



**HAL**  
open science

# How to protect and restore biodiversity? Exploring polarized citizens' expectations through embedding-based topic modeling with large language models

Magali Trelohan, Anne-Cécile Gay, David Moroz, Damien Chaney

## ► To cite this version:

Magali Trelohan, Anne-Cécile Gay, David Moroz, Damien Chaney. How to protect and restore biodiversity? Exploring polarized citizens' expectations through embedding-based topic modeling with large language models. *Technological Forecasting and Social Change*, 2026, 227, pp.124655. <10.1016/j.techfore.2026.124655>. <hal-05576392>

**HAL Id: hal-05576392**

**<https://hal.science/hal-05576392v1>**

Submitted on 2 Apr 2026

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



# How to protect and restore biodiversity? Exploring polarized citizens' expectations through embedding-based topic modeling with large language models

Magali Trelohan<sup>a,\*</sup>, Anne-Cécile Gay<sup>b</sup>, David Moroz<sup>a</sup>, Damien Chaney<sup>a</sup>

<sup>a</sup> EM Normandie Business School, Metis Lab, Paris Campus, 30-32 Rue Henri Barbusse, 92110, Clichy, France

<sup>b</sup> ACG Conseils, 2B impasse du Bel Air, 56860, Séné, France

## ARTICLE INFO

### Keywords:

Biodiversity  
Public consultation  
Environmental governance  
Large Language Models  
Polarization

## ABSTRACT

Despite increasing efforts to promote biodiversity protection, a persistent gap remains between policy measures and public expectations. Aligning biodiversity strategies with citizens' concerns is crucial to fostering support and ensuring effective implementation. Building on the environmental governance and ecosystem services literature, this study leverages Large Language Models for embedding-based topic modeling to analyze a large-scale public consultation involving 80,000 participants, 6000 proposals, and 1.8 million votes in France, identifying eight key biodiversity-related themes: Management of natural and urban spaces, Resource management and pollution control, Wildlife protection and species management, Environmental education and awareness, Agriculture and sustainable food practices, Environmental policy and regulatory measures, Sustainable mobility, energy, and marine conservation, and Emerging and niche concerns. Findings also reveal significant polarization in areas such as hunting regulations, taxation of polluting products, and energy transition, while topics like environmental education and urban greening enjoy strong consensus. These insights highlight the need for differentiated strategies: prioritizing immediate action in consensus areas while fostering dialogue and adaptive policymaking in highly polarized domains. Methodologically, this study demonstrates how Large Language Models can efficiently process and analyze large-scale citizen consultation datasets, offering a scalable approach to capturing public preferences and societal debates in a policy context.

## 1. Introduction

Biodiversity loss has emerged as a defining global issue, with species extinction rates accelerating to levels unprecedented in human history. According to the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, approximately 1 million species are currently at risk of extinction due to anthropogenic pressures, including habitat destruction, climate change, pollution, and overexploitation of resources (Brondízio et al., 2019). France, as a party to the Convention on Biological Diversity and a signatory of the European Union Biodiversity Strategy for 2030, has committed to ambitious conservation goals, such as protecting 30% of its terrestrial and marine areas and restoring degraded ecosystems (European Commission Directorate General for Environment, 2021). At the national level, initiatives like the French National Biodiversity Strategy (2021–2030) aim to integrate

biodiversity considerations into sectoral policies, while promoting stakeholder and citizen engagement in environmental governance (French Government - Directorate-General for Planning, Housing and Nature, 2023).

Despite these efforts, challenges remain in bridging the gap between policy objectives and public priorities. Aligning biodiversity strategies with citizens' expectations can enhance their legitimacy, effectiveness, and public support (Boiral and Heras-Saizarbitoria, 2017). Policies that reflect public concerns are more likely to be perceived as fair and representative, fostering a sense of ownership among citizens. This alignment also facilitates compliance and long-term engagement, as individuals are more inclined to support and adhere to measures they see as aligned with their values and interests (Barraclough et al., 2022; Mohri et al., 2013). However, understanding public priorities is a complex endeavor, particularly when addressing large-scale and

\* Corresponding author.

E-mail addresses: [mtrelohan@em-normandie.fr](mailto:mtrelohan@em-normandie.fr) (M. Trelohan), [annececile.gay@gmail.com](mailto:annececile.gay@gmail.com) (A.-C. Gay), [dmoroz@em-normandie.fr](mailto:dmoroz@em-normandie.fr) (D. Moroz), [dchaney@em-normandie.fr](mailto:dchaney@em-normandie.fr) (D. Chaney).

<https://doi.org/10.1016/j.techfore.2026.124655>

Received 3 April 2025; Received in revised form 4 March 2026; Accepted 14 March 2026

Available online 28 March 2026

0040-1625/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

heterogeneous populations. Public consultations (or citizen consultations), i.e. the process of gathering and integrating public feedback into decision-making by governments, organizations, or other entities (Menon and Hartz-Karp, 2019), have emerged as a key mechanism for incorporating public input into biodiversity policy-making, offering an avenue to capture diverse perspectives and identify actionable priorities (Leighninger, 2006; Zhao et al., 2022). Nevertheless, the methodological tools required to analyze such extensive and multifaceted datasets are still underdeveloped, often relying on traditional approaches that struggle to scale effectively.

Previous studies analyzing public consultation data have highlighted both the opportunities and the challenges of integrating citizen input into policy-making. For example, some work relied on case studies, qualitative coding and survey-based evaluations of deliberative processes (Bedessem et al., 2022; Leighninger, 2006; Menon and Hartz-Karp, 2019; Swette et al., 2023), which provided valuable insights but remained limited in scalability. Other recent research has applied dictionary-based or keyword-driven approaches to large-scale text corpora (Grimmer and Stewart, 2013), for instance in climate change (Fariello and Jemielniak, 2025). While these studies demonstrated the potential of computational methods to detect salient issues and citizen preferences, they also revealed significant shortcomings, such as oversimplification of nuanced arguments and difficulty in capturing polarization dynamics. Other contributions have focused on evaluating the deliberative quality of consultations (Bächtiger and Beste, 2017; Dryzek et al., 2019; Lafont, 2015), emphasizing legitimacy and inclusiveness but paying less attention to systematic large-scale textual analysis. Taken together, this literature underscores that although public consultations are increasingly used in participatory governance, methodological advances remain needed to adequately capture both the thematic complexity and the divisive nature of citizen input.

In 2022, *Make.org* conducted one of the largest public consultations ever organized in France, gathering input from over 80,000 participants on biodiversity-related issues. This consultation resulted in nearly 6000 proposals and 1.8 million votes, representing an unprecedented opportunity to explore citizen priorities, identify key concerns, and examine the divergences and polarities that characterize public debates on environmental topics. The consultation's scale and complexity highlight both its potential as a participatory governance tool and the methodological challenges involved in extracting actionable insights from such extensive datasets.

Analyzing these data requires methodologies capable of addressing both the qualitative richness of citizen proposals and the quantitative depth of voting patterns. Traditional methods, such as manual content analysis or basic keyword-based approaches, are ill-suited to capturing the nuances of citizen input at this scale. As Grimmer and Stewart (2013) notes, reliance on manual or rule-based methods often leads to oversimplified interpretations that fail to reflect the complexity of public discourse. Recent advances in Natural Language Processing (NLP) and machine learning, however, offer transformative possibilities for large-scale textual analysis, enabling researchers to extract meaningful patterns, themes, and relationships from vast datasets (Grybauskas and Cárdenas-Rubio, 2024).

This study leverages state-of-the-art artificial intelligence, computational tools, specifically Large Language Models (LLMs) and clustering algorithms, to analyze the consultation data. By employing text embeddings to represent proposals semantically, clustering techniques (K-means) to group related themes, and dimensionality reduction methods (t-SNE) to visualize the thematic structure, the research provides a robust framework for analyzing public consultations. Importantly, the integration of quantitative analyses—focused on voting patterns—adds a critical layer to the study, allowing for the identification of themes that are widely supported (consensual) or heavily debated (polarizing). This dual approach, blending textual and quantitative insights, ensures a comprehensive understanding of citizen priorities.

The originality of this work lies in its interdisciplinary methodology,

combining computational, statistical, and theoretical approaches from participatory governance. Specifically, the study pursues three key objectives:

1. To identify the key themes prioritized by citizens on biodiversity-related issues through advanced text analysis methods.
2. To uncover dynamics of consensus and polarization by integrating thematic clustering with quantitative voting data.
3. To propose a replicable framework for analyzing large-scale public consultations, contributing to methodological advancements in social science research.

To achieve these objectives, we build on the literature on environmental governance and ecosystem services to examine the role of public consultation in these processes. We then explore how advancements in methodologies using LLMs enable the analysis of such an extensive dataset, offering new perspectives on large-scale citizen engagement. Next, we detail our methodological approach and present the key findings derived from our analysis. Finally, we discuss the theoretical, methodological, and managerial contributions of our study, along with its limitations and future research directions.

## 2. Literature review

### 2.1. Biodiversity governance and polarization

Environmental governance encompasses the rules, laws, policies, institutions, structures, networks, and processes that guide the behavior of actors in the management and use of natural resources (Bennett and Satterfield, 2018). Environmental or biodiversity governance requires the collaboration of multiple stakeholders, including governments, non-governmental organizations, private entities, and local communities (Batterbury and Fernando, 2006; Newig et al., 2023, 2018; Turnhout et al., 2020). This multi-stakeholder approach has been increasingly recognized as essential for addressing the interconnected challenges of biodiversity conservation, climate change, and ecosystem degradation (Maczka et al., 2021). For instance, the governance of Natura 2000 areas in Europe demonstrates the importance of balancing conflicting interests between conservation objectives and socioeconomic activities, a process often mediated through stakeholder engagement. Moreover, decentralization and multi-level governance do not mechanically lead to improved environmental outcomes, but instead can generate ambivalent effects depending on institutional arrangements and power distribution (Cheng et al., 2021).

At the heart of these efforts lies the concept of ecosystem services—the benefits humans derive from ecosystems, including provisioning (e.g., food, water), regulating (e.g., climate, water purification), and cultural (e.g., recreational, aesthetic) services (Chen et al., 2024; Fisher et al., 2009; Haines-Young and Potschin, 2018) which appears to be an anthropocentric (Schröter et al., 2014) but still relevant approach to understand how citizens consider biodiversity (Matzek and Wilson, 2021). Citizens' perceptions of biodiversity are often shaped by their understanding of ecosystem services, influencing their priorities and support for conservation policies. The socio-cultural valuation of ecosystem services is a key determinant, and perceptions vary across demographic and geographic contexts (Tindale et al., 2023). Moreover, citizens exhibit different values regarding biodiversity, ranging from egocentric values (prioritizing personal benefits), to altruistic values (considering broader societal well-being), and biospheric values (focusing on the intrinsic value of nature and ecosystems) (De Groot and Steg, 2008). These diverse perspectives influence what individuals believe should be done to protect biodiversity but also the behaviors they adopt in response to environmental challenges. Understanding how these differences translate into public expectations and potential areas of consensus or contention is central to effective biodiversity governance.

Different individual values and perceptions of ecosystem services shape diverse views on biodiversity governance. These differences increasingly contribute to polarization in environmental issues, as contrasting priorities and interpretations solidify opposing positions in public discourse and policy debates (Casola et al., 2022). Defining polarization, however, is far from straightforward, as the term encompasses a wide range of social, political, and psychological phenomena (Jost et al., 2022). According to Jost et al. (2022), polarization involves both cognitive and motivational mechanisms that vary depending on context, making a universal definition challenging. Consequently, it manifests in three different forms with distinct implications. The first is idea-based polarization, which arises from deep ideological disagreements centered on rational debates about policies and governance strategies. The second is identity-based polarization, rooted in group affiliations where divisions are shaped more by collective identities than by reasoned policy disagreements (Bernaerts et al., 2023). The third is affective polarization, where emotional hostility toward out-groups deepens divisions and complicates consensus-building, particularly on contentious topics such as environmental policies (Hoffarth and Hodson, 2016). These dimensions often intersect, making policy debates framed in identity or affective terms more resistant to compromise, while idea-based divisions tend to be more negotiable through deliberation (Doornbosch et al., 2025).

The degree and nature of polarization vary also significantly across political and institutional contexts. Bernaerts et al. (2023) indicate that countries with consensus-oriented institutions, such as Denmark, Norway, Sweden, and Finland, tend to exhibit lower levels of both identity-based and idea-based polarization. These governance models prioritize deliberation, coalition-building, and institutional mechanisms that promote compromise. In contrast, majoritarian democracies, including the USA, Greece, France, and the UK, display higher levels of polarization, often due to adversarial political systems that reinforce binary divisions and limit opportunities for cross-cutting consensus (Bernaerts et al., 2023). These varying polarization patterns have direct implications for biodiversity governance and public expectations, as polarization shapes public attitudes and also the quality and stability of environmental decision-making processes. High levels of polarization tend to undermine deliberation, reduce the scope for compromise, and politicize environmental issues, which can weaken policy coherence and long-term commitment. In such contexts, environmental measures are more likely to face contestation, implementation delays, or policy reversals, thereby indirectly constraining environmental outcomes. Conversely, lower levels of polarization facilitate more constructive deliberation and incremental decision-making, creating institutional conditions that are more conducive to sustained environmental action (Judge et al., 2023).

The presence of polarization in environmental governance is particularly notable, as environmental policies often must balance competing priorities, such as economic growth and ecological sustainability. Although environmental polarization is well-documented, academic studies typically address it either holistically (Kennedy and Muzzerall, 2022; Özler and Obach, 2018) or through specific lenses, making a comprehensive understanding difficult to achieve. Climate change remains one of the most polarized issues due to its association with broader economic and political agendas (Guber, 2013; Hoffarth and Hodson, 2016). Similarly, the fossil fuel versus renewable energy debate highlights deep ideological and identity-based divisions, with certain communities perceiving the energy transition as a direct threat to their livelihoods (Hoffarth and Hodson, 2016). Other contentious topics, including biodiversity conservation (Lucas and Warman, 2018), deforestation (Picanço Rodrigues and Leonel Caetano, 2023), water resource management (Gibson et al., 2021), and sustainable agriculture (Boer and Aiking, 2023), further demonstrate how environmental issues often become focal points for polarization, each reflecting underlying economic, cultural, or ideological tensions. This fragmented approach makes it challenging to develop a unified framework that captures the

complexity of environmental polarization, where emotional, ideological, and identity-driven factors intersect.

However, while stakeholder engagement is essential at multiple levels, one of the most direct mechanisms for capturing citizen priorities is through structured participatory processes such as public consultations (Kiss et al., 2022). These initiatives provide a critical avenue for integrating societal values into biodiversity governance, helping to shape policies that are both effective and socially legitimate. The next section explores the role of public consultations in fostering participatory governance and aligning biodiversity strategies with public expectations.

## 2.2. Public consultation in participatory governance

Public consultations have increasingly emerged as pivotal mechanisms for fostering participatory governance (Zhao et al., 2022), particularly in addressing intricate societal challenges such as biodiversity loss. By enabling the engagement of diverse stakeholders, these processes enhance the legitimacy, inclusiveness, and responsiveness of policy decisions (Bedessem et al., 2022). This participatory dynamic bridges the gap between citizens and policymakers, fostering a shared sense of responsibility and a deeper public understanding of environmental priorities (Binley et al., 2021). Such engagement is particularly significant in biodiversity governance, where citizen input can shape policies that better align with societal values and ecological imperatives.

While participatory governance aims to integrate diverse perspectives, the degree to which citizen engagement contributes to effective decision-making depends on the capacity of these processes to manage both consensus-building and polarization. Public consultations serve as platforms where contrasting priorities and ideological orientations interact, sometimes fostering deliberation and compromise, but at other times exacerbating divisions (Swette et al., 2023). Polarization in participatory governance can manifest in various ways: ideological polarization, where disagreements stem from conflicting policy visions; identity-based polarization, where group affiliations shape attitudes toward biodiversity policies; and affective polarization, where emotional antagonism impedes constructive engagement (Bernaerts et al., 2023). The presence of these dynamics necessitates the careful design of consultation mechanisms that can mediate tensions while preserving meaningful dialogue.

When citizens perceive their involvement in governance as meaningful, they are more likely to support and engage with environmental initiatives (Kitamori, 2025). Participatory approaches heighten public support for biodiversity-related policies, especially when citizens believe their contributions are valued and effectively reflected in decision-making processes (Itten and Mouter, 2022). However, ensuring that consultations translate into tangible outcomes is critical for sustaining engagement. Deliberative democratic theory suggests that citizens are more willing to accept policy decisions—even those they disagree with—when they view the decision-making process as fair and inclusive (Dryzek et al., 2019). This underscores the importance of procedural legitimacy in participatory governance.

Despite their potential, large-scale public consultations present methodological challenges. Managing the sheer volume and diversity of citizen input requires methodologies that can scale efficiently while capturing the nuanced priorities of participants. The increasing reliance on digital consultation platforms has enabled broader participation but has also introduced concerns about representativity, as digitally mediated engagement may not capture all demographic groups equally (Townley and Koop, 2024). Furthermore, large-scale participatory processes must account for differences in expertise, interest, and motivation among participants, which can shape the depth and quality of contributions (Buntaine et al., 2021).

Beyond simply collecting citizen perspectives, policymakers and researchers must assess the degree of alignment or divergence in public opinion to inform more adaptive and inclusive decision-making. This

requires distinguishing between consensual themes, where there is broad agreement among participants, and polarizing issues, where public opinion is fragmented. Consensus-oriented discussions typically focus on widely accepted principles, such as the intrinsic value of biodiversity or the importance of nature conservation. In contrast, polarization often emerges when biodiversity policies intersect with competing economic, cultural, or political interests (Judge et al., 2023).

In highly polarized contexts, participatory governance faces additional challenges. Stakeholder engagement can become adversarial, reflecting entrenched divisions rather than fostering meaningful deliberation. In such cases, conventional consultation methods may be insufficient to bridge divides, necessitating deliberative mechanisms that prioritize trust-building and conflict mediation (Kiss et al., 2022). Research in democratic innovation suggests that structured deliberation—where citizens engage in facilitated discussions based on shared evidence—can mitigate the effects of polarization by promoting mutual understanding and constructive dialogue (Bächtiger and Beste, 2017; Caluwaerts et al., 2023).

The interplay between consensus and polarization in participatory biodiversity governance has direct implications for policy design. Understanding where public agreement exists allows policymakers to reinforce widely supported initiatives, while recognizing contentious issues enables them to design targeted interventions that address key divisions.

### 2.3. Advancing methodologies with LLMs for public consultation analysis

NLP techniques, including machine learning-based text analysis, have proven invaluable for extracting insights from large-scale textual datasets, such as public consultations and citizen engagement platforms. While NLP has been extensively applied in fields such as policy analysis (Grimmer and Stewart, 2013) and environmental discourse mining (Fariello and Jemielniak, 2025; Larosa et al., 2025), its use in participatory biodiversity governance remains largely unexplored. This presents both a methodological challenge and an opportunity to refine how citizen priorities are analyzed and incorporated into policy frameworks.

Recent advancements in text-mining methodologies have demonstrated their effectiveness in structuring complex qualitative data across diverse domains. For instance, in labor market analytics, text-mining techniques have been successfully employed to classify job advertisements and extract labor market trends (Grybauskas and Cárdenas-Rubio, 2024; Pejic-Bach et al., 2020). Similarly, in environmental research, machine learning models have been used to identify thematic patterns within large-scale text corpora (Fariello and Jemielniak, 2025; Larosa et al., 2025). However, despite these successful applications, the potential of LLMs to process and structure citizen input in biodiversity consultations remains underexplored. Unlike traditional NLP approaches, LLMs offer the ability to capture deeper semantic relationships and contextual nuances, making them particularly suited for analyzing open-ended participatory governance datasets.

One of the main limitations of previous text-mining approaches has been their reliance on predefined dictionaries, rule-based NLP models, or traditional machine learning classifiers, which struggle to capture the depth and nuance of human language in large, open-ended citizen datasets (Qaiser and Ali, 2018). These techniques often require extensive manual preprocessing and do not always generalize well across different consultation topics.

The emergence of distributed embeddings such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) represented a major advance over earlier symbolic approaches. In these models, words are encoded as dense vectors in which semantic information is distributed across multiple dimensions, and their training relies on co-occurrence patterns within a local context window. However, once trained, each word is still associated with a single static vector, regardless of the sentence in which it appears. This limitation prevents such models from fully capturing polysemy and context-dependent

meaning, which are especially prevalent in open-ended citizen discourse. In addition, the inherent biases of such lexical embedding models, along with their restricted contextual window, limit both the interpretability and generalizability of results—an issue highlighted in previous studies on text classification and sentiment analysis in participatory governance (Brown et al., 2020).

Contextual embeddings such as embeddings from language model (ELMo) (Peters et al., 2018) represented a major shift by generating word representations that vary according to their surrounding sentence, rather than assigning a single static vector. Built on bidirectional Long Short-Term Memory (LSTMs) trained as language models, ELMo was able to capture polysemy and context-dependent meaning more effectively. Nevertheless, the sequential architecture limited scalability and flexibility compared to later approaches. While earlier models like Word2Vec and ELMo attempted to leverage local or sequential context, these efforts remained constrained in scope and did not fully resolve the challenges of long-range dependencies.

The introduction of Transformer architectures (Vaswani et al., 2023) marked a decisive methodological leap, enabling models to capture long-range dependencies efficiently through attention mechanisms. This innovation facilitated the development of models such as Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019), which leveraged bidirectional attention to generate contextualized embeddings at scale, bridging the gap between sequential embeddings and the large-scale contextual understanding now achieved by LLMs. LLMs extend these capabilities further, achieving state-of-the-art performance in representing meaning, handling long contexts, and performing knowledge-intensive tasks (Bender et al., 2021; Brown et al., 2020). These methodological advances have important implications for the analysis of citizen discourse in policy contexts, allowing for more nuanced and scalable text-mining approaches in participatory governance.

Their ability to process large volumes of unstructured text enables a more comprehensive and dynamic identification of themes, consensus patterns, and polarizing issues within public consultations. Furthermore, LLMs have been successfully deployed in policy-relevant domains such as climate discourse analysis (Fariello and Jemielniak, 2025) or tested to propose climate policies (Larosa et al., 2025) and legal text interpretation (Ariai and Demartini, 2024; Greco and Tagarelli, 2024), suggesting their potential applicability in biodiversity governance.

However, we need to acknowledge the challenges associated with LLM-based analyses. Although the underlying architectures and algorithms of LLMs are well documented, their outputs rely on high-dimensional embedding spaces that lack direct interpretability for human observers. This intrinsic interpretative opacity poses concerns in policy contexts where outputs must be interpretable by domain experts and decision-makers, making it difficult to trace the reasoning behind generated outputs (Rudin, 2019). In our approach, we do not rely on the direct generation of outputs from the LLM. While embeddings themselves are derived from the LLM and inherit its complexity, our approach leverages them as intermediate representations, which are then analyzed statistically and interpreted qualitatively to provide actionable insights for experts and decision-makers. Specifically, we examine discriminative words and representative phrases near cluster centroids, which provides insight into the reasoning behind the results and ensures that the findings are comprehensible and actionable for experts and policymakers. Additionally, potential biases in training data could influence thematic identification, particularly when analyzing politically charged or value-laden environmental discussions (Weidinger et al., 2021). Addressing these challenges requires a careful methodological approach, combining quantitative clustering techniques (e.g., K-means) with qualitative validation to ensure the robustness of results.

Despite these challenges, the integration of LLMs into the analysis of biodiversity consultations represents an innovative step forward, offering a scalable and nuanced approach to understanding public priorities. Given the growing emphasis on data-driven policy design, exploring the

potential of LLMs in participatory governance is not only methodologically relevant but also essential for ensuring that citizen engagement efforts translate into actionable and well-informed biodiversity policies.

### 3. Methodology

#### 3.1. Data

France offers a relevant context for examining citizens' expectations regarding biodiversity protection, given its majoritarian political system that fosters higher levels of polarization (Bernaerts et al., 2023). This polarization challenges consensus-building on environmental issues, making France an ideal case to explore how citizen expectations can be managed in divided societies.

The data for this study come from a large-scale public consultation conducted in France in 2022. The consultation engaged 80,716 participants, generating 5554 proposals and 1,701,343 votes. Citizens were invited to participate in two complementary ways: (i) by submitting proposals in open-text format to suggest specific actions, measures, or ideas for biodiversity protection, and (ii) by evaluating others' proposals through a trichotomous voting system ("pro", "against", or "neutral"). This dual mechanism ensured that participants could both create content and position themselves toward the contributions of others. While no additional comment threads were collected beyond the proposals themselves, the open-text format allowed participants to express their opinions in their own words and provide nuanced input.

The data analyzed in this study were collected and pre-processed by [Make.org](https://make.org), a well-established platform specializing in large-scale public consultations. [Make.org](https://make.org) implements a multi-layer moderation and verification process to ensure the veracity and usability of the contributions. This process includes participant authentication mechanisms, automated detection of bots and duplicate submissions, and human moderation to remove off-topic, abusive, or non-compliant content, in accordance with the consultation's rules and objectives. Only validated proposals that passed these moderation procedures were included in the dataset made available for analysis. As a result, the corpus analyzed in this study consists of genuine human contributions that are directly relevant to the biodiversity consultation, providing a reliable basis for semantic analysis and polarization measurement.<sup>1</sup>

Key demographic variables were partially reported, with age information provided by 83% of participants and departmental location by 45.2%. The average age of participants was 49 years.

#### 3.2. Topic model process

A topic modeling approach combining advanced text representation techniques and clustering methods is implemented (Keraghel et al., 2024). The process follows these steps:

- **Embedding computation:** Transformation of textual data into vector representations using embeddings generated by a LLM.
- **Dimension reduction:** Application of Principal Component Analysis (PCA) and comparison with a no-reduction approach to assess its relevance.
- **Clustering:** Grouping of proposals into thematic clusters using the K-Means algorithm.
- **Cluster interpretation:** Identification of underlying themes through several complementary analyses:
  - o Projection of embeddings into 2D space using t-SNE to visualize cluster structures,
  - o Extraction of the most discriminative words per cluster,
  - o Analysis of proposals closest to the cluster centroids to refine the interpretation of emerging themes.

These four steps are presented in Fig. 1 and described in detail in the following sections (step 1 to 3 in the methodology sections following and the interpretation of clusters in the results section).

##### 3.2.1. Embedding computation

The textual data were transformed into numerical vectors known as embeddings (Mikolov et al., 2013). Embeddings are multidimensional vectors that encapsulate the semantic and contextual relationships of texts. Unlike traditional methods such as TF-IDF (Term Frequency-Inverse Document Frequency) (Bafna et al., 2016), which assign scores based on term frequency without considering context, embeddings capture the underlying meanings of words and phrases by accounting for the context in which words appear. These representations enable a richer and more nuanced modeling of relationships between texts.

An embedding can be formally defined as a function  $f : T \rightarrow R^d$  where  $T$  is the set of citizen proposals, and  $R^d$  is a  $d$ -dimensional vector space. Each text  $t \in T$  is projected into this vector space as a numerical vector  $f(t)$ .

In this  $R^d$  space, the proximity between two vectors reflects the semantic similarity between the corresponding texts (Fig. 2). Conversely, a large distance between two vectors indicates significant semantic differences. This vector projection reduces the complexity of textual data and makes it directly usable by clustering algorithms, enabling the identification of underlying themes within the set of citizen proposals.

In this study, embeddings were generated using OpenAI's text-embedding-3-large model (OpenAI, 2025), based on a LLM (Keraghel et al., 2024; Petukhova et al., 2025). These large-scale models are trained on diverse corpora spanning multiple domains, allowing them to learn complex linguistic structures and produce rich textual representations (Grybauskas and Cárdenas-Rubio, 2024). The text-embedding-3-large model projects each text into a 3072-dimensional vector space, where each vector encodes semantic and contextual relationships. Similar techniques for embedding computation and aggregation have been described in Gao et al. (2022) and Cer et al. (2018).

This model was selected for its superior performance in capturing semantic nuances compared to earlier generations, as demonstrated by recent results from the Massive Text Embedding Benchmark (MTEB: Muennighoff et al., 2023). In addition, the model relies on Matryoshka

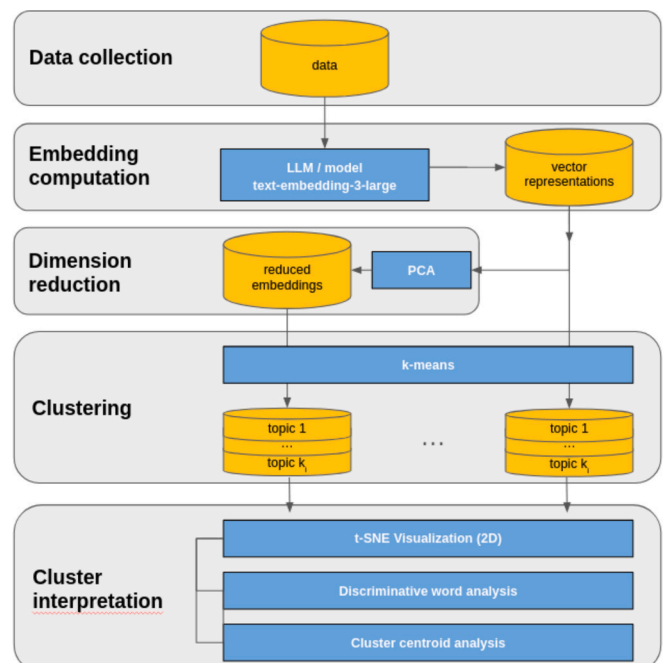


Fig. 1. Topic modeling pipeline based on embeddings generated by an LLM.

<sup>1</sup> <https://make.org/FR/moderation>.

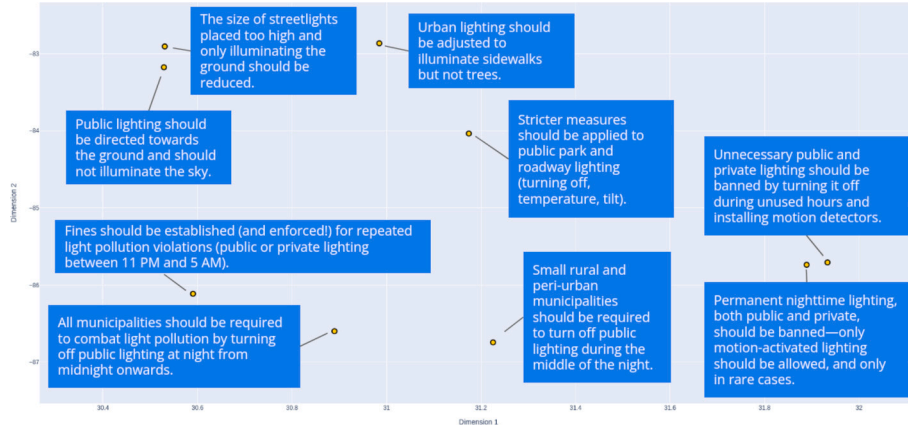


Fig. 2. 2D T-SNE projection (perplexity = 15): zoom on the structure of embeddings and their spatial-semantic proximity.

Representation Learning (Kusupati et al., 2024), which ensures that semantic information remains robust and dense even in the first embedding dimensions. This property is critical for the stability of our pipeline combining PCA and K-means.

We did not rely on predefined search terms or dictionaries. The approach adopted is fully unsupervised and bottom-up, as the entire set of 5554 proposals was analyzed in order to allow thematic structures to emerge organically from the data. The raw text provided by the consultation platform was processed through a Python-based pipeline to generate embeddings. No automated topic extraction or pre-classification engine was used at any stage, ensuring that the results were not biased by any pre-existing thematic framework.

Although OpenAI has not publicly disclosed the precise architecture of the text-embedding-3-large model, its embedding computation can be described following the general principles of Transformer-based models (Vaswani et al., 2023), which is at the core of modern LLMs. The architecture is composed of multiple layers of self-attention and feed-forward networks. The embedding computation can be summarized in the following steps:

(a) **Tokenization and positional encoding.**

The input text is first divided into subword units called tokens. Each token is mapped to an initial embedding vector  $e_i$ . Positional encodings  $P_{ii}$  are then added to preserve the sequential order:

$$z_i^{(0)} = e_i + p_i$$

(b) **Multi-head self-attention layers.**

Each transformer layer computes contextualized representations by relating each token to all others in the sequence. The attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where  $Q$ ,  $K$ , and  $V$  are linear projections of the input vectors, and  $\sqrt{d_k}$  is a scaling factor. Multi-head attention applies this operation in parallel across several heads, allowing the model to capture multiple types of contextual dependencies simultaneously.

(c) **Feed-forward and residual connections.**

The outputs of attention layers are passed through position-wise feed-forward networks, with residual connections and normalization ensuring stability during training. After  $L$  layers, each token  $t_i$  is associated with a contextualized vector representation  $\mathbf{h}_i^{(L)}$ .

(d) **Sentence-level embedding aggregation.**

To obtain a single embedding for the entire text, the model combines the vector representations of all tokens into one overall representation. Although OpenAI has not disclosed the exact aggregation method, a common approach in comparable models (e.g., SBERT: Gao et al., 2022; Reimers and Gurevych, 2019) is mean pooling, which averages the vectors of all tokens:

$$f(t) = \frac{1}{n} \sum_{i=1}^n h_i^{(L)}$$

This produces a single dense vector  $f(t)$  that summarizes the semantic content of the entire sequence. Fig. 3 illustrates the process by which embeddings are generated.

This ability to capture the richness of natural language makes these embeddings particularly suitable for advanced analyses such as topic modeling. The choice of OpenAI's GPT-based embeddings is motivated by their recognized performance in NLP tasks and their strong results in text clustering across a wide range of datasets and evaluation metrics (Harris et al., 2024; Petukhova et al., 2025). While no direct comparative experiments were conducted in this study, the decision to prioritize OpenAI's model was informed by consistent evidence in the literature of its superior performance across clustering and semantic similarity tasks (Petukhova et al., 2025). Nevertheless, BERT would also have constituted a solid alternative. While generally outperformed by OpenAI's proprietary models, BERT has long stood out as the strongest performer among open-source and lighter alternatives, and it remains a competitive choice in resource-constrained settings (Petukhova et al., 2025).

To link the abstract vector representations produced by embeddings to interpretable linguistic concepts, we implemented a deterministic lexical specificity analysis procedure. Rather than relying on standard TF-IDF, we used a relative frequency ratio, computed as follows:

$$R(w, c) = \frac{f_{w,c}/N_c}{f_{w,g}/N_g}$$

where  $f_{w,c}$  denotes the frequency of word  $w$  in cluster  $c$ ,  $N_c$  the total number of words in cluster  $c$ , and  $g$  the frequency of word  $w$  in the overall corpus.

A word  $w$  is considered discriminant for a cluster  $c$  if it satisfies two conditions:

1. A minimum occurrence frequency ( $f_{w,c} \geq 15$ ) to ensure representativeness;
2. A salience ratio  $R > 2$ , defined as the ratio between the word's relative frequency in the cluster and its relative frequency in the full corpus.

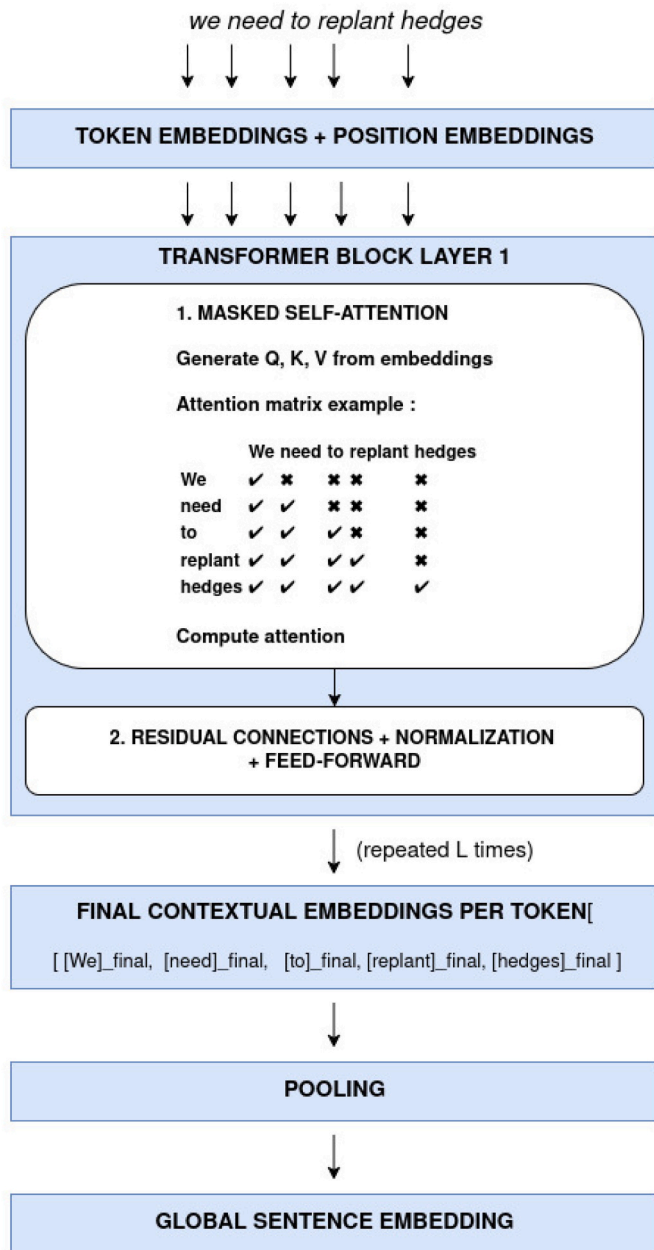


Fig. 3. Computation of contextual embeddings in a GPT-style Transformer block. Each token embedding passes through masked self-attention, followed by residual connections, normalization, and feed-forward layers.

A threshold value of 2—indicating that the word is at least twice as frequent within the theme as elsewhere in the corpus—was retained for its high level of managerial and policy interpretability. Unlike TF-IDF-based approaches, which may highlight rare or weakly informative terms, this method identifies the lexical pillars of the debate. The coherence of these lexical signatures was validated through a qualitative examination of the contributions closest to cluster centroids (see Appendix 1.1), ensuring a rigorous correspondence between vector geometry and the substantive meaning of citizen statements.

### 3.2.2. Dimension reduction

The high-dimensional vector space generated by embeddings offers substantial informational richness, but its high dimensionality can pose challenges for the efficiency of clustering algorithms. However, dimensionality reduction must be handled with caution, as its efficiency

can vary depending on the specific dataset (Petukhova et al., 2025). To simplify the vector space, we tested two approaches: with and without dimensionality reduction. Dimensionality reduction was performed using PCA, allowing the embeddings to be projected into a lower-dimensional subspace while preserving semantic relationships between texts as much as possible. The number of retained components was determined by analyzing the proportion of explained variance relative to the number of principal components. By selecting a cumulative variance threshold of 90%, we retained an informative subspace with 558 dimensions.

A comparison of results with and without PCA revealed better interpretability of clusters when using dimensionality reduction. PCA facilitated the identification of coherent thematic structures, leading us to adopt this approach for the final analysis.

### 3.2.3. Clustering

Based on Keraghel et al. (2024), K-Means algorithm is used to group the textual data into thematic clusters based on the embeddings generated by LLMs.

The selection of the optimal number of clusters ( $k$ ) is a critical step in unsupervised clustering, especially when dealing with textual data. Unlike traditional numerical data, textual data projected into a multi-dimensional vector space (via embeddings) do not always exhibit clear boundaries between groups. This complexity makes determining  $k$  more challenging and requires a combination of quantitative and qualitative approaches.

Thus, we adopted a combined approach: analyzing silhouette scores to obtain an initial objective estimate and applying human judgment to the generated clusters to ensure their thematic coherence and relevance, as statistical coherence alone does not guarantee semantic interpretability (Eklund and Forsman, 2022; Jacobi et al., 2016; Lau et al., 2014).

### 3.2.4. Clustering with K-Means++ initialization

We used the K-Means++ algorithm (Arthur and Vassilvitskii, 2006; Petukhova et al., 2025) to group the textual data into thematic clusters based on the generated embeddings. Unlike the classic K-Means initialization, where initial centroids are chosen randomly, the K-Means++ approach improves clustering quality by selecting initial centroids in an optimized manner. This method aims to maximize the distance between initial centroids, reducing the risk of convergence to local minima and enhancing the stability of the results.

The K-Means++ algorithm initializes the centroids through the following steps:

- The first centroid  $C_1$  is randomly selected from the data points  $x_1, x_2, \dots, x_n$ .
- For each data point  $x_i$ , the minimum distance to the nearest already selected centroid is calculated as:

$$D(x_i) = \min_{c \in \{C_1, \dots, C_k\}} |x_i - c|^2$$

where  $|x_i - c|$  the distance between  $x_i$  and centroid  $c$ .

- The centroid  $c_{j+1}$  is selected from the data points with a probability proportional to  $D(x_i)$ :

$$P(x_i) = \frac{D(x_i)}{\sum_{j=1}^n D(x_j)}$$

This step favors the selection of points that are distant from the existing centroids, ensuring better initial dispersion. The second and third steps are repeated until all  $k$  centroids are selected.

Once the centroids are initialized, the standard K-Means algorithm takes over, partitioning the data into  $k$  clusters by minimizing intra-cluster variance. At each iteration, data points are assigned to the nearest centroid, and the centroid positions are recalculated as the mean of the points assigned to them:

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in c_j} x_i$$

where  $C_j$  the centroid of the cluster  $j$  and  $|C_j|$  the number of points in the cluster.

### 3.2.5. Silhouette coefficient: a quantitative measure

The silhouette score (Rousseeuw, 1987) is a widely used metric for evaluating clustering quality. It measures intra-cluster cohesion and inter-cluster separation, providing an indication of how well data points are grouped. This metric, defined for each point  $i$ , is calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where:

- $a(i)$  is the average distance between point  $i$  and the other points within its own cluster (intra-cluster cohesion);
- $b(i)$  is the average distance between point  $i$  and the points in the nearest neighboring cluster (inter-cluster separation).

The overall silhouette score is obtained by averaging the individual scores. It ranges from  $-1$  to  $1$ :

- $s(i) \simeq 1$ : points are well clustered and clearly separated from other clusters;
- $s(i) \simeq 0$ : points are close to the boundary between two clusters;
- $s(i) < 0$ : points are poorly assigned, being closer to a different cluster than to their own.

As indicated in Fig. 4, the Silhouette Score (SS) varies across datasets and embedding-algorithm combinations, ranging from  $-0.002$  to  $0.170$  in recent comparisons (Petukhova et al., 2025). The analysis of the silhouette score for different configurations showed that the PCA-based approach provides better cluster separation and a more coherent thematic structure. Additionally, our silhouette scores fall within standard benchmark ranges from Petukhova et al. (2025), indicating satisfactory clustering performance. Based on this analysis, we selected the model with dimensionality reduction via PCA and  $k=32$  clusters for the final analysis. Its SS is  $0.048$ , which is close to the median score of  $0.056$  reported in the study by Petukhova et al. (2025), indicating comparable clustering performance. Importantly, in the context of LLM-based semantic embeddings and public consultation data, moderate silhouette values are expected and reflect semantic continuity rather than poor

topic distinctiveness (Ataman et al., 2025). Citizen contributions frequently link multiple issues, resulting in overlapping but interpretable thematic structures (Ryfe, 2005). Consequently, the silhouette score should be interpreted as an indicator of discourse interconnectedness rather than as evidence of algorithmic confusion between topics.

### 3.2.6. Selected number of clusters

Textual data, although transformed into numerical vectors, retain a complex semantic nature. Unlike traditional data (such as numerical data measured on well-defined variables), textual clusters do not always correspond to clearly distinct groups in the vector space. This ambiguity may arise from the semantic richness of the texts or the fluid nature of the themes discussed. Thus, even when the Silhouette Score provides an objective indication of the data structure, it may not perfectly reflect their thematic interpretation. For this reason, we complemented this analysis with an in-depth human review of the generated clusters, examining their textual content to validate their coherence and relevance. The qualitative analysis confirmed that the configuration corresponding to the peak Silhouette Score offered a relevant and interpretable segmentation, thereby reinforcing our methodological choice.

By applying this approach, we initially estimated the optimal number of clusters at  $k = 32$  based on the Silhouette Score. This value was confirmed through a qualitative analysis of the clusters. This combined approach resulted in groupings that were both mathematically coherent and interpretable, aligning with the objectives of the study.

Thirty-two distinct themes may seem like a high number of clusters. However, this is not unusual in topic modeling when dealing with diverse textual data. For example, the 20Newsgroups dataset is typically divided into 20 classes, and the Reuters dataset into 65 classes. Similarly, the MN-DS dataset includes up to 109 classes at level 2 (Petukhova et al., 2025). In comparison, our analysis achieved a peak silhouette score of  $0.048$  at thirty-two clusters, which is consistent with the thematic complexity of our dataset.

To assess the robustness of the thematic structure, we subjected the corpus to a control analysis using manifold learning techniques (UMAP) combined with a density-based clustering algorithm (HDBSCAN). A systematic grid search identified an optimal configuration (50 dimensions,  $\text{min\_cluster\_size} = 25$ ,  $\text{min\_samples} = 5$ ), associated with a DBCV validity index of  $0.234$ . The DBCV (Density-Based Clustering Validation) index assesses within-cluster compactness and between-cluster separation while accounting for density structure, and is a standard metric for evaluating partitions produced by HDBSCAN (Moulavi et al., 2014).

This approach confirms the existence of dense semantic cores but also results in the exclusion of 20.9% of the corpus, which the algorithm

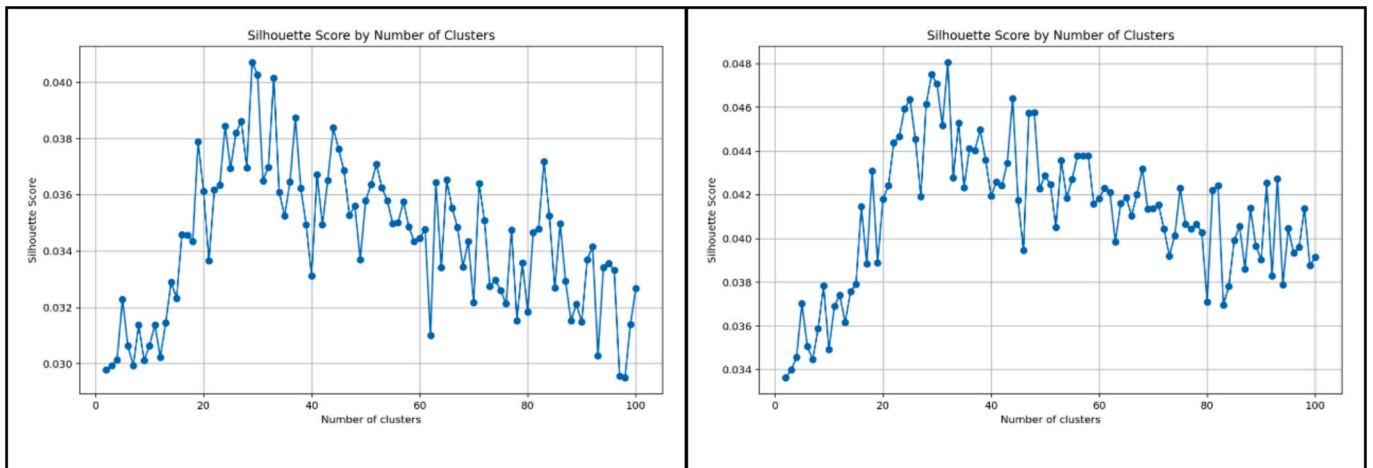


Fig. 4. Silhouette score as a function of the number of clusters - full embeddings (on the left) and embeddings after dimensionality reduction by PCA (on the right).

classifies as “noise”. This level of exclusion is relatively low compared to recent studies on large-scale text corpora, where density-based clustering methods often classify up to 50% of observations as noise (Zhao and Syed, 2026). A qualitative audit of these excluded documents shows that they correspond to valid but singular citizen voices, rather than irrelevant or aberrant data. As noted by Chang et al. (2025), non-linear dimensionality reduction methods such as UMAP favor a local interpretation of data structure, whereas linear approaches such as PCA provide a global representation, which is more appropriate when the scientific objective is to capture the overall organization of a corpus.

In light of this distinction, and given our objective of exhaustiveness and representativeness of citizen voices, we chose to retain the PCA + K-means framework for the final analysis. This deterministic linear approach ensures the inclusion of 100% of textual contributions and provides the stability and reproducibility required for a thematic analysis aimed at global interpretation rather than the sole identification of locally dense semantic cores.

### 3.3. Polarization analysis

#### 3.3.1. Data description

The second step of our study is to analyze the polarization of the participants, based on their votes for each proposition. Three voting options were provided: “in favor”, “against” or “neutral”.

Each proposal received an average of 306.3 votes (standard deviation: 100), with a minimum of 15 votes and a maximum of 4586. The first quartile of proposals received at least 294 votes, the median 296 votes, and the top quartile over 301 votes. Voting preferences were predominantly positive, with a median of 71.1% “in favor” votes per proposal. The median share of “against” votes was 11.5%, and “neutral” votes accounted for 15.3%. The distribution shows significant variability: “in favor” votes ranged from 4.5% to 96.5%, “against” votes from 0.4% to 87.0%, and “neutral” votes from 1.1% to 83.8%.

#### 3.3.2. Polarization measurement method

In polarization studies, existing indices have primarily been developed for continuous or ordinal variables (Adams et al., 2022; Bramson et al., 2017; Dalton, 2008; Esteban and Ray, 1994; Van der Eijk, 2001). These approaches rely on measuring the dispersion of values along a continuum, which makes them unsuitable for our categorical voting data, where individuals can only choose between “Pro” (in favor), “Against” (against), or “Neutral” responses. Standard polarization measures fail to account for the discrete nature of the data and the role of neutrality in mitigating polarization.

To address this limitation, we construct a Binary Polarization Index (BPI) adapted to categorical data. After the first step of topic modeling, our dataset comprises votes on biodiversity-related proposals, which were grouped into 32 clusters based on thematic similarity. The goal is to identify areas of strong consensus and polarization by analyzing the distribution of votes within each cluster.

The BPI is designed to be maximized when votes are evenly split between “Pro” and “Against”, reflecting a scenario of strong division, and minimized when there is either a clear consensus or a high proportion of neutral votes. The index is defined as:

$$BPI = 4 \times \left( \frac{Votes_{pro}}{Votes_{active}} \times \frac{Votes_{against}}{Votes_{active}} \right) \times \left( 1 - \frac{Votes_{neutral}}{Votes_{total}} \right)$$

where:

- $Votes_{pro}$  and  $Votes_{against}$  represent the number of affirmative and opposing votes.
- $Votes_{active} = Votes_{pro} + Votes_{against}$  accounts for engaged votes (excluding neutral votes).
- $Votes_{neutral}$  represents votes with no stance, which reduces polarization.

- $Votes_{total}$  is the total number of votes in the cluster.

The BPI is based on the normalized product of the proportions of “Pro” and “Against” votes among active participants (i.e., excluding neutral votes). This structure captures the idea that polarization is strongest when two opposing camps are equally represented. Mathematically, the product:

$$\left( \frac{Votes_{pro}}{Votes_{active}} \times \frac{Votes_{against}}{Votes_{active}} \right)$$

reaches its maximum value of 0.25 when both proportions equal 0.5—i.e., when the vote is evenly split. To ensure that the index ranges from 0 (no polarization) to 1 (maximum polarization), this product is multiplied by a factor of 4. This normalization facilitates interpretability and comparability across clusters.

We further introduce a dampening factor:

$$\left( 1 - \frac{Votes_{neutral}}{Votes_{total}} \right)$$

to account for disengagement. A high proportion of neutral votes indicates low participation in the debate and thus reduces the meaningfulness of any observed division. This adjustment ensures that polarization scores reflect not only antagonism between active opinions but also the level of engagement. It prevents overestimation of polarization in clusters where a significant share of participants remains neutral or disengaged.

This formulation ensures that the BPI reaches its maximum (1) when votes are evenly split between “Pro” and “Against” (i.e., 50%–50%), indicating a fully divided cluster. The index decreases when one side dominates (e.g., 90%–10%), and is further dampened by neutrality, as a high proportion of neutral votes reduces the intensity of division.

While conceptually simple, this construction aligns with prior work using the product of proportions to capture maximal opposition in binary distributions (Rehm and Reilly, 2010). It also resonates with the broader class of polarization measures developed in economics (Duclos et al., 2004; Esteban and Ray, 1994), which emphasize the co-presence of group size parity and attitudinal distance. However, unlike these indices, which are tailored to continuous or ordinal data, the BPI is specifically adapted to trichotomous categorical data with a neutral middle category. The dampening adjustment for neutrality introduces an engagement-sensitive component that enhances the robustness of the measure in participatory settings.

## 4. Results

### 4.1. Cluster interpretation

In this section, we focus on the interpretation of the 32 clusters generated by the topic model process. The objective is to better understand the distinctive characteristics of each cluster by examining the most frequent words, the most discriminative words, and the representative propositions closest to the centroids of the clusters. These elements help identify the dominant themes and better define the specificities of the text groups.

To achieve this, we first conducted a text preprocessing phase, including punctuation and accent removal, conversion to lowercase, exclusion of low-information words (“stop-words”), and lemmatization. Finally, we identified representative propositions for each cluster by extracting the sentences that are closest to the cluster centers in the vector space (see Appendix 1). This approach provides a more concrete overview of the dominant ideas within each group. Table 1 below presents the defined titles for the thirty-two clusters.

These thematic clusters highlight the large diversity in the way citizens approach the topic of biodiversity. The identified clusters, ranging from “Agricultural and Vegetal Planning” to “Limiting Human

**Table 1**  
Thematic clusters.

Cluster	Name
1	Agricultural and Vegetal Planning
2	Urban Gardens and Green Spaces
3	Waste and Plastic Reduction
4	Hunting Regulations and Game Protection
5	Protection of Natural Areas and Wildlife
6	Water Management and Quality
7	Wetland Preservation
8	Animal Protection and Livestock Farming
9	Urbanization and Land Artificialization
10	Urban Greening and Green Spaces
11	Waste Management and Pollution
12	Pesticide Ban
13	Species and Predator Protection
14	Biodiversity Awareness
15	Organic and Sustainable Agriculture
16	Sustainable Transport and Mobility
17	Solar and Renewable Energy
18	Biodiversity Protection and Environmental Actions
19	Human Activity Impact on the Environment
20	Sustainable Diet and Meat Reduction
21	Marine Environment and Fishing Protection
22	Domestic Animal Management
23	Environmental Education
24	Forest Management and Protection
25	Bird Protection and Nest Box Installation
26	Creation of Wildlife Corridors
27	Water Management and Conservation
28	Ecological Management of Green Spaces
29	Fight Against Light Pollution
30	Limiting Human Population
31	Taxation of Polluting Products
32	Environmental and Ecological Policies

Populations” reflect the wide array of perspectives and priorities expressed by participants. Consequently, this breadth may lead to misunderstandings in public discourse, differing perceptions of policy implementation, and varied expectations from citizens. Such findings emphasize the complexity of addressing biodiversity-related issues in policymaking and the importance of considering these nuances to foster public engagement and policy acceptance.

To better understand these clusters, we analyzed their relative importance based on the number of proposals within each theme (Table 2). The most prominent clusters, such as “Organic and Sustainable Agriculture” (5.58%), “Hunting Regulations and Game Protection” (5.51%), and “Urban Greening and Green Spaces” (4.72%), likely correspond to topics that citizens most commonly associate with biodiversity or those that generate the highest levels of interest and concern. In contrast, smaller clusters, such as “Limiting Human Population” (0.68%) and “Fight Against Light Pollution” (1.21%), represent themes that, while still relevant, appear to capture more specific or less widespread areas of interest. This distribution highlights how public engagement with biodiversity is shaped by both familiar issues and emerging concerns.

To facilitate the understanding of the 32 thematic clusters identified through topic modeling, a qualitative aggregation procedure was conducted by the research team. Each cluster was independently reviewed by two researchers, who examined the thematic labels and the representative proposals associated with the cluster. The researchers then assigned the clusters to higher-level categories based on semantic proximity and substantive similarity. Following this independent coding step, the results were compared and discussed collectively within the team. Discrepancies were resolved through deliberation until consensus was reached. This iterative procedure ensured intercoder reliability and minimized individual biases, following best practices in qualitative content analysis (Saldaña, 2021; Schreier, 2012).

Through this process, the 32 clusters were grouped into eight overarching thematic families. This methodological choice offers a balance

**Table 2**  
Number of proposals for each cluster.

Cluster	Name	No. of proposals	Proportion
15	Organic and Sustainable Agriculture	310	5,58%
4	Hunting Regulations and Game Protection	306	5,51%
10	Urban Greening and Green Spaces	262	4,72%
23	Environmental Education	257	4,63%
19	Human Activity Impact on the Environment	240	4,32%
18	Biodiversity Protection and Environmental Actions	236	4,25%
24	Forest Management and Protection	233	4,20%
3	Waste and Plastic Reduction	226	4,07%
32	Environmental and Ecological Policies	224	4,03%
2	Urban Gardens and Green Spaces	208	3,75%
9	Urbanization and Land Artificialization	208	3,75%
5	Protection of Natural Areas and Wildlife	196	3,53%
14	Biodiversity Awareness	193	3,47%
12	Pesticide Ban	192	3,46%
13	Species and Predator Protection	179	3,22%
28	Ecological Management of Green Spaces	169	3,04%
20	Sustainable Diet and Meat Reduction	166	2,99%
1	Agricultural and Vegetal Planning	155	2,79%
11	Waste Management and Pollution	155	2,79%
27	Water Management and Conservation	149	2,68%
8	Animal Protection and Livestock Farming	145	2,61%
31	Taxation of Polluting Products	145	2,61%
16	Sustainable Transport and Mobility	140	2,52%
25	Bird Protection and Nest Box Installation	132	2,38%
21	Marine Environment and Fishing Protection	121	2,18%
17	Solar and Renewable Energy	118	2,12%
6	Water Management and Quality	115	2,07%
7	Wetland Preservation	110	1,98%
26	Creation of Wildlife Corridors	82	1,48%
22	Domestic Animal Management	77	1,39%
29	Fight Against Light Pollution	67	1,21%
30	Limiting Human Population	38	0,68%

between granularity and interpretability: the original clusters are preserved for detailed analysis, while the aggregated families provide a clearer framework for policy-relevant interpretation. This dual-level perspective allows the results to be examined both at a fine-grained and a more holistic scale.

The need for such human interpretation is consistent with findings in the literature. Jacobi et al. (2016) emphasize that mathematical measures of model fit (e.g., perplexity) do not necessarily guarantee semantic interpretability, and that human judgment remains indispensable for aligning computational results with substantive research questions. Lau et al. (2014) similarly argue that coherence scores approximate but cannot replace expert evaluation in applied contexts. More recently, Eklund and Forsman (2022) confirmed that even advanced neural embedding-based clustering pipelines require human validation to ensure thematic coherence and analytical usefulness.

As a result of this combined computational and interpretative approach, we identified the following eight thematic families, reflecting diverse citizen concerns regarding biodiversity (Table 3). *Management of Natural and Urban Spaces* (22.90%) emerges as the largest category, emphasizing the role of green and natural areas. *Resource Management and Pollution Control* (15.43%) highlights the perceived impact of pollution and resource practices on biodiversity loss. *Wildlife Protection and Species Management* (15.11%) underscores strong concerns for conservation and ethical animal management. *Environmental Education and Awareness* (12.42%) reinforces the role of public understanding in fostering pro-biodiversity behaviors. *Agriculture and Sustainable Food Practices* (11.36%) links agricultural systems and dietary habits to biodiversity preservation. *Environmental Policy and Regulatory Measures*

**Table 3**  
Composition of the eight thematic families of citizen proposals on biodiversity.

Family	Total_Proposals	Percentage	Clusters_Details
Agriculture and Sustainable Food Practices	631	11.36	Cluster 15 (5.58%); Cluster 20 (2.99%); Cluster 1 (2.79%)
Wildlife Protection and Species Management	839	15.11	Cluster 4 (5.51%); Cluster 13 (3.22%); Cluster 8 (2.61%); Cluster 25 (2.38%); Cluster 22 (1.39%)
Management of Natural and Urban Spaces	1272	22.9	Cluster 10 (4.72%); Cluster 24 (4.2%); Cluster 2 (3.75%); Cluster 9 (3.75%); Cluster 28 (3.04%); Cluster 7 (1.98%); Cluster 26 (1.48%)
Resource Management and Pollution Control	857	15.43	Cluster 3 (4.07%); Cluster 11 (2.79%); Cluster 27 (2.68%); Cluster 31 (2.61%); Cluster 6 (2.07%); Cluster 29 (1.21%)
Environmental Education and Awareness	690	12.42	Cluster 23 (4.63%); Cluster 19 (4.32%); Cluster 14 (3.47%)
Environmental Policy and Regulatory Measures	612	11.02	Cluster 32 (4.03%); Cluster 5 (3.53%); Cluster 12 (3.46%)
Sustainable Mobility, Energy, and Marine Conservation	379	6.82	Cluster 16 (2.52%); Cluster 21 (2.18%); Cluster 17 (2.12%)
Emerging and Niche Concerns	38	0.68	Cluster 30 (0.68%)

(9.67%) reflects expectations for stronger governmental action. *Sustainable Mobility, Energy, and Marine Conservation* (6.42%) acknowledges the systemic connections between biodiversity, energy transition, and mobility. Finally, *Emerging and Niche Concerns* (0.68%) signal less prominent but noteworthy societal considerations. These families represent the main areas of concern expressed by participants and provide a more integrated framework for interpreting the dataset.

To assess the robustness of these groupings, a hierarchical clustering analysis was conducted (see [Appendix 1.2](#)). While the statistical structure validates several key proximities, the final classification prioritized conceptual and policy-oriented coherence over strict lexical adjacency, ensuring the analytical relevance of the eight thematic families.

## 4.2. Polarization

### 4.2.1. Procedure

To go further, we conduct a polarization analysis based on the votes for each proposal. The polarization analysis was conducted using our BPI, which quantifies the degree of divergence in opinions within each cluster. As a reminder (see methodology), the BPI was computed by considering the proportion of affirmative and opposing votes, while adjusting for the presence of neutral votes. Higher values indicate greater division, whereas lower values reflect broader consensus or minimal contention.

We apply this index to our 32 clusters ([Table 4](#)) and visualize the results using a bar plot, where clusters are mapped against their BPI values ([Fig. 5](#)). The plot includes labels to identify each cluster and a color gradient, with high polarization clusters in red and low polarization clusters in blue. This method allows us to distinguish between strongly polarized clusters, where opposing views are equally represented, and consensus-driven clusters, where one position dominates, or neutrality prevails.

**Table 4**  
Votes and BPI by thematic cluster.

Cluster	Total votes	Votes pro	Votes against	Votes neutral	Active votes	BPI
Agricultural and Vegetal Planning	46,333	36,888	4610	4835	41,498	0,35
Urban Gardens and Green Spaces	60,504	41,234	7120	12,150	48,354	0,40
Waste and Plastic Reduction	67,539	50,927	5957	10,655	56,884	0,32
Hunting Regulations and Game Protection	102,658	61,507	27,141	14,010	88,648	0,73
Protection of Natural Areas and Wildlife	67,980	48,503	10,450	9027	58,953	0,51
Water Management and Quality	33,490	24,306	3691	5493	27,997	0,38
Wetland Preservation	32,593	24,468	3090	5035	27,558	0,34
Animal Protection and Livestock Farming	44,414	26,957	8848	8609	35,805	0,60
Urbanization and Land Artificialization	60,507	41,481	7874	11,152	49,355	0,44
Urban Greening and Green Spaces	77,165	59,885	6561	10,719	66,446	0,31
Waste Management and Pollution	45,859	33,789	4556	7514	38,345	0,35
Pesticide Ban	57,710	39,982	8855	8873	48,837	0,50
Species and Predator Protection	65,779	41,516	14,408	9855	55,924	0,65
Biodiversity Awareness	56,735	41,813	5402	9520	47,215	0,34
Organic and Sustainable Agriculture	92,124	65,259	12,358	14,507	77,617	0,45
Sustainable Transport and Mobility	41,831	26,118	7400	8313	33,518	0,55
Solar and Renewable Energy	34,198	19,922	5785	8491	25,707	0,52
Biodiversity Protection and Environmental Actions	70,687	50,105	8613	11,969	58,718	0,42
Human Activity Impact on the Environment	69,940	44,423	8767	16,750	53,190	0,42
Sustainable Diet and Meat Reduction	50,212	30,920	10,182	9110	41,102	0,61
Marine Environment and Fishing Protection	36,737	26,291	4372	6074	30,663	0,41
Domestic Animal Management	23,007	13,260	4902	4845	18,162	0,62
Environmental Education	76,176	60,657	6085	9434	66,742	0,29
Forest Management and Protection	68,927	45,968	10,336	12,623	56,304	0,49
Bird Protection and Nest Box Installation	38,843	26,459	4969	7415	31,428	0,43

(continued on next page)

Table 4 (continued)

Cluster	Total votes	Votes pro	Votes against	Votes neutral	Active votes	BPI
Creation of Wildlife Corridors	26,118	18,743	3242	4133	21,985	0,42
Water Management and Conservation	43,131	29,727	5595	7809	35,322	0,44
Ecological Management of Green Spaces	53,580	35,863	8494	9223	44,357	0,51
Fight Against Light Pollution	31,463	26,462	2115	2886	28,577	0,25
Limiting Human Population	11,472	4500	4034	2938	8534	0,74
Taxation of Polluting Products	44,413	28,071	8325	8017	36,396	0,58
Environmental and Ecological Policies	69,166	44,758	10,930	13,478	55,688	0,51

To simplify the categorization of the clusters, we applied a tercile-based classification, ensuring an even distribution of clusters into three levels: high polarization, moderate polarization, and low polarization. Clusters with a BPI above 0.42 were categorized as highly polarized, those between 0.32 and 0.42 as moderately polarized, and those below 0.32 as weakly polarized. This approach provides a categorically grounded measure of polarization and extends traditional polarization metrics to discrete voting data.

4.2.2. Analysis of polarization by clusters

The highly polarized clusters (BPI > 0.42) display the greatest divergence of opinions, signaling contentious debates and strong divisions among the population. Notably, *Hunting Regulations and Game Protection* (0.73) and *Limiting Human Population* (0.74) exhibit the highest polarization levels, suggesting that these topics provoke intense ethical and societal debates. For *Hunting Regulations*, the tension may arise from conflicts between cultural practices and conservation needs. The *Limiting Human Population* cluster, although representing only 0.68% of total proposals, reveals significant societal concerns about the implications of population control for biodiversity. Other highly

polarized clusters, such as *Species and Predator Protection* (0.65), *Animal Protection and Livestock Farming* (0.60), and *Sustainable Diet and Meat Reduction* (0.61), highlight ethical considerations related to human-animal interactions and dietary practices, indicating deeply held but conflicting views. The high polarization in *Sustainable Transport and Mobility* (0.55) and *Solar and Renewable Energy* (0.52) suggests socio-technical challenges, where support for ecological transitions is tempered by economic or infrastructural concerns.

Clusters categorized under moderate polarization (BPI between 0.32 and 0.42) present a more balanced distribution of affirmative, opposing, and neutral votes, signaling potential for consensus-building. For example, *Urban Gardens and Green Spaces* (0.40) and *Water Management and Quality* (0.38) indicate that while opinions are divided, there is room for alignment, potentially through targeted communication and policy incentives. Clusters such as *Urbanization and Land Artificialization* (0.44) and *Marine Environment and Fishing Protection* (0.41) reflect the tension between economic development and biodiversity protection, where citizens recognize the importance of both but differ in how to balance these priorities. Interestingly, *Biodiversity Protection and Environmental Actions* (0.42) also falls within this category, suggesting general agreement on the need for biodiversity actions, though there remains debate on methods and priorities.

Finally, the low polarization clusters (BPI < 0.32) demonstrate strong consensus, either favorable or unfavorable. Importantly, the direction of consensus was assessed based on the proportion of supportive votes. For example, *Environmental Education* (0.29) and *Fight Against Light Pollution* (0.25) exhibit low polarization with a favorable consensus, indicating broad public agreement on the need for awareness programs and the mitigation of light pollution, respectively. These clusters reflect topics where citizen support is strong and stable, presenting low resistance to policy adoption. Similarly, *Urban Greening and Green Spaces* (0.31) and *Waste and Plastic Reduction* (0.32) also display low polarization and favorable consensus, highlighting citizens' recognition of the role that urban biodiversity and waste management play in sustainability.

These findings emphasize the varying degrees of public agreement on biodiversity-related issues. Highly polarized topics may require more nuanced communication strategies and targeted engagement efforts to bridge opinion gaps, whereas low-polarization topics represent opportunities for policy reinforcement and implementation. The categorization provides valuable insights for policymakers to prioritize

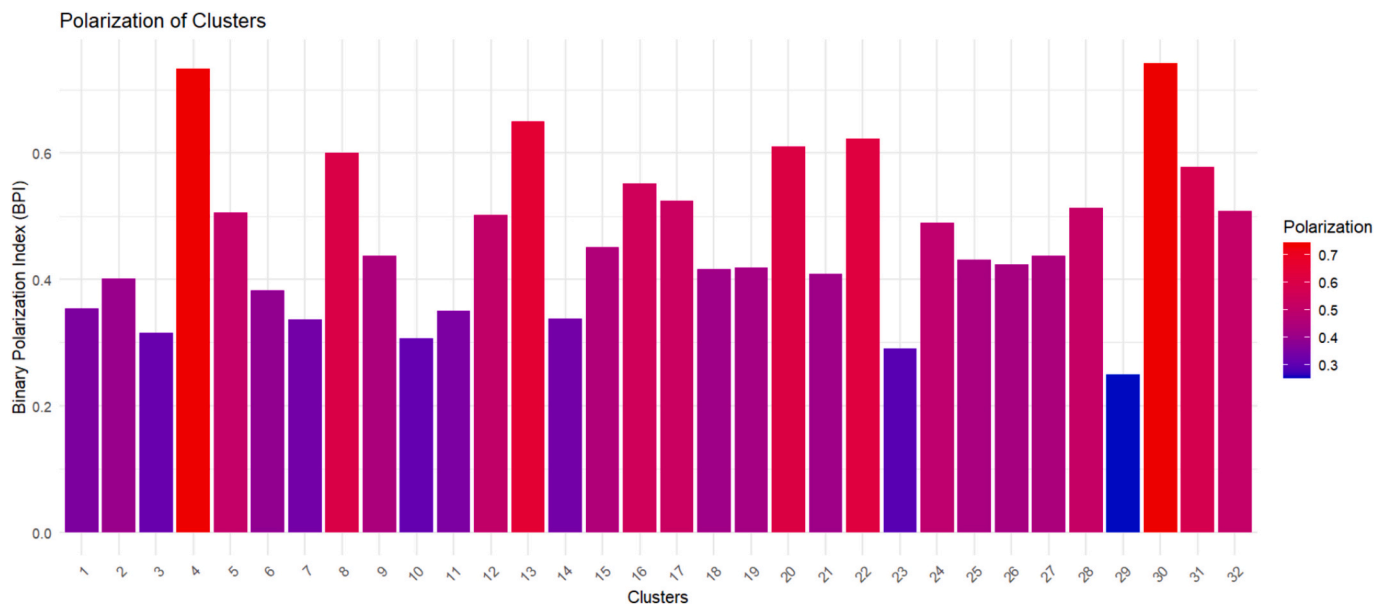


Fig. 5. Polarization levels by thematic cluster.

interventions based on levels of public contention or support.

#### 4.2.3. Analysis of polarization by families of clusters

The *Management of Natural and Urban Spaces* family encompasses clusters related to urban and natural spatial management, such as Urban Greening and Green Spaces, Urban Gardens, Forest Management, and Creation of Wildlife Corridors. Interestingly, polarization levels in this family are low to high, indicating both shared concerns and contested perspectives. Urban Greening and Green Spaces (BPI = 0.31) shows low polarization with a strongly favorable consensus (59,885 supportive votes out of 77,165). This suggests broad acceptance of urban greening initiatives. In contrast, Urbanization and Land Artificialization (BPI = 0.44) and Forest Management and Protection (BPI = 0.49) are highly polarized, revealing conflicting priorities between development needs and biodiversity preservation. The divergence may arise from economic interests linked to urban expansion and land use, particularly in peri-urban areas. Creation of Wildlife Corridors (BPI = 0.42) also falls within the high polarization category, underscoring contention over land reallocation for ecological purposes.

This family highlights how spatial management decisions are perceived as either opportunities for sustainable living or constraints on urban development, depending on the respondents' socio-economic perspectives.

Topics in the *Resource Management and Pollution Control* family, such as Waste and Plastic Reduction, Water Management, and Fight Against Light Pollution, directly address the impact of human activities on ecosystems. The polarization levels range from low to high, reflecting nuanced public views. Waste and Plastic Reduction (BPI = 0.32) and Water Management and Quality (BPI = 0.38) are moderately polarized, suggesting general agreement on the need for action but with disagreements over implementation strategies, such as bans on single-use plastics or the adoption of circular economy practices. The Fight Against Light Pollution cluster (BPI = 0.25) stands out with low polarization and a favorable consensus (26,462 supportive votes out of 31,463), indicating broad acknowledgment of the negative effects of light pollution on nocturnal biodiversity and human well-being. However, Taxation of Polluting Products (BPI = 0.58) exhibits high polarization, showing that economic instruments for environmental regulation remain contentious, possibly due to concerns about equity and consumer costs.

This family underscores that regulatory approaches related to pollution are more divisive than behavioral adjustments, which tend to generate consensus.

In the *Wildlife Protection and Species Management* family, clusters focus on species protection, hunting regulations, and ethical animal management, demonstrating high polarization in most cases. Hunting Regulations and Game Protection (BPI = 0.73) and Species and Predator Protection (BPI = 0.65) are among the most polarized clusters, indicating deep societal divisions. These likely stem from cultural traditions, rural-urban divides, and economic dependencies on hunting and related activities. Similarly, Domestic Animal Management (BPI = 0.62) and Animal Protection and Livestock Farming (BPI = 0.60) also show high polarization, reflecting ongoing debates over animal rights, industrial farming practices, and their environmental impacts. Bird Protection and Nest Box Installation (BPI = 0.43) sits on the borderline of high polarization, suggesting less contentious discussions when wildlife protection measures do not conflict with significant economic interests. Overall, this family illustrates that ethical considerations regarding wildlife often lead to highly polarized debates, especially when linked to traditional practices or economic activities.

The *Environmental Education and Awareness* family shows relatively lower polarization levels, pointing to broad agreement on the role of education in biodiversity protection. Environmental Education (BPI = 0.29) and Biodiversity Awareness (BPI = 0.34) exhibit low polarization with a strongly favorable consensus (e.g., 60,657 supportive votes for Environmental Education out of 76,176). The Human Activity Impact on

the Environment cluster (BPI = 0.42), however, sits at the threshold of high polarization, suggesting that while there is general acknowledgment of human responsibility, there may be disagreements on the extent of this responsibility and necessary lifestyle changes. This family demonstrates that while knowledge-sharing efforts are widely accepted, translating awareness into action remains a polarizing challenge.

The topics of the *Agriculture and Sustainable Food Practices* family link agricultural systems, dietary habits, and biodiversity preservation, showing polarization from moderate to high levels. Organic and Sustainable Agriculture (BPI = 0.45) and Sustainable Diet and Meat Reduction (BPI = 0.61) are highly polarized, reflecting strong divisions over agricultural transitions, meat consumption, and food sovereignty. The debate is likely influenced by economic concerns, cultural food practices, and misconceptions regarding plant-based diets. Agricultural and Vegetal Planning (BPI = 0.35) shows moderate polarization, suggesting that while there is broad recognition of the need for sustainable agricultural planning, implementation details (e.g., land-use regulations or subsidies) may cause disagreements.

This family underscores that agriculture-related transitions could be sensitive, touching on identity, economy, and tradition.

Clusters in the *Environmental Policy and Regulatory Measures* family focus on regulatory frameworks, showing high polarization, especially regarding state intervention. Environmental and Ecological Policies (BPI = 0.51) and Protection of Natural Areas and Wildlife (BPI = 0.51) both show high polarization, reflecting divergent opinions on government-led conservation initiatives. Pesticide Ban (BPI = 0.50) also falls within the high polarization category, revealing tensions between health concerns, agricultural practices, and industrial lobbying.

These results highlight that top-down environmental policies are controversial, often perceived as restrictive or misaligned with certain stakeholders' interests.

The *Sustainable Mobility, Energy, and Marine Conservation* family illustrates a notably high level of polarization, particularly concerning energy transition topics. Solar and Renewable Energy (BPI = 0.52) and Sustainable Transport and Mobility (BPI = 0.55) are highly polarized, showing substantial division regarding energy infrastructures and mobility reforms. Marine Environment and Fishing Protection (BPI = 0.41) is moderately polarized. This family reveals that while marine conservation attracts some consensus, systemic transitions in energy and mobility are highly divisive.

Finally, in the *Emerging and Niche Concerns* family, the Limiting Human Population cluster (BPI = 0.74) is the most polarized among all, reflecting ethical, cultural, and political sensitivities surrounding population control discussions. Despite its low volume of proposals (0.68% of total), this theme evokes deep divisions, possibly related to individual freedoms, global inequalities, and ethical considerations about demographic policies.

## 5. Discussion

### 5.1. Theoretical contributions

Our study makes two primary theoretical contributions. Firstly, our study introduces a novel perspective on ecosystem services by examining them through the lens of polarization. Existing literature has largely focused on the classification of ecosystem services into provisioning, regulating, and cultural categories (Fisher et al., 2009; Haines-Young and Potschin, 2018; Matzek and Wilson, 2021; Schröter et al., 2014), yet little attention has been given to their role in societal divisions. Our findings reveal a clear pattern: regulating services, such as habitat protection and species management, are associated with high levels of polarization due to conflicts between economic and environmental stakeholders. Issues like hunting regulations (BPI = 0.73) and predator protection (BPI = 0.65) create stark divisions between groups with distinct identity-driven values, such as hunters, farmers, and conservationists. In contrast, cultural services, including environmental

education (BPI = 0.29) and urban green spaces (BPI = 0.31), enjoy broader consensus, as they are perceived as common goods benefiting society at large. Provisioning services, particularly sustainable agriculture (BPI = 0.45) and water resource management (BPI = 0.38), exhibit moderate polarization, reflecting economic and agroecological transition debates. By linking ecosystem services to varying levels of polarization, our study enriches theories of environmental governance (Bennett and Satterfield, 2018; Maczka et al., 2021) by demonstrating how governance conflicts are structured not only around ecological trade-offs but also around value-laden societal divides. This perspective highlights the need for governance frameworks that account for the differentiated conflict potential of ecosystem services.

Secondly, we contribute to the literature on citizen participation by developing a polarization-sensitive framework for analyzing public consultations. Existing research on participatory governance has emphasized inclusiveness and procedural legitimacy (Bedessem et al., 2022; Dryzek et al., 2019; Kitamori, 2025), but often treats citizen input as aggregated preferences without sufficiently accounting for divisions within public opinion. Building on recent work on polarization (Bernaerts et al., 2023; Hoffarth and Hodson, 2016; Jost et al., 2022), our approach distinguishes between consensual domains (e.g., education, urban greening) and highly contested ones (e.g., taxation of polluting products, energy transition). This polarization-sensitive lens advances theories of participatory governance by showing that public consultations are not only tools for eliciting collective priorities but also arenas where underlying societal divides are expressed and negotiated (Swette et al., 2023). Such an approach provides a theoretical bridge between deliberative democracy and empirical studies of polarization, emphasizing that citizen participation should be understood not only in terms of inclusiveness but also in terms of its capacity to reveal and manage conflict.

Thirdly, our study contributes to the emerging literature on AI and policy studies by theorizing the role of LLMs in participatory governance. Whereas prior applications of computational text analysis in policy research have relied on dictionary- or keyword-based approaches (Grimmer and Stewart, 2013; Qaiser and Ali, 2018), our work shows how LLM-based embeddings reshape the way citizen input can be conceptualized and analyzed at scale. Rather than treating LLMs merely as technical tools, we highlight their potential to transform participatory governance by making it possible to capture consensus and conflict simultaneously within large-scale consultations. This points to an evolving theoretical agenda on the interplay between digital methods, democratic legitimacy, and polarization management (Rudin, 2019; Weidinger et al., 2021). Recent studies in adjacent fields, such as climate discourse analysis or automated policy generation (Fariello and Jemielniak, 2025; Larosa et al., 2025), confirm the growing relevance of this agenda. In this sense, LLMs are methodological devices but also catalysts for rethinking how public participation can be integrated into governance frameworks.

Finally, biodiversity research has often faced a dichotomy: studies tend to be either overly general, addressing biodiversity as a broad concept, or excessively fragmented, analyzing specific sectors such as conservation, agriculture, or urban planning (Tindale et al., 2023; De Groot and Steg, 2008; Judge et al., 2023). Our research addresses this gap by leveraging a large-scale citizen public analyzed through LLMs, providing a holistic view of biodiversity that integrates various dimensions into a single analytical framework. Unlike prior studies that focus on isolated sectors (Guber, 2013; Lucas and Warman, 2018; Judge et al., 2023), our methodology captures how biodiversity is perceived as an interconnected issue within public discourse. By applying advanced computational techniques, we identify how citizens simultaneously engage with biodiversity through concerns about species conservation, land-use planning, and sustainable food systems. This comprehensive perspective aligns with recent calls in the literature for a more integrated approach to biodiversity governance (Maczka et al., 2021; Young, 2016; Kiss et al., 2022) and provides empirical evidence that biodiversity-

related issues cannot be effectively addressed in isolation. Our study contributes to the literature by bridging the gap between sectoral and overarching approaches, demonstrating how computational techniques can facilitate a nuanced, large-scale understanding of biodiversity governance.

## 5.2. Methodological contributions

Our study offers also methodological advancements by integrating large-scale public consultation data with advanced NLP techniques, particularly embeddings from LLMs. Traditional methods for analyzing public participation in biodiversity governance have relied on manual coding or keyword-based approaches, which often fail to capture the complexity and scale of citizen input (Grimmer and Stewart, 2013). By employing LLMs for semantic clustering and polarization analysis, we provide a scalable and replicable framework that enables the extraction of nuanced insights from vast participatory datasets. Importantly, this framework is designed to bridge the interpretative opacity of high-dimensional embeddings with structured human judgment, through qualitative validation and transparent polarization metrics, thereby supporting informed and accountable decision-making.

Additionally, our integration of computational clustering techniques, such as K-means with PCA-based dimensionality reduction, enables the identification of emergent themes in public consultations without relying on predefined categories. This bottom-up approach ensures that thematic structures emerge organically from citizen input, enhancing the robustness and interpretability of results.

A key methodological contribution of our study is the development of a polarization metric, BPI, tailored for categorical voting data. While existing polarization indices primarily address continuous or ordinal variables (Esteban and Ray, 1994; Dalton, 2008), our BPI accounts for the presence of neutral votes and assesses the degree of division within citizen responses. This methodological innovation allows for a more precise measurement of contentious versus consensual themes in public discourse, aiding policymakers in identifying areas requiring targeted engagement strategies. For example, the BPI could be applied to the analysis of social media data, provided that individual messages or user positions are categorized into three distinct stances—support (“Pro”), opposition (“Against”), or neutrality. This trichotomous classification is increasingly common in studies of online discourse, particularly those investigating political debates, policy controversies, or public reactions to environmental issues. Once classified, the BPI can be computed for clusters of messages grouped by topic, hashtag, time period, or user community, enabling researchers to track the evolution of polarization over time, compare patterns across platforms, or assess the internal cohesion of digital communities. In this context, the BPI offers a lightweight yet robust alternative to more complex network-based or entropy-based measures, especially when engagement levels (reflected in neutral or disengaged responses) are analytically relevant. This suggests promising extensions of the BPI to the study of digital deliberation, online mobilization, and the dynamics of public opinion in algorithmically mediated spaces.

By combining large-scale citizen engagement with state-of-the-art NLP techniques, our study advances methodological practices in participatory governance research, offering a replicable framework for analyzing public expectations on biodiversity and environmental policy.

## 5.3. Managerial implications

Our findings have significant implications for policymakers, environmental organizations, and corporate sustainability initiatives. Highly polarized issues, such as hunting regulations, taxation of polluting products, or energy transition, require tailored strategic communication. Rather than imposing contested measures, it is necessary to clarify the benefits for each stakeholder group and counter misinformation. Additionally, actively engaging stakeholders in decision-making can

help integrate diverse interests and foster policy acceptance, reducing resistance and enhancing long-term effectiveness.

Conversely, widely accepted topics, such as urban greening or environmental education, should be prioritized for immediate action, as they enjoy strong support. Moderately polarized issues, like water management or forest protection, offer opportunities for policy experimentation. Pilot programs can be implemented to test different approaches, allowing for adjustments before broader implementation, thereby mitigating risks and fostering gradual consensus-building.

Additionally, our methodological approach offers a data-driven tool for identifying societal priorities in biodiversity governance. By leveraging LLMs and participatory consultations, organizations can refine policy designs that align with public sentiment, increasing both legitimacy and adoption rates. Businesses in sectors linked to sustainability and biodiversity can also use these insights to align their corporate social responsibility initiatives with public expectations, enhancing consumer trust and stakeholder engagement.

#### 5.4. Limitations and future research directions

This study presents certain limitations, particularly the lack of detailed sociodemographic data on consultation participants. This restricts our ability to distinguish polarization patterns based on individual characteristics such as age, education level, or professional background. Thus, future research would benefit from integrating individual-level data—such as demographic, socioeconomic, political and psychological attributes—drawing on insights from the literature on identity-based polarization (Hoffarth and Hodson, 2016; Jost et al., 2022) and values-driven public opinion (De Groot and Steg, 2008). For instance, studies on political polarization have shown that identity-driven divisions can shape attitudes toward policy adoption (Iyengar et al., 2012; Thota et al., 2023), a phenomenon that may also influence citizen responses in biodiversity governance. This would allow for a deeper investigation into identity-based and affective polarization (Shah, 2025). Understanding how specific social groups respond to biodiversity-related policies could inform the design of more inclusive and effective consultation processes. This approach would also help policymakers anticipate potential areas of conflict and tailor communication strategies accordingly.

Additionally, our analysis focuses on France, which may limit the generalizability of our findings to other national contexts. Future research should explore the application of our methodology across different political contexts to determine whether the discursive polarization patterns identified in France are consistent with broader global trends. Comparing public consultations in highly polarized countries, such as the United States, and in more consensus-oriented nations, like Sweden, would provide insights into how political structures influence

## Appendix 1

### 1.1. Appendix 1.1

Cluster	Rank	Proposal
1	1	Plant hedgerows of local species around fields and along paths.
1	2	We need to replant hedges on a massive scale and prevent farmers and local authorities from cutting back existing ones to 1.50m.
1	3	Hedges must be replanted on a massive scale between fields and meadows, along roads and around towns and villages.
1	4	Hedges need to be put back in place for birds and wildlife.
1	5	We need to replant hedges in the fields.
1	6	Plant hedges
1	7	We need to help farmers plant and care for the hedges bordering their fields.
1	8	The hedges that surround meadows and fields should no longer be removed.
1	9	We need to replant roadside trees and natural hedges between cultivated plots.
1	10	We must encourage the reconstruction of hedges in agricultural fields
2	1	We need to encourage people to create areas of biodiversity in their gardens.
2	2	We need to offer grants to individuals and municipalities to plant hedges and create ponds to encourage biodiversity.

(continued on next page)

the structure and intensity of public debates on biodiversity governance (Bernaerts et al., 2023).

In terms of methodological robustness, we acknowledge that the use of LLMs in policy-related text analysis raises critical questions of validation and reproducibility. Our study can be reproduced in principle, as we provide a detailed description of the data source, preprocessing, and analytical workflow. Validation relied on internal consistency checks, including iterative reviews of model outputs and systematic examination of the coding scheme to ensure stable categories. However, strict replication of the exact results is constrained by the reliance on proprietary LLMs, whose parameters and outputs may vary over time. We therefore suggest that future work could benefit from open-source models and benchmark datasets to improve replicability and comparability across studies.

Another promising direction is to apply this approach to other policy areas that may exhibit different polarization dynamics. Topics such as artificial intelligence governance, digital privacy regulations, and public trust in automation each have distinct socio-political dimensions that could reveal additional insights into the mechanisms driving public consensus and contention. By analyzing the extent to which polarization patterns persist across diverse topics, researchers can refine models of public deliberation and environmental governance.

Finally, it is important to acknowledge a key limitation related to the nature of public consultation data. Participation in public consultations is voluntary and reflects the views of a self-selected subset of the population rather than the general public. As a result, the data analyzed in this study capture only the perspectives of participating citizens, and LLM-based analyses are necessarily limited to expressed viewpoints. Accordingly, our findings should be interpreted as insights into the structure and diversity of articulated opinions within the consultation, rather than as population-level estimates.

#### CRediT authorship contribution statement

**Magali Trelohan:** Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Anne-Cécile Gay:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **David Moroz:** Validation. **Damien Chaney:** Writing – original draft, Validation.

#### Acknowledgments

The authors would like to thank [Make.org](https://www.make.org/), and in particular David Mas, Chief AI Officer, for sharing access to the citizen consultation data that made this research possible.

(continued)

Cluster	Rank	Proposal
2	3	We need to encourage shared gardens to recreate areas of biodiversity in the city.
2	4	We need to enable city dwellers to become more involved in their community's biodiversity through more accessible vegetable gardens.
2	5	We need to develop participatory gardens and green areas such as flower meadows.
2	6	We need to create community gardens to grow local plants.
2	7	We need to turn private gardens into biodiversity areas: create biodiversity specifications for gardens in local authorities.
2	8	Communities should offer a choice of flower seeds to encourage biodiversity, bird boxes, insect hotels...
2	9	We need to promote permaculture in the management of public and private gardens (rational and differentiated mowing, mulching, etc.).
2	10	We need to educate people about gardening and encourage them to grow things by promoting permaculture.
3	1	We need to reduce the use of disposable packaging within companies and reintroduce deposits and bulk packaging.
3	2	Stop using plastic packaging (return to deposits, reusable packaging, etc.).
3	3	We must force manufacturers to stop over-packaging and produce recyclable materials by developing recycling companies.
3	4	We need to ban unnecessary packaging, or replace it with bio-degradable materials.
3	5	We need to ban the use of petroleum-based plastics in packaging and require the use of materials that are easy to recycle.
3	6	We need to ban all production/imports of objects that cannot be repaired/recycled and are made of polluting products (plastics).
3	7	We need to consume and therefore produce fewer objects, in particular reduce to a strict minimum those that are single-use or made of plastic.
3	8	We need to ban plastic and develop packaging-free products and other alternatives.
3	9	We need to reduce our use of plastic to a strict minimum by forcing manufacturers to ban it from their packaging.
3	10	Companies must be forced to stop over-packaging goods to reduce plastic waste.
4	1	Hunting must be banned.
4	2	Hunting must be banned.
4	3	Hunting must be banned.
4	4	Hunting must be banned
4	5	Hunting must be banned
4	6	Hunting must be banned
4	7	We need to limit the number of hunters and ban hunting drives.
4	8	We must ban the breeding of animals for hunting and their release into the wild, as well as the staunching of game.
4	9	We need to limit hunting to «edible» species and ban recreational hunting.
4	10	We must ban «hunting» in enclosures and promote natural, ethical and environmentally-friendly hunting.
5	1	Protected areas must be created: national parks, nature reserves and other protected areas must be established and maintained.
5	2	We need to create more protected natural areas and ensure that the laws governing these areas are respected.
5	3	Wildlife reserves, free from human nuisances such as hunting, should be widespread to restore biodiversity.
5	4	We need to create sanctuaries in the last wilderness areas, where no human development or exploitation is allowed.
5	5	We need to set up more Integrated Reserves throughout the country. It is forbidden to enter these territories of at least 1000 ha.
5	6	We need to protect more wild spaces.
5	7	We need to protect our nature parks, turning them into biodiversity reserves and havens of peace for our wildlife.
5	8	We need to create wildlife reserves, natural areas where human activities are prohibited except for hiking on marked trails.
5	9	We need to develop wildlife reserves with no hunting or exploitation (including forests, hedges and/or wetlands).
5	10	We need to create wildlife reserves free from human activity.
6	1	We need to clean up our waterways (and ban/punish all chemical discharges into the environment).
6	2	We need to control the area around watercourses and penalize all illegal discharges (sanitary facilities not connected to the network, factories, etc.).
6	3	Increased regulation of pollutant discharges (agricultural, industrial, etc.) into watercourses
6	4	Polluters must be severely penalized! Individuals and companies alike (fight against illegal dumping and dumping in water).
6	5	We need strict regulation of eternal pollutants and plastics in our industrial production circuits.
6	6	Polluting industries must be banned.
6	7	It is imperative to work on water pollution, and to eliminate the chemical and pharmaceutical molecules it contains as a matter of urgency.
6	8	Rivers must be cleaned up and those who pollute them severely penalized
6	9	We need to counter those who pollute not only the land, but also the waters up to the ocean.
6	10	We need to stop using household products filled with solvents, phosphates and other harmful substances that end up in our river water.
7	1	Wetlands need to be restored and preserved, by returning them to grassland, re-meandering them, landscaping their banks and so on.
7	2	We need to restore wetlands, meandering rivers and riparian forests, which will preserve water resources and ecological continuity.
7	3	Wetlands (ponds, marshes, peat bogs, etc.) must be preserved, restored and maintained in good ecological condition.
7	4	The old map of watercourses and wetlands should be restored and extended, rather than reduced as was done years ago.
7	5	Wetlands must be protected and restored
7	6	Encourage private individuals and farmers to create and maintain ponds and other wetlands
7	7	Preserve wetlands (lakes, ponds, watercourses, etc.)
7	8	We need to recreate the wetlands that have disappeared since 1950 to promote biodiversity (insects, fish, amphibians) and prevent flooding.
7	9	Wetlands must be preserved.
7	10	More natural environments and wetlands need to be restored
8	1	We must BAN hunting everywhere, CORRIDAS and animal sacrifices. We must stop eating animals.
8	2	We must ban the capture and keeping of wild animals (dolphins, circuses, zoos, etc.).
8	3	We must ban intensive factory farming, which pollutes and exploits sick animals.
8	4	We must formally ban all forms of hunting, as well as all shows that include animals (circuses/corridas/water parks).
8	5	We must ban all forms of hunting and all shows using animals (bullfighting, circuses, dolphinariums, zoos, etc.).
8	6	We need to ban intensive livestock farming and the new decree issued by the Minister of Agriculture on June 10, 2024, which limits controls on such farming.
8	7	The export and import of live animals must be banned.
8	8	Intensive livestock farming must be banned, and the general public must be made aware of what goes on there.
8	9	This unacceptable industrialization must be banned for caged pigs and chickens.
8	10	Hunting and trapping must be banned in all their forms (bullfighting, cockfighting and dogfighting).
9	1	We need to stop urban sprawl and stop building on agricultural land.
9	2	We need to limit construction, which is over-concreting cities in favor of housing with land and garden.
9	3	Buildings should be limited, but existing buildings should be rehabilitated to avoid encroaching on natural areas.
9	4	We need to put an immediate stop to the artificialization of land (ZAN policy) and force people to densify and rethink urban planning.
9	5	We need to ban all new construction in natural areas, and rehabilitate and green existing buildings.
9	6	We need to stop the development of housing estates and suburbs, which are increasingly encroaching on wilderness areas.
9	7	We need to rebuild on existing living areas and stop granting new building permits on undeveloped land.
9	8	We need to densify town centers and already urbanized areas, and stop the consumption of natural spaces for construction.

(continued on next page)

(continued)

Cluster	Rank	Proposal
9	9	We need to stop destroying natural areas to build on them, and reduce taxes for construction on brownfield sites.
9	10	We need to stop building buildings that are too large and urbanized.
10	1	We need to plant large public squares in cities and replant trees everywhere (squares, avenues, roadsides).
10	2	We need to encourage the creation of parks, shared gardens and green islands in every city.
10	3	Cities need to be greened, whether by vertical plant structures for buildings or by trees.
10	4	We need to encourage cities to re-vegetate, leave land and green spaces, and avoid concreting over everything.
10	5	Buildings, sidewalks and parking lots must be greened
10	6	We need to plant more trees in schoolyards and parking lots, and green up our communities.
10	7	We need to green up our cities, plant trees and use flat roofs to create green spaces (shared gardens).
10	8	We need to green up our ugly business parks and create as many roof terraces as possible on our buildings, and so on.
10	9	Mayors must be obliged to create more green spaces to compensate for construction and concreting.
10	10	We need to green up building courtyards in cities. Ask each resident to plant a plant, a tree ...
11	1	Penalties must be imposed on those who dump their rubbish in the open air: compensation in the form of fines or community service, depending on the damage.
11	2	We need to track down and get tougher with those who use illegal garbage dumps and/or leave their garbage in the countryside.
11	3	We need to launch campaigns to collect rubbish in nature more often and everywhere, and regularly remind people of its impact.
11	4	We need to educate people to dispose of their waste in garbage cans and not in nature: severely punish illegal dumping.
11	5	We must severely punish those who dump waste in nature and feed illegal dumps.
11	6	The law must be enforced against illegal dumping and littering.
11	7	We need to monitor and punish littering of all kinds.
11	8	We need to penalize people who litter, including cigarette butts. With a fine to pay on the spot.
11	9	Cigarette butt throwers should really be fined, or even take part in litter clean-ups after X number of repeat offences.
11	10	We need to severely penalize people who dump their garbage in the countryside, or throw garbage along roadsides.
12	1	Pesticides must be banned.
12	2	Synthetic pesticides and fertilizers must be banned.
12	3	Phytosanitary products, insecticides and other chemical products that massively destroy flora and fauna must be banned.
12	4	We must ban all use of pesticides and other insecticides.
12	5	We need to ban chemical spraying in nature and encourage the use of green fertilizers on crops.
12	6	Ban all chemical pesticides with harmful effects on humans and/or biodiversity
12	7	Insecticides, herbicides, fungicides and pesticides that kill bees, insects, birds, farmers and biodiversity must be banned!
12	8	A total ban on the use of chemical plant protection products that harm pollinating insects
12	9	We must ban the marketing of chemical pesticides/fertilizers that have not been tested or deemed harmful by science.
12	10	Pesticides and PFAS must be banned
13	1	We need to stop destroying species classified as pests, such as badgers, foxes, etc.
13	2	We need to give back space to species such as the wolf (shepherds need to do their job instead of leaving flocks alone).
13	3	We need to create more protected natural habitats, reintroduce more predators and ban hunting.
13	4	We need to relearn how to live with wild animals, and do away with the list of species likely to cause damage.
13	5	We need to help farmers protect their land and animals without killing so-called predators.
13	6	Large predators (wolf, bear, lynx) must be protected (no slaughter) because they are essential to the proper functioning of ecosystems.
13	7	We need to encourage the return of predators to recreate the balance that has been broken by inappropriate hunting policies.
13	8	We need to encourage the elimination of pests that destroy protected areas, such as wild boar.
13	9	We need to ban hunting and reintroduce even more large predators (wolves, lynx) so that ecosystems can recover!
13	10	Let the large predators (wolves, bears and lynxes) live, and they will regulate themselves according to the number of prey.
14	1	We need to raise awareness/inform the general public about the impact of human behavior on biodiversity and how to reduce this impact.
14	2	We need to educate and inform people about the impact of pests (animals/plants) to limit their impact in order to preserve biodiversity.
14	3	Raising public awareness of biodiversity-friendly developments
14	4	We need to make as many people as possible aware of the need to protect natural areas (dry grasslands, wetlands, forests, etc.).
14	5	We need to help individuals learn about the actions that can help maintain biodiversity in their immediate environment.
14	6	We need to make people aware of the challenges of restoring biodiversity through awareness-raising sessions!
14	7	Biodiversity education campaigns are needed on a large scale and in all strata of society (businesses, schools, etc.).
14	8	We need more commercials to raise awareness of good and bad habits and French biodiversity (plants, insects, etc.).
14	9	What's needed is a public communication campaign to promote and develop empathy with biodiversity (insects, plants, animals, etc.).
14	10	We need to inform the general public about little-known endangered species and raise awareness about their conservation.
15	1	We need to increase agricultural subsidies for small-scale organic and sustainable farming.
15	2	We need to reward farmers who work organically and encourage the development of permaculture areas.
15	3	We need to encourage local and organic farming, and improve the marketing of its products, by paying for the ecological service it provides.
15	4	We must help our farmers and support them in making the transition to more respectful crops (organic, permaculture, free-range).
15	5	Organic producers (pesticide-free and with mixed crops) must be promoted. Stop huge monoculture plantations.
15	6	We need to promote sustainable agriculture by reducing pesticides and monocultures, and using techniques that promote resilience.
15	7	We need to promote agro-ecology for an agriculture that respects biodiversity.
15	8	We need to encourage and/or practice fair organic farming, with fair remuneration throughout the value chain, especially for farmers.
15	9	We need to promote farming techniques that respect ecosystems, such as agro-ecology and organic farming, so that they become the norm.
15	10	We need to drastically reduce the biodiversity-destroying capitalist agro-industry and help small farmers instead.
16	1	We need to promote soft traffic and public transport.
16	2	We need to develop cycling and public transport networks in cities, and limit car access to local residents.
16	3	We need to develop public transport, especially in rural areas, penalize polluting transport and ban new roads.
16	4	We need to promote soft, free public transport everywhere, to reduce the need for individual transport.
16	5	We need to create more cities or neighborhoods where there are no cars and very few buses, or create a tramway.
16	6	We need to develop public transport
16	7	We need to make public transport free of charge in all cities, to limit car use.
16	8	We need to promote public transport in cities, so that in the long term, cars have no place in cities.
16	9	Transport must be nationalized and green transport deployed (hydrogen buses, bicycle paths, SNCF in rural areas).
16	10	We need to reduce the number of cars in cities and offer less polluting means of transport.
17	1	Photovoltaic panels should be installed on all new and existing buildings and industrial wasteland.
17	2	Photovoltaic panels must be installed on existing buildings and the destruction of forests to make way for solar panels must be prohibited.
17	3	We need to stop building photovoltaic parks (to the detriment of nature and agriculture) and equip public roofs!
17	4	We need to build nuclear power plants and reduce the use of land by wind turbines and solar panels as much as possible.

(continued on next page)

(continued)

Cluster	Rank	Proposal
17	5	We need to install more solar panels for heating, etc., at a lower cost.
17	6	Wind turbines and solar panels should be banned on natural terrain, and businesses should be required to install solar panels on their roofs/parking lots.
17	7	We need to set up infrastructures in the open countryside and at sea (photovoltaic, wind power): we need to stop trawling at sea.
17	8	Wind turbine subsidies should be abolished and manufacturers and businesses should be obliged to plant vegetation or install photovoltaic panels/roofs.
17	9	We need to destroy existing wind farms and ban the development of land for solar panels.
17	10	We need to stop clear-cutting, wood-fired power plants, wood heating and solar power plants in the middle of nowhere.
18	1	We need to promote companies' efforts to integrate biodiversity into their activities.
18	2	Manufacturers/companies must commit to restoring biodiversity when their activity has an impact on nature.
18	3	All local authorities must be required to create a «biodiversity restoration» budget line.
18	4	Local authorities need to appropriate and mobilize all the biodiversity protection tools at their disposal.
18	5	We need an ambitious government biodiversity plan, drawn up in partnership with all nature protection associations.
18	6	Companies must be involved in restoring biodiversity
18	7	A share of corporate profits must be mobilized to restore biodiversity
18	8	Any company with a green space around its buildings should be encouraged to act in favor of biodiversity.
18	9	We need to make companies responsible for promoting biodiversity in their ecosystems, through their planning choices.
18	10	Increase public and private subsidies for biodiversity protection and restoration initiatives
19	1	We need to re-establish the value of setting an example of sound, sensible conduct that respects life and the earth's resources.
19	2	We need to become more collectively involved in good ecological gestures and instill a sense of civic responsibility towards nature.
19	3	We need to communicate with people about the consequences of our consumption patterns. To raise awareness.
19	4	We need a paradigm shift, breaking with the concept of punitive ecology, and working towards happy sobriety, consuming less and consuming better.
19	5	We need to do everything we can to change attitudes, at every age. We have to stop the spiral of safe consumption.
19	6	We all need to become aware of our personal and individual impact on the world in order to reduce it through everyday actions.
19	7	We need to reinject human values, dialogue and collective action, get away from the financial yoke and restore our natural environment to its rightful place.
19	8	We need to explain, discuss, share, spread our ideas, experiences and convictions, and above all act «ecologically» in our daily lives.
19	9	Every citizen needs to be informed about what they can do at their own level.
19	10	We all need to think about our individual relationship with this «Nature» of which we are a part (education) in order to become true ambassadors.
20	1	We need to promote quality plant-based food.
20	2	We must reduce the production of animal products, which are responsible for the exploitation of large areas of land and significant pollution.
20	3	We need to promote vegetarianism.
20	4	We absolutely must develop organic farming and plant-based food.
20	5	We must encourage the population to reduce its consumption of meat and fish (health, ecology, animal welfare, we have everything to gain).
20	6	We need to move towards a more plant-based, more local diet, with fewer synthetic inputs that harm the entire ecosystem.
20	7	We need to make it easier for people to eat plant-based diets. Meat and fish consumption are, according to IPBES, major threats.
20	8	Vegan cuisine needs to be improved (taste-wise) and promoted to encourage a reduction in meat consumption and intensive livestock farming.
20	9	We must drastically reduce our consumption of meat and animal products, which are terribly costly ecologically and polluting.
20	10	We need to plant our food!
21	1	Non-selective fishing methods (such as trawling) must be banned, and fishing in protected areas completely prohibited.
21	2	Stop overfishing, stop killing sharks and severely punish trawlers that destroy the seabed with their nets.
21	3	We must ban industrial fishing, which systematically destroys marine ecosystems, and protect species by avoiding overfishing.
21	4	We need to ban large trawlers, which destroy the seabed and fish stocks.
21	5	All marine territories must be protected and only a few (non-destructive) authorized fishing zones must be defined.
21	6	We need to create marine protected areas where fishing and destroying the seabed are effectively prohibited.
21	7	We need to create protected areas, on land and at sea, and defend access to them, as well as small-scale fishing zones that are off-limits to industrial fishing.
21	8	Marine protection zones must be extended and strengthened. Ban mooring for pleasure boats.
21	9	The most destructive fishing practices must be banned => Chalus destroys the seabed
21	10	We need to punish industrial fishing gear that doesn't respect our laws, and ban industrial fishing (demersal seining).
22	1	The cat population must be reduced through a sterilization/coercion program.
22	2	A major campaign to sterilize cats is needed to limit their impact on biodiversity (rodents, reptiles, birds).
22	3	Cats must be spayed or neutered, as they are major bird predators.
22	4	Cats must be spayed or neutered by private individuals, and a national campaign to capture stray cats must be carried out.
22	5	Cats must be sterilized and the proliferation of stray cats, the primary predators of birds and small fauna, must be combated.
22	6	We must penalize the failure to sterilize domestic cats and finance large-scale sterilization campaigns for abandoned cats.
22	7	We need to force cat owners to wear a collar with a name/tel and regulate stray cats, which kill far too many birds.
22	8	Massive sterilization campaigns for stray cats must be developed to limit their predation on birds and rodents.
22	9	A law must be passed requiring municipalities to carry out sterilization campaigns for stray cats (predators of wildlife).
22	10	Cat roaming (massive destruction of small animals and birds) must be banned and punished.
23	1	We need to raise awareness in schools of the need to respect nature and biodiversity.
23	2	We need to promote the importance of biodiversity from the earliest age at school, right through to high school.
23	3	We need to raise children's awareness of biodiversity in schools, with practical work in green spaces, vegetable gardens, mini-farms...
23	4	Biodiversity conservation needs to be integrated into school curricula from elementary school onwards, with nature work camps.
23	5	We need to introduce school courses on biodiversity and the conservation of wild species to better protect and manage their environment.
23	6	Children and teenagers need to be introduced to understanding flora and fauna through "green classes"/discovery days in the countryside.
23	7	Schools need to be encouraged to set up compulsory "green classes" programs to raise awareness of nature and its upkeep.
23	8	Children need to be taught at school from an early age to respect the natural world around them.
23	9	Young people need to be educated about biodiversity (forests, the role of water, insects) in school. Explanation/transmission.
23	10	We need to create and democratize lessons on biodiversity (benefits, dangers, best practices, etc.) from kindergarten onwards.
24	1	Natural forest regeneration must be encouraged, and clear-cutting and planting of single species prohibited.
24	2	We need to stop clear-cutting and over-pruning.
24	3	We need to ban clear-cutting of forests larger than one hectare, and return to reasonable maintenance plans.
24	4	We need to plant trees. Promote agroforestry.
24	5	Clear-cutting and conifer monocultures must be banned, and mixed species and age forests encouraged.
24	6	Ecological forest management must be generalized, with a ban on clear-cutting in favor of planted forest management.
24	7	Forests need to be replanted and placed under protection.
24	8	We need to stop clear-cutting our forests and give priority to respectful maintenance of our existing forests.
24	9	We must limit deforestation and encourage tree replanting
24	10	Woods and forests must be replanted

(continued on next page)

(continued)

Cluster	Rank	Proposal
25	1	Install nests in public gardens for small birds. Monitor and combat invasive animal species
25	2	Provide shelters for wildlife in public parks (bats, hedgehogs, birds).
25	3	Integrated nesting boxes should be compulsory for all new buildings and major renovations (kingbirds, chiropterans, birds of prey, etc.).
25	4	Artificial nesting boxes and towers for swallows, swifts and chiropterans must be installed in our towns and countryside.
25	5	Nesting boxes for birds and bats should be installed to encourage nesting and a return to biodiversity.
25	6	Install nest boxes and provide shelter for birds, hedgehogs, insects and other wildlife.
25	7	Architects must be trained to respect biodiversity and every house must be fitted with birdhouses and hedgehog/bat shelters.
25	8	Birds should be allowed to nest by keeping openings in bell towers, barns, hollow and dead trees, etc.
25	9	We need to make it compulsory for new buildings to include refuge areas for wildlife. E.g.: nesting boxes, shelters, etc.
25	10	Bird nesting boxes should be built into parks (e.g. titmice) and new buildings (e.g. bats).
26	1	We need to create safe migration corridors for wild animals.
26	2	We need to create passageways over freeways or on land to encourage the genetic mixing of large and small animals.
26	3	Wildlife crossings are needed to link fragmented natural areas.
26	4	Roads need to be upgraded on a large scale to allow wildlife to cross safely.
26	5	We need to create wide wilderness corridors so that wild animals can move around freely and with minimum risk.
26	6	We need to increase the number of protected crossings so that animals can cross our roads safely.
26	7	Ecological corridors must be restored to allow wildlife to move between their habitats.
26	8	We need to install more ecological corridors to allow wildlife to cross highways (example of Costa Rica).
26	9	Ecological corridors between natural areas must be developed nationwide to reduce animal mortality on roads
26	10	We need to create more passages under and over freeways and roads so that animals are no longer crushed.
27	1	Building permits must include a rainwater collection tank to supply toilets and washing machines.
27	2	Future houses/buildings must be required to incorporate underground rainwater collection tanks for toilets and watering.
27	3	Water from sinks, washbasins, showers, etc. must be recycled for use in toilets.
27	4	We need to encourage rainwater harvesting, by subsidizing water recovery systems.
27	5	Systematize the installation of water recovery systems in communities and individual homes.
27	6	We need to rethink our use of drinking water, reserving it for cooking and drinking: encourage rainwater harvesting and filtering.
27	7	We need to do away with drinking water in toilet bowls and install rainwater recovery systems in private homes.
27	8	A second water supply must be installed in homes, clean non-drinking water for WCs to limit its use.
27	9	We need to encourage people to recover and use rainwater for everyday activities.
27	10	Building permits for public buildings should be made conditional on the installation of a wastewater recovery system for WCs.
28	1	We need to encourage tall grass to grow in private and public gardens, and limit the use of lawnmowers!
28	2	Communes need to be encouraged to stop mowing grassed areas at the beginning of April. Let the grass grow!
28	3	We need to leave more wild areas, stop building all over the place and encourage private individuals to use differentiated mowing.
28	4	Town councils must be forced to maintain green spaces responsibly (stop mowing lawns with a rotary mower).
28	5	No more razing embankments, no more spring pruning in gardens and towns, no more constant mowing of lawns.
28	6	We should encourage people to stop mowing and replace sad, useless lawns with beds of wild flowers and local shrubs.
28	7	Vegetation should be allowed to grow wherever it doesn't get in the way (roadsides, traffic circles, etc.).
28	8	We need to inform people about the need to leave natural, unmown areas in each landscaped plot to encourage biodiversity.
28	9	Fallow and unmown areas should be left.
28	10	We need to encourage people to stop mowing their lawns, plant nourishing hedges and sow wild flowers.
29	1	Unnecessary public and private lighting must be banned, by switching off at unnecessary times and installing motion detectors.
29	2	We need to limit light pollution: streetlights aimed at the ground, no LEDs, lower intensity at night, etc.
29	3	We need to reduce lighting at night (shop windows lit non-stop, street lighting without passageways, etc.) because it disturbs wildlife at night.
29	4	Light pollution must be reduced.
29	5	We must reduce night-time light pollution and optimize public lighting at night.
29	6	Turn off signs and lights at night in cities to save energy, see the stars and avoid disturbing wildlife.
29	7	We need to ban night lighting, which serves no useful purpose and is harmful to nocturnal wildlife.
29	8	Permanent night-time public and private lighting must be banned (i.e. with motion detectors and only in rare cases).
29	9	Turn off all signs and store windows at night, and limit brightness during the day (too harsh for the eyes).
29	10	Small rural and suburban communities must be obliged to turn off their street lighting at the dead of night.
30	1	Above all, we need to reduce the world's population and abolish pro-birth policies. We need to become aware of our overpopulation!
30	2	We need to limit human demographics by ceasing to encourage births and rewarding couples who have only one child (or 0).
30	3	We must systematically limit the increase in the birth rate and reduce the world's population to a maximum of 3 billion.
30	4	We need to limit the world's population in order to reduce our consumption in the long term.
30	5	We need to limit infinite population growth. There are too many of us, and it makes no sense to have more than 3 children per couple these days.
30	6	We need to reduce the world's human population, the only one that has exploded while the others have collapsed.
		Abolition of benefits
30	7	We must put a stop to human proliferation, which is a harmful and invasive species !!!
30	8	We need to limit human reproduction, the main cause of biodiversity destruction.
30	9	We need to find solutions to limit the human population to 2 billion (maximum reached).
30	10	We need to fight poverty and develop all countries to reduce the number of births and reduce anthropogenic pressure on the Earth.
31	1	Industrial products should be taxed according to their environmental and social impact.
31	2	We need to tax airlines, cruise lines and cargo ships, which are largely responsible for the deterioration of our planet.
31	3	Tourist air travel should be heavily taxed, as it uses excessive amounts of kerosene and chemtrail for the amusement of the wealthy.
31	4	We need to introduce a major tax on people who pollute more than the rest of the population (carbon tax).
31	5	We need to create a real carbon tax: the more a product pollutes (raw material extraction, travel, recycling, etc.), the more it is taxed.
31	6	We need to create a substantial tax for the over-affluent, whose lifestyles consume far more energy and pollute far more than normal, sober people.
31	7	We must tax products that come from far away and/or that do not meet our environmental standards.
31	8	We need to introduce a carbon tax on the biggest polluters.
31	9	We need to (over)tax imports of products that already exist locally or are produced under permissive, ecologically damaging legislation.
31	10	Energy-hungry air travel should be taxed
32	1	The civil service needs to recruit more environmental inspectors, and legal texts need to evolve (sanctions).
32	2	We need to give more power to the environmental authority, whose opinions are merely consultative and ignored by project developers.
32	3	Prefectoral departments must be obliged to respond to whistle-blowers and investigate any failure to comply with the environmental code.
32	4	We need to create an environmental police force to check that all sorting and ecological practices are being carried out in companies.
32	5	We need effective environmental policing

(continued on next page)

(continued)

Cluster	Rank	Proposal
32	6	We need to deploy and fund a specialized police force to deal with environmental and ecological offenses.
32	7	The various environmental police forces must be given greater resources to enforce the environmental code.
32	8	More resources (financial, human, etc.) must be allocated to environmental policing.
32	9	We must limit the power of prefects who can grant multiple exemptions to circumvent environmental regulations.
32	10	We need ethical obligations (CHARTER) for political and administrative leaders in the field of ECOLOGY

## 1.2. Appendix 1.2

Note. Hierarchical Dendrogram of the 32 Topics (Ward's Linkage, Semantic Dissimilarity).

This dendrogram illustrates the hierarchical structure of the dataset based on the semantic distance between the 32 topics. The vertical axis represents the Ward distance, a measure of the loss of information when clusters are merged.

The research team used this visualization as a heuristic tool to compare mathematical proximity with qualitative interpretation. The alignment between the two methods varies:

- **Strong Convergence (Validation):** Several families are clearly reflected in the mathematical structure. For instance, Environmental Education and Awareness is strongly supported by the tight grouping of Clusters 23, 14, and 18 (visible in the grey/olive branch on the right). Similarly, Wildlife Protection shows high consistency, with Clusters 4, 8, 21, and 22 forming a distinct sub-hierarchy, confirming their shared focus on species management.
- **The Case of Niche Topics:** The dendrogram highlights the singularity of Cluster 30, which branches off very early (far left). This mathematical isolation justifies its classification as Emerging and Niche Concerns, as it shares very little vocabulary with the rest of the corpus.
- **Substantive Nuances:** In some instances, the researchers overrode the mathematical grouping to maintain policy relevance. For example, the Agriculture Family. While Clusters 15 and 20 are mathematically close (orange branch), Cluster 1 is grouped by the algorithm with urban planning topics (Clusters 24 and 28). This is likely due to shared technical vocabulary regarding "land use" and "soil". However, the research team kept Cluster 1 within the Agriculture family to prioritize its substantive focus on food systems. Resource Management vs. Urban Spaces: Clusters 7 and 27 are grouped together mathematically but were split qualitatively between Management of Natural Spaces and Resource Management because they address different levels of intervention (spatial planning vs. pollution control).

## Data availability

The authors do not have permission to share data.

## References

- Adams, J.A., White, G., Araujo, R.P., 2022. Mathematical Measures of Societal Polarisation. <https://doi.org/10.48550/ARXIV.2208.05080>.
- Ariai, F., Demartini, G., 2024. Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges. <https://doi.org/10.48550/arXiv.2410.21306>.
- Arthur, D., Vassilvitskii, S., 2006. k-Means++: The Advantages of Careful Seeding.
- Ataman, C., Perrault, S., Tunçer, B., 2025. Big data from online debates to urban design and planning actions: visualizing large-scale human-centric data for post-participation. *Environ. Plann. B Urban Anal. City Sci.*, 23998083251398062 <https://doi.org/10.1177/23998083251398062>.
- Bächtiger, A., Beste, S., 2017. Deliberative citizens, (non) deliberative politicians: a rejoinder. *Daedalus* 146, 106–118.
- Bafna, P., Pramod, D., Vaidya, A., 2016. Document clustering: TF-IDF approach. In: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 61–66. <https://doi.org/10.1109/ICEEOT.2016.7754750>.
- Barraclough, A.D., Cusens, J., Måren, I.E., 2022. Mapping stakeholder networks for the co-production of multiple ecosystem services: a novel mixed-methods approach. *Ecosyst. Serv.* 56, 101461. <https://doi.org/10.1016/j.ecoser.2022.101461>.
- Batterbury, S.P.J., Fernando, J.L., 2006. Rescaling governance and the impacts of political and environmental decentralization: an introduction. *World Dev.* 34, 1851–1863. <https://doi.org/10.1016/j.worlddev.2005.11.019>.
- Bedessem, B., Morère, L., Roblin, L., Dozières, A., Prévot, A.-C., 2022. Participatory biodiversity governance: a comparison of two french initiatives. *Sustainability* 14, 7715. <https://doi.org/10.3390/su14137715>.
- Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S., 2021. On the dangers of stochastic parrots: can language models be too big?. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21. Association for Computing Machinery, New York, NY, USA, pp. 610–623. <https://doi.org/10.1145/3442188.3445922>.
- Bennett, N.J., Satterfield, T., 2018. Environmental governance: a practical framework to guide design, evaluation, and analysis. *Conserv. Lett.* 11, e12600. <https://doi.org/10.1111/conl.12600>.
- Bernaerts, K., Blanckaert, B., Caluwaerts, D., 2023. Institutional design and polarization. Do consensus democracies fare better in fighting polarization than majoritarian democracies? *Democratization* 30, 153–172. <https://doi.org/10.1080/13510347.2022.2117300>.
- Binley, A.D., Proctor, C.A., Pither, R., Davis, S.A., Bennett, J.R., 2021. The unrealized potential of community science to support research on the resilience of protected areas. *Conserv. Sci. Pract.* 3, e376. <https://doi.org/10.1111/csp2.376>.
- Boer, J. de, Aiking, H., 2023. EU citizen support for climate-friendly agriculture (Farm) and dietary options (Fork) across the left-right political spectrum. *Clim. Pol.* 23 (4), 509–521.
- Boiral, O., Heras-Saizarbitoria, I., 2017. Managing biodiversity through stakeholder involvement: why, who, and for what initiatives? *J. Bus. Ethics* 140, 403–421. <https://doi.org/10.1007/s10551-015-2668-3>.
- Bramson, A., Grim, P., Singer, D.J., Berger, W.J., Sack, G., Fisher, S., Flocken, C., Holman, B., 2017. Understanding polarization: meanings, measures, and model evaluation. *Philos. Sci.* 84, 115–159. <https://doi.org/10.1086/688938>.
- Brondizio, E.S., Settele, J., Díaz, S., Ngo, H.T. (Eds.), 2019. The Global Assessment Report of the Intergovernmental Science-policy Platform on Biodiversity and Ecosystem Services. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), Bonn.
- Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language Models Are Few-shot Learners. <https://doi.org/10.48550/arXiv.2005.14165>.
- Buntaine, M.T., Hunnicutt, P., Komakech, P., 2021. The challenges of using citizen reporting to improve public services: a field experiment on solid waste services in Uganda. *J. Public Adm. Res. Theory* 31, 108–127. <https://doi.org/10.1093/jopart/maa026>.
- Caluwaerts, D., Bernaerts, K., Kesberg, R., Smets, L., Spruyt, B., 2023. Deliberation and polarization: a multi-disciplinary review. *Front. Polit. Sci.* 5. <https://doi.org/10.3389/fpos.2023.1127372>.
- Casola, W.R., Beall, J.M., Nils Peterson, M., Larson, L.R., Brent Jackson, S., Stevenson, K. T., 2022. Political polarization of conservation issues in the era of COVID-19: an examination of partisan perspectives and priorities in the United States. *J. Nat. Conserv.* 67, 126176. <https://doi.org/10.1016/j.jnc.2022.126176>.
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., Kurzweil, R., 2018. Universal Sentence Encoder. <https://doi.org/10.48550/arXiv.1803.11175>.
- Chang, A., Tang, T.M., Zikry, T.M., Allen, G.I., 2025. Unsupervised machine learning for scientific discovery: workflow and best practices. *arXiv*. <https://doi.org/10.48550/arXiv.2506.04553>, 5 June.
- Chen, Y., Wang, L., Yu, P., Cheng, C., 2024. Linkages between digital economy and ecosystem services: revealing the spatiotemporal relationship based on U-shaped effect. *Technol. Forecast. Soc. Chang.* 207, 123601. <https://doi.org/10.1016/j.techfore.2024.123601>.
- Cheng, Y., Awan, U., Ahmad, S., Tan, Z., 2021. How do technological innovation and fiscal decentralization affect the environment? A story of the fourth industrial

- revolution and sustainable growth. *Technol. Forecast. Soc. Chang.* 162, 120398. <https://doi.org/10.1016/j.techfore.2020.120398>.
- Dalton, R.J., 2008. The quantity and the quality of party systems: party system polarization, its measurement, and its consequences. *Comp. Pol. Stud.* 41, 899–920. <https://doi.org/10.1177/0010414008315860>.
- De Groot, J.I.M., Steg, L., 2008. Value orientations to explain beliefs related to environmental significant behavior how to measure egoistic, altruistic, and biospheric value orientations. *Environ. Behav.* 40, 330–354.
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Presented at the NAACL-HLT 2019. Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
- Doornbosch, L.M., van Vuuren, M., de Jong, M.D.T., 2025. Moving beyond us-versus-them polarization towards constructive conversations. *Democratization* 32, 96–122. <https://doi.org/10.1080/13510347.2024.2401127>.
- Dryzek, J.S., Bächtiger, A., Chambers, S., Cohen, J., Druckman, J.N., Felicetti, A., Fishkin, J.S., Farrell, D.M., Fung, A., Gutmann, A., Landemore, H., Mansbridge, J., Marien, S., Neblo, M.A., Niemeyer, S., Setälä, M., Slothuus, R., Suiter, J., Thompson, D., Warren, M.E., 2019. The crisis of democracy and the science of deliberation. *Science*. <https://doi.org/10.1126/science.aaw2694>.
- Duclos, J.-Y., Esteban, J., Ray, D., 2004. Polarization: concepts, measurement, estimation. *Econometrica* 72, 1737–1772.
- Eklund, A., Forsman, M., 2022. Topic modeling by clustering language model embeddings: human validation on an industry dataset. In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track*. Presented at the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track. Association for Computational Linguistics, Abu Dhabi, UAE, pp. 635–643. <https://doi.org/10.18653/v1/2022.emnlp-industry.65>.
- Esteban, J.-M., Ray, D., 1994. On the measurement of polarization. *Econometrica* 62, 819–851. <https://doi.org/10.2307/2951734>.
- European Commission. Directorate General for Environment, 2021. *EU Biodiversity Strategy for 2030: Bringing Nature Back Into Our Lives*. Publications Office, LU.
- Fariello, G., Jemielniak, D., 2025. The changing language and sentiment of conversations about climate change in Reddit posts over sixteen years. *Commun. Earth Environ.* 6, 1–9. <https://doi.org/10.1038/s43247-024-01974-8>.
- Fisher, B., Turner, R.K., Morling, P., 2009. Defining and classifying ecosystem services for decision making. *Ecol. Econ.* 68, 643–653. <https://doi.org/10.1016/j.ecolecon.2008.09.014>.
- French Government - Directorate-General for Planning, Housing and Nature, 2023. *National Biodiversity Strategy 2030*.
- Gao, T., Yao, X., Chen, D., 2022. SimCSE: Simple Contrastive Learning of Sentence Embeddings. <https://doi.org/10.48550/arXiv.2104.08821>.
- Gibson, K., Lamm, A., Woosnam, K., Croom, D., 2021. Engaging the public in water policy: do political affiliation and ideology matter? *J. Contemp. Water Res. Educ.* 173. <https://doi.org/10.1111/j.1936-704X.2021.3355.x>.
- Greco, C.M., Tagarelli, A., 2024. Bringing order into the realm of Transformer-based language models for artificial intelligence and law. *Artif. Intell. Law* 32, 863–1010. <https://doi.org/10.1007/s10506-023-09374-7>.
- Grimmer, J., Stewart, B.M., 2013. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Polit. Anal.* 21, 267–297. <https://doi.org/10.1093/pan/mps028>.
- Grybauskas, A., Cárdenas-Rubio, J., 2024. Unlocking employer insights: using large language models to explore human-centric aspects in the context of industry 5.0. *Technol. Forecast. Soc. Chang.* 208, 123719. <https://doi.org/10.1016/j.techfore.2024.123719>.
- Guber, D.L., 2013. A cooling climate for change? Party polarization and the politics of global warming. *Am. Behav. Sci.* 57, 93–115. <https://doi.org/10.1177/0002764212463361>.
- Haines-Young, R., Potschin, M., 2018. *Common International Classification of Ecosystem Services (CICES) V5.1 and Guidance on the Application of the Revised Structure*.
- Harris, N., Butani, A., Hashmy, S., 2024. Enhancing Embedding Performance Through Large Language Model-based Text Enrichment and Rewriting. <https://doi.org/10.48550/arXiv.2404.12283>.
- Hoffarth, M.R., Hodson, G., 2016. Green on the outside, red on the inside: perceived environmentalist threat as a factor explaining political polarization of climate change. *J. Environ. Psychol.* 45, 40–49. <https://doi.org/10.1016/j.jenvp.2015.11.002>.
- Iten, A., Mouter, N., 2022. When digital mass participation meets citizen deliberation: combining mini- and maxi-publics in climate policy-making. *Sustainability* 14, 4656. <https://doi.org/10.3390/su14084656>.
- Iyengar, S., Sood, G., Lelkes, Y., 2012. Affect, not ideology: a social identity perspective on polarization. *Public Opin. Q.* 76, 405–431. <https://doi.org/10.1093/poq/nfs038>.
- Jacobi, C., Van Atteveldt, W., Welbers, K., 2016. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digit. Journal.* 4, 89–106. <https://doi.org/10.1080/21670811.2015.1093271>.
- Jost, J.T., Baldassarri, D.S., Druckman, J.N., 2022. Cognitive-motivational mechanisms of political polarization in social-communicative contexts. *Nat. Rev. Psychol.* 1, 560–576. <https://doi.org/10.1038/s44159-022-00093-5>.
- Judge, M., Kashima, Y., Steg, L., Dietz, T., 2023. Environmental decision-making in times of polarization. *Annu. Rev. Environ. Resour.* 48, 477–503. <https://doi.org/10.1146/annurev-environ-112321-115339>.
- Kennedy, E.H., Muzzerall, P., 2022. Morality, emotions, and the ideal environmentalist: toward a conceptual framework for understanding political polarization. *Am. Behav. Sci.* 66, 1263–1285. <https://doi.org/10.1177/00027642211056258>.
- Keraghel, I., Morbieu, S., Nadif, M., 2024. Beyond words: a comparative analysis of LLM embeddings for effective clustering. In: *Advancing Intelligent Data Analysis*. IOS Press.
- Kiss, B., Sekulova, F., Hörschelmann, K., Salk, C.F., Takahashi, W., Wamsler, C., 2022. Citizen participation in the governance of nature-based solutions. *Environ. Policy Gov.* 32, 247–272. <https://doi.org/10.1002/eet.1987>.
- Kitamori, K., 2025. Citizen Engagement Is Key to a Successful Green Transition. *World Economic Forum*.
- Kusupati, A., Bhatt, G., Rege, A., Wallingford, M., Sinha, A., Ramanujan, V., Howard-Snyder, W., Chen, K., Kakade, S., Jain, P., Farhadi, A., 2024. Matryoshka Representation Learning. <https://doi.org/10.48550/arXiv.2205.13147>.
- Lafont, C., 2015. Deliberation, participation, and democratic legitimacy: should deliberative mini-publics shape public policy? *J. Polit. Philos.* 23, 40–63. <https://doi.org/10.1111/jopp.12031>.
- Larosa, F., Hoyas, S., Conejero, H., Garcia-Martinez, J., Nerini, F., Vinuesa, R., 2025. Large Language Models in Climate and Sustainability Policy: Limits and Opportunities. <https://doi.org/10.48550/arXiv.2502.02191>.
- Lau, J.H., Newman, D., Baldwin, T., 2014. Machine reading tea leaves: automatically evaluating topic coherence and topic model quality. In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Presented at the Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Gothenburg, Sweden, pp. 530–539. <https://doi.org/10.3115/v1/E14-1056>.
- Leighninger, M., 2006. *The Next Form of Democracy: How Expert Rule is Giving Way to Shared Governance-and why Politics Will Never be the Same*. Vanderbilt University Press.
- Lucas, C., Warman, R., 2018. Disrupting polarized discourses: can we get out of the ruts of environmental conflicts? *Environ. Plan. C. Polit. Space* 36. <https://doi.org/10.1177/2399654418772843>.
- Maczka, K., Matczak, P., Jeran, A., Chmielewski, P.J., Baker, S., 2021. Conflicts in ecosystem services management: analysis of stakeholder participation in Natura 2000 in Poland. *Environ. Sci. Pol.* 117, 16–24. <https://doi.org/10.1016/j.envsci.2021.01.001>.
- Matzek, V., Wilson, K.A., 2021. Public support for restoration: does including ecosystem services as a goal engage a different set of values and attitudes than biodiversity protection alone? *PLoS One* 16, e0245074. <https://doi.org/10.1371/journal.pone.0245074>.
- Menon, S., Hartz-Karp, J., 2019. Institutional innovations in public participation for improved local governance and urban sustainability in India. *Sustain. Earth* 2, 6. <https://doi.org/10.1186/s42055-019-0013-x>.
- Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. *Efficient Estimation of Word Representations in Vector Space*. arXiv Preprint arXiv:1301.3781.
- Mohri, H., Lahoti, S., Saito, O., Mahalingam, A., Gunatilleke, N., Irahm, Hoang, V.T., Hitinayake, G., Takeuchi, K., Herath, S., 2013. Assessment of ecosystem services in homegarden systems in Indonesia, Sri Lanka, and Vietnam. *Ecosyst. Serv.* 5, 124–136. <https://doi.org/10.1016/j.ecoser.2013.07.006>.
- Moulavi, D., Andretta Jaskowiak, P., Campello, R., Zimek, A., Sander, J., 2014. Density-based Clustering Validation. <https://doi.org/10.1137/1.9781611973440.96>.
- Muennighoff, N., Tazi, N., Magne, L., Reimers, N., 2023. MTEB: Massive Text Embedding Benchmark. In: Vlachos, A., Augenstein, I. (Eds.), *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Presented at the EAACL 2023. Association for Computational Linguistics, Dubrovnik, Croatia, pp. 2014–2037. <https://doi.org/10.18653/v1/2023.eacl-main.148>.
- Newig, J., Challies, E., Jager, N.W., Kochskämper, E., Adzersen, A., 2018. The environmental performance of participatory and collaborative governance: a framework of causal mechanisms. *Policy Stud. J.* 46, 269–297. <https://doi.org/10.1111/psj.12209>.
- Newig, J., Jager, N.W., Challies, E., Kochskämper, E., 2023. Does stakeholder participation improve environmental governance? Evidence from a meta-analysis of 305 case studies. *Glob. Environ. Chang.* 82, 102705. <https://doi.org/10.1016/j.gloenvcha.2023.102705>.
- OpenAI, 2025. Platform: models. URL, OpenAI. <https://platform.openai.com>. (Accessed 27 January 2025) (WWW Document).
- Özler, Ş.İ., Obach, B., 2018. Polarization and the environmental movement in Turkey. *J. Civ. Soc.* 14 (4), 311–327.
- Pejic-Bach, M., Bertonecel, T., Meško, M., Krstić, Ž., 2020. Text mining of industry 4.0 job advertisements. *Int. J. Inf. Manag.* 50, 416–431. <https://doi.org/10.1016/j.ijinfomgt.2019.07.014>.
- Pennington, J., Socher, R., Manning, C., 2014. Glove: Global Vectors for Word Representation. *EMNLP*. <https://doi.org/10.3115/v1/D14-1162>.
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep Contextualized Word Representations. <https://doi.org/10.48550/arXiv.1802.05365>.
- Petukhova, A., Matos-Carvalho, J.P., Fachada, N., 2025. Text clustering with large language model embeddings. *Int. J. Cognit. Comput. Eng.* 6, 100–108. <https://doi.org/10.1016/j.ijcce.2024.11.004>.
- Picanço Rodrigues, V., Leonel Caetano, M.A., 2023. The impacts of political activity on fires and deforestation in the Brazilian Amazon rainforest: an analysis of social media and satellite data. *Heliyon* 9, e22670. <https://doi.org/10.1016/j.heliyon.2023.e22670>.

- Qaiser, S., Ali, R., 2018. Text mining: use of TF-IDF to examine the relevance of words to documents. *IJCA* 181, 25–29. <https://doi.org/10.5120/ijca2018917395>.
- Rehm, P., Reilly, T., 2010. United we stand: constituency homogeneity and comparative party polarization. *Elect. Stud.* 29, 40–53. <https://doi.org/10.1016/j.electstud.2009.05.005>.
- Reimers, N., Gurevych, I., 2019. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. <https://doi.org/10.48550/arXiv.1908.10084>.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>.
- Ryfe, D.M., 2005. Does deliberative democracy work? *Annu. Rev. Polit. Sci.* 8, 49–71. <https://doi.org/10.1146/annurev.polisci.8.032904.154633>.
- Saldaña, J., 2021. *The Coding Manual for Qualitative Researchers*.
- Schreier, M., 2012. *Qualitative Content Analysis in Practice*. SAGE Publications Ltd. <https://doi.org/10.4135/9781529682571>.
- Schröter, M., van der Zanden, E.H., van Oudenhoven, A.P.E., Remme, R.P., Serna-Chavez, H.M., de Groot, R.S., Opdam, P., 2014. Ecosystem services as a contested concept: a synthesis of critique and counter-arguments. *Conserv. Lett.* 7, 514–523. <https://doi.org/10.1111/conl.12091>.
- Shah, T.M., 2025. Trends in political science research: affective polarization. *Int. Polit. Sci. Abstr.* 75, 1–14. <https://doi.org/10.1177/00208345251323534>.
- Swette, B., Huntsinger, L., Lambin, E.F., 2023. Collaboration in a polarized context: lessons from public forest governance in the American West. *Ecol. Soc.* 28. <https://doi.org/10.5751/ES-13958-280129>.
- Thota, H., Moh, M., Moh, T.-S., 2023. Towards Detecting and Quantifying Identity-based Polarization in Online Content: A Deep-learning Approach. Presented at the 2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). IEEE Computer Society, pp. 593–599. <https://doi.org/10.1109/WI-IAT59888.2023.00098>.
- Tindale, S., Vicario-Modroño, V., Gallardo-Cobos, R., Hunter, E., Miškolci, S., Price, P.N., Sánchez-Zamora, P., Sonneveld, M., Ojo, M., McInnes, K., Frewer, L.J., 2023. Citizen perceptions and values associated with ecosystem services from European grassland landscapes. *Land Use Policy* 127, 106574. <https://doi.org/10.1016/j.landusepol.2023.106574>.
- Townley, C., Koop, C., 2024. Exploring the potential and limits of digital tools for inclusive regulatory engagement with citizens. *Gov. Inf. Q.* 41, 101901. <https://doi.org/10.1016/j.giq.2023.101901>.
- Turnhout, E., Metzke, T., Wyborn, C., Klenk, N., Louder, E., 2020. The politics of co-production: participation, power, and transformation. *Curr. Opin. Environ. Sustain.* 42, 15–21. <https://doi.org/10.1016/j.cosust.2019.11.009>.
- Van der Eijk, C., 2001. Measuring agreement in ordered rating scales. *Qual. Quant.* 35, 325–341.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need. <https://doi.org/10.48550/arXiv.1706.03762>.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L.A., Isaac, W., Legassick, S., Irving, G., Gabriel, I., 2021. Ethical and Social Risks of Harm From Language Models. <https://doi.org/10.48550/arXiv.2112.04359>.
- Young, O.R., 2016. *On Environmental Governance*, 1st ed. Routledge. <https://doi.org/10.4324/9781315633176>.
- Zhao, J., Syed, T.A., 2026. Making sense of algorithms: workers' coping strategies for algorithmic management on digital platforms. *Technol. Forecast. Soc. Chang.* 223, 124443. <https://doi.org/10.1016/j.techfore.2025.124443>.
- Zhao, L., Zhang, L., Sun, J., He, P., 2022. Can public participation constraints promote green technological innovation of Chinese enterprises? The moderating role of government environmental regulatory enforcement. *Technol. Forecast. Soc. Chang.* 174, 121198. <https://doi.org/10.1016/j.techfore.2021.121198>.

**Magali Trelohan** is Associate Professor at Marketing at EM Normandie Business School. She works on consumer behavior, pro-environmental behaviors and gender issues. Within these themes, she is particularly interested in social norms, engagement, empowerment, and disengagement. She has published in international journals including *Journal of Business Research* and *Journal of Marketing Management*.

**Anne-Cécile Gay** is a data scientist, artificial intelligence engineer, and specialist in natural language processing. Her applied research focuses on the analysis and modeling of textual data, with a particular interest in hybrid approaches combining machine learning and linguistic methods. She is especially interested in techniques for interpreting NLP models, the structuring and exploitation of specialized corpora, and the integration of generative models into decision-support systems. She has published in international journals including *Journal of Business Research*.

**Damien Chaney** is Professor of Marketing at EM Normandie Business School. He works on consumer experience, consumer-brand relationship, and the social evaluation of markets. He has published in international journals including *Journal of Service Research*, *Journal of Business Ethics*, *Technological Forecasting and Social Change*, *Journal of Business Research* and *Annals of Tourism Research*.

**David Moroz** is an Associate Professor at EM Normandie Business School. His research focuses on alternative investments, second-hand markets, and bibliometric analysis. He has published in international journals such as *Economic Modelling*, *Defence and Peace Economics*, and the *Journal of Agricultural Economics*.