MASTER’S THESIS IN INFORMATION SCIENCE
FACULTY OF LIBRARIANSHIP, INFORMATION, EDUCATION AND IT
2025

Do we live in the Twilight Zone?

- a study of information resilience and filters of perception in times of climate change.

Lovisa Rydén

UNIVERSITY
OF BORÅS

English title:	Do we live in the Twilight Zone? - a study of information resilience and filters of perception in times of climate change.
Author:	Lovisa Rydén
Completed:	2025
Abstract:	This thesis analyses 205,792 English-language YouTube comments on “news” videos (2010–2024) to examine how perceptual filters shape public discourse. A mixed-method design uses fine-tuned RoBERTa (hybrid inference) to estimate seven indicators – sentiment, irony, trust/distrust, hostility, and the misinformation-related stances concept, belief, accusation – and qualitatively aligns yearly patterns with Climate Change in the American Mind (CCAM). Results show a shift toward negative sentiment, widening gaps between distrust and trust, persistently elevated hostility, and irony functioning as distancing. Belief and accusation rise in later years, while concept remains rare. The study argues for complementing information literacy with an information resilience lens that addresses conditions of engagement based on worldview and perception filters.
Keywords:	Information science, information resilience, perceptual filters, climate change, climate denial

Content

1 Introduction

Social media platforms amplify misleading or false content at unprecedented speed, while algorithmic infrastructures personalise and reinforce exposure to information that confirms existing views (Haider & Sundin, 2019, p.34–37). Hence, misinformation, disinformation, and malinformation have become defining challenges of the digital age. Bawden & Robinson (2020) describes three negative forms of problematic information that have a negative impact on an individual, group or society. These are misinformation, which is false (inaccurate) information; disinformation, which is deliberately false information (lies); and malinformation, which is true information (truth) used in a malicious way and/or to inflict harm (p.340-341). Research has shown that these types of information spread more rapidly and widely than factually accurate and objective information, driven in part by emotional resonance and novelty (Vosoughi et al., 2018, p.1146; Haider & Sundin, 2022, p.8-7). In such an environment, societies face what Haider and Sundin (2019) describe as a paradox: while access to information appears ubiquitous, the reliability and democratic quality of that information is increasingly uncertain. For decades, information literacy was regarded as the principal defence against misinformation. By strengthening individuals' critical skills in evaluating sources and claims, it was thought possible to build collective resistance to misinformation, disinformation and malinformation (p.102–104, 107–109). However, increasingly there have been indications that something else is causing the issues, which has sparked an interest in information resilience within the information science field (Haider & Sundin, 2022, p.30).

Librarians and other information specialists have more roles in society than the stereotypical roles associated with them. Ahlryd and Hanell (2024) show the important role that librarians played during the corona pandemic in sorting credible, good information from misinformation in the floods of information for the benefit of hospital staff. To ensure that the doctors and nurses had access to the most recent and accurate information, they sifted through the troves of information coming in, they created knew cataloguing systems, they incorporated the library's function into other hospital systems to ensure that information could still be sent throughout the hospitals and more (p.273-275). At the same time, public librarians got to work adapting the public libraries so that citizens would still have access to their democratic right to the library. This was done through minimizing contact between people, limiting the time-period each user got to spend in the library each visit, and moving as much of the daily operation of the library online (Ahlryd & Hanell, 2024, p.276-278). These are some of the roles and functions that librarians took on during the pandemic, and you see them do the same during similar during other societal crisis. It was also during the pandemic that the elected leaders of Sweden found out, apparently for the first time, that they were not allowed to close down the libraries, even during a public emergency (Lindström, 2020, December 23). This raises profound questions about how aware the Swedish government is, not just about the role that libraries have in democratic countries, but about the central role in emergencies for information specialists, like librarians, as well as the information science field at large.

Furthermore, while Russia's war in Ukraine and the rearmament of Europe is ongoing, and the discussions of how best to protect nations as well as allies are everywhere. Even as conversations take place within the fields of information science and library and information science, no serious dialogue about what role librarians and other information specialists will have in wartime are seems to be taking place in government – even after the political leaders have been asked. This despite the fact that Russia and other adversaries are already attacking all systems, including information systems, they can get access to, as well as trying to disrupt social cohesion and effect public opinion about the society in question through campaigns of disinformation (Haider & Sundin, 2022, p.126, 135-136). The suspected and alleged success of Russia's influence during both the 2016 US election, as well as the 2016 UK Brexit referendum, as well as paying American Youtubers for promoting ideas they agreed with during the 2024 US election, has likely only emboldened them ((U.S. Department of the Treasury, 2024, "RT EXECUTIVES' and RaHDit Members' Russian Government Influence Operations," para. 1-2). This is made clear by the sheer amount of bots and trolls in the comment sections and forums of any topic that Russia uses to sow division. The majority of people today get their news from social media, where countries like Russia remains ever-present and invest in widespread disinformation campaigns, using the algorithms to their benefit (Llewellyn et al, 2019, p.39-40; Haider & Sundin, 2022, p.126, 136). This explains why many people express having the feeling of living in the Twilight Zone – that the turbulence that is happening in the world today makes it seem impossible and unreal. At the same time, the research within the information science field into what causes vulnerability and how to counteract it is still lacking, while the one thing that is becoming increasingly clear is that information literacy isn't the full answer.

In few information contexts is this more palpable than in the context of climate change. Climate change is not only a scientific and political issue, but also an existential threat that threatens humans view of their control over their world. Myers (2014) argues that climate denial and avoidance are best understood not as a lack of knowledge but as coping mechanisms against the destabilising anxiety provoked by the recognition of climate change as an existential crisis (p.53–60, 66). Similarly, studies of climate information behaviour show that engagement is shaped less by objective assessments of information needs and more by affective and social factors such as trust, the norms of the social group (tribalism), and emotional needs (Choo, 2023, p.1087, 1097–1098). Hurley and Molloy (2025) further demonstrate that shifts in beliefs of individuals around climate change often depend on and happens subsequently to changes in social and experiential context, rather than on exposure to new information on its own (p.6–9, 14–16). These findings underscore that vulnerabilities to misinformation are not simply a matter of deficits in information literacy, but are embedded in the ways people experience, interpret, and filter information in relation to their perception and view of the world around them.

This thesis studies these challenges by focusing on the role of perceptual filters in shaping information behaviour under conditions of existential threat. Rather than assuming that individuals process information in a neutral manner, the study examines how public discourse reflects orientations of trust, distrust, hostility, irony, sentiment, and references to misinformation across time, and how these orientations might make societies more or less vulnerable to misleading narratives. The empirical case is drawn from YouTube comments on news-related videos spanning the years 2010 to 2024. YouTube was selected as a platform because of its dual role as both a mass media archive and a social media forum: it is simultaneously a site of information dissemination and of public commentary. Approximately 200,000 unique English-language YouTube comments were collected, pre-processed, and analysed using supervised NLP methods. The analysis is anchored in theoretically motivated indicators – general sentiment, irony distrust, hostility and misinformation-related indicators – treated as indicators for perceptual filters. Trust is reported descriptively for balance but is not used as a primary theoretical indicator. The analysis employs a mixed-method design combining large-scale natural language processing with comparative survey data. The modelling results are contextualised against survey findings from the Climate Change in the American Mind (CCAM) corpus, which provide longitudinal benchmarks of public opinion in the United States. By triangulating platform data with population-level survey results, the study examines both how online discourse evolves over time and how it relates to broader social trends in climate perception.

1.1 Problem formulation

For a long time, information literacy has been seen as the solution to the growing “infodemic” and information overload that today’s society is exposed to, as well as the best way to combat misinformation, disinformation, and malinformation. In recent years, holes in this safety wall have become more obvious, especially when external influences and events – such as the accelerating impacts of climate change, pandemics, or geopolitical crises – reshape people’s perception of the world around them. If the field of information science, as well as society as a whole, are not aware of how such events recalibrate everyday sense-making, they risk creating continual vulnerabilities in which patterns of interpretation harden and misleading narratives gain durability. One way to address this is to examine how perceptual filters shape what people can take in, and how this relates to information resilience in the face of sustained external pressures.

Building on the theoretical framework, these perceptual filters are understood as expressions of underlying worldviews: people are always already attuned in ways that filter how information appears to them. Such worldviews condition what is noticed as credible, threatening, or irrelevant, and prolonged external pressures can recalibrate them; if left unexamined, they may harden into ongoing patterns that perpetuate vulnerability over time.

The study therefore examines how perceptual filtering manifests in a longitudinal corpus of YouTube news comments (2010–2024) using four metric families that serve as discursive proxies for attunements: sentiment (including irony-aware sentiment), trust/distrust, hostility, and misinformation-related indicators. By analysing these metrics across time – and by juxtaposing them with external opinion corpus – the thesis investigates how the comment milieu's orientation evolves under external pressures and what this implies for information resilience.

1.2 Aim

The aim is to describe and interpret how perceptual filters show up in public discourse over time, and what this means for information resilience. The study traces year-by-year patterns in YouTube comments (2010–2024) for a small, transparent set of indicators—sentiment, irony, trust/distrust, hostility, and the misinformation-related categories concept, belief, and accusation—examines how they move in relation to one another, and situates those movements against *Climate Change in the American Mind* (CCAM) without making causal claims. A final step is to assess whether the experimental RoBERTa-based approach is adequate for drawing year-level conclusions. The purpose is a defensible way to read the discourse environment—reserved vs. open perceptual profiles—and to clarify what follows for information resilience.

1.3 Research questions

1. How do the indicators sentiment, irony, trust/distrust, hostility, concept, belief, and accusation in YouTube comments collected from news-related videos change over time from 2010 to 2024?

2. How do these indicators move in relation to one another, and how do they align with public perceptions of climate change found in *Climate Change in the American Mind* (CCAM)?

3. How do year-to-year shifts in these public perceptions relate to changes in public discourse on YouTube, and what do these relations indicate about people's perceptual filters and worldview?

4. What do the trajectories, the relations between indicators, and possible CCAM alignments indicate about the impact of perceptual filters – as well as climate change itself – on information resilience?

5. Given the experimental aspects of the study, how adequate and trustworthy is the methodological approach used and what limits follow for interpretation?

1.4 Delimitations and scope

This thesis analyses English-language YouTube comments posted on videos collected through the query "news" between 2010 and 2024. Music videos and obviously off-topic genres are excluded by design. The study focuses on comment text (not video content or user profiles) and reports aggregate patterns per calendar year. The benchmark comparison uses published CCAM results; no CCAM data is reanalysed. The aim is to examine discursive indicators (sentiment, irony, trust/distrust, hostility, and misinformation-related indicators) as proxies for perceptual filters, not to adjudicate the truth of specific claims.

1.5 Thesis disposition

The thesis follows the following structure:

- Chapter 1 – Introduction:

  Establishes the problem formulation and rationale, outlines the research questions, and situates the study within Information Science. It also states delimitations and scope, previews the contributions, and points forward to the analytical strategy used later in the thesis.

- Chapter 2 – Theoretical framework:

  Develops the conceptual lens used in the study. It introduces attunement as orientation in information environments, connects this to information resilience, and positions mis/dis/malinformation as discursive orientations rather than case-by-case truth adjudication. The chapter closes with a concise summary that maps constructs to the operational indicators used later (sentiment, irony; trust/distrust; hostility; and misinformation-related labels).

- Chapter 3 – Literature review:

  Surveys prior research in Information Science relevant to the study's focus, including platformed publics and information practices, risk information seeking and processing, and scholarship on misinformation and trust/distrust in socio-technical systems. The review identifies gaps that motivate the present design and clarifies how the thesis contributes to these debates.

- Chapter 4 – Method:

  Presents the mixed-method design. It specifies the data sources and sampling frame for YouTube comments (2010–2024), documents the experiment execution pipeline (pre-processing, language identification, deduplication, modelling and calibration), and details the protocol for compiling and harmonising CCAM toplines as a contextual benchmark. Ethical considerations (including non-retention of usernames), reliability/validity considerations, and limitations are made explicit. The chapter ends with a short closing that recaps methodological coherence.

- Chapter 5 – Results:

  Reports empirical findings without interpretation. It describes corpus characteristics, model diagnostics, and yearly trajectories for each indicator. It then presents descriptive alignments with CCAM at the year level, accompanied by tables and figures placed in the main text or appendix as appropriate.

- Chapter 6 – Analysis:

  Interprets the results through the theoretical framework. Indicators are treated as proxies for attunements, and co-movements with CCAM are read as contextual evidence for shifts in orientation under external pressures. The analysis addresses each research question directly and notes where evidence is strong, mixed, or absent.

- Chapter 7 – Discussion:

  Explicates implications for Information Science and practice, including how the findings speak to information resilience in platformed environments. It reflects on contributions, situates the work in relation to prior studies, reviews limitations and potential biases, and outlines avenues for future research.

- Chapter 8 – Conclusion:

  The conclusions of the study are presented and potential opportunities for future research are discussed.

2 Theoretical framework

The theoretical framework of this thesis builds on Otto Friedrich Bollnow's phenomenology of attunements (Stimmungen) to conceptualize how individuals filter information, construct meaning, and respond to disruptive events such as climate change and misinformation. This framework is integrated with research in information science on information literacy, perceptual filters, and information resilience, allowing an analysis of how affective and existential orientations condition information behaviour in complex media environments.

2.1 Worldview: attunement as the basis of experience

As Gerhard Thonhauser (2019) explains, the phenomenological framework by Otto Friedrich Bollnow is centred in the view that human beings are never neutral observers of the world. Instead, human beings exist in the "condition of an entity existing in the mode of being-in-the-world" (Thonhauser, 2019, p.1). This is something that Bollnow referred to as "being attuned". Attunements are not transient psychological states but the foundational background against which all cognition, perception, and action occur. They are not directed at specific objects, as emotions are, but suffuse the whole of experience. Cheerfulness, melancholy, anxiety, or despair are not simply feelings but global modes of being that structure what is noticed, how events are interpreted, and which responses appear possible. In their essence, attunements are a subconscious, primal form of self-awareness (Thonhauser, 2019, p. 1-2, 5-6). This means that attunements act as filters of perception that form each person's worldview. They determine, for instance, whether a piece of information is encountered as threatening, credible, or irrelevant, and whether it is safe for a person to consider a certain piece of information at all.

2.1.1 Plurality and Transformation of Attunements

As Thonhauser (2019) highlights, a central contribution of Bollnow's theory is his insistence on the plurality of attunements. Each attunement discloses a different dimension of existence and transforms the individual's relation to others, to the world, and to time. Elevated attunements, such as cheerfulness or gaiety, open individuals to community, sociability, and possibilities for action. Chastened attunements, such as despair or melancholy, can close individuals off from others, making the world appear resistant and hostile. Likewise, attunements shape temporal experience: boredom makes time drag, joy makes it seem to pass quickly, and in retrospect these proportions can reverse (p. 7–9). For this thesis, this plurality is important. It allows an analysis of how collective expressions of trust, distrust, hostility, irony, and sentiment in online discourse may be understood not merely as feelings of individual opinions, but as manifestations of underlying perception filters. In Bollnow's terms, these are existential orientations that filter how information is absorbed. In this thesis, the terms attunements, filters of perception, and perception filters are used interchangeably (Thonhauser, 2019, p.7-9).

Something that can be used as an example of an attunement that shapes a person's perception in relation to the world is climate change, as explained by Myers' (2014) account clarifies how attunements shape the uptake of information by linking climate-related fear to the protection of a person's worldview. When the possibility of loss or collapse is felt, people may avoid or downplay information that threatens their sense of security; this is not simply a lack of knowledge, but a way of keeping anxiety at bay (p.53–60). In that setting, denial can become a longer-term coping practice rather than a momentary reaction, and it can be reinforced in groups where shared narratives stabilise the same avoidance. Read through the present framework, this means that what appears in discourse as more negative tone, greater reliance on indirect expression (e.g., irony), rising distrust, and antagonistic exchanges can be understood as outward signs of an attunement that guards the worldview against perceived threat. In other words, the attunement does not only colour how information is felt; it conditions which information is admissible at all and how it is handled in public talk (Myers, 2014, p. 57–60, 66).

2.2 Information Resilience

The concept of resilience originates in the natural sciences, where it refers to the capacity of a substance to withstand external stresses and return to its original form (Vårheim, 2016, p.1-2). In psychology, resilience has been defined as the ability to adapt and recover in the face of adversity, trauma, or significant stress. In systems theory, it has been understood as the ability of a system to absorb disturbance and reorganize while retaining essential functions and structures. In the field of cybersecurity, resilience is conceptualized as the ability of systems to prepare for, respond to, and recover from attacks, ensuring continuity of operations (Munro, 2013, par.3-4). When entered into the context of the field of information science, resilience becomes a way of analysing and studying how individuals, communities, and institutions withstand and tolerate disruptions in their information environments. It extends beyond the skills and understanding of information literacy and addresses the broader capacity to withstand, adapt to, and recover from the destabilizing effects of misinformation, disinformation, and malinformation on both individual and societal levels (Haider & Sundin, 2022, p.30-31).

Ahlryd and Hanell (2024) describe information resilience as “a process where adaptive information capacities are combined and utilized to enable functioning and adaptation after a disturbance” (p.268). They identify three key attributes: robustness, referring to the durability of resources and practices; redundancy, referring to the presence of alternative pathways and channels; and rapidity, referring to the speed of accessing and mobilizing information. Since this study has perceptual filters and worldview as its main focus, studying the vulnerability that these might pose both to society as it is, as well as to the currently existing counter-measures (such as information literacy), the attributes in this study will focus on measurements connected to Bollnow’s attunements (Ahlryd & Hanell, 2024, p.270-271, 282-283; Thonhauser, 2019, p.3-5, 7-9).

When transposed into information science, resilience concerns the capacity of individuals and institutions to withstand, adapt to, and recover from disruptions in their information environments (Ahlryd & Hanell, 2024, p.270-271). In this thesis, that capacity is analysed by using measurable components of perceptual filters to indicate the state of perceptual filters. These components are sorted into the following indicator categories: sentiment, irony, trust/distrust, hostility, as well as the misinformation-related categories of concept, belief, and accusation. Sentiment captures the prevailing sentiment and its refracting through ironic expression. Irony captures how constructive the general discourse is, and can indicate a need for emotional distancing, emotional protection or the need to release frustration – especially when viewed in relation to the sentiment indicator. Trust/distrust is an indicator of whether commenters perceives that people or society is trustworthy or not. Hostility represents the amount of vile language used – not just angry or sarcastic comments, but comments that uses especially hostile and vile language. Concept is an indicator for neutral mentions of the concept of misinformation itself. Belief reflects mentions of believing something that is misinformation – such as conspiracy theories. Finally, accusation reflects accusations of something being or someone spreading misinformation. As for the separate kinds of negative information, disinformation and malinformation are not distinguished or treated as separate concepts. Hence, the term ‘misinformation’ is used as a representative term for all types of negative information.

2.3 Summary of the theoretical framework

This framework connects Bollnow's account of attunement – as the background that shapes perception, interpretation, and possible responses – with an information resilience perspective that examines how individuals, institutions, and societies handle disruptions in their information environments (Thonhauser, 2019, 1-2, 5-6; Ahlryd & Hanell, 2024, p.270-271). The analysis uses measurable components of perceptual filters to describe indications of worldview over time. These components are organised as the following indicators: sentiment, irony, trust/distrust, hostility, and misinformation-related categories: concept, belief, accusation. Taken together, the indicators form a picture of the perceptual filters and worldview found in the YouTube comment corpus and in the CCAM data corpus.

3 Background and previous research

This chapter reviews prior work on how people communicate and interpret information in platformed environments and how these phenomena have been studied to date. It brings together research on sentiment analysis and natural-language processing for social media, studies of information literacy and platform infrastructures that shape visibility and credibility, and research on information resilience in relation to climate-related communication, risk perception, and climate anxiety. The aim is to provide the background for this study and place it within its wider research context.

3.1 Sentiment analysis and social media

Social media platforms, in all their forms, are melting pots of human opinion, and as they have become central parts of most people's lives, huge amounts of human-generated data can now be found within them. Bravo-Marquez et al. (2014) describe the importance of emotion in human communication, pointing out that understanding the emotional load, not just in verbal but also in written communication, is essential in order to understand it's meaning, and – in turn – human interactions. This means that in any analysis of human behaviour, sentiment analysis is a crucial tool when conducing different forms of social analysis. These forms of analysis, used on a collective level, can be used to measure the impact and/or polarization of different situations on groups of people (p.87-88).

Sentiment analysis is conducted by using techniques from text analytics, NLP, machine learning, and linguistics and can be used on both smaller and larger datasets, as well as on both longer and shorter types of text. Sentiment is calculated through polarity which is a numeric score based on the positive and negative aspects of a text document, as well as subjectivity parameters calculating a score of whether it's more likely to be fact or opinion, based on specific words and phrases expressing feelings and emotion. Neutral sentiments typically have a 0 polarity, since it does not express any specific sentiment, while positive sentiments have polarity > 0, and negative sentiments are < 0. Specific threshold can be set for positive sentiment and negative sentiment, depending on the type of data set (Sarkar, 2019, p.567-568; Bravo-Marquez et al., 2014, p.87-88).

Since the beginning of social media, sentiment analysis research has become even more important, since it gives access to the analysis of Big Data datasets that were previously very difficult to achieve in research. However, this has brought new challenges, as the datasets are now too big to be processed manually, which is why computational sentiment analysis methods have increasingly become more relevant (Bravo-Marquez et al., 2014, p.87-88). Sarkar (2019) also addresses issues that have arisen since these huge datasets became part of research, such as managing to process and analyse them in a meaningful and accurate way. The majority of machine learning and data analysis methods that exist today struggle with reading and interpreting unstructured data, such as text, in a meaningful way, and are generally more efficient at analysing numbers (p.1-2). To combat text-related issues, methods and tools have been developed that are specifically built for analysing natural language. Natural language are languages that naturally developed in humans – such as English - unlike programming languages and similar artificial languages (Sarkar, 2019, p.1-4). This is done through Natural Language Processing (NLP), which are methods that enable computers to process and analyse natural language datasets, for example datasets taken from human interaction on social media. Each of these methods have frameworks and libraries for natural language processing and text mining with their own features and capacities which yield different outputs (Sarkar, 2019, p.62, 111).

3.1.1 Sentiment analysis approaches and machine learning methods

There are various methods for machine learning, including unsupervised learning, supervised learning, reinforcement learning, and deep learning. Each of these approaches have different techniques and algorithms, can be applied to different types of texts and datasets and have different advantages and disadvantages. A machine learning model, in turn, is a combination of data and algorithms, and once it is trained on a dataset, the model can be put to use analysing new datasets that the model has never analysed before. This makes sentiment analysis far more effective and means it can be applied to large datasets (Sarkar, 2019, p.134, 275-276).

Supervised machine learning

Supervised sentiment analysis builds on the availability of annotated datasets, where each text sample is labelled with its corresponding sentiment category. Machine learning models are trained to recognize patterns in linguistic features that correspond to these categories and that can then be applied to the unseen data. Early supervised approaches relied on models such as Support Vector Machines, and Logistic Regression, while more recent work frequently employs deep learning models and transformer-based architectures such as BERT or RoBERTa (Sarkar, 2019, p.194-197, 567-568; Camacho-Collados et al., 2022, p. 38–40).

The strength of supervised methods lies in their ability to achieve high accuracy when trained on sufficiently large and representative datasets. However, supervised approaches are constrained by the domain dependence of their training data. A model trained on movie reviews, for example, is less likely to perform well on political commentary or human rights discourse without retraining. Furthermore, the production of labelled datasets is labour-intensive, and also raises consistency and subjectivity in annotation Camacho-Collados et al., 2022, p. 38–40).

Unsupervised machine learning

Unsupervised sentiment analysis, by contrast, does not rely on labelled training data. Instead, it infers sentiment orientation from linguistic resources such as sentiment lexicons (e.g. VADER) or from statistical associations between words in a corpus. In these approaches, each word or phrase is assigned a polarity score, which is then aggregated to produce an overall sentiment value for a text. Some unsupervised approaches also rely on clustering or topic modelling, grouping texts based on similarity without predefined labels (Sarkar, 2019, p.p.111, 268, 567-568; Camacho-Collados et al., 2022, p. 38–40.

The key advantage of unsupervised methods is that they can be applied to large, unlabelled datasets without the resource-intensive process of manual annotation. This makes them attractive for rapidly exploring new domains, such as interactions on social media. However, their reliance on fixed lexicons or surface-level co-occurrence patterns means they often perform poorly in capturing contextual nuance, irony, or culturally specific sentiment expressions (Sarkar, 2019, p.268, 275-276).

3.1.1.1 VADER and TextBlob

According to Ahmad et al. (2025), VADER is a is a lexicon and rule-based sentiment analysis tool that is designed for short, informal text such as social-media posts; it uses a list of positive and negative words plus simple rules to estimate how positive or negative a message sounds (p. 6). TextBlob is a general-purpose Python library that makes common text tasks – like sentiment analysis, basic classification, noun-phrase extraction, and part-of-speech tagging – easy to run with a few lines of code (Ahmad et al., 2025, p. 5). Both tools are quick and transparent, which is why they are widely used; however, they can miss the mark when people write with sarcasm or irony, or when language use drifts from the dictionaries they rely on. Studies in other settings show that VADER and TextBlob can sometimes come close to human judgments (e.g., modelling peer-reviewer opinions and predicting funding outcomes) even without heavy setup (Alabi et al., 2021, p. 1274–1275).

3.1.1.2 TweetNLP

Camacho-Collados et al. (2022) present TweetNLP as a toolkit designed for the everyday realities of social-media language rather than for long, edited prose. In their account, most prior NLP effort has optimized models on curated sources – articles, blogs, Wikipedia – where sentences are complete, spelling is standard, and context is plentiful; Twitter, by contrast, is fast, noisy, and short (p. 37–38). The authors emphasize features that routinely trip up general-purpose models but are routine online: slang and abbreviations, multilingual mixing, technical jargon, typos and informality, immediacy, multimodal markers (emoji, GIFs, memes), and strict length limits that strip away context (Camacho-Collados et al., 2022, p. 37–38). Their core claim is simple: if systems are to perform well on standard NLP tasks in social-media settings – sentiment, part-of-speech tagging, normalization, named-entity recognition – they must be trained and evaluated with these constraints in view rather than treated as miniature versions of edited text (Camacho-Collados et al., 2022, p.37–38). Put plainly, TweetNLP is built for the kind of "messy real-time talk" people actually produce online, not the polished language of encyclopaedias.

At the level of use, TweetNLP prioritizes practicality and access. Camacho-Collados et al. (2022) describe a plain, high-level Python API that exposes ready-to-run models for common social-media tasks, deliberately avoiding heavyweight deployment requirements (p. 38–39). Instead of chasing ever-larger language models, the project targets models that ordinary users can run on typical laptops or free cloud tiers; the intention is that analysts and researchers can get reliable results without a GPU and without wrestling with complex infrastructure. In this sense, TweetNLP is framed as much as an applied resource – a set of usable, documented tools – as it is a set of underlying model checkpoints, with an emphasis on lowering barriers to entry for social-media analysis (Camacho-Collados et al., 2022, p. 38–40).

Under the hood, the approach is straightforward: start from well-known transformer backbones and adapt them to social media. Rather than training new models from scratch, TweetNLP builds on RoBERTa and XLM-R checkpoints and continues pre-training on large-scale Twitter datasets so that the resulting representations better reflect the platform's style and constraints (Camacho-Collados et al., 2022, p. 39). Camacho-Collados et al. (2022) are clear about their goal, the democratization of social-media-specialized language models, emphasising the importance of keeping model sizes comparatively small and architecturally close to RoBERTa/XLM-R "base" variants, trading a marginal performance ceiling for broad usability (p. 39-40). On top of these adapted backbones, TweetNLP offers task-specific output layers for the kinds of problems researchers routinely face in online text: sentiment analysis, emotion recognition, emoji prediction, irony detection, hate-speech and offensive-language identification, stance detection, topic classification, and named-entity recognition. The result is a coherent suite:

models are specialized enough to cope with the quirks of short, informal messages, but lightweight enough to be practical in ordinary research settings (Camacho-Collados et al., 2022, p. 38–40).

For an information science study concerned with platform discourse, this framing is directly relevant. TweetNLP's premise – that task performance improves when models are adapted to the communicative setting – speaks to the methodological challenge of analyzing user-generated comments at scale (Camacho-Collados et al., 2022, p. 37–40). In other words, the value here lies not only in the individual tasks it supports, but in the design principle it embodies: align tools with the linguistic ecology of the data and keep them usable by researchers who need to run robust analyses on real-world machines (Camacho-Collados et al., 2022, p. 38–39).

3.1.1.3 BERT and RoBERTa

Liu et al. (2019) describe RoBERTa as a refinement of BERT's pretraining method rather than a new architecture. Much of BERT's untapped capacity comes from how it is trained: if one removes the Next-Sentence Prediction (NSP) objective, trains longer on more text with larger batches, packs inputs as continuous "full sentences" up to the maximum length, and uses dynamic masking (the masked tokens are resampled each time a sequence is seen), downstream accuracy improves reliably (p.1-2). In ablation studies, dynamic masking performs at least as well as the static masking used by BERT and is better aligned with longer training schedules. The NSP removal paired with full-sentence packing matches or outperforms NSP setups on standard development benchmarks such as MNLI, SST-2, and SQuAD when other variables are held constant (Liu et al., 2019, p.2-3).

RoBERTa also expands and diversifies the pretraining corpus (e.g., CC-NEWS, OpenWebText, Stories, Wikipedia), showing that increasing data volume and training time yields steady gains with the same masked-LM objective (Liu et al., 2019, p.1-3). The authors further explore a byte-level subword vocabulary (so the model can represent any character sequence without "unknown" tokens), a practical advantage for noisy, user-generated text typical of social platforms, even if the raw performance differences are modest across tasks. Put simply, RoBERTa demonstrates that careful choices about objective configuration and scale – rather than inventing a new architecture – are sufficient to reach or surpass state-of-the-art results on widely used benchmarks, such as GLUE, RACE, and SQuAD (Liu et al., 2019, p. 1-4).

Finally, RoBERTa's contribution is also methodological: by releasing code, models, and clear ablations, it establishes a reproducible training recipe that later studies can adopt or adapt (Liu et al., 2019, p. 1-2). For literature reviews in information science, this positions RoBERTa as the robust baseline for transformer-based text classification in contemporary, heterogeneous corpora – especially where inputs are short, informal, and context-sparse.

Barbieri et al. (2020) writes that modern NLP systems that are commonly ill-equipped to handle noisy text generated by social media users, which is generally fast paced, idiosyncratic, and conversational in nature, especially when combined with platform-specific restrictions such as character limits. RoBERTa-based models as show comprehensive performance quality, outperforming a model that was trained solely on Twitter data on nearly all given tasks. They speculate that this could be ascribed to Twitter not only being noisy text but also contains a fair amount of formal text. The one task that had RoBERTa beat by a Twitter-trained model is irony detection, and it is theorised that this might stem from the fact that expressions of irony on social media might differ from those on standard text, combined with the factor of tweets in the training data being shorter than the ones used for other tasks. Barbieri et al. (2020) state that the efficacy of a RoBERTa-base is most benefited when accompanied by supplementary training on Twitter corpus (p.1647).

3.2 Information resilience and climate anxiety

> *"Our collective resilience against disinformation, misleading information and propaganda will prevent or pre-empt antagonists from influencing decisions, perceptions or behaviour. During peacetime and in a heightened state of alert or war alike, psychological defence shall ensure Sweden's freedom and independence, an open and democratic society with freedom of opinion and free media." (Psychological Defence Agency, 2024, "What is psychological defence?").*

3.2.1 Information literacy

The idea that information literacy can protect against misinformation isn't new. Information literacy as a concept was created in the second half of the 20th century, building upon literacy and springing from the increasing role of libraries and education in shaping how people find, evaluate, and use information. It had begun to become apparent that reading skills weren't enough to give people the ability to critically assess information and information sources. This idea gained even more traction with the internet and the immense expansion to public access to information, and the challenges that this created. Librarians, teachers and researchers began to argue that educating people in information literacy would foster a more informed public and a safer society as a whole (Haider & Sundin, 2019, p.2-3; Haider & Sundin, 2022, p.101-102, 126). This was particularly important in democratic societies, where the legitimacy of governance depends on that citizens are capable of making informed decisions, as well as the citizens' ability to trust public institution – their trustworthiness would be ensured by the ability of the people to hold them accountable. With the emergence of social media platforms and the increasing spread of misinformation, the argument has solidified. A new argument was added to the original one, which was the argument of information literacy as a form of cognitive self-defence. The idea is that if individuals are trained to question the information they encounter, as well as to verify both the claims and the sources on multiple levels, they would be less susceptible to manipulation and deception(Haider & Sundin, 2022, p.101-102, 126).

Today, society is an information system with more information circulating than ever, and with an increasing difficulty when it comes to assessing the information in that system. Jutta Haider and Olof Sundin (2019) argue that search engines and social media platforms have embedded search behaviour into daily life, something they call "search-ification". The algorithms associated with the digital platforms that enable the search-ification also shape users' perceptions of information availability and trustworthiness, with top-ranked results gaining disproportionate credibility regardless of their factual accuracy (p.2-3). Additionally, in a second study by Haider & Sundin (2022), they describe the structure of digital platforms as a framework that creates information bubbles and echo chambers, where people are exposed to one side of discussions, with repeated narratives that reinforce already existing worldviews. Algorithm-driven personalization on social media leads to selective exposure, making individuals less likely to encounter diverse perspectives and more susceptible to highly engaging but misleading information. This leads to an erosion of democratic discourse, fostered by media infrastructures that prioritize virality over truth and therefore creates and solidifies political polarization, distrust, as well as the effectiveness of misinformation campaigns (p.101-102, 126). Linvill et al. (2021) argue that platform algorithms enable the viral spread of disinformation by amplifying emotionally charged content. Their work demonstrates how misinformation can metastasize in algorithmic environments, especially with content related to moral urgency, conspiracy, danger and/or existential threat (p.1-3, 16-19).

3.2.1.1 The problem emerges

In recent years – especially after the 2016 US election, Brexit, and the 2020 US election, along with the January 6th Insurrection at the US capitol – it's become increasingly apparent that there is something that is making information literacy ineffective in combating misinformation on its own (Llewellyn et al, 2019, p.39-41; Hodges, 2023, p.137-140). Hodges (2023) study focuses on the information practices of the followers of QAnon and what these can reveal about the best ways to counteract the dissemination of inaccurate beliefs. He begins by confronting the common belief that people who believe in conspiracy theories are information illiterate. On the contrary, his research showed that followers of QAnon are good at searching for and finding legitimate sources (p.132-133). Instead, the problem lies in the biases and the worldview which governs how the followers interpret the information they are assessing. Instead of focusing on the important points of the information, they would focus on details – idiosyncrasies – that would be mentioned in the information, but often not even part of the focus of the information. These would be things like influential people – such as former president Obama, Hillary Clinton, and other public figures – being mentioned attending events and then tying this to a far-fetched narrative or conspiracy, which could only be achieved if the person digesting the information already had a perceptual filter that would give this information meaning (Hodges, 2023, p.137-140). Hodges (2023) argues that more research is needed regarding this type of information behaviour, and that this starts by recognizing that each conspiracy theory's believability originates from a grain of truth that is enveloped layers of fictions and misinformation, forming a false narrative. He argues that researching this further will help researchers and professionals explore and find more effective interventions in stopping the spread of and belief in misinformation. That it can help find information that might have a positive impact on people's epistemological filters, something that he considers essential in order to help fight misinformation (Hodges, 2023, p.139-141).

3.2.2 Information resilience

> *"resilience (noun)*
>
> *re·sil·ience ri-ˈzil-yən(t)s*
>
> *1: the capability of a strained body to recover its size and shape after deformation caused especially by compressive stress*
>
> *2: an ability to recover from or adjust easily to misfortune or change" (Merriam-Webster, nd).*

As described by Andreas Vårheim (2016), resilience theory has its origin in physics, describing the ability of a substance to endure outside forces and return to its former state (p.1-3). In psychology it's been viewed as the psychological adaptation in the face of adversity, trauma, or significant stress. In systems theory, resilience has been described as the capacity of a system to absorb disturbance and reorganize while undergoing change, retaining essentially the same function, structure, and feedback. In the context of cybersecurity, resilience pertains to the ability of cyber systems to prepare for, respond to, and recover from cyberattacks, ensuring the continuity of operations (Munro,2013, par.2-4). From these descriptions, it becomes evident that information resilience is an emerging interdisciplinary concept that addresses the capacity of individuals, system, and society to withstand, adapt to, and recover from disruptions in the information environment.

Ahlryd & Hanell (2024) define information resilience as "a process where adaptive information capacities are combined and utilized to enable functioning and adaptation after a disturbance" (p.268). Information resilience, as developed in this study, hinges on three core attributes: robustness, redundancy, and rapidity. These attributes describe how information systems or individuals withstand, adapt to, and recover from stress. Robustness refers to the durability of a resource or process, redundancy to the availability of alternative pathways, and rapidity to the speed at which information can be accessed and utilized. In the case of misinformation, these attributes also translate into the system's ability to resist manipulation: redundant access points prevent monopolization of information channels, and robustness supports resistance to discursive erosion caused by conspiracy theories or institutional failure (Ahlryd & Hanell, 2024, p.268-271).

3.2.3 Perceptions and climate change

Research studying the differences in perceptions of climate change and their causes go back two decades. Howe, Mildenberger, Marlon, and Leiserowitz (2015) demonstrate that public attitudes towards climate change diverge widely. In their study, which focuses on the US, some areas show strong belief and concern while others remain sceptical or disengaged – regardless of the physical location and how close these areas were to each other. Their study underlines that climate change opinions are not monolithic but shaped by surrounding political, cultural, and social environments (p. 596-603).

Akerlof et al. (2010) extend this understanding by showing how perceptions of climate change risk are filtered through social and cultural frames in different countries. Their comparative surveys of the United States, Canada, and Malta reveal significant differences in whether climate change is seen as an abstract environmental problem or as an immediate threat to human health. These results highlight that climate change is not only an environmental or scientific issue but also a matter of how risks are interpreted within specific social contexts (p.2559-2606).

Wang, Geng, and Rodríguez-Casallas (2021) take a psychological perspective to examine how perceptions of climate change risk translate into inaction. They show that higher levels of risk perception are associated with lower climate change inaction, when seen through the perceptual filters of climate change belief and environmental efficacy. In other words, perceiving risk strengthens belief in climate change, which in turn fosters a sense of efficacy, ultimately reducing inaction. They further demonstrate that mindfulness moderates this pathway, with individuals high in mindfulness more likely to convert risk perception into belief, efficacy, and action. This study is significant because it shows that climate change doesn't influence people only as an isolated issue but works through layered psychological mechanisms that shape how they respond to broader risks and responsibilities. For this thesis, its relevance lies in the demonstration that climate change impacts people indirectly, filtering into their attitudes toward information and action through underlying beliefs, even when not consciously acknowledged (Wang, Geng, & Rodríguez-Casallas, 2021, p.1–2).

3.2.4 Climate-anxiety

Climate change has been used as a tool for measurement to investigate information behaviour and information practices within information science. However, the start of the intersection between climate studies and information studies focused on the credibility and type of sources, on documenting and cataloguing the research area, and on the impact of climate change on information institutions – such as libraries – and the work they do (Vaughan, 2011; Line, 2006, p.iii-iv; Zhang et al., 2018, p.468-469). Previous research has also focused on the information structures related to climate data, on how information science could contribute to and benefit climate research, on moving research toward open data and open access research, digital libraries, interoperable metadata, and reuse of data using climate change as case studies. The studies also focused on the application of information science theories to climate change data and studies, and eventually the information research related to climate change began to focus on information behaviour and information practices (Zhang et al., 2018, p.468-469).

Climate-related information behaviour and information practices studies

It has becoming increasingly clear in climate research that climate change effects people's mental health and that their mental health, in turn, effects their view of climate change. Myers (2014) explains that people in society don't have what they need to confront climate change. He argues that the implications of climate change are received and internalized by people as an existential threat, which collides with what Myers (2014) refers to as their "lifeworld". In Merriam-Webster (n.d.), lifeworld is defined as "the sum total of physical surroundings and everyday experiences that make up an individual's world." It is the worldview of a person that is based on a person's experience of the world, rather than facts or scientific evidence. When the lifeworld collides with an existential threat, the person is confronted with the possibility of the entire lifeworld – their worldview and perception of the world they live in – collapsing. This threat of collapse, the threat of their worldview being wrong, triggers anxieties that people are deeply motivated to avoid. This is, according to Myers (2014), where denial, including climate denial, comes from. The denial is used as a tool to protect the integrity of the lifeworld for the sake of preserving the person's sense of security and shield them from the stresses and anxieties which the person fears would overwhelm and debilitate them (p.53-54). Myers (2014 also shows how important it is to see the connection between worldview and people's ability to act in situations where it's needed. If an issue confronts the worldview of a person, and the person doesn't have any other tools, it leads to a cognitive distancing of the ethical implications – which would normally make the person act to resolve whatever problem is causing the issue – which can become a collective action of working together to deny that the problem even exists (p.54). This makes people more susceptible to misinformation, as long as it serves to reinforce (and protect) their worldview and thus their sense of security, making information literacy ineffective. This means that the 'information deficit model', widely used for many years within information science, falls short of combating all forms of negative information. Myers (2014) argues that if information science presumes that it's a lack of knowledge of information literacy and access to scientific, credible information that is the issue, but the issue is actually how susceptible people are to misinformation – even more importantly that people will seek to avoid correct information because of these psychological blockages – this would mean that information science is fighting the wrong issue entirely, and that the methods put in place will not work. He argues that the information science field and society as a whole have to start looking at the relationship between people and information phenomenological approach in order to solve these issues (p.57-60,66).

In their study from 2023, Choo (2023) further adds to the social aspects of information behaviour and practices. They conducted a study focusing on climate change information seeking, basing their study on information behaviour and risk information seeking literature. Their model was based on the hypothesis one of the general theories within information behaviour that information seeking is "influenced by the individual's cognitive state, affective responses, and social context” (Choo, 2023, p.1087). The study used a number of cognitive, affective and social/situational variables to attempt to analyse what makes a person seek information. Their study found that social norms, affective response, and social trust were the most important variables to predict information seeking behaviour, while “knowledge need” was of small consequence. In addition to this, their study had two significant findings: the first finding was that in climate change information seeking behaviour, the social norms and expectations of the relevant people in a person’s social environment are of major consequence. As for the second finding, a person’s emotional response to the risk of climate change is “much more important” than the person’s perception about how much they need the information (Choo, 2023, p.1097-1098).

Adding to the understanding of climate change beliefs, a study by Emily A. Hurley and Micheal S. Molloy (2025) examines how individuals shift their beliefs about and process of changing their opinions about climate change. The findings reveal that before opinion change, participants often held beliefs which were grounded in ideological identity and sometimes reinforced by mistrust of the sources of information which opposed these identity-beliefs (p.6–7). The change process the authors describe begin with a distancing from ideological community, actively seeking out information, and experiencing gradual or sudden realizations that created and/or solidified new beliefs (p.8). Importantly, distancing that the participants described was not intentional, which the authors describe as follows

> “This distancing was not an active decision by participants to expose themselves to climate change information, but an experience they found themselves in that gave them perspective to newly question ideologies that they had come to internalize.”
> (Hurley and Molloy, 2025, p.8)

It could be caused by travel, moving or other obvious changes in the person’s life, but the experience could also be caused by less tangible things such as pursuing a new interest or having a conversation with a person within an already established social network. To the contrary of the first phase of the process, the seeking phase that follows this newfound insight is, as described by Hurley and Molloy (2025), is often intentional and involves reassessing previously trusted media and moving toward more scientific or direct evidence. Furthermore, the study shows that not all the change perspectives led to the same conclusions or views about what should be done next, but the study still shows that identifying the right information and the process of changing beliefs can be leveraged when communicating important information to the public (p.8-9). Hurley and Molloy (2025) also advocate for more quantitative studies into the subject of how people’s opinions change regarding issues like climate change, looking into “bridging ties” and using distancing experiences to constructively reconstruct inaccurate or problematic worldviews. They argue that this is especially important because of yet another finding in the study – that most of the participants only changed their opinions once they had searched for more information themselves, and that this occurred as an active need, not a passive action. This provided further context to how motivated reasoning and cognitive dissonance arise in the reconstruction of worldviews in relation to information behaviour in a way that is highly relevant for the context of information resilience. Hurley and Molloy (2025) argue the importance that facilitating the connection between the population and the actual information, so that they themselves can research the area in question and transforming their approach into quantitative studies (p.14-16). This is very important both when it comes to how to approach populations during times of high susceptibility to misinformation, what they need, as well as for assessing what role information specialists – such as librarians – could or should have during times of crisis.

4 Method

This is a mixed-method design that combines large-scale quantitative sentiment analysis on YouTube comments with a qualitative, documentary compilation of CCAM toplines for contextual triangulation. The chapter first clarifies design choices and data sources, then documents the experiment execution (from scraper to modelling and calibration), and finally details how CCAM was compiled and aligned year-by-year.

4.1 Research design

The study is a mixed-method study which employs quantitative supervised machine learning, using RoBERTa, to produce yearly indicators for sentiment, irony, trust/distrust, and hostility, as well as the misinformation indicators concept, belief, and accusation from YouTube comments (2010–2024). Model selection was conducted as a computational experiment (baselines → off-the-shelf transformers → fine-tuned model), which was evaluated on a held-out test split. A hybrid inference rule sources irony and sentiment from an irony-initialized encoder and other labels from a base encoder.

4.1.1 Data and data sources

YouTube was selected as the primary platform due to its significant role in shaping public discourse, and its accessibility through the YouTube Data API. The platform has increasingly become a site for both information dissemination and misinformation spread, making it an ideal environment for the study.

YouTube comments were selected as the data source because of the access they give to public opinion, especially for a quantitative data collection, spanning 15 years. Comments on videos can serve as timestamped expressions of audience reaction and opinion, capturing a snapshot of sentiment in time. Access via the YouTube Data API enabled automated retrieval of structured metadata – such as video identifiers, titles, publication dates, and comment threads – ensuring consistency across the fifteen-year timespan. Moreover, leveraging the API's filtering parameters (e.g., publication date ranges, ordering by view count or date) allows precise alignment of collected comments with specific temporal and thematic criteria.

4.1.2 Data collection

Data source, query and temporal frame

To observe high-visibility public discourse relevant to information resilience, the general search query “news” was implemented to discover videos, without topical filtering. This “news-general” strategy aligned with the aim of analysing general perceptual filters in a way that can give indications of an overarching worldview instead of the views of a specific topic. Including comments from videos about multiple topics (politics, economics, crises, science) was done to reduce topical siloing.

To enable year-to-year comparisons, a fixed yearly quota was implemented for each calendar year (2010–2024), targeting an equal-sized sample of comments. Both video search and comment retrieval were constrained to year-based time windows. This enabled the scraper to reach the desired sample-size, as well as making the sample more evenly spread across each year.

Only public comments were collected and applied no platform-side filters beyond the search query and year windows. Language, duplication, and text quality were handled in preprocessing. Usernames were converted into unique identifiers by the script to ensure that no personal identifiers were collected.

.

Yearly sampling and quota

To make year-to-year comparisons meaningful, data collection followed a fixed yearly quota design: for each calendar year from 2010 to 2024, an equal-sized sample of comments was targeted. Video search and comment retrieval were performed within year-specific time windows so that exposure is balanced across the period. The fixed-quota approach avoids overweighting recent years and supports the analyses in the Results chapter.

Units of analysis

The unit of analysis was the individual comment as returned by the API (top-level comments by default). For consistency, the comment text, timestamp, comment_id, video_id, and minimal structural metadata required for auditing were retained. The study uses both comments and replies in the dataset in order to achieve the desired sample size.

Hand-off to preprocessing

The YouTube comment scraper produced per-year NDJSON files, writing one comment per row. Before modelling, all the years were merged, schemas were validated, and rows without alphanumeric content were removed. The comments were also filtered to English via fastText LID (with threshold and fallback) and applied de-duplication by comment_id across years. Earlier exploratory runs applied only within-year de-duplication, which risked allowing cross-year duplicates to persist. In the final pipeline, global de-duplication after concatenation ensures that the analytic corpus reflects unique English comments for the entire period. The final model-input file comprises 205,792 comments for sentiment analysis and is the basis for all results reported in the thesis.

Reproducibility safeguards

To support replication, the collection process logs API parameters, request counts, and year-windows; random seeds are fixed where relevant (e.g., when sub-sampling). All scripts can be found in the GitHub repository linked to in the appendices.

4.1.3 Contextual triangulation (CCAM benchmark)

This is combined with a CCAM is used as a contextual benchmark to assess directional alignment of worldview with platform indicators rather than numerical agreement. “Positive/accepting” survey data (higher acceptance, human causation, scientist consensus, certainty among accepters) are read against “positive/accepting” comment data (higher trust, higher positive sentiment, lower hostility/distrust/accusation/irony). “Concerned/negative” survey data (worry, harm, harmed-now) are read against “concerned/negative” platform data (higher negative sentiment, hostility, distrust, accusation, irony). Alignment is evaluated qualitatively at the sign level (co-movement vs divergence) using the latest CCAM values for each year and per-year proportions from the YouTube corpus. No cross-dataset modelling is performed.

4.1.4 Ethics

The ethical considerations of this study concern both the nature of the data and the standards of validity and reliability required to ensure that results are trustworthy, transparent, and responsibly interpreted.

Use of Publicly Available Data

All data used in this thesis consist of publicly available YouTube comments, retrieved via the YouTube Data API. No attempts were made to identify individual users, and no personal identifiers were collected. This was ensured by allowing the script to convert usernames to unique identifiers, without collecting the usernames The study therefore complies with both GDPR principles and disciplinary standards in library and information science for the ethical use of digital trace data.

Transparency, Validity, and Reliability

A key ethical responsibility in research is to ensure that results genuinely address the questions that are being explored and are not distorted through methodological issues or biases. This relates directly to the concepts of validity and reliability (Trost, 2012, p. 61–64). Validity refers to the degree to which a study measures what it intends to measure. In this thesis, validity was supported by the careful operationalisation of indicators – trust, distrust, hostility, irony, sentiment – so that they connect meaningfully to the theoretical framework of attunements and information resilience. Validity was further ensured by grounding the annotation schema in multiple domains of expertise: information science, psychology, climate information behaviour studies and communication studies. This interdisciplinary grounding strengthened the credibility of the measures and ensured that the operationalisations reflected the multifaceted nature of misinformation and resilience.

Reliability concerns the stability and consistency of measurement across time and context (Trost, 2012, p. 58–63). In this study, reliability was enhanced by standardising the preprocessing pipeline and modelling procedures, ensuring that identical data would yield identical outcomes if re-analysed. Standardisation also applies to data collection: the scraper was executed with the same inclusion and exclusion criteria across all years, and all preprocessing scripts were structured to treat each year's data uniformly. This prevents accidental biases introduced by inconsistent handling (Trost, 2012, p. 79–80).

Reliability can also be challenged by the subjectivity of interpretation, a risk that Kvale and Brinkmann (2014, p. 97–106) emphasise in qualitative research. Although this study employs quantitative NLP methods, subjective judgement was unavoidable in defining the annotation schema and in resolving ambiguous cases of irony, sarcasm, or hostility. To mitigate this, an explicit annotation guide was created, with clear definitions and examples, and manual labels were applied consistently according to this guide. When uncertainty remained, the comment would be labelled neutral as a way of making sure that it didn't effect either of the polarity results. This level of reflexivity ensures that the influence of researcher subjectivity is visible and can be accounted for when interpreting the results. The annotation guide can be found in appendices.

Balancing Research Value and Ethical Risks

An additional ethical concern lies in balancing the societal importance of studying misinformation and climate change discourse with the risks of amplifying harmful narratives. To address this, the study refrains from reproducing or disseminating individual misinformation claims in isolation. Instead, analysis focuses on aggregated patterns and structural indicators of trust, distrust, irony, and hostility. By doing so, the study avoids amplifying disinformation while still making its dynamics visible.

4.2 Experiment execution

This section documents how the YouTube comment corpus was assembled, cleaned, and modelled. Search and retrieval used a single, general query ("news") to surface a broad stream of public discourse, and comments were constrained to calendar-year windows (2010–2024) so that each year stands on its own. Only public comments were retrieved via the YouTube Data API. Platform-side filters beyond the search query and year-specific windows were avoided, and choices about language, duplication, and text quality were deferred to preprocessing so they could be applied uniformly. The unit of analysis is the individual comment (top-level by default, replies included as needed), retaining only the minimal metadata required, as well as the text – usernames/handles were never stored. The scraper produced one NDJSON file per year, which were later merged to a single table for cleaning and modelling. After removing non-textual rows, filtering to English via fastText LID with langdetect fallback, and performing global de-duplication across years, the analysis file contained 205,792 unique English comments, which is the dataset used for all modelling and results. Each year was sampled to a comparable size. This reduces bias from later years having more comments due to platform growth and construct, so the trends reflect temporal change.

4.2.1 Scaling and Iterative Expansion

To reach the target-sample of 150,000 comments, the pipeline was scaled in stages. Initial test runs covered three years with just 5 videos and 10 comments per year. Subsequent iterations expanded the sample size: first to 50 videos and 100 comments per year, then further to 400 videos and 1,000 comments per year across all fifteen years.

A large-scale pilot collected 150,000 comments (10,000 per year, 2010–2024), confirming the pipeline's capacity before expanding to the full dataset.

4.2.2 Development of Scraper and Preprocessing Pipeline

The creation of the final YouTube comment-scraper was the result of multiple iterations developed to ensure validity, completeness, and ethical reliability in line with the study's design. The primary aim was to collect a dataset of YouTube comments from the years 2010 through 2024, excluding music videos, while filtering out spam, non-English text, and irrelevant material. Early prototypes of the scraper were tested for basic functionality using the YouTube Data API v3. However, these initial versions suffered from limitations such as rate-limiting issues, missing language detection, and inadequate filtering mechanisms. Iterative refinements were implemented in both logic and infrastructure. This included: rotating API keys to avoid request limits; limiting comments per video to prevent a skewed sample dominated by a few viral uploads; and normalizing all text inputs for language filtering using langdetect. Non-English comments were removed, as were those containing no alphanumeric characters, to ensure semantic relevance and prevent inaccurate skewing in downstream NLP pipelines.

The scraper was constructed using Python and outputs files in .ndjson format, with one file per year. Each comment record is tagged with metadata including video ID, comment ID, publication date, and the associated query. To ensure reproducibility, all processing steps – including deduplication, language detection, and per-year caps – are logged directly in the terminal and verified manually at each run. To further structure the data and prepare it for modelling, code for preprocessing was developed to convert .ndjson comment files into consolidated and analysis-ready CSV datasets. This code performed a range of tasks: grouping comments by year, formatting and validating text fields, filtering non-English content, and creating both full and subsampled datasets for diagnostic purposes. Importantly, these preprocessing steps integrate directly with the larger sentiment analysis pipeline, minimizing the risk of data leakage and reducing the need for redundant processing.

The NLP pipeline itself was designed to operate modularly, with separate but harmonized components for lexicon-based scoring (VADER and TextBlob), trust and distrust keyword detection, and transformer-based classification using TweetNLP. This includes sarcasm detection, and sentiment scoring. In the beginning, emotion or topic scoring were also included, but upon inspection of the output, these metrics were excluded due to a clear incompatibility with the dataset, resulting in the same label on all rows. Outputs from each model are saved to file for inspection and used to calculate per-year summary statistics. Where relevant, spot-check files (e.g., for flagged trust/distrust comments or high-confidence irony classifications) are generated to support manual auditing. All ethical safeguards were embedded upstream, such that downstream models would operate only on cleaned, relevant, and valid English-language user comments sampled according to clear inclusion criteria.

4.2.3 Experimental Pipeline: From Lexicon Baselines to a Fine-Tuned, Hybrid RoBERTa System

In the end, the experimental pipeline developed for the study to progress in three stages: (1) rapid lexicon-based sentiment baselines to establish an initial reference point, (2) off-the-shelf transformer models trained on social-media text to improve treatment of figurative language, and (3) a fine-tuned RoBERTa architecture tailored to the study's multi-label taxonomy with a dedicated strategy for irony. Each stage was retained in the record to preserve an auditable trajectory from quick diagnostics to the final, deployment-grade models.

Stage 1: Lexicon baselines (VADER and TextBlob)

We first ran lexicon baselines – VADER and TextBlob – because they are fast, widely used, and easy to understand. A lexicon method assigns positive/negative/neutral scores by looking up words in a dictionary with fixed weights. These baselines provide immediate sentiment references and offer a lower bound on performance. However, they expose a principal weakness: irony is often misread. For example, a sentence like "Great, just what we needed – another delay" is scored as positive by a lexicon because of "great," even though a human reader hears the opposite. This distorts extremes and can inflate positivity or negativity in ways that do not reflect how people actually mean their words. For this reason, the lexicon outputs were kept as sanity checks and calibration aids, but not as the final measurement used in the results. They remain useful as a transparent baseline and as a reference point when later models disagree.

Stage 2: Off-the-shelf social-media transformers (TweetNLP/TweetEval)

To address the limitations observed with lexicons, the pipeline next incorporated pretrained TweetNLP/TweetEval checkpoints for sentiment and irony. These models read the entire sentence in context rather than counting individual words, which helps with short, figurative expressions found in YouTube comments. In practice, this step improved recognition of informal phrasing, slang, emojis, and some irony-like cues (e.g., "yeah, right"). Importantly, it established a condition for sentiment-labelling, which was to label sentiment overall and conditional on irony. That is, sentiment is calculated for the full set of comments and also for the subset that a model marks as ironic. This ensures that the model doesn't interpret ironic comments as literal. Off-the-shelf transformers are stronger than lexicons, but they are not yet aligned to this study's specific labels and decision rules, which means that they are treated as an intermediate step that aims to reduce obvious errors.

Stage 3: Fine-tuned RoBERTa with multi-label output layer and a parallel sentiment output layer

The final modelling stage implemented a RoBERTa-based system designed around the study's inventory of constructs. The architecture comprised a shared encoder with two output layers: one output layer produces an independent probability for each binary indicator: irony, trust, distrust, hostility, concept, belief, accusation. A second output layer produces a three-way probability distribution for sentiment (negative, neutral, positive). The binary probabilities are thresholded (using the value selected on validation) to assign 0/1 labels; the sentiment label is labelled using the highest probability. All probabilities are show in the output. Inputs were tokenized with truncation and padding enabled at max_length = 128, meaning that since the model only analysed a set amount of characters, longer comments would be batched into separate sections, analysed section by section, and then the highest matching probability for each section was chosen as the final label. This configuration ensured stable batching during both training and inference.

The model training was done in a development environment, in this case the environment called tf210. The training loop was implemented directly in PyTorch to avoid upgrade conflicts in the tf210 environment. Training, validation, and test sizes in the curated labelled subset were 1,047, 225, and 225 rows, respectively. Per-label decision thresholds for the multi-label output layer were selected on the validation set from precision–recall trade-offs and then held fixed for the test set.

On the held-out test split, the multi-label classifier achieved micro-F1 = 0.305 and macro-F1 = 0.146, reflecting substantive class imbalance and scarce support for several rare labels. The sentiment output layer reached macro-F1 = 0.735 overall and 0.423 on the irony-positive subset. The gap between overall and irony-subset sentiment confirmed the necessity of handling irony explicitly in deployment.

Hybridization for Irony and Sentiment at Inference

Given the centrality of irony to correct polarity interpretation, inference in the full run used a hybrid rule. All irony decisions and sentiment outputs were sourced from a RoBERTa encoder initialized from an irony-focused checkpoint (cardiffnlp/twitter-roberta-base-irony), while all remaining labels were produced by a base RoBERTa encoder (cardiffnlp/twitter-roberta-base). This division exploited the benefits of irony-specialized pretraining where it matters most, while retaining a strong general encoder for the rest of the taxonomy. The hybridization occurred purely at inference: training data and loss geometry remained unchanged, and outputs were merged deterministically by comment identifier.

Input Corpus and Preprocessing for Deployment

Deployment operated on the merged output of the YouTube scraper. The scraper discovered candidate videos using the single query q='news' in the YouTube Data API, sorted by year; all comments and replies for discovered videos were retrieved via paginated endpoints and written to per-year NDJSON files with unmodified API fields. These NDJSON files were concatenated into a single archive of 485,504 comment-rows, preserving all original columns. The model-input dataset was then produced through a uniform, auditable sequence: removal of non-alphanumeric entries (11,688 rows), language identification with fastText LID and a langdetect fallback yielding an English-only subset (62,232 removed), and deduplication by comment_id (eliminating 205,792 duplicates). The resulting analytic file contained 205,792 unique English comments. Video-level context (title and publication year) was retained for each row.

4.2.4 Dual Inference Passes and Hybrid Assembly

Two complete inference passes were executed over the output file: one with the base model and one with the irony-focused model. Each pass produced probability columns and corresponding binary decisions at the development thresholds. A consistency check verified row alignment and the presence of required columns. The hybrid file was then assembled by taking irony and all sentiment columns from the irony-focused pass and all remaining labels from the base pass. This unified CSV served as the substrate for calibration, summarization, and reporting.

4.2.5 Calibration of Irony at Corpus Scale

Direct transfer of the validation-tuned irony threshold to the full corpus produced a disproportionally high prevalence, indicating threshold drift under domain shift. To solve this issue, re-thresholding on saved probabilities was applied. The operational decision threshold for irony was chosen to reproduce the test-set prevalence of 15.6%, which yielded $\tau^* = 0.254$ and a stable full-run irony rate of 0.156. All probability columns were retained so that alternative thresholds can be evaluated without re-running inference.

4.2.6 Full-run Estimation and Temporal Aggregation

The calibrated hybrid file was summarized as per-year proportions for all binary labels and as mean probabilities for the three sentiment classes, both overall and conditional on irony. Overall prevalence rates in the full corpus were 0.377 for distrust, 0.131 for trust, 0.103 for hostility, and accuse_misinfo registered 0.003, while spam_flag, concept_misinfo, believe_misinfo, and cogni_dis_manual were near zero at the chosen thresholds. Mean sentiment probabilities across the corpus were 0.519 negative, 0.239 neutral, and 0.242 positive. Reporting at the proportion level ensured that calendar years with larger raw volumes did not dominate trend interpretation, since the underlying scraper did not enforce quota balancing in the full run.

Year-specific associations are summarised using Pearson's correlation coefficient (r), with two-sided p-values and 95% confidence intervals computed via Fisher's z-transformation ($n = 15$ years); for interpretation, $|r| \geq 0.50$ is treated as substantively meaningful, while exact estimates are always reported. Pearson's r shows how the two yearly series move together on a straight line scale from −1 to +1. Values near +1 mean the two rise and fall together; values near −1 mean they move in opposite directions; values near 0 mean there is no clear linear link. Here it is calculated on the paired yearly values (2010–2024; $n = 15$). A standard way to add a margin of error to r. The value of r is first converted to a scale where the usual confidence-interval math works well, the interval is computed, and the result is converted back to r. With 15 years, this gives a sensible 95% confidence interval around r.

4.2.7 Outputs, Integrity Checks, and Sensitivity Analysis

All summaries were exported to an analysis workbook (fullrun_results.xlsx) together with intermediate CSVs and configuration files. A compact integrity script confirmed one-to-one row correspondence between the model-input and hybrid outputs and verified that required columns were present prior to summarization. Sensitivity analysis focused on decision-rule robustness: small perturbations around the calibrated irony threshold (±0.02) did not alter findings, and relaxing the language filter produced similar patterns for high-prevalence labels. Probability-level summaries were reported for rare labels to avoid over-interpretation of threshold artifacts.

4.2.8 Robustness and Sensitivity Checks

Two simple robustness checks strengthen the interpretability of the full-run results:

Threshold sensitivity

Because binary prevalence depends on decision thresholds, modest shifts were verified around the calibrated irony threshold (±0.02) do not qualitatively change the main conclusions. For rare labels with near-zero prevalence, probability-level summaries (mean <label>_ft_prob) are reported alongside binary rates to avoid overinterpreting threshold artifacts.

Preprocessing audit

The preprocessing script emits counts for each filter stage; repeating the pipeline with --lang-only any (no language filtering) yields similar qualitative patterns for high-prevalence labels, indicating that the English-only constraint is not solely driving the results.

Verifying outputs

Every run of the script was followed by an audit of both the output in Visual Studio Code, as well as an audit of the script outputs, verifying that the results sought after were achieved, and that the script output matched the output in the original NJSON-files.

Challenges and Adjustments

Several issues arose during the experiment, which required adaption of the script. In the beginning, analysis of the outputs showed the inadvertent inclusion of irrelevant videos (e.g., music content), resolved by keyword-based exclusion. Language filtering required refinement when early iterations allowed some non-English comments to pass through. This was resolved by implementing more advanced language detection algorithms. Temporal misalignment initially resulted in comments from different years being included, which was corrected through stricter timestamp cross-validation. In addition to this, the years 2010 and 2011 couldn't reach 10.000 comment, so a separate script with more queries was made for the years, successfully collecting the required sample. It is also safe to presume that MAX_COMMENTS didn't work and must've accidentally been excluded from the main part of the script, since the scraper collected a sample that was substantially larger than the intended 150 000 comments. This was mitigated by making sure that the output included results using a mean per year, to prevent the results from becoming slanted by the amount of comments by the study in itself.

API quota limitations initially disrupted data collection frequently. The introduction of API key rotation and exponential backoff protocols significantly mitigated this issue, ensuring continuous data collection. Adding tiers to the scraping quotas and adding the amount of pages that the scraper could run through resolved most of these issues. However, the scraper script and the final sentiment script took 16 hours and 20 hours respectively to finish their runs, likely in part because of these issues.

After the entire collection had run, due to a misunderstanding of how to use the API keys, quota warnings were received from Google. Due to communication issues, as well as time limitations, the script was unable to run on an extended quota, which is what should've been used. If the script is ever reused, and this is applied to the study, this will also remove the need for multiple API keys.

The final issue, which remains unresolved, was an error in the final script for the RoBERTa models. During the creation and testing of the first versions of the VADER/TextBlob/TweetNLP, it was discovered that while comment_IDs remained unique within each year, that was not always the case within the entire sample. Due to the entire sample already being collected, and the inability to run the script again – both due to time constraints and the before-mentioned alerts from Google about the use of the API keys – it was not feasible to give all the comments unique IDS now, due to the potential of giving duplicates different IDs which would have the opposite effect of what was intended. Instead, it was decided that the script should analyse both the comment_ID and the text columns belonging to each comment before labelling it as a duplicate and removing it. This was, it would prevent comments with the same text but possibly different commenters, and the same comment from the same commenter made at different times or on different videos, from being labelled as duplicates. In the last sentiment analysis script, this was accidentally limited to the comment_id column, likely resulting in the significant addition of "duplicates" - more than 100 000 - that were removed in the full-run.

4.3 Compiling the results from the benchmark study

This section documents how the *Climate Change in the American Mind* (CCAM) corpus was compiled and structured for use as a population-level benchmark. The goal is to create a reproducible pipeline from public CCAM releases to yearly indicators that can be aligned with platform-level discursive measures from the YouTube comment corpus. The description covers data source and scope, variable selection and collapsing rules, harmonization across years, year-to-year alignment with the platform series, and the statistical procedures and diagnostics used in the Results chapter.

Citation note: in order to avoid superfluous citations and citations that risk being more confusing than helpful, it was decided that any reference to CCAM and a specific year in the thesis specifically refers to the report of that specific year and the results within that report. For example, the text "In 2010, the CCAM results show(…)" refers to the results within the report of 2010 found in the list of the CCAM reports used in the reference list. This was done because the repetitiveness of years and "CCAM" became repetitive and confusing instead of helpful when citing the reports. When referring to all the years, it will be cited as follows: (CCAM, 2010-2024).

4.3.1 Data source and scope

CCAM is a recurring, nationally representative U.S. survey program that tracks public orientations toward climate change across multiple domains, including belief that global warming is happening, worry about climate impacts, perceived personal harm, perceived scientific consensus, subjective certainty, and issue importance (CCAM, 2010-2024). For this thesis, the benchmark corpus is drawn from the compiled workbook prepared for analysis (CCYM.xlsx), which records the published, year-coded percentages for each indicator between 2010 and 2024. The compilation operates at the level of yearly proportions as reported in CCAM public materials; no attempt is made to re-estimate values from microdata. This approach ensures that the benchmark reflects the official toplines while avoiding discrepancies due to alternative weighting or post-stratification choices. Because CCAM is an independent, probability-based survey program, its estimates serve as a population reference point rather than as a direct mirror of the platform's commenter population (CCAM, 2010-2024).

4.3.2 Variable selection and collapsing rules

Six CCAM constructs were selected to cover the principal facets of public orientation relevant to the comment indicators: acceptance that global warming is happening, worry about climate change, perceived harm to self, subjective importance of the issue, recognition of scientific consensus, and certainty that global warming is happening. Each construct was collapsed into an interpretable "high" category to support alignment with platform indicators and to limit year-to-year noise from fine-grained response options.
Acceptance is defined as the proportion answering "Yes" to the question "Is global warming happening?". Worried is defined as the sum of the "Very worried" and "Somewhat worried" response categories. Harm (to you) is defined as the sum of "A great deal" and "A moderate amount" of harm to "you," capturing perceived personal exposure rather than impersonal or national harm. Importance (high) is defined as the sum of "Extremely important" and "Very important" to the respondent. Consensus (true) records the share indicating that there is a scientific consensus among scientists about global warming. Certainty (high) is defined as the sum of "Extremely sure" and

“Very sure” that global warming is happening; only positive certainty is used in order to align the benchmark with acceptance rather than with disbelief. The rationale for these collapses is twofold. First, combining adjacent “high” categories improves comparability over time when response distributions shift modestly but the substantive orientation does not. Second, the collapses increase stability for comparison with yearly platform indicators by reducing variance associated with small categories. In sensitivity diagnostics, stricter variants were also evaluated – such as defining worry with “Very worried” only or defining importance with “Extremely important” only – to test whether results depend on collapse breadth; as reported in Results, these stricter versions attenuate correlation magnitudes but do not change signs.

4.3.3 Harmonization and quality checks

The CCAM workbook was harmonized to a common schema with one row per calendar year and one column per collapsed indicator. The harmonization standardized year fields, verified that sheet names and variable labels matched their intended constructs, and checked that all values represented proportions in percentage points. Where CCAM offered multiple items tapping similar constructs (for example, harm framed at different social distances), the “you” framing was chosen for Harm because it most directly parallels the personal stance and tone visible in comments. No interpolation or smoothing was applied; years with missing items for a given construct were left blank and excluded from pairwise computations that required that construct. Rounding for reporting in text was set to one decimal place to match the native precision of the compiled workbook while keeping figures directly traceable to the underlying values.

4.3.4 Alignment with the YouTube comments

Alignment is performed at the year level. For each calendar year between 2010 and 2024, the CCAM row is matched to the corresponding year’s prevalence estimates from the YouTube comment corpus. The platform corpus reports proportions of comments by year for sentiment classes and for the discursive indicators (hostility, distrust, trust, irony, and three misinformation-related indicators). Because both corpuses are year-specific, the alignment does not require time shifting or lag adjustments. For correlation analyses, only years where both the CCAM indicator and the platform indicator are available are used. No reweighting is applied to adjust for differences in the number of comment observations per year; instead, uncertainty handling is addressed in the diagnostic procedures.

4.3.5 Statistical procedures

Two kinds of quantities are computed from the aligned corpuses. The first are descriptive year-by-year proportions, reported in the Results chapter as numerically anchored trajectories for both CCAM and the platform indicators. The second are linear associations computed as Pearson's correlation coefficients on matched yearly values across 2010–2024 (or the maximal overlapping subset for a given pairing). Correlations are accompanied, where appropriate, by Fisher z–based confidence intervals to communicate uncertainty about association magnitudes. For proportions, uncertainty is handled with binomial or Wilson intervals at the yearly level; these intervals are used in tables and appendices to support the prose summaries. Because the CCAM values are taken from published toplines, no additional survey-weight re-estimation is applied, and the uncertainty associated with CCAM estimates is treated as given by the survey program's own methodology.

4.3.6 Diagnostics and sensitivity analyses

Diagnostics were pre-specified to test the stability of observed associations. First, event-year sensitivity was assessed by repeating correlation calculations with the year 2022 excluded, as several platform indicators exhibit their peaks in that year. The expectation was that exclusion would reduce peak magnitudes but preserve directions if the associations are not driven solely by a single year. Second, collapse sensitivity was assessed by substituting stricter "high" definitions for worried and importance. Third, an interaction diagnostic was conducted on the platform side by recomputing yearly sentiment after removing comments flagged as ironic to test whether negative sentiment trends were an artefact of irony prevalence. Fourth, rare-label behaviour was explicitly monitored. Because the three misinformation-related indicators have low base rates and were modelled with conservative decision thresholds, they are more susceptible to under-detection than high-prevalence indicators; as a result, their observed yearly proportions should be read as lower-bound estimates. All diagnostics were designed to preserve the alignment logic – same years, same constructs – so that changes in results could be traced to the diagnostic condition rather than to a change in data.

4.3.7 Reproducibility notes

The entire procedure is reproducible from the compiled CCAM workbook and the yearly platform outputs used elsewhere in the thesis. Each collapsed indicator maps to a single sheet and set of columns in the CCAM file, and each platform indicator maps to a single sheet in the comment-results workbook. The alignment is deterministic by calendar year, and the statistical procedures are standard. Any subsequent updates to CCAM toplines or extensions to additional years can be incorporated by appending rows to the CCAM table and re-running the alignment and correlation steps without modification to the core logic.

5 Results

This chapter presents empirical findings from the study, and begins with corpus composition and preprocessing outcomes, then summarize model performance on the held-out test split used for development and finally provide the full-run descriptive results (prevalence of labels and sentiment dynamics across years). The objective is to answer the research questions with a transparent chain from raw data to reported indicators.

5.1 Corpus and preprocessing outcomes

The full archive combined all per-year NDJSON files (2010–2024), preserving original API fields and adding a year variable. Preprocessing removed rows with no alphanumeric content and filtered to English using fastText LID with a threshold and fallback, then applied global de-duplication by comment_id after concatenating all years. Early baseline runs operated on per-year files (with de-dup limited to within-run). After the final consolidation step, the analytic file contained 205,792 unique English comments (the dataset used for modelling and all full-run summaries).

5.2 Model evaluation on development test split

Performance is reported on the held-out test split from the curated, labelled subset. For the multi-label output layer, micro-F1 = 0.305 and macro-F1 = 0.146, reflecting class imbalance and sparse support for rare labels. The sentiment output layer achieved macro-F1 = 0.735 overall and 0.423 on the irony-positive subset. These results motivate using an irony-specialised encoder for sentiment at inference.

5.2.1 Sentiment

The sentiment distribution shifts decisively across the period, with negative sentiment rising year-on-year from a minority share at the start to a clear majority in the early 2020s. In 2010, negative sentiment accounts for 32.0% of comments, increasing to 41.6% in 2011 and 45.7% in 2012. A temporary easing appears in 2013 at 40.4%, followed by renewed growth in 2014 (43.6%) and 2015 (47.2%). The threshold to a consistently negative-dominant profile is crossed in 2016 at 51.2%; values remain elevated in 2017 (47.4%) and 2018 (48.5%), before climbing further in 2019 (54.8%) and 2020 (54.2%). There is a brief moderation in 2021 (51.6%), after which the series reaches its maximum in 2022 at 62.8%, followed by only partial reversion in 2023 (60.6%) and 2024

(59.5%). Over the full window the cumulative increase is +27.5 percentage points, with the steepest two-year rise occurring between 2021 and 2022.

Positive sentiment declines correspondingly. It begins at 41.7% in 2010 and drops to 34.2% in 2011 and 22.1% in 2012; after a modest recovery in 2013 (24.3%) and 2014 (29.0%), it is 28.6% in 2015 and dips below thirty again in 2016 (26.5%). A short-lived rebound in 2017 (29.4%) and 2018 (28.2%) is followed by sustained contraction through 2019 (23.9%) and 2020 (22.6%), a slight uptick in 2021 (24.0%), and then the corpus minimum in 2022 at 16.4%. Subsequent years remain low, at 17.8% in 2023 and 19.5% in 2024. The total change from start to end is –22.2 percentage points, and the 2022 trough coincides with the negative peak.

Neutral sentiment narrows without collapsing. From 26.3% in 2010 it moves to 24.2% in 2011 and 32.2% in 2012 (the only year in which neutral exceeds thirty percent), then 35.3% in 2013 before declining to 27.4% in 2014 and 24.2% in 2015. It continues edging down in 2016 (22.2%), holds steady in 2017 and 2018 (23.3% in both years), and falls to 21.4% in 2019. The early 2020s show 23.2% (2020) and 24.3% (2021), followed by 20.8% (2022), 21.6% (2023), and 21.1% (2024). The corpus thus contracts by –5.2 percentage points across the period, with volatility early in the decade giving way to a tighter band in later years.

Taken together, these trajectories indicate a structural rebalancing of the comment space from a near parity of positive and negative sentiment in 2010 to a durably negative-majority profile from the mid-2010s onward, culminating in an intensified negativity in 2022 and only partial normalization thereafter. The sentiment shift is not the result of a single shock year; the transition to negative dominance is already visible by 2016 and remains evident through 2024.

5.2.2 Hostility

Hostility increases over the observation window, with pronounced volatility early in the decade and a clear peak in the early 2020s. The corpus begins at 5.6% in 2010 and rises to 8.9% in 2011 before dipping to 6.9% in 2012 and briefly to 3.5% in 2013. From 2014 onward the trajectory is upward overall, moving to 7.1% in 2014 and 8.4% in 2015, crossing into double digits in 2016 at 11.0%. The mid-period holds a relatively elevated plateau – 10.2% in 2017 and 10.0% in 2018 – followed by 11.6% in 2019 and 11.8% in 2020. A modest easing appears in 2021 at 11.2%, after which the corpus reaches its maximum in 2022 at 15.4%. The final years show partial reversion, with 12.2% in 2023 and 10.6% in 2024.

Measured from the 2010 baseline to the 2022 apex, hostility increases by +9.8 percentage points, indicating a shift in antagonism and aggressive language. This shift is not confined to a single shock year: hostility is already approximately double its 2010 level by 2016 and remains at or above ten percent for eight consecutive years (2016–2023), before settling only slightly lower in 2024. The pattern is consistent with a structural hardening of discourse, with 2022 concentrating the period's maximum antagonism and subsequent years failing to return to early-decade levels.

5.2.3 Distrust and Trust

The balance between distrust and trust shifts decisively over time, with distrust becoming increasingly dominant and trust contracting to single digits over time. In 2010, distrust and trust are near parity at 22.4% and 22.6%, respectively. Distrust climbs to 30.6% in 2011 and 34.9% in 2012, dips to 28.9% in 2013, and then resumes growth in 2014 (31.7%) and 2015 (33.0%). The corpus reaches 37.6% in 2016 and remains elevated in 2017 (34.0%) and 2018 (34.3%). A renewed rise occurs in 2019 at 39.8% and in 2020 at 39.7%, with a slight easing in 2021 to 37.8% before the peak in 2022 at 48.3%. The final years show partial reversion to 44.1% in 2023 and 41.7% in 2024. Measured from the start to the 2022 apex, distrust increases by +25.9 percentage points; relative to 2010, the 2024 level is +19.3 points higher, indicating a persistent elevation beyond early-decade conditions.

Trust follows the opposite trajectory. From 22.6% in 2010 it declines to 19.2% in 2011 and reaches an early trough of 8.8% in 2012, after which it stabilizes at low levels: 9.5% in 2013 and 13.9% in 2014, followed by 15.7% in 2015 and 15.3% in 2016. The mid-period holds in the mid-teens with 17.7% in 2017 and 16.6% in 2018, then falls to 14.1% in 2019 and 12.1% in 2020. A minor uptick appears in 2021 at 13.6%, but the corpus minimum occurs in 2022 at 8.5%, after which it recovers slightly to 9.3% in 2023 and 11.5% in 2024. The net change from 2010 to 2024 is –11.1 percentage points. The widening separation between the two corpuses is stark: distrust exceeds trust by roughly 0 points in 2010, by 22.3 points in 2016, and by 39.8 points at the 2022 inflection, with a remaining gap of 30.2 points in 2024.

Taken together, these trajectories show a durable change in the comment space toward suspicion over openness. Trust never returns to its early-decade parity with distrust, and the 2022 low point for trust coincides with maximums in other adversarial indicators reported elsewhere in this chapter.

5.2.4 Irony

Irony becomes a steadily more prominent feature of discourse over the period, with a clear intensification in the early 2020s and a high plateau thereafter. The corpus opens at 8.3% in 2010 and rises to 12.1% in 2011 and 11.9% in 2012, followed by a temporary contraction in 2013 to 9.1%. From 2014 onward the overall direction is upward: 10.8% in 2014 and 14.1% in 2015, crossing into mid-teens in 2016 at 16.3%. A brief softening appears in 2017 (13.9%) and 2018 (14.3%), after which the corpus strengthens again in 2019 (17.2%) and remains elevated in 2020 (16.4%) and 2021 (15.9%). The high point occurs in 2022 at 20.5%, followed by 18.8% in 2023 and 18.9% in 2024. Relative to the 2010 baseline, irony increases by +12.2 percentage points at the 2022 apex and by +10.6 points by 2024, indicating not only a peak year but also a lasting shift to a higher level of ironic engagement. The co-occurrence of the irony maximum with the maximums in negative sentiment and hostility underscores a broader hardening of the discursive style that persists beyond 2022 rather than reverting to early-decade norms.

5.2.5 Misinformation-related indicators

Three indicators related to misinformation are tracked: mentions of misinformation as a concept, mentions of believing something that is generally accepted as being misinformation (such as popular conspiracy theories), and accusations that others are spreading misinformation. All three operate at lower absolute prevalence than sentiment, hostility, distrust, or irony, but they display coherent trajectories that align with the broader hardening of discourse in the early 2020s.

Concept remains uncommon across the entire period and exhibits only small movements year to year. The corpus begins at 0.4% in 2010 and remains 0.4% in 2011, dips to 0.3% in 2012 and 0.2% in 2013, and then fluctuates narrowly between 0.2% and 0.4% through 2020. A modest increase appears in 2021 at 0.5%, followed by the corpus maximum of 0.6% in 2022 and again 0.6% in 2023, before returning to 0.3% in 2024. The pattern indicates that explicit conceptual framing of misinformation is rare even when other adversarial indicators intensify, or alternatively that the models struggled with distinguishing this type of label in the dataset.

Belief is consistently the most prevalent of the three indicators and shows a sustained upward movement into the 2020s. It registers 2.7% in 2010 and rises to 4.3% in 2011, with a temporary decline to 3.2% in 2012 and a trough of 1.0% in 2013. Thereafter, it increases to 2.6% in 2014 and 3.1% in 2015 and reaches 4.9% in 2016. Mid-period levels remain elevated – 4.8% in 2017, 4.6% in 2018, and 5.5% in 2019 – before the corpus attains 5.7% in 2020 and 5.5% in 2021. The maximum appears in 2022 at 8.3%, after which the rate declines to 6.0% in 2023 and 5.2% in 2024. The cumulative change from 2010 to the 2022 peak is +5.6 percentage points, and the 2024 level remains nearly double the 2010 baseline.

Accusations that others are spreading misinformation also rise over time but remain at lower levels than belief. The corpus starts at 2.0% in 2010, increases to 3.0% in 2011, and then dips to 2.3% in 2012 and 0.7% in 2013. It returns to 1.9% in 2014 and 2.2% in 2015, reaching 3.7% in 2016 and holding near that level thereafter: 4.2% in 2017, 3.9% in 2018, 4.0% in 2019, and 4.2% in 2020 and 2021. The maximum occurs in 2022 at 5.5%, followed by 4.2% in 2023 and 3.7% in 2024. Despite low absolute levels, the upward drift into 2022 mirrors the pattern seen in hostility and negative sentiment, suggesting greater reliance on negative discourse during the period's discursive peak.

Viewed together, these indicators show that explicit discussion about misinformation remains rare, while concrete belief in debunked claims and accusations against others are more common and responsive to the same pressures that elevate negativity and antagonism. The 2022 concentration of maxima in belief and accusation, and the minor crest in concept, further supports the characterization of 2022 as the period's inflection year for adversarial modes of expression. Given their low base rates and the stricter decision thresholds applied, these three indicators are more susceptible to under-detection than high-prevalence indicators; observed values should therefore be read as conservative estimates of stance prevalence.

5.3 CCAM results and comparisons

To begin the section with a short summary of the explanation of CCAM in section 4.3. Over 2010–2024, CCAM indicators show broadly rising orientations with early-2020s deviations. Acceptance increases from 61% at the start to 73% by 2024, interrupted by a 2022 dip to 70%. Worried climbs from 53% to a sustained high plateau, peaking at 70% in 2021 and settling at 64–65% thereafter. Harm (to you) moves from 31% to the mid-40s by the end of the period, with a local high around 50% in 2021 and a 2022 softening that partially recovers. Importance (high) shifts from the low 20s early on to the mid–high 30s in the early 2020s, with a salience crest in 2019–2021. Recognition of scientific consensus strengthens from the mid-30s to the upper-50s, easing slightly in 2023 before recovering in 2024. Certainty (high) reaches a late-2010s plateau at 73% (2019–2020) and then declines in the early 2020s, bottoming at 49–52% (2022–2024).

5.3.1 Sentiment and CCAM Acceptance/Worried

The relationship between the comment-space tone and population benchmarks is clearest when sentiment is aligned with CCAM acceptance and worried. Over the fifteen-year window, acceptance rises from 61.0% in 2010 to 73.0% in 2024, with a single deviation in 2022 at 70.0%. Worried follows a similar upward trajectory, moving from 53.0% in 2010 to a sustained high plateau in the late 2010s and early 2020s – 66.0% in both 2019 and 2020, peaking at 70.0% in 2021 and remaining elevated at 64.0–65.0% thereafter. Against these benchmarks, the distribution of sentiment in comments shifts toward a durable negative-majority profile: negative sentiment increases from 32.0% in 2010 to 59.5% in 2024, reaching 62.8% at the 2022 maximum, while positive sentiment declines from 41.7% to 19.5% with a 2022 minimum of 16.4%, and neutral tightens from 26.3% to 21.1% with modest fluctuations around that band.

The year-by-year co-movement is reflected in strong linear associations. Acceptance correlates positively with negative sentiment ($r = .79$) and negatively with positive ($r = -.69$) and neutral ($r = -.51$) sentiment. Worried exhibits the same pattern: a positive association with negative sentiment ($r = .77$) and negative associations with positive ($r = -.58$) and neutral ($r = -.64$) tone. The timing of inflections is informative. When acceptance attains early consolidation in 2016 at 70.0%, the comment space has already crossed into negative-majority at 51.2%; when acceptance recovers from the 2022 dip to 73.0% in 2024, negative sentiment remains elevated at 59.5% and positive sentiment remains depressed at 19.5%. The alignment with worried is similarly concrete. As worried rises from 58.0% in 2016 to 66.0% in 2020, negative sentiment increases from 51.2% to 54.2%; when worried peaks at 70.0% in 2021, negative sentiment moderates briefly to 51.6% but then intensifies to 62.8% in 2022 while worried remains high at 64.0%. In other words, across the full corpus the consolidation of climate reality and sustained public concern coincide with a progressively harsher tone in open comment space, and the 2022 maxima in negativity occur alongside continued high levels of concern rather than during a trough.

The magnitude of the shift is substantive when framed in absolute changes. Between 2010 and 2024 the growth of acceptance by twelve percentage points accompanies a twenty-seven-point rise in negative sentiment and a twenty-two-point decline in positive sentiment; between 2016 and 2024, during which acceptance remains at or above seventy percent, negative sentiment remains above fifty percent every year and never returns to its pre-2014 levels. The same holds for worried: once the CCAM corpus stabilizes above sixty percent (2017 onward), the comment space does not revert to a balanced sentiment profile; instead, negative sentiment predominates and culminates in 2022, while positive sentiment remains constrained even after the post-2022 easing. These patterns do not assign causal direction; they show that when the survey record indicates a public more settled on the reality of global warming and more persistently worried about it, the discursive output of commenters is not more constructive or neutral but more discouraging in tone, with fewer affirming responses and only limited space for a neutral relationship to the information.

5.3.2 Hostility and CCAM Worried/Harm

Hostility in the comment space aligns closely with CCAM indicators that capture affective pressure and perceived threat. The strongest relationship is with worried. From 2010 to 2024 the worried corpus rises from 53.0% to 64.0%, with a sustained high plateau from the late 2010s onward – 66.0% in both 2019 and 2020 and a peak of 70.0% in 2021 – followed by 64.0% in 2022 and 2024 and 65.0% in 2023. Over the same period hostility increases from 5.6% in 2010 to a mid-2010s step-change at 11.0% in 2016, holds near that elevated level through 2020 at 11.8%, and then reaches its corpus maximum at 15.4% in 2022 before easing to 12.2% in 2023 and 10.6% in 2024. The linear association between the two corpuses is strong ($r = .80$), reflecting the fact that higher public worry is consistently accompanied by a more antagonistic interactional style in comments. The timing reinforces this reading: when worried moves from 58.0% in 2016 to 66.0% in 2020, hostility remains on a high shelf around eleven to twelve percent; when worried remains elevated in 2022 at 64.0% following its 2021 apex, hostility not only persists but climbs to its peak of 15.4%. The cumulative change relative to the corpus start is substantial – hostility is nearly double the 2010 level by the mid-2010s and almost triple at the 2022 apex – indicating that concern registered in surveys is mirrored by a durable thickening of antagonism in open discourse.

Perceptions of harm to the respondent personally as asked in the CCAM surveys show a similar, though more moderate, alignment with hostility. The collapsed harm (to you) indicator increases from 31.0% in 2010 to 46.0% in 2024, with notable waypoints at 42.0% in 2012, 49.0% in 2017, and 50.0% in 2021, followed by a deviation to 44.0% in 2022 and recovery to the mid-forties thereafter. The correlation between harm (to you) and hostility is moderate

($r = .52$), which is consistent with co-movement impaired by the 2022 dip in perceived harm despite peak hostility that year. Even so, when the public reports greater personal exposure to climate-related harm – around the late-2010s lift and into 2021 – the comment space routinely registers hostility at or above ten percent. Read together with worried, the pattern suggests that both generalized concern and personal threat appraisal are reflected in a sharper interactional temperature online, with Worried providing the tighter linear coupling and harm (to you) corroborating the direction of travel despite year-specific deviations.

5.3.3 Distrust and CCAM Acceptance

The comparison between distrust in comments and acceptance in CCAM shows a strong, positive association, indicating that consolidation of climate reality at the population level coincides with a discursive shift toward suspicion in open comment space. Acceptance increases from 61.0% in 2010 to 73.0% in 2024, with a sustained consolidation in the late 2010s (70.0% in 2016, 71.0% in 2017, 70.0% in 2018, 72.0% in 2019) and a momentary deviation in 2022 at 70.0%. Over the same period, distrust rises from 22.4% in 2010 to 37.6% in 2016 and 39.7% in 2020, reaching a maximum of 48.3% in 2022 before easing to 44.1% in 2023 and 41.7% in 2024. The linear association across years is strong ($r = .77$). The year-by-year anchors clarify the nature of this coupling. When acceptance first stabilizes at ≈70% in the mid-2010s, distrust has already moved into the mid-to-high thirties; as acceptance returns to 73.0% in 2024 after the 2022 dip, distrust remains ~20 points above its 2010 baseline despite partial reversion from its peak. The widening separation between distrust and trust documented earlier reinforces the point that mainstreaming of climate acceptance does not soften the comment environment; rather, commenters express more suspicion toward institutions, sources, and interlocutors as acceptance becomes the survey norm.

5.3.4 Irony and CCAM Certainty

The alignment between survey-based certainty and ironic style in comments is best characterized as a contrast in trajectories rather than a tight linear coupling. CCAM's certainty (high) – the share who are very or extremely sure that global warming is happening – rises through the late 2010s to a plateau of 73.0% in 2019–2020 after steady gains from 57.0% in 2010. It then drops sharply in the early 2020s, reading 58.0% in 2021, 51.0% in 2022, 49.0% in 2023, and 52.0% in 2024. Over the same period, irony in comments follows a different path: it increases from 8.3% in 2010 to mid-teen levels by the mid-2010s (16.3% in 2016), softens briefly in 2017–2018 (≈14%), and then climbs again to 17.2% in 2019 and 16.4% in 2020. The early 2020s bring a further intensification, with 15.9% in 2021 and a corpus maximum of 20.5% in 2022, followed by persistently high levels of 18.8% and 18.9% in 2023–2024.

When assessed as a year-by-year linear association, the pairing yields a near-null correlation ($r = -.12$), indicating that higher or lower levels of survey certainty do not scale proportionally with irony prevalence in comments. The temporal anchors clarify why: irony reaches its peak precisely when certainty is near its trough (2022), and irony remains elevated even as certainty stabilizes at lower levels in 2023–2024. The pairing remains substantively informative as a contrast. CCAM records a population that became very sure of the reality of global warming by 2019–2020 and then less sure in the early 2020s, while the comment space shows a durable shift toward ironic style, culminating in 2022 and not reverting to early-decade baselines thereafter. This indicates that ironic engagement is prevalent regardless of certainty levels and is likely shaped by dynamics other than linear changes in survey-reported confidence.

5.3.5 CCAM scientific consensus and accusations of misinformation

Recognition of scientific consensus in CCAM strengthens across the corpus, and accusations of misinformation in the comment space – while low in absolute terms – become more frequent over the same period. The consensus (true) indicator rises from 34.0% in 2010 to 41.0% in 2011 and 44.0% in 2012, stabilizes near the mid-forties in 2013–2014, and then advances through 51.0% in 2016, 56.0% in 2017, and 57.0% in 2018, with a slight easing to 55.0% in 2019 and 56.0% in 2020. It reaches 59.0% in 2021, records 58.0% in 2022, dips to 53.0% in 2023, and recovers to 57.0% in 2024. Over the same window, accusations of misinformation in comments move from 2.0% in 2010 to 3.0% in 2011, contract to 2.3% in 2012 and 0.7% in 2013, then rebuild through 1.9% in 2014 and 2.2% in 2015, reaching 3.7% in 2016. Mid- to late-decade values are 4.2% in 2017, 3.9% in 2018, 4.0% in 2019, and 4.2% in both 2020 and 2021. The corpus maximum occurs in 2022 at 5.5%, after which it declines to 4.2% in 2023 and 3.7% in 2024.

Viewed as a yearly linear association, consensus (true) and accusations co-move to a noticeable degree ($r = .84$). The practical reading is that acknowledgement of a scientific consensus expands at the population level at the same time as accusations remain available and are increasingly deployed within comment discourse, particularly around the 2022 discursive apex. Because accusations are a low-prevalence label, the upward co-movement should not be mistaken for high absolute frequency; rather, accusations function as a delegitimization-action that becomes more common precisely when survey measures suggest the issue is settled among respondents. This pattern does not attribute causality to either corpus. It indicates that the period in which consensus recognition strengthens is also the period in which commenters more often contest others' claims as misinformation, culminating in 2022 and easing thereafter without returning to early-decade baselines.

5.3.6 CCAM issue importance and negative sentiment

Issue salience in CCAM aligns with a harder tone in the comment space. The importance (high) indicator begins at 24.0% in 2010, dips to 19.0% in 2011, and remains near the low twenties through 2014. It edges upward to 25.0% in 2015 and 26.0% in 2016, rises to 31.0% in 2017, and reads 28.0% in 2018. A pronounced lift appears in 2019 at 42.0%, followed by 37.0% in 2020 and 43.0% in 2021. The early-2020s show 35.0% in 2022, 37.0% in 2023, and 36.0% in 2024. Over the same years, negative sentiment in comments moves from 32.0% in 2010 to 41.6% in 2011 and 45.7% in 2012, reaches 51.2% in 2016, holds in the high forties in 2017–2018, and then increases to 54.8% in 2019 and 54.2% in 2020. It moderates briefly in 2021 at 51.6% and then reaches 62.8% in 2022, with 60.6% in 2023 and 59.5% in 2024.

The year-by-year alignment yields a strong positive correlation between importance (high) and negative sentiment ($r = .74$). The practical meaning is straightforward: as more respondents rate global warming as extremely or very important, the comment environment tends to be more negative in tone, with the strongest concentration of negativity appearing in 2022 when importance remains in the mid-thirties. Salience in surveys therefore does not manifest as a more supportive or affirming discussion climate on the platform; rather, the tone is more discouraging precisely in periods when the issue is judged most important. This is consistent with the broader shift documented earlier from near parity of positive and negative sentiment in 2010 to a persistent negative-majority profile from the mid-2010s onward, culminating in the early-2020s hardening of discourse.

5.3.7 Uncertainty and sensitivity

The alignments above compare yearly trajectories from representative surveys (CCAM) with yearly trajectories of discursive prevalence among commenters. They are descriptive, indicating co-movement and contrast across independent systems. The precision of yearly proportions varies by prevalence and comment volume: high-prevalence indicators (e.g., negative sentiment, distrust) show narrower uncertainty than low-prevalence indicators (e.g. concept). Given conservative decision thresholds for the rare indicators, their reported proportions should be read as lower-bound estimates (cf. 6.2.5).

Planned sensitivity checks confirm that the main findings do not hinge on a single year or a specific collapse choice. Excluding 2022 reduces peak magnitudes in platform indicators but does not change their direction: negative sentiment remains a majority from 2016 onward; hostility and distrust remain elevated; irony had already doubled relative to 2010. Stricter CCAM collapses (e.g., “Very worried” only; “Extremely important” only) attenuate correlation magnitudes by reducing variance but preserve signs. Recomputing sentiment without irony-flagged comments lowers negative levels modestly at each point yet retains the upward slope and the 2022 maximum. Pair-specific outcomes remain differentiated: certainty (high) vs irony shows a near-null linear association, whereas consensus (true) vs accusations co-move strongly even though accusations are low in level.

6 Analysis

This chapter analyses the empirical results through the lens of the theoretical framework set out in Chapter 2. In that framework, attunements (filters of perception) shape how information is met, evaluated, and acted upon, while information resilience concerns the capacity to cope with disruptions in the information environment. The analysis follows a structure that is based on the research questions: reading the trajectories of the indicators in the YouTube corpus (2010–2024), examining relations among indicators and their alignment with year-level CCAM benchmarks of public perceptions, interpreting what these relations indicate about perceptual filters and worldview, and drawing implications for information resilience. A final section analyses the success of the experimental method used in the study.

6.1 Trajectories as manifestations of perceptual filters

The first question focuses on how the indicators – sentiment, irony, trust/distrust, hostility, and the misinformation-related indicators of concept, belief, and accusation – change across the period. The results show a clear rebalancing of tone toward negative sentiment, a durable elevation of distrust, a rise in hostility, and a steady intensification of irony, with belief and accusation cresting in the early 2020s albeit at lower absolute levels. These shifts are not isolated shocks; they are structural in the sense that mid-2010s changes persist thereafter, culminating in a 2022 peak across several indicators and followed by only partial easing. Read through the lens of attunements, these trajectories depict an environment in which incoming information and people are increasingly met with negative expectation, suspicion, and adversarial framing (Thonhauser, 2019, p.1-2, 5-6).

6.1.1 Tone: from near parity to a negative-majority profile

The corpus moves from near parity of positive and negative sentiment in 2010 to a negative-majority profile from 2016 onward, peaking in 2022 (negative ≈ 62.8%) and only partly easing by 2024 (≈ 59.5%). Positive sentiment declines correspondingly and remains low at the end of the window. Neutral sentiment narrows to a tighter band. In attunement terms, this signifies an increasingly constrictive worldview – a way of meeting information that anticipates trauma-triggers or the possibility of the person becoming emotionally overwhelmed and unable to regulate those emotions, making neutral perception and absorption of the information, as well as affirming responses, less available as default options (Thonhauser, 2019, p.5-6; Munro, 2013, par.4).

6.1.2 Openness to information, people and society: widening gap between distrust and trust

Distrust increases from early-decade levels into the high thirties by the mid-late 2010s, reaching a maximum in 2022 (≈ 48.3%), and remains well above its 2010 baseline by 2024. Trust, in contrast, declines into single digits around 2012 and never recovers to parity. The gap between distrust and trust widens markedly and persists. As a filter of perception, this pattern signals a form of self-protection against perceived danger: information is received, but the presumed credibility of sources narrows, which constrains possibilities for and willingness to broaden perspectives and correcting possible inaccuracies in already held belief (Myers, 2014, p.57-60,66).

6.1.3 Interactional stance: elevated hostility

Hostility – language that enacts antagonism rather than disagreement – roughly doubles by the mid-2010s and peaks in 2022 (≈ 15.4%), then partially reverts without returning to early-decade norms. Hostility isn't an indicator of perceptual filters as much as it is an indicator of changing norms. It's the visible expression of when behaviours that are generally considered to be unacceptable become normalized. Hostility encompasses the language and behaviour that crosses the boundaries of the otherwise "normal" behaviour norms and social codes of society, including harmful language and hate speech.

6.1.4 Style of expression: irony as distancing

Irony rises from single digits to the high-teens and also peaks in 2022 (≈ 20.5%), remaining elevated thereafter. Irony is interpretively important because it mediates tone: negative stance can be voiced indirectly, and positive wording can carry negative intent. The observed intensification of irony therefore compounds the shift in tone – messages travel, but they do not land plainly – and increases ambiguity in how the information is absorbed.

6.1.5 Misinformation-related indicators: belief and accusation crest, concept remains rare

Among the three misinformation-related indicators, belief (mentions of someone believing debunked claims) is more prevalent than accusation, and both rise into the early 2020s, reaching their peak in 2022 before easing; concept (neutral mention) remains rare across the period. Given conservative thresholds and low base rates, reported levels are lower-bound estimates; the substantive point is that endorsement and delegitimization become more available coping moves in the same temporal window that tone hardens and distrust rises (Myers, 2024, p.57-60,66).

Together, these trajectories form a picture of the background orientation of people's perceptual filters, in which public talk is more negative, more suspicious, more adversarial, and less straightforward over time, with an inflection around 2022. In the framework of Chapter 2, this constitutes a shift in perceptual filters: the same kinds of information are increasingly predisposed to be met as threat, as untrustworthy, and as opportunities for boundary-drawing rather than for cooperative problem-solving.

6.2 Relations among indicators and alignment with CCAM

The second research question asks how the indicators move in relation to one another and how those patterns align with public perceptions of climate change as captured by Climate Change in the American Mind (CCAM). The empirical picture is twofold: first, within-corpus relations that cluster tone, stance, and style during periods of pressure; and second, between-corpus alignments in which year-to-year movement in CCAM co-occurs with changes in the comment space. The analysis reads these as perceptual profiles rather than single-variable effects.

6.2.1 Within-corpus relations: converging adversarial perceptual profiles

The co-occurrence of negative sentiment, distrust, hostility, and irony is most pronounced from the mid-2010s into the early 2020s, culminating in 2022. In practical terms, when sentiment becomes more negative, hostility is elevated, distrust exceeds trust by a large margin, and irony is more often chosen as the delivery style. These indicators become mutually reinforcing: indirectness lowers interpretability, hostility raises the cost of engagement, and distrust narrows who can be heard and who dares to express themselves. The relative rarity of the misinformation indicators means they contribute less to the overall volume but are directionally consistent with the perceptual profile in the early 2020s (belief and accusation both crest then).

6.2.2 Between-corpus alignments: CCAM and platform indicators

The strongest alignments pair CCAM measures of acceptance, worry, importance, and (to a lesser extent) harm (to you) with negative sentiment and hostility, while acceptance also aligns with distrust. Certainty shows a near-null linear relation with irony; nonetheless, irony remains elevated even when certainty is lower in the early 2020s. Accusations of misinformation co-move with recognition of scientific consensus despite low absolute levels. These patterns are descriptive co-movements, not causal attributions. They indicate that years in which survey respondents consolidate or elevate concern about climate change are also years in which the open comment space is harder in tone, less open in posture, and more prone to accusatory stances.

The internal clustering of negative tone, distrust, hostility, and irony forms a discursive perceptual profile that tends to be more present in years when CCAM indicates stronger public acceptance, worry, or awareness.

6.3 What relations indicate about perceptual filters and worldview

The third research question asks what the year-level relations imply about how people's perceptual filters and worldview shape talk in the comment space over time. The theoretical starting point is that people are never neutral observers: attunements filter what feels credible, threatening, or irrelevant – they condition how readily information can be taken up at all. Sustained external pressures – here, an issue like climate change that remains chronically salient – can recalibrate these filters. From this perspective, the indicators trace how orientations are enacted in public talk rather than measuring private beliefs directly.

6.3.1 Negative tone with indirect expression: constrictive orientation and emotional distance

When acceptance and worry are higher in the survey record, the comment space is more negative and more ironic. Read through the attunement-lens, this suggests a possible worldview that anticipates conflict and uses indirect communication as a coping mechanism. In terms of perceptual filters, this is a protective posture: the discursive space keeps emotional distance while treating incoming claims as risky. But, especially in written form like YouTube comments, this can lead to misinterpretations and misunderstandings in a societal discourse that already is perceived as concerning or maybe even hostile to some degree.

6.3.2 Openness and suspicion: defensive absorption under consolidation

The alignment between rising acceptance in CCAM and rising distrust in comments indicates that agreement with climate facts in the population context does not automatically translate into a more open comment environment.

Instead, suspicion of institutions, sources, and interlocutors widens even as the issue is settled in surveys. The perceptual filter in this context can be interpreted as that the awareness and belief in the existence of societal disruptions (in this case climate change) doesn't just create fear, but that the fear causes people to be more distrusting of others as well. This clearly has implications for how misinformation is to be counteracted and fought, since it indicates the possibility that learning the truth can cause more disruption and division in society, not less.

6.3.3 Enacting disagreement as antagonism: interactional hardening

The co-movement of worry and hostility points to enactment rather than internal appraisal alone. Contexts with higher levels of concern indicates that disagreement increasingly is expressed through antagonism rather than as problem-oriented debate. As a filter, this frames others as opponents to be overcome rather than co-participants. The discursive environment becomes less forgiving of missteps and less oriented to solution-seeking, both of which lower resilience by raising the interactional cost of sense-making.

6.3.4 Coping mechanisms in contested information: belief, accusation, concept

The low-prevalence but directionally informative indicators, belief and accusation crest in the very years where the adversarial perceptual profile is strongest. Belief means adopting a narrative because it supplies a simple framework that eases uncertainty, even if the story is false. Accusation is a delegitimating action: it confronts other people's beliefs in antagonistic ways when the social cohesion is already on shaky ground. Concept remains rare; when present without belief or accusation, it marks a form of meta-talk about misinformation – discussing the concept instead of expressing opinions about something being or not being misinformation, rather than endorsement or blame. It's a comparatively open coping channel, but it does not dominate the period. In attunement terms, the availability of belief and accusation as coping mechanisms under pressure signals a shift in the approaches when engaging contested claims.

The relations point to a background orientation that, in periods of higher concern and awareness, is more suspicious (distrust), more adversarial (hostility), more distant (irony), and more prone to delegitimating actions (accusation). Even as public perceptions consolidate on the reality of climate change, the comment environment does not become more straightforward or cooperative; instead, it is harder to engage in and be a part of. These are filters of perception made visible at scale.

6.4 Implications for information resilience

The fourth research question synthesises what the trajectories and relations indicate about the impact of perceptual filters – and climate change as sustained external pressure – on information resilience. In the framework used here, resilience concerns the capacity to withstand, adapt to, and recover from disruptions in the information environment. That capacity depends not only on skills (information literacy) but also on the background conditions under which information is received. The findings above allow two complementary readings: a reserved perceptual profile and an open perceptual profile.

6.4.1 A reserved perceptual profile (reduced ease of engagement)

A reserved perceptual profile is characterised by negative-majority tone, elevated distrust, elevated hostility, and elevated irony, with belief/accusation more available as coping mechanisms. This perceptual profile is most present in the early 2020s and peaks in 2022; only partial easing follows. In resilience terms, reserved means:

- Fewer open channels: distrust narrows who is given the opportunity to speak, and who dares to express their opinions, and accusation delegitimates not just opposing opinions, but people with different opinions.

- Higher interactional costs: hostility transforms disagreement into a battlefield, and repairing any damage caused requires more effort.

- Lower interpretability: irony makes signals harder to read, even supportive messages can be misinterpreted. This can cause unnecessary disagreements and distrust, especially since irony is often connected to distrust, creating an environment where it can even be difficult to be allowed to sort out or even explain the misunderstanding.

- Greater persistence of strain: because tone, posture, and style shift together, the system has less slack to absorb misunderstanding, controversy and disagreements without hardening further.

Under a reserved perceptual profile, information literacy alone is insufficient. People may know how to evaluate sources and claims, but filters of perception – defensive, adversarial, indirect – can pre-empt absorption of information. This is particularly clear in years when CCAM indicates consolidation of acceptance and sustained worry: exactly when collective coordination is needed, the ambient conditions of discourse become less conducive to it.

6.4.2 An open perceptual profile (greater ease of engagement)

By contrast, an open perceptual profile would feature a less negative tone, lower distrust, lower hostility, and reduced reliance on irony, with concept (neutral mentions) more visible than belief/accusation when misinformation is discussed. The corpus contains few extended periods that match this profile, but relative improvements (e.g. partial easing after 2022) show that an opening perceptual profile is possible. In resilience terms, an open perceptual profile means more pathways for communication, coordination and correction, lower costs for participation, clearer signals, and minimizing risk of misinterpretation (Ahlryd & Hanell, 2024, p.270-271).

6.4.3 Predominant perceptual profile 2010-2024

The comparative scarcity of open periods suggests that, over 2010–2024, the default orientation migrated toward the reserved perceptual profile. This indicates that, across an open comment environment like the YouTube comment section, years with a predominantly reserved perceptual profile – more negative sentiment, higher distrust/hostility, greater reliance on irony – provide less favourable conditions for information resilience than open perceptual profile-years.

6.4.4 What climate change contributes to this picture

This thesis does not claim that indication and/or correlation with climate perceptions also implies causation – although it could argued that this is something that warrants further study – but the co-movement between CCAM indicators (acceptance, worry, importance) and the adversarial perceptual profile suggests that a sustained external pressure – in this case, climate change as a long-running public concern – travels into background orientation and perceptual filters. That is, as the issue becomes more accepted and more salient, public talk in this space becomes less straightforward and less open, not more. In terms of resilience, this indicates an environment where coping is real, but coping via protective/defensive filters dominates over coping via openness, and that it has concerning implications in relation to information resilience (Myers, 2024, p.57-60,66).

6.5 Adequacy and limits of the methodological approach

The fifth and final research question asks how adequate and trustworthy the experimental approach is and what limits follow for interpretation.

6.5.1 Strengths for the present questions

Temporal comparability by design

The year-specific retrieval and per-year summarisation support comparisons that are not simply volume artefacts; the analysis relies on proportions and means within year windows, not raw counts. This design choice is essential for the other research questions.

Alignment with social-media language

The modelling stages move from lexicon baselines to transformer models and finally to a fine-tuned RoBERTa system, with a hybrid inference strategy for irony and sentiment. This pipeline directly targets natural language text and figurative expressions, which are central to the indicators in question.

Explicit handling of irony

Treating irony as a first-class indicator and conditioning sentiment interpretations on irony (where relevant) reduces a known source of polarity error and is congruent with the theoretical focus on style as part of filter expression.

Transparent calibration and sensitivity

Re-thresholding irony to match test-set prevalence, perturbation checks around the chosen threshold, and probability-level summaries for rare labels all improve interpretability and guard against threshold artefacts.

Contextual triangulation, not fusion

Using CCAM as a benchmark rather than as a pooled dataset respects the independence of the corpora and keeps the alignment strictly at the year level, appropriate for orientation-level inferences rather than individual-level claims.

6.5.2 Limits and their implications

Domain shift and rare indicators

Even with trained NLP models, domain shift from the original training sources to YouTube comments persists. Rare indicators (concept, belief, accusation) are particularly sensitive to thresholding and class imbalance. The study mitigates this by reporting probabilities and acknowledging lower-bound prevalence, but measurement error remains a plausible issue that needs to be taken into account. Interpretation therefore prioritizes direction and perceptual profiles over exact magnitudes for these indicators. Something that would likely improve this for future studies is a larger training set and more time to train the models, since the test-samples were smaller and had to be completed within a short amount of time.

Hybrid inference and model separation

The hybrid assembly uses an irony-specialised model for irony and sentiment and a base model for other indicators. This setup works better in practice, but it can be assumed that the different models have different type of errors – like two different types of thermometers measuring the temperature. They will be similar, but if they use different measurements their errors will likely look different, even if they both measure the temperature. Because of that, the thesis avoids making detailed claims about what happens at the exact same moment inside a single comment (in other words, it avoids looking at a specific measurement during a specific minute) and instead focuses on patterns across whole years.

Sampling frame and representativeness.

The retrieval used the “news” query to achieve a broad sample, while excluding music and obvious off-topic content. This captures a wide discourse but is not representative of all social-media discourse, nor of any national population. As a result, the conclusions concern the structure of open comment discourse over time, not national opinion or platform-general behaviour. The CCAM alignment is interpreted accordingly (as co-movement, not match).

Duplicates and preprocessing constraints

The pipeline ultimately deduplicates globally by comment_id; an earlier decision to combine comment_id and text to avoid false positives was not used in the last run, possibly removing more near-duplicates than intended. Summaries nonetheless operate on proportions, which reduces sensitivity to absolute counts, but id-only deduplication could slightly deflate prevalence in years with repeated IDs across videos. The analysis emphasises directional findings and cross-indicator configurations that are robust to moderate prevalence shifts.

Held-out performance and generalisation

On the held-out test set, performance varies by indicator. Sentiment and irony are the most stable. Distrust and hostility also perform reliably and are treated as core anchors in the analysis. By contrast, the three misinformation indicators (concept, belief, accusation) have a low base rate and use conservative decision thresholds; their yearly proportions are therefore interpreted as directional context, not on par with higher-prevalence indicators. Accordingly, interpretations prioritise sentiment, irony, distrust, and hostility as the primary measurements, with the misinformation indicators used to qualify those patterns.

Implication for the fifth research question

The approach is adequate to for the purpose of the study. The study is not designed to find the causes of the patterns found or study individual comments, which means findings should be read as interpretations of patterns within public discourse across a time-period of 15 years, and what perceptual filters these patterns point to. That means, the finding must be read with awareness of the modelling limits and the sampling frame. That said, practical improvements could benefit future studies, such a improvements to the training of the model, the deduplication, and the time it took to run the scripts.

6.6 Synthesis

Bringing the pieces together, the analysis supports four integrative points.

6.6.1 From skills to conditions

The period 2010–2024 shows a migration from a comment space where positive and negative sentiment were near parity to one where negative tone is durably dominant, distrust rises and trust contracts, hostility increases, and the use of irony becomes frequent. These conditions make information literacy harder to exercise effectively because the costs of uptake are higher: claims are met defensively, interlocutors are encountered antagonistically, and messages are expressed in ways that are open for miscommunications. Perceptual filters thus constrain the value of skills unless ambient conditions soften.

6.6.2 Sustained pressure and defensive coping

When CCAM records higher acceptance, higher worry, or greater importance, the comment space is not more supportive; it is harder – more negative, more hostile, more suspicious. In the language of the framework, a sustained external pressure (climate change as lived experience) appears to travel into the background orientation of society. Coping is visible, but coping via defensive filters (indirectness and fear-based communication) is more available than coping via openness.

6.6.3 Resilience as a property of the environment

The configuration that peaks in 2022 – negative tone + distrust + hostility + irony with belief/accusation available – represents a tightening of the discursive environment. Under such configurations, resilience is reduced: there are fewer pathways for coordination, greater costs for engagement, and slower recovery after shocks. Periods of relative loosening exist but are limited; the default has shifted toward tightness.

6.7 Summary of the Analysis chapter

Across fifteen years of platform discourse, the indicators reveal a reorientation of background conditions under which public information is encountered: harder tone, narrower openness, adversarial interaction, and indirect delivery. These are not idiosyncrasies of particular topics, but structural features of the environment in which sense-making occurs. In conjunction with independent survey evidence that climate change has become more accepted, more salient, and more worrying, the analysis indicates that filters of perception have tightened in ways that lower the ease of resilient engagement. For practice and policy in information science, this underscores that strengthening information resilience requires attention not only to individual competencies, but also to ambient discourse conditions – making talk more straightforward, less hostile, and more open – so that skills have room to operate constructively.

7 Discussion

This chapter interprets the trajectories of the seven indicators through the lens of perceptual filters and information resilience. Platform-level shifts, pattern alignment with CCAM, the roles of the different indicators indicators, and the differentiation between environmental constraints and skills-centric explanations are discussed. Practical implications follow for design and policy that reduce antagonistic configurations and support more open perceptual profiles. The chapter closes by outlining alternative readings, scope limits, and transferability beyond YouTube and the 2010–2024 window, keeping conclusions proportionate to the evidence.

7.1 Perceptual filters, platform dynamics, and identity

Myers (2014) describes climate denial as a coping mechanism for preserving a person's sense of security and as a way of walling of anxieties and fears that would overwhelm them. He describes how this fear of information won't just make them more susceptible to misinformation but causes them to actively avoid accurate information that threatens their worldview, and how this can become a collective action among people to practice as a group to make their actions feel more morally acceptable (p.53-60, 66). The findings in this thesis seem to support his findings, indicating that there may be a relationship between perceptual filters and worldview in relation to climate change, and the state of online discourse on YouTube – especially when it comes to sentiment, distrust and hostility. These findings indicate that when concern about climate change is higher, the discourse environment becomes less constructive. The more people believe in that a crisis is occurring, the more they close themselves off – the more they believe that the crisis isn't happening, the more open and constructive the discourse becomes. This seemingly supports information behaviour and tribal behaviour described by Myers (2014) and raises concerns about the information resilience people currently have and the effectiveness of information literacy (p.53-54).

Haider & Sundin (2022) describe how social media platforms create information bubbles through algorithm-driven personalization, which in turn reinforces already existing worldviews. This leads to an erosion of democratic discourse and solidifies political polarization, distrust, as well as the effectiveness of misinformation campaigns (p.101-102, 146-150). In this study hostility and distrust are found to be mutually reinforcing, which in part could be seen as the findings substantiating their perspective, but also something that intensifies the issues they describe by creating a circle of distrust and hostility within the discourse that is already polarized by algorithms. It might be worth repeating that the hostility indicator marked comments that went beyond negative, beyond insults, and beyond curse words and profanity. Comments that included serious threats, attacks, and comment with language so degrading and humiliating that they would be described as 'vile' in society at large. These were not just negative comments, insults or profanity. Seeing this indicator be mutually reinforcing with distrust raises additional concerns on their own, not just regarding the impact of this relationship on information resilience, but also on the misinformation campaigns that Haider and Sundin (2022) describe. The only thing those with bad intentions towards democracies – such as Russia – would need to do to exploit this would be to add bots and trolls that amplify distrust and hostility, which they seemingly have already realized (Haider & Sundin, 2022, p.101-102, 120-122; Llewellyn et al, 2019, p.39-41). Furthermore, Hurley and Molloy (2025) reveal that before changing beliefs, not only were the participants beliefs directly connected to identity but was also reinforced by mistrust of the sources of information promoting opposing beliefs (p.6–7). This adds to the complexity of the increase in distrust between 2010-2024 in the findings of this study, since this indicates that the participants of their study used mistrust of sources as a coping mechanism in the same way that Myers (2014) argues that they use denial as a coping mechanism, which becomes the same kind of reinforcing relationship as the one found between distrust and hostility in this study (p.57-60, 66). Based on the findings in this study, these reinforcing loops create conditions that promote the reserved perceptual filter, which in turn creates the opposite of what a society need for information resilience – a constructive connection between information and people (Hurley & Molloy, 2025, p.14-16).

7.2 Irony and indicator dynamics

A further point concerns irony. In the results, irony rises in the same years that negativity, distrust, and hostility increase. Read alongside the diagnostics – where recomputing yearly sentiment after removing comments flagged as ironic did not remove the negative trend – indicating that the trend is not an artefact of irony prevalence – this suggests that irony did not create the pattern so much as accompany it. In Myers's (2014) terms, when an issue confronts the worldview and triggers anxiety, indirect expression becomes a way to maintain distance from the threat while still speaking about it; the distancing protects the lifeworld (p.53–60, 66). In an open comment environment, that distancing requires interpretation from everyone else. When irony co-occurs with heightened distrust and hostility, the costs of engagement rise further, because it becomes harder to know which assertions are straightforward and which are not. This can become an obstacle for information resilience, because it obstructs clarity in communication when it’s needed most (Haider & Sundin, 2022, p.146–150).

7.3 CCAM alignment and the shift toward information resilience

The year-based movement of indicators relative to CCAM yielded interesting findings. The analysis shows that when the participants are more aware of climate change, the YouTube comment environment tends to exhibit more distrust, more hostility, and more irony, with a more negative score in sentiment. This does not establish causation, but it does indicate that public concern over climate change and the character of discourse move together over time. In Myers’s (2014) framing, such co-movement this would be consistent with the idea that climate change is interpreted as an existential threat impacts perceptual filters. Furthermore, in Haider and Sundin’s (2022) account, it also points to that platform conditions amplify emotionally charged, polarising material when attention spikes (Myers, 2014, p.53–60, 66; p.101–102, 146–150). The implications of seeing these findings side-by-side becomes fairly straightforward: when concern is high, the environment in which people meet information is less forgiving, and the pathways that normally enable mediation and clarification become narrower.

These findings add to the strength of the argument for moving the focus from information literacy alone to one including information resilience by shifting emphasis from only competencies (e.g. information literacy) to including the way people perceive the world around them. Resilience depends on the social environment in which skills are exercised: whether discourse is straightforward or indirect, open or adversarial, trusting or suspicious. The reserved perceptual profiles identified in this thesis constitutes a kind of environmental obstacle for information resilience: it raises the costs of participation, increases division, and narrows pathways for cooperation and communication. Equally, an open perceptual profile would lower friction, enhance acceptance to diversity in opinions and perspectives and increase cooperation, enabling societies to absorb shocks and attacks, and to retain stability and function in their information environment no matter what (Ahlryd & Hanell, 2024, p.273-278; Linvill et al., 2021, p.1-3, 16-19).

7.4 Practical and policy implications

The findings also underscore how quickly distrust and hostility can lock into a loop. As discussed above, distrust reinforces hostility, and hostility in turn normalises a presumption of bad faith that feeds distrust. Hurley and Molloy's (2025) participants describe mistrust of opposing sources as part of an identity-focused stance before opinion change. When translated into a large, public comment space, that stance scales into an underlying expectation that other people or sources are not to be believed (p.6–7). In such conditions, even routine disagreements become far more strained, because each side reads the other through perception filters that anticipate deception and/or animosity. The results here show that this loop is not rare or episodic across 2010–2024; it appears repeatedly, and when it does, it aligns with the reserved perceptual profile that is incompatible with functioning information resilience (Hurley & Molloy, 2025, p.14–16).

There are also practical implications for platformed news spaces. If distrust and hostility move together and co-occur with a more negative sentiment balance and higher irony, then the timing and format of public communication, such as the news, matter even more. Haider and Sundin (2019) emphasize that ranking and personalisation shape perceived credibility and exposure (p.11-12). In a year that tilts toward the reserved perceptual profile, interventions that would normally be sufficient – posting clarifications, linking sources, or reminding audiences to source-check new information – will predictably have less effect. Reducing interpretive burden (for example, by keeping official replies neutral in tone, and by structuring evidence in a way that is easy to audit) is not cosmetic; it lowers the ambient costs of engagement in precisely the years where those costs are highest (Haider & Sundin, 2022, p.30–31, 101–102).

The three misinformation-related indicators are equally interesting. Their base rates are low and thresholds conservative by design and are therefore read as directional rather than exhaustive. Even so, belief and accusation tend to be more visible in the same years that distrust and hostility rise. This is congruent with the accounts above: when anxiety is high and trust is low, it becomes easier to take on identity positions by embracing simplified narratives and viewpoints, and/or by delegitimising opponents and sources (Myers, 2014, p.53–60, 66; Hurley & Molloy, 2025, p.6–7). Haider and Sundin's (2022) observation that platform infrastructures prioritize engagement further suggests why these signals surface when they do: content that fuses alarm and antagonism trigger more engagement, and the platforms automatically posts that reproduce that mix (p.7-9). Taken together, the directional evidence for belief and accusation strengthens – not weakens – the broader interpretation, even if these are still too low to be seen as anything more indications.

The policy implications are fairly straightforward. If the aim is to maintain or restore information resilience in society, then times when the reserved perceptual profile is more prevalent are the times where routine measures will underperform. In such years, the priority is to lower the emotional costs of engagement and working to achieve a safe environment for information communication. This can be done by promoting clear communication, presenting evidence in ways that are easy to verify, keeping official messaging and moderation neutral, and anticipating that clarification will take longer and require more patience. Linvill et al. (2021) show that emotionally charged content has the advantage of being prioritized in algorithmic environments; the results in this study show what that advantage looks like in practice when measured through negativity, distrust, hostility, and irony (p.1–3, 16–19).

7.5 Alternative readings, generalisation, and design guardrails

At the same time, alternative readings remain possible and should be taken seriously. One alternative interpretation is compositional: shifts in who comments could produce the same visible pattern without a shift in the average person's perceptual filters. Another is platform infrastructural: changes in moderation or recommendation could select for comments that look the way the results describe, even if the underlying expression among all viewers is more mixed (Bravo-Marquez et al., 2014, p.87-88; Haider & Sundin, 2022, p.7-8, 30-31). The design of this thesis addresses these risks by focusing on year-specific movement and by using the same procedures throughout, but the risks are important to consider. The interpretation offered here – reserved vs. open perceptual profiles as compositions of indicators that align with periods of higher concern – fits the data and is grounded in prior research, but it stops short of asserting causation.

These results generalise conditionally to large, open, English-language comment spaces with similar affordances. The corpus consists of comments on news-related videos on YouTube across 2010–2024. The results therefore travel best to large, open, English-language comment spaces with similar affordances. They do not necessarily transfer to settings outside of this scope. Within that scope, the patterns reported here – particularly the co-movement of negativity, distrust, hostility, and irony in harder years – can be expected to appear elsewhere, because the underlying mechanisms are not platform-specific: anxiety about an ongoing existential threat, identity-protective reading, and an infrastructure that rewards engagement (Sarkar, 2019, p.567-568; Liu et al., 2019, p. 1-4).

8 Conclusion

The introduction described a situation in which perceptual filters and platform conditions together shape how people meet information; the analysis shows that this is what occurs in practice over a sustained period. The research questions asked how indicators move, how they align with a population benchmark, what that movement implies for information resilience, and whether the method is adequate for those claims. The study delivers qualified but consistent answers on each point. By tying the results back to the initial problem and by setting out clear next steps, the thesis communicates its core contribution: a defensible way to describe and reason about the changing conditions under which public sense-making occurs in open, platformed environments .

The study shows that perceptual filters can be examined at scale through a small, transparent set of indicators visible in public discourse, and that their year-level co-movement can be situated relative to population benchmarks without overfitting. The contribution is therefore both conceptual and practical: it reframes the problem from competencies alone (information literacy) to competencies and the perceptual filters through which people experience the world (information resilience), and it provides a workable description – reserved vs. open perceptual profiles – for communicating the state of an information environment and, in turn, how vulnerable the population being studied is.

8.1 Implications for information resilience

Within the limitations of the study discussed in previous chapters, the implications of the findings are significant. A reserved perceptual profile raises the costs of participation in discourse, narrows who is heard and/or willing to speak, and increases the chance that people misunderstand each other. In practical terms, periods when public perception is primarily characterized by the reserved perceptual profile are the times when standard approaches and routines measures – such as information literacy – are less effective than expected. In simple terms, being able to search for, retrieve, assess, and use reliable sources in a skilful way doesn't protect people, nations or societies from internal vulnerabilities such as denial induced by fear and avoidance. The fear of the trauma experienced when a person's worldview collapses, and the coping mechanisms used to manage that fear, overrides their information literacy competencies. It's not just a temporary reaction to fear-driven denial, but can become a long-term coping mechanism, and one that is even performed in groups and reinforced by that group behaviour. An open perceptual profile lowers friction and creates more paths for communication, correction, and cooperation. If information literacy is to be effective when it matters most, attention must also be given to creating an information environment that people trust and one where they feel safe to express their opinions. This will increase the information resilience of the society to which that information environment belongs.

8.2 Broader perspective and continuation

Set against the wider literature, the results support the shift from assuming that the knowledge of and access to reliable information alone will resolve problems related to vulnerability to and susceptibility to negative information, to recognising that perceptions, social context and sense of security in the information environment and discourse also need attention. Not only this, but that information specialists – especially those that already are trusted by the public to some extent, for example librarians – can play an important part in this. This is especially true when it comes to creating a safe information environment, something that librarians are already working hard to create, but that the public is often unaware of. The findings also highlight why it is vital that the field of information science recognizes there are functions in the human brain – such as the flight/fight system, tribalism, and similar subconscious survival systems – that override logical thinking. This doesn't just disable any information literacy competencies that a person has but enables the person to use their information literacy skills to further reinforce their denial and spread misinformation, as well as engage in pack behaviour where these abilities are used to reinforce the avoidance and denial of entire groups. Finally, the contribution is as much reflexive as it is empirical (Ahlryd & Hanell, 2024, p.270-271). It proposes a vocabulary that lets information science describe changes in public discourse in a straightforward way. It shows how to hold together three things at once: that worldview matters, that platform conditions matter, and that there are things within human subconscious behaviour that need to be studied within information science to provide the context for how these phenomena should be handled that only a field like that of information science can provide. This will give information specialists and the field of information science have an informed perspective about the roles of both, which they can pass on to lawmakers. Importantly, this needs to be done before a crisis situation occurs that can disable the information resilience and disrupt social cohesion within a country. On that basis, this study encourages more research into the extent of the problem, what type of countermeasures that are needed, how these are to be implemented, laws and policies that would be needed to implement them, and what the roles of different actors – such as information specialists, lawmakers, teachers, the media, libraries, and so on –

should be, and when and how these roles take effect. This research will enable society to make informed decisions, which in turn will enable people to know what to do, and have better tools for coping with stressors and pressure, during a period dominated by a reserved perceptual profile – especially during times of crisis or war.

References

Ahlryd, S. and Hanell, F. (2024). Mitigating the infodemic of the pandemic: hospital librarians' enactment and development of information resilience in healthcare organisations. *Journal of Documentation, 80(*7), 267-286. https://doi.org/10.1108/JD-12-2023-0258

Ahmad, S., Saqib, S. M., Syed, A. H., Alromema, N., & Kararay, A. (2025). Exploring the best fit: A comparative analysis of AFINN, Textblob, VADER, and Pattern on Arabic reviews for optimal dictionary extraction. *Mehran University Research Journal of Engineering and Technology*, *44*(2), 197–216. https://doi.org/10.22581/muet1982.3449

Akerlof, K., DeBono, R., Berry, P., Leiserowitz, A., Roser-Renouf, C., Clarke, K.-L., Rogaeva, A., Nisbet, M. C., Weathers, M. R., & Maibach, E. W. (2010). Public Perceptions of Climate Change as a Human Health Risk: Surveys of the United States, Canada and Malta. *International Journal of Environmental Research and Public Health*, *7*(6), 2559–2606. https://doi.org/10.3390/ijerph7062559

Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1644–1650. https://doi.org/10.18653/v1/2020.findings-emnlp.148

Bawden, D., & Robinson, L. (2022). *Introduction to information science* (2 ed.). London: Facet Publishing.

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, *69*, 86–99. https://doi.org/10.1016/j.knosys.2014.05.016

Camacho-Collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., Boisson, J., Espinosa-Anke, L., Liu, F., Martínez-Cámara, E., Medina, G., Buhrmann, T., Neves, L., &

Barbieri, F. (2022). TweetNLP: Cutting-edge natural language processing for social media. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–49). Association for Computational Linguistics. https://aclanthology.org/2022.emnlp-demos.5

Choo, C. W. (2023). Climate change information seeking. *Journal of the American Society for Information Science and Technology*, *74*(9), 1086–1099. https://doi.org/10.1002/asi.24805

Haider, J., & Sundin, O. (2019). *Invisible search and online search engines: The ubiquity of search in everyday life.* Routledge.

Haider, J., & Sundin, O. (2022). *Paradoxes of media and information literacy: The crisis of information.* Routledge.

Hodges, J. A. (2023). Taxonomizing information practices in a large conspiracy movement: Using early QAnon as a case study. *Information & Culture*, *58*(2), 129–144. https://doi.org/10.7560/IC58201

Hurley, E. A., & Molloy, M. S. (2025). How do people change their beliefs about climate change? A qualitative study on opinion shift in the U.S. Midwest. *Climatic Change*, *178*(1), Article 8. https://doi.org/10.1007/s10584-024-03835-x

Kvale, S. & Brinkmann, S. (2014). *Den kvalitativa forskningsintervjun*. (3rd [revised] ed.). Studentlitteratur.

Line, M. B. (2006). Global warming and national libraries. *Alexandria, 18*(3), iii–iv. http://proxy.lnu.se/login?url=https://www.proquest.com/scholarly-journals/global-warming-national-libraries/docview/57643335/se-2

Lindström, A. (2020, December 23). Folkhälsomyndigheten backar om biblioteken – olagligt att stänga. *Sveriges Radio*. https://www.sverigesradio.se/artikel/7633561

Linvill, D., Chambers, M., Duck, J., & Sheffield, S. (2021). Mapping the messenger: Exploring the disinformation of QAnon. *First Monday*, *26*(12). https://doi.org/10.5210/fm.v26i12.11745

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv*. https://arxiv.org/abs/1907.11692

Llewellyn, C., Cram, L., Hill, R. L., & Favero, A. (2019). For whom the bell trolls: Shifting troll behaviour in the Twitter Brexit debate. *Journal of Common Market Studies*, *57*(5), 1148–1164. https://doi.org/10.1111/jcms.12882

Luo, J., Feliciani, T., Reinhart, M., Hartstein, J., Das, V., Alabi, O., & Shankar, K. (2021). Analyzing sentiments in peer review reports: Evidence from two science funding agencies. *Quantitative Science Studies, 2*(4), 1271–1295. https://doi.org/10.1162/qss_a_00156

Merriam-Webster. (n.d.). *Lifeworld*. In *Merriam-Webster.com Dictionary*. Retrieved October 13, 2025, from https://www.merriam-webster.com/dictionary/lifeworld

Merriam-Webster. (n.d.). *Resilience*. In *Merriam-Webster.com Dictionary*. Retrieved October 13, 2025, from https://www.merriam-webster.com/dictionary/resilience

Munro, A. (2013). Discursive resilience. *M/C Journal, 16*(5). https://doi.org/10.5204/mcj.710

Myers, T. C. (2014). Understanding climate change as an existential threat: Confronting climate denial as a challenge to climate ethics. *De Ethica, 1*(1), 53–70. https://doi.org/10.3384/de-ethica.2001-8819.141153

Psychological Defence Agency. (2024, April 3). *What is psychological defence?* In *Frequently asked questions*. https://mpf.se/psychological-defence-agency/knowledge-and-support/frequently-asked-questions

Sarkar, D. (2019). *Text analytics with Python: A practitioner's guide to natural language processing* (2nd ed.). Apress. https://doi.org/10.1007/978-1-4842-4354-1

Trost, J. (2012). *Enkätboken* (4th, updated & expanded ed.). Studentlitteratur.

U.S. Department of the Treasury. (2024, September 4). *Treasury takes action as part of a U.S. government response to Russia's foreign malign influence operations* [Press release]. https://home.treasury.gov/news/press-releases/jy2559

Vaughan, K. T. L. (2001). Science and technology sources on the Internet: Global warming and climate change science. *Issues in Science and Technology Librarianship, (32).* https://journals.library.ualberta.ca/istl/index.php/istl/article/view/1874/1785

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vårheim, A. (2016). A note on resilience perspectives in public library research: Paths towards research agendas. *Proceedings from the Document Academy, 3*(2). https://doi.org/10.35492/docam/3/2/12

Wang, C., Geng, L., & Rodríguez-Casallas, J. D. (2021). How and when higher climate change risk perception promotes less climate change inaction. *Journal of Cleaner Production*, *321*, Article 128952. https://doi.org/10.1016/j.jclepro.2021.128952

Zhang, Q., Xue, H., & Tang, H. (2018). Knowledge domain and emerging trends in vulnerability assessment in the context of climate change: A bibliometric analysis (1991–2017). *Knowledge Organization, 45*(6), 467–483. https://doi.org/10.5771/0943-7444-2018-6-467

CCAM reports

2010
Leiserowitz, A., Maibach, E., Roser-Renouf, C., & Hmielowski, J. D. (2010). Climate Change in the American Mind: Americans' global warming beliefs and attitudes in June, 2010. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication.

2011
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Smith, N., & Dawson, E. (2011). Climate Change in the American Mind: Americans' global warming beliefs and attitudes in November 2011. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication.

2012
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Howe, P. (2012). Climate Change in the American Mind: Americans' global warming beliefs and attitudes in September, 2012. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication.

2013
Leiserowitz, A., Feinberg, G., Howe, P., Rosenthal, S., Roser-Renouf, C., Maibach, E., & Atkinson, A. (2013). Americans' Global Warming Beliefs and Attitudes in April 2013. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication.

2014
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., Rosenthal, S., & Marlon, J. (2014). Climate Change in the American Mind: October, 2014. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication.

2015

Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Rosenthal, S. (2015). Climate Change in the American Mind: October, 2015. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2016
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Feinberg, G., & Rosenthal, S. (2016). Climate Change in the American Mind: March, 2016. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2017
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Rosenthal, S., Cutler, M., & Kotcher, J. (2017). Climate Change in the American Mind: October 2017. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2018
Leiserowitz, A., Maibach, E., Roser-Renouf, C., Rosenthal, S., Cutler, M., & Kotcher, J. (2018). Climate Change in the American Mind: March 2018. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2019
Leiserowitz, A., Maibach, E., Kotcher, J., Bergquist, P., Ballew, M., Goldberg, M., Gustafson, A., Rosenthal, S., Wang, X., & Lacroix, K. (2019). Climate Change in the American Mind: November 2019. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2020
Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Bergquist, P., Ballew, M., Goldberg, M., Gustafson, A., & Wang, X. (2020). Climate Change in the American Mind: April 2020. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2021
Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Carman, J., Neyens, L., Marlon, J., Lacroix, K., & Goldberg, M. (2021). Climate Change in the American Mind: September 2021. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2022
Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Carman, J., Verner, M., Lee, S., Ballew, M., Uppalapati, S., Campbell, E., Myers, T., Goldberg, M., & Marlon, J. (2023). Climate Change in the American Mind: Beliefs & Attitudes, December 2022. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2023
Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Goddard, E., Carman, J., Ballew, M., Verner, M., Marlon, J., Lee, S., Myers, T., Goldberg, M., Badullovich, N., & Thier, K. (2023). Climate Change in the American Mind: Beliefs & Attitudes, Fall 2023. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

2024
Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Goddard, E., Carman, J., Ballew, M., Verner, M., Marlon, J., Lee, S., Myers, T., Goldberg, M., Badullovich, N., Thier, K., & Wang, X. (2024). Climate Change in the American Mind: Beliefs & Attitudes, Fall 2024. Yale University and George Mason University. New Haven, CT: Yale Program on Climate Change Communication.

Appendix A - Scripts and models

All code and configuration used in this study are available in the public GitHub repository:

Repository: info.resilience.twilight
URL: https://github.com/LovisaRyden/info.resilience.twilight

Contents of the repository (submission snapshot):

- Main_Scraper.py — YouTube comment scraper (news-related queries).
- Scraperforfirstyears.py — supplemental scraper used for the earliest years.
- finetune_multitask_roberta.py — training script for the multi-task RoBERTa model (7 binary indicators + 3-class sentiment).
- evaluate_ft.py — evaluation on the held-out test split; writes metrics JSON.
- merge_hybrid.py — merges predictions (irony+sentiment from the specialist model; other indicators from the base model).
- label_config.json — label set and decision thresholds used at inference.
- class_frequencies.json — label frequencies used during development/training.
- prediction_meta.json — provenance for single-model inference runs.
- prediction_meta_hybrid.json — provenance for hybrid merged predictions.
- Labelling guide.docx — operational definitions and labelling guidance.

No raw user data (comments, IDs, or CCAM tables) are distributed. Reproduction requires recollecting public data under platform terms and aligning CCAM via the public reports by year, as described in the Methods.

Release tag: v1.0.0

Archived on 26/10/2025.

Suggested citation:
Rydén, L. (2025). *Do we live in the Twilight Zone? – a study of information resilience and filters of perception in times of climate change.* (v1.0.0) [Source code]. GitHub. https://github.com/LovisaRyden/info.resilience.twilight

Appendix B – CCAM questions and results from reports

1. Question: Do you think that global warming is happening?

Results:

Year:	Yes	No	Don't know
2010	61	18	21
2011	63	17	20
2012	70	12	18
2013	63	16	20
2014	66	16	18
2015	67	16	18
2016	70	11	18
2017	71	13	15
2018	70	14	16
2019	72	12	16
2020	73	10	16
2021	76	12	11
2022	70	16	14
2023	72	15	11
2024	73	14	14

2. Questions: How sure are you that global warming is happening?
How sure are you that climate change is not happening?

Results:

Year:	Extremely sure - positive	Very sure positive	Somewhat sure positive	Not at all sure positive	Extremely sure negative	Very sure negative	Somewhat sure negative	Not at all sure negative
2010	20	37	40	3	20	31	44	4
2011	22	35	39	5	21	37	37	5
2012	27	30	40	3	15	27	45	13
2013	27	33	37	4	18	28	42	12
2014	24	35	38	4	23	30	35	11
2015	28	32	37	3	21	32	36	12
2016	30	31	34	4	20	37	34	9
2017	34	32	32	2	19	33	39	9
2018	41	29	27	3	15	39	34	12
2019	43	30	24	3	32	24	38	5
2020	44	29	24	3	17	38	36	9
2021	35	23	17	2	3	4	4	1
2022	31	20	18	2	5	5	4	1
2023	29	20	21	2	3	5	5	2
2024	31	21	19	2	3	5	5	1

3. Assuming global warming is happening, do you think it is...
* Caused mostly by human activities
* Caused mostly by natural changes in the environment
* None of the above because global warming isn’t happening
* Other

Results:

Year:	Humans responsible	Natural changes	Caused by both	GW isn't happening	Other
2010	50	34	7	6	1
2011	50	33	6	8	2
2012	54	30	6	7	2
2013	49	33	10	8	1
2014	50	32	-	9	9
2015	53	33	5	8	1
2016	53	34	5	7	1
2017	54	33	6	6	1
2018	58	28	5	7	1
2019	59	30	6	5	1
2020	62	29	5	5	1
2021	60	26	-	9	8
2022	58	23	6	8	0
2023	58	29	4	6	1
2024	60	28	-	6	6

4. Question: Which comes closest to your own view?
* Most scientists think global warming is happening
* There is a lot of disagreement among scientists about whether or not global warming is happening
* Most scientists think global warming is not happening
* Don’t know enough to say

Results:

Year:	Consensus GW is happening	Consensus GW is not happening	Lots of disagreement	Don't know enough to say
2010	34	4	45	17
2011	41	3	39	18
2012	44	3	36	18
2013	42	4	33	20
2014	44	3	31	22
2015	40	3	32	25
2016	51	3	27	18
2017	56	2	27	14
2018	57	3	25	19
2019	55	4	24	17
2020	56	3	24	16
2021	59	3	21	17
2022	58	3	20	19
2023	53	3	25	19
2024	57	3	21	17

5. Question: How worried are you about global warming?

Results:

Year:	Very worried	Somewhat worried	Not very worried	Not worried at all
2010	12	41	30	18
2011	12	42	31	15
2012	16	42	25	17
2013	14	37	30	19
2014	11	45	26	18
2015	16	41	27	16
2016	16	42	25	18
2017	22	42	23	14
2018	21	41	22	16
2019	30	36	19	15
2020	26	40	22	12
2021	35	35	18	12
2022	27	37	19	16
2023	29	36	20	15
2024	28	36	21	15

6. Question: How much do you think global warming will harm you personally?

Results:

Year:	A great deal	Moderate amount	Only a little	Not at all	Don't know
2010	10	21	27	24	18
2011	9	21	28	25	17
2012	16	26	25	20	13
2013	15	25	26	23	11
2014	12	24	26	27	11
2015	14	27	26	20	12
2016	12	28	24	23	12
2017	16	33	24	19	7
2018	14	28	27	20	11
2019	18	26	24	22	11
2020	13	29	26	21	10
2021	21	29	21	20	9
2022	16	28	23	22	12
2023	16	29	24	22	10
2024	15	31	24	21	9

7. Question: How important is the issue of global warming to you personally?

Results:

Year	Extremely important	Very important	Somewhat important	Not too important	Not important at all
2010	6	18	39	24	14
2011	6	13	41	25	15
2012	6	17	36	24	17
2013	8	13	36	24	18
2014	5	16	37	24	18
2015	7	18	37	21	17
2016	9	17	35	21	17
2017	12	19	37	19	14
2018	10	18	35	22	15
2019	18	24	26	18	16
2020	13	24	29	20	13
2021	19	24	28	15	15
2022	14	21	31	16	18
2023	14	23	30	13	20
17	17	17	17	17	17