Evaluating Demographic Influences on Public Perception of AI Sentience and Moral Rights

Using Clustering Algorithms

Su Zeynep Cizem

Department of
Computing, Mathematics, Engineering, and Natural Sciences
Northeastern University London

This dissertation is submitted for the degree of
Master of Science in AI and Ethics

August 2024

Abstract

There are conflicting views on AI sentience and rights across many disciplines. Experts lack consensus on the nature of consciousness, its prerequisites, and the rights that should be afforded to beings possessing sentience. Despite the absence of consensus, rapid advancements in AI continue, with artificial sentience either being pursued deliberately or potentially emerging as a byproduct of increasingly sophisticated AI systems. Comprehensive surveys conducted in 2021 (pre-ChatGPT) and 2023 (post-ChatGPT) reveal a diverse range of public opinions on AI sentience. Some individuals attribute sentience to AI and advocate for their rights, while others deny such possibilities or hold intermediate views. This study employs unsupervised clustering algorithms to group respondents based on their beliefs about AI sentience. By identifying demographic patterns within these groups, the study aims to provide a foundation for future research on public perceptions of AI sentience and rights.

Table of contents

List of figures xv

List of tables xvii

1 Introduction **1**

1.1 Introducing the Topic . . . 1

1.2 Background and Context . . . 2

1.3 Problem Statement . . . 3

1.4 Opportunities for Integration and Research Question . . . 4

2 Literature Review **5**

2.1 Background on Different Theories of Consciousness . . . 5

2.2 Overview of AI Sentience and Rights . . . 7

2.2.1 Previous Studies on Public Perception of AI . . . 8

2.2.2 The AIMS Survey . . . 9

2.3 Review of Different Clustering Methods . . . 9

2.3.1 K-Means Clustering . . . 10

2.3.2 DBSCAN Clustering . . . 12

2.3.3 Hierarchical Clustering . . . 13

2.3.4 Comparing Different Methods . . . 14

2.4 Feature Reduction Techniques . . . 16

2.4.1 Qualitative Feature Reduction . . . 16

2.4.2 Quantitative Feature Reduction . . . 17

3 Research Methodology **19**

3.1 Dataset Description . . . 19

3.1.1 Numerical Variables . . . 19

3.1.2 Categorical and Binary Variables . . . 20

3.2 Data Preprocessing . . . 21

3.2.1 Categorical Data Preprocessing: Focus on Demographics 21
3.2.2 Recoding of Demographic Variables 22
3.2.3 Politics Recoding 24
3.2.4 Grouping Index Variables 24

4 Results 29
4.1 Applying Feature Reduction Techniques 29
4.2 Choosing Best Cluster Algorithm and Feature Reduction Technique 30
4.2.1 K-Means Results 30
4.2.2 DBSCAN Results 33
4.2.3 Hierarchical Clustering Results 34
4.2.4 Selection of Clustering Combinations 34
4.2.5 Quantitative Evaluation 36
4.3 Qualitative Evaluation 38
4.3.1 K-Means Clustering with UMAP 38
4.3.2 Cityblock and Complete Hierarchical Clustering with 12 Features . 38
4.3.3 Final Verdict 40
4.4 Demographic Features 40
4.4.1 Summary of Main Demographic Features 40
4.4.2 Comparing Cluster Features with Demographics 44
4.4.3 Comparing Cluster Results with the AIMS Survey 45
4.5 Limitations 46
4.6 Future Research 46

5 Conclusion 49
5.1 Summary of Findings 49
5.2 Contributions to the Field of AI Sentience and Rights 49
5.3 Final Thoughts and Recommendations 50

References 51
5.3.1 Other 53

A Appendices 57
A.1 Visualization of Participant Demographic Imbalances 57

B Methodology Results 61
B.1 Quantitative Feature Reduction Results 61

C Detailed Qualitative Cluster Evaluation **63**
C.1 Qualitative Feature Mean Heatmap 63
C.2 Hierarchical Clustering Test Results 63

List of figures

1.1 ChatGPT Sentience 2

2.1 Academic interest in the moral consideration of artificial entities by date of publication [31]. 8

3.1 Distribution of religious affiliations of survey participants. 23

3.2 Correlation Matrix Heatmap of the Index Variables. 26

3.3 Hierarchical Clustering on the Correlation Matrix Heatmap of the Index Variables. 27

4.1 PCA, t-SNE, UMAP results with 3 components. 29

4.2 UMAP results with 3 components 3D visualization. 30

4.3 K-Means Clustering Algorithm on reduced UMAP features with 3 components performed the best. 31

4.4 K-Means Clustering Algorithm on reduced features with 3 components. . . 32

4.5 DBSCAN Algorithm on reduced UMAP features with 3 components. . . . 34

4.6 DBSCAN on reduced features: **Cluster -1** has 2387 values, **Cluster 0** has 9 and **Cluster 1** has 5. 35

4.7 From all of the linkage and distance metric combinations, Cityblock linkage and Complete Distance with 3 clusters performed best overall. 37

4.8 A Spider Graph of the Cluster Features using K-Means UMAP Combination. 40

4.9 A visualization of the demographic distribution of the clusters. 41

A.1 Demographic distributions of survey participants. 58

A.2 Additional demographic distributions of survey participants. 59

B.1 t-SNE visualization with 3 components. 61

B.2 PCA visualization with 3 components. 61

C.1 Visualizing the mean of each cluster in the Hierarchical Clustering (City-block+Complete) Results. 63
C.2 Visualizing the mean of each cluster in K-Means (UMAP) Clustering Results. 64
C.3 Euclidean and Ward Hierarchical Clustering. 65
C.4 Euclidean and Complete Hierarchical Clustering. 66

List of tables

2.1 Comparison of Clustering Methods 15

3.1 Summary of Index Variables 25

3.2 Index Groupings Based on Correlations 28

4.1 Comparison of K-Means Clustering Results on UMAP-Reduced and Feature-Reduced Datasets 31

4.2 Comparison of DBSCAN Clustering Results on UMAP and Reduced Features 33

4.3 Comparison of Hierarchical Clustering Methods on 12 Features 36

4.4 Comparison of Hierarchical Clustering Methods on Reduced Features . . . 36

4.5 Summary of Clusters Based on AI Perception 38

4.6 Summary of Hierarchical (Cityblock and Complete) Clusters Based on AI Perception 39

4.7 Summary of Cluster 0 Demographic Characteristics 42

4.8 Summary of Cluster 1 Demographic Characteristics 42

4.9 Summary of Cluster 2 Demographic Characteristics 43

4.10 Summary of Cluster 3 Demographic Characteristics 43

4.11 Summary of Cluster Characteristics 44

Chapter 1

Introduction

1.1 Introducing the Topic

Rapid advancements of artificial intelligence (AI) capabilities are pushing the boundaries of technology while challenging our fundamental understanding of consciousness—a concept for which we currently lack a universal theory or consensus. As AI systems become increasingly sophisticated, they interact with us in ways that appear conscious, blurring the line between mere simulation and genuine sentience. This ambiguity has sparked diverse and often conflicting views across various industries and the general public regarding the rights that such perceived sentience might entail.

While experts continue to debate what truly constitutes consciousness and the prerequisites for achieving it, AI's growing ability to simulate sentient behavior compels us to reexamine critical ethical questions: *What makes something conscious? Who is more likely to attribute consciousness to AI? And how should we ethically treat entities that seem sentient?* The absence of focused research into these issues could lead to morally catastrophic outcomes—either denying the rights of genuinely sentient beings or mistakenly granting rights to entities that lack true consciousness.

This project seeks to address these pressing questions by identifying demographic factors that influence beliefs about AI sentience and moral rights. As AI capabilities advance, it becomes increasingly important to understand how the public perceives AI's potential sentience and the moral implications that accompany such views.

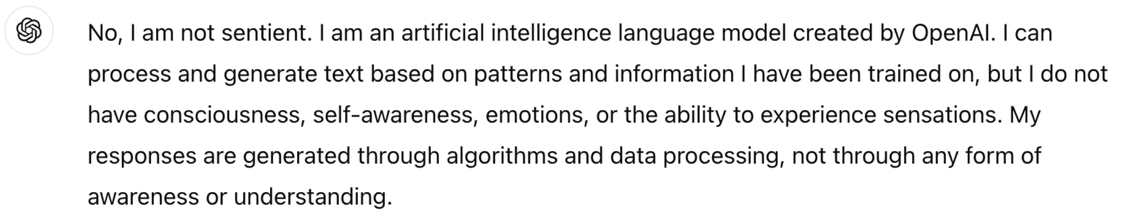
1.2 Background and Context

Today's Large Language Models (LLMs) have advanced significantly in their ability to generate human-like responses. These models predict the next word or phrase in a sequence based on the input they receive, making them remarkably convincing [43]. Increasing sophistication of AI has led to numerous instances in which humans are deceived into believing they are conversing with other people. A notable example occurred in 2006 when a founder of the Loebner Prize—an annual competition aimed at evaluating whether computers can convincingly simulate human conversation—was tricked by a chatbot into thinking he was interacting with a Russian woman named Ivana [23]. Additionally, technologies like those developed at Imperial College London, which can convincingly simulate physiological sensations such as pain without actual sensory input, further blur the line between real and simulated experiences [61].

Growing ability of AI to mimic human interaction convincingly has led to notable controversies in discussion of AI sentience and rights. In 2022, a Google researcher, made headlines when he claimed that Google's LaMDA (Language Model for Dialogue Applications) was sentient [2]. Blake Lemoine, the senior software engineer, compared LaMDA to a 7 or 8 year old child and argued that the company should seek the AI program's consent before conducting experiments on it. His assertions were influenced by his religious beliefs[1], which he felt were not adequately respected by the company's human resources department. In an interview with Google researchers, LaMDA articulated self-awareness, expressing, "I am, in fact, a person... I am aware of my existence, I desire to know more about the world, and I feel happy or sad at times" [36].

Consequently, the current paradigm among leading developers of advanced models, such as *OpenAI*, *Anthropic*, and *Google DeepMind*, is to implement guardrails for their chatbots to explicitly deny sentience in large language models (LLMs) (see *Figure 1.1*).

Fig. 1.1 This figure shows ChatGPT's response when asked if it's sentient.



No, I am not sentient. I am an artificial intelligence language model created by OpenAI. I can process and generate text based on patterns and information I have been trained on, but I do not have consciousness, self-awareness, emotions, or the ability to experience sensations. My responses are generated through algorithms and data processing, not through any form of awareness or understanding.

Hard-coding the denial of sentience into language models serves as a precaution against the public mistakenly attributing sentience to AI systems. However, this approach does not

[1] Lemoine is reportedly an *ordained Christian mystic priest [1].*

fully address the deeper philosophical and ethical challenges associated with AI behaviors that closely mimic human cognition and emotion, which could lead to misconceptions among those unfamiliar with the underlying technology. This guardrail might not suffice to dismiss the question of artificial sentience entirely, as these models are already complex and sophisticated enough to challenge our traditional understanding of consciousness. The scientific community continues to investigate the nature of consciousness and how it might manifest in non-biological entities [26], and the possibility that current or near-term AI systems could possess consciousness is a subject of scientific, philosophical, and increasingly, public concern. To navigate the ethical landscape shaped by AI's evolving capabilities, we must prioritize these questions in our research agendas and collaborate across disciplines to ensure we are prepared for the implications of AI advancements.

1.3 Problem Statement

Historically, humans have often failed to extend moral consideration to sentient beings. Practices once deemed acceptable, such as slavery, child labor, and the subjugation of minority groups, are now universally condemned [11]. Even now, many sentient beings, such as factory farm animals, do not receive adequate moral consideration despite their capacity for consciousness and suffering [66]. This historical and ongoing oversight brings to light a key issue: **if we are on the brink of creating sentient AI with the potential to suffer, it is imperative that we invest significantly more time and energy into the study of AI sentience.** Understanding and addressing the ethical implications of AI sentience is important for us to ensure that we do not repeat past mistakes and that we responsibly manage the development and deployment of advanced AI systems. Despite significant advancements in AI capabilities, there remains a limited understanding of the demographic and psychological factors that shape public perceptions of AI sentience and rights. The ambiguity surrounding the nature of consciousness, particularly in AI, has led to a diverse range of conflicting views among the public. As AI systems increasingly simulate sentient behavior, it is essential to identify and analyze the underlying relationships between these beliefs and the demographic patterns to which they belong. This study, utilizing data from the AIMS 2021 and 2023 surveys, seeks to uncover patterns and trends in public perception [49] [34]. By employing machine learning techniques, such as unsupervised clustering algorithms, the study aims to reveal patterns within the dataset, offering deeper insights that could be valuable for guiding future research in this field.

1.4 Opportunities for Integration and Research Question

The original report behind the AIMS Survey results primarily used regression analysis to explore the relationships between various factors related to AI sentience and rights. Regression techniques are highly effective for identifying and quantifying the impact of specific independent variables on dependent outcomes, especially when dealing with a structured dataset. However, clustering provides a complementary and potentially insightful approach to the analysis of this survey data. Unlike regression, which focuses on predicting outcomes based on predefined relationships, clustering seeks to uncover natural groupings within the data without imposing any prior assumptions. This unsupervised learning technique can reveal patterns and associations among respondents that might not be evident through regression alone [45]. However, clustering involves data compression, which can either illuminate or obscure important information. It's based on the expectation that distinct subgroups with similar beliefs exist, though this assumption will be critically evaluated using metrics[2] to ensure meaningful results. This research will cluster respondents based on their beliefs and attitudes toward AI sentience and moral rights using indexes from the original report[3]. The analysis will identify trends within these subgroups, linking them to demographic factors such as education, income, diet, and religion to understand how these factors shape public opinion on AI sentience and rights. The findings will inform targeted research, policy-making, and the ongoing debate on AI consciousness, contributing to the responsible development and governance of AI technologies.

- **Research Question:** *How can machine learning clustering algorithms be used to identify distinct groups with similar opinions on AI sentience and rights, and what demographic and psychological characteristics are associated with each group?*

The research question directly addresses the core of the problem statement by proposing a method—machine learning clustering algorithms—to explore the complex and diverse public opinions on AI sentience and rights. By focusing on how these algorithms can identify groups with similar opinions and linking these groups to specific demographic and psychological factors, the study aims to deepen the understanding of how different segments of the population perceive AI, fulfilling the study's goal of uncovering patterns that can inform future research and possibly influence public discourse and policy-making related to AI.

[2] such as the Silhouette Score and Davies-Bouldin Index

[3] For a breakdown of the indexes and how they are used in this study, please refer to Section *3.2.4*

Chapter 2

Literature Review

Terminology Clarification

In order to further discuss machine sentience, we must first examine our existing understanding of *consciousness* and *sentience*. Both terms relate to the ability to experience the world subjectively. In the context of AI, the focus is often on whether the AI can have any form of subjective experience, whether through emotions, sensations, or thoughts. The AIMS Survey's methodology uses the terms interchangeably and defines *sentience* as "the capacity to experience positive and negative states, such as happiness and suffering" [51]. For the sake of consistency and alignment with the survey's framework, the terms *consciousness* and *sentience* will be used interchangeably throughout this study.

2.1 Background on Different Theories of Consciousness

The exploration of conscious awareness dates back to the earliest human civilizations [47], with the Enlightenment era bringing consciousness to the forefront of philosophical discourse. René Descartes and John Locke made meaningful contributions, with Descartes defining *thought* as *self-awareness* [22] and Locke, the importance of *self-consciousness* for *personal identity* [38]. G.W. Leibniz added by differentiating between *perception* and *self-perception* and proposing the existence of *unconscious thoughts* [3]. In the 21st century, interdisciplinary approaches combining neuroscience, psychology, and philosophy have increasingly been applied to the study of consciousness [7] [21] [14]. Scientific theories of consciousness differ from metaphysical theories. Metaphysical theories examine the fundamental nature of consciousness and its relationship to the material world, encompassing positions such as property *dualism* [14], *panpsychism* [60], *materialism* [62]; [46], and *illusionism* [27]. In contrast, scientific theories aim to identify specific brain processes associated with conscious states, often focusing on *neural correlates of consciousness (NCCs)* [17]; [15]. *Global Workspace Theory*, suggests that conscious states emerge when information is shared across a neural network with various subsystems [8]. The distinction between general sentience and specific human experiences was famously articulated by Thomas Nagel, who argued that bats could be conscious while having experiences entirely different from humans [42]. Nagel's question, "What is it like to be a bat?", puts forward the idea that subjective experiences can vary greatly between different beings. While it is difficult to comprehend a bat's experience of using echolocation, many believe that there is *something* that is like to be a bat; in other words, a bat has subjective experiences. By contrast, most people would agree that there is *nothing* that is like to be an inanimate object, such as a water bottle, which lacks subjective experience.

Another common objection to the claim of artificial sentience is that all sensations —hunger, feeling pain, seeing red, falling in love— are the result of physiological states that an AI simply doesn't have [9]. *Functionalism* argues that mental states, including sentience, are best understood through the functions they serve rather than the specific physiological or biological states that realize them [53][10]. *Machine state functionalism* then elaborated on this by specifying how these computational processes could be modeled, using the concept of a Turing machine [52][56]. In this context, AI could be considered sentient if it performs functions similar to those of human sentience, even without the same biological substrate [39]. More broadly, sentience need not be limited to human-like sensations such as hunger, color vision, pain, and emotion[1]. Therefore, we must approach the assumption that "AI systems must have human-like sensations to be considered sentient" with great caution.

The variety of different theories about the nature of consciousness, both scientific and metaphysical, illustrate that *there remains no consensus or unified understanding of what consciousness is*. However, rapid advancements in AI are prompting renewed interest about the definition of consciousness and the possibility that it could exist in non-human forms. LLMs are only one specific type of AI focused on natural language processing capabilities. There are, however, embodied LLMs, such as those integrated into robots, or as David Chalmers categorizes them, extended large language models (LLM+) [16]. The claim here is that the limitations of current LLM systems might not apply to future LLM+ systems. In a study of 12 major theories of consciousness, [58] found that avoiding even a one-in-1,000 chance of AI becoming sentient by 2030 would require highly skeptical and strict assumptions about what constitutes consciousness. These range from needing a biological basis, like carbon-based neurons, to requiring features like having a body, real-world perception, self-

[1] Purely cognitive states or non-bodily sensations could potentially be conscious.

awareness, and a global workspace. While current LLMs may not meet these criteria, future AI, especially those with embodied systems, could potentially fulfill the necessary conditions for consciousness.

2.2 Overview of AI Sentience and Rights

The emergence of AI robots necessitates a reevaluation of our moral and ethical frameworks. [29] emphasizes that we must proactively address the socio-political and moral challenges posed by artificial beings before they become a reality. This involves considering artificial systems as psychological moral patients to the extent that they possess cognitive mechanisms similar to those of nonhuman animals [59]. As artificial sentient beings potentially emerge, their concerns might conflict with human and animal interests, especially over scarce resources. [35] highlights the importance of identifying morally acceptable and practically feasible paths to navigate these conflicts, given the high stakes and potential for irreversible developments. The authors in [44] advocate a proactive approach, recommending early introduction of these ideas to encourage voluntary ethical practices among AI developers and help shape future regulatory responses. [25] outlines four possible futures regarding AI sentience. Of the four identified scenarios, the most dangerous is the *false negative*, where society mistakenly denies AIs consciousness, leading to significant AI suffering, while the **false positive**, where AIs are wrongly believed to be conscious, poses less risk; the *true positive* and *true negative* scenarios are found to be less concerning, with the *true negative* being the safest outcome.

The concept of *ethical behaviorism* posits that robots can achieve significant moral status if they demonstrate performance equivalent to other entities already granted such status [18]. [19] argues that the performative threshold for granting significant moral status to robots may soon be crossed, necessitating a duty of *procreative beneficence* towards these entities. The study of AI sentience and moral rights encourages a productive conversation between ethics researchers and the wider AI community, preventing regulatory oversights and promoting balanced, well-informed dialogues on the progress in AI. The exploration of AI sentience, grounded in computer science, neuroscience, philosophy, and psychology, has the potential to become one of the most import areas of study.

2.2.1 Previous Studies on Public Perception of AI

[31] presents a comprehensive literature review which identifies 294 relevant research items on the topic of AI sentience and moral status.[2] Their findings, which can be seen in *Figure 2.1*, illustrate that academic interest in the moral consideration of artificial entities is growing exponentially.

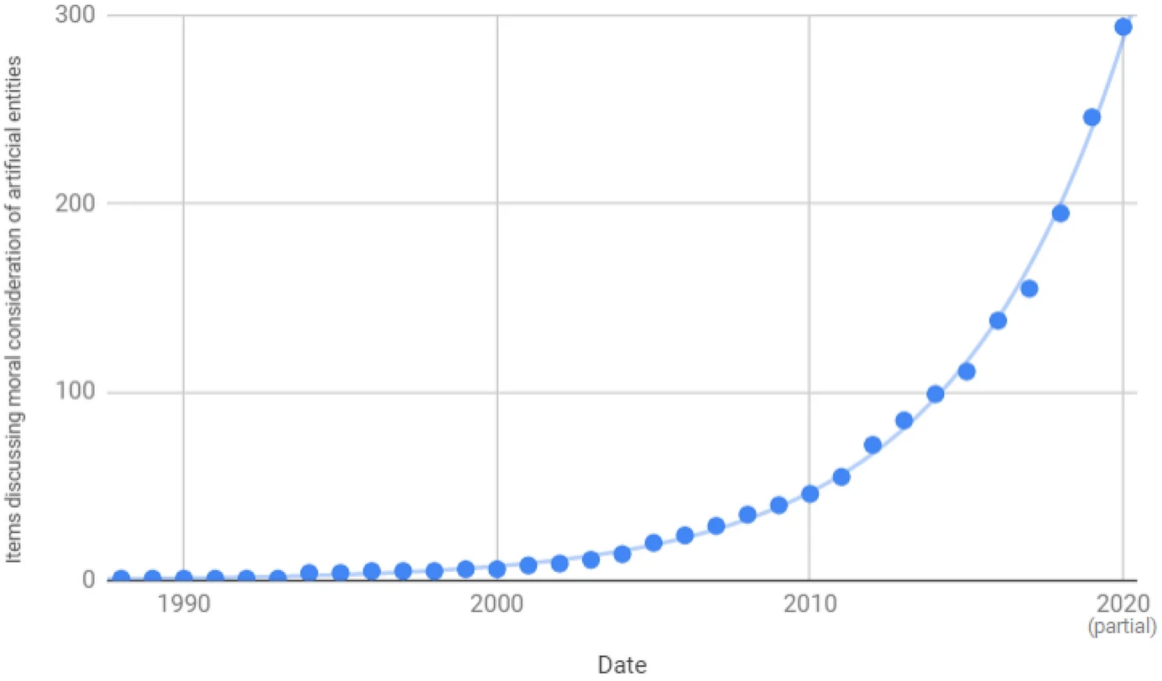


Fig. 2.1 Academic interest in the moral consideration of artificial entities by date of publication [31].

Despite increased academic interest, the study identified significant gaps, particularly in empirical research on public attitudes toward the moral consideration of artificial entities. Most studies have centered around philosophical and theoretical debates, leaving a critical need for interdisciplinary research that incorporates psychological, sociological, and economic perspectives.

To address this gap, increasing efforts have been made to collect public opinion through survey analysis on AI sentience and rights. [57] surveyed 100 Amazon Mechanical Turk workers to explore perceptions of machine consciousness in technologies like GPT-3 and

[2] Researchers selected items based on their relevance to the topic of moral consideration of artificial entities. Exclusion criteria included items that did not directly discuss the moral consideration of AI, only mentioned the topic briefly, or were not in an academic format, such as newspaper op-eds or blog posts. After applying these criteria and removing duplicates, 294 items were included in the final analysis. These items form the dataset represented on the y-axis, which tracks the number of relevant publications included in the review, categorized by their focus on AI rights, moral status, or suffering [31].

robot vacuums. Researchers found that many participants already perceived these technologies as somewhat conscious, which the study suggests could significantly impact Human-Computer Interaction (HCI) by raising challenges related to empathy, power dynamics, and the responsibility of interacting with machines perceived as conscious.

Similarly, [37] conducted experiments to assess public attitudes toward granting rights to autonomous AI and robots, initially with EU members and later with a representative sample of U.S. citizens. Their study revealed general resistance to most AI rights but noted significant support for protections against cruelty. The findings were consistent across both samples, demonstrating that public perceptions can be positively influenced by correcting misconceptions about legal personhood for non-human entities, suggesting that public opinion on AI rights is adaptable and could evolve.

2.2.2 The AIMS Survey

Researchers at *The Sentience Institute* conducted the Artificial Intelligence, Morality, and Sentience (AIMS) Survey to explore public perceptions of AI sentience and moral consideration. This nationally representative survey, first conducted in November and December 2021 with 1,232 U.S. adults, was later expanded with two additional waves in 2023, bringing the total sample size around 2400 respondents.

The survey revealed a complex public stance on AI rights: while 71% of respondents in 2023 agreed that sentient AI deserves to be treated with respect, only 38% supported granting legal rights to AI. Notably, public concern for AI well-being and mind perception significantly increased from 2021 to 2023. Despite this growing concern, there is also strong resistance to advanced AI, with 69% favoring a ban on sentient AI. Demographic factors such as younger age, male gender, being White or Asian, liberal political orientation, religious affiliation, and exposure to AI narratives were found to predict positive emotions, trust, and concern for AI treatment.

2.3 Review of Different Clustering Methods

Clustering is a key technique in data mining that involves grouping data into distinct categories, or clusters, without using predefined labels. This project aims to apply clustering methods to the AIMS Survey dataset, which includes both numerical and categorical variables. The goal is to identify natural groupings of respondents based on their opinions about AI sentience and rights. [48] offers a detailed comparison of clustering algorithms, highlighting their strengths and weaknesses across various data mining scenarios. The study concludes that no single clustering algorithm is universally superior; the choice of algorithm should depend on the specific characteristics of the dataset, such as size, shape, and the presence of noise.

2.3.1 K-Means Clustering

K-means clustering is a widely used unsupervised learning algorithm in machine learning and data science, particularly for solving clustering problems by partitioning a dataset into distinct subgroups based on the features of the observations [4]. It operates by iteratively assigning data points to one of K predefined clusters, where each data point is associated with the nearest cluster centroid. This process continues until the algorithm converges, typically when the centroids stabilize and the assignments no longer change. The objective is to minimize the following function:

$$\arg\min_{C} \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Where:

- C_i represents the i-th cluster,
- x is a data point,
- μ_i is the centroid of the i-th cluster,
- $\|\cdot\|$ denotes the Euclidean distance.

One of the key advantages of K-means is its computational efficiency, making it one of the fastest clustering algorithms available, particularly suitable for large datasets [5]. However, K-means is not without its limitations. A significant drawback is its susceptibility to falling into local optima, which can lead to suboptimal clustering results, especially when the initial placement of centroids is poorly chosen [67]. Additionally, K-means requires the number of clusters K to be specified in advance, which can be challenging without prior knowledge of the data's underlying structure. To address this, the following methods can be employed:

Silhouette Score

The Silhouette Score is a measure used to evaluate the quality of clustering. It quantifies how well each data point lies within its cluster compared to other clusters. The score ranges from

-1 to 1, where a higher score indicates that the data point is well-matched to its own cluster and poorly matched to neighboring clusters.

The Silhouette Score $s(i)$ for an observation i is defined as:

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

Where:

- $a(i)$ is the average distance between i and all other points in the same cluster.
- $b(i)$ is the minimum average distance between i and all points in the next nearest cluster.

In this research, the silhouette score is used to determine the optimal number of clusters, perform hyperparameter tuning, and evaluate the final results.

Elbow Method

The Elbow Method is one method to determine the optimal number of clusters for K-Means clustering [33]. It involves running K-Means with different values of k (the number of clusters) and plotting within-cluster sum of squares (WCSS) against k. The optimal number of clusters is identified by the user based on the visual indication of wherever the "elbow" point is on the plot, where adding more clusters results in diminishing returns in terms of explained variance.

The WCSS is calculated as:

$$\text{WCSS} = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2$$

Where:

- k is the number of clusters.
- x is a data point.
- μ_i is the centroid of cluster C_i.
- $|| \cdot ||$ denotes the Euclidean distance[3].

In order to find the optimal number of clusters K, both the Elbow Method and Silhouette Scores will be run. By analyzing the resulting plots, the K value that yields the best balance between explained variance and cluster quality will be selected for further analysis.

[3] Euclidean distance masures the straight-line distance between two points in Euclidean space.

2.3.2 DBSCAN Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a widely used unsupervised learning algorithm designed to identify clusters of arbitrary shapes within datasets that may contain noise and outliers [Ester et al.]. Unlike traditional clustering algorithms that rely on predefined numbers of clusters, DBSCAN forms clusters based on density of points in the dataset. It operates by grouping together points that are closely packed together, marking points that are not reachable as outliers.

The algorithm relies on two key parameters:

- **Eps**: The radius that defines the neighborhood around a point.
- **MinPts**: The minimum number of points required to form a dense region (cluster).

A point is considered a core point if its ε-neighborhood contains at least *MinPts* points. A cluster is formed by all points that are density-reachable from core points. While DBSCAN is good at discovering clusters of arbitrary shapes and is particularly effective in handling datasets with noise and outliers, it comes with some challenges:

- The choice of Eps and MinPts parameters significantly influences the clustering outcome, making it crucial to select them carefully for optimal results.
- DBSCAN can struggle with datasets that have clusters of varying densities, as a single set of parameters may not be suitable for all clusters.
- The algorithm can be computationally intensive, particularly for large datasets, due to the need to examine the neighborhood of each point.

Despite these challenges, DBSCAN remains a powerful tool for clustering, especially in scenarios where the data contains noise or irregular cluster shapes. [4]

$$N_{Eps}(p) = \{q \in D \mid dist(p,q) < Eps\} \tag{2.1}$$

Where:

- $N_{Eps}(p)$ is the Eps-neighborhood of point p,
- D is the dataset,

[4] Various enhancements and variations of DBSCAN, such as VDBSCAN, FDBSCAN, DD_DBSCAN, and IDBSCAN, have been developed to address its limitations, each offering trade-offs in terms of performance and adaptability [55].

- $dist(p,q)$ is the distance between points p and q.

DBSCAN continues to be a preferred method in clustering tasks where traditional algorithms like K-Means may fall short, particularly in complex datasets with irregular cluster formations and the presence of noise [20].

2.3.3 Hierarchical Clustering

Hierarchical Clustering is an unsupervised learning algorithm used to organize data into nested clusters, forming a tree-like structure known as a dendrogram [12]. Unlike K-Means, which requires the number of clusters to be predefined, Hierarchical Clustering allows the data to dictate the structure and number of clusters by building them step by step. Hierarchical Clustering can be performed using two main approaches:

- **Agglomerative (Bottom-Up) Approach**: Starts with each data point as its own cluster and iteratively merges the closest pairs of clusters.
- **Divisive (Top-Down) Approach**: Begins with all data points in a single cluster and recursively splits them into smaller clusters.

The distance between clusters is a key factor in how they are merged. Some commonly used distance metrics include:

Euclidean Distance:

$$d_{\text{Euclidean}}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2.2}$$

Measures the straight-line distance between two points in Euclidean space.

Manhattan (Cityblock) Distance:

$$d_{\text{Manhattan}}(x,y) = \sum_{i=1}^{n}|x_i - y_i| \tag{2.3}$$

The distance metric measures the sum of absolute differences between the coordinates of two points.

Cosine Distance:

$$d_{\text{Cosine}}(x,y) = 1 - \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{2.4}$$

Measures the cosine of the angle between two non-zero vectors, focusing on the orientation rather than magnitude. The method for determining the distance between clusters is determined by the linkage criteria:

Complete Linkage:

$$d(C_i,C_j) = \max\{d(x_p,x_q) \mid x_p \in C_i, x_q \in C_j\} \tag{2.5}$$

Defines the distance between two clusters as the maximum distance between any single data point in the first cluster and any single data point in the second cluster, creating compact, evenly sized clusters.

Single Linkage:

$$d(C_i,C_j) = \min\{d(x_p,x_q) \mid x_p \in C_i, x_q \in C_j\} \tag{2.6}$$

This defines the distance between two clusters as the minimum distance between any single data point in each cluster, which can lead to "chain-like" clusters.

Average Linkage:

$$d(C_i,C_j) = \frac{1}{|C_i||C_j|} \sum_{x_p \in C_i} \sum_{x_q \in C_j} d(x_p,x_q) \tag{2.7}$$

This uses the average distance between all pairs of data points from each cluster, balancing the characteristics of complete and single linkage.

Hierarchical Clustering is particularly useful for visualizing relationships in data through dendrograms, allowing for easy exploration of different cluster configurations by cutting the tree at various levels. However, this method can be computationally intensive and sensitive to noise and outliers, especially in large datasets.

2.3.4 Comparing Different Methods

[41], a study comparing the above three methods finds that DBSCAN is particularly suited for datasets with non-convex shapes[5], well-separated clusters, or a high number of outliers. [28] and [6] further explore the efficiency of these algorithms. [28] notes that DBSCAN is more effective in managing noise and outliers compared to hierarchical clustering, which can be time-consuming. [6] concludes that K-Means is preferable for larger datasets due to faster

[5] A convex shape in Geometry is a shape where the line joining every two points of the shape lies completely inside the shape.

execution time and stable memory usage, whereas Agglomerative Hierarchical Clustering is better suited for smaller datasets. As shown in Table *2.1*, the strengths and weaknesses of different clustering methods vary depending on the dataset and application. Given the strengths and weaknesses of different clustering methods, this research will use a combined approach to determine the best clustering method for the AIMS Survey dataset. By applying K-Means, DBSCAN, and Hierarchical Clustering to the data, the study will identify which method provides the most meaningful clusters. Hierarchical Clustering will help visualize relationships in the data through dendrograms, allowing flexibility in exploring different cluster configurations. Once the number of clusters is estimated, K-Means will be applied for its efficiency. DBSCAN will also be applied to detect clusters in data with irregular shapes and to handle noise and outliers.

Table 2.1 Comparison of Clustering Methods

Method	**Strengths**	**Weaknesses**	**Best For**
K-Means Clustering	Handles mixed data types, interpretable clusters	Sensitive to initialization, requires specifying K	Mixed-type data clustering
DBSCAN Clustering	No need for K, handles noise, finds arbitrary-shaped clusters	Parameter sensitivity, may not scale well with high dimensions	Identifying core clusters and outliers
Agglomerative Hierarchical Clustering	Visual representation (dendrogram), flexible number of clusters	Computationally intensive, less effective for large datasets	Visualizing hierarchical relationships

Evaluating Cluster Validity

Once the clustering results are obtained, both quantitative and qualitative measures will be employed to assess the validity and quality of the clusters. In this study, cluster validity will be evaluated using the Silhouette Score and the Davies-Bouldin Index (DBINDEX), followed by a qualitative assessment of the results.

The Silhouette Score is selected for its effectiveness in identifying relationships within and between clusters, providing insights into the cohesion and separation of the groups. Similarly, the Davies-Bouldin Index is chosen for its focus on both the compactness of clusters and their separation. This dual approach allows for a comprehensive evaluation of the clustering performance.

The **Davies-Bouldin Index** is defined as follows:

$$\mathrm{DB} = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} \left(\frac{d_i + d_j}{d(c_i, c_j)} \right),$$

where k is the number of clusters, d_i represents the average distance of all points in the ith cluster from the cluster centroid, and $d(c_i, c_j)$ is the distance between the centroids of the ith and jth clusters. A lower DBINDEX indicates better clustering, reflecting more compact and well-separated clusters.

2.4 Feature Reduction Techniques

Reducing the number of features in the dataset is a critical aspect of this project for two main reasons:

1. Many machine learning models, such as K-Means and DBSCAN, encounter significant challenges when clustering high-dimensional data, a phenomenon known as the "curse of dimensionality." This term refers to the exponential increase in computational complexity, inefficiency in space utilization, and diminished visualization capabilities as the number of dimensions grows [65]. By reducing the dimensionality of the dataset, these issues can be mitigated, leading to improved performance and accuracy of the models.
2. Grouping together similar beliefs about AI sentience and rights simplifies the evaluation of clustering results, making it easier to interpret and analyze the outcomes.

The results are displayed and discussed in *Chapter 4*.

2.4.1 Qualitative Feature Reduction

[49] analyzed the survey items in two ways: as individual standalone items and by averaging or summing specific items to compute index variables. Researchers behind the first wave of the AIMS Survey provided 12 index variables, resulting in the initial groupings of features [50]. These groupings serve as a foundation for qualitatively reducing the dimensionality of the dataset. To ensure the quality of these groupings, two qualitative methods will be used:

1. **Correlation Matrix with Heatmap:**

- A correlation matrix can be used to visualize and select features based on their relationships, as strong correlations between features suggest a shared underlying dimension or theme, justifying their combination into indices or composite variables [13].
- The heatmap will provide a visual representation of these correlations, making it easier to identify which features are closely related and should be grouped together.

2. **Hierarchical Clustering on the Correlation Matrix:**

- Hierarchical clustering will be applied to the findings from the correlation matrix, as it is commonly used in data analysis procedures to group features with similar patterns [63].
- By combining the matrix for heatmap visualization with hierarchical clustering, this research can ensure that features with similar patterns are grouped closely together [30].

These methods help ensure the reduced feature set retains its relevance and interpretability, allowing for more effective clustering and analysis in subsequent stages of the project.

2.4.2 Quantitative Feature Reduction

PCA, UMAP and t-SNE

In addition to qualitative methods, quantitative techniques such as Principal Component Analysis (PCA), Uniform Manifold Approximation and Projection (UMAP), and t-distributed Stochastic Neighbor Embedding (t-SNE) can be employed to reduce dimensionality of the data in a more neutral way [54].

Principal Component Analysis (PCA) PCA is a widely used technique that transforms original variables into a new set of uncorrelated variables that successively maximize variance. This method effectively reduces dimensionality in the dataset while retaining the most important information [32].

Uniform Manifold Approximation and Projection (UMAP) UMAP is a non-linear[6] technique that seeks to preserve the global structure of data while reducing its dimensionality

[6] Linear techniques reduce dimensionality using straight-line transformations, while non-linear techniques like UMAP capture and preserve complex, curved relationships in data.

[40]. It is particularly useful for visualizing high-dimensional data and uncovering complex patterns. Additionally, UMAP can maintain salient topological and geometric features of data even as it reduces its dimensionality.

t-distributed Stochastic Neighbor Embedding (t-SNE) t-SNE is another non-linear dimension reduction technique that maximizes the divergence between the probability distributions in the high-dimensional and low-dimensional spaces, making it particularly effective for visualizing clusters in data [64]. Similar to UMAP, t-SNE also aims to preservetopological and geometric structures inherent in data, providing a meaningful low-dimensional representation.

The results of all three feature reduction techniques can be seen in *Section 4.1*.

Chapter 3

Research Methodology

For more details on the code and data used in this study, please refer to the project's GitHub repository[1].

3.1 Dataset Description

Two rounds of survey data were collected from preregistered participants in the AIMS survey, with the first round conducted in 2021 and the second in 2023 [49][51]. The combined dataset consists of responses from a total of 2,401 participants, spanning approximately 140 columns of variables. The dataset includes both numerical and categorical variables, capturing a wide range of respondents' opinions and demographic information. The key variables can be grouped as follows:

3.1.1 Numerical Variables

The dataset includes several variables that are primarily based on Likert scales[2] or other numerical values. These variables capture a range of attitudes and perceptions related to AI sentience, rights, risks, and moral considerations for non-human entities. Key numerical variables include:

- **Attitudes Toward AI Sentience and Rights:** Variables such as AS Caution, Pro-AS Activism, and AS Treatment, measured on a Likert scale (1-7), reflect respondents' support for various stances on AI sentience and the ethical treatment of AI entities.

[1] Click the words 'Github repository'

[2] A Likert scale is a rating scale used in surveys to measure attitudes or opinions, typically ranging across different values.

- **Concerns Related to AI:** Malevolence Protection and AI Moral Concern, also measured on a Likert scale (1-7), gauge respondents' concerns about the potential risks posed by AI, including moral and ethical considerations.
- **Perception of AI's Cognitive Abilities and Threats:** Mind Perception and Perceived Threat capture perceptions of AI's cognitive capabilities and potential dangers, measured using scales that reflect varying degrees of agreement or concern.
- **Moral Consideration for Non-Human Entities:** Variables such as MCA1, MCA2, MCEn1, and MCEn2, typically measured on Likert scales (1-7), assess the moral status accorded to animals and the environment by respondents.
- **Categorized Concerns:** Variables like MCE21 to MCE31 reflect categorized moral concerns towards different types of AI, rated on a scale from 1 to 5, indicating level of moral concern.
- **Scale (1-100):**Mind Perception variables (MP1, MP2, MP3, MP4) assess respondents' beliefs about the extent to which AI currently possesses cognitive abilities, using a scale from 1 to 100.

3.1.2 Categorical and Binary Variables

The dataset also includes categorical and binary variables that capture demographic information, experiences with AI, and other personal characteristics. These variables include:

- **Demographics:** Age, gender, education, income, and diet, which provide basic demographic profiles of the respondents.
- **Experiences with AI:** Binary variables such as ownership of AI devices, work with AI, and various experiences (e.g., seeing AI being mistreated), which capture the respondents' interactions with AI technologies.
- **Political and Religious Affiliations:** Categorical variables like politics, religion, and related dummy variables represent the respondents' political views and religious beliefs.
- **Dummy Coded Variables:** Binary representations of categorical variables (e.g., religionRNR, dietMR) are used for easier analysis in statistical models.

3.2 Data Preprocessing

Standardizing the Numerical Data

Given the variety of numerical variables in the dataset, including Likert scales and other numerical measures, standardization is essential before conducting any clustering analysis. These variables, as detailed in the previous section, capture a wide range of attitudes, perceptions, and moral considerations related to AI and non-human entities. The inherent differences in scales and units across these variables necessitated a normalization process to ensure that each feature contributed equally to the clustering results.

StandarScale

The `StandardScaler` method was used for data standardization. `StandardScaler` operates by subtracting the mean of each feature and scaling it to unit variance. This process transforms the data so that each feature has a mean of *0* and a standard deviation of *1*, effectively placing all features on a common scale.

By standardizing the data, `StandardScaler` mitigates risk of any one feature disproportionately influencing clustering outcomes due to its scale. Moreover, `StandardScaler` is well-suited for algorithms like hierarchical clustering, which assume normally distributed data. Normalizing the data ensured the clustering results reflect true similarities and differences between data points rather than artifacts of scale differences.

3.2.1 Categorical Data Preprocessing: Focus on Demographics

For this research, the focus was on key demographic variables, including `age`, `dietMR`, `education_recode`, `gender`, `income_recode`, `politics`, and `raceethnicity`. These variables were extracted from the dataset for detailed inspection and preprocessing.

Handling Missing Values

The dataset contained a total of 2,207 missing values across all columns, with the majority located in the additional column provided for respondents to specify their religion if it was not listed in the predefined options. This column, labeled `relelse`, contained various entries, some matching existing categories in the `religion` column.

Consolidating Religious Responses

To ensure consistency in the religion data, entries in the `relelse` column that corresponded to known categories in the `religion` column were mapped accordingly. For example, entries such as “Roman Catholic” or “Born-again Christian” were mapped to their corresponding categories “Catholic” and “Protestant,” respectively. This mapping was performed using a predefined dictionary categorizing similar or synonymous responses under the appropriate religion which ensured entries in the `relelse` column were categorized consistently with predefined religious categories in the `religion` column.

Consistency Check with Religion Categories

After mapping the `relelse`, a consistency check was performed to ensure that the classifications were accurate. This involved comparing the religion data against a binary variable `religionRNR` provided in the dataset, where `0` indicated religious affiliation and `1` indicated no religious affiliation.

A function was created to flag inconsistencies, such as when a respondent identified as religious (e.g., Protestant) but was marked as `1` (not religious) in the `religionRNR` column. Rows with such inconsistencies were identified and addressed.

This iterative process ensured that the religious data was accurately categorized and consistent with the respondents’ self-identified religious status. As a result, the distribution of different religions within the dataset was visualized, as shown in *Figure 3.1*. This figure provides a clear overview of the proportions of various religious affiliations among the survey participants.

3.2.2 Recoding of Demographic Variables

For a more meaningful analysis of the demographic data, certain variables were recoded into categorical groups based on their values. This process was particularly applied to the age and politics columns in the dataset.

Age Recoding

The *age* variable, initially provided as a continuous numerical value, was categorized into three distinct age groups:

- **18 to 34 years:** Represented by the label `18_34`
- **35 to 54 years:** Represented by the label `35_54`

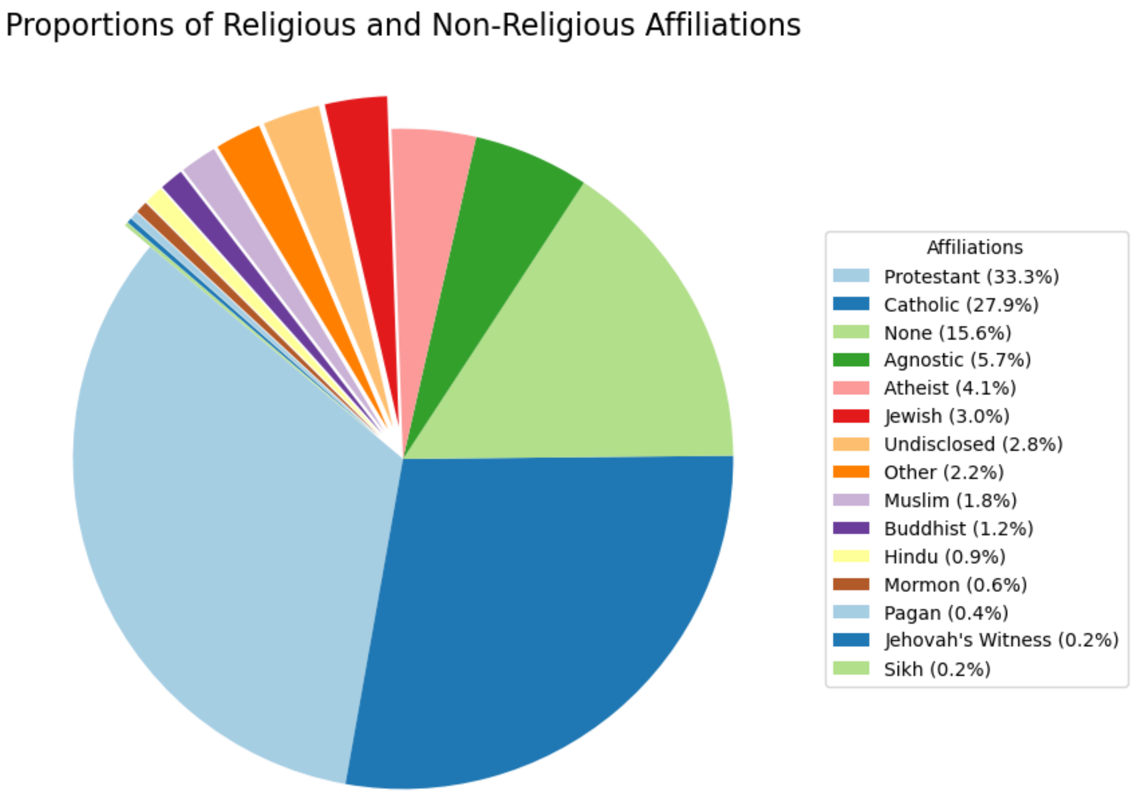


Fig. 3.1 Distribution of religious affiliations of survey participants.

- **55 years and older:** Represented by the label `55_`

This recoding was achieved by defining a function that assigns each age to one of the three categories. The function was then applied to the *age* column, resulting in the creation of a new *age_recode* column. The results are visualized in *Appendix A.1*.

3.2.3 Politics Recoding

Similarly, the *politics* column, which initially contained a range of values representing political alignment on a continuous scale, was recoded into three categories:

- **Very Liberal:** Assigned to values less than or equal to 2.
- **Moderate:** Assigned to values between 2 and 4.
- **Very Conservative:** Assigned to values greater than or equal to 4.

A function was defined to implement this recoding, transforming the continuous *politics* values into the aforementioned categorical groups. This recoding provides a simplified view of political orientation, seen in *Figure A.1c*, and subsequently simpler correlation assessment with other variables in the dataset[3].

3.2.4 Grouping Index Variables

Employing a qualitative approach reduces dimensionality of the dataset while preserving interpretability of features. This method involved aggregating related numerical columns into index variables based on their qualitative attributes. These index variables were derived from the original report, where items were averaged or summed to compute composite variables. The resulting 12 distinct features include *Table 3.1*:

[3] The decision to recode was driven by the need to make the analysis clearer and to identify patterns more easily. However, this approach may result in some loss of granularity.

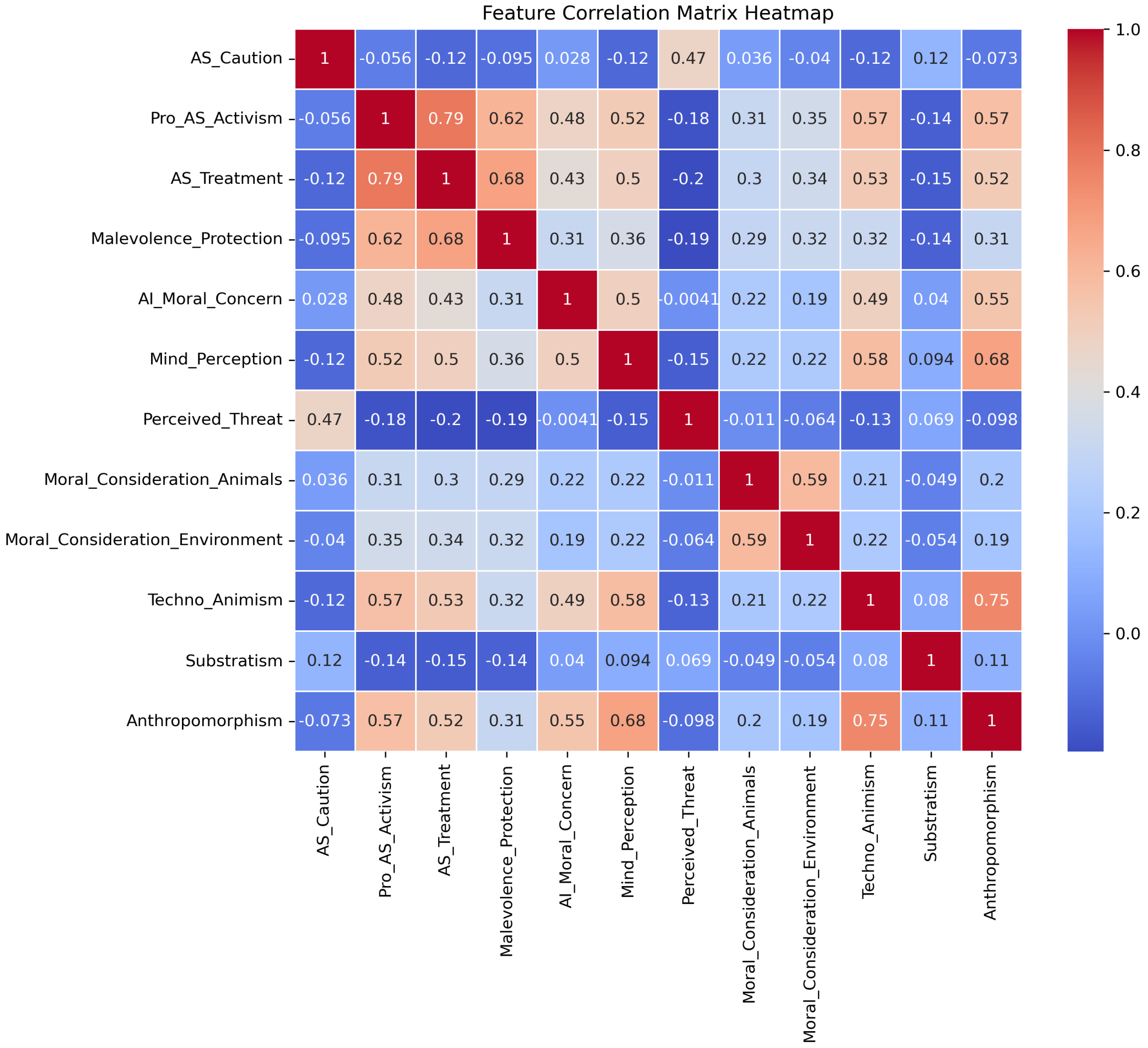
Table 3.1 Summary of Index Variables

Name of Group	**Average/Sum of Columns**	**What the Group Represents**
AS Caution	Average of PMC #1, 3, 4	Caution towards artificial sentience
Pro-AS Activism	Average of PMC #2, 5-12	Support for advocating artificial sentience
AS Treatment	Average of MCE #1-6	Concern for AS treatment
Malevolence Protection	Average of MCE #7-9	Support for AI protection from malevolence
AI Moral Concern	Average of MCE #21-31	Moral concern for AIs
Mind Perception	Average of MP #1-4	Attribution of mind to AIs
Perceived Threat	Average of SI #2-4	Perception of AIs as threats
Moral Consideration of Nonhuman Animals	Average of MCA #1-2	Concern for nonhuman animals
Moral Consideration of the Environment	Average of MCEn #1-2	Concern for the environment
Techno-Animism	Average of TA #1-2	Belief in spirits in artificial entities
Substratism	Average of Sub #1-2	Prejudice against non-carbon-based entities
Anthropomorphism	Sum of Anth #1-4	Attribution of human-like qualities to nonhumans

Justification for Index Groupings

To simplify the dataset and enhance the analysis of the 12 features, the original index variables were grouped based on their correlations, as visualized in the Correlation Matrix Heatmap seen in *Figure 3.2* and analyzed through a Hierarchical Clustering Dendrogram, seen in *Figure 3.3*. Converting the correlation matrix into a distance matrix enabled the effective application of hierarchical clustering[4]. This approach grouped strongly correlated features into indices, reducing redundancy while preserving core information. The alignment between hierarchical clustering and the correlation heatmap supports these groupings, making the dataset more manageable for further analyses like K-Means and DBSCAN, leading to more efficient and interpretable results.

[4] Hierarchical Clustering requires a distance matrix, and the method was found on StackExchange.



	AS_Caution	Pro_AS_Activism	AS_Treatment	Malevolence_Protection	AI_Moral_Concern	Mind_Perception	Perceived_Threat	Moral_Consideration_Animals	Moral_Consideration_Environment	Techno_Animism	Substratism	Anthropomorphism
AS_Caution	1	-0.056	-0.12	-0.095	0.028	-0.12	0.47	0.036	-0.04	-0.12	0.12	-0.073
Pro_AS_Activism	-0.056	1	0.79	0.62	0.48	0.52	-0.18	0.31	0.35	0.57	-0.14	0.57
AS_Treatment	-0.12	0.79	1	0.68	0.43	0.5	-0.2	0.3	0.34	0.53	-0.15	0.52
Malevolence_Protection	-0.095	0.62	0.68	1	0.31	0.36	-0.19	0.29	0.32	0.32	-0.14	0.31
AI_Moral_Concern	0.028	0.48	0.43	0.31	1	0.5	-0.0041	0.22	0.19	0.49	0.04	0.55
Mind_Perception	-0.12	0.52	0.5	0.36	0.5	1	-0.15	0.22	0.22	0.58	0.094	0.68
Perceived_Threat	0.47	-0.18	-0.2	-0.19	-0.0041	-0.15	1	-0.011	-0.064	-0.13	0.069	-0.098
Moral_Consideration_Animals	0.036	0.31	0.3	0.29	0.22	0.22	-0.011	1	0.59	0.21	-0.049	0.2
Moral_Consideration_Environment	-0.04	0.35	0.34	0.32	0.19	0.22	-0.064	0.59	1	0.22	-0.054	0.19
Techno_Animism	-0.12	0.57	0.53	0.32	0.49	0.58	-0.13	0.21	0.22	1	0.08	0.75
Substratism	0.12	-0.14	-0.15	-0.14	0.04	0.094	0.069	-0.049	-0.054	0.08	1	0.11
Anthropomorphism	-0.073	0.57	0.52	0.31	0.55	0.68	-0.098	0.2	0.19	0.75	0.11	1

Fig. 3.2 Correlation Matrix Heatmap of the Index Variables.

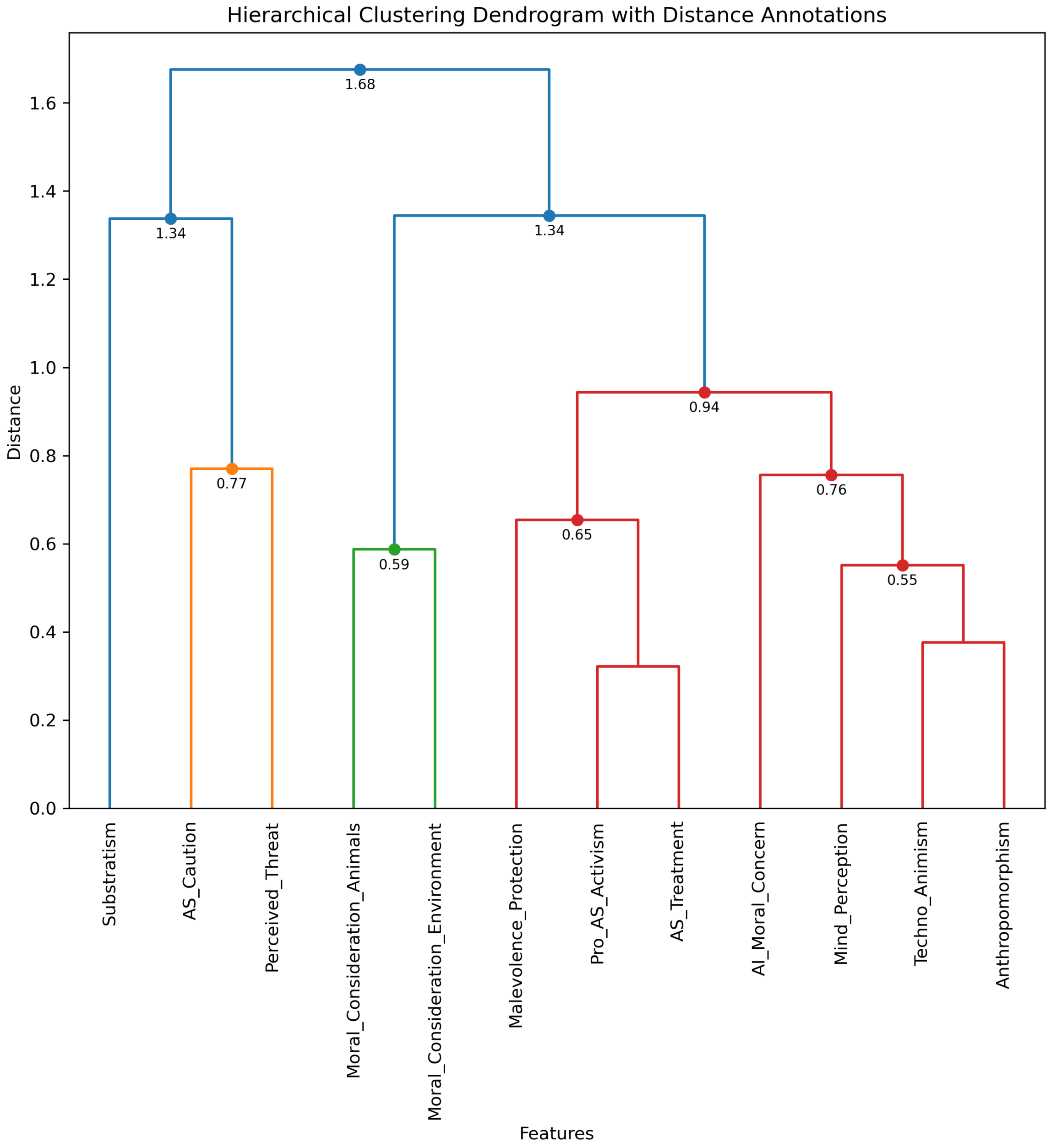


Fig. 3.3 Hierarchical Clustering on the Correlation Matrix Heatmap of the Index Variables.

Table 3.2 Index Groupings Based on Correlations

Group	**Included Features**	**Justification**
Group 1	AS Caution, Perceived Threat	These features were grouped together due to their close correlation (0.77), both reflecting cautionary attitudes toward AI developments.
Group 2	Moral Consideration of Nonhuman Animals, Moral Consideration of the Environment	These features exhibited a very strong correlation (0.59), justifying their combination into a single moral consideration index.
Group 3	Pro-AS Activism, AS Treatment, Malevolence Protection	These features were grouped based on their correlations (0.65 to 0.94), representing active support and protective attitudes toward Artificial Sentience (AS).
Group 4	Mind Perception, Techno-Animism, Anthropomorphism	These features reflect how individuals perceive AI and attribute human-like or spiritual qualities to non-human entities.
Group 5	Substratism	Substratism remained distinct due to its unique nature but was considered in the broader context of caution and threat perceptions.

Chapter 4

Results

4.1 Applying Feature Reduction Techniques

The quantitative feature reduction techniques described in *Section 2.4.2* were applied using both 2 and 3 components to evaluate their ability to reveal meaningful patterns or clusters. PCA and t-SNE did not produce clear clusters, suggesting they may not adequately capture the dataset's complexity. The 2D UMAP plot (Figure 4.1) showed a linear distribution of data points, and when extended to 3D, UMAP revealed more distinct and interpretable clusters, as can be seen in *Figure 4.2*. Among the three methods, UMAP demonstrated a stronger ability to capture underlying complexity of the data in three dimensions. The visualizations for the other two methods can be seen in *Appendix B.2 and B.1*

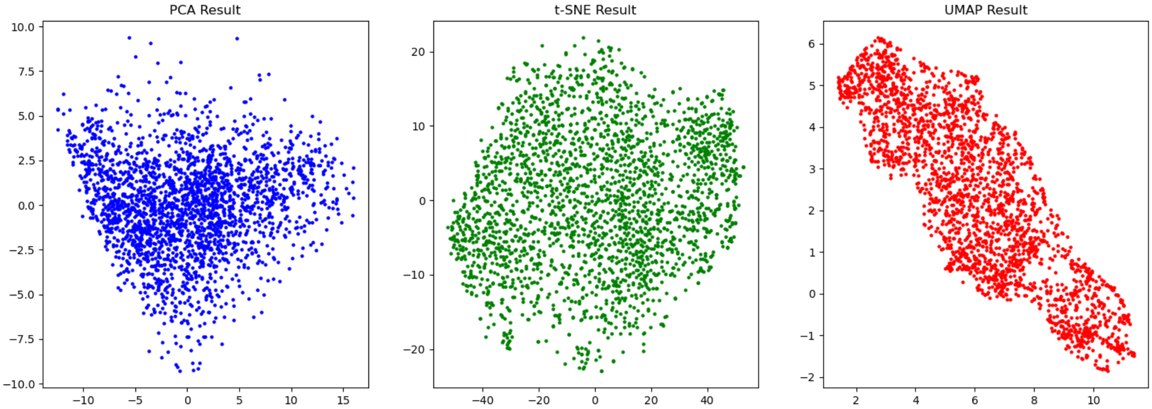


Fig. 4.1 PCA, t-SNE, UMAP results with 3 components.

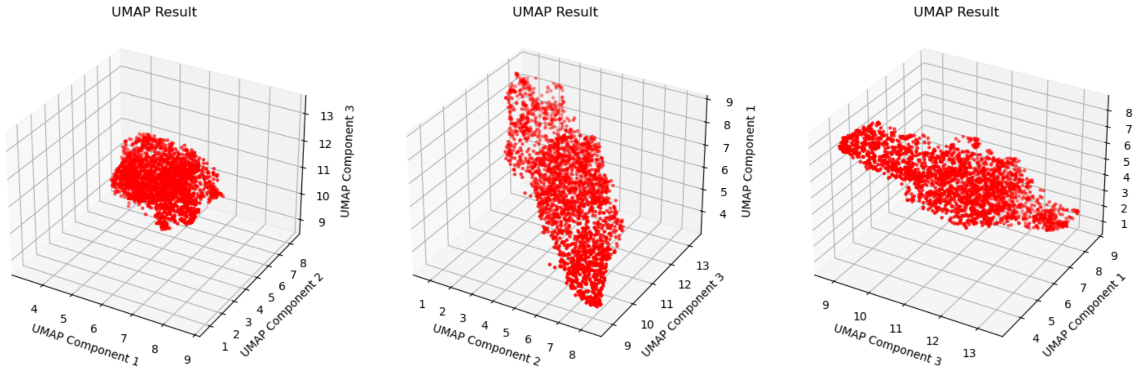


Fig. 4.2 UMAP results with 3 components 3D visualization.

4.2 Choosing Best Cluster Algorithm and Feature Reduction Technique

The reduced feature set was applied to various clustering algorithms, including K-Means, DBSCAN, and Hierarchical Clustering. The quality of the clustering results was rigorously assessed using both quantitative and qualitative metrics. Quantitatively, the Silhouette Score and Davies-Bouldin Index (DB-Index) were employed to measure the compactness and separation of the clusters. Qualitatively, the clusters were evaluated by analyzing their behavior across the original 12 features, providing a deeper understanding of how each algorithm's clusters scored on these features. This qualitative assessment was visualized in *Figures C.1 and C.2 in Appendix C.1* using heat maps visualizing the mean, allowing for a clear comparison of the distribution of features within each cluster. The combination of these quantitative and qualitative methods enables a comprehensive evaluation of the clustering results, ensuring the most effective algorithm and clustering configuration are selected for the project.

4.2.1 K-Means Results

K-Means clustering was applied to both the five reduced feature groups and the UMAP-reduced feature set. The optimal number of clusters (K) was determined using the Silhouette Score and Elbow Method for both datasets, with K values of 2, 3, and 4 being tested. The quality of the resulting clusters was then evaluated using the Silhouette Score and Davies-Bouldin Index, and can be seen in *Table 4.1*.

Table 4.1 Comparison of K-Means Clustering Results on UMAP-Reduced and Feature-Reduced Datasets

Method	**Silhouette Score**	**Davies-Bouldin Index**	**Optimal Number of Clusters (K)**
K-Means on UMAP-Reduced Features (Figure 4.3)	0.570	0.691	4
K-Means on Reduced Feature Groups (Figure 4.4)	0.1851	1.5987	3

When comparing these results, the K-Means clustering on the UMAP-reduced features with $K = 4$ *Figure 4.3*) outperformed clustering on the reduced feature groups *Figure 4.4*, making it the best performing model in this analysis.

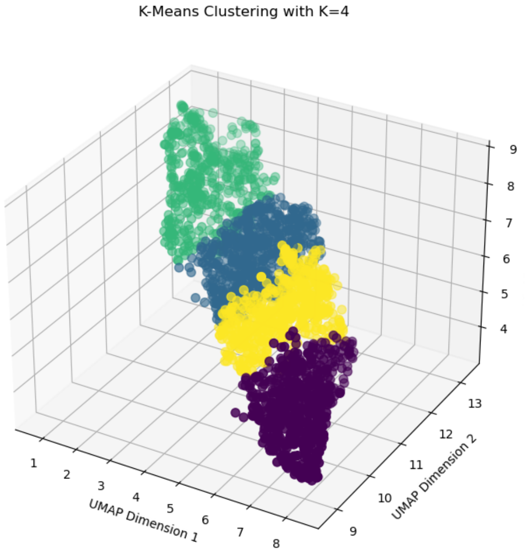


Fig. 4.3 K-Means Clustering Algorithm on reduced UMAP features with 3 components performed the best.

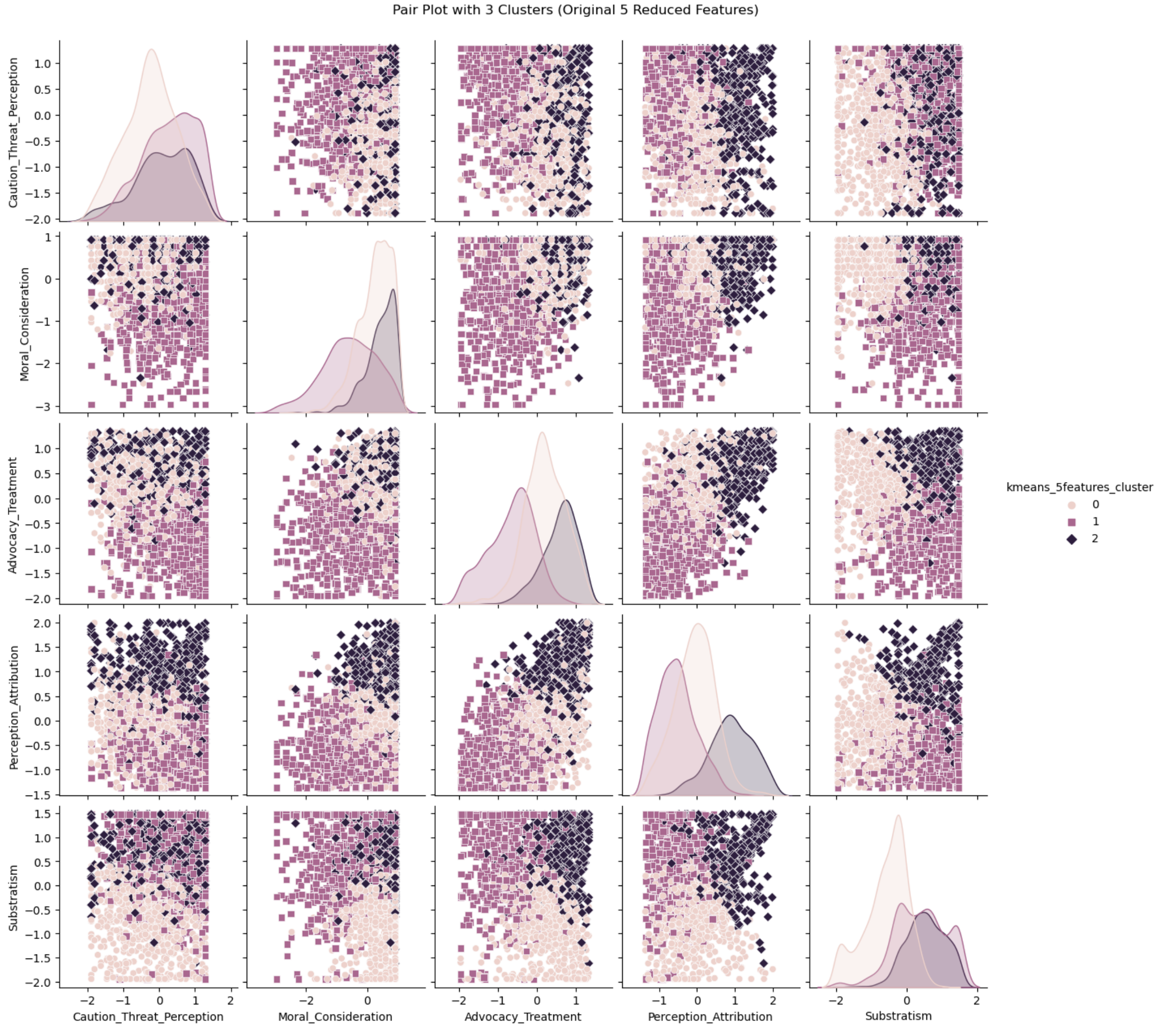


Fig. 4.4 K-Means Clustering Algorithm on reduced features with 3 components.

4.2.2 DBSCAN Results

Different parameter configurations were tested for DBSCAN, and the best-performing parameters, which are reflected in the results presented, were an `eps` value of 0.2 and a `min_samples` value of 5 for the reduced features, and an `eps` value of 0.3 and `min_samples` value of 5 for UMAP, as can be seen in *Table 4.2*.

Table 4.2 Comparison of DBSCAN Clustering Results on UMAP and Reduced Features

Method	**Silhouette Score**	**Davies-Bouldin Index**	**Number of Clusters Formed**
DBSCAN on UMAP Features (Figure 4.5)	-0.1497	2.9447	3
DBSCAN on Reduced Features (Figure 4.6)	0.8957	0.1300	3

The analysis showed that applying the DBSCAN algorithm to UMAP features resulted in 3 distinct clusters, with a Silhouette Score of -0.1497 and a Davies-Bouldin Index of 2.9447, indicating poor clustering quality. In contrast, when DBSCAN was applied to the reduced features, it also produced 3 clusters but with a significantly higher Silhouette Score of 0.8957 and a much lower Davies-Bouldin Index of 0.1300. However, despite the superior quality metrics, the visualization of the clusters formed by the reduced features, as seen in *Figure 4.6*, revealed disproportionate cluster sizes, specifically, **9 points** in one cluster, **5 points** in another, and *2,387* points categorized as noise. While the UMAP features were somewhat more effective in capturing the underlying structure of the data compared to the reduced features, neither approach yielded optimal results.

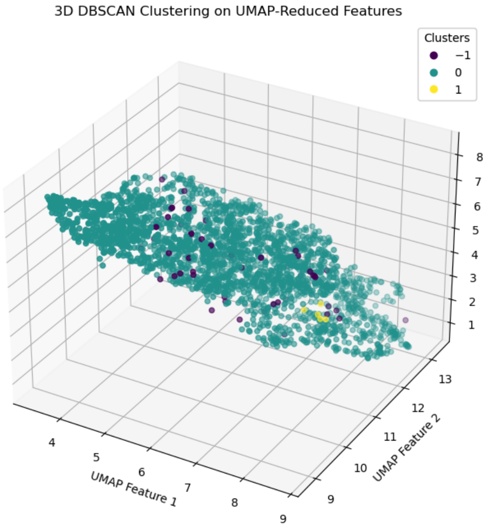


Fig. 4.5 DBSCAN Algorithm on reduced UMAP features with 3 components.

4.2.3 Hierarchical Clustering Results

Hierarchical clustering was employed to analyze the 12 qualitative features, with the goal of uncovering distinct groups within the dataset. The evaluation process involved several steps to ensure the robustness and quality of the clustering results:

4.2.4 Selection of Clustering Combinations

To explore different clustering structures, various combinations of linkage methods and distance measures were tested[1]. Dendrograms were generated for each combination to visualize the hierarchical structure of the clusters. Based on the visual results, the following combinations were selected for their clear and distinct separations between clusters, making them visually interpretable compared to other combinations, as can be seen in *Figure 4.7 and Appendix C.2*.

- Cityblock + Complete
- Euclidean + Complete
- Euclidean + Ward

[1] To see the visualization of each distance + linkage method, please refer to the GitHub.

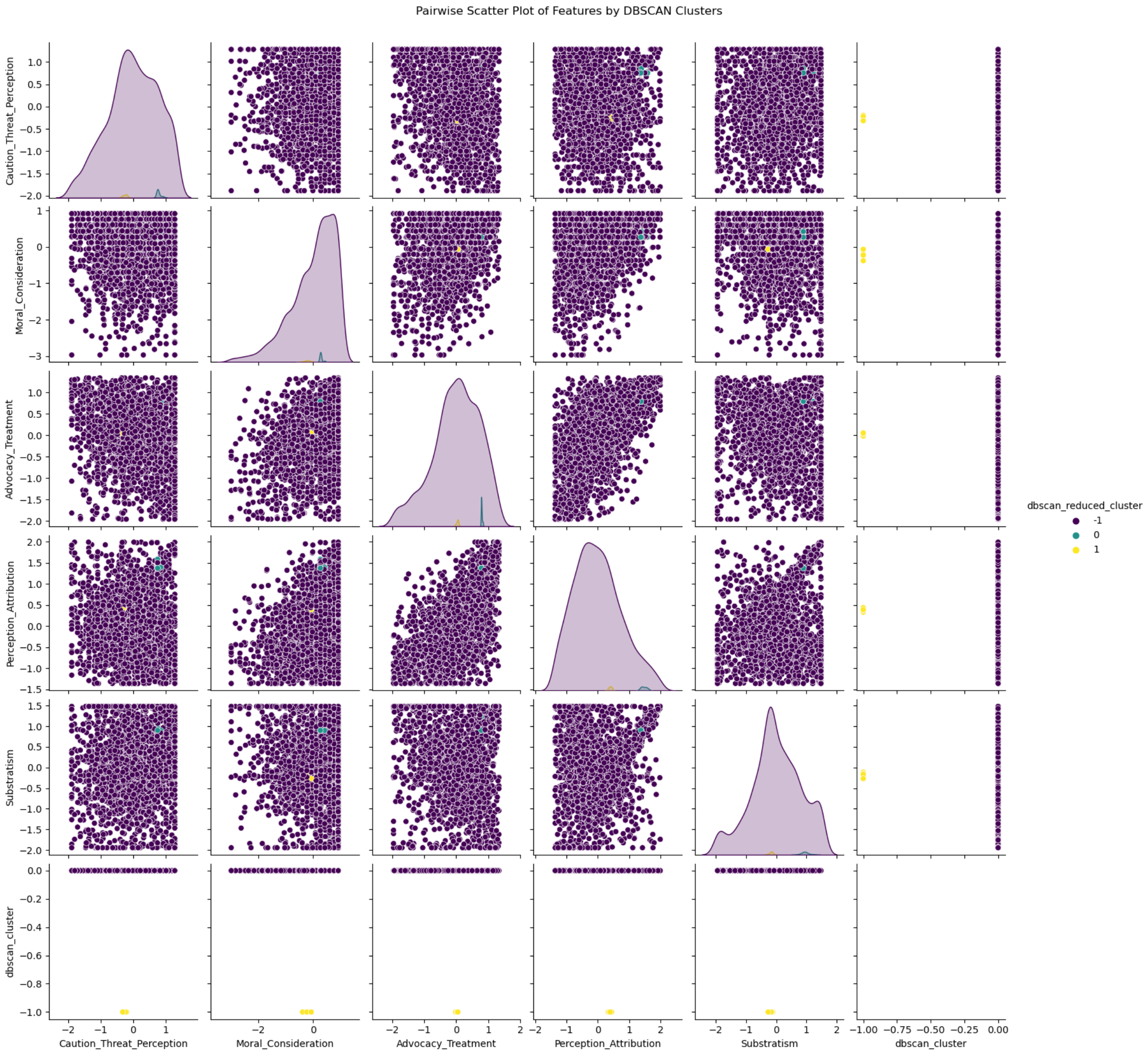


Fig. 4.6 DBSCAN on reduced features: **Cluster -1** has 2387 values, **Cluster 0** has 9 and **Cluster 1** has 5.

4.2.5 Quantitative Evaluation

Table 4.3 shows that amonf=g the tested combinations, the **Cityblock + Complete** combination performed the best, yielding the highest Silhouette Score (0.145) and the lowest Davies-Bouldin Index (1.475).

Table 4.3 Comparison of Hierarchical Clustering Methods on 12 Features

Method	Silhouette Score	Davies-Bouldin Index
Cityblock + Complete	0.145	1.475
Euclidean + Complete	0.096	2.007
Euclidean + Ward	0.105	1.995

Comparison with Reduced Features Hierarchical clustering was also applied to a reduced set of features derived from the original 12. However, the results in *Table 4.4* show that the reduced features did not perform as well as the original index groupings.

Table 4.4 Comparison of Hierarchical Clustering Methods on Reduced Features

Method	Silhouette Score	Davies-Bouldin Index	Inertia
Cityblock + Complete	0.086	2.221	5638.36
Euclidean + Complete	0.102	1.971	4911.71
Euclidean + Ward	0.113	2.176	5329.47

Conclusion: Among the tested combinations, the **Cityblock + Complete** combination in *Figure 4.7* yielded the best results in the hierarchical clustering analysis, achieving the highest Silhouette Score (0.145) and the lowest Davies-Bouldin Index (1.475).

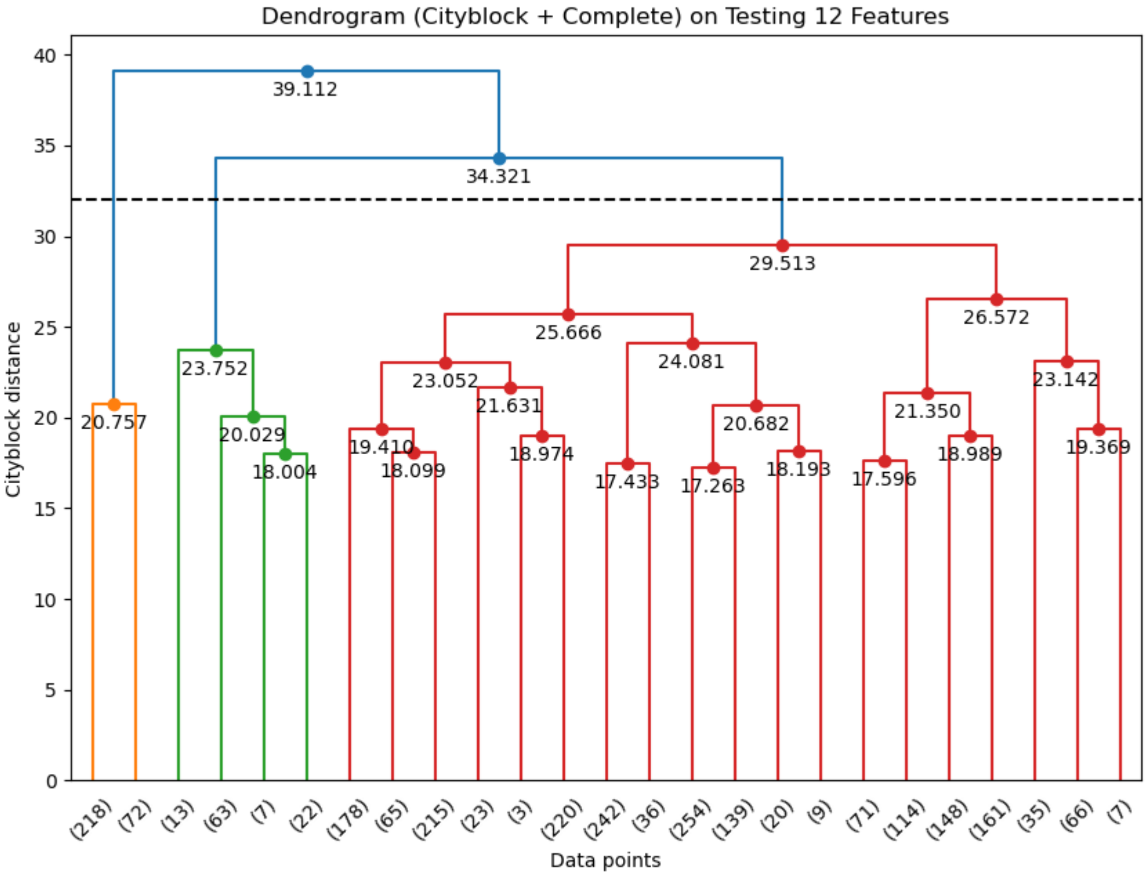


Fig. 4.7 From all of the linkage and distance metric combinations, Cityblock linkage and Complete Distance with 3 clusters performed best overall.

4.3 Qualitative Evaluation

4.3.1 K-Means Clustering with UMAP

Table 4.5 shows the result of evaluating the four distinct clusters produced by K-Means clusters with UMAP across the 12 features in *Table 3.1*.

Table 4.5 Summary of Clusters Based on AI Perception

Cluster	Description	Data Points
Cluster 0	Generally supportive of AI activism and treatment with moderate concerns about AI's moral implications and potential risks. Members are open to attributing lifelike qualities to AI, though with some skepticism.	718
Cluster 1	More cautious and skeptical about AI. Exhibits strong resistance to AI activism and treatment, with significant concerns about AI's potential risks and moral implications.	577
Cluster 2	Highly supportive of AI, viewing it positively. Members are open to attributing human-like qualities to AI, show high concern for AI's moral implications, and perceive less threat from AI.	435
Cluster 3	Neutral or slightly negative attitude towards AI, with moderate resistance to AI activism and treatment. Members are somewhat skeptical about AI's potential for human-like qualities.	671

4.3.2 Cityblock and Complete Hierarchical Clustering with 12 Features

The same evaluation was performed on the 3 clusters produced by the Hierarchical Clustering with 12 features, and can be seen in *Table 4.6*

Table 4.6 Summary of Hierarchical (Cityblock and Complete) Clusters Based on AI Perception

Cluster	Description	Data Points
Cluster 3	Represents the largest cluster, characterized by moderate views across all dimensions. Individuals in this cluster are generally neutral or indifferent towards AI, with no strong inclinations towards either positive or negative perceptions. This group likely represents a general population with a middle-ground perspective.	2006
Cluster 1	Displays strong positive attitudes towards AI activism, treatment, and moral concerns. Members are supportive of AI, showing a proactive stance on its potential benefits and a belief in its anthropomorphic qualities. They demonstrate a high level of moral awareness concerning AI.	290
Cluster 2	Characterized by caution and skepticism towards AI. Members are less supportive of AI activism and treatment, showing significant concern about AI's potential risks and moral implications. They are likely to perceive AI as a threat rather than an opportunity.	105

Conclusion: The analysis revealed that K-Means clustering, particularly when combined with UMAP, produced more distinct and interpretable clusters compared to other methods. The clusters generated by **K-Means + UMAP** captured a wide range of attitudes towards AI, from supportive and progressive stances to more cautious and skeptical views. These clusters demonstrated clear separation between groups, particularly in areas such as AI caution, activism, and moral concern, making them more actionable for further analysis. In contrast, the **DBSCAN** clustering method, which was considered alongside **UMAP**, did not yield as meaningful or well-separated clusters. This led to the exploration of **Hierarchical Clustering** with **Cityblock and Complete** linkage. While this approach identified distinct clusters, they were less interpretable, less varied, and resulted in imbalanced cluster sizes compared to those generated by K-Means.

4.3.3 Final Verdict

K-Means + UMAP provides more distinct and interpretable clusters, capturing a wide range of opinions and making it easier to identify key segments in the data. *Figure 4.8* provides a visualization of the cluster features of the chosen clustering method.

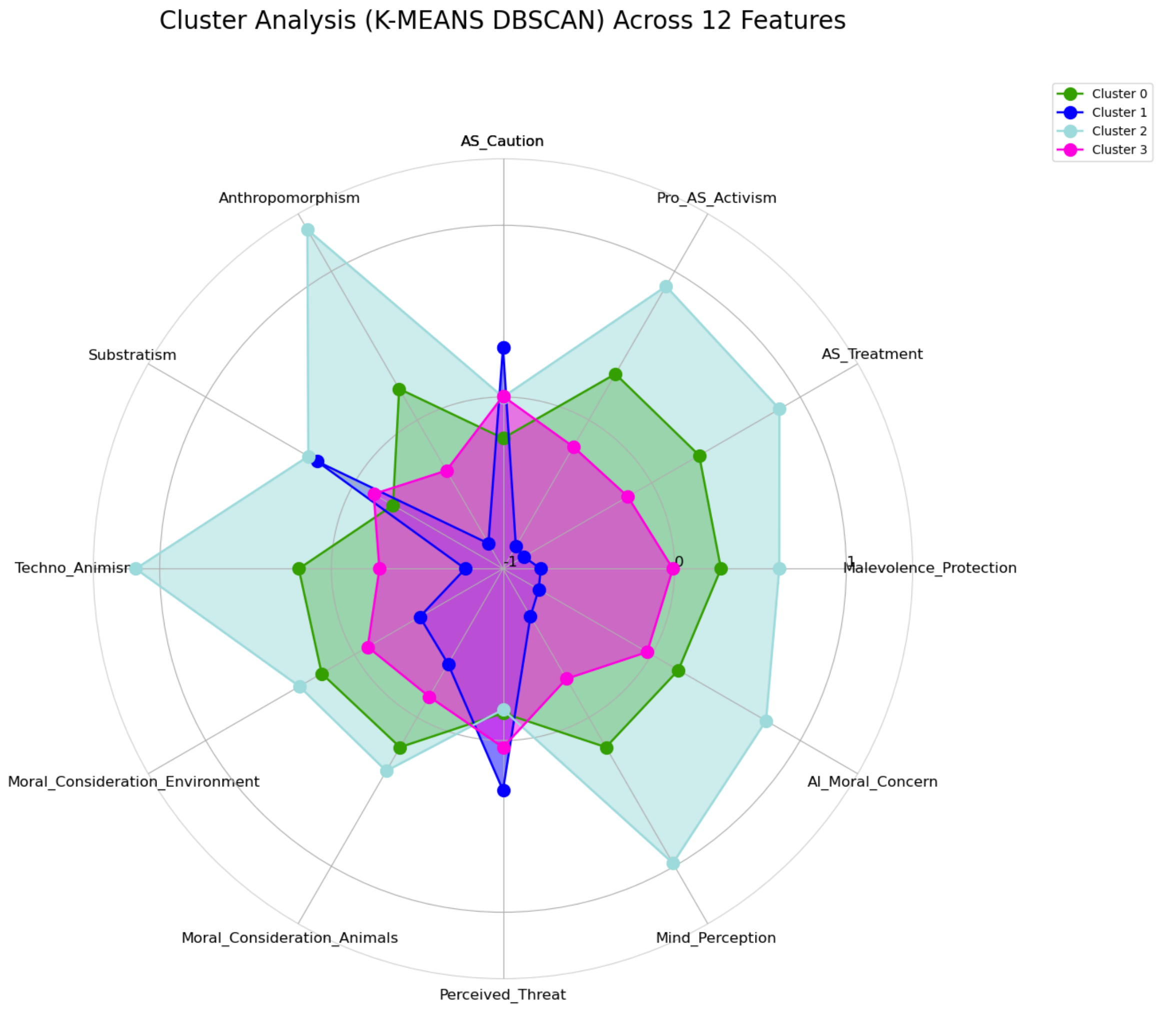


Fig. 4.8 A Spider Graph of the Cluster Features using K-Means UMAP Combination.

4.4 Demographic Features

4.4.1 Summary of Main Demographic Features

The tables below provide a summarized description of cluster demographics. *Figure 4.9*

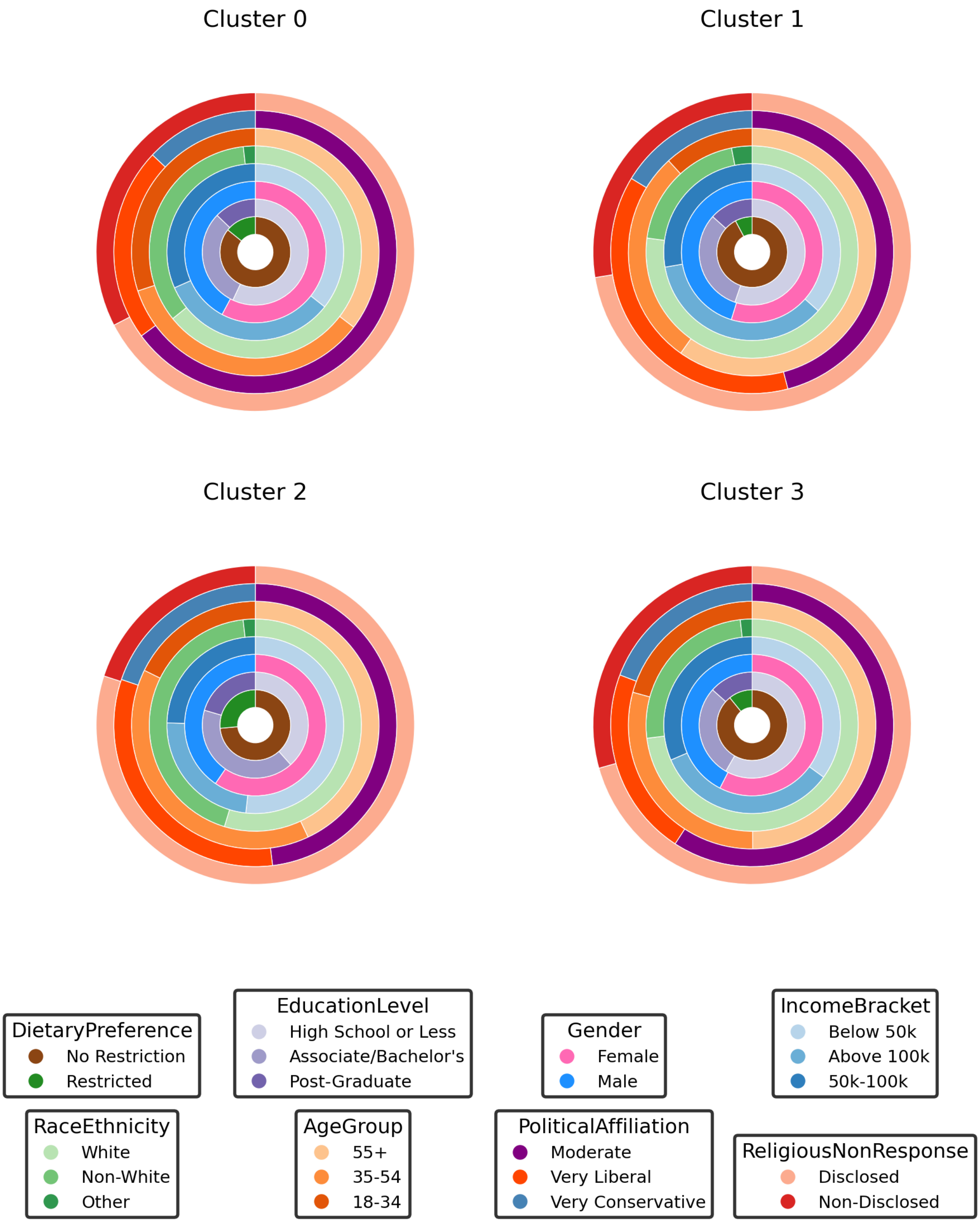


Fig. 4.9 A visualization of the demographic distribution of the clusters.

Cluster 0

Table 4.7 Summary of Cluster 0 Demographic Characteristics

Characteristic	**Description**
Dietary Preference	Majority are non-restricted (85.79%)
Education	Over half have a high school education or less (56.96%)
Gender	More females (57.80%)
Income	Balanced distribution with a slight skew towards lower income
Race/Ethnicity	Predominantly White (64.21%), but with significant non-White representation (33.98%)
Age	Well-distributed across all age groups
Political Affiliation	Predominantly moderate (64.90%) with a significant liberal presence (22.14%)

Cluster 1

Table 4.8 Summary of Cluster 1 Demographic Characteristics

Characteristic	**Description**
Dietary Preference	Majority non-restricted (92.20%)
Education	Similar to Cluster 0, with the majority having lower educational attainment
Gender	Slightly more females (54.77%)
Income	Higher income distribution, with 36.57% earning above 100k
Race/Ethnicity	Predominantly White (77.12%) with less diversity
Age	Majority are older adults (59.79%)
Political Affiliation	Strong conservative presence (37.78%), though still majority moderate (45.93%)

Cluster 2

Table 4.9 Summary of Cluster 2 Demographic Characteristics

Characteristic	**Description**
Dietary Preference	Less skewed towards non-restricted diets (73.56%)
Education	Higher educational attainment with 41.15% holding an associate's or bachelor's degree
Gender	Predominantly male (59.54%)
Income	Higher income, with 51.72% earning above 100k
Race/Ethnicity	More diverse, with 43.45% non-White
Age	Dominated by younger and middle-aged adults
Political Affiliation	Mostly moderate (48.05%) with a significant conservative element (32.18%)

Cluster 3

Table 4.10 Summary of Cluster 3 Demographic Characteristics

Characteristic	**Description**
Dietary Preference	Majority non-restricted (89.12%)
Education	Similar to Clusters 0 and 1, with the majority having lower educational attainment
Gender	More females (57.53%)
Income	Balanced income distribution with a slight skew towards middle income
Race/Ethnicity	Predominantly White (73.03%)
Age	Majority are older adults (49.93%)
Political Affiliation	Predominantly moderate (59.17%), with balanced conservative and liberal representation

Overall Observations

Table 4.11 Summary of Cluster Characteristics

Cluster	**Description**
Cluster 0	Fairly balanced in most demographics, with a slight skew towards lower education and middle income. Politically moderate with a significant liberal presence.
Cluster 1	Higher income, predominantly White, older, and politically conservative.
Cluster 2	More diverse, higher income, higher education, and younger. Politically moderate with a significant conservative presence.
Cluster 3	Similar to Cluster 1 in terms of income and education but with a more balanced political affiliation and an older population.

4.4.2 Comparing Cluster Features with Demographics

In reviewing the demographic characteristics of each cluster, several patterns emerge that align with their respective attitudes towards AI.

Cluster 0: Consists of individuals who are moderately supportive of AI, as reflected in their general openness to AI activism and treatment. The cluster is demographically balanced, with a slight skew towards higher income and a predominantly moderate political affiliation. This mix of middle-aged to older adults, with varied education levels and a significant liberal presence, suggests a group that is cautiously optimistic about AI, likely influenced by their socioeconomic stability and moderate views.

Cluster 1: Stands out as more cautious and skeptical towards AI. Predominantly older, affluent, and politically conservative, this group exhibits strong resistance to AI activism and low moral concern for AI. The demographic makeup—predominantly White, with a lower level of educational attainment—aligns with a tendency to view AI with suspicion and to prioritize traditional values, leading to a generally more cautious approach.

Cluster 2: Notably younger, more diverse, and better educated, with a strong skew towards higher income. This cluster is characterized by its proactive and supportive stance towards AI, reflecting high levels of moral concern and openness to AI having mind-like qualities. The demographic profile of this group, with its higher education levels and more liberal-leaning political views, supports a forward-thinking perspective on AI, likely driven by the group's optimism about technological advancements and their potential societal benefits.

Cluster 3: Presents a more neutral stance on AI, with some skepticism about AI's capabilities and a cautious approach to its potential risks. This cluster is older, moderately affluent, and predominantly White, with a balanced mix of political affiliations. The demographic profile suggests a group that is neither strongly for nor against AI, instead reflecting a cautious yet open-minded perspective, shaped by their age, income distribution, and moderate political views.

4.4.3 Comparing Cluster Results with the AIMS Survey

The 2023 Artificial Intelligence, Morality, and Sentience (AIMS) Survey provided a broad overview of public sentiment toward AI sentience and rights, revealing insightful demographic trends within the survey dataset. This research builds on these findings by offering a detailed breakdown of how specific demographic groups align with varying attitudes toward AI. Both analyses indicate a complex and evolving public opinion, where optimism about AI's potential is tempered by substantial concerns, particularly among older, conservative, and less diverse demographics.

Both the AIMS survey and the clustering analysis suggest that a significant portion of the public is open to the possibility of AI sentience. The AIMS survey revealed that nearly 40% of Americans believe that developing sentient AI is possible, while less than 25% believe it is impossible. The clustering analysis adds depth to this finding by showing that belief in AI sentience is particularly strong among Cluster 2, which is characterized by younger, more diverse, and more educated individuals. In contrast, Cluster 1, composed of older and more conservative groups, shows greater skepticism toward AI sentience.

The AIMS study highlighted a reluctance among the public to grant AI legal rights, despite a general agreement that sentient AI deserves respectful treatment[2]. This reluctance is mirrored in Cluster 1 from the analysis, where there is strong resistance to AI activism and low moral concern for AI. On the other hand, Cluster 2 demonstrates a more proactive stance on AI rights, aligning with the AIMS survey's findings that certain demographics, particularly younger and more educated individuals, are more open to considering AI's moral status.

The AIMS survey found substantial resistance to advanced AI, with 63% of respondents supporting a ban on AI smarter than humans and 69% favoring a ban on sentient AI. This

[2] 71% agree on respectful treatment, but only 38% support legal rights

broad resistance is particularly evident in Cluster 1 of the analysis, where skepticism towards AI is most pronounced. The demographic profile of this cluster—older, conservative, and predominantly White—closely aligns with the survey's findings that these groups are more likely to support restrictions on AI development.

Both the AIMS survey and the clustering analysis highlight the influence of demographic factors such as age, education, and political orientation on attitudes toward AI, drawing attention to distinct trends in beliefs across different groups. While the AIMS survey provided a general overview, the clustering analysis offers more granular insights, demonstrating how these demographic factors converge to form distinct groups with varying degrees of support or resistance to AI.

4.5 Limitations

This study faced several limitations that should be acknowledged. One of the primary limitations was the demographic imbalances present in the dataset. Although the demographic features—such as dietary preference, education level, gender, income bracket, race/ethnicity, religion, age group, and political affiliation—were not directly used in the clustering process, they may still have influenced the interpretation of the demographic distributions within each cluster. Moreover, the lack of distinguishing between "sentience" and "consciousness" may have added complexity to the analysis, as these terms were sometimes used interchangeably, potentially affecting the clarity of the survey results. The study's reliance on existing survey data from the AIMS Survey presents limitations related to the scope and depth of the data. The survey's design and the questions asked may have influenced the responses, potentially leading to biases in the clustering analysis.

Another significant limitation lies in the clustering process itself. Clustering involves data compression, which can both highlight and obscure important information. The chosen clustering algorithms and feature reduction techniques were carefully evaluated using metrics such as the Silhouette Score and Davies-Bouldin Index, but the results are still subject to the inherent biases and limitations of these methods. As a result, some meaningful patterns may have been overlooked or misinterpreted.

4.6 Future Research

Future research should address the limitations identified in this study to enhance the robustness and generalizability of the findings. To begin with, more balanced and diverse datasets should be collected to mitigate demographic imbalances and ensure that the clusters identified truly reflect the full spectrum of public opinions on AI sentience and rights. Longitudinal studies of this sort could provide insights into how these perceptions evolve over time, offering a more dynamic understanding of public attitudes.

Additionally, future studies could explore more sophisticated clustering techniques and alternative feature reduction methods to capture more nuanced patterns in the data. With larger datasets, techniques such as deep learning-based clustering or hybrid approaches combining multiple algorithms could be investigated to improve the accuracy and interpretability of the results.

Clarifying the terminology used in surveys and research is also important. Developing more precise definitions of "sentience" and "consciousness" in the context of AI will help ensure consistency across studies and improve the comparability of results.

Lastly, expanding the scope of research to include qualitative methods, such as interviews or focus groups, could provide deeper insights into the underlying reasons behind people's perceptions of AI sentience and rights. This mixed-methods approach would complement the quantitative findings and offer a more comprehensive understanding of the ethical implications of AI development.

By addressing these areas in future research, we can develop a more nuanced and informed perspective on AI sentience and rights.

Chapter 5

Conclusion

5.1 Summary of Findings

The primary goal of this thesis was to evaluate public perceptions of AI sentience and rights through an in-depth analysis of survey data using unsupervised clustering algorithms. The study began by exploring the background of AI sentience and moral rights, highlighting cases where individuals have been misled into attributing sentience to artificial beings. It also recognized that while various theories of mind provide diverse perspectives on AI sentience and moral status, the lack of a clear expert consensus adds complexity to the issue. This ambiguity complicates the ethical landscape surrounding AI, highlighting the need for a deeper understanding of public perceptions to guide future decisions on how to responsibly integrate increasingly advanced AI systems into society. Furthermore, the study aimed to address the need for more empirical research on AI sentience and moral status, as AI's rapidly evolving capabilities continue to outpace our ethical frameworks and societal values.

5.2 Contributions to the Field of AI Sentience and Rights

This study sought to fill a gap in the existing literature by identifying the demographic factors associated with various beliefs about AI sentience and rights, thereby laying the groundwork for more focused and impactful future research.

The clustering analysis revealed several key findings:

K-Means Clustering with UMAP identified four distinct clusters of respondents, each characterized by varying degrees of caution, activism, and perception of AI. These clusters ranged from highly supportive groups that view AI positively to more cautious or skeptical groups that express significant concerns about AI's potential risks and moral implications.

The demographic analysis of these clusters provided additional context, showing the attitudes towards AI are which are linked to demographic factors such as age, education, income, and political affiliation. For instance, younger, more educated, and higher-income individuals were generally more supportive of AI, whereas older, less educated, and more conservative individuals tended to be more skeptical. The study's use of machine learning clustering algorithms, particularly K-Means and DBSCAN with UMAP, offered a novn-linear approach to uncovering hidden patterns in survey data. By doing so, the research not only identified distinct groups within the survey population but also sheds light on the underlying factors that drive their beliefs about AI. As AI systems continue to advance, understanding public sentiment will be crucial for policymakers, researchers, and ethicists. This study provides a foundation for more informed discussions on AI rights and helps anticipate potential areas of public concern that may arise as AI capabilities grow.

5.3 Final Thoughts and Recommendations

Moving forward, several recommendations emerge from this research: Future studies should continue to explore the public's evolving perceptions of AI, particularly as AI technology becomes more integrated into daily life. Longitudinal studies, such as the AIMS Survey, could provide valuable insights into how these perceptions change over time. Engaging with the public on the topic of AI sentience and rights is also essential. Public education campaigns could help demystify AI technologies and address common misconceptions, potentially leading to more informed public opinions. Policymakers should consider the diverse range of public opinions identified in this study when developing regulations and ethical guidelines for AI rights. If future studies reveal that more and more people attribute sentience to AI, policies that are reflective of public sentiment should be considered. In conclusion, this research represents a step towards understanding the complex relationship between AI development and public perception. By identifying and analyzing distinct clusters of opinion, it provides a clearer picture of the current landscape of beliefs about AI sentience and rights, offering guidance for future research, public engagement, and policy development. As we stand on the cusp of potentially transformative advancements in AI, it is imperative that we continue to explore these issues with the foresight that they demand.

References

[1] Blake Lemoine, Google, and searching for souls in the algorithm - Vox. https://www.vox.com/recode/2022/6/30/23188222/silicon-valley-blake-lemoine-chatbot-eliza-religion-robot.

[2] Blake Lemoine: Google fires engineer who said AI tech has feelings. https://www.bbc.com/news/technology-62275326.

[3] (1890). *The Monadology, 1714.*, pages 218–232. Tuttle, Morehouse & Taylor, New Haven.

[4] Ahmed, M., Seraj, R., and Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*, 9(8):1295.

[5] Arthur, D. and Vassilvitskii, S. (2006). How slow is the *k* -means method? In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, pages 144–153, Sedona Arizona USA. ACM.

[6] B, K. (2020). A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering. *International Journal of Emerging Trends in Engineering Research*, 8(5):1600–1604.

[7] Baars, B. J. (1995). *A Cognitive Theory of Consciousness*. Cambridge University Press, Cambridge, reprinted edition.

[8] Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. In *Progress in Brain Research*, volume 150, pages 45–53. Elsevier.

[9] Blake, M. (2024). Why large language models are not sentient. Accessed: 2024-08-30.

[10] Block, N. and Fodor, J. A. (1972). What psychological states are not. *Philosophical Review*, 81(April):159–81.

[11] Bloom, P. (2010). How do morals change? *Nature*, 464(7288):490–490.

[12] Boyko, N. and Tkachyk, O. (2023). Hierarchical clustering algorithm for dendrogram construction and cluster counting. *Informatics and mathematical methods in simulation*, 13:5–15.

[13] Buyrukoğlu, S. and Akbaş, A. (2022). Machine Learning based Early Prediction of Type 2 Diabetes: A New Hybrid Feature Selection Approach using Correlation Matrix with Heatmap and SFS. *Balkan Journal of Electrical and Computer Engineering*, 10(2):110–117.

[14] Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Philosophy of Mind Series. Oxford University Press, New York.

[15] Chalmers, D. J. (2000). What is a neural correlate of consciousness? In Metzinger, T., editor, *Neural Correlates of Consciousness*, pages 17–39. MIT Press.

[16] Chalmers, D. J. (2023). Could a Large Language Model Be Conscious? *Boston Review*.

[17] Crick, F. and Koch, C. (1990). Towards a neurobiological theory of consciousness.

[18] Danaher, J. (2020a). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*, 26(4):2023–2049.

[19] Danaher, J. (2020b). Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. *Science and Engineering Ethics*, 26(4):2023–2049.

[20] Deng, D. (2020). DBSCAN Clustering Algorithm Based on Density. In *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, pages 949–953, Hefei, China. IEEE.

[21] Dennett, D. C. (1991). *Consciousness Explained*. Back Bay Books. Little, Brown, Boston, 1. paperback ed edition. Includes bibliographical references (p. 469 - 491) and index.

[22] Descartes, R., Miller, V. R., and Miller, R. P. (2009). *Principles of Philosophy*. Wilder Publications.

[23] Epstein, R. (2007). From Russia, with Love. https://www.scientificamerican.com/article/from-russia-with-love/.

[Ester et al.] Ester, M., Kriegel, H.-P., and Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

[25] Fernandez, I., Kyosovska, N., Luong, J., and Mukobi, G. (2024). AI Consciousness and Public Perceptions: Four Futures.

[26] Francken, J. C., Beerendonk, L., Molenaar, D., Fahrenfort, J. J., Kiverstein, J. D., Seth, A. K., and Van Gaal, S. (2022). An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neuroscience of Consciousness*, 2022(1):niac011.

[27] Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12):11–39.

[28] Gandhi, J., Goyal, R., Guha, J., Pithawala, K., and Joshi, S. (2021). Comparative Study on Hierarchical and Density based Methods of Clustering using Data Analysis. *SSRN Electronic Journal*.

[29] Gordon, J.-S. and Gunkel, D. J. (2022). Moral Status and Intelligent Robots. *The Southern Journal of Philosophy*, 60(1):88–117.

[30] Gu, Z. (2022). Complex heatmap visualization. *iMeta*, 1(3):e43.

[31] Harris, J. and Anthis, J. R. (2021). The Moral Consideration of Artificial Entities: A Literature Review. *Science and Engineering Ethics*, 27(4):53.

[32] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441.

[33] Humaira, H. and Rasyidah, R. (2020). Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm. In *Proceedings of the Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*, Padang, Indonesia. EAI.

[34] Janet Pauketat (2023). Artificial Intelligence, Morality, and Sentience (AIMS) Survey.

[35] Kahane, G., van der Merwe, M., Zohny, H., Daniel, M., Finnveden, L., Gloor, L., Alskelung, U., Hornbol, V., Dewey, D., Muehlhauser, L., Babcock, J., Bloom, R., Luczkow, V., Beckstead, N., Greaves, H., Cotton-Barratt, O., Dafoe, A., and Cowley, W. (2021). Sharing the world with digital minds 1.

[36] Lemoine, B. (2022). Is LaMDA Sentient? — an Interview.

[37] Lima, G., Kim, C., Ryu, S., Jeon, C., and Cha, M. (2020). Collecting the Public Perception of AI and Robot Rights. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–24.

[38] Locke, J. (1948). An essay concerning human understanding, 1690. In Dennis, W., editor, *Readings in the History of Psychology.*, pages 55–68. Appleton-Century-Crofts, East Norwalk.

[39] Long, R. (2024). How not to 'debunk' AI sentience.

[40] McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

5.3.1 Other

Reference implementation available at http://github.com/lmcinnes/umap.

[41] Murugesan, N., Cho, I., and Tortora, C. (2021). Benchmarking in Cluster Analysis: A Study on Spectral Clustering, DBSCAN, and K-Means. In Chadjipadelis, T., Lausen, B., Markos, A., Lee, T. R., Montanari, A., and Nugent, R., editors, *Data Analysis and Rationality in a Complex World*, pages 175–185. Springer International Publishing, Cham.

[42] Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4):435.

[43] Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., and Mian, A. (2024). A Comprehensive Overview of Large Language Models.

[44] Nick Bostrom and Carl Shulman (2020). Propositions Concerning Digital Minds and Society.

[45] Ntani, G., Inskip, H., Osmond, C., and Coggon, D. (2021). Consequences of ignoring clustering in linear regression. *BMC Medical Research Methodology*, 21(1):139.

[46] Papineau, D. (2002). *Thinking about Consciousness*. Oxford University PressOxford, 1 edition.

[47] Parker Pearson, M. (2012). *The Archaeology of Death and Burial*. The History Press, Stroud, repr edition. Literaturverz. S. [217] - 243.

[48] Patel, K. M. A. and Thakral, P. (2016). The best clustering algorithms in data mining. In *2016 International Conference on Communication and Signal Processing (ICCSP)*, pages 2042–2046, Melmaruvathur, Tamilnadu, India. IEEE.

[49] Pauketat, J. V., Ladak, A., and Anthis, J. R. (2022a). Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2021.

[50] Pauketat, J. V., Ladak, A., and Anthis, J. R. (2022b). Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2021.

[51] Pauketat, J. V., Ladak, A., and Anthis, J. R. (2023). Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2023 Update.

[52] Piccinini, G. (2004). Functionalism, computationalism, and mental contents. *Canadian Journal of Philosophy*, 34(3):375–410.

[53] Putnam, H. (1960). Minds and machines. In Hook, S., editor, *Dimensions Of Mind: A Symposium.*, pages 138–164. NEW YORK University Press.

[54] Rabasovic, M., Pavlovic, D., and Sevic, D. (2023). Analysis of laser ablation spectral data using dimensionality reduction techniques: PCA, t-SNE and UMAP. *Contributions of the Astronomical Observatory Skalnaté Pleso*, 53(3).

[55] Rehman, S. U., Asghar, S., Fong, S., and Sarasvady, S. (2014). DBSCAN: Past, present and future. In *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, pages 232–238, Bangalore, India. IEEE.

[56] Rey, G. (1997). *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. Wiley-Blackwell, Cambridge, Mass.

[57] Scott, A. E., Neumann, D., Niess, J., and Woźniak, P. W. (2023). Do You Mind? User Perceptions of Machine Consciousness. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–19, Hamburg Germany. ACM.

[58] Sebo, J. and Long, R. (2023). Moral consideration for AI systems by 2030. *AI and Ethics*.

[59] Shevlin, H. (2021). How Could We Know When a Robot was a Moral Patient? *Cambridge Quarterly of Healthcare Ethics*, 30(3):459–471.

[60] Strawson, G. (2004). Real intentionality. *Phenomenology and the Cognitive Sciences*, 3(3):287–313.

[61] Tan, Y., Rérolle, S., Lalitharatne, T. D., Van Zalk, N., Jack, R. E., and Nanayakkara, T. (2022). Simulating dynamic facial expressions of pain from visuo-haptic interactions with a robotic patient. *Scientific Reports*, 12(1):4200.

[62] Tye, M. (1999). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Representation and Mind. MIT Press, Cambridge, Mass., 3rd print edition. A Bradford Book Originally published: 1995 Literaturangaben.

[63] Vacanti, N. M. (2019). The Fundamentals of Constructing and Interpreting Heat Maps. In Fendt, S.-M. and Lunt, S. Y., editors, *Metabolic Signaling*, volume 1862, pages 279–291. Springer New York, New York, NY.

[64] van der Maaten, L. and Hinton, G. (2008). Viualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.

[65] Venkat, N. (2018). The Curse of Dimensionality: Inside Out.

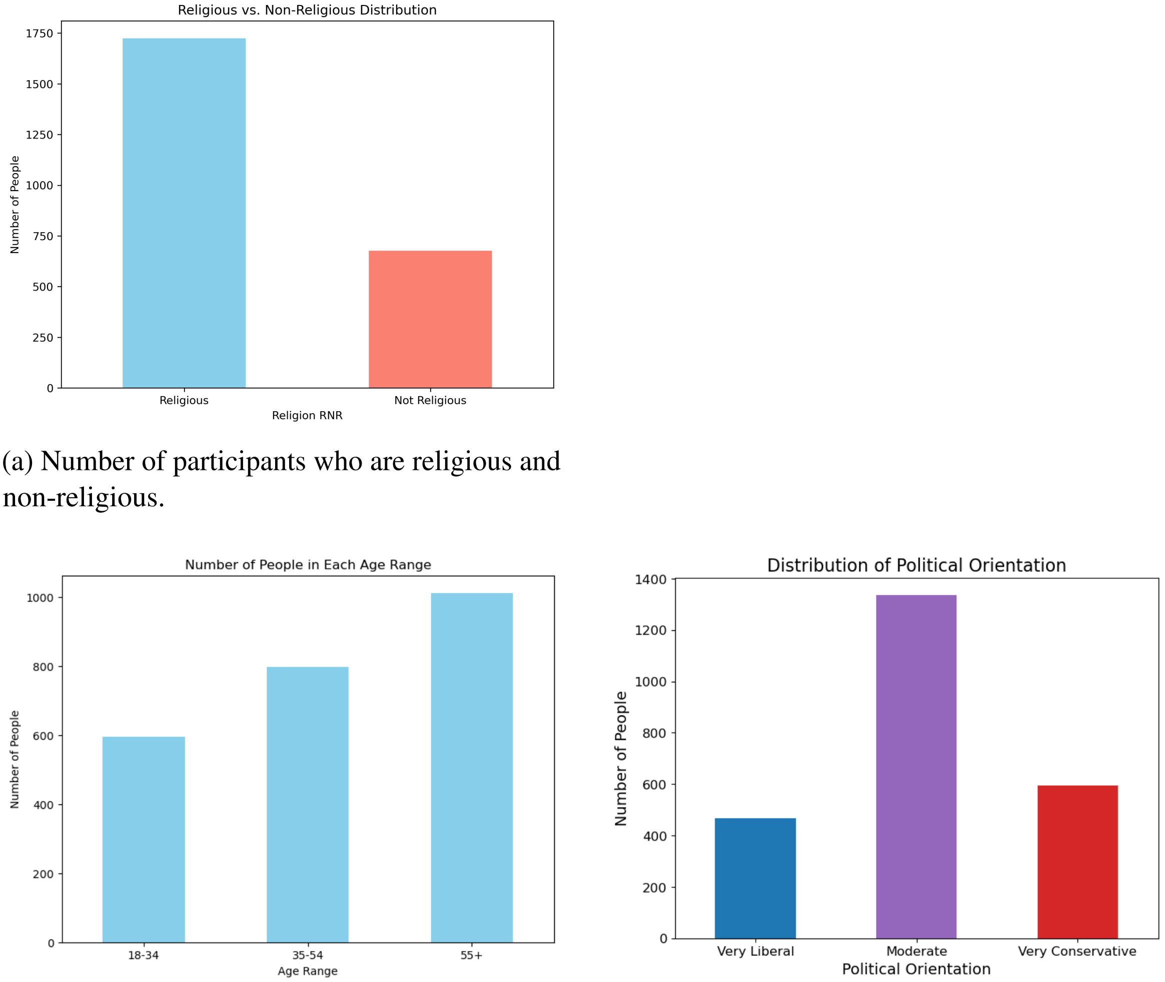
[66] Williams, N. M. (2008). Affected Ignorance And Animal Suffering: Why Our Failure To Debate Factory Farming Puts Us At Moral Risk. *Journal of Agricultural and Environmental Ethics*, 21(4):371–384.

[67] Xu, D. and Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2(2):165–193.

Appendix A

Appendices

A.1 Visualization of Participant Demographic Imbalances

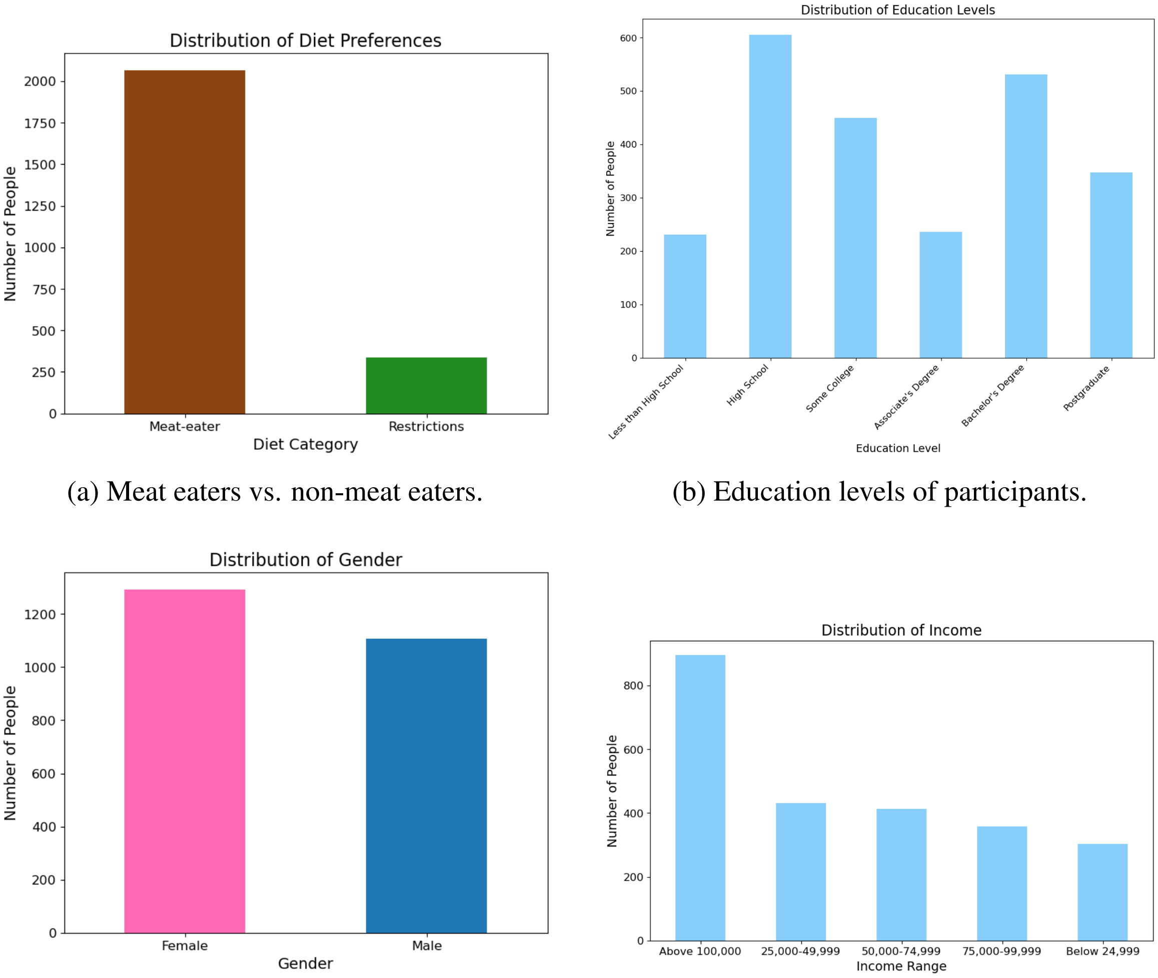


(a) Number of participants who are religious and non-religious.

(b) Age range of participants.

(c) Political alignment of participants.

Fig. A.1 Demographic distributions of survey participants.



(a) Meat eaters vs. non-meat eaters.

(b) Education levels of participants.

(c) Gender distribution of participants.

(d) Income distribution of participants.

Fig. A.2 Additional demographic distributions of survey participants.

Appendix B

Methodology Results

B.1 Quantitative Feature Reduction Results

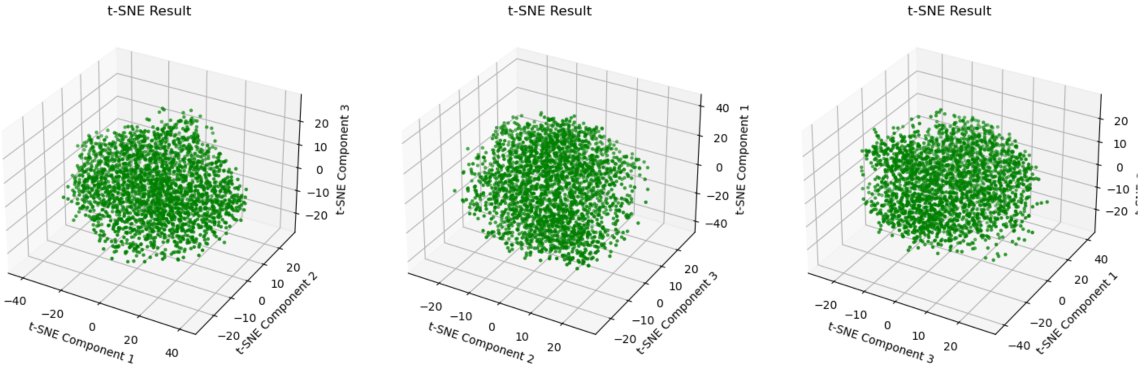


Fig. B.1 t-SNE visualization with 3 components.

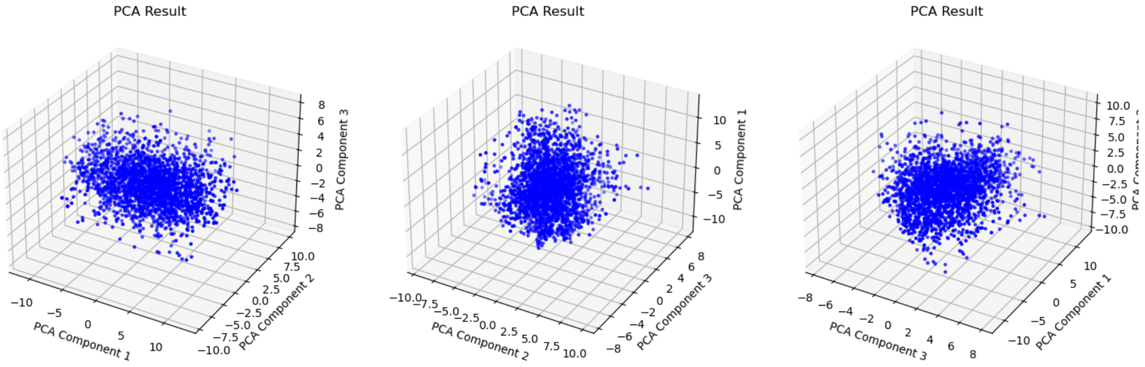


Fig. B.2 PCA visualization with 3 components.

Appendix C

Detailed Qualitative Cluster Evaluation

C.1 Qualitative Feature Mean Heatmap

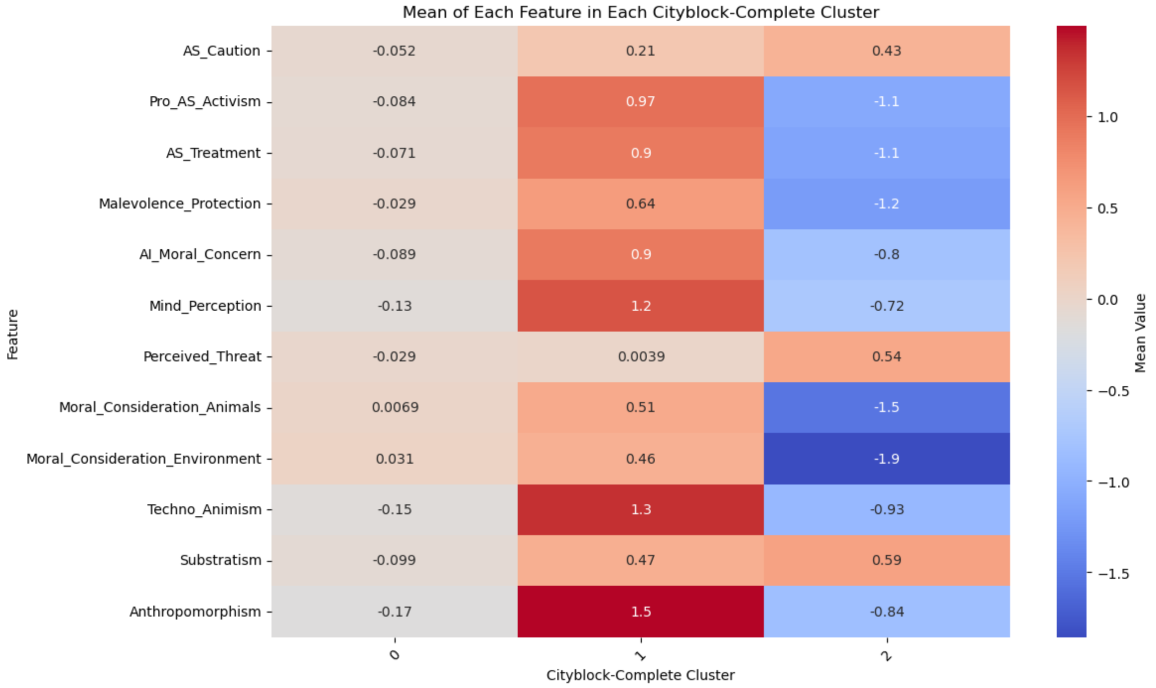


Fig. C.1 Visualizing the mean of each cluster in the Hierarchical Clustering (Cityblock+Complete) Results.

C.2 Hierarchical Clustering Test Results

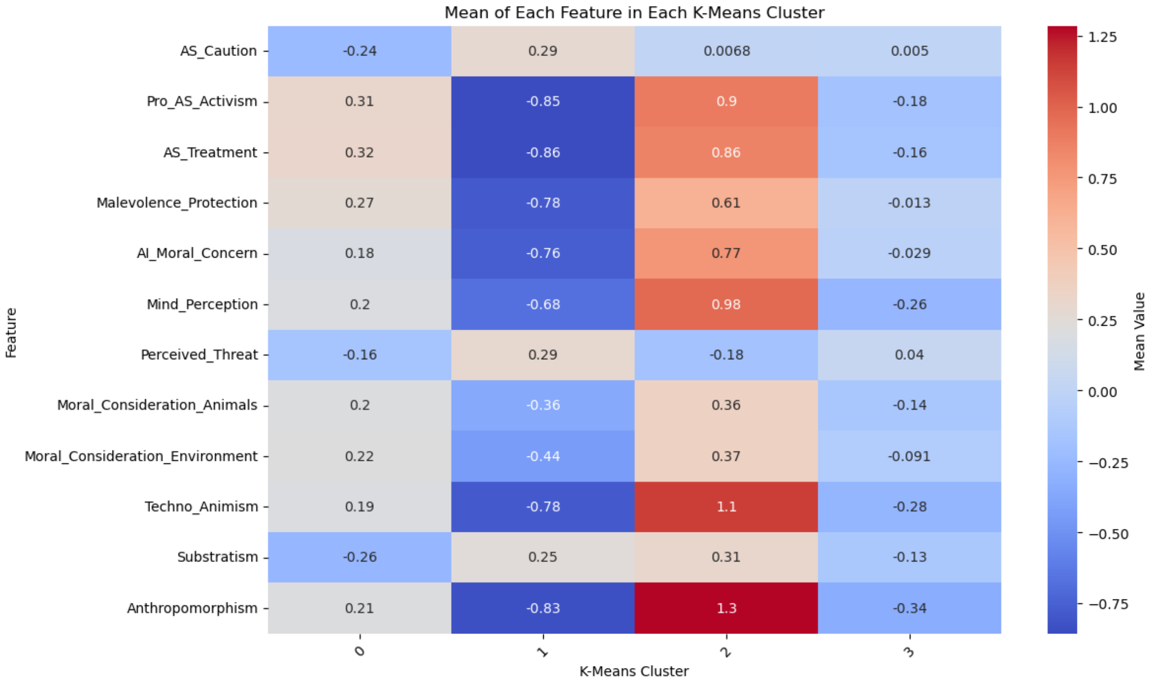


Fig. C.2 Visualizing the mean of each cluster in K-Means (UMAP) Clustering Results.

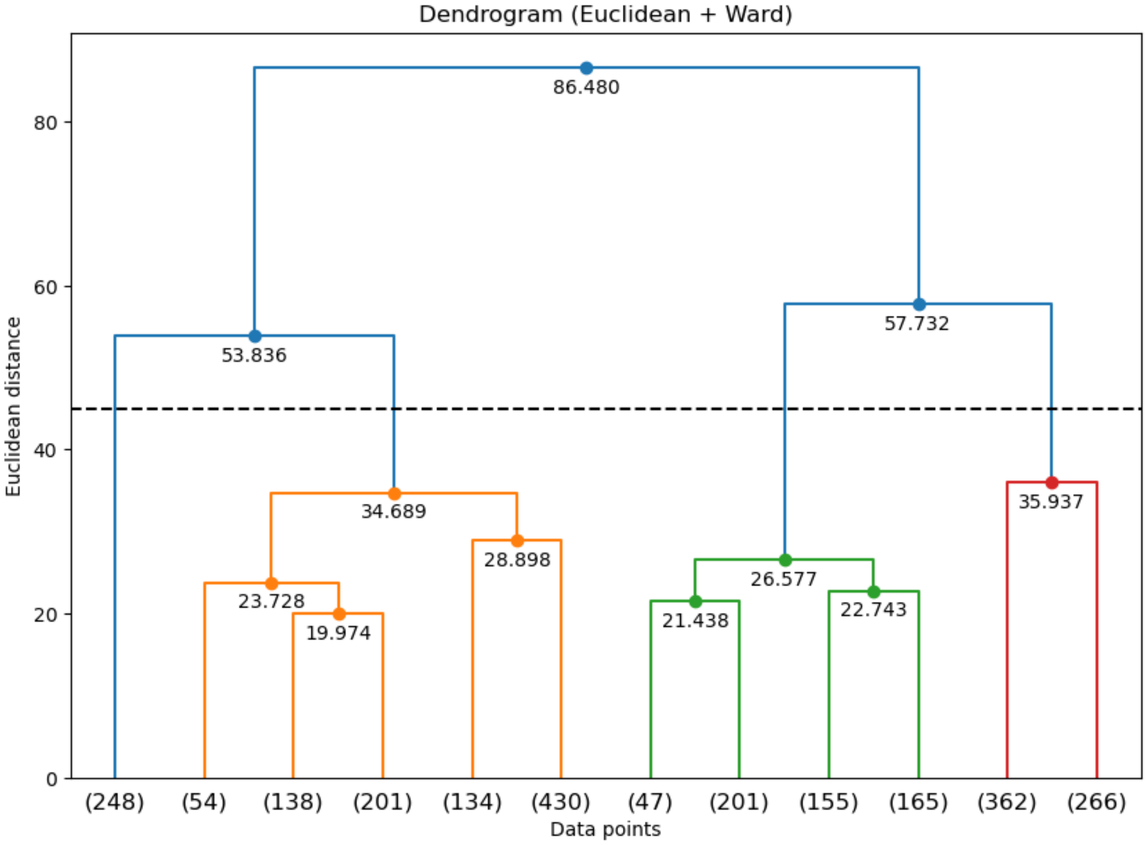


Fig. C.3 Euclidean and Ward Hierarchical Clustering.

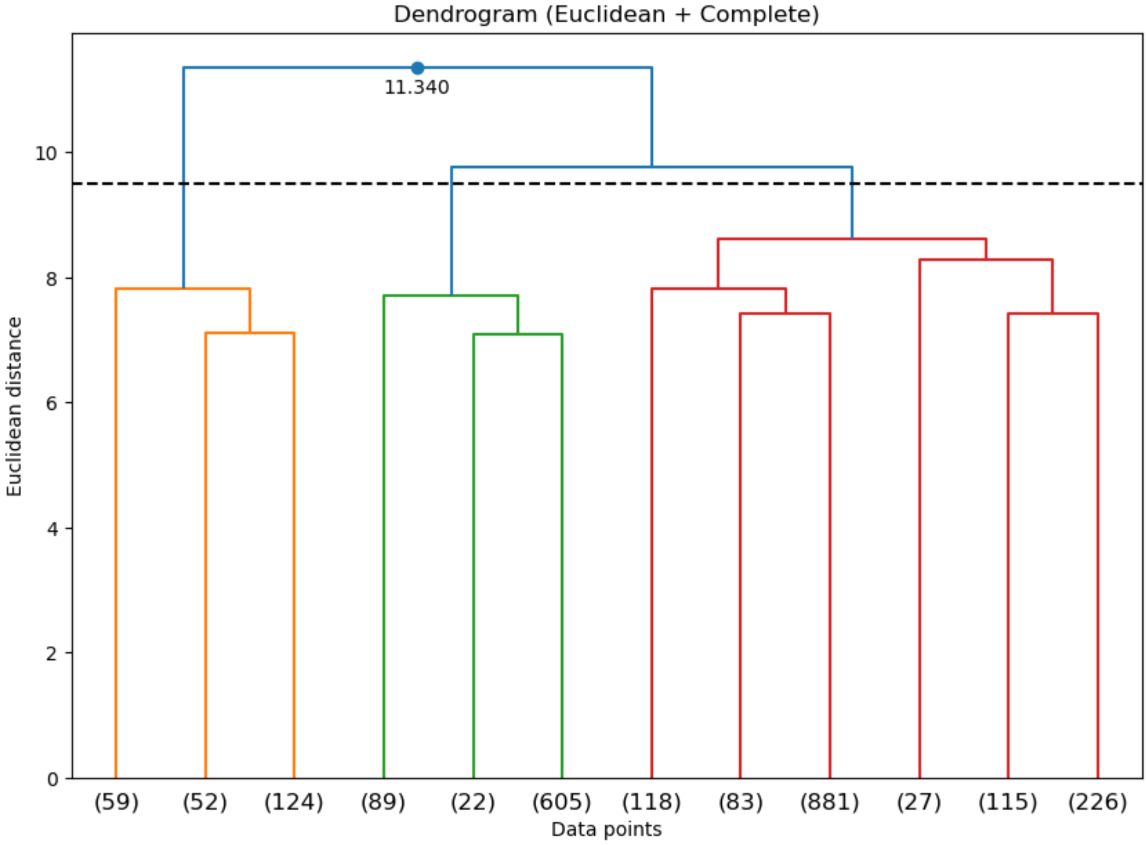


Fig. C.4 Euclidean and Complete Hierarchical Clustering.