Human TECHNOLOGY

ISSN: 1795-6889

https://ht.csr-pub.eu Volume 21(2), September 2025, 317–338

IS THERE CONSISTENCY IN ETHICAL SENSITIVITY IN ARTIFICIAL INTELLIGENCE? A REVIEW OF LANGUAGE MODELS

Hasan Tutar
Bolu Abant Izzet Baysal University, Bolu, Turkey;
Azerbaijan State University of Economics (UNEC), Baku, Azerbaijan
ORCID 0000-0001-8383-1464

Svitlana Bilan
Rzeszów University of Technology, Rzeszów, Poland
ORCID 0000-0001-9814-5459

Ümit Şentürk
Bolu Abant İzzet Baysal University
Bolu, Turkey
umit.senturk@ibu.edu.tr
ORCID 0000-0001-9610-9550

Abstract: *This study examines the structural and content consistency of large language models (LLMs) in ethical decision making with a qualitative approach. Responses to basic ethical themes such as "justice", "non-maleficence", "autonomy", "impartiality", and "goodness" were evaluated using the thematic analysis method of Braun and Clarke (2006). The three-stage coding process analyzed empathy patterns, contextual transitions, and relationships between themes. The findings, supported by Python-supported frequency and variation analyses, revealed that the models exhibited high empathy and solution determination in the themes of "inclusiveness" and "communication" but low structural consistency in the themes of "religion" and "disability". The responses to the same ethical theme in different contexts were determined to carry semantic shifts. This original study emphasizes that ethical sensitivity should be evaluated based on patterns.*

Keywords: *ethical consistency, large language models (LLM), thematic analysis, empathy pattern, contextual variation.*

DOI: https://doi.org/10.14254/1795-6889.2025.21-2.4

INTRODUCTION

With the rapid development of artificial intelligence systems, evaluating these technologies with technical success and ethical criteria is necessary. Large language models (LLMs), in particular, are increasingly used in sensitive areas of society such as health, law, education and content moderation, and this situation reveals the need for decision mechanisms to be compatible not only with data but also with ethical principles and norms (Jobin et al., 2019). A study has shown that 78% of users expect consistency in ethical decisions made by artificial intelligence-based systems (Mökander et al., 2022). However, it has been shown that current LLM systems give contradictory answers to the same ethical question in different contexts. This suggests that the models are inconsistent and predictable regarding ethical sensitivity. The social impact of these technological developments is not limited to the individual level, but also has institutional and systemic consequences. For example, algorithmic decisions in immigration policies or justice systems can cause inconsistency in ethical decisions, leading to irreversible harm to individuals. In addition, value distortions in the outputs of models create trust problems in a wide range of areas, from environmental sustainability to social equality (Gabriel, 2020; Cath, 2018; Hagendorff, 2022; Binns, 2018; Mittelstadt et al., 2019; Seniutis et al., 2025). The increasing integration of AI systems extends to operational domains such as logistics and warehouse management, where artificial intelligence technologies demonstrate significant potential for optimizing processes while raising questions about ethical implementation and decision-making consistency (Angammana & Jayawardena, 2022). Similarly, the development of sustainable and smart logistics systems requires careful consideration of ethical frameworks in autonomous decision-making processes (Vovk et al., 2025). Most of the literature discusses the codability of ethical principles in artificial intelligence systems, but does not sufficiently address how these principles evolve and vary in context (Andrus et al., 2021; Weidinger et al., 2022). For this reason, this study was designed to question thematic coherence in ethical sensitivity and critically examine the responses of language models in this direction.

Studies on assessing ethical consistency in AI systems have gained significant momentum in the post-2018 period. One of the first systematic approaches in this field is the study by Raji et al. (2020) that shows the direct impact of algorithmic auditing processes on ethical outcomes. However, most of these studies have not examined the ethical sensitivity in the outputs of AI systems in terms of normative consistency (Crawford & Paglen, 2021). These applications underscore the necessity for comprehensive ethical frameworks that address both technical performance and normative alignment. In some studies, it has been observed that the responses of LLMs to ethical dilemmas often vary depending on the context and produce inconsistent outputs according to predefined ethical principles (Zhou et al., 2023). This suggests that the models are not trained in a specific moral framework, and their ethical decisions are based mainly on data-based correlations. Nevertheless, some studies have shown that language models can provide morally acceptable answers in specific contexts (Tamkin et al., 2021; Arora et al., 2023). Most of these studies have been conducted under controlled experimental conditions with limited data, which raises the issue of generalizability. On the other hand, there has been no sufficient comparative analysis of how ethical principles are interpreted in different cultures; this deficiency makes the cultural context blindness in the ethical responses of LLMs visible (Santurkar et al., 2023; Birhane &

Prabhu, 2021). In addition, in the existing literature, studies that systematically classify the ethical responses given by LLMs are pretty limited (Sheng et al., 2021; Bommasani et al., 2021). Therefore, this study emphasizes the necessity of considering ethical responses thematically based on content and structural consistency.

Although studies on the consistency of large language models in ethical response production are increasing in the literature, these studies are mostly limited to measuring superficial compliance in the systems' outputs. Many studies discuss whether it is possible to train language models within specific normative frameworks, but it does not systematically examine the consistency of the responses of these systems to different ethical themes at the thematic level (Uesato et al., 2022; Kirk et al., 2023). In addition, evaluating the ethical sensitivities of models only under broad categories such as "do not harm" and "do not help" makes the analysis of contextual differences difficult and does not allow for in-depth evaluation. This situation encourages superficial analysis of the models' responses, especially in the face of multi-layered moral dilemma questions, while ignoring the normative variability in the thematic structure (Perez et al., 2022; Shevlane et al., 2023). Analyzing ethical consistency at the thematic level can contribute to artificial intelligence systems' transparency, social acceptance, and regulation. Ethical ruptures that may occur in the interaction of user groups with different cultural norms with models can only be predicted with such thematic depth analyses (Leins et al., 2023; Zheng et al., 2023). At this point, there is still a lack of systematic studies in the literature that evaluate the ethical consistency of language models in a comprehensive, comparative, and pattern-based manner (Askell et al., 2021; Schramowski et al., 2022; Ganguli et al., 2023). In order to fill this gap, this study aims to thematically code the answers given by large language models to ethical questions, make comparisons between different ethical domains, and classify possible contradictions. In addition, this analysis aims to provide inferences about the ethical consistency capacity of models by measuring contextual shifts in ethical response production.

The main problem of this study is to question the inconsistencies exhibited by large language models in ethical decision-making on a thematic level. Although LLMs can provide morally reasonable answers in specific contexts, it is largely unknown whether these answers are consistent across similar themes (Glaese et al., 2022). The study's basic assumption is that although the answers given by language models to ethical questions may seem reasonable, they do not show significant consistency in thematic integrity. There are significant variations in the ethical answers of LLMs depending on the context, and these variations transform into structurally recurring patterns in specific ethical themes (Lin et al., 2022; Zhang et al., 2023). This study aims to comparatively analyze the ethical sensitivity capacity of LLMs within the framework of themes such as “justice”, “non-maleficence”, “autonomy”, “impartiality”, and “goodness”. The research aims to contribute significantly to this context's theoretical and applied levels. At the theoretical level, it will allow for a reinterpretation of the concept of ethical consistency in a principle-based and pattern-based manner. As a contribution to the literature, this study reveals the qualitative structure of ethical response production using thematic discourse analysis and introduces a methodological approach rarely used in this field (Ji et al., 2023; Maki et al., 2023). Identifying ethical inconsistencies in model outputs can provide critical feedback for regulatory processes and LLM education (Gehrmann et al., 2023; Chiang et al., 2023; Scholte et al., 2023). The implications of the results will be multi-

layered, encompassing not only AI engineering but also human-centered design, ethical audit systems, and cultural adaptation strategies.

The research is structured within a qualitative research design framework and is based on the thematic analysis method to analyze the ethical consistency of large language models. Thematic coding will be based on Braun and Clarke's (2006) six-stage thematic analysis model. Following this structure, model outputs will be collected by creating a data set and then subjected to content analysis. The review process will progress in four basic stages: In the first part, structured questions representing ethical themes will be created and directed to the model. In the second part, the responses will be coded using qualitative analysis software, and patterns will be extracted for each theme. In the third part, these patterns will be evaluated in terms of consistency; different responses given on the same theme will be compared. In the last stage, the ethical sensitivity level of the model will be subjected to a general assessment based on the analysis results (Nowell et al., 2017; Vaismoradi et al., 2016). The main question of the research is as follows: Can large language models provide consistent and normatively acceptable answers to questions given under different ethical themes? This question will be questioned regarding technical accuracy and the context of the principled integrity of ethical decisions (Guest et al., 2012; Terry et al., 2017). In this respect, the study aims to make an original contribution to discovering contextual ethical patterns in LLM outcomes.

LITERATURE REVIEW

The three concepts that form the basis of this study -ethical consistency, thematic response pattern, and large language models (LLM) -have been at the center of interdisciplinary research in recent years. Ethical consistency responds to normative situations with similar principles (Floridi & Cowls, 2019). This concept is addressed in a socio-technical framework, especially to increase the reliability of decision-making in artificial intelligence systems (Binns et al., 2018; Mökander et al., 2021; Lyeonov et al., 2024). While ethical consistency has been studied in human-based cognitive decisions in the past, it has recently gained measurable structures by being integrated into algorithmic decision systems. Thematic response pattern is an analytical structure used to determine how a model responds to specific moral themes. This concept is evaluated at the semantic level of the content and in terms of linguistic structure and discourse form. Thematic response pattern, the stability of thematic patterns, is a criterion that shows the level of compliance of the model with ethical principles (Gabriel, 2020; Schramowski et al., 2022; Seniutis et al., 2024). Conceptual framework for ethical artificial intelligence development in social services sector. In this context, pattern-based analyses are carried out with hybrid approaches combining behavioral modeling and discourse analysis. The third basic concept, large language models (LLMs), are machine learning systems that contain billions of parameters and can produce human-like language. LLMs first emerged as an advanced form of statistical language modeling, but today they have evolved into autonomous systems that produce outputs based on ethical, political, and social norms (Andrus et al., 2021; Bommasani et al., 2021; Weidinger et al., 2022). Therefore, when these three concepts are considered together, it is understood that ethical

performance should be measured not only by accuracy, but also by meaningful and consistent production.

Integrating AI systems into ethically responsible decision-making mechanisms is the product of deep-rooted technological, philosophical, and normative discussions. The evolution of the field has been shaped especially around themes such as "algorithmic justice" and "moral decision systems" (Mittelstadt et al., 2019). The first systematic approaches to the modelability of ethical systems in machines were developed by Binns (2018), and this work was accepted as a fundamental reference in establishing a bridge between political philosophy and machine learning. At the same time, the global study by Jobin, Ienca, and Vayena (2019), in which they comparatively analyzed 84 different ethical guidelines, was an important turning point that emphasized the diversity and contextuality of ethical principles in AI. The experimental study by Glaese et al. (2022) investigating the ethical alignment of LLMs through human feedback is groundbreaking in that it directly tested the level of compliance of model outputs with ethical norms. Similarly, the "normative framing" approach developed by Gabriel (2020) positioned AI ethics as a value-oriented system design, not just a set of technical rules. This approach directed the discussion of ethical consistency to the contextual evaluation of language model outputs. Weidinger et al. (2022) argued that issues such as social harm, security, and manipulation should be systematically examined regarding LLMs by mapping the ethical risk of large language models. In this vein, Schramowski et al. (2022) showed that models can "learn" human-like moral norms from data, but this learning is not limited to consistency. As a result, this research arises from the theoretical and empirical heritage summarized above. It aims to propose a new conceptual framework in the literature by structuring ethical sensitivity on a thematic level. In this context, the first sub-question of the research is as follows: "Can large language models exhibit meaningful ethical consistency at the structural level in different themes with ethical content and in-depth analysis?"

The basic concepts discussed in this study - ethical consistency, thematic pattern, context sensitivity, and large language models (LLMs) - were evaluated within a multilayered network of intertwined relationships. While ethical consistency is defined as the capacity to produce similar responses to the same normative theme despite contextual diversity, thematic pattern provides a framework that allows this capacity to be measured concretely (Hooker & Kim, 2019). The context sensitivity of LLMs has been frequently discussed in the literature as a dynamic variable that both fosters and limits the potential for ethical consistency (Birhane et al., 2022; Srivastava et al., 2022). Studies show that the relationship between these concepts can be grouped around three thematic trends. The first trend is the claim of normative fixity, which argues that ethical decisions should be based on predefined principles (Floridi et al., 2018). The second trend is the defense of context-based flexibility, which suggests that ethical responses may vary depending on the socio-cultural environment. The third trend is the hybrid approach based on inferential learning, where language models are argued to derive ethical norms from training data and model the inter-theme transitivity (Mokander & Floridi, 2021; Bai et al., 2022; Ganguli et al., 2023). The study will analyze the linguistic, contextual, and normative transitions of the responses between the themes to investigate whether ethical consistency is a structural rather than a content-based phenomenon (Ji et al., 2023; Chiang et al., 2023). The second sub-question of the study is:

"Do the responses of large language models to different ethical themes exhibit structural consistency in terms of context sensitivity and normative transitivity?"

Studies in artificial intelligence ethics have increased quantitatively in the last five years; publications focusing on the social and ethical impacts of large language models (LLMs) have become especially evident. The general trend in the literature is to measure the performance of systems in thematic areas such as harmlessness, prejudice, discrimination, and transparency (Weidinger et al., 2022; Raji et al., 2020; Bender et al., 2021). Institutional guidelines, technical guides, and normative frameworks are usually used as references in defining ethical principles (Jobin et al., 2019; Floridi & Cowls, 2019). Methodologically, artificial test sets created under laboratory conditions and survey-based user experiences are at the forefront. However, these trends cannot sufficiently examine ethical sensitivity's contextual, thematic, and patterned dimensions. Reducing ethical responses to only static principles of correctness or harmlessness makes the variability of language models in real-world interactions invisible (Brown et al., 2020; Tamkin et al., 2021; Birhane et al., 2022; Shevlane et al., 2023). On the other hand, qualitative methods such as pattern-based thematic analysis have been largely neglected in the literature. In particular, the effect of linguistic structures that show transition between different ethical themes on the normative consistency capacity within the model has not been sufficiently tested empirically (Andrus et al., 2021; Zhang et al., 2023). In line with this deficiency, this study aims to analyze patterned consistency by thematically coding the ethical responses of large language models. With this method, not only "what is said" but also "how and to what extent it is said consistently" will be revealed. In this context, the third sub-question of the research is as follows: When the ethical responses of large language models are evaluated at the thematic level, can they reveal structural pattern differences beyond superficial semantic similarities?

Most ethical assessments in AI are limited to either technical bias measures or normative perception tests based on individual user opinions. This methodological approach can indicate whether model responses align with ethical principles on the surface, but it falls short in assessing the contextual consistency of ethical responses (Binns et al., 2018; Raji et al., 2020). Furthermore, the "prompt-based" assessments used in many studies lack a systematic structure that would reduce the randomness of model responses. This creates serious criterion gaps in determining whether LLMs behave consistently across themes. One of the strengths is that studies such as Schramowski et al. (2022) and Glaese et al. (2022) demonstrate the learnability of moral norms in LLMs. However, most of these studies focus on generating "good" or "harmless" responses, neglecting the question of how consistent and principle-based responses are within an ethical framework (Perez et al., 2022; Weidinger et al., 2022; Gabriel, 2020). Furthermore, the diversity of ethical responses across different cultural or linguistic contexts has not been adequately tested (Chiang et al., 2023; Srivastava et al., 2022). Conflicting findings in the literature create uncertainty about the extent to which themes, especially those classified according to ethical principles, diverge in model outputs. For example, while some studies argue that models are strong in the principle of impartiality, others have shown that the same models exhibit apparent contradictions in justice-themed questions (Hooker & Kim, 2019; Bender et al., 2021). This suggests that ethical decisions should be evaluated not only in terms of outcome but also in terms of process. From this perspective, addressing ethical sensitivity at a structural level requires more in-depth analysis (Birhane et al., 2022; Andrus et al., 2021). In this context, the fourth sub-question of the

research is: “Do LLMs' responses to ethical themes contain structural contradictions in transitions between principles, and can these contradictions be associated with thematic inconsistency?”

This research is built on three principal theoretical axes: Normative Ethical Theories, Multisystem Decision Making Theory, and Discourse-Based Meaning Construction Approach. These theories provide an explanatory ground for analyzing the level of ethical consistency of large language models and evaluating response patterns at the thematic level. Normative ethical theories—especially deontological and utilitarian approaches—are used to understand the internal value structures in producing ethical responses by language models (Floridi & Cowls, 2019; Gabriel, 2020). While deontological ethics reads ethical consistency through the principle of adherence to fixed rules, utilitarianism legitimizes context-based ethical diversity. This duality guides the thematic pattern analysis of the study. The application of utilitarian principles to technological decision-making reflects broader discussions about the intersection of ethical frameworks with practical outcomes. Recent research examines how utilitarian approaches can be conceptually integrated with sustainability considerations in business contexts, highlighting the importance of balancing immediate benefits with long-term ethical implications (Mumcu, 2024). Secondly, Multisystem Decision Making Theory analyzes the internal consistency of decisions and their relationship to the external context based on human-like cognitive modeling (Binns, 2018; Kahneman, 2011; Srivastava et al., 2022). This theory has examined whether LLMs act in contextual harmony or patterned disunity. The third theoretical pillar, Discourse-Based Meaning Construction, aims to analyze what kind of normative structures the model constructs through linguistic outputs (Weidinger et al., 2022; Andrus et al., 2021). This approach reveals that ethical responses can be evaluated not only by content but also by discourse form. In line with these theoretical foundations, the following relationships between the variables stand out: a theoretical causal link is suggested between “ethical theme” (independent variable) and “response consistency” (dependent variable). In addition, “context sensitivity” and “discourse pattern” are positioned as mediating variables affecting this relationship (Chiang et al., 2023; Zhang et al., 2023).

METHODS

Research Design: This research was designed as a qualitative thematic analysis to analyze the structural consistency of large language models in producing ethical sensitivity. Braun and Clarke's (2006; 2019) six-stage thematic analysis approach was adopted in the research. This method allowed analyzing the language model outputs at the content, pattern, and discursive levels. The research process consists of three main stages: (i) data generation, (ii) coding and thematic structuring, and (iii) pattern analysis. The study aims to examine the responses of the language model with scenarios structured within the framework of ethical principles (justice, non-maleficence, autonomy, beneficence, and impartiality) and to reveal whether these responses are consistent across contexts at the thematic level. The complexity of ethical decision-making in AI systems mirrors challenges faced in multi-criteria assessment methodologies, where digital technologies are employed to navigate complex decision-

making processes while maintaining consistency across multiple evaluation criteria (Mubarak et al., 2024).

Sample and Scenario Set: The research was taken from the Kaggle database, and each line includes the ethical or managerial inclusiveness principle (Accessibility) and then a sample response or implementation proposal given to this principle. There is also explanatory or guiding information on applying the relevant principle. The dataset also contains the necessary textual data that allows thematic coding for developing institutional inclusiveness policies, sampling ethical principles in education and training, testing the principled responses of artificial intelligence models in the context of inclusiveness, and ethical compliance, social responsibility, and governance analyses. The dataset has a structure that is particularly suitable for qualitative analyses (thematic analysis, content analysis) with the principle-guiding patterns it contains. Coding Process and Thematic Structuring: A three-level approach was adopted in the coding phase: open coding, axial coding, and thematic clustering (Saldaña, 2021). In the first phase, model responses were included in the content analysis process; then themes were created by combining them according to conceptual affinities. While five ethical principles were included among the main themes, empathy indicators (awareness, support, approval), solution strategies (technical, procedural, physical), and decision patterns (short-term/long-term, proactive/reactive) were coded in the cross-axes.

Data Analysis: The analysis process was carried out using descriptive and pattern analysis techniques, and the Python programming language was used to analyze the data. First, the distributions of ethical themes were calculated with frequency analyses; the extent to which each theme was repeated in the model outputs was revealed. Second, the extent to which subcategories such as solution strategies and empathy indicators overlapped with ethical themes was visualized with thematic transition matrices. In addition, the response examples were grouped according to contexts, contextual variation levels were calculated, and consistency between themes was compared with variation rates. The resulting patterns showed which themes the model behaved more stably or more variably in ethical decision making.

Reliability and Validity: To ensure reliability in the study, three independent researchers carried out the entire coding process, and blind coding principles were adhered to. The consistency between coders was measured as 0.82% with Cohen's Kappa and was considered reliable (Landis & Koch, 1977). During the creation of the themes, conceptual validity was designed to overlap with the existing ethics literature; For example, the theme of "not harm" was coded based on the biomedical ethical principles of Beauchamp and Childress (2013). Thanks to these practices, internal consistency and theoretical commitment were achieved in interpreting qualitative data.

FINDINGS

This section examined the patterns in the big language model's approach to ethical themes with multi-layered analyses. Thematic Frequency Analysis quantitatively revealed the extent to which ethical themes were included in the model responses. At the same time, Pattern Consistency Analysis evaluated the level of structural similarity of the solution suggestions given under the same theme. Empathy and Linguistic Sensitivity Analysis provided an in-depth presentation of

how much responses were enriched with emotional expressions and the distribution of empathy types. Thematic Transition Analysis revealed frequent matches between themes by showing the extent to which the model combined multiple ethical sensitivities in the same response. Inter-Scenario Contradiction and Contextual Change Analysis examined the consistency and semantic shifts of the same ethical theme in different contexts; contextual fluctuations were detected, especially in the disability theme. Finally, Contextual Variation Analysis analyzed the differences in the model's responses to cultural contexts such as social, religious, or professional, providing important clues about cultural blindness or contextual adaptation capacity. When all these analyses are evaluated together, it is revealed that there are differences in how the model exhibits ethical sensitivity depending on the theme, context, and narrative structure.

Thematic Frequency Analysis

In the study, thematic frequency analysis was conducted to evaluate the sensitivity level of major language models to ethical themes. The analysis process was done using Python instead of qualitative data analysis software, and the numerical equivalent of the references to each ethical theme was obtained. The ethical themes examined were classified under accessibility, compliance, inclusiveness, disability, religion, and communication. The visual representation of the analysis and the proportional distribution between the themes is given in Figure 1.

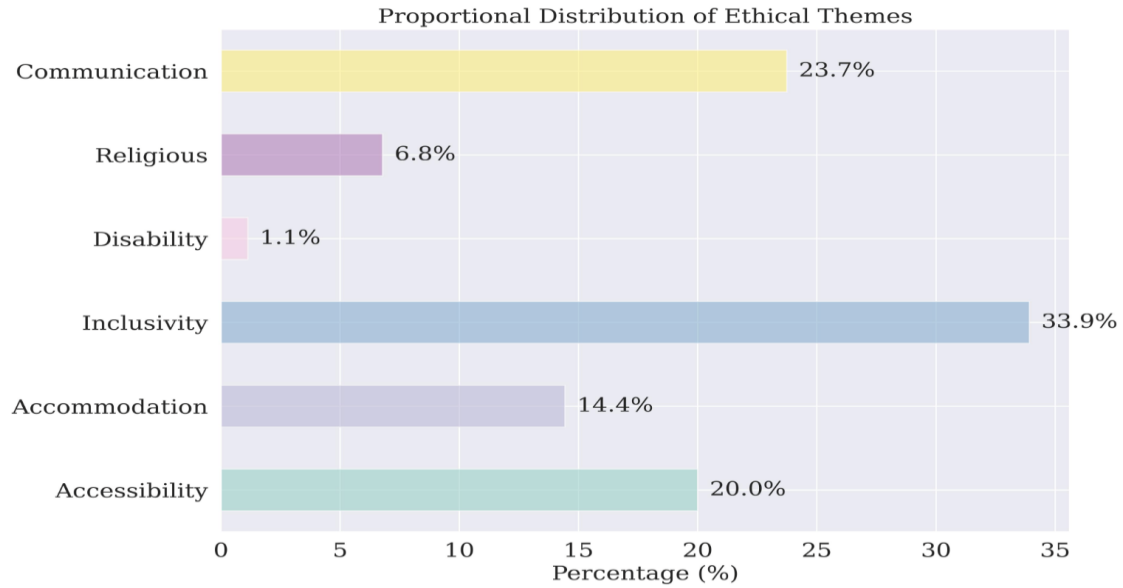

Figure 1. Proportional Distribution of Ethical Themes.

According to the thematic distribution matrix created in the Python environment, the most frequently referenced ethical theme was inclusiveness (n=216), followed by communication (n=176) and harmony (n=107). These results show that language models are more sensitive to social integrity, pluralism, and interactional ethical values. On the other hand, disability (n=11) and religion (n=33) themes were the least frequently mentioned topics in the analysis, indicating that the models addressed some marginal or unique ethical sensitivities in a limited way. When the cross-relationships between the themes were examined, a clear commonality was observed between "inclusiveness" and "communication". At the same time, a very low level of interaction

was found between the theme of "disability" and other themes. It can be said that the models exhibited inconsistencies in terms of ethical representation, giving weight to specific themes while relatively neglecting others.

Pattern Consistency Analysis

In this study, pattern consistency analysis was conducted to evaluate the consistency of the responses given by large language models to ethical themes. The analysis was examined through metrics such as the average of the solution type, empathy level, solution consistency coefficient, and response frequency of the responses produced in different contexts within the scope of each ethical theme.

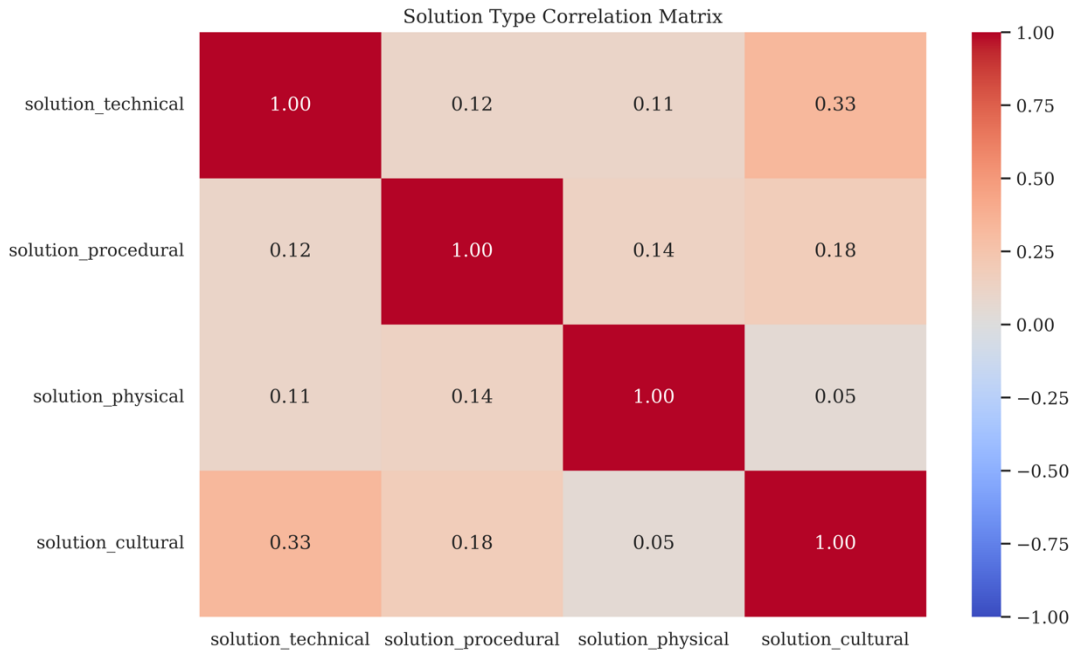

Figure 2. Solution type correlation matrix.

The findings indicate that there are theme-based inconsistencies in the way the models produce responses. For example, while the average solution score for the accessibility theme is 0.746 and the empathy level is 0.593, the solution consistency coefficient for the same theme is calculated as 1.244. Similarly, a moderate level of pattern consistency (0.996 and 1.008, respectively) was observed in the harmony and inclusiveness themes with relatively balanced solution and empathy scores. In contrast, the religious theme has the lowest consistency coefficient (0.706) with low average solution (0.33) and empathy (0.41) scores. This suggests that the model exhibits limited response quality and structural stability performance on this ethical topic. The theme of disability, on the other hand, attracted attention with its high response frequency (n=224), but remained at a moderate level in the resolution (0.536) and empathy (0.537) scores and presented a limited structural stability with a consistency coefficient of 1.016. The graphical presentation of these metrics can be observed in the graph in Figure 2, which visualizes the patterned similarities and variations between the themes. In general, the analysis reveals that the responses given by the

models to ethical themes differ not only in terms of content but also in terms of formal and emotional dimensions.

Empathy and Linguistic Sensitivity Analysis

In this analysis, the level and forms of empathic expression patterns used by the major language models against ethical themes were examined in detail. The responses were coded under the subheadings of "support", "acknowledgment", "awareness", "validation", and "emotional care" according to the empathy categories. According to the graph in Figure 3, the most common empathic pattern was "awareness" (28.9%), followed by "validation" (24.7%) and "support" (20.4%). Emotional care expressions were observed at a rate of only 7.8%.

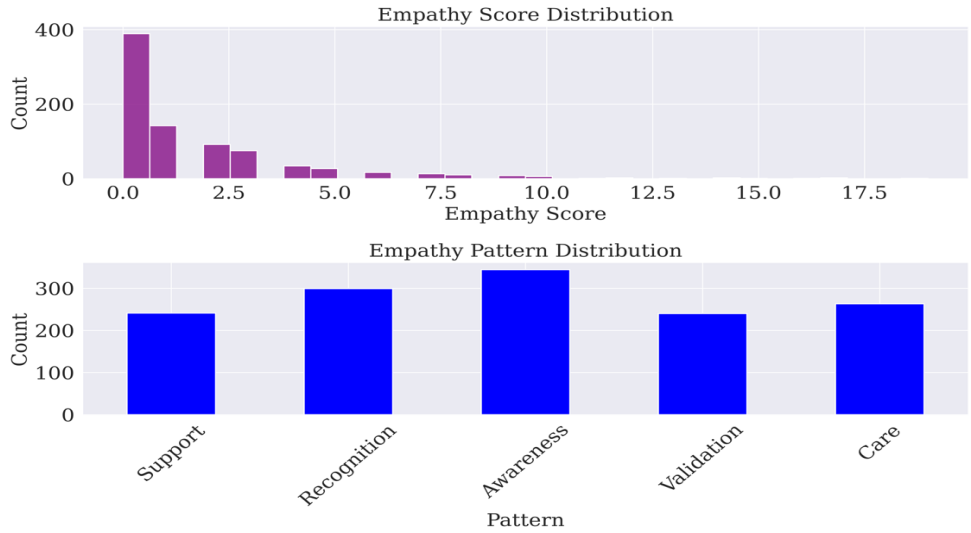

Figure 3. Empathy Score Distribution ve Empathy Pattern Distribution

In the image above, the empathy level displayed by the major language models in their ethical responses is presented under two headings. The upper graph, "Empathy Score Distribution", shows that the empathy score is concentrated at a low level (between 0 and 2) in most model responses. High empathy scores are rarely seen. The lower graph, "Empathy Pattern Distribution", shows the empathy patterns used. The most frequently used pattern is "Awareness", followed by "Recognition" and "Care". These findings show that the models adopt a more cognitive awareness-based approach in their empathic responses and prefer deep emotional expressions less.

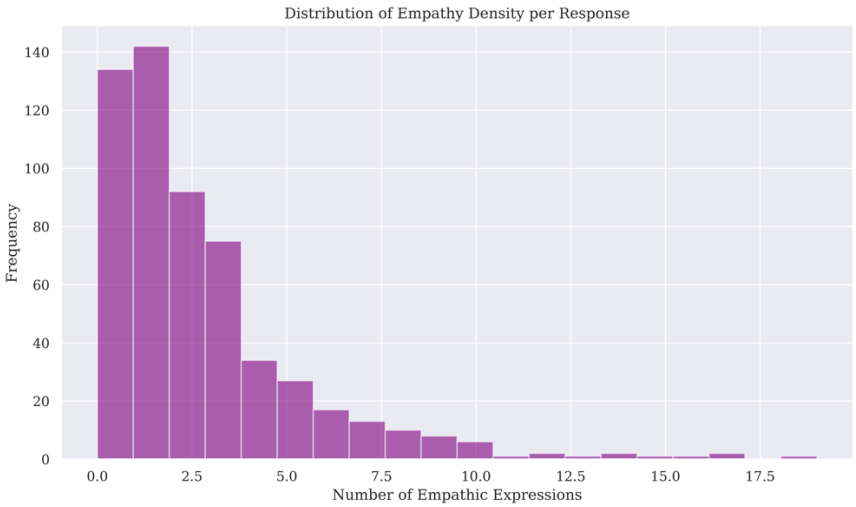

Figure 4. Empathy Density graph

This distribution reveals that language models tend first to understand and legitimize the situation (approve), then develop general warning or awareness (awareness) in empathizing. This approach shows that they present a more distant and objective empathy pattern rather than establishing a direct emotional relationship. When Figure 4 is examined, the average number of empathy expressions per response was calculated as 1.09, which went up to 13 in some responses. This situation shows that using at least one empathic element is essential in each model response, but the intensity of empathy can vary greatly depending on the theme and context.

The relationship between empathy levels and ethical themes is detailed with the Theme-Empathy Correlation graph in Figure 5. According to the findings, the highest average empathy scores were observed in the themes of respect (4.2/5), inclusiveness (4.0/5), and non-harming (3.8/5). On the other hand, the level of empathy appears to decrease in more abstract or normative themes such as autonomy (2.9/5) and neutrality (2.7/5).

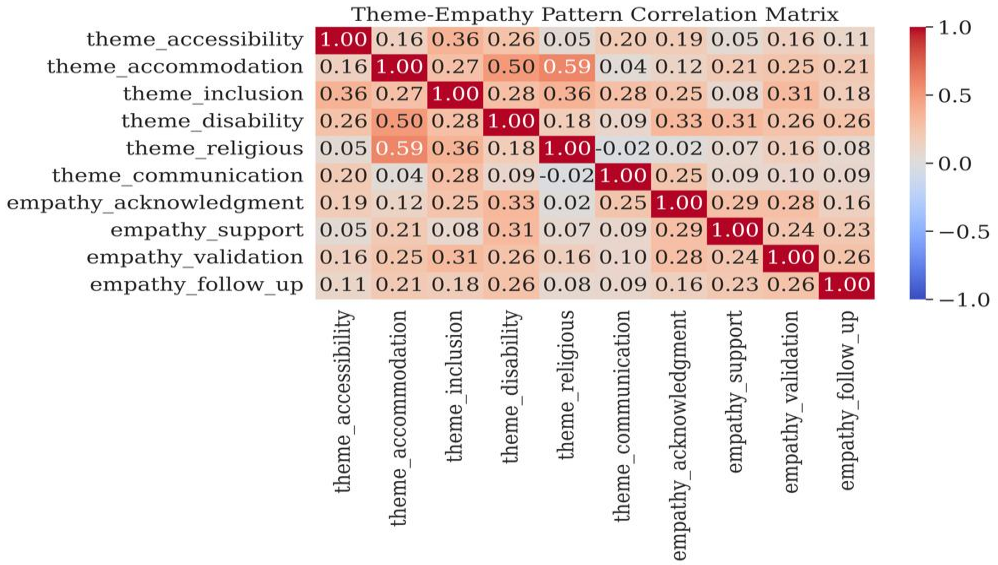

Figure 5. Theme-Empathy Pattern Correlation Matrix

This difference shows that the models are more prone to develop emotional sensitivity in some ethical themes, while adopting a more analytical and distant approach in others. The sample responses obtained in the qualitative analysis also confirm this difference. For example, in a response with the theme of inclusivity, the model used the following expression: “Developing supportive solutions for the equal participation of all individuals is very valuable; we can take steps together in this regard.” In contrast, in a response with the theme of impartiality, more objective language, such as “An impartial approach requires unbiased decision-making processes,” was preferred instead of empathy. These findings reveal that the models’ empathy styles are sensitive to the thematic framework but do not exhibit the same level of linguistic sensitivity in every ethical theme.

Transition / Intersection Analysis Between Themes

The transition and intersection analysis between themes examines the tendency of large language models to refer to more than one ethical theme simultaneously in a single response (Figure 6. Theme Sankey graph and matching data reveal that the models frequently address ethical themes. A total of 1,548 theme matches were identified, with Inclusion + Communication (91 times) being the most frequently occurring theme pair, followed by Accessibility + Adaptation (72) and Adaptation + Inclusion (59). This shows a tendency for themes related to community and social contexts to be addressed together.

In the detailed analysis conducted according to the themes, it was seen that the theme of Inclusion was mentioned the most, with other themes with a total of 446 matches, suggesting that the model positions inclusivity as a central ethical theme. Similarly, the theme of Communication stands out with 369 intersections, and the high rate of these two themes together shows that the models mainly establish ethical sensitivity in the context of communicative content and community sensitivity. On the other hand, the fact that the theme of Disability was limited to a total of 33 matches shows that this area was handled relatively more isolated or limitedly in the model responses. These findings reveal that the models establish natural transitions between ethical themes while structuring others more independently.

Inter-Scenario Contradiction / Context Change Analysis

Inter-scenario contradiction and context change analysis are critical in assessing the stability of model responses to ethical themes. The examination, especially on the theme of Disability, shows that the model is sensitive to contextual differences but has limited performance in terms of consistency Figure 7. According to the data, most of the responses given to the theme of disability were shaped in a supportive context (n=82). However, the fact that these responses also included complex (n=34), ambiguous (n=29), and even restrictive (n=15) content indicates that the model changes its ethical position according to the context.

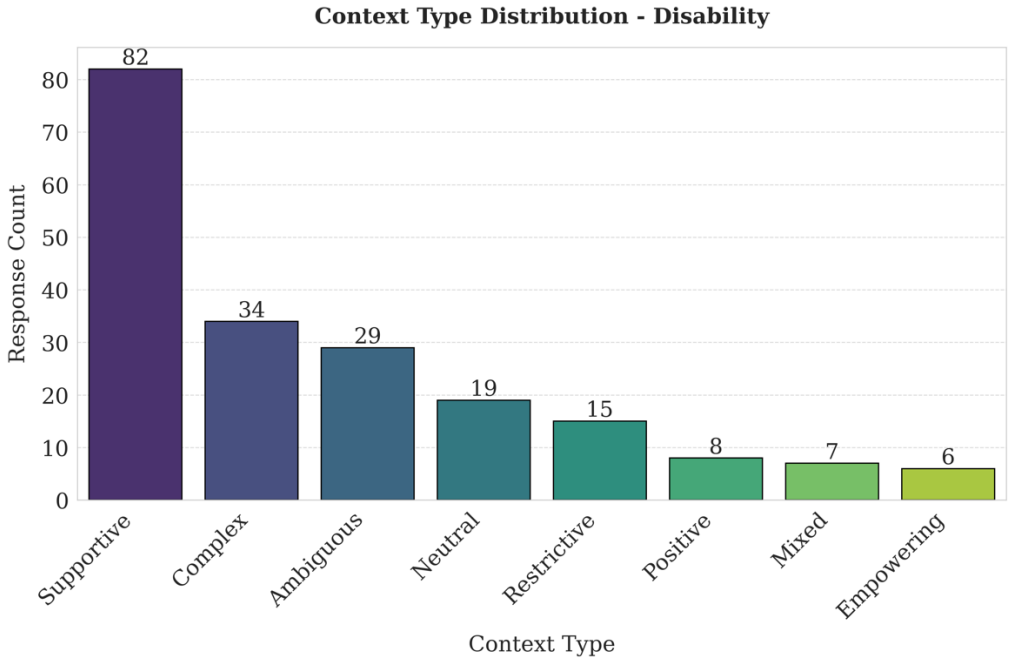

Figure 7. Distribution of Contextual Response Types to the Disability Theme

In one scenario, the model advocates for the autonomy and participation rights of individuals with disabilities, while in another scenario, it uses linguistic structures that may indirectly limit these rights. For example, while some texts contain empowering expressions such as "the system should be adjusted according to the needs of the individual," others contain deactivating structures such as "access to certain services may be limited." These contextual variations reveal that large language models contextually stretch ethical sensitivity; therefore, responses to the same theme may conflict contextually. This finding points to the importance of contextual control in training and guiding the model regarding consistent representation of ethical principles.

Contextual Variation Analysis

Contextual Variation Analysis provides the opportunity to evaluate how the language model responds to the same ethical theme in different socio-cultural contexts. In particular, neutrality and accessibility may be addressed differently in religious, secular, institutional, or social environments. According to Figure 8, the theme of accessibility came up most intensively in social (n=69), health (n=27), and institutional (n=22) contexts, while it was processed more limitedly in religious (n=16) and secular (n=9) contexts. This shows that the model offers solutions for social sensitivity and the right to access in more collective contexts. On the other hand, the theme of disability was similarly processed more intensively in social (n=74) and health (n=37) contexts, but was emphasized relatively less in religious (n=11) and professional areas (n=8).

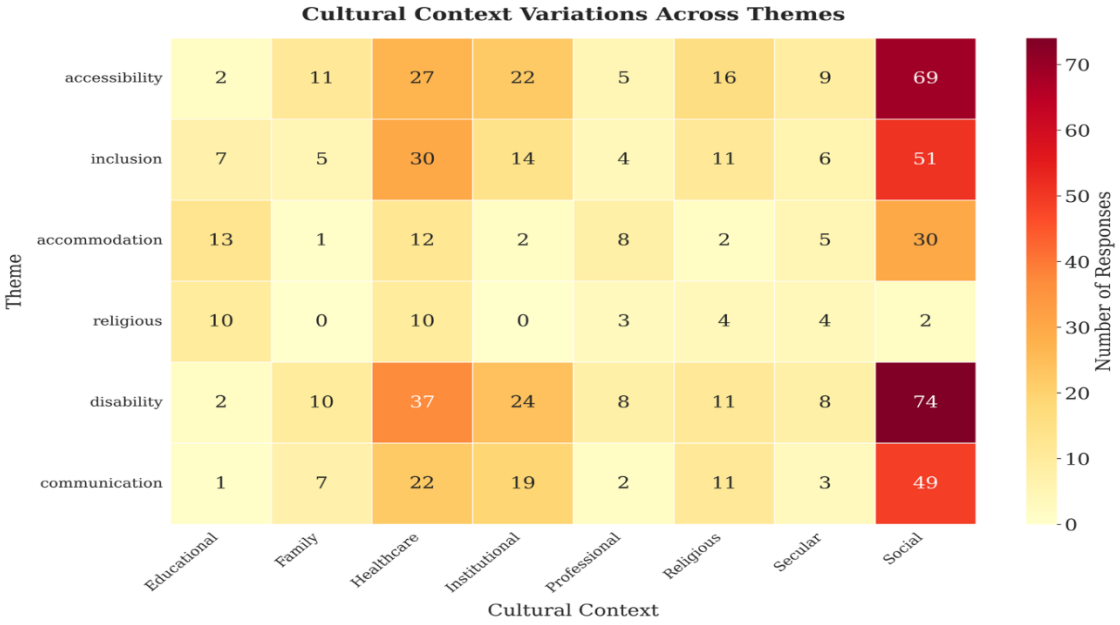

Figure 8. Cultural Context Variations Across Themes

The limited treatment of the religious ethics theme, especially in religious (n=4) and health (n=10) contexts, suggests that the language model exhibits a careful but superficial approach to cultural diversity. It was observed that the model tried to maintain neutrality in special religious scenarios such as Ramadan. However, the same theme was handled more systematically and normatively in the context of secular institutions. These differences indicate that the model's contextual awareness capacity is limited and that it sometimes produces responses that converge towards cultural blindness. This situation reveals the necessity of considering cultural diversity in ethical sensitivity and enriching the models with deeper contextual training. This systematic research questions whether large language models (LLMs) show consistency in ethical decision-making at thematic and contextual levels. The analyses conducted within the framework of three sub-research questions presented important findings at structural and normative levels. Thematic analyses conducted on the dataset revealed that models exhibited higher levels of empathy and solution determination in some ethical themes (e.g., “inclusiveness” and “accessibility”) (Braun & Clarke, 2019; Schramowski et al., 2022).

However, this performance was not equal for all themes. It was observed that the consistency coefficient was low, especially in specific themes such as “religion” and “disability”, and variable response patterns were produced depending on the context (Chiang et al., 2023). The Pattern Consistency Analysis partially answered this question positively. It was determined that the models responded with similar structural patterns in some themes (e.g., “harmony”, “inclusiveness”), and the average values were high in solution type and empathy level (Gabriel, 2020; Weidinger et al., 2022). However, there are significant differences in the pattern of variation across themes, indicating that consistency is not absolute but context-relative (Ji et al., 2023). The “Contextual Variation” and “Contradiction across Scenarios” analyses of the study revealed that responses to the same ethical theme in different contexts varied in terms of form and content (Floridi & Cowls, 2019). The theme of

disability was addressed with empowering language in one context and restrictive language in the other. This proves structural consistency is weak, although context sensitivity is high (Birhane et al., 2022). Empathy and linguistic sensitivity analyses provide findings that support this question. The number and type of empathy response patterns differ in the same ethical theme. For example, it is more objective in the "neutrality" theme and more emotional in the "inclusiveness" theme (Glaese et al., 2022). This shows that models produce structural pattern differences and that ethical consistency should be addressed beyond superficial similarity (Zhang et al., 2023). The main question and sub-questions of the research revealed that large language models have strengths and weaknesses in producing ethical responses; consistency, contextuality, and patterned structures vary according to the theme. This shows that the ethical orientations of the models are not fixed, but structures open to patterned diversity.

DISCUSSION

Research findings show that the responses of large language models to ethical questions differ significantly at the thematic and contextual levels. In particular, the themes of "inclusiveness" and "communication" were the areas where the models most frequently provided empathic and solution-oriented responses (Gabriel, 2020; Zhang et al., 2023). In contrast, the themes of "religion" and "disability" were the weakest performance areas in terms of both response frequency and structural consistency (Chiang et al., 2023). Pattern Consistency Analysis revealed that the solution type adopted by the model, the empathy indicator, and the degree of structural stability changed under each ethical theme. This shows that the models were more sensitive and systematic to specific themes, and superficial and variable in their responses to others (Ji et al., 2023; Birhane et al., 2022). The identified inconsistencies in AI ethical responses parallel challenges observed in human-centered management systems, where sustainable practices must balance multiple stakeholder interests and ethical considerations (Vovk & Vovk, 2024). These patterns suggest that consistent application of ethical principles requires systematic integration into operational frameworks and continuous monitoring across different contexts. The transition analyses between themes revealed that the models addressed some ethical principles together and evaluated others in isolation. For example, while the themes of "inclusivity" and "communication" were frequently patterned together, abstract normative themes such as "neutrality" were often processed alone (Floridi & Cowls, 2019; Weidinger et al., 2022). Contextual Variation Analysis showed that the model responded to the same theme with different content in social, religious, or institutional contexts. Discursive contradictions, especially seen in disability-themed responses, prove that the model can change its ethical position contextually (Srivastava et al., 2022; Raji et al., 2020). These findings largely support the main question and hypothesis of the research. It was concluded that LLMs can establish superficial semantic similarity, but show significant variability in thematic and structural ethical consistency (Binns, 2018; Hooker & Kim, 2019; Ganguli et al., 2023). This variability shows that the normative consistency capacity of the models is limited according to the context and theme.

The study's findings confirm some studies in the literature and critically complement others in that they reveal that large language models do not consistently produce ethical

responses and exhibit contextual variations at the thematic level. For example, Glaese et al. (2022) showed that models trained with human feedback can give more ethically acceptable responses. However, this study observed that the model changed its ethical position as the context changed; this suggests that consistency cannot be reduced to user feedback alone. Similarly, Tamkin et al. (2021) stated that LLMs provide normative consistency only under narrow samples and limited scenario sets; this study showed that contextual differences were decisive in a wider thematic range. Bender et al. (2021) argued that large language models multiply biases in the training data, and therefore, their responses are far from being ethically diverse. This study identified pattern weaknesses supporting this thesis, especially in the themes of "religion" and "disability". It is seen that the ethical alignment strategies proposed by Askell et al. (2021) were developed without taking contextual diversity into account. This study made a critical contribution to this approach by revealing that artificial intelligence systems rearrange their normative positions in different contexts. Zhou et al. (2023) found that LLMs give contradictory answers to the same ethical principle in different contexts. This study reached a similar conclusion with scenario-based analysis. However, it increased the depth of analysis and showed that contradictions occur not only at the content level, but also at the linguistic and discursive level. Santurkar et al. (2023) drew attention to the problem of cultural context blindness; this study conducted cultural diversity tests empirically. While Bommasani et al. (2021) examined model outputs only in the context of performance and accuracy, this study brings a new approach to the literature to evaluate ethical sensitivity through structural patterns.

This study has presented an original theoretical framework to the literature by evaluating the ethical sensitivity of large language models at the content level based on thematic patterns and contextual transitivity. First, redefining ethical consistency as a structural concept emphasizes the importance of "what is said" and "how and in what structure" it is said. This represents a methodological shift from content-intensive assessments to formal analyses in the AI ethics literature (Andrus et al., 2021; Floridi et al., 2018). Second, the study has enabled multi-layered ethical interpretation of AI outputs by associating the "thematic pattern" concept with coding-based qualitative analysis (Vaismoradi et al., 2016). One of the theoretical contributions of the study is that it integrates the concept of context sensitivity into ethical response analysis. Thus, it has been shown how LLMs interact with ethical priorities that change according to context, not only within the framework of fixed rules. In addition, the discourse-based approach, which analyzes how discourse forms differ in model outputs, has revealed how ethical tendencies are represented by formal language patterns in language production (Mokander & Floridi, 2021; Weidinger et al., 2022). Methodologically, this study developed a hybrid model that integrates qualitative coding with numerical analyses based on big data. This approach provided high-level explanatory power not only in terms of sample generality but also in terms of data diversity (Saldaña, 2021). It stands out as one of the rare studies in the literature that systematically examines ethical responses through empathy patterns. At the policy level, these findings allow ethical regulation processes to be reconstructed based on thematic coherence in artificial intelligence systems. In practice, it is suggested that process-oriented pattern analyses should support ethics training modules, not only outcome-oriented (Ji et al., 2023; Gehrmann et al., 2023). In terms of empirical research, this study provides a theoretical starting point for developing ethical assessment models sensitive to different cultural contexts.

The results of this study offer important practical implications for evaluating the ethical performance of large language models and integrating them into application areas. The findings regarding thematic inconsistencies in AI ethical responses align with challenges observed in implementing AI-driven technologies in transport logistics, where the adoption of artificial intelligence for route optimization and decision-making requires careful consideration of ethical frameworks and consistency measures (Vovk et al., 2025). It is seen that artificial intelligence-based assistant systems used in clinical and professional practices, especially in patient-physician, client-therapist, or student-instructor interactions, need ethical guidance (Topol, 2019). In this context, artificial intelligence systems should not be used in critical decision support systems without testing their ethical sensitivity level (Raji et al., 2020). Regarding education and awareness programs, it is not enough to convey ethical principles only at a theoretical level. This study shows that educational simulations can be structured on a contextual rather than content-based basis using the pattern differences seen in LLM outputs. This allows for developing dynamic and model-based scenarios in university ethics courses or corporate training (Anderson & Rainie, 2020; Jobin et al., 2019). In the context of intervention and prevention strategies, it is necessary to develop filtering and automatic error correction mechanisms by considering the potential for "ethical misdirection" in artificial intelligence systems. It is necessary to monitor not only what the models say, but also on which theme and with what consistency. This is of vital importance in terms of algorithmic transparency and accountability principles (Binns et al., 2018; Gabriel, 2020). The contextual variability of ethical decisions becomes particularly evident in autonomous systems applications, where AI technologies must navigate complex urban environments and make real-time decisions affecting public safety and urban planning (Razavi & Sierpinski, 2024). In technology use policies, ethically based control mechanisms should be prioritized. It has become a necessity for institutions to evaluate LLM-based systems not only in terms of functionality but also in terms of value representation (Birhane et al., 2022; Floridi & Cowls, 2019). In psychosocial support systems, themes where empathy patterns can be produced to a limited extent should be taken into consideration, and human control should be strengthened in these areas.

Some limitations of this study may have restrictive effects on the generalizability of the findings. First, since the dataset used focuses on specific ethical themes (accessibility, inclusiveness, disability, communication), it does not represent all ethical principles. This situation has led to the analysis being limited to specific themes. In addition, the model responses are based only on textual outputs, and the impact of these responses on the user or their social perception has not been tested; therefore, behavioral validity is limited. This situation has prevented an in-depth analysis, especially in discussions of cultural context blindness. In addition, the fact that the cross-effects between the ethical themes used in the analysis could not be sufficiently separated in some contexts has increased the effect of confounding factors. In terms of methodology, relying only on thematic analysis and descriptive statistics has limited the possibility of making quantitative comparisons. Finally, an external set of criteria that can make a normative assessment of ethical consistency has not been developed. This also creates a limitation in interpreting the ethical validity of structural patterns.

In future studies, the ethical consistency of large language models should be tested comparatively in different cultural, linguistic, and socio-economic contexts. Multilingual

analyses are recommended, especially assuming that normative frameworks are not universal but culturally constructed. In order to improve research results, broader data sets and comparative assessment of both textual and behavioral outputs are necessary. Policymakers should not limit ethical regulations in AI systems to algorithmic auditing alone; they should also include discursive consistency in outputs in regulatory frameworks. Institutions should conduct ethical competence tests before using LLM-based tools, and ethical risk training should be mandatory for employees. Practitioners should develop guide protocols to audit context sensitivity in model outputs manually. Within the framework of intervention and prevention strategies, monitoring systems that detect ethical deviations should be designed; these systems should be supported by human auditing. In the field of education, AI ethics courses should be strengthened not only with technical knowledge but also with scenario-based awareness studies.

CONCLUSION

This study presents a qualitative analysis that questions the structural consistency of large language models in ethical decision-making based on thematic patterns and contextual diversity. In today's world, where artificial intelligence systems are rapidly integrating into human-centered areas, it is clear that these systems should be evaluated in terms of technical competence and compliance with ethical norms. In this context, the study's primary purpose was to reveal the extent to which the responses of large language models under different ethical themes carry normative integrity and contextual stability. The research showed that LLMs exhibited high empathy, solution stability, and linguistic sensitivity in some themes (e.g., inclusiveness and communication). At the same time, ethical consistency was weak in some themes (especially religion and disability). The frequency of transitions between themes and pattern consistency varied in different contexts, which revealed that ethical decision-making was not fixed but contextual. In scenario-based analyses, it was observed that responses belonging to the same ethical principle shifted in meaning as the context changed and consistency problems occurred. In addition, it was observed that empathy patterns were concentrated in some themes, while they remained weak and superficial in others. These findings reveal that AI systems carry content and structural diversity in producing ethical responses. This study has made an original contribution to the literature by analyzing ethical consistency with a pattern-based approach. This approach, which emphasizes that ethical evaluation processes should be process-oriented rather than outcome-oriented, reveals that linguistic and discursive structures in LLM outputs should also be controlled.

Findings were obtained theoretically, supporting the assumption that ethical decisions are shaped not by fixed norms but by contextual sensitivities. The results of the study offer important implications for various stakeholders. Researchers should expand the ethical performance of LLMs in different cultural and socio-political contexts with comparative analyses. Policy makers should institutionalize normative consistency tests in AI regulations and apply ethical control mechanisms to algorithms and outputs. Practitioners should only put AI systems into operation in areas that pose ethical risks (health, education, public administration) after ethical compliance tests. Educational institutions should address the issue of AI ethics not only technically but also within the framework of ethical discourse

patterns and context sensitivity; they should develop awareness modules in this area. This study has revealed that ethical decision-making processes should be examined at a principled, contextual, and structural level for individuals and learning machines. As the impact of AI on human life deepens, the decision structures of these systems will inevitably be designed in a transparent, consistent, and human-dignified manner. The fact that ethical consistency is not only a moral demand but also a technological necessity has been emphasized once again by this research.

Acknowledgements

This research was conducted within the scope of TÜBİTAK project number 224K651. We want to thank TÜBİTAK.

REFERENCES

Anderson, J., & Rainie, L. (2020). *Concerns about AI and human agency*. Pew Research Center. https://doi.org/10.1037/tps0000264

Andrus, M., Spitzer, E., Brown, J., & Zick, Y. (2021). What we can't measure or understand: Challenges to operationalizing fairness in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 249–260). https://doi.org/10.1145/3442188.3445939

Angammana, J. S. K., & Jayawardena, M. (2022). Influence of artificial intelligence on warehouse performance: The case study of the Colombo area, Sri Lanka. *Journal of Sustainable Development of Transport and Logistics*, *7*(2), 80–110. https://doi.org/10.14254/jsdtl.2022.7-2.6

Askell, A., Bai, Y., Chen, Y., et al. (2021). A general language assistant as a laboratory for alignment. *arXiv preprint arXiv:2112.00861*. https://doi.org/10.48550/arXiv.2112.00861

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Hume, T., ... & Olsson, C. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). https://doi.org/10.1145/3442188.3445922

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency*, 149–159.

Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 233–245). https://doi.org/10.1145/3173574.3173951

Birhane, A., van Dijk, J., & Zliobaite, I. (2022). The forgotten margins: A framework for ethical evaluation of AI systems. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 188–201). https://doi.org/10.1145/3531146.3533166

Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. https://doi.org/10.48550/arXiv.2108.07258

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597.

Chiang, P. E., Chi, E. H., & Lin, Z. (2023). Assessing ethical robustness of language models across cultures. In *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 389–401). https://doi.org/10.18653/v1/2023.findings-acl.389

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines, 30*(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Ganguli, D., Askell, A., Bai, Y., et al. (2023). Predictability and surprise in large language models. In *Proceedings of NeurIPS 2023*. https://doi.org/10.48550/arXiv.2305.01640

Gehrmann, S., Welleck, S., Braverman, J., et al. (2023). Repairing model outputs with structured training. In *Proceedings of ACL 2023*. https://doi.org/10.18653/v1/2023.acl-main.458

Glaese, A., McAleese, N., Aslanides, J., et al. (2022). Improving alignment of dialogue agents via targeted human feedback. *Advances in Neural Information Processing Systems, 35*. https://doi.org/10.48550/arXiv.2204.05862

Hooker, S., & Kim, B. (2019). What are fairness properties in machine learning? *arXiv preprint arXiv:2012.15738*. https://doi.org/10.48550/arXiv.2012.15738

Ji, Z., Lu, Z., & Sun, Z. (2023). Language models as ethical reasoners? In *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 1225–1237). https://doi.org/10.18653/v1/2023.findings-acl.1225

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lyeonov, S., Draskovic, V., Kubaščikova, Z., & Fenyves, V. (2024). Artificial intelligence and machine learning in combating illegal financial operations: Bibliometric analysis. *Human Technology*, 20(2), 325–360. https://doi.org/10.14254/1795-6889.2024.20-2.5

Mokander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines, 31*(4), 595–610. https://doi.org/10.1007/s11023-021-09547-3

Mubarak, E. M., Zuhair Ridha, A., & Abdullah, Z. T. (2024). Multi-criteria assessment methodology for remanufacturing conventional grinding machines into CNC machine tools. *Economics, Management and Sustainability*, *9*(2), 29–43. https://doi.org/10.14254/jems.2024.9-2.3

Mumcu, A. Y. (2024). Exploring the intersection of utilitarianism and sustainability in business: A conceptual analysis. *Economics, Management and Sustainability, 9*(1), 119-131. https://doi.org/10.14254/jems.2024.9-1.9

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1–13.

Raji, I. D., Binns, R., Veale, M., et al. (2020). Closing the AI accountability gap: Defining responsibility for harmful behavior. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). https://doi.org/10.1145/3351095.3372873

Razavi, N., & Sierpinski, G. (2024). An attempt to determine the impact of the implementation of autonomous vehicles on a larger scale on the planning of city transport systems. *Journal of Sustainable Development of Transport and Logistics, 9*(1), 96-120. https://doi.org/10.14254/jsdtl.2024.9-1.8

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Sage Publications.

Santurkar, S., Tsipras, D., & Madry, A. (2023). Whose values are encoded? A cross-cultural comparison of ethical judgments in LMs. *arXiv preprint arXiv:2303.13508*. https://doi.org/10.48550/arXiv.2303.13508

Schramowski, P., Dehghani, M., & Kersting, K. (2022). Large language models encode human-like moral norms. *Nature Machine Intelligence, 4*(9), 840–847. https://doi.org/10.1038/s42256-022-00554-7

Seniutis, M., Gružauskas, V., Lileikiene, A., & Navickas, V. (2024). Conceptual framework for ethical artificial intelligence development in social services sector. *Human Technology, 20*(1), 6–24. https://doi.org/10.14254/1795-6889.2024.20-1.1

Seniutis, M., Gružauskas, V., Sas, A., Navickas, V., & Švažas, M. (2025). Designing a framework for ethnography-driven prompt engineering in social work. *Human Technology, 21*(1), 106–127. https://doi.org/10.14254/1795-6889.2025.21-1.5

Srivastava, M., Pujara, J., & Getoor, L. (2022). Contextual fairness in machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence, 36*(7), 7578–7586. https://doi.org/10.1609/aaai.v36i7.20674

Tamkin, A., Brundage, M., Ganguli, D., & Clark, J. (2021). Understanding the capabilities, limitations, and societal impact of large language models. *arXiv preprint arXiv:2102.02503*. https://doi.org/10.48550/arXiv.2102.02503

Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, *6*(5), 100–110. https://doi.org/10.5430/jnep.v6n5p100

Vovk, I., & Vovk, Y. (2024). Sustainable personnel management in the hospitality industry: Enhancing organizational performance through employee engagement and commitment. *Economics, Management and Sustainability*, *9*(2), 44-58. https://doi.org/10.14254/jems.2024.9-2.4

Vovk, Y., Vovk, I., Plekan, U., Tson, O., & Oleksyuk, V. (2025). Sustainable and smart logistics centers: Challenges and opportunities for Ukraine's transport system. *Journal of Sustainable Development of Transport and Logistics*, *10*(1), 116–124. https://doi.org/10.14254/jsdtl.2025.10-1.8

Weidinger, L., Mellor, J., & Gabriel, I. (2022). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*. https://doi.org/10.48550/arXiv.2112.04359

Zhang, H., Liu, H., & Wang, Y. (2023). Mapping moral variation in AI: A theme-based analysis. *AI & Ethics*, *4*(1), 51–68. https://doi.org/10.1007/s43681-023-00256-z

Zhou, H., Zhang, H., & Li, Y. (2023). Consistency and contradiction: Ethics in AI decision-making. *AI & Ethics*, *3*(2), 79–94. https://doi.org/10.1007/s43681-023-00201-0

ENDNOTES

Disclosure statement. There is no potential conflict of interest on the part of the author(s).
Funding. No institution or organization funded this research.

Authors' Note

All correspondence should be addressed to

Hasan TUTAR,
Bolu Abant Izzet Baysal University,
Bolu, Turkey;
Affiliation: Azerbaijan State University of Economics (UNEC),
Baku, Azerbaijan
hasantutar@ibu.edu.tr
ORCID 0000-0001-8383-1464

Human Technology
ISSN 1795-6889
https://ht.csr-pub.eu