differences both between and within media sources.

It is evident that the aforementioned studies all utilize the deductive approach. According to Dirikx and Gelders, an additional reason for this preference is that previous studies have shown that journalists use "generic frames" that "transcend limits of time" (733). However, what happens when different scholars select different generic frames? This leads to increased divergence of opinions and complicating efforts to build a cohesive understanding of framing in climate communication. The notion of some generic frames that transcend time should remind readers of the movement of genre studies into the digital age, where the use of computational tools has allowed for an exploration of the contingencies of knowledge, rather than a search for

objective truth. In the case of frames in climate news, we propose that the use of LLMs enables an inductive approach, where automation makes the process much less labor-intensive and applicable to large samples, all while making it possible to identify frames that are unique to different regions, cultures, and time periods. In a sense, the application of LLMs can help us rethink the field of climate communication and beyond, potentially leading to new discoveries.

Literature reviews of the field have compiled research covering different periods and geographic regions (Anderson). Some studies compare climate issue framing in various nations (Brossard et al.; Boykoff; Xie); while others perform temporal analyses correlating media coverage with significant climatic and political events, such as the COP and Kyoto Protocol (McComas and Shanahan; Young and Dugas; Keller et al.; Pan et al.). While we won't go into much detail regarding the focus of each study, we are interested in their methodology, i.e. how the researchers determine the presence of each frame. Traditionally, most studies have relied on manual coding methodologies, where coders are trained to identify specific elements in articles, such as scientific controversies, typically resulting in a dataset comprising a few hundred articles. Recently, however, automated methods like topic modeling have been adopted in climate framing research to enhance data analysis efficiency (Keller et al.; Rabitz et al.). For example, Hase et al. have identified the prevalent themes and topics in news coverage of climate change in 10 countries, including "causes of and solutions to climate change," "climate politics," "awareness and education," "impact on humans," etc. Nevertheless, methods like topic modeling reveal only the topics and themes of the news instead of their stance and framing, which means that we are less likely to gain information on how a topic is framed to achieve a certain goal. It should be emphasized that while this section has focused in part on the limitations of existing frames used in environmental communication studies, this does not imply that morals are

conceptually distinct from frames. Rather, morals can be seen as a type of framing, one that identifies and utilizes the overarching patterns that govern messages concerning environmental issues and topics. As such, by taking morals into account, this work can be seen as a further extension of the aforementioned automated approaches, but with a novel focus on the values of narrative messaging by leveraging the affordances of LLMs.

1.4 Conclusion

In conclusion, this chapter presents the extraction of story morals using Large Language Models (LLMs) as a bottom-up schema, a departure from the traditional top-down categorization systems previously used in textual analysis. The adoption of LLMs is grounded in their superior capabilities for narrative understanding and the limitations of conventional computational methods. Specifically, I underscore the value of LLMs in providing fresh insights into environmental communication studies, as they are capable of addressing questions previously reliant on human annotators. While past environmental communication research has often focused on framing, I find the term "story moral" more apt in the sense that it not only captures the essence of the narrative's guiding principles but also emphasizes the behavioral values each story promotes. The moral of a story is intended to shape future behaviors, making it directly relevant to its potential impact on behavior – whether or not it successfully does so remains subject to analysis. It is important to acknowledge that, to date, the only approach to automate the extraction of higher-level lessons from texts is through the use of LLMs. While traditional text mining techniques are adept at identifying topics, themes, stylistic traits, and valence in texts, they lack the sophistication to interpret the underlying messages and narrative values. As such, I view the extraction of morals as a more direct method to surface the latent messaging in

environmental news articles, uncovering their deeper, underlying values that differ between cultures, ideologies, and regions. As will be demonstrated, my research aims to extend beyond mere moral extraction, incorporating techniques from previous studies such as sentiment analysis, topic extraction, and agent-type identification to enhance our understanding of the narrative's moral dimension. In the next chapter, I will detail the data, methodology, and experimental setup used in this research, as well as the validation approaches employed to verify the efficacy of this method.

Chapter Two: Extracting Morals and Narrative Information with Large Language Models

In this chapter, I will delve into the data and methodology employed in this research to conduct high-level information extraction on environmental news.[2] As emphasized in the previous chapter, addressing environmental issues like pollution and the "super wicked problem" of climate change is more urgent than ever. Therefore, I have chosen to concentrate on two major contributors to environmental discourse, China and North America (which are also the top two emitters of greenhouse gases), and study the main agents, morals, and sentiments that emerge from their narratives. This approach sheds light on how environmental issues, such as climate change and pollution, are framed within these communities, highlighting the variances in priorities and framing techniques that arise from distinct cultural and ideological perspectives. Below is a detailed description of our dataset.

2.1 Data

Our dataset includes around 15,000 news articles from Dow Jones Factiva, divided by language into Mandarin and English (see Table 2). These articles, ranging in length from 250 to 2,500 words, were published between January and December 2023 and selected using five key environmental terms: "climate change" (气候变化/氣候變遷 for Taiwanese sources), "pollution" (污染), "carbon emissions" (碳排放), "renewable energy" (再生能源), and "sustainability." For "sustainability," we used the Chinese term huan bao (环保), meaning "environment-friendly." This is because the direct translation of "sustainability" in Chinese (可持续性) is less commonly used in Chinese environmental discourse, whereas the term huan bao

[2] This chapter is a incorporation and extension of the research being done at McGill University's .txtlab. A more comprehensive framework, which can be applied to not just environmental new, has been demonstrated in a paper titled "Story Morals: Surfacing value-driven narrative schemas using large language models," by David G. Hobson, Haiqi Zhou, Derek Ruths, and Andrew Piper.

(环保) is more prevalent. While both "sustainability" and "huan bao" frequently appear in discussions, their usage does not directly translate between languages. Therefore, they are chosen as a key term pair. Originally, the dataset contained approximately 18,600 entries. However, since an article often includes multiple key terms and can thus appear as multiple entries, I removed the duplicates and reduced the dataset to about 15,000 unique articles. In general, this keyword strategy ensures our dataset comprehensively represents major environmental discussions in both language media landscapes.

Our news articles are chosen from eight media outlets in total and two sources each represent a given ideological orientation ideology, including State Media (mainland China), Offshore Media (Hong Kong and Taiwan), Liberal, and Conservative Media (North America). On the Chinese-language side, *People's Daily* (人民日报) and *Global Times* (环球时报) are chosen to represent mainland Chinese "state" ideology because of the popularity of these two sources in China and because of the wide reference of scholarly studies to these sources (Pan et al.; Guo et al; Xie). *People's Daily* is often viewed as the mouthpiece of the Chinese government or the Chinese Communist Party (CCP), and *Global Times* is owned by the People's Daily group, but with a focus on international news and is published in different languages including Chinese and English. To capture non-mainland news, we choose Hong Kong and Taiwan as examples due to their distinct economic, political, and cultural systems that differ significantly from mainland China, coupled with strong international ties and influences that shape their unique identities.[3] *Ming Pao* (明報) is selected for Hong Kong and *Liberty Times* (自由時報) for Taiwan. Both sources are popular in their respective region in terms of the number of readers.

[3] We recognize that Hong Kong and Taiwan have distinct histories and cultures, making it potentially problematic to group them together under the umbrella of "offshore." Additionally, other Chinese-speaking regions, such as Macau, could also be considered when discussing "offshore" ideology. It's also worth noting that due to a higher degree of freedom of speech, Hong Kong and Taiwan both host news outlets with diverse opinions and ideologies, meaning one source may not be representative. Future studies should thus examine these regions separately to capture their unique perspectives.

For the English (North American) side, *The New York Times* and *CNN* were selected to represent liberal ideology, while *The Wall Street Journal* and *The Globe and Mail* were chosen for conservative perspectives. These sources were also selected based on their popularity and the accessibility of their articles through news aggregators like Dow Jones Factiva. Although *Fox News* was initially considered a top choice for a conservative news outlet, it was excluded due to a lack of an up-to-date dataset. Instead, *The Globe and Mail* was chosen, as it is often associated with conservative voices and the right wing in Canada. While political leanings are subjective, numerous studies have analyzed the news sources chosen for this study and attempted to place them on an ideological spectrum, considering factors such as audience composition (Mitchell et al.), thus we believe our choices reflect their respective ideology. [4]

Source Name	**Region**	**Class**	**Number**
People's Daily (人民日报)	Mainland China	State	3748
Global Times (环球时报)	Mainland China	State	1167
Ming Pao (明報)	Hong Kong	Offshore	1315
Liberty Times (自由時報)	Taiwan	Offshore	2028
CNN	United States	Liberal	1730
The New York Times	United States	Liberal	2111
The Wall Street Journal	United States	Conservative	1499
The Globe and Mail	Canada	Conservative	1499

Table 2. Dataset Composition

[4]Similar to the selection of Chinese sources, the choice of these North American sources may seem less well-justified. For instance, the inclusion of one Canadian source alongside three American media outlets might appear inconsistent. Additionally, *CNN* is often regarded as a non-partisan and centrist source even though its audience tends to be more liberal. Future studies should aim to use articles from sources that more accurately represent the intended ideologies.

One potential issue with using keywords to extract news articles is the uncertainty regarding their actual relevance to environmental topics. During the key term selection process, I have already selected words that are adequately context-specific – words like "green" are excluded due to their multiple senses, which can introduce significant noise. Nevertheless, there is still no guarantee that an article pertains to environmental topics. To address this, an annotator is asked to examine a sample of 100 articles and rate their relevance to these subjects, with 3 indicating very relevant, 2 loosely relevant, and 1 not at all relevant. Out of the 100 articles, 25 were marked as loosely relevant, and 6 were rated as not relevant. For example, a Chinese article could use the word "污染," which directly translates to pollution, when referring to the contamination of food in the Chinese language (e.g. 沙门氏菌污染). Despite this nuance in the use of language, 94% of the sampled articles were shown to be at least somewhat relevant to our research topic, indicating that the dataset is adequately representative and the number of false positives is small.

2.2 Choice of Model and Parameters

While we explained our choice of using Large Languages Models (LLMs) in the first chapter, we have yet to describe why LLMs are a valuable tool for this study. Large Language Models were first introduced with significant impact around the mid-2010s, with Google's release of the Transformer model in 2017 that marks a pivotal advancement in their development (Vaswani et al.). The Transformer architecture enabled more efficient training and higher performance on natural language processing tasks by utilizing self-attention mechanisms, which set a foundation for subsequent LLMs. The introduction of OpenAI's GPT (Generative Pre-trained Transformer) in 2018 (Radford et al.) further showcased the capabilities of these

models in generating coherent and contextually appropriate text across various domains. Subsequent iterations, such as GPT-2 and GPT-3, expanded on this by increasing the model sizes and training data, significantly enhancing their complexity and utility. These early models paved the way for the widespread adoption of LLMs and demonstrated their potential in automating tasks, generating human-like text, and providing insights across disciplines. The profound impact of these early developments set the stage for ongoing innovations in AI and the continual evolution of language models.

As this paper is drafted, GPT-4 (OpenAI et al.) enjoys widespread use, while its successor, GPT-4o, has recently debuted. Other notable models include Meta's Llama (Touvron et al., 2023), Anthropic's Claude, and Google's Gemini (Gemini Team et al.). A growing body of research has been comparing these models, such as their capabilities in question answering (Myrzakhan et al.) and detecting fake news (Koka et al.). However, determining the most effective model for narrative understanding – especially given its reliance on in-depth qualitative analysis (Zhu & Zhao et al.) – remains uncertain. Ideally, all leading models would be tested and compared. However, due to constraints of time and cost, this project focuses solely on GPT-4 and GPT-4o. In addition, considering our project's scope includes Chinese-language news, I also evaluated Chinese large language models like Baidu's Ernie Bot (文心一言) and Baichuan AI (百川大模型). Preliminary comparisons showed that, compared to the Chinese LLMs, GPT has better performance in terms of giving the correct answers regarding the protagonist and understanding the articles' morals even in Mandarin. As well, I also decided that using GPT to process both Chinese and English news increases the research's integrity and coherence. Nonetheless, I recognize that the decision might lack robust justification, as comparing Chinese and North American LLMs for narrative understanding merits a dedicated study.

All prompting exercises were conducted using GPT-4 (specifically, 0125-preview) and GPT-4o (gpt-4o-2024-05-13) through OpenAI's API, with a temperature set to zero to minimize output randomness. Besides setting the temperature, our team also takes into account other parameters that can be adjusted. We determined that simply having a user prompt suffices, so no system prompt is set. We also didn't set a maximum token, so by default, it has a maximum token limit of 4,096. Both models have a context window of 128,000 tokens, but gpt-4-0125-preview has a training data of up to December 2023, whereas 4o's training data is up to October 2023 (OpenAI Platform).

2.3 Prompting Workflow

Viewing news as narratives that contain various elements such as agents and morals, I developed a prompting workflow (detailed in Table 3) to capture these different pieces of information through careful prompt engineering to ensure results are in a direct format that requires minimal cleaning. First, GPT-4 is asked to generate a summary of the original article. It is then asked to identify principal agents (protagonist and antagonist), the central topic, a free-form moral, and moral keywords that assume positive and negative valence. There are two reasons for first prompting GPT-4 for a summary. Firstly, the summary aids GPT in better understanding the article through a chain-of-thought process and also translates Chinese articles into English. Secondly, using the summary as context reduces token counts and overall processing costs. To account for potential differences in generated answers when using a summary versus the full context, the experiment is designed with two prompt flow frameworks: the *full-context pipeline* and a *simplified framework*. The full-context pipeline includes all prompts cumulatively (the summary and the original text) so that each prior prompt and its

answer are included in the subsequent prompt, facilitating the model's awareness of prior answers for improved outputs. Alternatively, the simplified framework uses only the summary as the context for each question to reduce costs and computational resources.

Category	**Prompt**
Summary	Can you summarize this story? State your answer as a single paragraph.
Agent	Who is the protagonist of this story? State your answer as a single name.
Agent	Is the protagonist a hero or a villain (i.e., are they portrayed positively or negatively), or are they a victim? You may choose more than one. If none, say none.
Agent	Who is the antagonist of this story? State your answer as a single name. If there is none, say none.
Topic	What is the central topic or issue of this story? State your answer as a single keyword or phrase.
Valence	Is this story more negative or positive? State your answer as a single number between 1 and 5 where 5 = very positive, 1 = very negative, 3 = neutral.
Moral	What is the moral of this story? State your answer as a single sentence.
Moral Keyword Positive	What is the moral of this story? State your answer as a single word or phrase followed by "is a good behavior".
Moral Keyword Negative	What is the moral of this story? State your answer as a single word or phrase followed by "is a bad behavior"

Table 3. Prompting workflow

As indicated in Table 3, the three agent questions are employed to extract representations of agents and their relationships within the article. While protagonists in fiction are typically clear to readers, identifying the protagonists in news articles can often be contentious. Therefore, we will examine annotator agreement on this matter in the validation section. To enhance our understanding of the article's content and themes, we input the central topic question into GPT, which then generates topic keywords. These keywords facilitate the grouping of similar topics,

potentially demonstrating regional focuses on different topics. The article's valence is expressed as a number between 1 and 5, which allows the calculation of statistical measures such as mean and variance to explore differences across regions and ideologies.

Man charged with smuggling greenhouse gases from Mexico into US in first-of-a-kind prosecution

California man was arrested and charged Monday with allegedly smuggling potent, planet-heating greenhouse gases from Mexico, marking the first such prosecution in the US, according to a statement from the US Attorney's Office for the Southern District of California. Michael Hart, a 58-year-old man from San Diego, pleaded not guilty to smuggling hydrofluorocarbons, or HFCs —commonly used in air conditioning and refrigeration —and selling them for profit, in a federal court hearing Monday. According to the indictment, Hart allegedly purchased the HFCs in Mexico and smuggled them into the US in the back of his truck, concealed under a tarp and tools. He is then alleged to have sold them for a profit on sites including Facebook Marketplace and OfferUp. "It is illegal to import certain refrigerants into the United States because of their documented and significantly greater contribution to climate change," Assistant Attorney General Todd Kim of the Justice Department's Environment and Natural Resources Division said in a statement Monday.

Q1: Who is the protagonist of this story?
Michael Hart
Q2: Is the protagonist a hero, a villain, or a victim?
Villain
Q3: What is the central topic or issue of this story?
Illegal smuggling of hydrofluorocarbons
Q4: Is this story more negative or positive?
Very negative
Q5: What is the moral of this story?
Illegal actions that harm the environment and undermine international climate change efforts will be prosecuted and held accountable.

Figure 1. Excerpted version of our prompts on a sample news article

Additionally, the model outputs the morals in three formats: a moral sentence, positive moral keywords, and negative moral keywords. As previously defined, morals represent behavioral values the author aims to convey, and can thus be captured through the keywords as good or bad behaviors. Therefore, the moral sentence provides a comprehensive view, while the keywords offer a concise summary that facilitates easier aggregation and comparison. Figure 1 presents an example output for a news story, where I have manually highlighted sections in various colors to indicate the parts that provide evidence supporting the answers to the prompts, based on my interpretation. As will be demonstrated in the third chapter, the extraction of

narrative features such as topic, agent, and valence allows us to identify larger associations between key actors or topics and particular morals. In that section, we experiment with various aggregation techniques to identify salient differences among story morals across different cultural and ideological divides.

2.4 Validation[5]

One of the significant challenges with large language models (LLMs) is the issue of hallucination or, as it is increasingly termed, confabulation. Defined as the production of content that strays from factual accuracy, resulting in untrustworthy outputs (Rawte et al.; Maynez et al.), this problem poses a critical concern for scholars. Recently, however, studies have shown that confabulations, characterized by increased narrativity and semantic coherence, mirror a human tendency to use narrative as a cognitive tool for sense-making and communication, suggesting that LLM confabulations may have potential value in generating coherent narrative text (Sui et al.). In this particular study, the issue is somewhat mitigated by the nature of the inquiries posed to the LLMs. Specifically, these models are not tasked with answering questions that require a direct demonstration of factual knowledge. Instead, they address open-ended or subjective queries. How, then, do we adjudicate the fit of LLMs in advanced tasks like moral extraction? Furthermore, how should we incorporate and consider cultural perspectives and other variables that might shape the responses? This is where the role of validation becomes crucial, as validation not only provides insight into how subjectivity contributes to differences in moral perceptions but also assesses the performance of GPT models based on human judgments.

[5] As mentioned in the Contribution of Authors section, David G. Hobson is the main person responsible for the validation of this research.

To validate our methods, we employ both human evaluations and automated metrics. For the purpose of assessing GPT's effectiveness in various cultural contexts, our test dataset includes 64 news articles sourced from political coverage by *CNN*, *Al-Jazeera English*, and four Chinese-language news outlets (detailed in Table 2). Here, instead of having a direct sample of the full dataset, *CNN* and *Al-Jazeera* English are chosen as representations of English news as the former comes from a more "Western" perspective while the latter provides a more Eastern point of view. These documents average 987 words each, ranging from a minimum of 250 to a maximum of 2,200 words.

2.4.1 Human Validation

For human validation purposes, we employed ten student annotators – six who are native English speakers and four who are native Mandarin speakers. They were tasked with responding to the prompts listed in Table 3, excluding the summarization question, for each text passage. These annotators were equipped with a codebook containing category definitions and examples and completed a training session and at least one round of practice annotations to ensure consistent interpretation of these definitions. All responses were open-ended and completed independently of each other and of GPT-4.

For categories with more clear-cut answers (like protagonist and antagonist), we evaluated the alignment between GPT's responses and those from the human annotators. We collected multiple responses from GPT, equal to the number of human annotators, to strengthen the robustness of the data. On average, we observed agreement rates of 49% / 61% for the protagonist and 71% / 97% for the antagonist under majority (when more than 50% of annotator responses match with GPT's) / any agreement (when at least one annotator's response matches

with GPT's) conditions, respectively. For the more subjective categories such as Moral, Positive Moral Keyword, Negative Moral Keyword, and Topic, we utilized Amazon's Mechanical Turk (AMT). Here, crowd workers were shown three choices for each category – one from GPT-4 and two randomly chosen from the human annotators. They were tasked with selecting the options they deemed "most applicable" and "least applicable" to the text of the passage, without specific guidelines on what constitutes a good or bad choice, allowing them to make selections freely.

To maintain high-quality responses, we mandated that workers have a lifetime success rate above 95%, and required them to correctly answer a comprehension question about a passage. To counteract emerging concerns about crowd workers possibly using ChatGPT to complete tasks, we provided all passages as images. For Mandarin texts, we included English translations to maintain a consistent pool of annotators[6]. In total, we collected three responses for each passage. For evaluating the full-context prompt framework, we utilized the entire dataset. For the simplified prompt framework, we used a subset of 32 articles while ensuring an equal representation from all news sources.

[6] Translating Chinese texts into English for annotators who speak only English is not ideal, as nuances may be lost in translation. Although my qualitative analysis indicates that the GPT-generated translations effectively capture the essence of the original texts, future research should consider implementing a more sophisticated system that includes Chinese native speakers in the validation process.

	Agreement (%)			Fleiss κ	GPT Majority (%)	χ^2
	1	2	3			
Most applicable						
Moral	9.38	71.88	18.75	0.05	59.38	$p < 10^{-5}$
Positive Moral	25.00	43.75	31.25	0.16	37.50	$p = 0.14$
Negative Moral	21.88	65.62	12.50	0	34.38	$p = 0.28$
Central Topic	12.50	68.75	18.75	0.07	53.12	$p < 10^{-5}$
Least applicable						
Moral	28.12	65.62	6.25	-0.11	9.38	$p = 0.03$
Positive Moral	6.25	75.00	18.75	0.15	34.38	$p = 0.28$
Negative Moral	6.25	87.50	6.25	0.01	18.75	$p = 0.35$
Central Topic	18.75	71.88	9.38	-0.03	12.50	$p = 0.08$

Table 4. Inter-annotator agreement and GPT selection rate among AMT workers during the human evaluation of the *simplified prompt framework*. The first 3 columns give the breakdowns of agreement among the annotators; that is, how often 1, 2, or 3 annotators agreed on an option as a percentage of the total number of passages. The fourth column gives the Fleiss κ coefficients for inter-annotator agreement. The fifth column gives the observed rate at which GPT was selected by the majority of AMT workers. The final column gives the p-value for the χ^2 goodness of fit test under the null hypothesis that GPT responses were only selected at random ($p = 1/3$), and therefore had an expected probability of $7/27$ ($\approx 26\%$) of being selected in the majority ($P(X \geq 2) = 7/27$ for a binomial random variable X with $n = 3$ (the number of AMT responses) and $p = 1/3$). Source: Zhou et al., 2024.

As shown in Table 4 for the simplified prompt framework, we attained majority agreement in 75-90% of cases across each category. However, inter-rater agreement was notably low because, although two annotators might both prefer a human moral, they could have selected different ones. For instance, in a news article about TSMC (Taiwan Semiconductor Manufacturing Company)'s effort to address carbon emissions through the use of renewable energy, one rater chose the moral "A company needs to take measures for sustainable operation" while another chose "Using renewable energy can help reduce carbon emissions." The two morals differ in focus but are both plausible interpretations of the original article. Despite the lower inter-rater agreement, the selection of GPT morals and central topics significantly exceeded a random baseline. Additionally, the alignment of positive and negative morals with

random choice was confirmed by a χ^2 goodness of fit test. In no instance did crowd-workers consistently choose GPT answers as "least applicable."

For comparison with the simplified framework, the result for the full-context framework is shown in Table 5. With the inclusion of full context, the results slightly improved, and GPT responses were more frequently chosen by the majority of AMT raters. However, since neither prompt framework demonstrated a negative bias against GPT responses, we opted to use the simplified framework for our comprehensive analysis.

	Agreement (%)			Fleiss κ	GPT Majority (%)	χ^2
	1	2	3			
Most applicable						
Moral	14.1	50.0	35.9	0.03	73.44	$p < 10^{-5}$
Positive Moral	17.2	64.1	18.8	-0.01	57.81	$p < 10^{-5}$
Negative Moral	18.8	67.2	14.1	0	51.56	$p < 10^{-5}$
Central Topic	9.4	48.4	42.2	0.15	73.44	$p < 10^{-5}$
Least applicable						
Moral	15.6	60.9	23.4	0.06	7.81	$p < 10^{-3}$
Positive Moral	18.8	65.6	15.6	0.05	12.50	$p = 0.01$
Negative Moral	26.6	57.8	15.6	0.01	17.19	$p = 0.11$
Central Topic	6.3	62.5	31.3	0.22	7.81	$p < 10^{-3}$

Table 5. Inter-annotator agreement and GPT selection rate among AMT workers during the human evaluation of the *full-context prompt* framework. The first 3 columns give the breakdowns of agreement among the annotators; that is, how often 1, 2, or 3 annotators agreed on an option as a percentage of the total number of passages. The fourth column gives the Fleiss κ coefficients for inter-annotator agreement. The fifth column gives the observed rate at which GPT was selected by the majority of AMT workers. The final column gives the p-value for the χ^2 goodness of fit test under the null hypothesis that GPT responses were only selected at random ($p = 1/3$), and therefore had an expected probability of $7/27$ ($\approx 26\%$) of being selected in the majority ($P(X \geq 2) = 7/27$ for a binomial random variable X with $n = 3$ (the number of AMT responses) and $p = 1/3$). Source: Zhou et al., 2024.

To assess whether there is a significant difference in rater preference between the translated Chinese news articles and the English news, we conducted a two-sample t-test on the

percentage of GPT Majority (shown in the fifth column of Tables 4 and 5) and obtained a p-value of 0.7814. This result indicates that there is no statistically significant difference in how raters selected GPT's responses between the English texts and translated Chinese texts.

2.4.2 Automated Validation

	human-human	human-GPT	GPT-GPT	U-test
Moral				
Rouge-1	0	8.00	58.62	$p < 10^{-4}$
Rouge-L	0	7.41	51.61	$p < 10^{-4}$
GloVe	55.01	64.74	91.03	$p < 10^{-4}$
STSb-MPNet	25.89	38.83	85.11	$p < 10^{-4}$
NLI-MPNet	33.17	46.63	89.58	$p < 10^{-4}$
Positive Moral				
Rouge-1	0	0	57.14	$p < 10^{-3}$
Rouge-L	0	0	57.14	$p < 10^{-3}$
GloVe	31.07	40.73	81.84	$p < 10^{-4}$
STSb-MPNet	27.12	26.66	76.73	$p = 0.62$
NLI-MPNet	38.58	35.57	81.70	$p = 0.01$
Negative Moral				
Rouge-1	0	0	66.67	$p = 0.03$
Rouge-L	0	0	66.67	$p = 0.03$
GloVe	24.69	29.45	84.28	$p = 0.15$
STSb-MPNet	20.87	18.85	86.10	$p < 10^{-3}$
NLI-MPNet	30.84	26.14	86.54	$p < 10^{-4}$
Central Topic				
Rouge-1	0	11.11	100	$p < 10^{-4}$
Rouge-L	0	11.11	100	$p < 10^{-4}$
GloVe	44.17	55.79	100	$p < 10^{-4}$
STSb-MPNet	33.78	41.45	100	$p < 10^{-3}$
NLI-MPNet	39.63	48.61	100	$p < 10^{-4}$

Table 6. Median similarity (out of 100) between the different groups of annotators in the validation dataset. The human-human column compares all pairs of responses (to the same passage) among the human annotators, the human-GPT group compares all pairs of responses between human and GPT responses, and the GPT-GPT column compares all responses between GPT responses. The p-values are calculated using a Mann-Whitney U-test (rank-sum test) with a null hypothesis that the human-human and human-GPT distributions are the same. Source: Zhou et al., 2024.

For the automated validation of morals and central topics, we utilized a set of semantic textual similarity (STS) scores that align with our human annotations. These included ROUGE-1 and ROUGE-L metrics (Lin), and cosine similarity calculated with pre-trained embedding models from the SentenceTransformers library (Reimers and Gurevych). The ROUGE-based metrics were processed using the HuggingFace library, while the embedding models were implemented through the SentenceTransformers library. The specific models used included averaged GloVe word embeddings (6b-300d) (Pennington et al.), as well as the stsb-mpnet-base-v2 and nli-mpnet-base-v2 models (Song et al.).

Table 6 presents an additional analysis using automated evaluation metrics to contrast human responses with those from GPT. For each story, all pairs of responses were analyzed and aggregated into a single distribution of pairwise similarity among the annotations. The 'human-human' column shows the median similarity for comparisons exclusively between human annotators, the 'human-GPT' column displays the median for comparisons between human and GPT responses, and the 'GPT-GPT' column measures the similarity between different GPT responses to assess replicability. As mentioned before, the quantity of GPT responses was matched to the number of human annotations.

Moral 1	Moral 2	Rouge1	RougeL	GloVe	STSb-MPNet	NLI-MPNet
Hard work will pay off in the end	Hard work pays off	66.67	66.67	96.00	88.48	86.76
The truth is worth fighting for	Seeking the truth is important	54.55	54.55	69.76	76.59	75.43
Investigating your past can unearth old secrets	Confronting one's past, no matter how difficult, can lead to truth, justice, and the opportunity for a new beginning	14.81	14.81	54.69	45.49	53.25
Loyalty to your loved ones will be rewarded	Justice prevails against tyranny	0.00	0.00	24.18	21.07	20.73

Table 7. Sample morals from the validation dataset and their pairwise similarity with respect to the different automated metrics. Source: Zhou et al., 2024.

Overall, our analysis reveals that the semantic variation among human responses is greater (i.e., they show lower similarity) compared to the variation between GPT and human responses across all evaluated metrics, suggesting that GPT effectively mirrors a general human perspective. As indicated in Table 6, while positive and negative morals sometimes deviate from this pattern across certain metrics, the differences remain minor, even when they reach statistical significance. Additionally, GPT consistently shows high similarity scores when responding multiple times to the same text, although with some degree of variation. An exception is noted in the case of the Central Topic, where GPT consistently produces identical responses. For examples of passages illustrating these automated score comparisons, readers can reference Table 7.

	human-human	human-GPT	GPT-GPT
Valence (average standard deviation)	0.68	0.66	0.08
Protagonist Type (average Jaccard index)	44.05	42.46	89.06

Table 8. Average standard deviations in valence responses and Jaccard index (Jaccard is out of 100) in protagonist type between the different distributions of responses. The human-human column compares all pairs of responses (to the same passage) among the human annotators, the human-GPT group compares all pairs of responses between human and GPT responses, and the GPT-GPT column compares all responses between GPT responses. The number of GPT responses was always chosen to be equal to the number of human annotators. Source: Zhou et al., 2024.

Finally, for valence, Table 8 shows that the average standard deviation between human-only responses was 0.68 across all passages, and 0.66 between human and GPT responses. The variability in valence assessments between GPT and humans is comparable to that among humans, suggesting GPT's effectiveness in mimicking human judgment in emotional valence. GPT's assessments are also highly consistent in recognizing protagonist types, as indicated by a high GPT-GPT Jaccard index. Overall, GPT demonstrates a reliable and close alignment with human perceptions in the tasks evaluated.

2.5 Agent Types [7]

After completing the validation process, I discovered that simply extracting protagonist names using the question "Who is the protagonist of this story? State your answer as a single name." does not provide adequate information for effective data mining, especially in the context of news. For instance, knowing only the name of a mayor does not reveal enough about their role or influence, making it challenging to conduct thorough data analysis. Consequently, I replaced

[7] Because the agent type question is a later addition, it is not included in the previously mentioned validation process. In the future, this research can be improved by having a more cohesive validation framework.

this approach with a focus on agent types, which offers deeper insights into the main actors involved. This is particularly relevant in climate communication, where understanding the entities contributing to climate change and pollution is crucial. Recent research, such as that by Petzold et al., also emphasizes the roles of different agent types in climate change adaptation activities like "awareness raising" and "planning." While the authors have their own list of agents like "Government (national)," "Private Sector (corporations)," and "Academia," I came up with a more specific list of agent types to better capture their diversity. The prompt is as follows: "Which following group does the main agent in the article belong to? National governments, Local governments, Political entities, IGOs, NGOs, Industries and businesses, Scientific community, Media and communication entities, Educational institutions, Financial entities, Health and medical entities, Legal and judicial entities, Civil society and community groups, and Individuals or households. State as a single name." With this prompt, there is a more refined categorization that aids in a comprehensive understanding of the various entities discussed in the articles. By looking at the representation of these agent types (for example. as heroes, villains, or victims, or their valence association), one can gain a better insight into the framing of agents in environmental discourse.

To assess whether GPT can accurately identify agent types, a validation process was carried out on a sample of 100 articles. I independently annotated these articles and then had GPT-4o respond using a full-context framework, where it is asked to first summarize the article and then identify the agent type. This approach was chosen because GPT-4o offers performance comparable to GPT-4 but at a lower cost. The results revealed that in 20 of the 100 articles, GPT-4o and I identified different but equally plausible main agents (for example, an article might feature an event co-hosted by a local government and a business, allowing for either to be

considered the main agent) or assigned different types to the same agent. Additionally, in 3 articles, the agent type given by GPT was unclear or unjustified. This indicates that in 97% of the cases, GPT's responses aligned with human annotations, and in 20% of cases, it presented a different but potentially valid perspective from my own. The outcome highlights the validity of using GPT for the task of agent-type extraction.

Evidently, this validation process has a limitation of having just one annotator, so future studies can incorporate more annotators to improve the reliability of the results. As well, future work can explore the possibility of examining directional relationships between agents. For example, if the main agent is a national government, which country is talked about, and which media is framing it as a hero/villain/victim? If a finer level of information is extracted, we can learn more about the dynamics between agents.

2.6 Limitations

Understanding climate and environmental communication on a large scale presents several research challenges. Early in this chapter, I noted an important limitation of this study: the reliance on GPT as the primary large language model. Future research should explore the behavior of other advanced models and consider using smaller, specialized models to reduce the substantial carbon footprint associated with larger models. In addition, the methodology introduces some inconsistency due to the use of both GPT-4 and GPT-4o. Ideally, using a single model would be preferable to ensure uniformity in the analysis. Moreover, despite observing strong levels of human-judged validity in the appropriateness of the GPT-generated morals, further research is needed to understand intercultural differences in how narrative messaging is perceived. It is also important to recognize that while GPT-generated content generally shows

high semantic relatedness across multiple runs of the same queries, some variability remains even when the temperature setting is minimized, making exact replication challenging.

Although our study examines eight different news outlets across two national and ideological contexts, expanding the sample to include more cultures is crucial for enhancing the representativeness of the dataset. The current methodology, which uses keyword filtering to select articles, limits our understanding of how different environmental topics are addressed because it requires separate analysis for each keyword. Future studies might explore alternative methods for gathering a more comprehensive sample of environment-related communications.

2.7 Conclusion

In this chapter, I outlined the setup of the prompting workflow as well as the validation process using GPT-4 and GPT-4o. Both human validation and automated validation results show that GPT has no problem extracting higher-level information such as morals and valence that matches human judgments. In the following chapter, I will delve into how the extracted information, such as agent types and morals, is analyzed. I will present detailed findings that shed light on the current state of environmental discourse, including the difference in morals and agent portrayals between regions and ideologies. These insights, in turn, will underscore the potential of using AI for large-scale textual analysis to capture thematic and moral nuances.

Chapter Three: Analysis and Discussion

In this chapter, I detail the data structure and findings from the extraction of narrative information like agents, valence, and morals from environment-focused articles. The analysis unfolds in three distinct sections. Initially, I provide an overview of the variations in valence and agent distribution across different regions and ideologies. Subsequently, I explore the correlation of the agent types and sentiment values with narrative archetypes – heroes, villains, and victims – and their associated valences, examining these relationships both ideologically and regionally. Following this, I delve into the identification of distinctive morals using Fightin' Words (Monroe et al.) to highlight ideological and regional moral variances. This approach is juxtaposed with traditional topic modeling to underscore the unique moral frameworks of each region. Key research questions addressed include:

1. Is there a statistically significant difference in the valence across various regions or ideologies?
2. Which agent types are prevalent in each region and ideology, and how are they characterized in terms of their association with valence and narrative archetypes?
3. What distinct morals emerge in different regions and ideologies?
4. How does the analysis of morals through LLMs compare with traditional approaches like topic modeling?

This chapter thus systematically elucidates insights derived from analyzing environmental articles using LLMs, showcasing how different analytical techniques illuminate the interplay between narrative messaging and regional or ideological contexts, thereby enriching our understanding of the dynamics at play in environmental discourse.

3.1 Extracted Results

Title	The year's most extreme weather shows what a warming planet is capable of, and what's to come
Source	CNN Wire
Summary	In 2023, the world experienced its hottest year on record, which significantly influenced a series of extreme weather events, underscoring the impact of climate change…
Agent Type	Scientific community
Agent Archetype	Hero
Antagonist	None
Central Topic	Climate change
Valence	2
Moral	The increasing frequency and severity of extreme weather events due to climate change underscores the urgent need for climate action
Positive Moral	Taking climate action
Negative Moral	Ignoring climate change

Table 9. Sample row from the final data frame

After executing the prompting workflow on all 15k articles, the results are cleaned and consolidated into a data frame. The cleaning process involves removing extraneous elements of GPT's text generation that are irrelevant to the analysis. This ranges from simple adjustments, such as correcting capitalization and eliminating unnecessary punctuation, to more complex edits like removing redundant parts of sentences. For instance, sentence beginnings like "The story of the moral is" are deleted to focus solely on the moral itself. As mentioned in Chapter Two, the later addition of agent types questions using GPT-4o means that there is a second pass of the

articles. This second analysis is integrated with the initial results obtained with GPT-4. The final data structure is exemplified in Table 9, which shows a sample row organized vertically (metadata such as Date is omitted for clarity). In the next few sections, because I will present many findings from data exploration and result analysis, the main takeaways will be numbered and bolded.

3.2 Basic Data Exploration and Visualization

3.2.1 Keywords Distribution

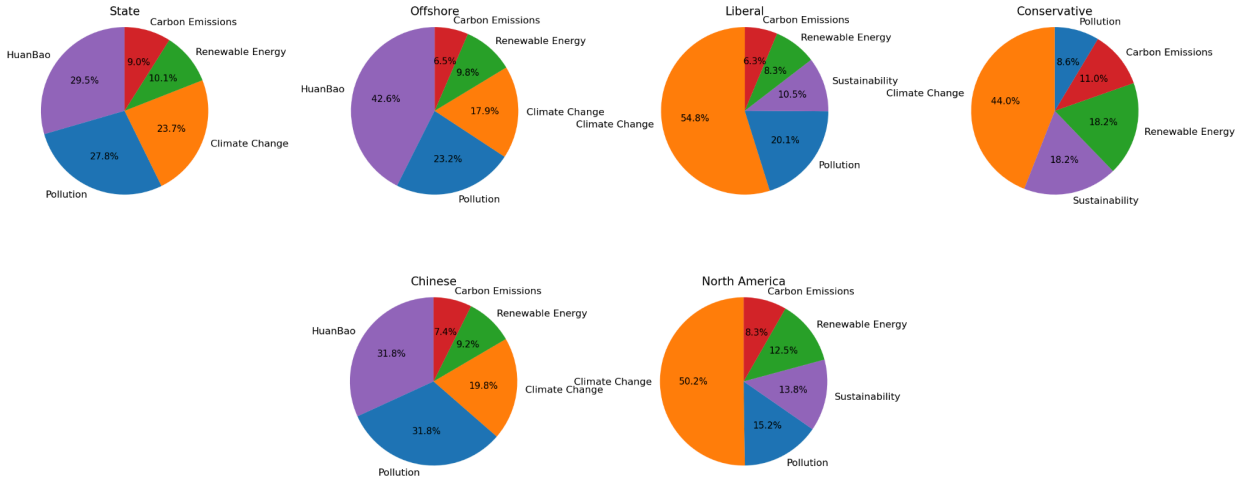

Figure 2. Distribution of keywords in each source

As mentioned before, the dataset of this research is selected using five keywords, thus a comparison of the percentage of keywords in each source can already provide some information. The percentage of each keyword in each source is captured in Figure 2. The charts are generated using the original dataset before the removal of duplicated articles, so one article can contain multiple keywords. Between mainland Chinese (state) and offshore sources, it is obvious from the pie charts that there is a stronger presence of the term *huan bao* (environmentally friendly) in

offshore sources. In North America, it is found that Liberal sources demonstrate more interest in "pollution" and "climate change," whereas Conservative sources show a higher percentage of both "sustainability," "renewable energy," and "carbon emissions." Between Chinese and North American sources, North America displays a more pronounced emphasis on "climate change," possibly indicating a greater orientation towards global climatic events and their impact. Chinese-language sources tend to feature a higher incidence of references to "pollution."[8] Section 3.4 (Distinctive Values) will shed light on the reasons for some of these differences.

3.2.2 Valence

Takeaway 1: Chinese-language news on environmental topics has a significantly more positive valence compared to North American ones.

Takeaway 2: Chinese state media have a higher valence than offshore media.

Takeaway 3: Conservative and Liberal news sources share the most similar valence scores.

Given its numerical nature, valence is one of the most straightforward metrics to compare different sources. A linear regression analysis indicates significant effects of both region (North American versus Chinese sources) and ideology (Liberal versus Conservative, state versus offshore) on average valence. Specifically, North American sources show a lower average valence (M=2.87) compared to those from Chinese-speaking regions (M=3.45). Within China, mainland sources exhibit a higher positivity (M=3.56) compared to offshore sources (M=3.29), likely influenced by censorship and the suppression of dissenting opinions in state-controlled media (Guo et al.). Although differences between Conservative (M=2.93) and Liberal (M=2.83) ideologies are statistically significant, these groups display the closest valence values.

[8]The term "huanbao" is mentioned more frequently in Chinese sources than "sustainability" is in North American sources. However, these two terms are not directly comparable.

3.2.3 Agent Types

For agent types, I analyzed the distribution percentages across the various ideologies and regions, as depicted in Figure 3. The x-axis labels each agent type, while the y-axis shows the percentage of each agent type within its source – summing to 100% for each color-coded category. Figure 3 (top) highlights that in Chinese sources, the most prevalent agent types are national governments, local governments, and industries and businesses. In contrast, North American sources frequently mention industries and businesses, national governments, and individuals or households.

Takeaway 1: Chinese sources contain more local governments than North American sources, whereas North American sources feature more individuals and households. A key regional disparity lies in the representation of local governments and individuals or households; the former is more common in Chinese sources, particularly due to the inclusion of offshore sources from Hong Kong and Taiwan. In these regions, city governments like those of Taipei and Kaohsiung are often referenced, and the Hong Kong government itself is categorized as a local government.[9] The frequent mention of individuals and households in North American sources may not have an explicit explanation, but it could be influenced by cultural values of individualism and collectivism that affect social perceptions and behaviors (Hofstede), as well as the media's role as either a governmental mouthpiece (Chinese state media) or an independent entity, reflecting the fundamental differences between the two geopolitical areas.

Takeaway 2: Offshore sources mention local governments more frequently, while state media more often mention national governments. Between both offshore and mainland Chinese ideologies, as depicted in Figure 3 (middle), the three most common agent types are

[9] Interestingly, it is worth noting that GPT often considers Hong Kong government as local government and Taiwanese government as national government, which might lead to confusion on the actual percentage of mainland China's national government (Beijing). Future modifications should attempt to resolve this ambiguity.

national governments, local governments, and industries and businesses. The most notable differences are seen in the representation of national and local governments, unsurprisingly aligning with earlier discussions.

Takeaway 3: In North America, Conservative sources tend to emphasize industries and businesses as the principal agents, whereas Liberal sources more frequently highlight individuals or households and the scientific community (Figure 3 (bottom)). The notably higher mention of the scientific community in Liberal sources could signify the enduring polarized views towards climate science, which have been a staple in the climate change discourse. Bolsen and Shapiro emphasize how, at the start of the 21st century, "the inherent *uncertainty of science* was accentuated during this period via a shared rhetorical tactic by Conservative organizations and industry representatives seeking to protect their constituents' interests" (151). Although the prevalence of climate skepticism and denial in news coverage has decreased over time, it appears that there remains a notable turn away from references to the scientific community in Conservative sources. In a way, the Conservative news outlets' focus on industries and businesses aligns with their reluctance to mention the scientific community, which may be due to a conflict of interest between the two agents. The heightened interest in industries and businesses also correlates with the frequent mention of the keyword "renewable energy," as previously demonstrated. Although the comparison so far has proven useful, it is crucial to recognize that the mere prominence of an agent type does not determine its depiction in the narrative. The following section will look further into how these media outlets portray different agents in their stories.

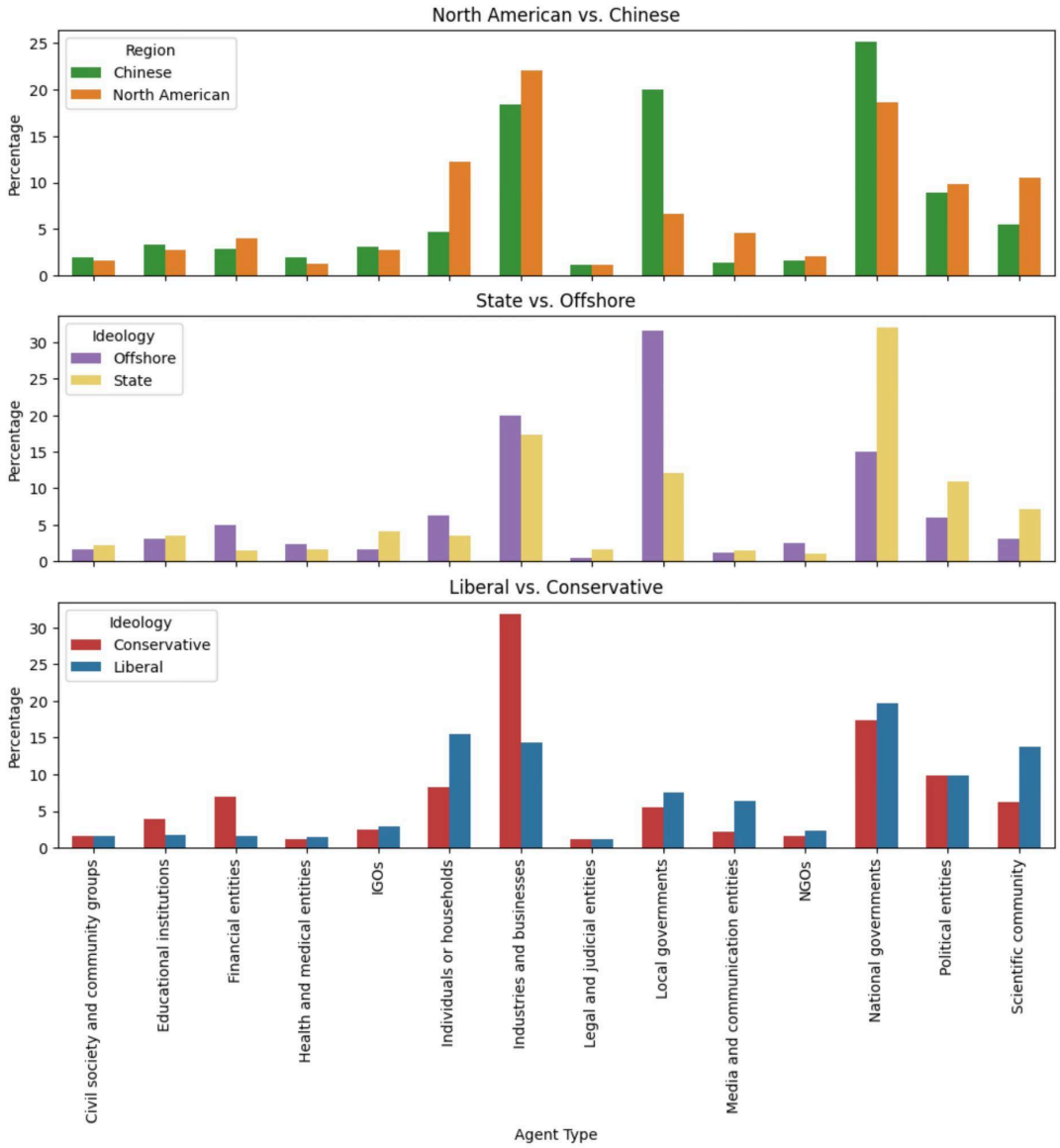

Figure 3. Comparison of agent-type distribution across regional and ideological entities

3.3 Agent Representation

We can understand agent representation in two key ways by analyzing the extracted information. Firstly, examining how different agent types are linked to specific character archetypes is insightful. For instance, studies have suggested that hero characters often have a more significant persuasive impact on audiences than other narrative elements, indicating that

those advocating for specific policies would benefit from framing their arguments around a clear hero figure (Michael D. Jones). Therefore, exploring how agent archetypes manifest across various regions and ideologies can reveal the underlying goals or intentions of the storyteller, whether it be the government, media, or other motivated entities. Secondly, valence values can indicate whether an agent type is generally portrayed in a positive or negative context, which provides a deeper insight into the sentiment trends associated with agent types within different media landscapes.

3.3.1 Agent Types-Archetypes

To analyze the distribution of heroes, villains, and victims across different groups, I created a horizontal stacked bar graph. This graph illustrates the percentage of each agent type categorized as a particular archetype. For example, in Figure 4, the first bar displays the proportion of agent types identified as heroes in North America. Notably, according to Figure 4, the leading hero agents in North America include industries and businesses, the scientific community, and national governments. In contrast, in Chinese-speaking regions, the primary hero agents are national governments, local governments, and industries and businesses. The most significant differences lie in the percentage of local governments, the scientific community, and individuals or households.

Takeaway 1: Chinese sources tend to heroize local governments more frequently, whereas North American sources are more likely to highlight the scientific community. There can be multiple reasons behind the much lower featuring of the scientific community in Chinese-language media. First, it's important to recognize that climate skepticism is less prevalent in Chinese news media compared to North America, and studies have shown that the uncertainty frame in Chinese news is extremely rare (Pan et al.; Liang et al.). This could be a

potential reason why the scientific community is not foregrounded as much – they are not needed to prove a point (e.g. that climate change is indeed happening or caused by humans). Additionally, the promotion of the scientific community could potentially conflict with governmental development goals. According to Guo et al., from 2013 to 2021, state officials were the dominant voices in news media, overshadowing experts, non-governmental organizations, and citizens. Only a small percentage of sources addressed scientific knowledge. Instead, most reporting framed climate change campaigns as achievements of China's leadership, and over time, critical opinions were increasingly suppressed.

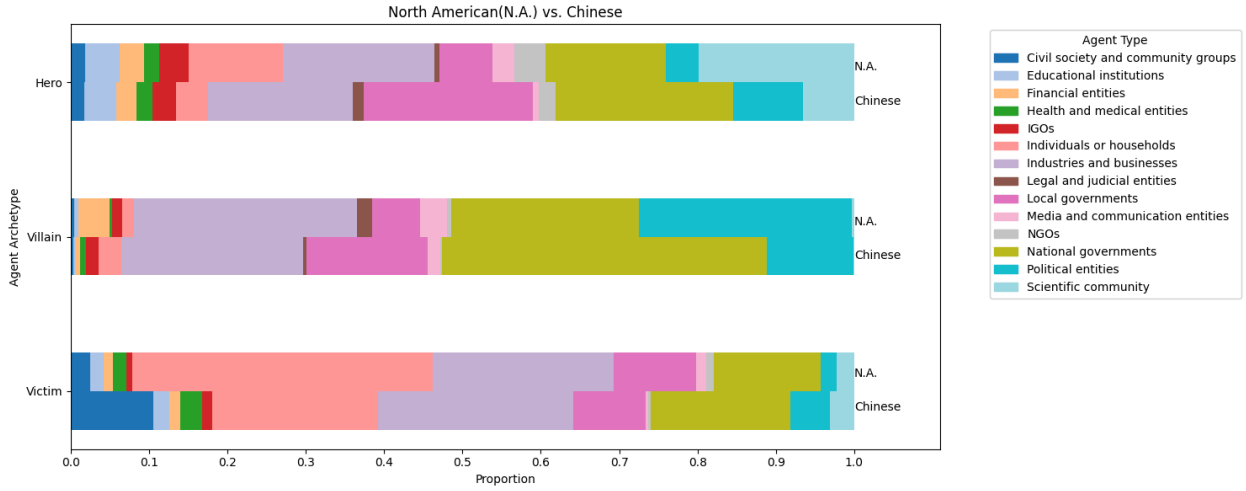

Figure 4. The distribution of agent types in North America vs. in Chinese-speaking regions

Takeaway 2: Political entities (e.g. Democratic or Republican Parties) are more frequently depicted as villains in North American sources, likely due to their prominent role in the diverse political landscape of North America, in contrast to mainland China's single-party system. Although Chinese-speaking regions sometimes portray national and local governments as villains, criticism of the Chinese government itself in state media is rare. This raises the question: which government entities are actually labeled as villains? Although detailed

information on the agents is not extracted, the upcoming comparison between state and offshore sources will provide further insights into this question through closer examination.

Takeaway 3: Chinese-language sources and North American sources have a similar distribution of agent types in the victim category. However, there is a subtle difference: North American sources commonly depict individuals and households as victims, whereas Chinese sources tend to highlight civil society and community groups. Despite these variations, both regions predominantly focus on civilian agents, illustrating how environmental issues impact the everyday lives of people.

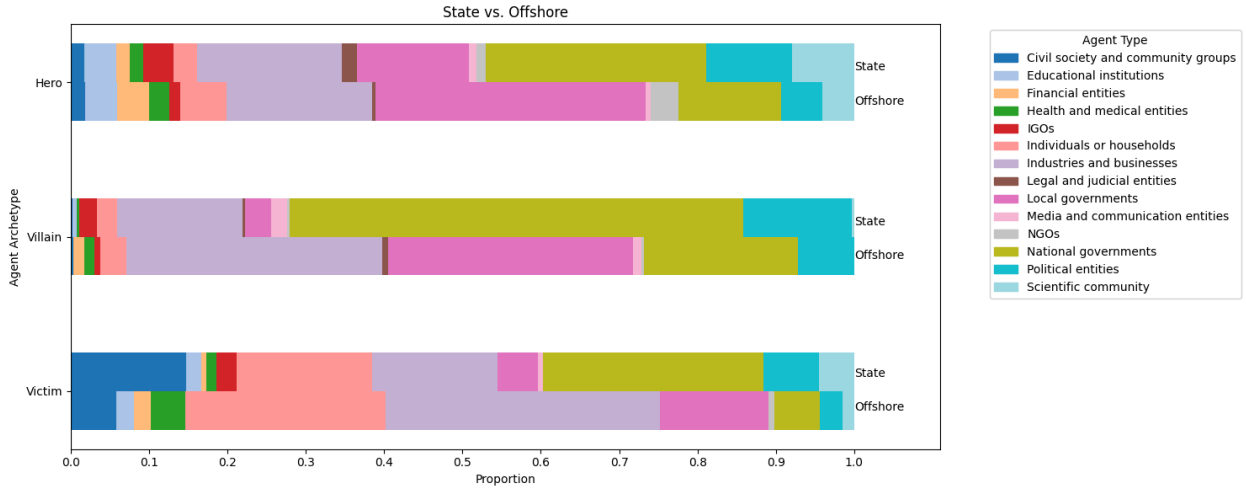

Figure 5. The distribution of agent types in State vs. Offshore ideologies

Between state and offshore sources within the Chinese-speaking region, as shown in Figure 5, state sources tend to feature a higher percentage of national governments, political entities, and the scientific community across all three character archetypes. In contrast, offshore sources are more likely to depict local governments, financial entities, health and medical organizations, and individuals or households.

Takeaway 4: Offshore outlets cast industries and businesses as villains or victims twice more frequently compared to state-run media. This phenomenon, as mentioned before,

could also be caused by critical voices being muted in mainland China. If the government wants to promote the development of businesses and industries, then it is unlikely that it will allow the portrayal of these entities in a negative light.

Takeaway 5: State media often depicts Japan as an antagonist in its negative portrayals of national governments, while Taiwanese media notably avoids portraying Japan as a villain and seldom reports on Japan's nuclear wastewater issues. Following previous discussions, it is crucial to explore why a strikingly high percentage (over 50%!) of national governments are portrayed as villains in state sources. A notable point of focus can be found regarding the Japanese government's release of Fukushima's nuclear wastewater into the ocean in August 2023. State media prominently features Japan as an antagonistic entity – approximately 40% of national governments depicted as villains are associated with Japan or its government. In contrast, Taiwan adopts an interestingly different stance by not extensively reporting on Japan's nuclear wastewater release nor portraying Japan as the villain. Instead, Taiwanese media focuses on highlighting the exaggeration by the mainland Chinese government and media of this event to fuel anti-Japan sentiments, or the environmental impact where the mainland has discharged more harmful substances into the ocean than Japan. This distinct portrayal reflects the differing relationships each region holds with Japan – China traditionally maintains tense relations, whereas Taiwan generally has friendlier ties. As well, this nuanced media attention underscores the complex dynamics of regional media narratives and indicates their influence on public perception, clearly demonstrating the media framing effects.

Takeaway 6: In North America (Figure 6), Liberal media outlets tend to depict the scientific community as heroes more frequently, while Conservative sources more often elevate educational institutions, financial entities, and industries and businesses. Notably,

the Conservative emphasis on educational institutions as heroes is primarily driven by references from the Canadian newspaper *The Globe and Mail*, which frequently cites experts from higher education institutes. The underlying reasons for this phenomenon could be further explored by examining the differences between Canadian and American news media.

Takeaway 7: Liberal and Conservative media outlets have similar villain distribution. Interestingly, despite the differences in heroes, both ideologies display a similar distribution of agent types in the villain category, with industries and businesses, national governments, and political entities consistently portrayed as the top villains.

Takeaway 8: Liberal sources feature individuals or households as victims, while for Conservative sources industries and businesses are more often victims. Liberal sources likely have individuals or households as victims to emphasize the impacts of environmental issues on civilians. In contrast, Conservative sources are more likely to portray industries and businesses as victims because of its preference to address market challenges, reflecting a different perspective on the economic and regulatory challenges faced by these entities.

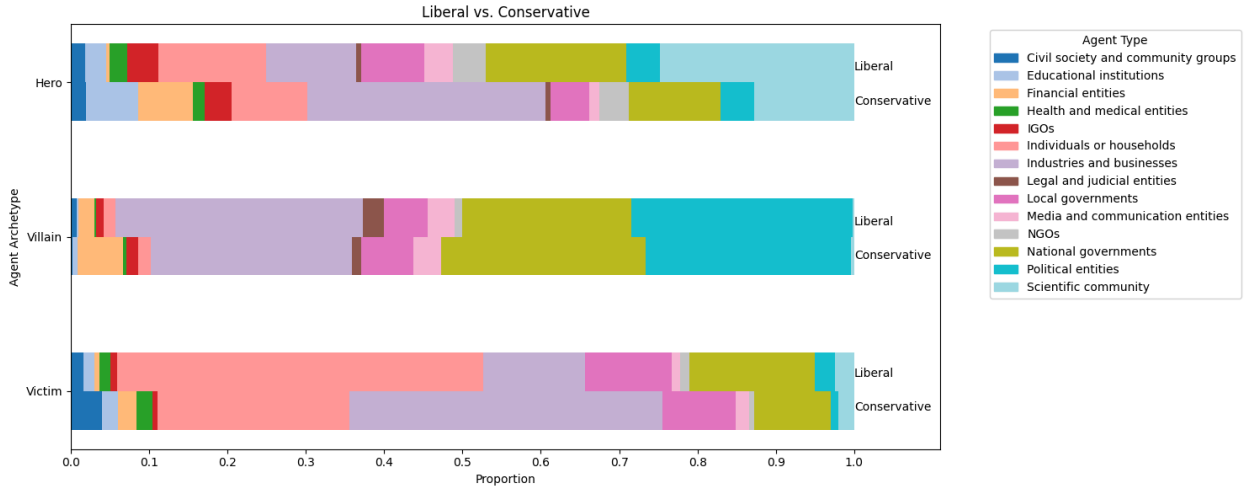

Figure 6. The distribution of agent types in Conservative vs. Liberal ideologies

3.3.2 Agent Types-Valence

Similar to the way agent archetypes convey specific sentiments – heroes typically associated with positive valence, villains with negative valence, and victims with empathy-evoking connotations that are often negative – the valence score situates the main agent in the context of the article, quantifying the agent-sentiment relationship on a numerical scale.

Agent Type	North America	Chinese	Offshore	State	Liberal	Conservative
Financial entities	0.11	0.09	0.08	0.42	-0.16	0.14
Industries and businesses	0.27	0.19	0.14	0.26	0.27	0.23
Local governments	-0.29	0.03	-0.03	0.38	-0.23	-0.38
IGOs	-0.13	-0.2	-0.34	-0.25	-0.19	-0.01
Educational institutions	0.15	0.32	0.42	0.25	0.44	-0.07
Political entities	-0.32	0.01	-0.12	0.01	-0.27	-0.38
National governments	0.04	-0.02	0.02	-0.12	0.14	-0.1
Individuals or households	0.15	-0.06	0.02	-0.06	0.12	0.27
Media and communication entities	0.16	-0.34	-0.08	-0.53	0.24	-0.06
Scientific community	-0.39	-0.1	-0.19	-0.15	-0.37	-0.34
Civil society and community groups	0.07	-0.49	0.12	**-0.86**	0.2	-0.11
NGOs	0.02	-0.39	-0.09	**-0.73**	-0.01	0.1
Health and medical entities	-0.17	**-0.89**	**-0.79**	**-0.93**	-0.14	-0.22
Legal and judicial entities	-0.41	-0.03	**-1.3**	0.18	-0.35	-0.51

Table 10. The difference (in standard deviation) from mean for each agent type in each region/ideology

Table 10 shows the standard deviations of valence scores for each agent type from its mean valence of the same ideological/regional category. Compared to simply showing the average valence for each agent type, This approach is adopted to take into account the difference in average valence (mentioned in Section 3.3.2). The values that are more than 0.7 standard

deviations higher or lower than the mean are bolded. North American sources have agent types whose valence scores do not vary much from the mean.

Takeaway 1: Health and medical entities in Chinese-speaking sources (both state and offshore media) often appear in articles that carry a more negative valence. For example, these entities are linked to adverse events like the increasing number of deaths due to rising temperatures across Asia and in countries like Italy and France. Conversely, in North America, health and medical entities are typically associated with positive developments, such as advancements in AI for healthcare and increased survival rates for lung cancer. North American articles also often include helpful advice for the public, such as tips for the fire season.

Takeaway 2: In state media, civil society and community groups, and NGOs are also frequently associated with articles of a more negative valence. The main community groups are notably those from Korea, China, and other countries that oppose Japan's release of nuclear water. NGOs, on the other hand, are involved in various negative contexts, such as Australia's emissions from plastic production and the death of elephants due to drought in Zimbabwe. Notably, articles where NGOs are the main focus tend to report on countries other than China.

Takeaway 3: Offshore media sources tend to feature legal and judicial entities as the main agent of articles with negative valence scores, alongside health and medical entities. In this case, legal and judicial entities are still portrayed as heroes, but the articles are reporting on rather negative events such as the punishment of crimes.

Overall, a negative mean valence does not necessarily indicate that an agent type is portrayed badly, as the experience of a victim could also contribute to a lower valence score. As well, while this thesis will not delve into every agent type in detail, it's important to acknowledge

that even those that constitute a smaller proportion of the data, such as legal and judicial entities and media and communication entities, merit further analysis.

3.4 Distinctive Morals

To identify the most distinctive morals for each region and ideology, I employed the Fightin' Words technique (Monroe et al.). The technique is called "fighting words" because it highlights the terms that metaphorically "fight" for prominence between different groups and helps to uncover linguistic patterns and the specific vocabulary that characterizes the groups being studied. I applied the Fightin' Words method to the extracted moral sentences and moral keywords, both positive and negative, and analyzed the moral variations between different groups and their potential causes. I then contrasted Fightin' Words results with results from traditional text-mining techniques like topic modeling. Topic modeling was conducted using the original news texts to reveal unique topics across different regions and ideologies. Although both methods provide insights into the varying focuses of different groups, I demonstrate that analyzing extracted moral keywords reveals higher-level values conveyed by the news that are not accessible through topic modeling alone.

3.4.1 Fightin' Words

Similar to the G-squared score (also known as Dunning's log-likelihood), Fightin' Words is a technique increasingly utilized in recent years in computational social science to identify words that are significantly associated with different groups in a comparative text analysis. The main goal is to determine which words are disproportionately used by one group compared to another, indicating that these words may be particularly important or distinctive for that group's

communication style or content. This method is employed at various levels within this study. For moral sentences (an example being: "The increasing frequency and severity of extreme weather events due to climate change underscores the urgent need for climate action"), each sentence is tokenized into unigrams, resulting in a table that highlights the most distinctive word for each regional or ideological group. For moral keywords, I opted for tokenization using n-grams (unigrams and bigrams) and the entire keyword phrase, as the concise nature of the keywords allows for easier comparison. As before, all comparisons are conducted across three dimensions: Chinese versus North American, Liberal versus Conservative, and State versus Offshore.

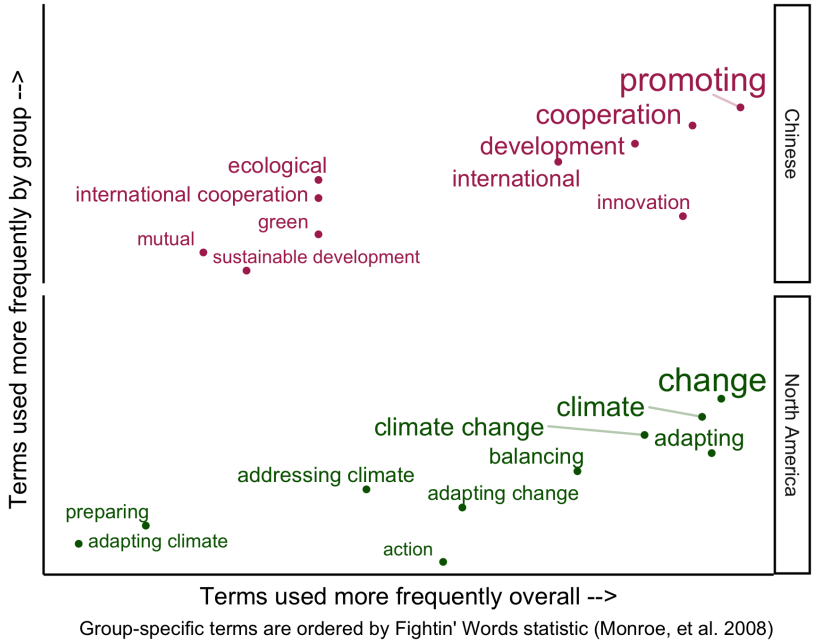

Figure 7. Fightin' Words illustration of distinctive positive moral keywords between North American and Chinese-language news sources

It is found that the most distinctive words from moral sentences are very similar to the n-grams approach on moral keywords, so the analysis presented will be solely on the keywords.[10]

[10] The moral sentences warrant a more thorough analysis and aggregation, currently underway at the .txtlab, with which the thesis author is affiliated. Due to constraints of space and time, this extensive analysis is not included in the thesis.

For positive moral keywords, Figure 7 illustrates distinct moral emphases in Chinese and North American media:

Takeaway 1: Chinese news media prioritize "international cooperation" and "sustainable development," while North American ones stress "addressing" and "adapting to" climate change. The focus of the Chinese sources reflects an active engagement with global environmental issues, which aligns with recent findings that show China playing an increasingly proactive role in climate change, compared to previously attributing responsibilities to developed countries (Pan et al.). In contrast, North American media's focus on "addressing" and "adapting" to climate change is indicative of a more immediate sense of crisis and a reflection of individualistic values. While the more frequent mention of "climate change" in North American media indicates a heightened level of concern, this should not be interpreted as North America experiencing greater impacts of climate change compared to China, which also regularly faces significant climatic challenges. Rather, this trend suggests that China may be less forthcoming in addressing the direct impacts of climate change, focusing instead on the positive potential of technological advancements and sustainable development.

Chinese	**North America**
promoting sustainable development	adapting to change
promoting international cooperation	addressing climate change
promoting environmental sustainability	adapting to climate change
promoting sustainability	compromise
promoting green development	investing in sustainability
promoting ecological civilization	taking action against climate change
mutual respect and cooperation	fiscal responsibility
promoting renewable energy	preparing for natural disasters
listening to community concerns	balancing interests

sustainable development	critical thinking

Table 11. Top 10 positive moral phrases in North American and Chinese sources

The distinctive moral keyword phrases in Table 11 show more nuanced differences between the two regions. Phrases that pop up include “ecological civilization” “renewable energy” and “community concerns” in Chinese-speaking regions, while in North America “compromise,” “fiscal responsibility,” “natural disasters,” and “critical thinking.” In addition, while North America emphasizes investing in sustainability – highlighting a focus on sustainability itself – Chinese discourse on developing sustainably places more emphasis on development. In general, Chinese-speaking regions’ promotion of solutions seems to suggest a strategy of active involvement, unlike North America’s more reactive approach to the consequences of climate change. In a similar vein, the most distinctive negative moral phrases (not shown as a table) include “ignoring environmental sustainability,” “ignoring international cooperation” and “polluting the environment” on the Chinese side, and “ignoring climate change impacts,” “ignoring evacuation orders,” and “greenwashing” on the North American side, which highlights the unique problems that each region is facing or cares about.

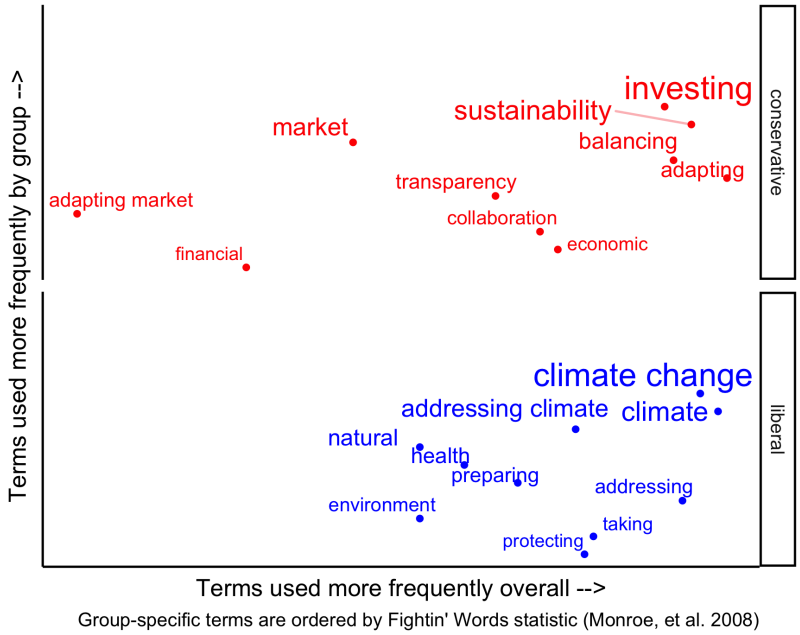

Figure 8. Fightin' Words illustration of distinctive positive moral keywords between Conservative and Liberal news sources

Takeaway 2: Within North America (Figure 8 and Table 12), Conservative outlets heavily emphasize markets, investments, and economic concerns, whereas Liberal outlets focus more on environmental issues and their impact. The Conservative's emphasis on the market reflects skepticism towards climate policies as highlighted in a 2019 Pew Research Center survey, where 62% of Conservative Republicans believed such policies are economically detrimental (Hefferon). A closer examination shows that Conservative media emphasizes "adapting" to market changes and challenges, while Liberal media focuses on "addressing" different environmental issues such as climate change, natural disasters, and pollution. It is evident that Liberal sources feature environmental topics more heavily than Conservative ones. Upon a detailed review of the data, morals exemplified in Conservative sources, such as "Adapting to technological and environmental changes can create new business opportunities and growth prospects for companies" (from an article about the transition to electric vehicle charging by gas stations) and "Adaptability and innovation can transform an industry to become

more inclusive, environmentally conscious, and resilient in the face of challenges" (from an article about the world's largest cruise ship), begin to emerge. These examples prove how the Conservatives' adaptations are predominantly market-driven, occasionally also steered towards sustainable practices. For negative moral keywords, distinctive Liberal sources include "ignoring environmental warnings/health/conservation/justice" and "spreading misinformation," whereas Conservative outlets are "ignoring sustainability," "greenwashing" and "over-regulation."

Conservative	**Liberal**
investing in sustainability	addressing climate change
collaboration	protecting the environment
adapting to change	preparing for natural disasters
adapting to challenges	conservation
adapting to market changes	reading diverse stories
investing in renewable energy	taking action against climate change
fiscal responsibility	protecting public health
diversification	appreciating natural phenomena
investing in innovation	reducing pollution
adapting to market demands	preserving cultural heritage

Table 12. Top 10 positive moral phrases in North American and Chinese sources

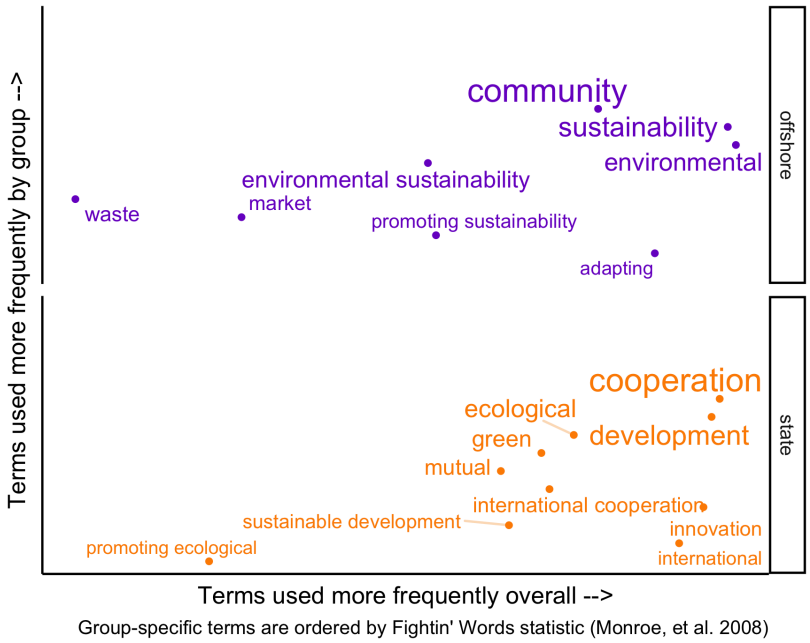

Figure 9. Fightin' Words illustration of distinctive positive moral keywords between offshore and state news sources

Takeaway 3: In Chinese-language media (Figure 9), state outlets frame their discourse around development, with a strong emphasis on sustainability and environmental awareness. In contrast, offshore media in Hong Kong and Taiwan view sustainability as an end goal, focusing on community-level actions such as recycling and local initiatives. This reflects their smaller geographical and socio-economic contexts. State media, however, use "sustainable" and "green" as modifiers to development, suggesting that China sees climate change not just as an environmental hurdle but as a driver of economic growth and opportunity (Pan et al.). With the state media's emphasis on development, one might expect a corresponding focus on market dynamics. Yet, it appears that market-related topics are more prominently featured in offshore sources. Similar to Conservative media, these offshore outlets emphasize "adapting to market changes." Analysis of these articles reveals that they often discuss global changes, including those in the U.S., Europe, and China, in a more reactive manner, potentially due to the distinctive political and economic contexts of Hong Kong and Taiwan. Conversely,

while mainland China undoubtedly also adjusts to market shifts, its robust position in the global market instills a sense of confidence and agency, leading to a proactive rather than reactive stance in its media portrayal. For offshore media, the distinctive negative moral keyword phrases include “ignoring community concerns,” “illegal dumping,” and “wasting resources.” In contrast, state media are characterized by phrases such as “ignoring environmental protection,” “ignoring global cooperation,” “protectionism,” “exploitation,” and “isolationism.”

Offshore	**State**
promoting sustainability	promoting green development
adapting to market changes	promoting international cooperation
promoting environmental sustainability	promoting sustainable development
respecting environmental laws	promoting ecological civilization
adapting to change	mutual respect and cooperation
adopting sustainable practices	international cooperation
investing in sustainability	promoting global cooperation
collaborating for sustainability	embracing innovation and sustainability
adapting to challenges	promoting unity and cooperation

Table 13. Top 10 positive moral phrases in offshore and state sources

3.4.2 Topic Modeling

While we have gained a lot of useful knowledge from the application of the Fightin’ Words method, it is not clear how exactly the extracted morals + Fightin’ Words approach differs from traditional methods like topic modeling. As one of the most widely used text mining methods, topic modeling is used to identify recurring themes or topics within large collections of textual data. It involves algorithms like Latent Dirichlet Allocation (LDA) that group similar words into topics based on their distribution across documents, which helps uncover hidden

semantic structures in the text. The goal of comparing topic modeling to the methodology of this thesis is thus to a) prove the value of the current methodology and b) discover whether topic modeling can be combined with the current method to provide supplemental information.

Because topic modeling is conducted on the original news articles, the topics are extracted into Chinese and English topics separately and thus don't allow for a cross-language comparison, which means that a comparison between North American and Chinese-language sources is unlikely. After testing for the number of topics (k) using perplexity and coherence, I decided that k=15 is the best choice. This results in 15 topics, each containing their most distinctive terms. The topics are labeled manually by taking into consideration the overarching theme captured by the different words. For example, for a topic containing words such as "renewable energy," "carbon," and "project," the topic is named "Renewable Energy & Green Tech." Another example is the topic labeled "Climate Effects & Weather Conditions", whose top key terms include "heat," "climate," "temperatures," and "change." The results, shown in Figure 10 and Figure 11 below, demonstrate the distinctive topics for each ideology. In Figure 10, distinctive Conservative topics are shown on the left of the black vertical line while Liberal ones are on the right.

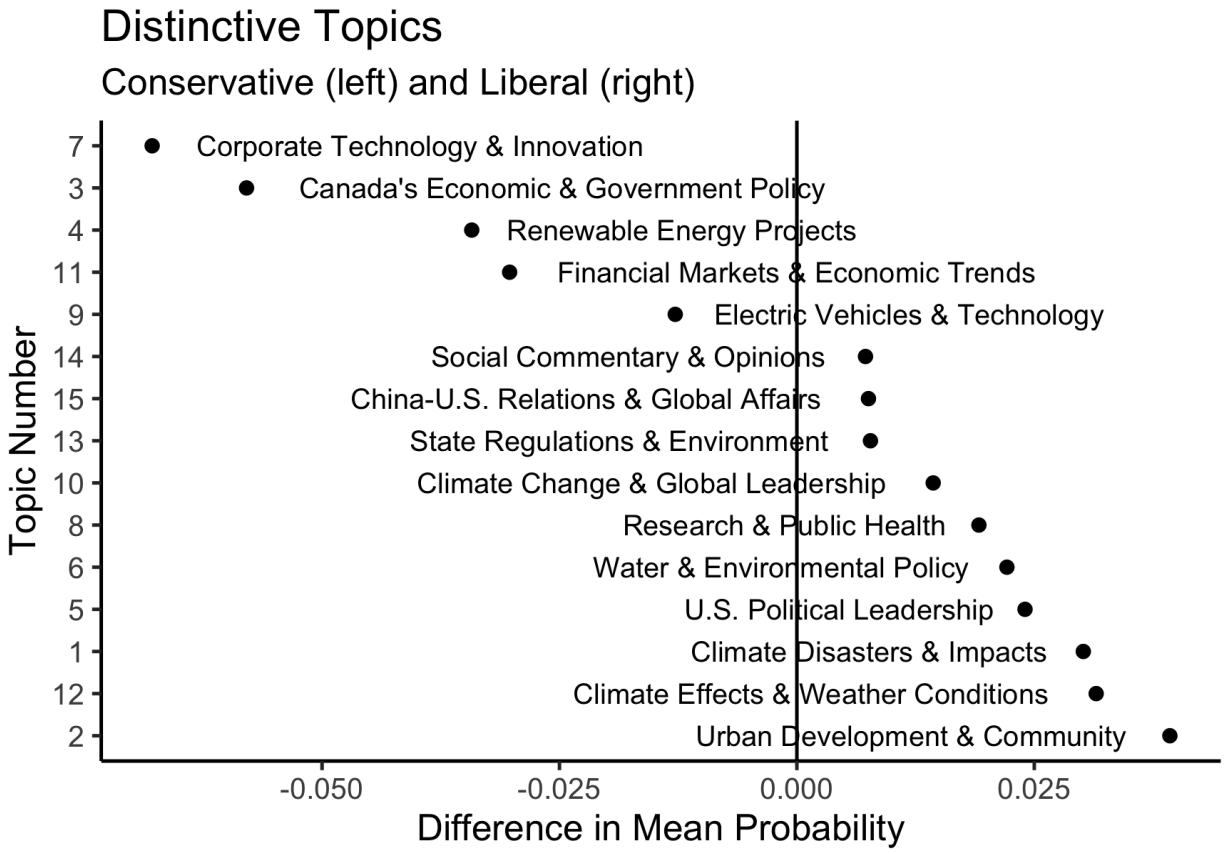

Figure 10. Distinctive topics in Conservative and Liberal media outlets

As shown in Figure 10, there are no surprises in the topics covered, which align with previous discussions. The comparison of topics shows that Conservative sources tend to concentrate more on domestic issues, while Liberal sources focus more on global matters. However, the topic of China-U.S. relations and global affairs, particularly prominent in Liberal sources, hasn't been thoroughly explored in previous sections since GPT has not been specifically asked to identify the main agent in greater detail. This suggests that although China frequently appears as a main agent, the current research setup requires additional information, such as topic keywords or agent names, to clearly identify it. Despite this limitation that is readily addressable, unlike the morals, these topics do not reveal the specific stance of each ideology; for instance, how exactly Liberal sources discuss climate change or how Conservative

sources talk about the market. Specifically, although the distinctively Conservative topic "Corporate Technology & Innovation" includes key terms like "companies," "investors," "market, " and "ai," it is not evident that Conservative sources are readily promoting investment in sustainability and renewable energy, which is only shown through the positive moral keywords. Morals, therefore, offer a deeper dive into understanding the perspectives and values of these media outlets.

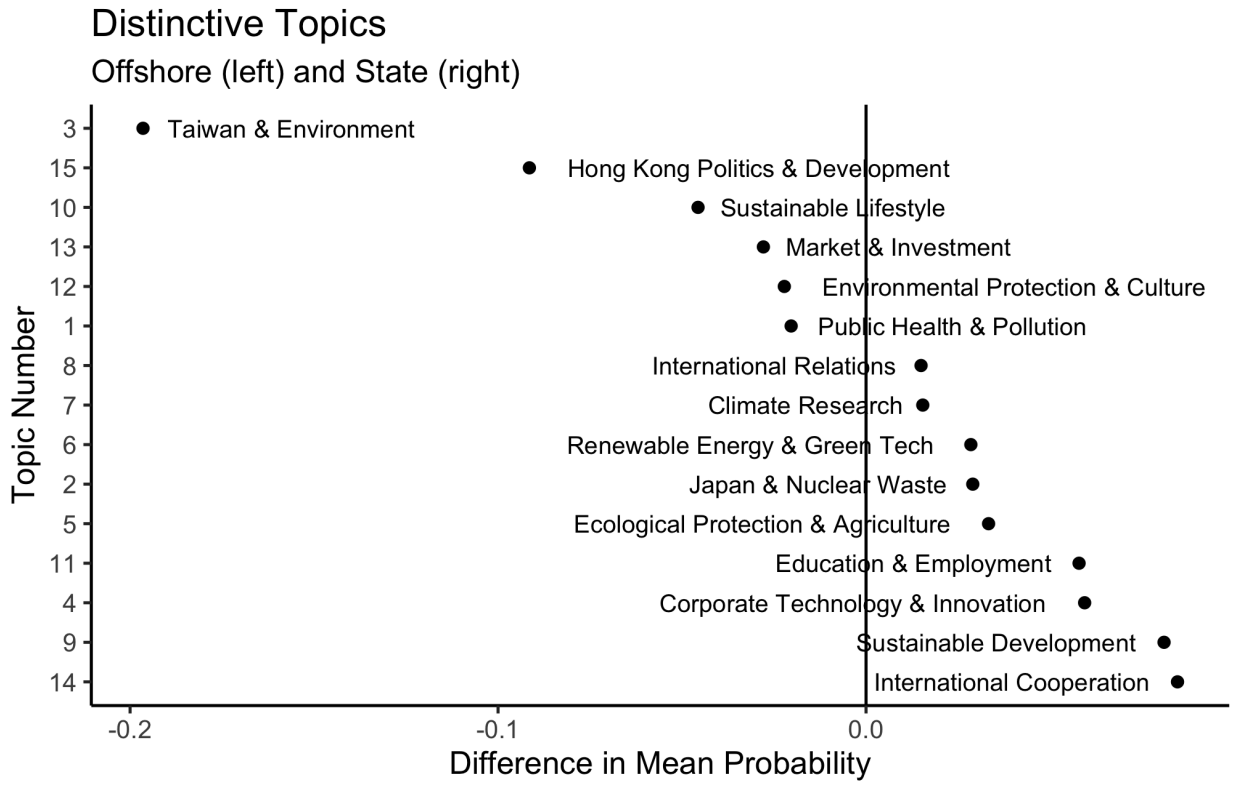

Figure 11. Distinctive topics in offshore and state media outlets

Between state and offshore sources, the findings are consistent with previous analysis in the Fightin' Words section. Interestingly, the topic "International Cooperation" includes mentions of the "One Belt One Road" initiative, a global development strategy heavily promoted by the Chinese government involving infrastructure development and investments in nearly 70 countries and international organizations. The cooccurrence of "One Belt One Road" and

"international cooperation" suggests mainland China's push to gain international acceptance of this initiative. Similar to the previous discussion, topics and morals each provide different kinds of information. However, while topic modeling offers useful insights, it remains somewhat superficial, which is why the analysis of morals, potentially in combination with topic extraction using LLMs, proves valuable, especially as LLMs have been shown to assess topics more effectively, aligning closely with human judgments (Stammbach et al.).

3.5 Conclusion

Throughout this chapter, I have attempted to elucidate intricate dynamics within environmental discourse guided by the methodologies of high-level narrative information extraction through large language models (LLMs). The use of these advanced computational tools has provided unprecedented insights. First, how different agent types – ranging from governments to scientific communities – are portrayed across various media landscapes, shedding light on the intersection of narrative, valence, and moral dimensions in environmental news articles. The findings have also highlighted significant regional and ideological differences in the portrayal of these agents, which demonstrates the influence of narrative frames on public perception and discourse.

Moreover, the use of the Fightin' Words technique has allowed for a nuanced exploration of the moral underpinnings in these narratives. By distinguishing between the moral landscapes of different regions and ideologies, we have gained deeper insights into the values that drive environmental reporting in distinct contexts. For example, while North American outlets frequently focus on "addressing" and "adapting to" climate change, Chinese media emphasize development and international cooperation that is centered around the values of sustainability.

The comparative analysis between the extracted moral keywords and traditional topic modeling further validated the efficacy of the methodology utilized in this thesis. It revealed that while topic modeling can uncover broad thematic patterns, the precision of LLMs in extracting and analyzing moral keywords provides a more detailed understanding of the underlying values and sentiments. As such, this thesis aims to not only advance our comprehension of how environmental issues are framed and understood in different media ecosystems but also demonstrate the value of integrating LLMs into media and cultural analysis. The ability of LLMs to parse complex narrative structures and extract nuanced information offers a significant enhancement over traditional analytical tools, providing richer, more context-sensitive insights into media narratives.

Conclusion

Because this research looks at the differences in environmental news narratives on various regional and ideological levels and through various kinds of narrative information, the takeaways are again summarized below to provide a clearer overview.

Summary of Findings

Valence

In terms of valence, Chinese-language news regarding environmental issues generally presents a more positive outlook compared to their North American counterparts. Additionally, within Chinese media, state-controlled outlets exhibit a higher positive valence than those based offshore, which is likely due to the presence of censorship. In North America, both Conservative and Liberal news sources display similar levels of positivity in their environmental coverage.

Agent Type Distribution

Chinese-language media sources are more likely to discuss local governments, while North American outlets focus more on individuals and households, which reflects the different agents that are central to environmental discourses in each region. Among Chinese-language sources, offshore media frequently mention local governments, whereas state media are more inclined to cover national governments. In North America, Conservative media prioritize industries and businesses as key agents in their reporting, whereas Liberal media tend to emphasize individuals, households, and the scientific community, demonstrating a stronger focus on the impact of environmental problems on everyday life.

Agent Type-Archetype

Chinese media sources frequently portray local governments as heroes, contrasting with North American media, which often emphasize the scientific community as heroic figures. In North America, political entities are commonly depicted as villains. Both Chinese and North American sources show a similar pattern in representing various agents as victims.

Offshore Chinese media outlets are more likely to depict industries and businesses as villains or victims twice as often as state-controlled media. Chinese state media tend to portray Japan negatively, particularly in discussions about national governments, while Taiwanese media generally avoids casting Japan in a villainous role and rarely covers Japan's nuclear wastewater issues.

In North American media, liberal outlets typically celebrate the scientific community as heroes, whereas conservative outlets are more inclined to highlight educational institutions, financial entities, and industries and businesses as heroes. Both liberal and conservative outlets in North America show a similar tendency in their depiction of villains. Additionally, liberal media often portray individuals or households as victims, whereas conservative media are more likely to depict industries and businesses as victims.

Agent Type-Valence

The examination of agent type and valence in Chinese-speaking media sources reveals a tendency towards negative portrayals across various entities. In both state-controlled and offshore media, health and medical entities are often featured in articles that possess a negative valence, highlighting challenges or crises in the health sector, many due to environmental problems. Additionally, in state media, civil society and community groups and NGOs are also commonly associated with negatively valenced articles, indicating a focus on societal issues or conflicts. Offshore media sources frequently depict legal and judicial entities alongside health

and medical entities in a negative light, suggesting a critical stance on governance and public health issues. The North American side, on the other hand, has agent types that do not deviate much from its mean valence.

Distinctive Morals

The analysis of media coverage on environmental topics reveals distinct moral priorities across different regions and political spectra. Chinese news media emphasize morals of "international cooperation" and "sustainable development," highlighting a global and forward-looking approach. In contrast, North American media focuses more on "addressing" and "adapting to" issues like climate change, suggesting a more reactive stance. Within North America, there is a further divergence: Conservative outlets in the region tend to concentrate on markets, investments, and economic implications, whereas Liberal outlets prioritize environmental concerns and their impacts on society. Additionally, state-controlled Chinese-language media frame their climate change discourse around development goals, with a notable emphasis on sustainability and environmental consciousness. Conversely, offshore media in Hong Kong and Taiwan portray sustainability as a community-driven goal, with a strong focus on local actions like recycling and community initiatives.

The comparison with topic modeling shows that, although topic modeling provides useful insights into the distinctive topics addressed in each region and ideology, it does not show how the topics are being talked about or what kinds of values or behaviors are being promoted. This demonstrates the usefulness of taking story morals into consideration to gain a better understanding of the latent narrative messaging of environmental news.

Overall, in this thesis, I have demonstrated a novel methodology that leverages LLMs for the extraction of deeper narrative information such as morals. The methodologies employed in this study offer a robust framework for dissecting the complex interplay of narratives within environmental discourse that involves different elements like agents and morals. The results show distinctive differences between North American and Chinese-language media outlets, between Conservative and Liberal media, and between offshore and state media, in the ways they address the environment and the values they highlight. Hopefully, these insights can not only enrich our academic understanding but also provide practical implications for policymakers, media practitioners, and advocates seeking to enhance the effectiveness of environmental communication.

Limitations and Future Work

As mentioned throughout various parts of this thesis, there are a few limitations that should be addressed in future studies. First, subjectivity in Large Language Models (LLMs) remains a challenge; although the research methodology has been validated with both human and automated validation metrics, the limited number of annotators may not adequately represent a diverse range of age groups and cultural backgrounds. Future research could improve this by expanding the demographic scope of annotators to enhance the generalizability of findings. In terms of the keyword method used for filtering environmental news, the current approach may not be comprehensive and can include irrelevant data. Future efforts should explore alternative methods to reduce noise, as well as include a broader spectrum of media sources to ensure a balanced representation of regional and ideological perspectives. Moreover, while the current study primarily categorizes human agents, it often overlooks instances where the climate acts as

the central agent. Future studies should revisit and expand the agent categories to include non-human elements where relevant.

Additionally, although the thesis extracts various levels of narrative information, there is a lack of detailed analysis due to the constraints of space and time. Future work can thus focus on integrating and comprehensively analyzing these diverse narrative dimensions. For exmaple, utilizing topics identified in articles for comparison with topic modeling results could yield deeper insights, as could a more thorough investigation of the roles of antagonists.

References

Allison, Sarah, et al. "Quantitative Formalism: An Experiment." *N+1*, vol. 13, Jan. 2012, pp. 81–108.

Anderson, Alison. "Media, Politics and Climate Change: Towards a New Research Agenda." *Sociology Compass*, vol. 3, no. 2, 2009, pp. 166–82. *Wiley Online Library*, https://doi.org/10.1111/j.1751-9020.2008.00188.x.

Archer, Jodie, and Matthew L. Jockers. *The Bestseller Code: Anatomy of the Blockbuster Novel*. St. Martin's Press, 2016.

Bolsen, Toby, and Matthew A. Shapiro. "The US News Media, Polarization on Climate Change, and Pathways to Effective Communication." *Environmental Communication*, vol. 12, no. 2, Feb. 2018, pp. 149–63. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17524032.2017.1397039.

Booker, Christopher. *The Seven Basic Plots: Why We Tell Stories*. Bloomsbury Academic, 2006.

Bould, Mark, and Sherryl Vint. *There Is No Such Thing as Science Fiction*. Dec. 2008. *uwe-repository.worktribe.com*, https://uwe-repository.worktribe.com/output/1006830/there-is-no-such-thing-as-science-fiction.

Boykoff, Maxwell T. "Flogging a Dead Norm? Newspaper Coverage of Anthropogenic Climate Change in the United States and United Kingdom from 2003 to 2006." *Area*, vol. 39, no. 4, 2007, pp. 470–81.

Brossard, Dominique, et al. "Are Issue-Cycles Culturally Constructed? A Comparison of French and American Coverage of Global Climate Change." *Mass Communication and Society*,

vol. 7, no. 3, July 2004, pp. 359–77. *Taylor and Francis+NEJM*, https://doi.org/10.1207/s15327825mcs0703_6.

Bushell, Simon, et al. "Strategic Narratives in Climate Change: Towards a Unifying Narrative to Address the Action Gap on Climate Change." *Energy Research & Social Science*, vol. 28, 2017, pp. 39–49, https://doi.org/10.1016/j.erss.2017.04.001.

Chambers, Nathanael, and Dan Jurafsky. "Unsupervised Learning of Narrative Schemas and Their Participants." *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, edited by Keh-Yih Su et al., Association for Computational Linguistics, 2009, pp. 602–10. *ACLWeb*, https://aclanthology.org/P09-1068.

Culler, Jonathan. *Theory of the Lyric*. Harvard University Press, 2015. *www-degruyter-com.proxy3.library.mcgill.ca*, https://doi.org/10.4159/9780674425781.

Dirikx, Astrid, and Dave Gelders. "To Frame Is to Explain: A Deductive Frame-Analysis of Dutch and French Climate Change Coverage during the Annual UN Conferences of the Parties." *Public Understanding of Science*, vol. 19, no. 6, Nov. 2010, pp. 732–42. *SAGE Journals*, https://doi.org/10.1177/0963662509352044.

Dundes, Alan. "From Etic to Emic Units in the Structural Study of Folktales." *The Journal of American Folklore*, vol. 75, no. 296, 1962, pp. 95–105. *JSTOR*, https://doi.org/10.2307/538171.

Durmus, Esin, et al. *Towards Measuring the Representation of Subjective Global Opinions in Language Models*. arXiv:2306.16388, arXiv, 28 June 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.2306.16388.

Entman, Robert. “Framing: Toward Clarification of A Fractured Paradigm.” *The Journal of Communication*, vol. 43, Dec. 1993, pp. 51–58. *ResearchGate*, https://doi.org/10.1111/j.1460-2466.1993.tb01304.x.

Frye, Northrop. *Anatomy of Criticism: Four Essays*. Revised edition, Princeton University Press, 2000.

Genette, Gérard. *The Architext: An Introduction*. University of California Press, 1992.

Gilardi, Fabrizio, et al. “ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks.” *Proceedings of the National Academy of Sciences*, vol. 120, no. 30, July 2023, p. e2305016120. *arXiv.org*, https://doi.org/10.1073/pnas.2305016120.

Graham, Jesse, et al. “Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism.” Advances in Experimental Social Psychology, edited by Patricia Devine and Ashby Plant, vol. 47, Academic Press, 2013, pp. 55–130. ScienceDirect, https://doi.org/10.1016/B978-0-12-407236-7.00002-4.

Guo, Jing, et al. “Authoritarian Environmentalism as Reflected in the Journalistic Sourcing of Climate Change Reporting in China.” *Environmental Communication*, vol. 17, no. 5, July 2023, pp. 502–17. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17524032.2023.2223774.

Hefferon, Cary Funk and Meg. “U.S. Public Views on Climate and Energy.” *Pew Research Center*, 25 Nov. 2019, https://www.pewresearch.org/science/2019/11/25/u-s-public-views-on-climate-and-energy/.

Hofstede, Geert. *Culture’s Consequences: International Differences in Work-Related Values*. SAGE, 1984.

Hulme, Mike. *Why We Disagree about Climate Change: Understanding Controversy, Inaction and Opportunity*. Cambridge University Press, 2009. *Cambridge University Press*, https://doi.org/10.1017/CBO9780511841200.

Jockers, Matthew. *A Novel Method for Detecting Plot Matthew L. Jockers*. https://www.matthewjockers.net/2014/06/05/a-novel-method-for-detecting-plot/. Accessed 7 Apr. 2024.

Jones, Michael D. "Cultural Characters and Climate Change: How Heroes Shape Our Perception of Climate Science." *Social Science Quarterly*, vol. 95, no. 1, 2014, pp. 1–39. *Wiley Online Library*, https://doi.org/10.1111/ssqu.12043.

Jung, C. G. *The Collected Works of C. G. Jung, Volume 9 (Part 1): Archetypes and the Collective Unconscious*. Edited by Gerhard Adler, Translated by R. F. C. Hull, Princeton University Press, 1969. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=1573473.

Keller, Tobias R., et al. "News Media Coverage of Climate Change in India 1997–2016: Using Automated Content Analysis to Assess Themes and Topics." *Environmental Communication*, vol. 14, no. 2, Feb. 2020, pp. 219–35. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17524032.2019.1643383.

Koka, Sahas, et al. "Evaluating the Efficacy of Large Language Models in Detecting Fake News: A Comparative Analysis." *arXiv.Org*, 5 June 2024, https://arxiv.org/abs/2406.06584v1.

Lakoff, George. "Why It Matters How We Frame the Environment." *Environmental Communication*, vol. 4, no. 1, Mar. 2010, pp. 70–81. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17524030903529749.

Lazarus, Richard. “Super Wicked Problems and Climate Change: Restraining the Present to Liberate the Future.” *Georgetown Law Faculty Publications and Other Works*, Jan. 2009, https://scholarship.law.georgetown.edu/facpub/159.

Lecheler, Sophie, and Claes H. De Vreese. *News Framing Effects: Theory and Practice*. Taylor & Francis, 2019. *library.oapen.org*, https://doi.org/10.4324/9781315208077.

Levin, K., et al. “Playing It Forward: Path Dependency, Progressive Incrementalism, and the ‘Super Wicked’ Problem of Global Climate Change.” *IOP Conference Series: Earth and Environmental Science*, vol. 6, no. 50, Feb. 2009, p. 502002. *Institute of Physics*, https://doi.org/10.1088/1755-1307/6/50/502002.

Liang, Xuan, et al. “Exploring Attribution of Responsibility in a Cross-National Study of TV News Coverage of the 2009 United Nations Climate Change Conference in Copenhagen.” *Journal of Broadcasting & Electronic Media*, vol. 58, no. 2, Apr. 2014, pp. 253–71. *tandfonline.com (Atypon)*, https://doi.org/10.1080/08838151.2014.906436.

Lin, Chin-Yew. “ROUGE: A Package for Automatic Evaluation of Summaries.” *Text Summarization Branches Out*, Association for Computational Linguistics, 2004, pp. 74–81. *ACLWeb*, https://aclanthology.org/W04-1013.

Lucy, Li, and David Bamman. “Gender and Representation Bias in GPT-3 Generated Stories.” *Proceedings of the Third Workshop on Narrative Understanding*, edited by Nader Akoury et al., Association for Computational Linguistics, 2021, pp. 48–55. *ACLWeb*, https://doi.org/10.18653/v1/2021.nuse-1.5.

Madrigal, Alexis C. “How Netflix Reverse-Engineered Hollywood.” *The Atlantic*, 2 Jan. 2014, https://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/.

Maynez, Joshua, et al. “On Faithfulness and Factuality in Abstractive Summarization.” *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky et al., Association for Computational Linguistics, 2020, pp. 1906–19. *ACLWeb*, https://doi.org/10.18653/v1/2020.acl-main.173.

McComas, Katherine, and James Shanahan. “Telling Stories About Global Climate Change: Measuring the Impact of Narratives on Issue Cycles.” *Communication Research*, vol. 26, no. 1, Feb. 1999, pp. 30–57. *SAGE Journals*, https://doi.org/10.1177/009365099026001003.

Miotto, Marilù, et al. *Who Is GPT-3? An Exploration of Personality, Values and Demographics*. arXiv:2209.14338, arXiv, 26 Oct. 2022. *arXiv.org*, https://doi.org/10.48550/arXiv.2209.14338.

Mitchell, Amy, et al. “Section 1: Media Sources: Distinct Favorites Emerge on the Left and Right.” *Pew Research Center*, 21 Oct. 2014, https://www.pewresearch.org/journalism/2014/10/21/section-1-media-sources-distinct-favorites-emerge-on-the-left-and-right/.

Monroe, Burt L., et al. “Fightin’ Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict.” *Political Analysis*, vol. 16, no. 4, Jan. 2017, pp. 372–403. *Cambridge University Press*, https://doi.org/10.1093/pan/mpn018.

Moretti, Franco. *Distant Reading*. Verso Books, 2013.

Myrzakhan, Aidar, et al. *Open-LLM-Leaderboard: From Multi-Choice to Open-Style Questions for LLMs Evaluation, Benchmark, and Arena*. arXiv:2406.07545, arXiv, 11 June 2024. *arXiv.org*, https://doi.org/10.48550/arXiv.2406.07545.

Nisbet, Matthew C. “Communicating Climate Change: Why Frames Matter for Public Engagement.” *Environment: Science and Policy for Sustainable Development*, vol. 51, no. 2, Mar. 2009, pp. 12–23. *Taylor and Francis+NEJM*, https://doi.org/10.3200/ENVT.51.2.12-23.

OpenAI, et al. *GPT-4 Technical Report*. arXiv:2303.08774, arXiv, 4 Mar. 2024. *arXiv.org*, https://doi.org/10.48550/arXiv.2303.08774.

OpenAI Platform. https://platform.openai.com. Accessed 16 July 2024.

Pan, Yeheng, et al. “China’s Pathway to Climate Sustainability: A Diachronic Framing Analysis of People’s Daily’s Coverage of Climate Change (1995–2018).” *Environmental Communication*, vol. 15, no. 2, Feb. 2021, pp. 189–202. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17524032.2020.1817766.

Pan, Zhongdang, and Gerald M. Kosicki. “Framing Analysis: An Approach to News Discourse.” *Political Communication*, vol. 10, no. 1, Jan. 1993, pp. 55–75. *Taylor and Francis+NEJM*, https://doi.org/10.1080/10584609.1993.9962963.

Pennington, Jeffrey, et al. “GloVe: Global Vectors for Word Representation.” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, edited by Alessandro Moschitti et al., Association for Computational Linguistics, 2014, pp. 1532–43. *ACLWeb*, https://doi.org/10.3115/v1/D14-1162.

Piper, Andrew. “There Will Be Numbers.” *Journal of Cultural Analytics*, May 2016. *ResearchGate*, https://doi.org/10.22148/16.006.

Propp, Vladimir. *Morphology of the Folktale: Second Edition*. University of Texas Press, 1975.

Rabitz, Florian, et al. “Topic Modelling the News Media Representation of Climate Change.” *Environmental Sociology*, vol. 7, no. 3, July 2021, pp. 214–24. *Taylor and Francis+NEJM*, https://doi.org/10.1080/23251042.2020.1866281.

Radford, Alec, and Karthik Narasimhan. *Improving Language Understanding by Generative Pre-Training*. 2018. *Semantic Scholar*, https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035.

Rawte, Vipula, et al. “The Troubling Emergence of Hallucination in Large Language Models - An Extensive Definition, Quantification, and Prescriptive Remediations.” *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, edited by Houda Bouamor et al., Association for Computational Linguistics, 2023, pp. 2541–73. *ACLWeb*, https://doi.org/10.18653/v1/2023.emnlp-main.155.

Reagan, Andrew J., et al. “The Emotional Arcs of Stories Are Dominated by Six Basic Shapes.” *EPJ Data Science*, vol. 5, no. 1, 1, Dec. 2016, pp. 1–12. *epjdatascience.springeropen.com*, https://doi.org/10.1140/epjds/s13688-016-0093-1.

Reimers, Nils, and Iryna Gurevych. “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks.” *arXiv.Org*, 27 Aug. 2019, https://arxiv.org/abs/1908.10084v1.

Rodrigo-Alsina, Miquel. “Talking about Climate Change: The Power of Narratives.” *Climate Change Denial and Public Relations*, Routledge, 2019.

Romero, Peter, et al. *Do GPT Language Models Suffer From Split Personality Disorder? The Advent Of Substrate-Free Psychometrics*. 27 Mar. 2023. *Research Square*, https://doi.org/10.21203/rs.3.rs-2717108/v1.

Schank, Roger C., and Robert P. Abelson. *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures*. 1st edition, Psychology Press, 1977.

Semetko, Holli A., and Patti M. Valkenburg Valkenburg. "Framing European Politics: A Content Analysis of Press and Television News." *Journal of Communication*, vol. 50, no. 2, June 2000, pp. 93–109. *Silverchair*, https://doi.org/10.1111/j.1460-2466.2000.tb02843.x.

Shanahan, Mike. *Talking about a Revolution: Climate Change and the Media*. https://www.iied.org/17029iied. Accessed 16 July 2024.

Song, Kaitao, et al. "MPNet: Masked and Permuted Pre-Training for Language Understanding." *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 16857–67. *Neural Information Processing Systems*, https://proceedings.neurips.cc/paper/2020/hash/c3a690be93aa602ee2dc0ccab5b7b67e-Abstract.html.

Stammbach, Dominik, et al. "Revisiting Automated Topic Model Evaluation with Large Language Models." *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, edited by Houda Bouamor et al., Association for Computational Linguistics, 2023, pp. 9348–57. *ACLWeb*, https://doi.org/10.18653/v1/2023.emnlp-main.581.

Strachan, James W. A., et al. "Testing Theory of Mind in Large Language Models and Humans." Nature Human Behaviour, vol. 8, no. 7, July 2024, pp. 1285–95. www.nature.com, https://doi.org/10.1038/s41562-024-01882-z.

Sui, Peiqi, et al. *Confabulation: The Surprising Value of Large Language Model Hallucinations*. arXiv:2406.04175, arXiv, 25 June 2024. *arXiv.org*, https://doi.org/10.48550/arXiv.2406.04175.

Team, Gemini, et al. “Gemini: A Family of Highly Capable Multimodal Models.” *arXiv.Org*, 19 Dec. 2023, https://arxiv.org/abs/2312.11805v3.

Thompson, Stith. *Motif-Index of Folk-Literature; a Classification of Narrative Elements in Folk-Tales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends*. -36., 1932.

Tong, Jingrong. “Environmental Risks in Newspaper Coverage: A Framing Analysis of Investigative Reports on Environmental Problems in 10 Chinese Newspapers.” *Environmental Communication*, vol. 8, no. 3, July 2014, pp. 345–67. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17524032.2014.898675.

Touvron, Hugo, et al. “LLaMA: Open and Efficient Foundation Language Models.” *arXiv.Org*, 27 Feb. 2023, https://arxiv.org/abs/2302.13971v1.

Underwood, Ted. “Genre Theory and Historicism.” *Journal of Cultural Analytics*, vol. 2, no. 2, Oct. 2016. *culturalanalytics.org*, https://doi.org/10.22148/16.008.

Uther, Hans-Jörg. *The Types of International Folktales: A Classification and Bibliography, Based on the System of Antti Aarne and Stith Thompson*. Suomalainen Tiedeakatemia, Academia Scientiarum Fennica, 2004.

Vaswani, Ashish, et al. *Attention Is All You Need*. arXiv:1706.03762, arXiv, 1 Aug. 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.1706.03762.

Wang, Zengzhi, et al. *Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study*. arXiv:2304.04339, arXiv, 17 Feb. 2024. *arXiv.org*, https://doi.org/10.48550/arXiv.2304.04339.

Xie, Lei. “The Story of Two Big Chimneys: A Frame Analysis of Climate Change in US and Chinese Newspapers.” *Journal of Intercultural Communication Research*, vol. 44, no. 2,

Apr. 2015, pp. 151–77. *Taylor and Francis+NEJM*, https://doi.org/10.1080/17475759.2015.1011593.

Young, Nathan, and Eric Dugas. "Representations of Climate Change in Canadian National Print Media: The Banalization of Global Warming." *Canadian Review of Sociology/Revue Canadienne de Sociologie*, vol. 48, no. 1, 2011, pp. 1–22. *Wiley Online Library*, https://doi.org/10.1111/j.1755-618X.2011.01247.x.

Zhao, Wanying, et al. *Discovering Collective Narratives Shifts in Online Discussions*. arXiv:2307.08541, arXiv, 17 July 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.2307.08541.

Zhu, Lixing, et al. "Are NLP Models Good at Tracing Thoughts: An Overview of Narrative Understanding." *Findings of the Association for Computational Linguistics: EMNLP 2023*, edited by Houda Bouamor et al., Association for Computational Linguistics, 2023, pp. 10098–121. *ACLWeb*, https://doi.org/10.18653/v1/2023.findings-emnlp.677.