ARTICLE | Article in Press

Ethics in natural language processing: a systematic meta-analysis

Rosa Vicari · Nadejda Komendantova

Received: 28 March 2025 / Accepted: 13 August 2026
© The Author(s) 2026

Abstract

This study investigates the ethical considerations of Natural Language Processing (NLP) technologies, which are increasingly utilized in decision-making, content moderation, and creative fields. While significant research examines ethical issues in specific NLP applications, there is a lack of comprehensive frameworks covering ethics across the full range of NLP applications. To address this gap, we conducted a meta-analysis of the academic literature on NLP ethics. Our findings reveal a rapid expansion of research in this area, driven largely by the emergence of Large Language Models (LLMs). The literature is especially concentrated in healthcare and is primarly focused on identifying ethical challenges and defining ethical standards, whereas comparatively less attention is given to the design and implementation of practical ethical tools and standards. The most frequently discussed ethical values include accuracy, privacy, transparency and traceability, and bias mitigation. In contrast, environmental sustainability and explainability remain relatively underexplored areas. The analysis also suggests that geographic concentrations of research fundings, particularly in Western countries, may influence the ethical priorities reflected in the literature. By providing a systematic cross-domain synthesis of research trends, ethical concerns, and emerging gaps, this study contributes to a more comprehensive understanding of NLP ethics and highlights the need for broader ethical frameworks and stronger efforts to translate ethical principles into practice.

Introduction

In recent years, the ethical dimensions of Natural Language Processing (NLP) have garnered increasing attention as the technology rapidly expands in scope and influence. As NLP systems become integral to various domains—ranging from legal and financial sectors to creative industries and healthcare—the potential ethical concerns surrounding their application become more pronounced.

According to Jurafsky and Martin's terms (Jurafsky and Martin, 2009), NLP, a branch of artificial intelligence, is the study of how computers can be used to process and understand human languages at various levels (syntax, semantics, pragmatics). It includes subfields such as Natural Language Understanding (NLU), which enables machines to comprehend meaning and context and respond appropriately, and Natural Language Generation (NLG), which allows them to produce text that mimics human writing. Additionally, Machine Translation automates text conversion between different languages, while LLMs are sophisticated systems that are trained on vast datasets and use deep learning techniques to perform a variety of language tasks such as translation, summarization, and dialogue generation.

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Published online: 18 September 2026

By enabling computers to analyze, interpret, and generate human language in various forms, NLP bridges the gap between human communication and machine understanding. Its applications span across industries and influence not only individual user interactions but also broader societal systems. One example is content recommendation (Renuka et al., 2021) and moderation (Llansó et al., 2020), where NLP algorithms process large amounts of text to assist in the categorization and understanding of content, which is then used to suggest relevant items to users and filter inappropriate content. Additionally, NLP is utilized in digital marketing (Hartmann and Netzer, 2023), supporting tools that perform user engagement and sentiment analysis, enabling businesses to adjust their strategies based on real-time feedback and trends. The rise of social media platforms has brought with it the challenge of misinformation, through which false or misleading information can spread quickly. In this context, NLP plays a dual role. On one hand, NLP techniques are employed in the development of hoax bots, which are automated systems that generate and distribute false information. On the other hand, through techniques such as sentiment analysis, topic modeling, and text classification, NLP can detect suspicious language use, flagging or filtering content that might be harmful or misleading (Saquete et al., 2020). The wide range of applications of NLP affects how content is produced, shared, and managed in today's digital environment.

NLP plays a critical role in automating complex decision-making tasks in artificial intelligence (AI) systems across various sectors. For code generation and debugging, tools like TabNine (Tabnine Ltd., 2024) use NLP to predict and suggest code corrections and improvements. Such tools help developers to streamline coding and debugging processes. In the legal field, NLP is used for contract analysis, with software like Kira Systems (Kira Inc., 2024) and Luminance (Luminance Technologies Ltd., 2024) leveraging machine learning to review legal documents (Zhong et al., 2020). These tools can extract clauses, identify risks, and ensure compliance with legal standards, in order to reduce the time required for contract review and drafting. Moreover, in financial decision-making and compliance checking, NLP is applied to automate the processing of financial prediction and analysis and risk assessment (Gao et al., 2021). Systems like AlphaSense (Alphasense Inc., 2024) and ComplyAdvantage (ComplyAdvantage, 2024) use NLP to monitor financial documents, analyze risk, and ensure compliance with evolving legal frameworks, enabling institutions to process large volumes of data. These applications reflect the growing reliance on NLP to handle large volumes of text and data in decision-making processes.

Estimates indicate that, by 2025, up to 50% of digital work will be automated through applications utilizing LLMs (Ball, 2024). LLMs are increasingly being deployed across different fields, contributing to sector-specific developments. LLMs, such as OpenAI's GPT models (OpenAI, 2024), are designed to achieve deeper language comprehension by understanding context, semantics, and intent beyond simple word matching. In creative writing, LLMs are being used to generate stories, articles, and even poetry, offering tools that assist writers in brainstorming and drafting content (Gómez-Rodríguez and Williams, 2023; Kreminski and Martens, 2022). In healthcare, LLMs are being applied to analyze medical records, assist in diagnostics, and even generate personalized patient communication (Yang et al., 2023). Furthermore, in education, LLMs are aiding both students and teachers by providing tutoring, grading assistance, and educational content generation, while also supporting language learning and academic research (Alhafni et al., 2024). These broad applications highlight the growing role of LLMs in transforming various sectors, where they are increasingly integrated into workflows.

In conclusion, the expanding role of NLP in various fields brings with it both opportunities and ethical challenges that require careful consideration. As NLP technologies become integral to decision-making, content moderation, and creative tasks, understanding their ethical implications is crucial for ensuring responsible deployment. Despite the rapid expansion of NLP applications, research on their ethical implications remains fragmented across domains, technologies, and perspectives. This fragmentation highlights the need for a systematic synthesis of trends, gaps, and priorities in NLP ethics research. This sets the stage for a deeper examination of the ethical frameworks that guide NLP applications. The following background section will explore the regulatory and academic perspectives on these issues, providing a foundation for our systematic analysis of ethical challenges in NLP technologies.

Background

Regulatory and ethical Frameworks for NLP

The rapid development of NLP technologies has brought with it a host of ethical challenges that regulators and scholars are increasingly addressing. Anchoring ethical considerations in NLP within the framework of International Human Rights Law (United Nations, 1948) is essential for ensuring that technological advancements align with fundamental human rights, such as freedom of speech, access to information, equality and non-discrimination. NLP systems, often employed in content moderation, information dissemination, and decision-making, must respect the delicate balance between enabling free expression and mitigating harmful content. These systems need to uphold the right to access information while safeguarding against the spread of hate speech, misinformation, and other forms of digital harm.

European regulations, such as the General Data Protection Regulation (GDPR) (European Parliament and European Council, 2016), the EU Digital Service Act (DSA), the EU AI Act (European Commission, 2024a), and the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and Rule of Law (CETS No. 225) (Council of Europe, 2024), provide a legal foundation for regulating AI, including NLP technologies. The GDPR ensures data privacy and protection; it grants individuals the ability to review their data, rectify any mistakes, and request its removal, commonly called the "right to be forgotten". The DSA is a regulation that sets out responsibilities for online platforms to increase transparency, combat illegal content, and better protect users' rights. The EU AI Act seeks to establish a risk-based regulatory framework for AI systems, aiming to ensure safety, transparency, and accountability. The Council of Europe Framework Convention covers the entire lifecycle of AI and promotes AI progress and innovation, while emphasizing the need for AI systems to operate in ways that guarantee transparency and accountability and uphold human dignity and fundamental freedoms.

Literature review and research gaps in NLP ethics

Existing scientific literature has explored ethical issues through numerous review papers covering a wide range of NLP applications. Several key themes emerge in this body of work, highlighting the ethical dilemmas that arise in the deployment of NLP technologies. Many of the existing studies focus on the ethical implications of using NLP in sensitive domains like healthcare, consumer science and education. Hutchings et al. (Hutchings et al., 2023) examined the role of social media analysis as a potential substitute for traditional research methods, emphasizing the ethical implications of relying on NLP in the field of sensory-consumer science. Similarly, (Xu et al., 2021) highlighted ethical issues in using AI-powered chatbots for oncology, stressing that chatbots should be viewed as supportive tools to enhance efficiency without compromising the patient-caregiver relationship. In mental health, (Omarov et al., 2023) noted that AI-enabled chatbots show promise in delivering mental health interventions but face ethical challenges due to the lack of established frameworks for their use. Farrelly and Baker (Farrelly and Baker, 2023) also raised ethical issues related to AI-generated texts and AI-based assessments in higher education, underscoring the importance of AI literacy and ethical considerations in the adoption of NLP technologies to guarantee fair access and beneficial outcomes for all students.

Ethical concerns surrounding generative and conversational AI are a recurring theme, largely due to the recent popularization of these technologies through ChatGPT. Fatani (2023) raised concerns about generative AI role in scientific writing, calling for clearer guidelines on AI-generated content. Kim et al. (2023) proposed guidelines for using LLMs in academic publications in the field of medicine, highlighting the need for human verification to maintain accountability. In human resources management, Budhwar et al. (Budhwar et al., 2023) stressed that the long-term impact of generative AI on workers remains uncertain, with risks including privacy concerns, bias, and context insensitivity.

An issue discussed in review papers on NLP ethics is the need for regulation, particularly in healthcare applications, where AI cannot replace human empathy and critical thinking in the physician-patient relationship. Wojcik

et al. (Wójcik et al., 2023) pointed out that legal responsibility for diagnoses is a limiting factor in the use of chatbots for medical consultations. Srivastav et al. (Srivastav et al., 2023) and Nakaura et al. (Nakaura et al., 2024) similarly addressed the risks associated with AI automation in diagnosis, such as inaccuracy, bias, and hallucinations, which can undermine the reliability of diagnoses and compromise patient safety.

While substantial work has been done on certain ethical aspects of NLP, gaps remain in comprehensive frameworks for addressing ethics across the full spectrum of NLP applications. Existing literature often emphasizes application-specific concerns (e.g., ChatGPT in education, content moderation in social media) but does not provide cross-application ethical principles. Many papers examine isolated ethical challenges like bias, accountability, or data privacy, but there is a lack of research that synthesizes all these aspects together. The ethical literature on NLP remains fragmented across domains, with few meta-analyses synthesizing existing knowledge and providing a unified view of the ethical challenges in NLP technologies. This study aims to fill these gaps by providing a comprehensive, cross-application review of the ethical challenges and solutions associated with the development and deployment of NLP technologies. A systematic meta-analytic synthesis is essential not only to consolidate dispersed insights but also to enable a comparative and structured understanding of how ethical priorities, risks, and governance approaches evolve across applications, technologies, and regions. Building on this need, the study addresses the following research questions: (1) How is the debate on NLP ethics distributed over time, across research areas, NLP applications, and regions? (2) What research scopes and ethical values dominate the literature? (3) What gaps and emerging priorities can be identified across the NLP ethics landscape?

Answering these questions provides a structured and comparative perspective on the evolution of NLP ethics research, supporting the development of more comprehensive and evidence-based ethical frameworks.

Methodology

To systematically examine the research landscape concerning ethics in NLP, this methodology section details our approach to data selection, corpus refinement, and analysis criteria. Following the background discussion of regulatory frameworks, current studies, and application-specific considerations, we compiled a dataset of academic abstracts that address ethics in NLP through a focused set of queries. Focusing on abstracts enables a consistent and scalable comparison across a large corpus, as they provide standardized summaries of each study's objectives, methods, and key themes, making them particularly suitable for structured cross-study synthesis. This section provides an outline of our approach, designed to capture a wide range of ethical themes across diverse NLP applications. The methods described here establish the foundation for our meta-analysis. The present meta-analysis is based on the systematic extraction and quantitative aggregation of predefined variables across studies, and should be distinguished from effect-size-based meta-analysis, as no statistical pooling of outcomes is performed due to the heterogeneity of the corpus.

Data for analysis

We gathered our corpus of abstracts from Web of Science, which provides access to abstracts, references, and citation data from academic journals. The abstracts included studies published up to 27 July 2024. The corpus was compiled based on key terms relevant to research on ethics related to AI technologies involving NLP.

Corpus selection based on keyterms included in abstracts (Web of Science)

1st keyterm query

Abstract = (ethic) AND ((natural language processing) OR (large language models) OR (natural language understanding) OR (natural language generation))*

2nd keyterm query

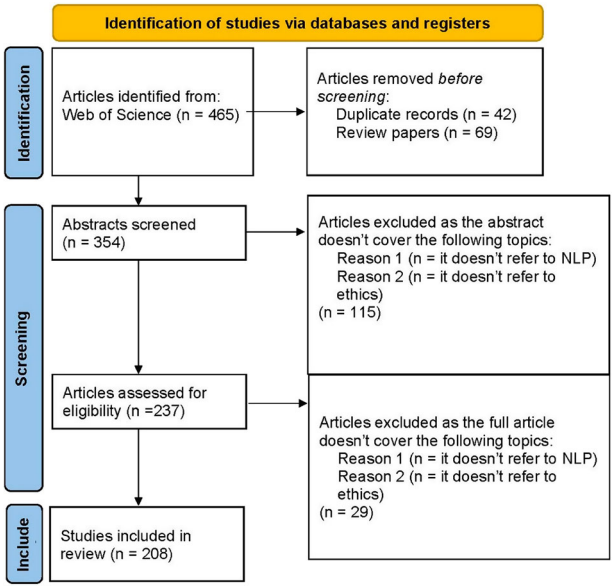

Fig. 1 PRISMA flow diagram: Our data selection process followed the PRISMA 2020 flow diagram, which details the steps for conducting a systematic meta-analysis and the associated flow of information.

Abstract =(ethic) AND (artificial intelligence) AND ((chat*) OR (social media) OR (bot*) OR (website))*

The Boolean operator "AND" indicates that the chosen abstracts must include both search queries. Meanwhile, the Boolean operator "OR" indicates that the selected abstracts should contain at least one of the key terms. Each search query begins with "Abstract = " to specifically search for key terms within the abstracts.

The selection of the keyword queries in Web of Science is designed to ensure comprehensive coverage of research focused on ethics in NLP and its related subfields. The use of explicit ethical terminology in the search strategy was intended to ensure that the selected studies directly engage with ethical considerations, thereby maintaining a focused and coherent corpus for systematic analysis. The first query specifically targets studies discussing ethical considerations in NLP, including its subfields such as natural language understanding, generation, and large language models. The second query expands the search to include broader references to artificial intelligence (AI) in contexts where NLP might not be explicitly mentioned, but is relevant, such as in chat applications, social media interactions, and bots.

We intentionally excluded terms that refer to NLP applications but are too generic, such as "text," "translation," or "record," as they could yield results unrelated to the ethical dimensions of natural language processing. Additionally, we avoided including terms that might skew the search toward specific disciplines, such as "marketing," "law", "finance," or "education." This exclusion is important because one of the objectives of our systematic analysis is to capture the distribution of published studies across various disciplines without biasing the search toward particular fields.

We employed the PRISMA 2020 flow diagram (Fig. 1) (Page et al., 2021) to document the selection process and inclusion of papers in our study. Initially, we extracted 465 abstracts from Web of Science. After removing 42 duplicate records and 69 review papers, we manually screened the remaining 354 abstracts. At this stage, we excluded 115 abstracts that did not focus on our central subjects: ethics and natural language processing. Subsequently, we manually assessed 237 articles for eligibility, excluding an additional 29 papers that did not treat the aforementioned topics as key subjects of study. Ultimately, 208 studies were included in the meta-analysis.

Table 1 Coding protocol: categories, definitions, and decision rules for each variable.

Variable	Definition	Decision rules
Year of Publication	Year in which the study was published	Extracted directly from metadata; no interpretation required
Research Area	Disciplinary field(s) associated with the study	When multiple areas are listed, group into predefined categories; assign all relevant categories
Type of Technology	NLP or AI technology addressed (e.g., LLMs, chatbots)	Assign based on primary technology discussed; allow multiple codes if clearly addressed
Co-occurring Topics	Keywords/themes appearing together within a study	Use author keywords; consolidate synonyms using thesaurus (Table 2); include if co-mentioned
Research Scope	Main objective of the study (e.g., identifying challenges, defining standards, implementation)	Assign based on dominant aim; allow multiple scopes if equally emphasized
Ethical Values	Ethical principles discussed (e.g., privacy, bias, transparency)	Code all explicitly mentioned values; group synonymous terms under unified labels
Funding Sponsor Location	Geographic origin of funding organization	Assign based on reported funding source; if multiple, include all; if unavailable, mark as missing

Methods of analysis

To analyze our corpus of 208 studies, we adopted various quantitative approaches. Our meta-analysis seeks to examine the dataset on the basis of the following variables: *Year of Publication*, *Research Area*, *Type of Technology*, *Co-occurring Topics*, *Research Scope*, *Ethical Values* and the *Location of the Funding Sponsor*.

To ensure consistency and transparency, we developed a predefined coding protocol specifying categories, definitions, and decision rules for each variable (see Table 1). These rules guided coding decisions in ambiguous cases. They included

prioritizing a study's primary focus, allowing multiple codes for co-occurring themes, and standardizing synonymous terms. Two independent coders applied this protocol, and inter-coder reliability was assessed on a subset of the corpus using Cohen's kappa coefficient ($\kappa = 0.72$), indicating substantial agreement. Remaining disagreements were resolved through discussion and consensus, resulting in a harmonized dataset.

Descriptive statistics were applied to address each research variable, and we used network representation to analyze author keyword co-occurrences, providing a visual depiction of thematic relationships.

The *Year of Publication* and *Research Area* are available in two dedicated columns within the abstract dataset from Web of Science, which could be exported in CSV format (Vicari and Komendatova, 2023). This facilitated a straightforward extraction of annual publication counts, represented in a bar chart. However, *Research Area* data required further processing, as studies were often classified under multiple research areas, resulting in an extensive list of categories. To manage this complexity, we consolidated related fields, streamlining the research areas into a manageable list of 17 categories, which we then visualized in a bar chart to reveal the thematic distribution of studies.

To delve into the *Type of Technology* each study addressed, we manually reviewed the abstracts and articles, identifying ten distinct technology types. For each type, we calculated the number of articles and documented the distribution across the dataset, offering insight into patterns in technological focus. This categorization of technology types provided an additional lens through which we could observe thematic trends, complementing the broader analysis of research areas and enhancing our understanding of focal shifts within the literature.

To further uncover relationships within the dataset, we explored the *Co-occurrences of Topics* in the studies, which refers to the frequency with which specific themes appear together within single studies across a corpus. This variable reveals patterns and relationships between topics. It highlights thematic pairings or clusters that may signal interdisciplinary connections, emerging research trends, or areas where topics are often studied in tandem (Vicari et al., 2019). This process involved several steps. First, we extracted the author keywords provided in a specific column of the Web of Science abstract dataset. Then, we consolidated synonymous terms as outlined in Table 2. We used VOSviewer (Centre for Science and Technology Studies, 2022) to create a network visualization derived from the compiled list of keywords and their instances of co-occurrence across the papers. In this visualization, each node

Table 2 Thesaurus table designed to consolidate similar terms and reduce redundancy.

Label	Replace by
Large language models	LLMs
Large language models (llm)	LLMs
Large language model (llm)	LLMs
Large language model	LLMs
LLMs	LLMs
LLM	LLMs
Machine learning	ml
Generative artificial intelligence	Generative AI
AI chatbot	Chatbots
Chatbot	Chatbots
Conversational agents	Chatbots
AI chatbots	Conversational AI

Each entry under "Label" is replaced by the corresponding standardized term in "Replace by".

symbolizes a specific keyword, with the node's size reflecting the frequency of its appearance—the larger the node, the more papers reference that keyword. A connection between nodes (representing two keywords) is formed when both keywords are mentioned in the same paper. To distinguish different clusters of nodes, a color coding system is applied. Only keywords with a minimum of two co-occurrences are included in the network map.

Through further manual screening, we identified eight distinct *Research Scopes* within the sample, many of which reflected overlapping objectives across studies. Additionally, this process revealed 30 distinct *Ethical Values* mentioned across the corpus, each of which we tallied to assess prominence. This dual categorization of research scope and ethical focus enriched our understanding of thematic nuances, providing a more nuanced perspective on the ethical underpinnings prevalent in the literature.

Finally, we analyzed the *Location of the Funding Sponsor*, as recorded in the Web of Science abstract dataset. By examining the location of sponsors, we aimed to understand the geographical influences on ethical perspectives within NLP research, as funding sources may shape the focus, values, and priorities of scientific research. Out of the 208 papers in our corpus, 67 cited a funding organization. We manually labeled these 67 papers with the *Location of the Funding Sponsor*, identifying whether the funding came from a specific country or, in cases like European Union funding, a region. In total, we identified 43 different countries. We then tallied the number of studies associated with each sponsor's location and categorized them into seven value ranges: (i) 14–10 papers, (ii) 6–9 papers, (iii) 4–5 papers, and (iv) 1–3 papers. To visually represent this distribution, we created a color-coded map, with seven distinct colors corresponding to each of these value ranges.

Our methodology applied a systematic approach to gather a comprehensive and relevant selection of studies on ethics in NLP. Key term queries guided the construction of a dataset of 208 research papers, with exclusion criteria designed to maintain a focus on ethical considerations across various NLP research contexts. The PRISMA flow diagram documented the data refinement process and resulted in a final sample suitable for detailed analysis. The study then examines publication trends, research areas, technology types, co-occurring topics, research scopes, ethical values, and funding sponsor locations. This structure provides a basis for understanding ethical discourse in NLP research, outlining the scope and distribution of ethical issues and technological applications within the selected literature, as well as identifying trends in funding support across geographic regions. The Results and Discussion section will explore these findings in detail.

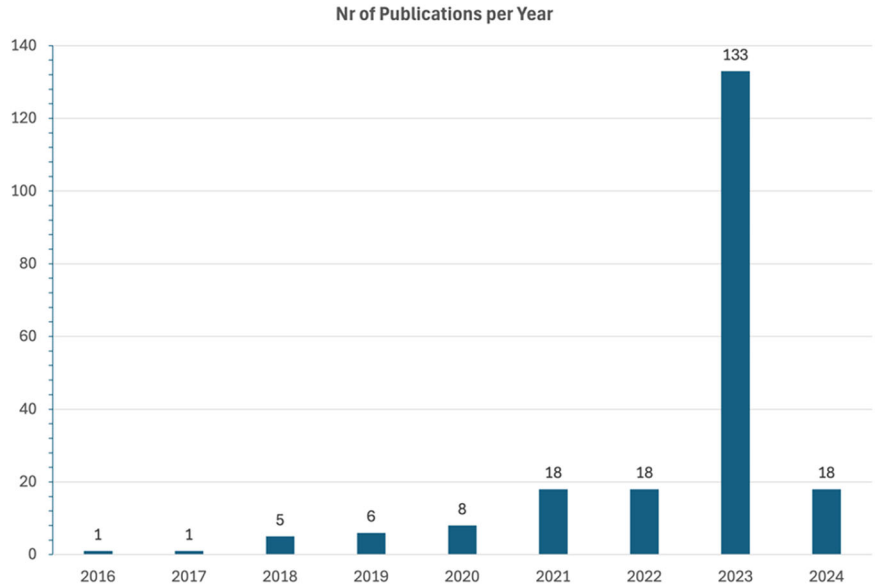

Fig. 2 Number of Publications per Year: This bar chart illustrates the increase in NLP ethics research publications, peaking in 2023, which shows a significant rise in attention towards ethical considerations in the last year.

Results and discussion

Drawing from the prior methodological framework, the *Results and Discussion* section presents a detailed analysis of the selected corpus, exploring trends, patterns, and key themes in research on NLP ethics. The data analysis delves into *Year-by-Year Publication Trends*, *Research Areas Distribution within the Corpus*, *Distribution of Studies by Technology Type*, *Co-occurrence of Topics in the Corpus*, the *Research Scopes across the Corpus*, specific *Ethical Values addressed across Studies*, and the *Location of the Funding Sponsors*.

By beginning with a chronological overview of publication rates, the section highlights shifts in research focus over time, followed by an in-depth look at the primary disciplines contributing to the field. Subsequent analyses cover the types of NLP technologies that dominate ethical discussions, a network of frequently co-occurring keywords, and a breakdown of research scopes. Finally, a closer examination of specific ethical values and the geographic distribution of research funding provides insight into global priorities and focal areas within NLP ethics.

Each subsection addresses a specific dimension of the research questions by analyzing, respectively, temporal trends (Year-by-Year Publication Trends), disciplinary distribution (Research Areas), technological focus (Technology Type), thematic relationships (Co-occurring Topics), analytical focus (Research Scopes), ethical priorities (Ethical Values), and geographical patterns (Funding Sponsor Location), thereby providing a structured and comparative understanding of how NLP ethics research evolves across contexts.

Year-by-year publication trends

The data analysis reveals a significant rise in publications addressing NLP ethics, culminating in a peak of 133 papers in 2023 (Fig. 2). This sharp increase indicates a growing focus on ethical considerations in the field of NLP, likely driven by recent advancements in language models and widespread concerns over potential ethical pitfalls, such as biases and misuse of language technologies. Prior to 2023, publication counts remained relatively low and stable, with only minor growth observed between 2016 and 2022. This trend suggests that the ethical implications of NLP only recently became a major topic of interest within the research community, reflecting the rapid pace of technological evolution and societal demands for responsible AI practices. Interestingly, despite the peak in 2023, publications in 2024 have not yet sustained this heightened output, which may indicate a consolidation phase where existing research is evaluated and contextualized. In academic research, it's common for a peak in publication activity

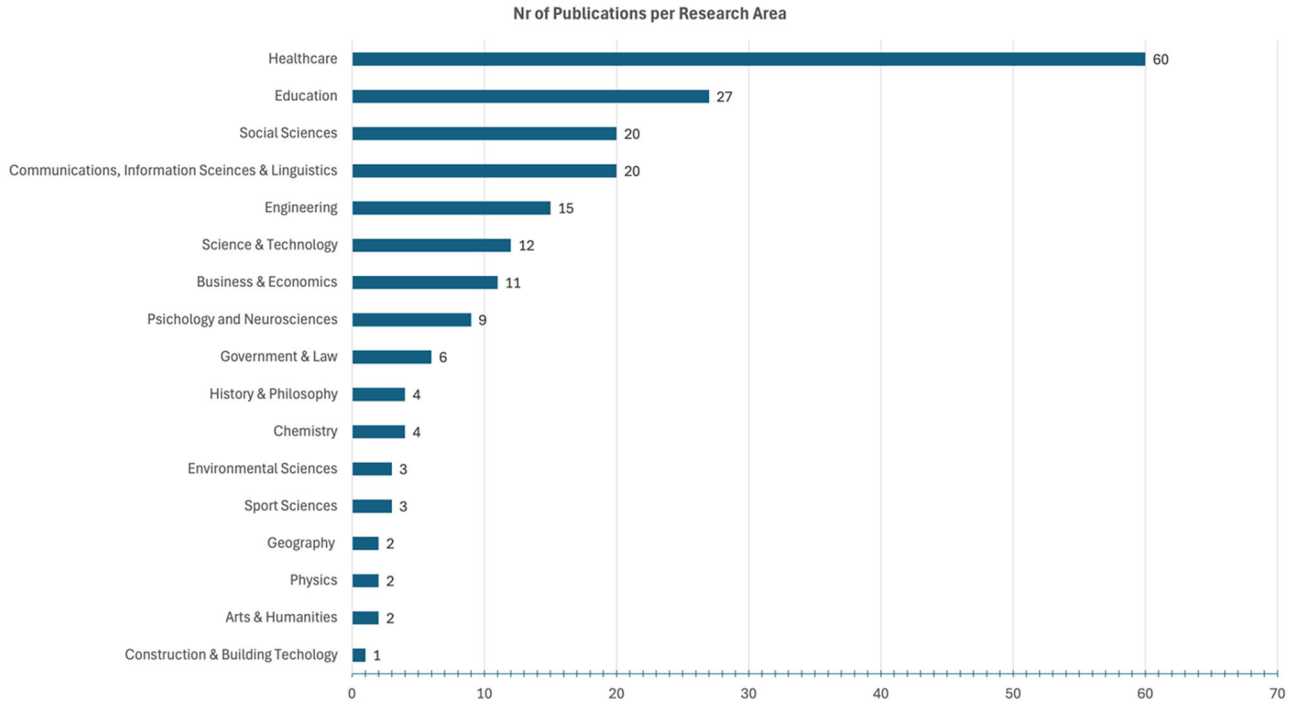

Fig. 3 Number of Publications per Research Area: This chart outlines the focus of NLP ethics research in different academic areas, with Healthcare being the most prominent research field, followed by Education, Social Sciences, Communication, Information sciences and Linguistics.

to be followed by a period of reflection, during which researchers review existing findings, address identified challenges, and integrate these insights into theoretical frameworks, methodologies, or practical applications (Bastos et al., 2021).

Research areas distribution within the corpus

Following the observed peak in NLP ethics publications in 2023, an examination of research areas highlights Healthcare as the predominant focus, with 60 studies (Fig. 3). This focus likely reflects the unique ethical challenges within healthcare, where NLP applications interact directly with sensitive patient data and assist in clinical decision-making. Education follows with 27 papers, while Social Sciences has 20 publications, and Communication, Information Sciences, and Linguistics contribute another 20 studies. In education, ethical concerns center on bias in NLP-driven assessments, NLP generated content and personalized learning systems. Social sciences explore NLP's consequences on community cohesion, social tensions and conflicts, along with its influence on personal identity and human creativity. Meanwhile, communication and linguistics emphasize the ethical challenges NLP poses in shaping public discourse and influencing the cultural sector.

Distribution of studies by technology type

Shifting focus to the types of technology, LLMs emerge as the most researched area, with 92 publications (Fig. 4). This increased focus likely stems from the widespread adoption of user-friendly conversational tools, like ChatGPT, which popularized LLMs after OpenAI released it in November 2022. Although LLMs existed since 2017, this mainstream adoption revealed the practical potential of LLMs for a broader audience and generated an intense focus on their ethical implications, sparking a rapid response in the research community. The widespread use of LLMs has underscored the need for robust, user-centered ethical guidelines as these technologies permeate daily life and impact diverse user groups across society.

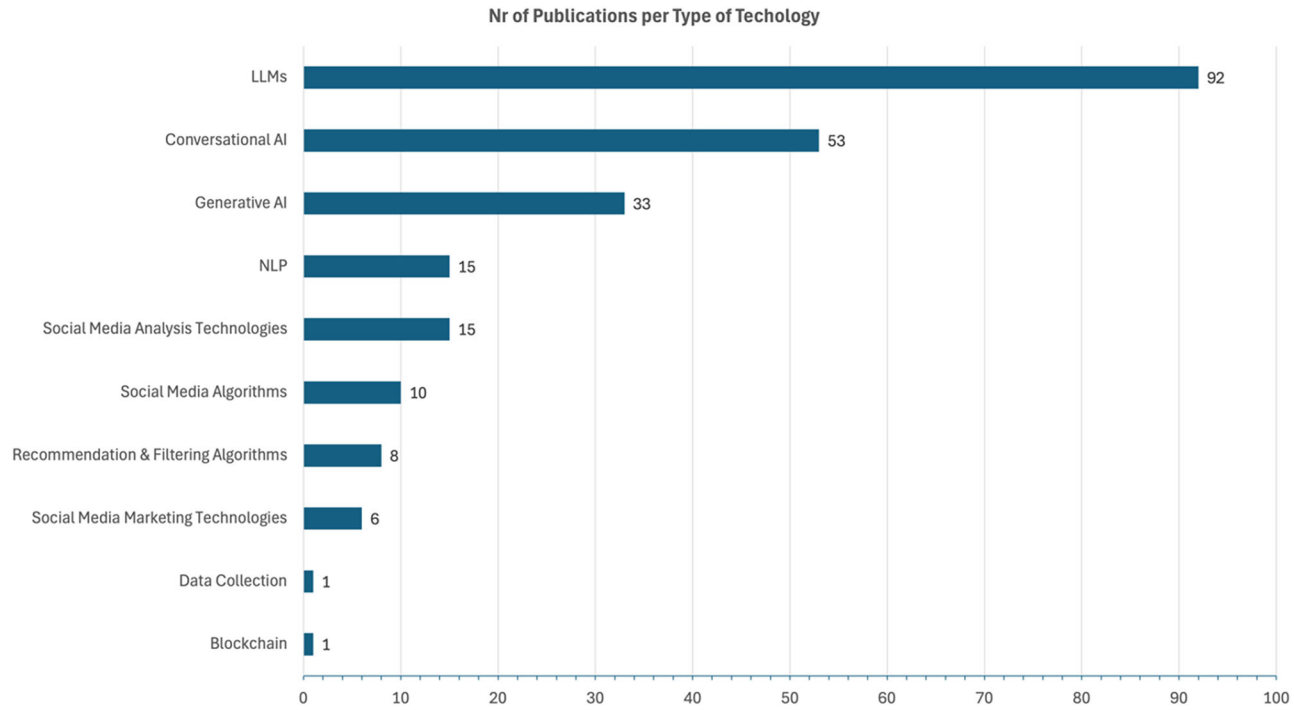

Fig. 4 Number of Publications per Type of Technology: This bar chart shows the distribution of research publications across different types of technologies like LLMs, conversational AI, and generative AI, with LLMs leading in the number of studies published.

Co-occurrence of topics in the corpus

To further explore thematic trends, a network analysis of author keywords was conducted, revealing the most central topics in NLP ethics discussions. Figure 5 visualizes frequently co-occurring keywords across papers, with focal points like *ChatGPT*, C*hatbots*, *LLMs*, *Machine Learning (ML), Education and Generative AI*. The prominence of *ChatGPT, Chatbots*, *LLMs* and *Generative AI* reflects the recent surge in ethical discussions related to large-scale conversational models, especially following the widespread public adoption of ChatGPT in 2023. This adoption has heightened interest in ethical issues, especially in critical fields such as *Education*, where AI technologies have a direct impact on shaping knowledge. The abundance of keywords related specifically to healthcare—such as *Medical Education*, *Public Health*, *Mental Health Services*, *Precision Medicine*, *Medicine*, and *Health Care*—confirms that also this field is particularly prominent within our corpus of studies.

Different clusters are color-coded to represent thematic groups based on frequently co-occurring author keywords. The network analysis reveals a structure of 12 small clusters, each representing distinct focal points within the co-occurrence of keywords. This distribution reflects a wide variety of topics without a single dominating cluster of keywords.

The blue cluster [1], including terms like *Chatbot*, *Health care*, *Covid-19*, *Data Privacy*, and *Professional Ethics*, and the light green cluster [2], including terms like *Language Models, Communication*, *Precision Medicine* and *Medical Education*, center on the ethical implications of Conversational AI in healthcare, particularly regarding the patient data and professional integrity. Nearby, the light blue cluster [3] delves into advanced NLP technologies like *ChatGPT* and *Generative Adversarial Networks*, touching on fields as diverse as *Blockchain* and *Language Learning*, illustrating the breadth of applications in the generative AI space. Moving to the essential principles of ethical governance, the green cluster [4] weaves together themes like *Bias*, *Transparency*, and *Informed Consent* with the legal and regulatory considerations of using AI responsibly.

Shifting towards societal impacts, the cyan cluster [5] highlights values like *Trust* and *Gender*, alongside the influence of *Journalism* and *Deep Learning*, pointing to the need for ethical reflection as AI systems mediate public

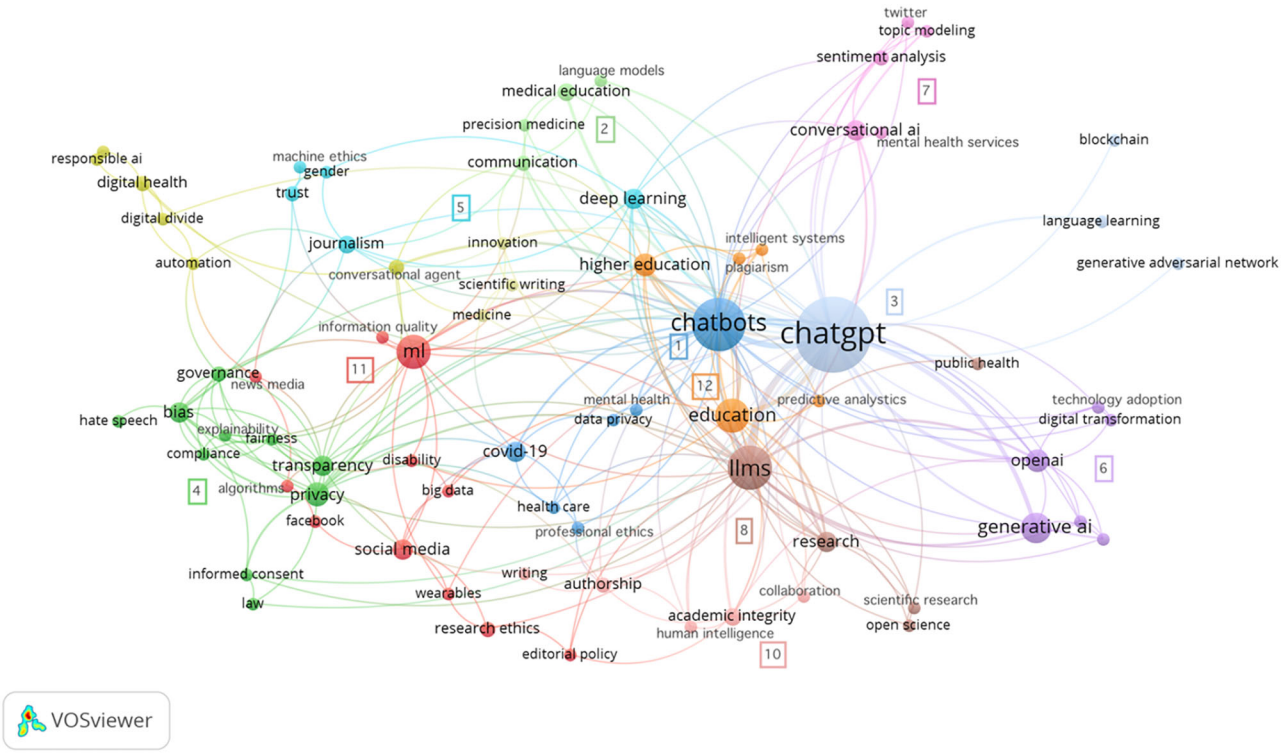

Fig. 5 Author Keywords Network: A network representation of author keywords that frequently co-occur across papers in our corpus, highlighting central topics like ChatGPT, Chatbots, LLMs, and Machine Learning in the discourse.

information. The purple cluster [6] addresses the broader landscape of *Technology Adoption* and *Digital Transformation*, with *Generative AI* and *OpenAI* as pivotal players driving change across industries.

The pink cluster [7] explores NLP's role in understanding public sentiment on *Twitter* and supporting *Mental Health Services*, with applications in *Conversational AI*, *Topic Modeling*, and *Sentiment Analysis*.

In the brown cluster [8], we see a focus on *LLMs* and *Open Science*, which speaks to the ethical considerations surrounding accessibility in academic *Research*, while the pale red cluster [9] emphasizes challenges related to academic *Authorship*, *Integrity*, and *Collaboration* in research. The complementary yellow cluster [10] reflects *Innovation*-driven efforts in fields like *Scientific Writing* and *Medicine*, signaling NLP's impact on advancing academic knowledge.

Meanwhile, the red cluster [11] examines the deployment of *ML* in social domains like *Disability*, *Social Media*, *News Media* and *Wearables*, emphasizing the importance of ethical standards, such as privacy and accuracy, in these applications.

Finally, the orange cluster [12] connects *Education* and *Higher Education* with *Predictive Analytics* and issues of *Plagiarism*, capturing the ethical responsibilities inherent in using AI across educational platforms.

Together, these clusters highlight the diverse ethical terrain in NLP, showcasing both the sector-specific applications and the overarching principles needed to guide responsible development in this rapidly evolving field.

Research scopes across the corpus

While the keyword network illustrates the core themes and applications in NLP ethics, Fig. 6 provides a complementary view by categorizing the primary research scopes within this field. The largest number of studies addresses *Identifying Ethical Challenges* (146 papers) and *Defining Ethical Standards* (143 papers), highlighting a foundational focus on outlining the specific ethical issues within NLP and establishing guiding principles to address

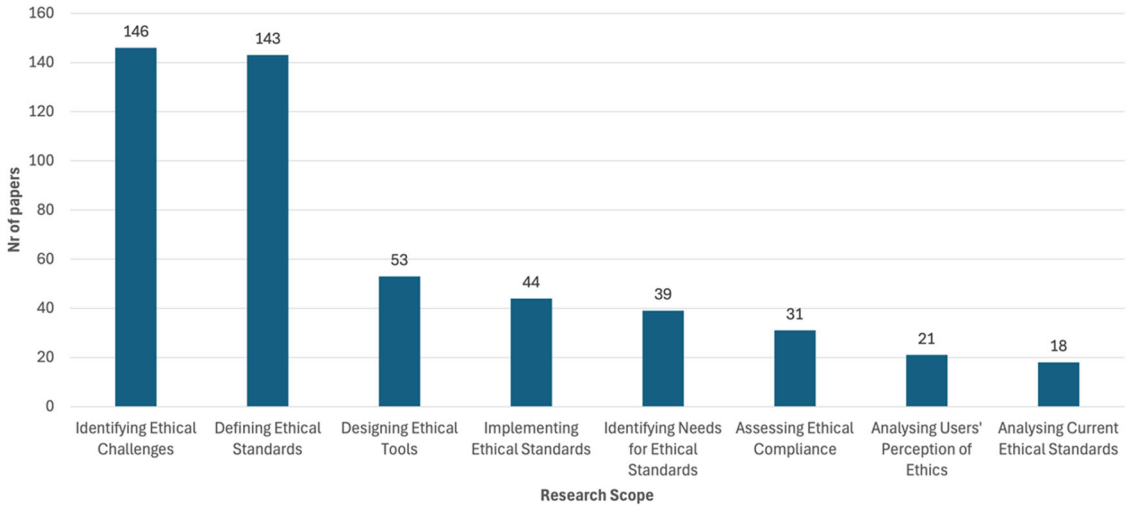

Fig. 6 Overview of Research Scope in Ethics of NLP Studies: The bar chart displays the number of academic papers that address different research scopes within ethics in NLP.

them. This emphasis suggests that researchers prioritize an understanding of potential risks and the development of ethical frameworks as NLP technologies continue to evolve.

Fewer studies concentrate on *Designing Ethical Tools* (53 papers) and *Implementing Ethical Standards* (44 papers), pointing to efforts aimed at translating ethical principles into practical solutions. The research areas with the fewest studies include *Assessing Ethical Compliance* (31 papers) and *Analyzing Users' Perception of Ethics* (21

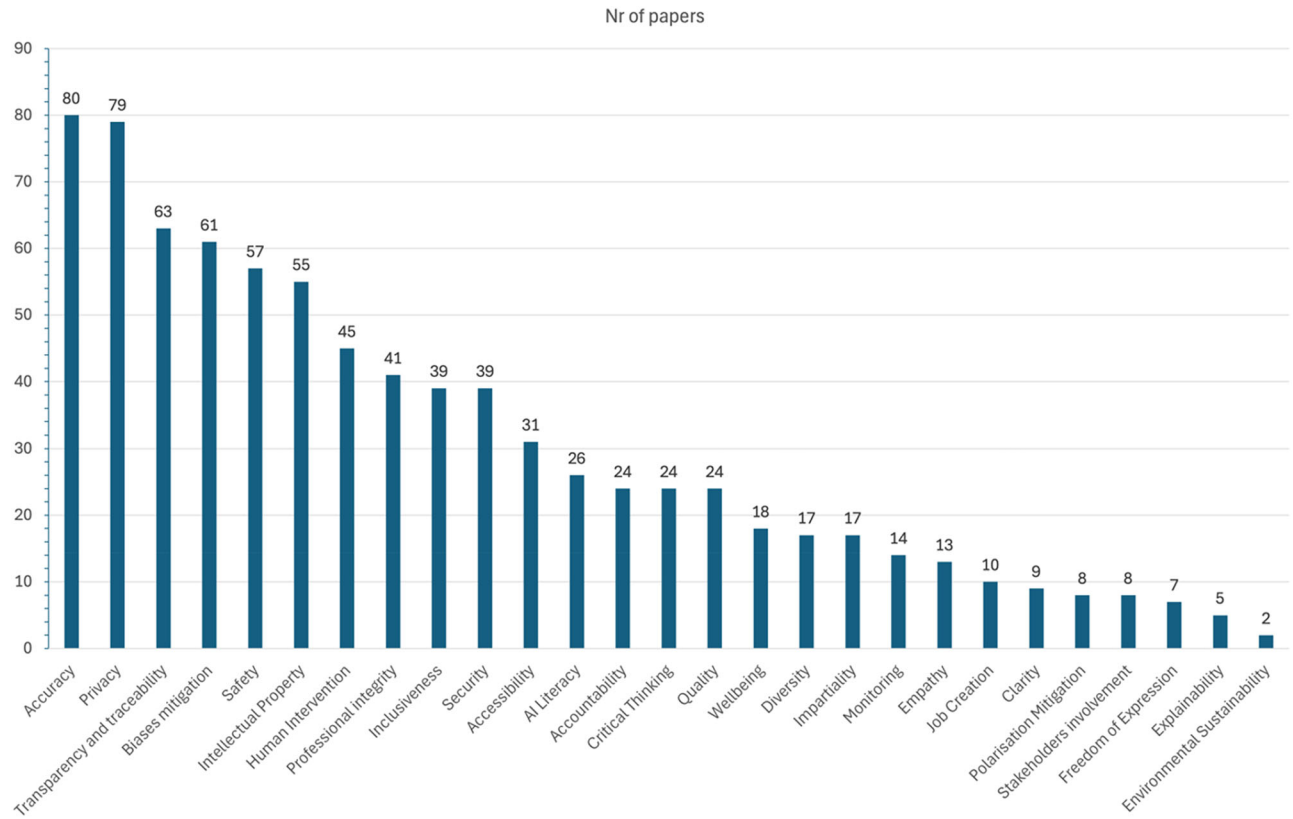

Fig. 7 Distribution of Ethical Values Addressed in NLP Ethics Research: The bar chart illustrates the number of academic papers that address various ethical values in the context of NLP.

Table 3 Ethical values addressed in NLP Ethics Research: This table presents a list of key ethical values and their definitions, as identified through a comprehensive review of 208 papers on ethics in NLP.

Ethical Value	Definition
Accessibility	Ensuring that NLP technologies are available and usable for all individuals, regardless of disabilities or limitations.
Accountability	Holding developers and stakeholders responsible for the outcomes and impacts of NLP technologies.
Accuracy	Ensuring that NLP systems provide correct and reliable information without errors or misleading content.
AI Literacy	Supporting users in developing the skills to understand, interpret, and effectively engage with NLP tools, to enhance informed and responsible use.
Bias Mitigation	Reducing and preventing biases in NLP models to promote fair and equitable outcomes across diverse populations.
Clarity	Providing clear, understandable outputs from NLP systems that avoid ambiguity and promote transparency in communication.
Critical Thinking	Encouraging the development of NLP systems that support and promote critical analysis rather than passive consumption of information.
Diversity	Ensuring that NLP systems respect and represent diverse cultural, social, and linguistic backgrounds.
Intellectual Property	Protecting the rights of creators and ensuring that NLP systems respect copyright and ownership of original content.
Environmental Sustainability	Minimizing the environmental impact of NLP systems, particularly regarding energy consumption and carbon emissions from training large models.
Empathy	Designing NLP systems that understand and respond to human emotions in a sensitive and considerate manner.
Explainability	Ensuring that the decision-making processes of NLP models are understandable and interpretable to users.
Freedom of Expression	Protecting the right of individuals to express themselves freely without censorship or suppression by NLP systems.
Human Intervention	Allowing human oversight and input in NLP processes to maintain control and prevent automated errors.
Inclusiveness	Creating NLP systems that consider the needs of all users, fostering inclusivity in technology use.
Impartiality	Ensuring that NLP systems remain neutral and unbiased in their analyses and outputs.
Job Creation	Exploring opportunities for NLP technologies to create new job roles rather than solely automating existing ones.
Monitoring	Establishing mechanisms to observe and regulate the functioning of NLP systems to ensure compliance with standards.
Polarization Mitigation	Preventing NLP technologies from contributing to societal divides or reinforcing polarized viewpoints.
Privacy	Safeguarding user data and ensuring that NLP systems do not infringe on personal privacy rights.
Professional Integrity	Ensuring that NLP technologies are developed and deployed with a commitment to honesty, reliability, and responsibility, maintaining the trust and standards expected in professional environments.
Quality	Ensuring high standards in the performance, and reliability of NLP systems.
Safety	Protecting users from potential harm by ensuring that NLP systems operate securely and predictably.
Security	Protecting NLP systems from cyber threats and unauthorized access to safeguard user data and system integrity.
Stakeholder Involvement	Engaging all relevant stakeholders in the development and implementation of NLP technologies to consider diverse perspectives.
Transparency & Traceability	Providing clear information on how NLP systems work and tracking decision-making processes for accountability.
Wellbeing	Designing NLP technologies that promote the physical, mental, and social wellbeing of users.

papers), highlighting an emerging dimension of how users perceive NLP ethics and how well existing frameworks are applied. These results suggest that current research places greater emphasis on identifying ethical issues and defining standards than on practical implementation and understanding user perspectives.

Ethical values addressed across studies

Moving from general research scopes to specific ethical concerns, Fig. 7 and Table 3 reveal which values dominate research studies and which may be overlooked. The bar chart in Fig. 7 shows that *Accuracy* is the most frequently addressed ethical principle, with 80 papers dedicated to this theme. This reveals a frequent core goal of maintaining information reliability and credibility, particularly as NLP systems influence public discourse and decision-making. *Privacy*, which appears in 79 papers, is another current ethical value, reflecting ongoing concerns about safeguarding user data.

Table 3 further defines each ethical value based on our review of 208 studies in the field of NLP ethics. For instance, *Bias Mitigation* is defined as promoting fairness across diverse populations, ensuring NLP models do not

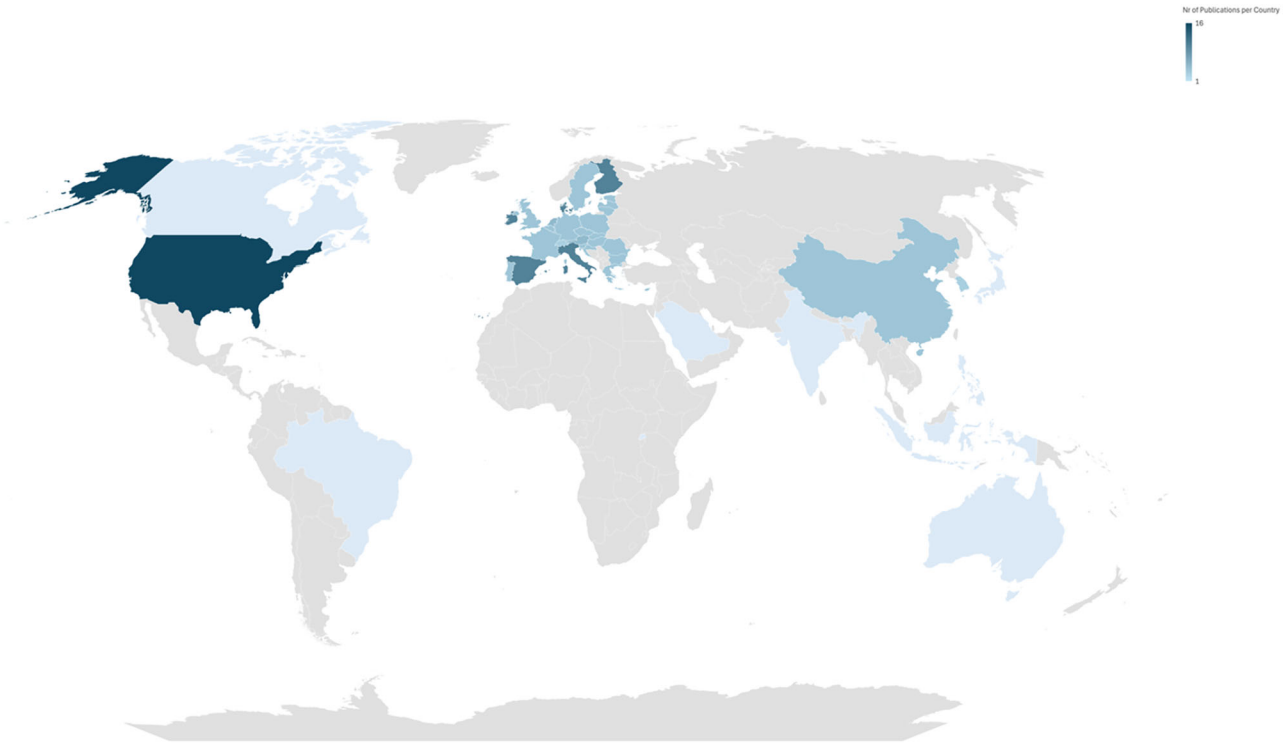

Fig. 8 Sponsor Location: A global heatmap that shows the geographical distribution of research sponsors in NLP ethics, with the majority of research sponsored by institutions based in the United States, followed by European countries and several Asian nations.

propagate discriminatory outcomes. This value, discussed in 61 papers, reflects the emphasis on equitable AI and the ethical imperative to prevent harm due to biases embedded in training data. *Transparency and Traceability* (addressed in 63 papers) are essential for accountability, as outlined in Table 3, and involve making NLP decision-making processes clear to users and stakeholders, enhancing trust and understanding. *Intellectual Property*, addressed in 55 papers, highlights the importance of respecting creators' rights, ensuring that NLP models honor copyright and ownership of original content. This is particularly relevant as AI models become increasingly influential in the cultural sector, raising complex ethical and legal questions about the ownership and authorship of AI-generated content.

Notably, values like *Environmental Sustainability* and *Explainability* receive less attention, with only 2 and 5 papers, respectively, focusing on these areas. *Environmental Sustainability* addresses the environmental impact of NLP systems, as large-scale AI models contribute to a significant carbon footprint. Greater focus on this value could encourage development practices that are more aligned with global sustainability goals. *Explainability* is crucial for making NLP systems' decisions understandable to users. As these systems increasingly influence real-world decisions, *Explainability* can ensure that users can make informed interactions with NLP tools.

Similarly, values like *Stakeholders Involvement* and *Freedom of Expression*, with 8 and 7 studies raise limited interest. *Stakeholder Involvement* ensures diverse perspectives guide NLP development, which can reduce bias and improve social relevance. *Freedom of Expression* is key in maintaining the public debate open, especially as NLP models play a growing role in moderating and generating content online.

The limited focus on these values indicates an opportunity for future studies to adopt a more comprehensive approach to NLP ethics, addressing not only core principles but also emerging societal and ecological concerns.

Table 4 Geographical distribution of funding sponsors for NLP ethics research studies, categorized by color to represent the number of papers from each region.

Funding Sponsor's Location	N° of Studies	Corresponding Color
United States	10–16 papers	
Denmark, Ireland, Finland, Italy, Spain	6-9 papers	
Austria, Belgium, Bulgaria, China, Croatia, Cyprus, Czech Republic, Estonia, France, Germany, Greece, Hungary, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Sweden, UK	4-5 papers	
Australia, Brazil, Canada, India, Indonesia, Japan, Philippines, Rwanda, Saudi Arabia, Singapore, South Korea, Switzerland,	1–3 papers	

The darkest blue indicates the United States with 10–16 papers. A medium blue shade represents Denmark, Ireland, Finland, Italy, and Spain with 6–9 papers each. A lighter blue shade covers countries contributing 4–5 papers, while the lightest blue denotes regions with 1–3 papers.

Location of the funding sponsors

After examining the ethical values prioritized in NLP ethics research, it is also relevant to consider the geographic distribution of sponsorship, as funding sources can significantly influence research directions and priorities. The global distribution of research sponsorship in NLP ethics, as illustrated in Fig. 8 and Table 4, reveals a strong concentration of funding from institutions in the United States, which supports the majority of studies in this field. With 16 publications funded by U.S.-based sponsors, the United States emerges as a central player, likely due to the concentration of major tech companies (Kollnig and Li, 2023) and research institutions dedicated to advancing AI ethics, such as the Partnership on AI(Partnership on AI, 2024), the AI Now Institute (AI Now Institute, 2024), and the National Science Foundation (U.S. National Science Foundation, 2024).

Following the U.S., several European countries, including Denmark (7 studies) and Ireland (7 studies), contribute notably to sponsorship, collectively reflecting Europe's growing commitment to responsible AI practices and ethical frameworks in technology development. This European focus is in line with recent regulatory initiatives, such as the European Union's AI Act (European Commission, 2024a), which prioritizes transparency, accountability, and ethical standards in AI applications.

In addition to North America and Europe, several Asian countries, including China, India, and Saudi Arabia, have also sponsored research, contributing to a more diversified geographic landscape.

The geographical concentration of funding, particularly in the United States and Europe, with Western countries collectively supporting 44 studies, likely influences current priorities in NLP ethics research. The strong U.S.-based funding presence may drive an emphasis on healthcare ethics, as the U.S. has a prominent healthcare technology sector that prioritizes data privacy and security in medical applications (Kiel, 2022). Additionally, the regulatory cultures in both the U.S. and Europe tend to prioritize identifying ethical challenges and establishing standards over direct implementation. This emphasis is evident in initiatives like the EU's AI Act (European Commission, 2024a), DSA (European Commission, 2024b) and GDPR (European Parliament and European Council, 2016), in Europe, and Blueprint for an AI bill of Rights (The White House, 2024) and the NIST Risk Management Framework (U.S. Department of Commerce, 2024) in the United States. Consequently, this funding distribution may have directed attention toward ethical values like accuracy, privacy, bias mitigation, and transparency—values that align with Western concerns around misinformation, data protection, discrimination, social equity and accountability.

In this section, we have examined key patterns and focal areas within NLP ethics research, identifying trends in publication volume, primary research domains, prevalent technologies, and specific ethical values addressed. The analysis highlights a strong emphasis on foundational ethical principles like accuracy, privacy, and bias mitigation, while also revealing gaps in areas such as environmental sustainability, explainability, and stakeholder involvement.

The geographic distribution of funding further suggests that Western countries, particularly the U.S. and Europe, play a central role in shaping research priorities. Together, these findings address the research questions by providing a comparative and structured understanding of how the debate on NLP ethics is distributed over time, across research areas, technologies, and regions, as well as which research scopes and ethical values emerge most prominently in the literature.

In the following *Conclusions and Perspectives* section, we will synthesize these insights, discussing how current trends inform the state of NLP ethics research and proposing directions for future work aimed at creating more balanced and globally inclusive ethical frameworks.

Conclusions and perspectives

As NLP technologies become embedded in decision-making processes, content moderation systems, and creative tasks in various fields, assessing their ethical implications becomes increasingly important for ensuring responsible and transparent use. This study addresses the following research question: while considerable research exists on isolated ethical issues within specific NLP applications, comprehensive frameworks addressing ethics across the full range of NLP applications are limited. To address this, we conducted a meta-analysis of 208 studies on NLP ethics, systematically selected from an initial pool of 465 abstracts sourced from Web of Science. Using both quantitative and qualitative approaches, we analyzed seven variables, including *Year of Publication*, *Research Area*, *Type of Technology*, *Co-occurring Topics*, *Research Scope*, *Ethical Values* and the *Location of the Funding Sponsor*, to provide a structured overview of the ethical landscape in NLP research.

The findings of this meta-analysis underscore intensifying focus on the ethical considerations surrounding NLP technologies. A notable finding is the sharp rise in publications on LLMs in 2023, underscoring a growing focus on NLP ethics as this technology permeates high-impact fields. However, this initial surge appears to have entered a period of reflection, signaling a shift towards consolidating existing research and refining theoretical frameworks. As NLP technologies permeate everyday decision-making, content generation, and user interactions, this phase may facilitate the development of more robust ethical frameworks that balance the technological potential of NLP with social accountability.

The healthcare sector, in particular, stands out, reflecting the unique ethical concerns associated with patient care and decision-making integrity. The prominent focus on healthcare ethics within NLP research suggests a need for sustained exploration of other sectors. Leveraging existing ethical frameworks from healthcare, researchers could adapt these tools for use in other fields.

While significant emphasis is placed on identifying ethical challenges and setting standards, the limited number of studies on the practical implementation of these standards reveals a disconnection between theory and application. This space for development suggests a valuable opportunity for applied research to translate theoretical guidelines into real-world tools and practices, making ethical standards actionable.

Our findings also reveal that the most common ethical values addressed—accuracy, privacy, bias mitigation, and transparency—dominate the discourse, indicating a focus on ensuring trustworthy and fair AI systems. However, the limited attention to other values, such as environmental sustainability, explainability, stakeholder involvement and freedom of expression, points to an opportunity for expanding the ethical framework to address less visible considerations. As NLP applications continue to evolve, integrating these overlooked values could create a more holistic ethical framework.

The *Topic Co-occurrence* analysis further corroborates our findings on *Research Area Distribution*, *Technology Type Distribution*, and *Ethical Values*. Our network representation of co-occurring keywords reveals that the recent surge in *Conversational AI*, particularly with models like *ChatGPT*, has intensified ethical discussions around *Bias*, *Privacy*, and *Transparency*, in particular in the fields of healthcare and education.

The geographical distribution of research funding, primarily concentrated in the United States and Europe, likely shapes the ethical priorities reflected in this body of work. The funding landscape indicates an alignment between ethical research agendas and Western regulatory priorities, such as data privacy, transparency, and bias prevention.

Expanding sponsorship beyond these regions, especially in emerging markets, could diversify ethical perspectives in NLP, enriching the research field with culturally contextualized insights.

These results suggest a need to move beyond fragmented, domain-specific analyses toward more integrated, cross-sectoral ethical frameworks. They also highlight the importance of translating ethical principles into practical guidelines, tools, and governance mechanisms. Compared to previous studies that focus on specific applications or isolated ethical issues, this work contributes a systematic and cross-application synthesis of NLP ethics, offering a structured, multi-dimensional perspective across research scopes, ethical values, technologies, domains, and regions. This study is subject to limitations, including potential selection bias due to the use of explicit ethical terminology in the search strategy, the predominance of English-language and Western-centered publications, the under-representation of conference outputs in NLP, and the limited availability of funding data; future work should address these constraints by broadening search criteria, incorporating more diverse and multilingual sources, and improving data coverage to enable more balanced and robust analyses. This analysis identifies directions for expanding the ethical landscape in NLP research. Future studies could focus on translating ethical principles into practical guidelines, tools, and best practices that make these standards actionable in real-world applications. Additionally, fostering collaborations beyond current regional and disciplinary boundaries may enhance the inclusivity of ethical frameworks by integrating a larger panel of values such as sustainability and stakeholder involvement. By addressing these areas, future research can support the development of NLP technologies that are aligned with evolving societal expectations and ethical standards. Cross-disciplinary studies on the long-term impacts of NLP in various fields would also help monitor how ethical concerns evolve over time, informing adaptable and context-sensitive guidelines that align NLP's development with broader societal and cultural priorities.

Overall, this study provides a systematic and cross-application synthesis of NLP ethics research, revealing a rapidly growing field shaped by the rise of LLMs and concentrated in specific domains such as healthcare. The findings highlight both dominant ethical priorities—such as accuracy, privacy, and bias mitigation—and critical gaps, particularly in implementation, underexplored values, and geographic diversity, with implications for developing more inclusive and actionable ethical frameworks. By examining trends across time, domains, technologies, and regions, this study provides a structured and comparative perspective that supports more coherent governance approaches.

Author contributions R.V. Concept, Methodology, Data curation, Vizualization, Writing; N.K. Supervision, Reviewing, Funding Management

Funding This work was supported by European Union's Horizon 2020 research and innovation program under grant agreement No. 101021746, CORE (science and human factor for resilient society).

Data availability The data that support the findings of this study are available from the authors upon reasonable request, subject to the access and usage restrictions of the source database.

Declarations

Competing interests The authors declare no competing interests.

Ethics approval and consent to participate Ethical approval and informed consent were not required, as no human participants were involved in this research.

Declaration of generative AI and AI-assisted technologies in the writing process During the preparation of this work the author(s) used ChatGPT 3.5 in order to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by

statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

AI Now Institute (2024) AI Now Institute. https://ainowinstitute.org/
Alhafni B, Vajjala S, Bannò S, Maurya K, Kochmar E (2024) LLMs in education: novel perspectives, challenges, and opportunities. Preprint at. https://arxiv.org/abs/2409.11917
Alphasense Inc. Available at: https://www.alpha-sense.com/. Accessed 27 Oct 2027
Ball S (2024) LLM use-by data – who's keeping the data up to date? ERP Today. https://erp.today/llm-use-by-data/
Bastos EC, Sengik AR, Tello-Gamarra J (2021) Fifty years of University-industry collaboration: a global bibliometrics overview. Sci Public Policy 48(2):177–199. https://doi.org/10.1093/scipol/scaa077
Budhwar P, Chowdhury S, Wood G, Aguinis H, Bamber GJ, Beltran JR, Boselie P, Lee Cooke F, Decker S, DeNisi A, Dey PK, Guest D, Knoblich AJ, Malik A, Paauwe J, Papagiannidis S, Patel C, Pereira V, Ren S, Varma A (2023) Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. Hum Resour Manag J 33(3):606–659. https://doi.org/10.1111/1748-8583.12524
Centre for Science and Technology Studies (2022) VOSviewer. https://www.vosviewer.com/
ComplyAdvantage. Available at: https://complyadvantage.com/. Accessed 27 Oct 2027
Council of Europe (2024) Framework convention on artificial intelligence and human rights, democracy and the rule of law. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence
European Commission (2024a) AI Act. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
European Commission (2024b) The Digital Services Act. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en
European Parliament, & European Council (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504
Farrelly T, Baker N (2023) Generative artificial intelligence: implications and considerations for higher education practice. Educ Sci 13(11):1109. https://doi.org/10.3390/educsci13111109
Fatani B (2023) ChatGPT for future medical and dental research. Cureus. https://doi.org/10.7759/cureus.37285
Gao R, Zhang Z, Shi Z, Xu D, Zhang W, Zhu D (2021) A review of natural language processing for financial technology. In: Nakashima S, Mu S, Lu H (eds.) International Symposium on Artificial Intelligence and Robotics 2021. SPIE, p 54. https://doi.org/10.1117/12.2604371
Gómez-Rodríguez C, Williams P (2023) A Confederacy of models: a comprehensive evaluation of LLMs on creative writing. Findings of the Association for Computational Linguistics: EMNLP 2023, 14504–14528. https://doi.org/10.18653/v1/2023.findings-emnlp.966
Hartmann J, Netzer O (2023) Natural language processing in marketing, pp 191–215. https://doi.org/10.1108/S1548-643520230000020011
Hutchings SC, Dixit Y, Al-Sarayreh M, Torrico DD, Realini CE, Jaeger SR, Reis MM (2023) A critical review of social media research in sensory-consumer science. Food Res Int 165: 112494. https://doi.org/10.1016/j.foodres.2023.112494
Kira Inc. Available at: https://kirasystems.com/. Accessed 27 Oct 2027
Jurafsky D, Martin JH (2009) Speech and Language Processing. Pearson Prentice Hall
Kiel JM (2022) Data privacy and security in the US: HIPAA, HITECH and beyond. pp. 427–435 https://doi.org/10.1007/978-3-030-91237-6_28
Kim JK, Chua M, Rickard M, Lorenzo A (2023) ChatGPT and large language model (LLM) chatbots: the current state of acceptability and a proposal for guidelines on utilization in academic medicine. J Pediatr Urol 19(5):598–604. https://doi.org/10.1016/j.jpurol.2023.05.018
Kollnig K, Li Q (2023) Exploring antitrust and platform power in generative AI. Preprint at https://arxiv.org/abs/2306.11342
Kreminski M, Martens C (2022) Unmet creativity support needs in computationally supported creative writing. In Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022), 74–82. https://doi.org/10.18653/v1/2022.in2writing-1.11
Llansó E, van Hoboken J, Leerssen P, Harambam J (2020) Artificial Intelligence, Content Moderation, and Freedom of Expression. The Transatlantic Working Group Papers Series
Luminance Technologies Ltd. Available at: https://www.luminance.com/. Accessed 27 Oct 2027
Nakaura T, Ito R, Ueda D, Nozaki T, Fushimi Y, Matsui Y, Yanagawa M, Yamada A, Tsuboyama T, Fujima N, Tatsugami F, Hirata K, Fujita S, Kamagata K, Fujioka T, Kawamura M, Naganawa S (2024) The impact of large language models on radiology: a guide for radiologists on the latest innovations in AI. Jpn J Radiol 42(7):685–696. https://doi.org/10.1007/s11604-024-01552-0

Omarov B, Narynov S, Zhumanov Z (2023) Artificial intelligence-enabled chatbots in mental health: a systematic review. Comput Mater Contin 74(3):5105–5122. https://doi.org/10.32604/cmc.2023.034655
OpenAI. ChatGPT. Available at: https://openai.com/chatgpt/. Accessed 27 Oct 2027
Page MJ, McKenzie JE, Bossuyt PM et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71
Partnership on AI (2024) Partnership on AI. https://partnershiponai.org/
Renuka S, Raj Kiran GSS, Rohit P (2021) An Unsupervised Content-Based Article Recommendation System Using Natural Language Processing. pp. 165–180. https://doi.org/10.1007/978-981-15-8530-2_13
Saquete E, Tomás D, Moreda P, Martínez-Barco P, Palomar M (2020) Fighting post-truth using natural language processing: a review and open challenges. Expert Syst Appl 141: 112943. https://doi.org/10.1016/j.eswa.2019.112943
Srivastav S, Chandrakar R, Gupta S, Babhulkar V, Agrawal S, Jaiswal A, Prasad R, Wanjari MB (2023) ChatGPT in radiology: the advantages and limitations of artificial intelligence for medical imaging diagnosis. Cureus. https://doi.org/10.7759/cureus.41435
Tabnine Ltd. Available at: https://www.tabnine.com/. Accessed 27 October 2027)
The White House (2024) Blueprint for an AI Bill of Rights. https://www.whitehouse.gov/ostp/ai-bill-of-rights/
United Nations. (1948) Universal Declaration of Human Rights. https://www.ohchr.org/en/universal-declaration-of-human-rights
U.S. Department of Commerce, N. I. of S. and T. (2024) NIST Risk Management Framework. https://csrc.nist.gov/projects/risk-management/about-rmf
U.S. National Science Foundation (2024) Artificial Intelligence. https://new.nsf.gov/focus-areas/artificial-intelligence
Vicari R, Komendatova N (2023) Systematic meta-analysis of research on AI tools to deal with misinformation on social media during natural and anthropogenic hazards and disasters. Humanit Soc Sci Commun 10(1):332. https://doi.org/10.1057/s41599-023-01838-0
Vicari R, Tchiguirinskaia I, Tisserand B, Schertzer D (2019) Climate risks, digital media, and big data: following communication trails to investigate urban communities' resilience. Nat Hazards Earth Syst Sci 19(7):1485–1498. https://doi.org/10.5194/nhess-19-1485-2019
Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz, J (2023) Beyond ChatGPT: what does GPT-4 add to healthcare? The dawn of a new era. Cardiol J. https://doi.org/10.5603/cj.97515
Xu L, Sanders L, Li K, Chow JCL (2021) Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 7(4):e27850. https://doi.org/10.2196/27850
Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DSW, Liu N (2023) Large language models in health care: development, applications, and challenges. Health Care Sci 2(4):255–263. https://doi.org/10.1002/hcs2.61
Zhong H, Xiao C, Tu C, Zhang T, Liu Z, Sun, M (2020) How Does NLP benefit legal system: a summary of legal artificial intelligence.In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5218–5230. https://doi.org/10.18653/v1/2020.acl-main.466

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Rosa Vicari[1] · **Nadejda Komendantova**[1]

[1] International Institute for Applied Systems Analysis, Laxenburg, Austria