



LUDOVIKA
UNIVERSITY PRESS

Volume 14. (2026)
No. 1.

ISSN 2063-7330 (online)

KOME

**An International Journal
of Pure Communication Inquiry**

KOME – An International Journal of Pure Communication Inquiry

Vol. 14. No. 1. 2026

EDITOR-IN-CHIEF:

János Tóth

EDITORIAL ASSISTANT:

Ákos Tussay

EDITORIAL BOARD:

Jessica Brophy (*Bridgewater College, USA*)
Péter Cseke (*Babes-Bolyai University, Cluj-Napoca, Romania*)
Márton Demeter (*Ludovika University of Public Service, Hungary*)
Katalin Fehér (*Budapest Business School, Hungary*)
Sarah Gehlert (*Washington University, St. Louis, USA*)
Özséb Horányi (*Corvinus University of Budapest, Hungary*)
Imre Lázár (*Semmelweis University, Hungary*)
Sue McGregor (*Mount Saint Vincent University, Canada*)
Claudia Mellado (*Pontificia Universidad Católica de Valparaíso, Chile*)
Basarab Nicolescu (*CNRS, Paris, France*)
Natalie Pang (*Nanyang Technological University, Singapore*)
Panayiota Tsatsou (*University of Leicester, UK*)
Tudor Vlad (*Georgia University, Athens, USA*)
Ed Vollans (*Middlesex University, London, UK*)

Published by the Ludovika University of Public Service
Ludovika University Press
Ludovika tér 2. 1083 Budapest, Hungary
Responsible for publishing
Gergely Deli, Rector
Copy editors: Zsuzsánna Gergely, Ágnes Resofszi

ISSN 2063-7330 (online)

Reproduction is authorized, provided the source is acknowledged.



LUDOVIKA
UNIVERSITY PRESS

Contents

Parallel Empires of Knowledge

Márton Demeter, Ferenc Petruska, Gergő Háló, Erik Palusek 3

From Tools to Teammates: Reimagining Moral Agency and Human–Robot Partnerships in Journalism Practice

Wenxi Wen, Ashraf Sadat Ahadzadeh, Kam-Fong Lee 43

Perception of Information and Disinformation on Social Media

Beatriz Catalina-García, Antonio García-Jiménez,
Manuel Montes-Vozmediano 73

Hijacked Victimhood: The Political Communication Strategy of Vice President Sara Duterte at Congress Hearings

Audrey Buenavista Morallo 95

Political Communication on Telegram

Rubén Rivas-de-Roca, Concha Pérez-Curiel 115

Data Pollution as Resistance

János József Tóth 137

Parallel Empires of Knowledge

AI and the Fracturing of Global Science

Márton Demeter,^{*} Ferenc Petruska,^{**} Gergő Háló,^{***} Erik Palusek^{****}

^{*} Universidad Complutense de Madrid, Facultad de Ciencias de la Información, SPAIN
E-mail: mdemeter[at]ucm.es

^{**} Department of Defence Law and Administration, Ludovika University of Public Service, HUNGARY
E-mail: petruska.ferenc[at]uni-nke.hu

^{***} Office for Science Strategy, Ludovika University of Public Service, HUNGARY
E-mail: halo.gergo[at]uni-nke.hu

^{****} Office for Science Strategy, Ludovika University of Public Service, HUNGARY
E-mail: palusek.erik[at]uni-nke.hu

This study examines shifting patterns in global academic knowledge production through the lens of world-systems theory, focusing on the role of generative artificial intelligence (AI) in reshaping epistemic hierarchies. Drawing on longitudinal bibliometric data (1996–2023), it analyses publication volumes, international collaborations and open access trends across core (U.S., U.K.), semi-peripheral and peripheral regions, with special attention to China’s emergence as a leading scientific producer. The paper highlights a growing divergence in AI infrastructure and access: Western scholars increasingly rely on open tools like ChatGPT, while China’s closed ecosystem is governed by distinct political and epistemic norms. These developments may entrench parallel scientific systems, exacerbating inequalities in visibility, legitimacy and collaboration. As AI becomes a central driver of research practices, this paper argues for inclusive, interoperable knowledge infrastructures to avoid deepening global academic fragmentation. The findings offer a critical perspective on the geopolitics of knowledge in the digital age.

Keywords: world-systems theory, generative artificial intelligence, global knowledge production, academic bifurcation

Address for correspondence: Márton Demeter, e-mail: mdemeter[at]ucm.es.

Article received on 24 June 2025. Article accepted on 1 August 2025.

Conflict of interest: The authors declare no conflicts of interest.

Funding: The authors received no financial support for the research, authorship, or publication of this article.

Introduction

Generative Artificial Intelligence (AI) technology has rapidly become a transformative force in scholarly publishing, causing excitement and concern among academics (Lund et al., 2024; Yin et al., 2025). The development of AI technologies has led to a significant increase in scientific publications, with China leading in the number of its publications followed by the United States of America (the U.S.) and India (Al-Marzouqi & Arabi, 2024). AI tools are widely used for literature searches, organization of scientific publications, and visualization of research results (Filetti et al., 2024). Specifically, the public release of OpenAI's ChatGPT in late 2022 sparked fierce debates among researchers and publishers about the appropriate role of AI in scientific writing, with issues raised ranging from factual accuracy and hallucinated references to questions of authorship and transparency (Fehér & Demeter, 2025; Lund & Naheem, 2024; Oladokun et al., 2025). Proponents argue that large language models (LLM) can streamline the writing process and even help non-native English-speaking authors to overcome language barriers in academic communication (Hwang et al., 2023). Indeed, early evidence suggests that some academics have begun to use chatbots when preparing manuscripts “to save time or to bolster their command of English”. In an era of publish or perish (Stokel-Walker, 2024) this is an unsurprising trend. At the same time, critics highlight the pitfalls of relying on generative AI – from biased outputs reflecting predominantly Western training data to the risk of further eroding trust in scientific integrity through automated, opaque content generation (Mollema, 2024; Qu & Wang, 2024).

These technological developments are unfolding against a backdrop of persistent global inequality in knowledge production (Demeter, 2020). Scholarly publishing has long been dominated by a few high-income countries, institutions and languages, leading to what some have called a *geopolitics of knowledge* that privileges the Global North while marginalising researchers in the Global South (Demeter, 2020; Demeter & Goyanes, 2021). The advent of generative AI thus raises critical questions: Will AI tools democratise academic writing and help bridge the gap between peripheral and core knowledge producers? Or will they exacerbate existing disparities by giving well-resourced scholars yet another advantage?

Similarly, the evolving landscape of Open Access (OA) publishing – with its promise of free knowledge for readers but through several business models; and with high article processing charges (APCs) for authors – presents a double-edged sword for global equity in scholarship (Demeter & Istratii, 2020). On the one hand, OA could decolonise knowledge by making research universally accessible; on the other hand, costly pay-to-publish models may reinforce the two-tier system of haves and have-nots in publishing (Pooley, 2020). In this article, we examine how generative AI is impacting scientific publishing practices, including open access models, and how these impacts intersect with global inequalities in academic knowledge production.

Theoretical framework: A world-system perspective on academic publishing

Inequalities in scientific publishing can be understood through the lens of world-system analysis, which views global academia as a stratified system of core and peripheral regions (Demeter, 2019; Wallerstein, 2004). Demeter (2019) argues that academic knowledge production operates within a centre–periphery structure that is analogous to the economic hierarchy of the world-system. In this model, a small number of hegemonic “core” countries concentrate the bulk of academic prestige, resources and transnational academic capital, while a multitude of other countries occupy semi-peripheral or peripheral positions with significantly less influence. Crucially, the core’s knowledge is almost automatically treated as having global relevance, whereas research from peripheral regions is often deemed of local or regional interest only. This hierarchy is reinforced by academic institutions and journals: elite universities and well-indexed journals (mostly based in North America and Western Europe) set the standards and gate-keep what counts as international knowledge, often privileging English-language output and Western theoretical frameworks (Demeter, 2020).

Demeter’s (2019) world-systemic model, building on both Wallerstein’s world-systems analysis (Wallerstein, 1990, 2004) and Pierre Bourdieu’s field theory (Bourdieu, 2009, 1998), introduces a three-dimensional perspective to capture both geographic and social stratification in science. He distinguishes between horizontal inequalities (core vs. periphery differences between countries or regions) and vertical inequalities (hierarchies within each country’s academic field, such as elite versus non-elite institutions or researchers). This provides a detailed view in which even peripheral countries have internal elites, and core countries have internal stratifications – yet the global flows of academic capital are skewed toward the centre. A few English-speaking countries occupy the very top tier (the United States of America [U.S.] and the United Kingdom [U.K.] in particular); their national academic systems are in essence the international system’s centre of gravity. Demeter (2019) and others identify several ideal-typical tiers in the hierarchy of global science. Core countries such as the U.S. and the U.K. hold a privileged position in which national research outputs automatically attain international significance. Scholars based in these countries dominate editorial boards and citation networks, and affiliation with the Anglo-American academic sphere effectively grants access to the global scientific elite. Academic capital accumulated in these contexts – such as publications in top journals or prestigious degrees – is widely recognised and legitimised without question.

Semi-peripheral major powers, including France, Germany, Japan and, to a lesser extent, Spain, maintain strong national scholarly traditions that exist alongside the international, Anglophone-dominated academic field. In these countries, researchers often build careers by either producing nationally recognised work in their own language or by engaging with the international research community through English-language publications. However, these two tracks generally remain separate, and achievements within national academic systems do not automatically confer international status. As a result, scholars frequently face an early strategic decision: whether to pursue visibility within the global academic hierarchy or to remain embedded in their national scholarly networks.

In semi-peripheral but small advanced countries such as Switzerland and the Netherlands, international science is the dominant mode of academic practice, and local scholarly circuits play a comparatively minor role. Researchers in these countries typically focus on publishing in English-language international journals to advance their careers. While their national education systems are of high quality, academic capital generated domestically does not easily translate into global recognition. Consequently, many ambitious scholars pursue doctoral training or establish collaborative ties in core countries, especially the U.S., to boost their visibility and credibility on the global stage.

Peripheral countries – comprising the majority of nations, often in economically developing or politically marginalised regions – face profound structural barriers in achieving global academic integration. Research in these contexts is often constrained to local concerns, and attempts to enter high-impact international venues are frequently met with systemic exclusion. While many scholars in these regions aspire to contribute meaningfully to global science, their work is often undervalued or overlooked by dominant academic centres (Wu & Zha, 2018). Talented individuals from peripheral countries often migrate to institutions in the core in search of better infrastructure and professional opportunities, a phenomenon widely known as “brain drain” (Gerhards et al., 2017). This pathway, however, tends to be accessible primarily to those with privileged socioeconomic backgrounds, thereby reinforcing existing inequalities along both geographic and class lines.

This world-systemic theoretical framework makes it clear that global academic publishing is not a level playing field, but rather a stratified hierarchy in which structural power imbalances are continually reproduced. As Immanuel Wallerstein (2004) observed, knowledge production is an integral part of the world-system itself, intertwined with geopolitical and economic dominance. Academic ideas, much like capital, flow disproportionately from the centre to the periphery, and the core retains hegemony in setting research agendas and standards (Galtung, 1971, 1980). Demeter (2020) further argues that this inequitable distribution of what he terms “transnational academic capital” (prestige, citations, funding, etc.) not only disadvantages Global South scholars but impoverishes global science as a whole by impeding the circulation of diverse and innovative scholarship. Biases such as the dominance of English-language journals, Western-centric theories, and the underrepresentation of Global South researchers on editorial boards serve to peripheralise non-Western knowledge (Demeter, 2020). In summary, the world-system approach highlights that any new development in academic publishing – including technological innovations like AI or policy shifts like open access – must be analysed in light of these entrenched centre–periphery dynamics. The question is whether such developments will challenge the stratification of global knowledge, or inadvertently reinforce it?

Generative AI in scientific writing

The recent explosion of interest in generative AI for scientific writing has led to a burgeoning literature on its potential benefits and pitfalls. On the positive side, AI writing assistants (such as GPT-3.5/4, Bard and other large language models) offer novel support tools for researchers during manuscript preparation. Early adopter reports and user studies indicate

that these tools can make scientific writing more efficient and accessible. For instance, AI-driven text generation and editing can help authors overcome writer's block, refine their grammar and style, and even format text according to academic conventions (Liao & Zhang, 2024; Thomas, 2023; van Dis et al., 2023). Perhaps most significantly, generative AI has been hailed as a "fire of Prometheus" for non-native English-speaking researchers, providing on-demand assistance with translation, phrasing and polishing of academic prose (Hwang et al., 2023). By serving as "round-the-clock English tutors" (Hwang et al., 2023, p. 952), large language models can lower language barriers in scholarly communication, allowing researchers to focus more on their domain-specific contributions rather than worrying about idiomatic writing. In a world where publishing in English-language journals is often essential for career advancement, such AI tools have the potential to democratise authorship by empowering scholars whose first language is not English. Indeed, anecdotal evidence suggests that many academics are already using ChatGPT or similar models to edit and refine portions of their manuscripts, finding that it improves clarity and readability when used responsibly (Stokel-Walker, 2024).

However, alongside these benefits come serious concerns. A recurring warning in the literature is that generative AI systems are prone to generating fabricated or inaccurate content if used naively (Oladokun et al., 2025). LLMs do not possess genuine understanding: they can produce text that sounds plausible but may include erroneous facts or entirely hallucinated references (Bik, in Stokel-Walker, 2023). Studies have documented instances of chatbots inventing bibliographic citations for papers that do not exist, which poses obvious risks if such output is incorporated into a submission unnoticed (Thorp, 2023). Even when not outright fabricating, AI-generated text might subtly misrepresent scientific nuance or gloss over uncertainties, potentially misleading readers about the state of evidence (Silverman et al., 2023).

Ethical and transparency issues are another major theme. Because LLMs draw on vast training data (much of it from previously published literature), there are concerns about plagiarism or lack of originality in AI-composed passages (Gupta & Pruthi, 2025; Lee et al., 2023). Researchers must grapple with the questions of how to give proper credit (Should AI tools be acknowledged in authors' notes?) and of what constitutes acceptable use. A consensus is emerging that AI cannot be an author – it lacks accountability and cannot consent – and that human authors must take responsibility for any text generated by AI (Stokel-Walker, 2024; Thorp, 2023). A bibliometric analysis of policies at top journals found virtually unanimous agreement that AI tools must not be listed as authors and that any use of generative AI in the writing process should be fully disclosed in the manuscript (Fehér & Demeter, 2025). In response to these issues, major publishers and editorial bodies have released guidelines: for example, the journal *Science* implemented an explicit ban on text or images produced by AI unless explicitly approved by editors, and requires authors to affirm that no portion of the manuscript was generated by AI. Other publishers (e.g. Elsevier and PLoS) allow AI-assisted writing but mandate detailed disclosure of what tools were used and how, often in a dedicated section of the paper.

Beyond accuracy and ethics, scholars have also voiced concerns about bias and inequality in the use of AI for scientific writing. Generative models like ChatGPT have been shown to reflect the biases of their training data, which is heavily skewed toward English-language and

Western sources. For instance, Qu & Wang (2024) revealed that when ChatGPT simulates public opinion using World Values Survey data, its outputs consistently align with Western (particularly U.S.) norms, even in scenarios centred on non-Western countries, underscoring a persistent bias toward English-speaking, developed societies. This suggests that an LLM may implicitly favour Western perspectives or mainstream scientific paradigms, potentially marginalising diverse viewpoints. If researchers rely on such tools for literature reviews or to frame their arguments, there is a risk of reinforcing the epistemic dominance of the Global North.

Additionally, access to cutting-edge AI may itself become a new differentiator between well-resourced and resource-poor researchers. While models like ChatGPT are widely available now, effective use often requires robust internet access, computational resources for any custom model and digital literacy to know the limits of the tools. There is speculation of an emerging AI divide in which institutions with more funding and training will leverage AI to accelerate research and publications, whereas others may fall still further behind (Ahmed & Wahed, 2020; OECD, 2024).

The evolving role of open access publishing

Parallel to the rise of AI, scholarly communications have been transformed over the past two decades by the growth of OA publishing (Demeter et al., 2021). OA is intended to make research findings freely available to readers everywhere, removing paywall barriers that traditionally limited access to science for those without institutional subscriptions. Recent studies document an ongoing acceleration in OA publishing: by 2023, approximately 38% of global journal articles, reviews and conference papers were published under Gold OA, up from just 11% in 2013; while subscription-only content dropped from 70% to 52% (STM, 2024). The proportion of articles that could be published via Gold OA also swelled from 45% to 79% over the same period (STM, 2024). This expansion has been catalysed by funder mandates – such as Plan S, whose implementers report OA rates of around 80% for funded research alongside a surge in transformative agreements encouraging hybrid journals to transition toward OA (Jahn, 2025). Additionally, many large publishers now offer extensive OA options: for instance, Springer Nature reported that 50% of its primary research articles were OA in 2024 (Lauer et al., 2025) and Cambridge University Press had reached a 63% OA output by 2023, largely due to institutional agreements (Moran, 2024).

From the perspective of global equity, OA is often championed as a force for democratisation of knowledge. When anyone with an internet connection can read scientific articles for free, scholars and students in low-income or peripheral regions – whose universities cannot afford expensive journal subscriptions – ostensibly gain the ability to access the latest research on an equal footing with their peers in wealthier countries. This reader-side benefit of OA has been documented in increased usage and citation of OA articles in developing regions, and it aligns with calls to “decolonise” scholarly knowledge by breaking the monopoly of Western publishers over information dissemination (Mwambari et al., 2022).

However, the literature also provides sobering evidence that the current dominant models of Open Access can exacerbate other inequalities, particularly on the producer side of

research. The most common model for OA in high-profile journals is the author-pays model in which journals charge authors an Article Processing Charge (APC) to cover publication costs instead of charging readers. These APCs can be substantial – often ranging from \$1,500 to over \$5,000 (USD) for a single article in a top journal. Researchers with generous grant funding or those at affluent institutions can usually pay these fees (or their libraries arrange transformative read-and-publish deals to cover them). In contrast, authors from the Global South or less-funded universities may find APCs to be an insurmountable barrier, effectively pricing them out of publishing in the same venues (Smith et al., 2021). A large-scale study by Demeter and Istratii (2020) confirmed a positive correlation between journal prestige (impact factor) and APC levels, noting that many high-impact OA journals charge high fees that researchers from lower-income settings struggle to afford. This raises the concern that OA, while removing paywalls for readers, “lowers barriers to readers only to raise them for authors” (Pooley, 2020). The result can be a two-tier system of publication: well-resourced scholars enjoy seamless OA publication through institutional deals or personal funds, whereas those lacking funding are either forced to publish in lower-tier journals with no/low APCs or not at all (Brainard, 2024). In this way, the APC model can inadvertently mirror and entrench the core–periphery structure: authors from North America and Western Europe make up a disproportionate share of OA articles in prestigious journals, while authors from Africa, Latin America, or South Asia remain underrepresented (Smith et al., 2021).

The literature also discusses mitigating strategies and alternative models to make Open Access more equitable. Many journals offer waiver programs for authors from certain low-income countries, but researchers note that these waivers are not always well-publicised, may exclude upper-middle income countries (where funding gaps still exist), and can also carry a stigma or additional administrative burden. Another model is “Diamond” Open Access, in which neither readers nor authors pay fees – the publication costs are covered by institutions, consortia or governmental subsidies. Diamond OA journals and platforms (such as SciELO in Latin America or journals funded by universities) are highlighted as important outlets that empower researchers in the Global South to publish without charge, although they often operate with limited resources and face sustainability challenges (Demeter et al., 2022). Additionally, preprint servers and repositories provide a route for papers to be freely shared regardless of journal status, which can at least ensure access if not formal recognition. Still, as of 2025, the mainstream of scientific publishing – including most elite journals – has embraced an APC-based Open Access model, which means the resources of the author continue to play a role in who gets to publish openly (Brainard, 2024). The inequities arising from this have prompted calls for a rethinking of the academic reward system: for instance, less emphasis on venue prestige (which is tied to expensive journals) and more on the quality and openness of the work itself, or greater global investment in publication funds and infrastructure in the South (Heuritsch, 2024).

Implications of global epistemic inequality

Bringing together the discussions on generative AI and OA, a central question emerges: are these developments reducing or reinforcing the entrenched inequalities in global knowledge

production? Recent scholarship is mixed in its response. On the one hand, both AI tools and OA publishing are often portrayed as democratising forces. Generative AI has the potential to level the playing field in academic writing by giving every researcher a capable assistant for editing and composition – something that formerly might have required hiring a skilled (and expensive) English editor or having a well-networked mentor. In principle, a researcher in a peripheral context could use a tool like ChatGPT to produce a manuscript that meets international style norms and thereby improve their chances in peer review (Agarwal et al., 2025). Similarly, OA has the potential to amplify voices from the periphery by making their work visible and accessible to all, rather than hidden behind paywalls and accessible only to rich universities. Optimistically, one could imagine a scenario in which a scholar in sub-Saharan Africa uses AI assistance to write a high-quality paper and publishes it in an open-access journal, allowing peers worldwide to read and build on it – a virtuous cycle enhancing the visibility and impact of research from traditionally underrepresented regions.

On the other hand, without deliberate attention to equity, these same trends could perpetuate or even worsen gaps. Generative AI might create new dependencies: if cutting-edge AI tools (or the skills to use them effectively) concentrate in the hands of the core, researchers on the periphery could lag behind in productivity or quality of writing. There is also the risk that AI-generated text homogenises academic writing, pushing it towards a particular normative style (likely Western-centric, given the AI training data), and thereby undermining diversity in scholarly voices. As mentioned above, biases in AI outputs could sideline indigenous perspectives or context-specific knowledge, especially if researchers lean on AI for literature reviews or to frame arguments – the algorithmic filtering might inadvertently preferentially surface mainstream, core publications. Meanwhile, OA in its current form may be skewing the geography of knowledge production. Wealthy institutions in Europe and North America are striking big read-and-publish deals that allow their researchers to publish OA in major journals at no personal cost, while institutions in poorer countries often cannot afford such deals (or APCs directly), leaving their researchers on the outside (Brainard, 2024; Smith et al., 2021). In effect, the global North's dominance might actually increase, with them not only producing the lion's share of publications but also with a higher proportion of those publications being OA (hence garnering more citations and influence); whereas many Global South researchers could remain confined to either less visible national journals or struggling to finance international publication.

Some scholars have begun to use the term *epistemic inequality* or *epistemic injustice* to describe these compounded disadvantages in who can produce and disseminate knowledge (De Sousa Santos, 2016). The integration of AI and the economics of publishing are new layers atop the old patterns of inequity. To ensure that these innovations mitigate rather than amplify inequalities, several proposals surface in the literature: for AI, calls for capacity-building are common – training programs to ensure that researchers globally can use AI tools effectively and ethically; also, development of language models in diverse languages to reduce the Anglophone bias. For publishing, there are calls to subsidise APCs or expand diamond OA models in the Global South, and to reform evaluation metrics so that researchers are not penalised for publishing at local or lower-cost venues. Ultimately, as Demeter (2020) argues, a more inclusive global knowledge ecosystem will require conscious efforts to redistribute academic capital and challenge the notion that excellence is the monopoly of

the centre. Generative AI and Open Access each offer a promise to open up science – but with the realisation that the promise equitability will depend on policy choices, community norms, and possibly new forms of international collaboration to support researchers in underrepresented regions. The coming years will be critical in observing whether these trends lead to a more polycentric world of knowledge production, or whether the “hegemonic countries” will simply use these tools to further consolidate their advantage in the global academic system (Demeter, 2019).

Set against the backdrop of the corresponding literature, this study focuses on the interplay between publication trends, OA trends and GenAI usage with a specific focus on its possible impact on global patterns in academic research production. With this in mind, we address the following research question:

RQ: To what extent do structural geopolitical factors shape the interplay between generative AI usage, academic knowledge production, internationalisation and open access publication models in the global academic system?

Methods

Data collection

LLM usage

Generative AI adoption is accelerating worldwide, but the pace remains highly uneven across regions, sectors and demographic groups. Peer-reviewed surveys show that national uptake already spans a range of two-to-one: in a seven-country study (Australia, Denmark, Germany, Israel, South Korea, Taiwan and the U.S.) the share of adults who had used ChatGPT for science information searches in 2024 varied from 28% in Denmark to 57% in the U.S. (Greussing et al., 2025). Outside the Organization for Economic Co-operation and Development (OECD), adoption is often even faster. India tops global consumer tables, with 73% of respondents reporting GenAI use, while corresponding figures are 45% for the U.S., 49% for Australia and 29% for the U.K. according to the 2024 Salesforce State of the Connected Consumer survey (Salesforce, 2025). China’s Ministry of Industry and Information Technology counted 230 million registered users of home-grown GenAI services by June 2024 – roughly 16% of the national population (Kaiwei & Wenxing, 2024). A similar split is visible at organisational level. The 2024 McKinsey Global AI Survey shows that 65% of firms world-wide now run at least one GenAI use-case in production, up from 33% in 2023 (Singla & Sukharevsky, 2024), while a May 2025 U.S. corporate poll found that 95% of U.S. companies have deployed GenAI somewhere in their workflows (Webb, 2025). Sector-specific studies echo that pattern: German university communication departments reported a sharp jump in use of text-generation tools between 2023 and 2024, especially in private institutions (Henke, 2025).

However, these statistics should be interpreted with caution due to major methodological inconsistencies. Current global data on GenAI use are fragmented and sporadic, gathered

through disparate commercial surveys, industry reports and media investigations, often relying on self-reported behaviour or undefined user bases. There is no centralised, academically vetted platform that systematically collects, compares and updates cross-national GenAI usage statistics. This lack of standardisation across sources, definitions and measurement periods limits the comparability and generalisability of available data. Moreover, many surveys focus on specific platforms – most notably ChatGPT – whose accessibility is not global. In China, for instance, ChatGPT is not officially available, and users instead rely on domestic alternatives such as Ernie Bot or Kimi Chat. As a result, Chinese figures on ChatGPT-centric usage reports are misleading because they fail to capture the scope and nature of GenAI engagement within China's closed AI ecosystem.

The China case

Between 2022 and 2025, China's approach to generative artificial intelligence (GenAI) has been shaped by a comprehensive regulatory framework, a deliberate restriction on foreign AI platforms, and the accelerated development of domestic alternatives. The implementation of the Interim *Administrative Measures for Generative Artificial Intelligence Services* in August 2023 by the Cyberspace Administration of China (CAC) marked a critical regulatory milestone that requires GenAI providers to align outputs with socialist core values, ensure lawful data sourcing and undergo algorithmic security reviews (He, 2023). The 2025 Labeling Measures introduced by multiple government agencies further mandated explicit and implicit labelling of AI-generated content to bolster transparency and traceability (Wang & Yu, 2025).

Access to foreign platforms such as OpenAI's ChatGPT has been systematically blocked in mainland China, Hong Kong and Macau since July 2023, on grounds of data security, technological sovereignty and ideological control (He, 2023). These restrictions have led to a burgeoning domestic GenAI ecosystem, with major technology firms like Baidu, Alibaba, Tencent, and academic institutions such as Tsinghua University introducing models like Ernie Bot, Qwen, Hunyuan and ChatGLM. These models are designed not only to compete technologically with Western counterparts but also to comply with state censorship and regulatory standards (McMorrow & Hu, 2024).

In the academic sector, China has actively promoted the integration of GenAI in scientific research, especially in prioritised domains such as biomedicine and climate modelling. Government-led initiatives such as those at the Chinese Academy of Sciences and Tsinghua University have employed GenAI to accelerate tasks like protein structure prediction (Yang et al., 2025) and weather forecasting (Gao et al., 2025). However, stringent ethical and operational guidelines govern this usage. GenAI tools are prohibited from being credited as co-authors, and their use must be disclosed in research publications. AI-generated content cannot be cited as original literature unless accompanied by an explanatory note (Chinese Academy of Sciences & Cyberspace Administration of China, 2025).

China's regulatory approach has also introduced challenges for international academic collaboration. Data localisation laws, strict content controls and ideological mandates complicate cross-border research workflows and constrain the scope of AI-assisted inquiry, particularly in politically sensitive fields such as public policy or human rights (Liu, 2020;

Silver, 2025). Studies have found that Chinese AI-generated publications often avoid topics that could conflict with state narratives, contributing to self-censorship and ideological filtering (Chen, 2025). Moreover, despite widespread informal use of tools like ChatGPT via VPNs, formal academic outputs typically rely on domestically approved models due to institutional compliance pressures (Li et al., 2025).

Statistical data highlight the scale and strategic depth of China's GenAI ecosystem. By June 2024, the number of users of generative AI services in China had reached approximately 230 million – roughly one in six internet users – according to the China Internet Network Information Center (Kaiwei & Wenxing, 2024). In addition, industry reports indicate that China hosts over 130 distinct large language models, accounting for nearly 40% of the global total (Global Times, 2023). While domestic models continue to improve, they still lag behind leading Western systems in areas like cross-disciplinary reasoning and factual accuracy, posing challenges to global interoperability and academic parity (White & Case LLP, 2025). Nevertheless, China's investment in GenAI infrastructure and talent has positioned it as a formidable parallel AI power, shaping the international norms and ethical standards of AI governance.

Publication trends

We used the Scopus database to check the publication trends because it is considered to be the most inclusive international database for peer reviewed publications (Burnham, 2006; Visser et al., 2020). We worked with Scimago, which reports world statistics on publication trends across various world regions on both country and continent levels. The analysed time period covers the years from 1996 to 2024, starting with the earliest to the most recent years that Scimago reports. To provide a global comparison, we conducted both country-level and region-level analysis with the following categories.

- “Big three”: the single countries with the highest publication record: The U.S., the U.K. and China
- Europe: The European countries with the highest publication record (Germany, France, Spain, Italy and the Netherlands)
- Asia: The Asian countries (other than China) with the highest publication record (India, Japan, Korea, Indonesia and Taiwan)
- LATAM: The Latin American countries with the highest publication record (Brazil, Mexico, Argentina, Chile and Colombia)
- Global: Key regions with the highest publication records (Ibero-America, Western Europe, North America and Asia)

For each category, we analysed the following trends:

- publication records across the analysed time period (number of Scopus-indexed publications)
- international collaboration: the number of papers published in international collaboration
- open access: the number of open access publications

Analysis strategy

To analyse publication trends, we provided longitudinal descriptive statistics that shows trends in the analysed categories mentioned above. To estimate future trends in scientific publication output, we conducted a time series analysis using ARIMA (AutoRegressive Integrated Moving Average) models. ARIMA is a widely used statistical method for forecasting time-dependent data that is capable of capturing autoregressive patterns, differencing to ensure stationarity, and accounting for moving average components (Box & Jenkins, 1976).

Results

“Big three”

Publication

Figure 1 illustrates the longitudinal growth in the number of scientific documents published by the U.S., the U.K. and China between 1996 and 2024. A striking trend emerges: while the U.S. (green line) maintained a consistently high volume of publications, its growth has plateaued since approximately 2016. The U.K. (blue line) shows modest but steady growth throughout the period, with a relatively stable output compared to the other two countries. Most notably, China (orange line) exhibits exponential growth, surpassing both the U.K. by around 2005 and the U.S. by 2020. By 2024, China’s publication output exceeded 1.2 million documents annually – more than double that of the U.S. This rapid expansion reflects China’s strategic prioritisation of scientific output as a national objective and underscores its emerging dominance in the global academic landscape. However, as previous sections have noted, this surge in volume does not necessarily correspond to proportional gains in international collaboration or open access adoption, suggesting a complex decoupling of productivity and epistemic integration.

Figure 2 presents the total number of citations received annually by scientific documents. The trajectory reveals a peak-and-decline pattern across all three countries, with notable temporal and structural distinctions. The U.S. (green line) maintained global dominance in citation impact from the late 1990s through the 2010s, peaking around 2011–2013 with over 25 million citations per year. However, since then, a marked decline is visible, accelerating sharply after 2020. The U.K. (blue line) follows a similar but less pronounced pattern, peaking around 2012 and showing a gradual decline thereafter. China’s citation trajectory (orange line), by contrast, shows steady growth beginning in the early 2000s, peaking just after 2020. However, by 2024, all three countries exhibit a steep drop in total citations.

This recent convergence and collective decline likely reflect both structural and technical shifts: the lag effect of citation accumulation, delayed indexing for recent publications and possible changes in database aggregation methods post-2020. Additionally, China’s ascent in citation volume underscores its growing integration into global citation networks, although the subsequent drop raises questions about long-term visibility and citation lag in newer

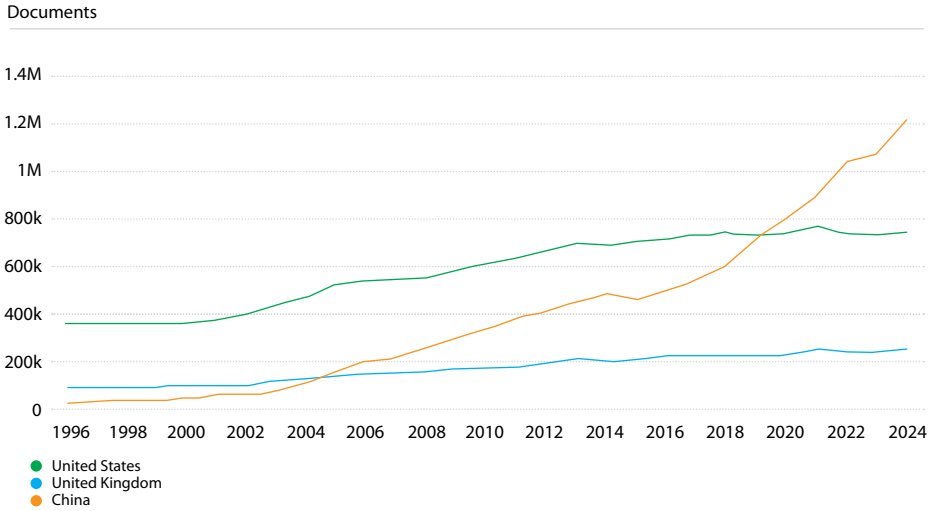


Figure 1:
Publication measures in Scopus for the U.S., the U.K. and China (1996–2024)
 Source: Compiled by the authors

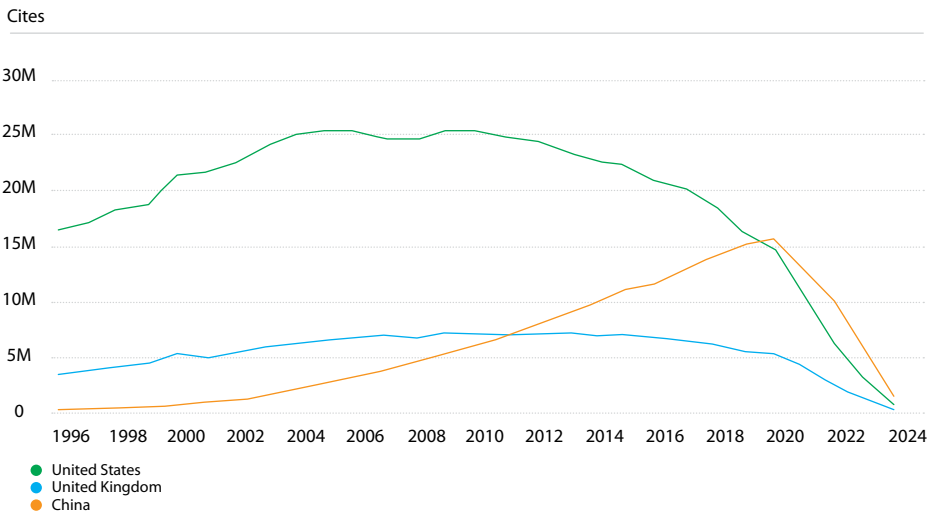


Figure 2:
Citation measures in Scopus for the U.S., the U.K. and China (1996–2024)
 Source: Compiled by the authors

publications. The decline across all countries suggests that citation metrics alone may be insufficient to capture contemporary scientific influence, especially in an era of rapidly changing digital and AI-driven publication practices.

International collaboration

Figure 3 displays trends in international scientific collaboration measured as a percentage of publications co-authored with at least one international partner. The U.K. (blue line) leads consistently and markedly in international collaboration, rising from approximately 27% in the mid-1990s to over 63% by 2024. This trajectory reflects the U.K.'s strong institutional integration into global academic networks and its historical openness to cross-border research initiatives. The U.S. (green line) exhibits a more moderate but steady increase in collaboration, from around 18% in the late 1990s to about 37% by 2024. This trend aligns with the country's growing emphasis on global partnerships, particularly in large-scale scientific infrastructures and consortia, despite its dominant domestic research capacity.

China (orange line), by contrast, shows relatively stagnant collaboration rates, remaining near 20% throughout the period, with minor fluctuations. This suggests that despite China's surge in publication volume and increasing global scientific visibility, its integration into international co-authorship networks remains limited. The divergence highlights a structural asymmetry: while China is ascending in output, its research is still disproportionately domestically oriented, potentially reflecting linguistic, institutional, or geopolitical barriers to broader collaboration.

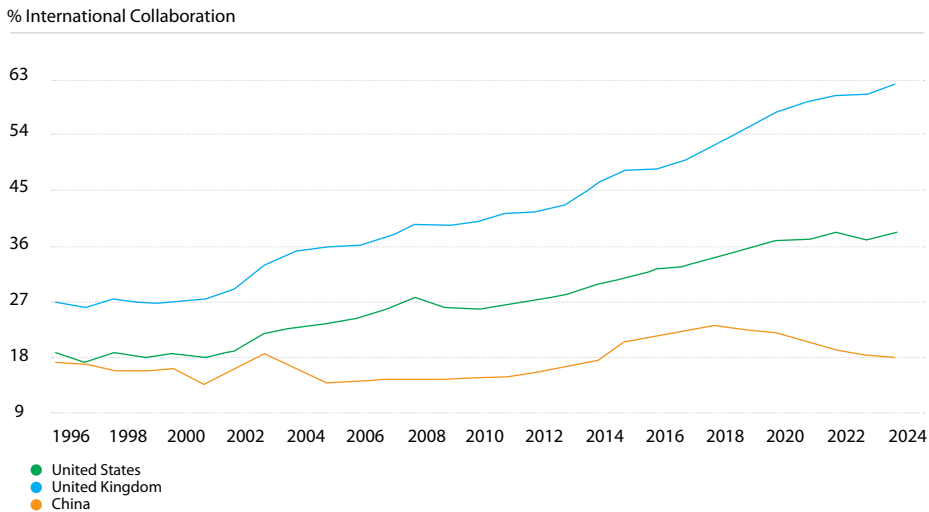


Figure 3:
International collaboration measures in Scopus
for the U.S., the U.K. and China (1996–2024) in percentages

Source: Compiled by the authors

Open access

Figure 4 illustrates the percentage of open access (OA) publications. The U.K. (blue line) demonstrates a remarkable upward trajectory, particularly after 2014, peaking at around 70% in 2022 before a slight decline in 2023–2024. This sharp rise aligns with the introduction of strong national mandates and funding requirements – such as those from U.K. Research and Innovation (UKRI) and the implementation of Plan S – which have made OA publishing a normative standard in the U.K. research ecosystem. The U.S. (green line) shows steady growth in OA adoption, rising from just over 20% in the late 1990s to approximately 55% by 2022, followed by a similar recent decline. Despite lacking a single national OA policy, the U.S.’s increase likely reflects a combination of institutional mandates, public funding conditions and growing participation in hybrid and green OA models across disciplines.

China (orange line) reveals a slower but consistent upward trend, increasing from below 20% in the early 2000s to over 45% by 2022. However, its OA share remains lower than that of the U.K. and the U.S., indicating institutional or systemic constraints. The more modest rise may stem from differing incentives in China’s academic promotion systems, limited alignment with Western OA policy frameworks, and reliance on domestic publication venues less integrated with global OA standards. The post-2022 dip in all three countries may reflect systemic lags in indexing, changes in OA categorisation, or broader disruptions in the publishing ecosystem – possibly influenced by shifting APC models, economic pressures, or the saturation of low-barrier OA outlets.

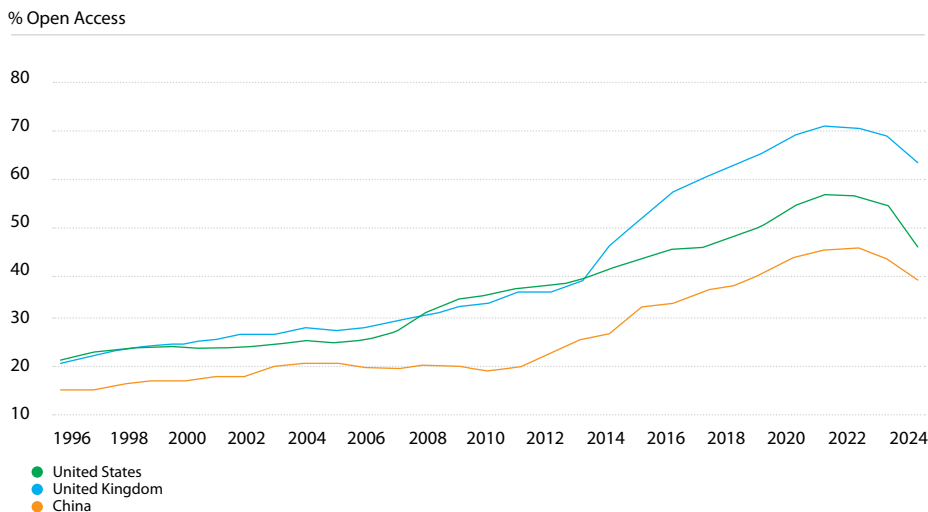


Figure 4:
Open access measures in Scopus
for the U.S., the U.K. and China (1996–2024) in percentages
Source: Compiled by the authors

European Union

Publication

Figure 5 charts the annual number of scientific documents produced by Germany, France, Spain, Italy and the Netherlands between 1996 and 2024. Germany (green line) consistently leads the group, with steady growth in publication output from around 80,000 documents in 1996 to a peak above 220,000 around 2021, followed by a slight decline in more recent years. This plateau may reflect structural saturation or a shift toward quality over quantity in research evaluation policies. France (blue line) and Italy (purple line) follow with relatively similar trajectories until the mid-2010s. Italy then shows a notable surge, overtaking France by 2018 and reaching over 160,000 publications by 2023. This upward shift suggests intensified investment in research or improved indexing practices for Italian outputs. Spain (orange line) demonstrates steady and consistent growth throughout the period, closely converging with France and Italy in recent years. The Netherlands (red line), while exhibiting the lowest total output among the five, maintains a steady growth curve with fewer fluctuations, likely reflecting its smaller research system but sustained productivity relative to population size.

Overall, the figure indicates a general upward trend in scientific output across continental Western Europe, albeit with national variations in scale and growth dynamics. The recent stabilisation or slight decline in some countries' output may signal shifting institutional priorities, funding cycles, or publication model transformations – such as the pivot toward open access or a reorientation around quality metrics.

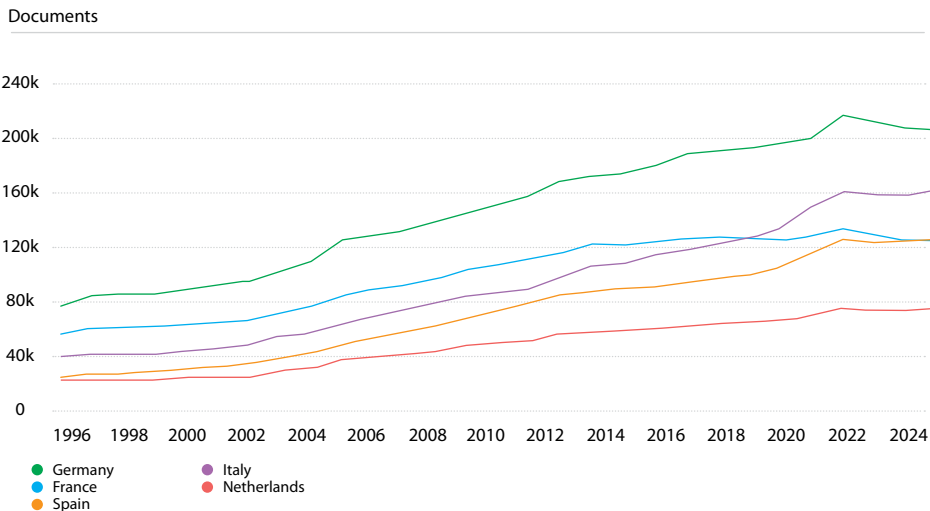


Figure 5:
Publication measures in Scopus
for Germany, France, Spain, Italy and the Netherlands (1996–2024)
 Source: Compiled by the authors

International collaboration

Figure 6 displays the percentage of scientific publications involving international co-authorship. All five countries show a clear upward trend, reflecting increased integration into global research networks over the past three decades. The Netherlands (red line) stands out as the consistent leader in international collaboration, beginning the period at around 33% and rising steadily to exceed 68% by 2024. This exceptional growth underscores the Netherlands’ strong international orientation and its strategic emphasis on open, collaborative science – despite its comparatively smaller research system. France (blue line) and Germany (green line) both follow upward trends, with France reaching around 60% and Germany approximately 53% by 2024. These patterns highlight the central role both countries play in European and global science, benefiting from their position as hubs within large-scale EU research frameworks. Spain (orange line) and Italy (purple line) started with lower collaboration rates, but show robust convergence with the other countries over time. Both reach over 45% by 2024, suggesting increased participation in EU-funded programs and greater cross-border integration.

Collectively, the data reflects a strong European trajectory toward internationalisation, likely driven by structural support through EU initiatives such as Horizon 2020 and Horizon Europe, as well as the expansion of multilingual and cross-national research consortia. The upward trend across all five countries signals a clear policy shift prioritising global research cooperation as a pathway to academic visibility and scientific impact.

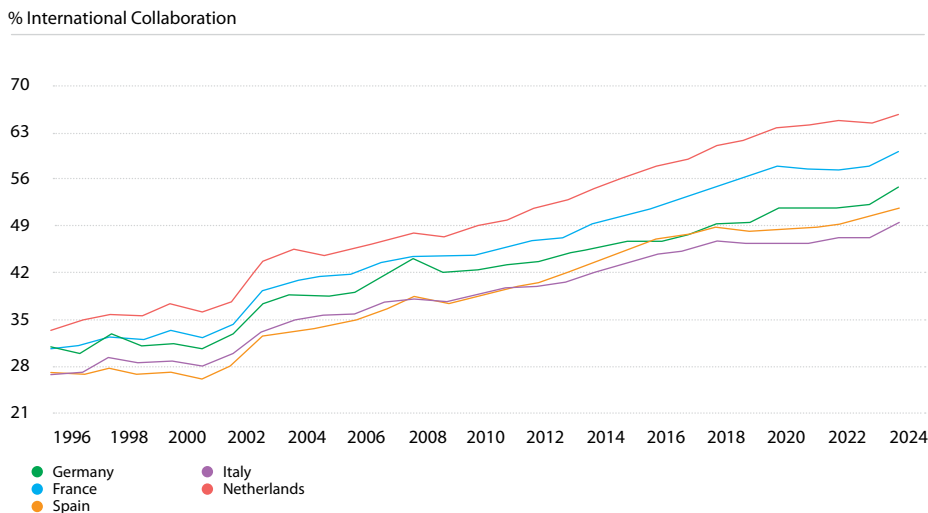


Figure 6:
International collaboration measures in Scopus
for Germany, France, Spain, Italy and the Netherlands (1996–2024) in percentages
 Source: Compiled by the authors

Open access

Figure 7 tracks the percentage of scientific publications made available through OA. A clear collective trajectory emerges: all five countries demonstrate a long-term rise in OA adoption, with particularly sharp increases after 2014 that were probably catalysed by the implementation of national and EU-level OA mandates. The Netherlands (red line) leads consistently, beginning the period at an already elevated level (~30%) and rising steeply after 2015 to a peak above 80% around 2022. This leadership reflects early national investment in OA infrastructure, a culture of policy-driven openness and compliance with initiatives such as Plan S. Spain, France and Italy (orange, blue and purple lines, respectively) show closely aligned trajectories, with all three surpassing the 60% threshold by 2022. These countries benefited from strong national policies and increasing alignment with European Commission OA frameworks. Germany (green line), while also trending upward, appears to have lagged slightly behind the others until recent years, though by 2022 it had reached comparable OA levels (~65%). The uniform drop in OA percentages across all countries in 2023–2024 suggests either a lag in indexing, temporary classification inconsistencies, or a recalibration of what constitutes OA in bibliometric databases.

Overall, the data points to successful regional coordination in advancing OA practices, while also highlighting how national contexts (such as funding mechanisms, institutional mandates and digital infrastructures) affect the pace and sustainability of OA implementation.

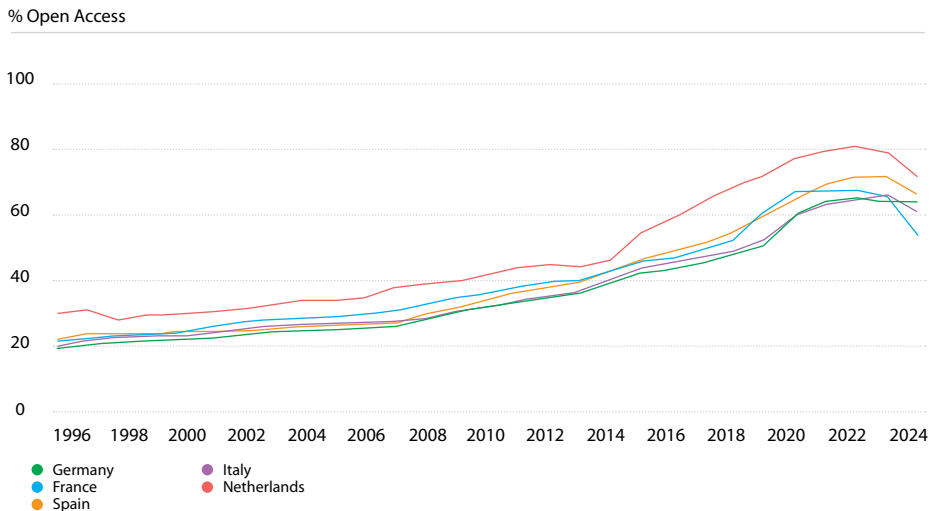


Figure 7:
Open access measures in Scopus
for Germany, France, Spain, Italy and the Netherlands (1996–2024) in percentages
Source: Compiled by the authors

Asia

Publications

Figure 8 illustrates the evolution of scientific publication output for five Asian countries – India, Japan, South Korea, Taiwan and Indonesia – between 1996 and 2024. One key trend is the dramatic rise of India (green line), which moves from approximately 25,000 publications in 1996 to nearly 350,000 by 2024, overtaking all others in the region by a substantial margin. This sharp increase likely reflects India’s expanding research infrastructure, increased investment in STEM education and national incentives for scientific publishing. Japan (blue line), once the clear regional leader, shows a stagnating trend. Its publication output remained relatively flat from 1996 through the mid-2010s, with a slight decline in recent years. This plateau may indicate a mature scientific system shifting focus from quantity to quality, or broader demographic and funding constraints. South Korea (orange line) shows steady growth, climbing from around 20,000 in 1996 to approximately 140,000 in 2024. Taiwan (purple line) and Indonesia (red line) also exhibit upward trajectories, with Taiwan’s growth tapering in the late 2010s and Indonesia demonstrating a steep rise after 2015, reflecting the latter’s increasing participation in global research networks and domestic science policy reforms.

The diverging curves suggest a rebalancing within Asia’s academic publishing landscape, with India emerging as a dominant regional force and Indonesia showing potential for future acceleration. Meanwhile, Japan’s relative decline signals a possible repositioning of leadership within the Asian scientific ecosystem.

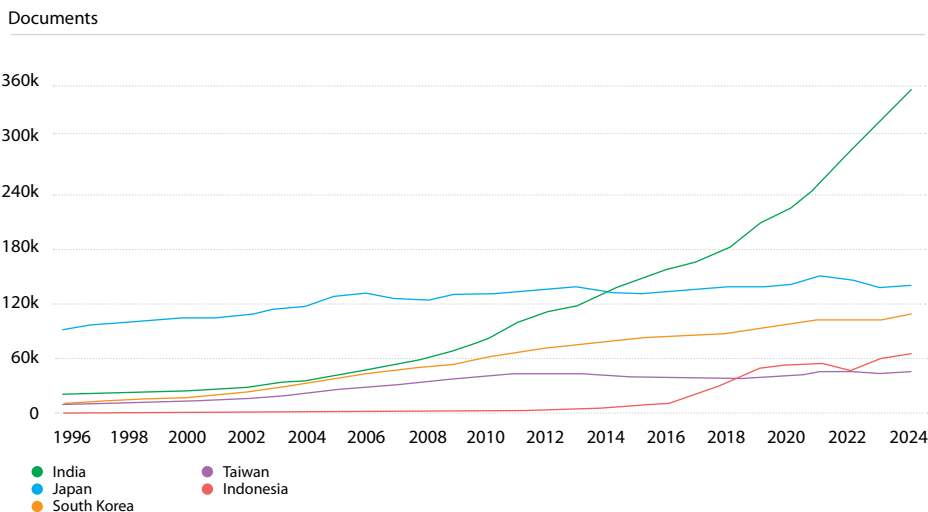


Figure 8:
Publication measures in Scopus
for India, Japan, South Korea, Taiwan and Indonesia (1996–2024)
 Source: Compiled by the authors

International collaboration

Figure 9 tracks the percentage of scientific publications involving international co-authorship. The patterns reveal distinct trajectories in regional integration into global research networks. Indonesia (red line) begins as the clear outlier, maintaining extremely high international collaboration rates (above 65%) until 2010. However, a sharp and sustained decline follows, dropping to below 20% by 2018 before stabilising and slightly rebounding in recent years. This dramatic shift likely reflects the expansion of domestic research capacity and a growing share of nationally driven publications.

Taiwan (purple line) exhibits the most dramatic growth in collaboration, rising from under 20% in the early 2000s to over 45% by 2024. This trend suggests deliberate policy efforts to internationalise its science system, possibly through bilateral agreements, joint research programs and regional partnerships. South Korea (orange line), Japan (blue line) and India (green line) all show relatively modest but steady increases over time. South Korea climbs from 25% to about 36%, while Japan increases from 18% to 31%. India remains the lowest throughout most of the period, although its collaboration rate has improved in recent years, rising from around 15% to over 25% by 2024.

Figure 9 reflects broader asymmetries in how Asian countries are positioning themselves in global science: while some (e.g. Taiwan) are becoming more outward-looking, others (e.g. India and South Korea) are strengthening domestic production capacity first, with international integration proceeding more cautiously. Indonesia's inverted trajectory further demonstrates how growing national capacity can reduce reliance on international partnerships, but also raise concerns about epistemic insularity if not balanced with global engagement.

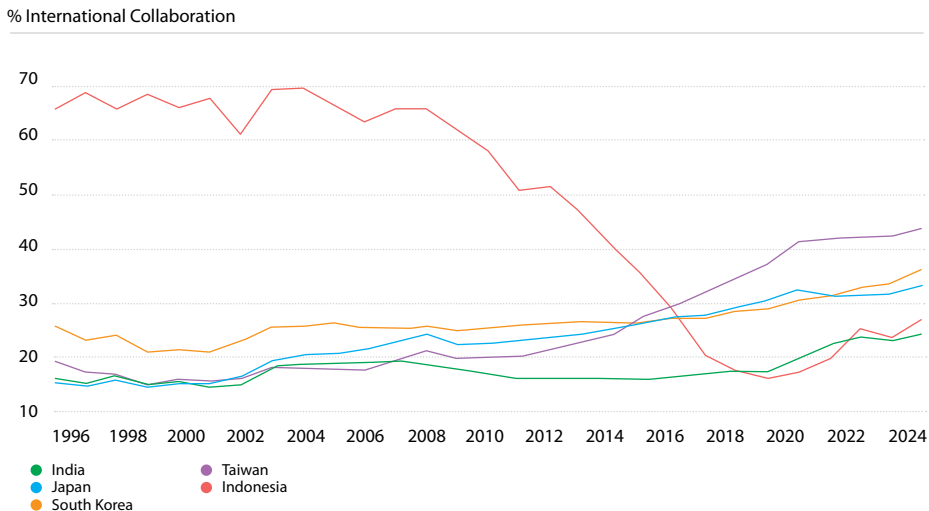


Figure 9:
International collaboration measures in Scopus
for India, Japan, South Korea, Taiwan and Indonesia (1996–2024) in percentages
 Source: Compiled by the authors

Open access

Figure 10 presents the share of scientific publications made available via open access in the countries analysed. Overall, the data reveals a regional trend toward increased OA adoption, although the pace and stability of growth differ markedly across countries. Indonesia (red line) stands out with a sharp rise in OA adoption beginning around 2016, peaking at nearly 80% by 2021. This rapid growth may reflect government-driven OA mandates, national repository expansion and integration with regional indexing platforms. Although the rate dips slightly after 2021, it remains the highest among the five countries. South Korea (orange line), Japan (blue line) and Taiwan (purple line) all exhibit relatively similar trajectories. These countries began with OA shares of around 20%–30% and steadily increased to roughly 55%–60% by 2021–2022, with minor declines in 2023–2024. These patterns suggest a regional convergence in OA policy alignment and journal participation in international OA infrastructures such as the Directory of Open Access Journals (DOAJ) and Scopus-indexed repositories. India (green line), in contrast, lags behind the regional trend. Although its OA share increased gradually until around 2020, it then plateaued and began to decline after 2021, falling to below 30% by 2024. This suggests systemic challenges in implementing large-scale OA practices, possibly related to APC affordability, institutional readiness, or the dominance of national journals not fully integrated into global OA platforms.

In summary, while open access has gained momentum across the Asian region, national contexts – including policy support, infrastructure and funding models – continue to shape the extent and sustainability of OA adoption. Indonesia’s rapid rise and India’s relative stagnation exemplify how divergent trajectories can emerge even within similar economic or regional contexts.

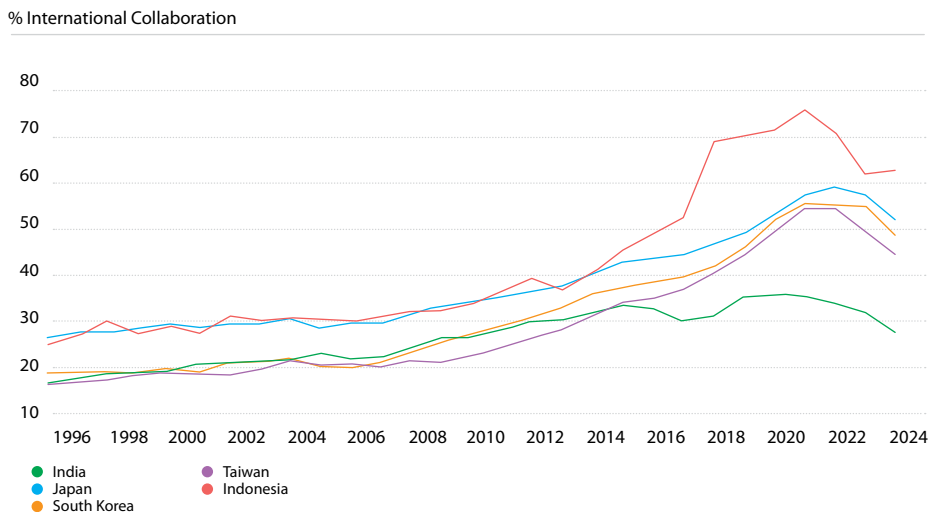


Figure 10:
Open access measures in Scopus
for India, Japan, South Korea, Taiwan and Indonesia (1996–2024) in percentages
 Source: Compiled by the authors

LATAM

Publication

Figure 11 presents the evolution of scientific publication output for Colombia, Brazil, Mexico, Argentina and Chile between 1996 and 2024. Brazil (blue line) emerges as the clear regional leader, with output rising steeply from fewer than 10,000 publications in the late 1990s to a peak of over 110,000 around 2021, followed by a notable decline in more recent years. This trajectory reflects Brazil's strong public investment in higher education and research during the 2000s and 2010s, but also reveals recent stagnation or contraction likely tied to economic and political shifts. Mexico (orange line) shows steady growth over the same period, reaching nearly 40,000 publications by 2024, although the pace has levelled off since around 2020. Argentina (purple line) and Chile (red line) have followed similar patterns, with gradual increases and moderate recent plateaus. Colombia (green line), while still the lowest in total output, has demonstrated strong relative growth, particularly after 2015, signalling expanding national research capacity.

Collectively, Figure 11 illustrates the continued development of scientific systems in Latin America, although at varying rates. Brazil's leadership is clear, but its recent decline raises concerns about the sustainability of its research model. Meanwhile, the upward momentum in countries such as Colombia points to the diversification of regional contributions to global science. These trends underscore the interplay between national science policies, regional funding structures and political stability in shaping long-term publication trajectories.

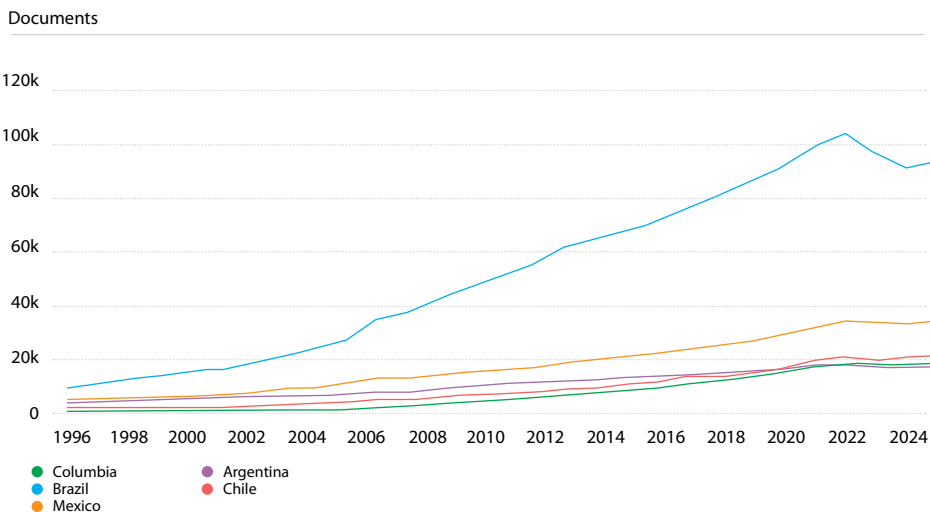


Figure 11:
Publication measures in Scopus
for Colombia, Brazil, Mexico, Argentina and Chile (1996–2024)
 Source: Compiled by the authors

International collaboration

Figure 12 presents the percentage of scientific publications involving international co-authorship. The data reveals both longstanding patterns and emerging shifts in regional research connectivity. Chile (purple line) consistently leads the region in international collaboration, increasing from approximately 43% in 1996 to nearly 65% by 2024. This sustained growth reflects Chile’s strategic emphasis on internationalisation, likely driven by its relatively small research system and the role of international partnerships in enhancing visibility and impact. Colombia (green line) also shows high and generally stable levels of collaboration, fluctuating around 50%–55% for most of the period and rising more clearly after 2016. Argentina (orange line) and Mexico (blue line) exhibit moderate but stable collaboration rates, both hovering between 40%–45% by the end of the period. Their flatter trajectories suggest long-standing institutional connections to global networks but without major recent expansions. Brazil (red line), despite being the region’s publication leader, consistently has the lowest international collaboration rate, starting below 30% in 1996 and gradually increasing to around 38% by 2024. This pattern likely reflects Brazil’s substantial domestic research capacity, which enables high-volume output without proportional reliance on international partners, and the fact that Brazil does not share a common language – Spanish – with the other countries analysed means intra-regional collaboration can be more challenging for Brazilian researchers. However, this also points to a relative insularity that may limit integration into global knowledge flows.

In sum, Figure 12 reveals a mixed landscape: while countries like Chile and Colombia are increasingly outward-facing, others – especially Brazil – still operate with a stronger

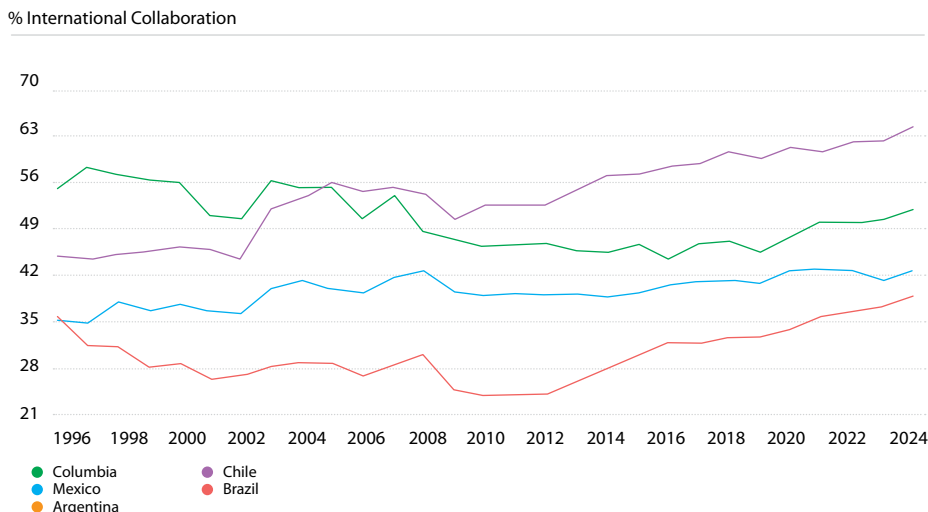


Figure 12:
International collaboration measures in Scopus
for Colombia, Brazil, Mexico, Argentina and Chile (1996–2024) in percentages
 Source: Compiled by the authors

domestic orientation. These disparities illustrate how national science systems, size, language and policy environments shape the nature and depth of international scientific engagement in Latin America.

Open access

Figure 13 presents data that reveals Latin America's strong and early commitment to OA publishing, with all five countries showing significant upward trends. Mexico (orange line) shows the most striking trajectory, rapidly increasing its OA share after 2010 and peaking above 70% around 2017–2019 before experiencing a slight decline in recent years. This surge likely reflects national mandates, the influence of institutional repositories and integration with regional platforms such as RedALyC and SciELO. Chile (purple line) and Argentina (red line) also display consistent growth, both surpassing 60% by 2020. Argentina's early and steady increase reflects long-standing public policies in support of OA, while Chile's trajectory aligns with its broader commitment to internationalisation and digital science infrastructure. Colombia (green line) and Brazil (blue line) both show marked improvement in OA adoption, reaching around 65% and 60% respectively by 2022. Brazil's growth is especially significant given its high volume of publications, underscoring its dual status as both a regional research powerhouse and a major proponent of OA publishing through SciELO.

The slight dip across most countries in 2023–2024 may reflect database reclassifications, temporary reporting lags, or changes in how OA is defined or indexed. Despite this, the overall pattern confirms Latin America's global leadership in open access, driven by regional platforms, public infrastructure and a strong normative commitment to equitable access to knowledge.

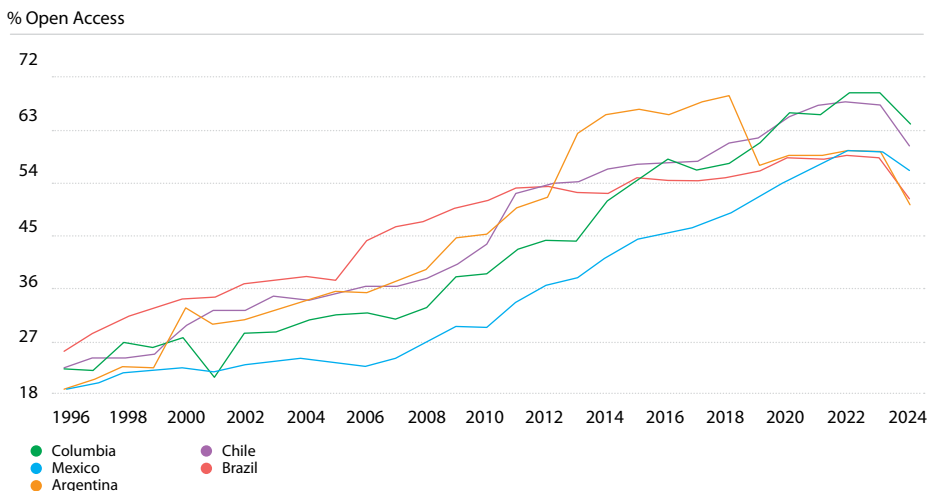


Figure 13:
Open access measures in Scopus
for Colombia, Brazil, Mexico, Argentina and Chile (1996–2024) in percentages
 Source: Compiled by the authors

Global trends

Publications

Figure 14 presents the total number of scientific publications produced annually across four major world regions with the highest research production: Ibero-America, Western Europe, Northern America and the Asiatic Region, between 1996 and 2024. The data reveals both long-standing hierarchies and significant shifts in global knowledge production. The Asiatic Region (purple line) shows the most dramatic transformation, with exponential growth beginning around 2005. Surpassing all other regions by 2014, it continues its steep ascent, reaching over 2 million documents by 2024. This trajectory is largely driven by the rise of China, alongside significant contributions from India, South Korea and other emerging Asian research powers. Western Europe (blue line) and Northern America (orange line) maintain relatively stable trajectories, with Western Europe overtaking Northern America around 2017. These trends reflect mature, high-capacity research systems with incremental growth, likely constrained by demographic and funding ceilings, as well as a growing emphasis on research quality over quantity. Ibero-America (green line), while still producing fewer publications overall, exhibits steady and consistent growth across the entire period. Its progress highlights the expansion of regional scientific capacity, supported by open access infrastructures and national science policies that emphasise global visibility.

The widening gap between the Asiatic Region and other regions of the world suggests a reconfiguration of the global research order. While the Global North continues to lead in infrastructure, influence and institutional prestige, the Global East is asserting dominance in terms of output volume, signalling a potential shift in the balance of global epistemic power.

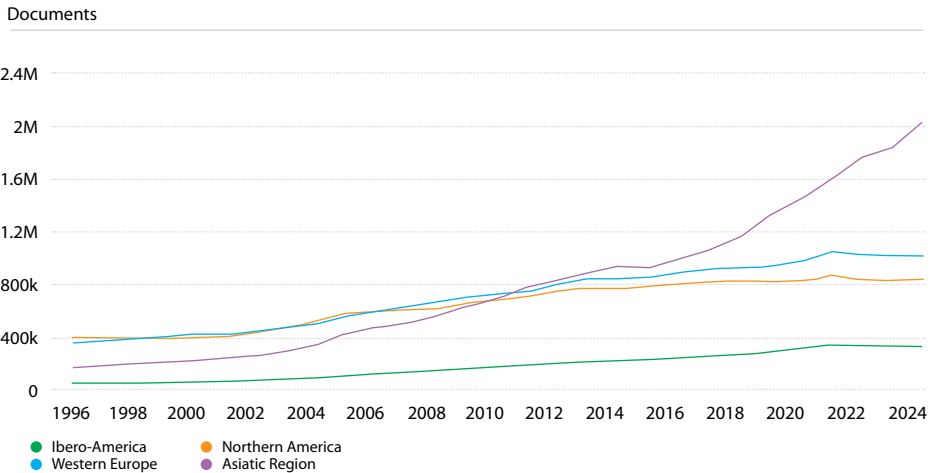


Figure 14:
Publication measures in Scopus
for the four most productive world regions (1996–2024)
Source: Compiled by the authors

Citations

As discussed above, the recent decline in citation counts across regions from 2021 to 2024 is most likely due to structural factors rather than a genuine drop in scientific influence. Citations naturally accumulate over time, so recent publications have had less opportunity to be cited. Additionally, delays in indexing by bibliometric databases and shifts in how citations are tracked can temporarily suppress citation counts. Post-pandemic disruptions, the rise of preprints and non-traditional publishing formats, and the growing use of alternative metrics may also contribute to lower recorded citation activity in the short term.

Figure 15 reflects these dynamics clearly. Northern America and Western Europe, long-standing leaders in citation volume, show steady rises through the early 2000s before peaking around 2014–2016 and then declining sharply after 2020. The Asiatic Region, despite its continuing growth in publication output, also experienced a rapid fall in citation counts post-2020, suggesting a decoupling of volume and impact, at least in the short term. Ibero-America follows a similar pattern to a lesser degree. Together, these trends highlight a systemic temporal lag and underscore the need to interpret recent citation data cautiously, particularly when comparing across regions with different publishing and indexing dynamics.

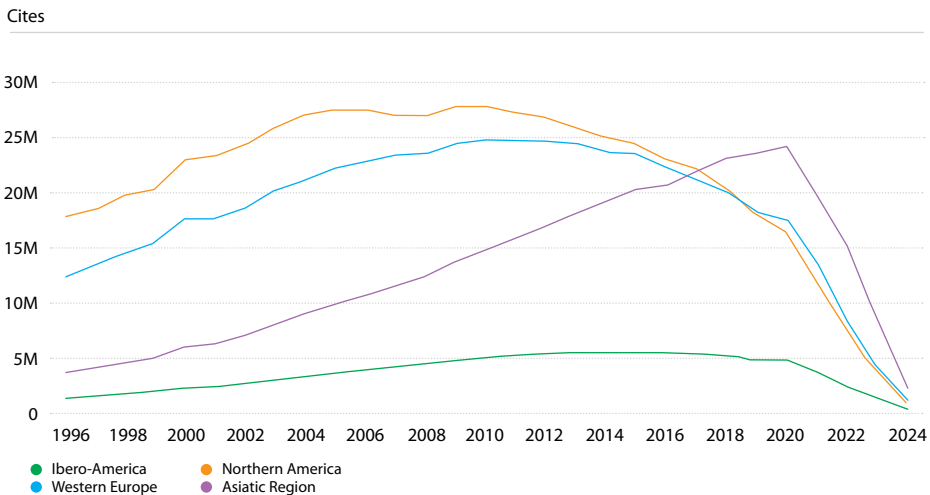


Figure 15:
Citation measures in Scopus
for the four most productive world regions (1996–2024)
 Source: Compiled by the authors

International collaboration

Figure 16 depicts trends in international scientific collaboration measured as a percentage of co-authored publications involving foreign institutions, from 1996 to 2024. Western Europe shows a steady rise from around 25% to nearly 50%, reflecting deep integration within transnational research frameworks such as the European Union’s funding programs. Northern America also increases over time, though more modestly, reaching just over 40% by 2024. Ibero-America follows a similar upward path, approaching 40%, indicating successful regional efforts to globalise research through OA platforms and networked institutions. The Asiatic Region, in contrast, remains significantly lower throughout the period, increasing only from about 15% to 25%. Despite its dramatic rise in publication volume, it continues to operate with limited international engagement, suggesting a more domestically driven research model. The regional gap in collaboration highlights structural asymmetries in global science, where high-output regions may still lag in epistemic integration.

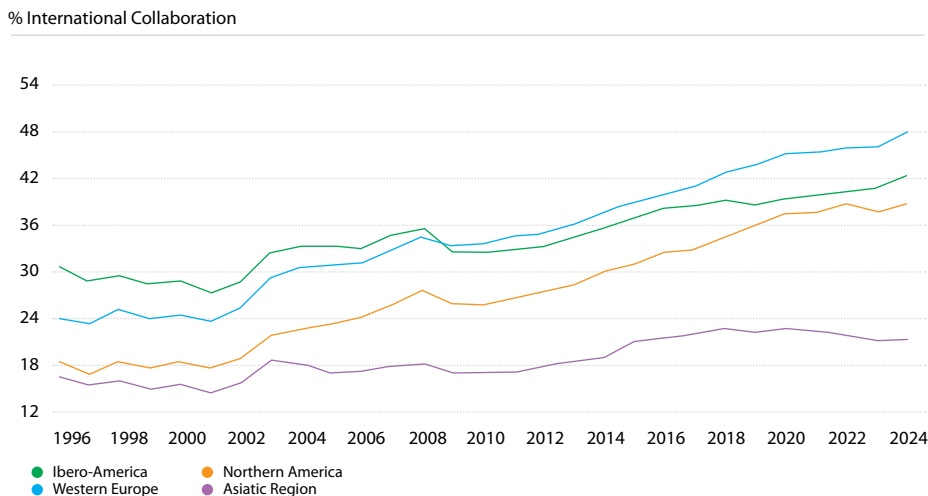


Figure 16:
International collaboration measures in Scopus
for the four most productive world regions (1996–2024) in percentages
 Source: Compiled by the authors

Open access

Figure 17 displays the evolution of open access publishing as a percentage of total scientific output. Over the nearly three-decade period, all regions exhibit a consistent upward trajectory, underscoring a global shift toward more accessible models of scholarly communication. Ibero-America (green line) emerges as an early and persistent leader in OA adoption. Its steady rise, culminating in over 65% OA by 2022, reflects the success of regional platforms such as SciELO and RedALyC, which have institutionalised OA without author-facing fees in line with public and governmental support for democratising knowledge. Western Europe (blue line) shows the most rapid acceleration after 2014, eventually surpassing other regions and peaking near 70% by 2022. This surge aligns with the implementation of Plan S and other policy mandates by European funders that require publicly funded research to be published in OA formats. The trend also reflects the influence of strong institutional OA mandates and comprehensive national repositories. Northern America (orange line), while slower to adopt in the early 2000s, steadily increases its OA share and follows closely behind Western Europe. By 2022, it reached over 60%, driven largely by a combination of institutional repositories, hybrid journal models and increased awareness of the limitations of pay-walled research – especially during the Covid-19 pandemic, which heightened the perceived need for open access to scientific knowledge. The Asiatic Region (purple line) demonstrates a clear upward trend but maintains the lowest OA percentages throughout the period, peaking around 50% in 2022. While countries such as China and India have expanded their OA presence, structural barriers – such as reliance on high-impact pay-walled journals for academic evaluation – continue to constrain full transition to open models.

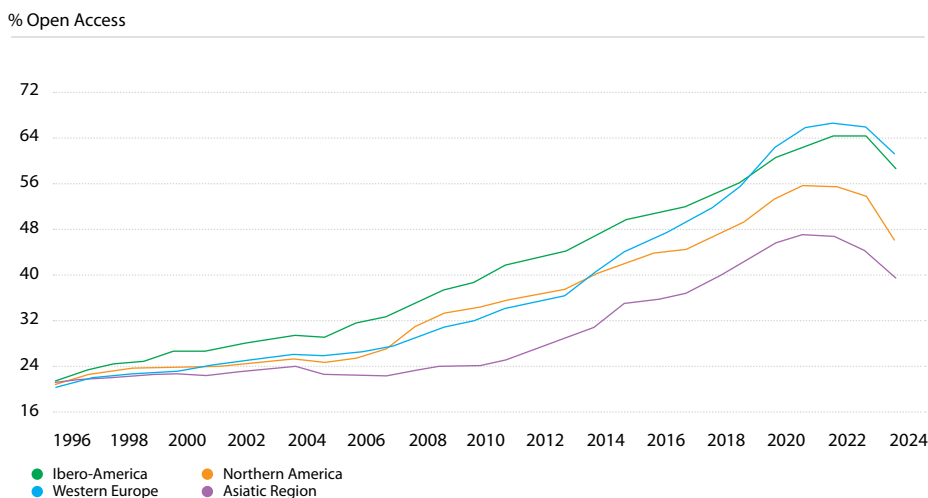


Figure 17:
Open access measures in Scopus
for the four most productive world regions (1996–2024) in percentages
 Source: Compiled by the authors

The slight downturn across all regions after 2022 may be attributed to data indexing delays, shifting OA classification criteria, or changes in the inclusion of repository content in bibliometric databases. Nevertheless, the overall trajectory illustrates a global realignment in the politics of academic publishing in which open access is no longer a fringe movement but an institutionalised norm, albeit unevenly distributed across geopolitical and infrastructural lines.

Trend setting (ARIMA)

“Big three”

To estimate future trends in scientific publication output for China, the U.S. and the U.K., we conducted a time series analysis using ARIMA (AutoRegressive Integrated Moving Average) models. ARIMA is a widely used statistical method for forecasting time-dependent data, capable of capturing autoregressive patterns, differencing to ensure stationarity and accounting for moving average components (Box & Jenkins, 1976; Cicero, 2025). The analysis used longitudinal publication data from 1996 to 2024 ($N = 29$ years per country) obtained from the Scimago database. Because the underlying data exhibited upward trends with potential non-stationarity, first-order differencing ($d = 1$) was applied to each series. After visually inspecting the autocorrelation and partial autocorrelation functions and considering model parsimony, an ARIMA(2,1,0) model was fitted for each country. This configuration implies a second-order autoregressive process on the first-differenced data, without a moving average term. Model fitting and forecasting were implemented in Python 3.11 using the statsmodels library. The models were estimated using maximum likelihood. Forecasts were generated for six future years (2025–2030), and results were visualised alongside historical data to assess growth trajectories.

Model diagnostics indicated a good fit across all three countries. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were low, suggesting model adequacy: China (AIC = 663.51, BIC = 667.50), the U.S. (AIC = 646.63, BIC = 650.63) and the U.K. (AIC = 590.55, BIC = 594.54). Residual autocorrelation was tested using the Ljung–Box Q-test at lag 10. All models returned non-significant p-values (China: $p = .18$; the U.S.: $p = .999$; the U.K.: $p = .996$), indicating no significant autocorrelation and well-specified models.

The forecasts revealed diverging trajectories among the three countries (Figure 18). Publication output in China is projected to increase steeply, reaching over 1.75 million publications by 2030. In contrast, output in the U.S. is expected to remain stable at around 744,000, while the U.K.'s output is projected to slightly increase, stabilising near 254,000 publications annually. These results confirm prior observations of China's rapid expansion in scientific output and highlight continuing structural differences in global academic productivity.

To examine future trajectories in the internationalisation of scientific collaboration, we conducted a time series analysis of international collaboration rates (Figure 19). Here we also employed AutoRegressive Integrated Moving Average (ARIMA) models. Initial exploratory data analysis revealed upward and non-linear trends, justifying the use of first-order differencing ($d = 1$). After assessing autocorrelation functions and balancing

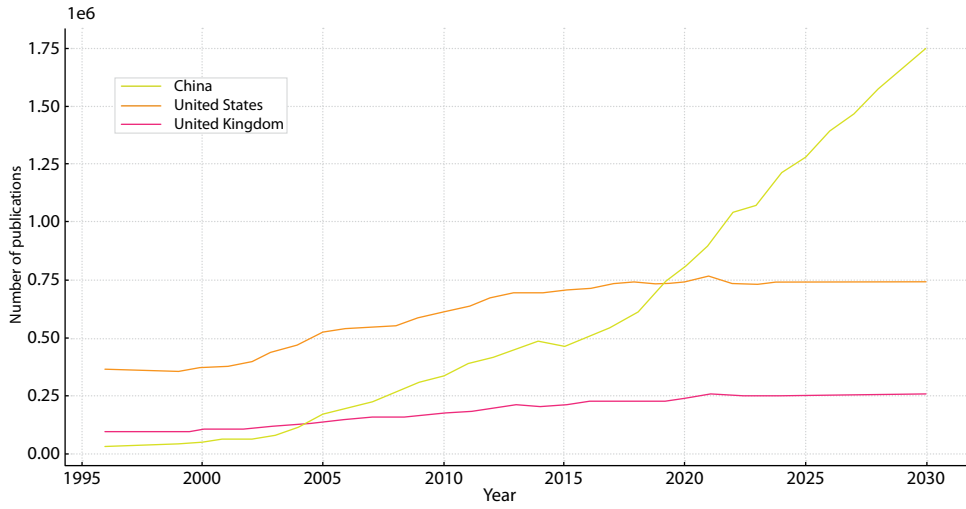


Figure 18:
Research production forecast for China, the U.S. and the U.K.
Source: Compiled by the authors

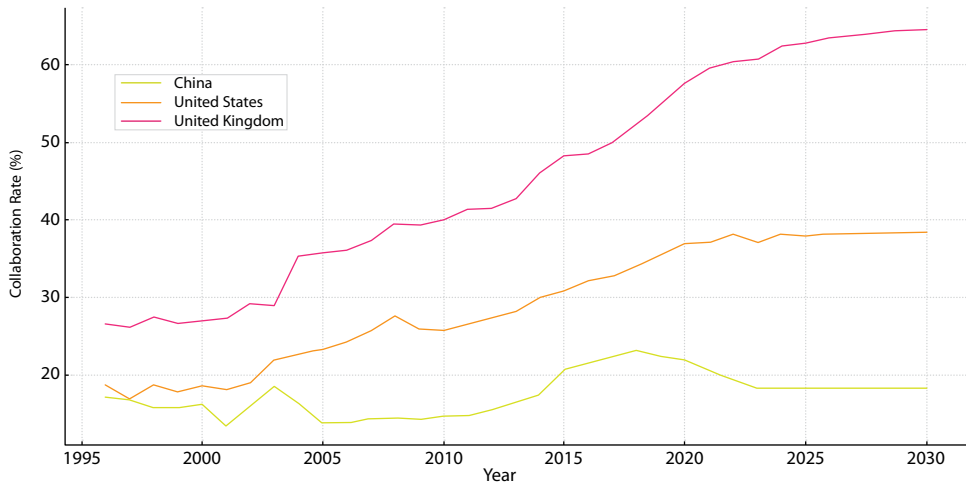


Figure 19:
International collaboration forecast for China, the U.S. and the U.K.
Source: Compiled by the authors

model parsimony with predictive power, an ARIMA(2,1,0) model was specified for each country. This model includes two autoregressive terms and one differencing step, but no moving average term. The analysis was implemented in Python 3.11 using the statsmodels library. Historical and forecasted values were plotted jointly to visualise national trajectories.

Model diagnostics indicated a good fit and specification. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were as follows: China (AIC = 663.51, BIC = 667.50), the U.S. (AIC = 646.63, BIC = 650.63) and the U.K. (AIC = 590.55, BIC = 594.54). Residuals were tested for autocorrelation using the Ljung–Box Q-test at lag 10; all returned non-significant p-values (China: $p = .18$; the U.S. $p = .999$; the U.K.: $p = .996$), suggesting no significant autocorrelation and supporting model adequacy.

The forecasts indicate that collaboration rates will continue to grow, albeit at varying paces. For China, the proportion of internationally co-authored publications is projected to increase gradually, continuing a steady upward trajectory. The U.S. is expected to maintain a relatively stable collaboration rate, with marginal year-to-year variation. In contrast, the U.K.'s already high collaboration rate is projected to rise further, albeit modestly, suggesting consolidation rather than expansion. These results reflect broader geopolitical and institutional dynamics in global science. China's increasing collaboration rates align with its broader international science diplomacy agenda and integration into global research networks. The U.S., while still a central node, appears to have reached a plateau, possibly reflecting internal shifts in funding or policy. The U.K. continues to demonstrate high engagement, likely supported by institutional structures that incentivise international networking, even in the post-Brexit landscape.

To investigate national-level developments in OA publishing, we conducted a time series forecast of OA publication rates in three major scientific producers: China, the U.S. and the U.K. (Figure 20). OA rates were defined as the percentage of annual scholarly outputs published in openly accessible formats. The historical data covered the period 1996 to 2024 ($N = 28$ years), and forecasts were generated through 2030. For consistency with prior modelling strategies we applied AutoRegressive Integrated Moving Average (ARIMA) models.

The forecasted trends reveal divergent national trajectories in OA publishing. The U.K. continues to lead globally, with OA rates projected to exceed 50% by 2030. This is consistent with sustained national-level mandates, funder requirements (e.g. UKRI, Wellcome Trust), and long-term compliance with international initiatives such as Plan S. The U.S. exhibits more modest projected growth. While a substantial share of research in the U.S. is published as open access, the decentralised policy environment and reliance on hybrid journals may limit more dramatic increases in the short term. China, which has shown relatively lower OA adoption compared to its output volume, is forecast to experience continued, if moderate, growth. This aligns with the country's gradual policy alignment with open science norms, as well as increasing international engagement that encourages openness. However, the institutional and infrastructural transition required to support systemic OA may explain the slower pace.

Taken together, these forecasts suggest that while OA adoption is increasing globally, its national trajectories are shaped by distinct policy environments, publishing ecosystems and strategic priorities. ARIMA-based forecasting offers a valuable quantitative tool to

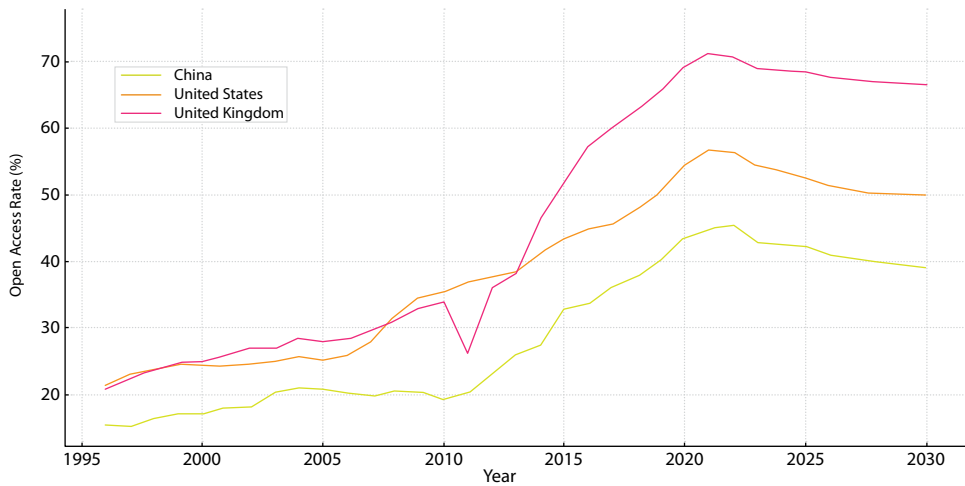


Figure 20:
Open access forecast for China, the U.S. and the U.K.
 Source: Compiled by the authors

anticipate future developments, although its predictive scope is limited with respect to sudden policy shifts or structural innovations in scholarly communication.

Discussion and conclusions

The findings of this study reinforce and complicate the theoretical framework articulated in the world-systems approach to global academic publishing. As Demeter (2019) and Wallerstein (2004) have argued, scientific knowledge production operates within a structurally unequal system in which core countries dominate not merely in volume, but in the recognition, prestige and global visibility of their research. Our analysis of publication trends, open access expansion and the uneven diffusion of generative AI tools suggests that this core-periphery hierarchy is not static, but in flux – subject to both reproduction and transformation through technological, geopolitical and institutional dynamics.

The rise of generative AI in academic writing offers a potent example of this dual character. On the one hand, AI tools such as ChatGPT have begun to lower traditional barriers to participation in global science, particularly for scholars in semi-peripheral and peripheral contexts who face linguistic and stylistic gatekeeping in English-language publishing. These tools promise a degree of epistemic empowerment: an algorithmic assistant that levels the linguistic playing field, accelerates manuscript preparation and supports broader access to publication venues. However, as our theoretical framing reminds us, technological access and capability are not evenly distributed. The most advanced large language models remain concentrated in the hands of core-based corporations and researchers, with training data

that overwhelmingly reflects Global North perspectives, language and epistemologies. Our findings on China's parallel AI ecosystem underscore this bifurcation: while the country boasts massive user bases and a sophisticated state-led AI infrastructure, its models remain circumscribed by regulatory, ideological and infrastructural constraints. The blocking of Western platforms such as ChatGPT, combined with the prioritisation of state-aligned research topics, illustrates how core-periphery dynamics in AI are not simply a matter of technological diffusion but are embedded in national strategies of control and legitimacy.

These geopolitical cleavages are mirrored in our ARIMA-based forecasts of scientific output. China is projected to continue its exponential growth in publication volume, outpacing both the U.S. and the U.K. by 2030. However, this surge in quantity does not translate into equivalent gains in open access adoption or international collaboration. Our data suggests that while Chinese science is becoming more productive, it remains relatively insulated, with slower increases in cross-border co-authorship and more modest open access trajectories. This supports Demeter's (2019) concept of "semi-peripheral hegemony": an actor that dominates in scale but does not yet shape global epistemic norms.

In contrast, the U.K. – although smaller in output – is forecast to retain leadership in international collaboration and OA. This reflects the country's embeddedness in transnational academic networks and its alignment with progressive publication policies such as Plan S. The U.S. appears to occupy a plateaued position, maintaining high output but with less dynamism in collaboration or OA growth, indicative of its mature, yet somewhat inward-facing scientific system. Taken together, these trends suggest a realignment of the academic world-system: while the Global North retains its central role, the Global East – China in particular – is exerting increasing gravitational pull in terms of the sheer scale of production. The question is whether this quantitative dominance will evolve into qualitative influence over what counts as legitimate knowledge.

Our regional analyses further substantiate these observations. The Asiatic Region, led by China, India and emerging Southeast Asian powers, is on track to surpass traditional hubs in publication volume. However, internationalisation and open access adoption remain uneven, raising concerns that these regions may replicate the same inequalities they seek to escape, albeit from a position of growing economic and political strength. Western Europe and Ibero-America, meanwhile, show high and increasing levels of collaboration and OA publishing, suggesting that regional strategies – rooted in shared infrastructure, linguistic networks and policy alignment – can counteract core-periphery asymmetries. Yet even within these regions, the disparities between elite and non-elite institutions, between national and global circuits of knowledge persist.

A key insight emerging from this study is that global academic inequalities are no longer simply geographic but are becoming stratified along technological, institutional and epistemic axes. Access to generative AI tools, the ability to pay article processing charges, and the integration into citation and editorial networks now function as the new currencies of academic capital. This raises the spectre of a multi-layered hierarchy: core countries with core institutions that concentrate not only material resources but also algorithmic and epistemological authority. The result may be a more intricate, yet still unequal, academic order – what we might term *platform stratification* in science, where visibility, credibility and productivity hinge on access to both infrastructures and networks of validation.

And yet, the future need not be one of deterministic stratification. The very tools and practices that risk reinforcing inequalities also contain the seeds of transformation. If global institutions commit to expanding equitable access to AI training and infrastructure, if funders rethink the dominance of APC-based open access in favour of diamond or subsidised models, and if scholarly communities revalue diversity of voice over conformity to Anglophone norms, then a more polycentric and inclusive system of knowledge production becomes imaginable. Regional OA platforms, multilingual LLMs and international collaborations that decentre the Global North can all serve as pathways toward a reconfigured world-system in science.

Indeed, the AI revolution could herald not just a shift in how research is written, but in who writes it, who reads it and who is recognised for it. Just as earlier waves of digitisation opened the gates of scientific communication, the current convergence of AI, open access and global networking holds the potential for a “great unbundling” of prestige and productivity. In this future, recognition would be less tied to institutional affiliation or national origin, and more to contribution, transparency and accessibility. The democratisation of writing and publishing – if harnessed inclusively – could help rewire the circuits of global science, creating a space for peripheral perspectives to shape global paradigms rather than merely echoing them.

At the same time, caution is warranted. Our forecasts are based on extrapolations from historical trends and current trajectories. They cannot account for disruptive events such as geopolitical conflicts, funding crises, or paradigm shifts in evaluation metrics. Moreover, our analysis relies on aggregate indicators – publication volume, collaboration rates, OA percentages – that, while being informative, obscure intra-national disparities and epistemic nuances. We have not disaggregated by discipline, institutional tier, or language of publication – factors that critically shape academic opportunities and exclusions.

In short, the academic world-system remains deeply stratified, but is also dynamically evolving. Technological advances such as generative AI and policy shifts like open access introduce new variables into the equation. Whether these forces will entrench or unsettle the existing hierarchy depends on choices yet to be made by institutions, funders, governments and scholars themselves. The challenge ahead is not merely to predict the future of academic publishing, but to shape it toward equity, inclusiveness and epistemic justice. Our vision is of a global knowledge ecosystem where excellence is not confined to the centre, but diffused across a pluralistic, interconnected scientific community – one in which ideas rise not by geography, but by their capacity to enlighten, challenge and transform.

Limitations and future research directions

This study, while comprehensive in its scope and methodological consistency, is not without limitations. First, the reliance on aggregate national and regional indicators – such as total publication counts, international collaboration rates and open access percentages – may obscure significant intra-national disparities and disciplinary variations. Elite institutions often drive national performance, while peripheral universities within the same country may face vastly different constraints and opportunities.

Second, the forecasts are based on ARIMA models that, by design, extrapolate from historical trends without incorporating exogenous shocks, non-linear transformations, or abrupt policy changes. As such, our projections are probabilistic, not deterministic and should be interpreted as scenario-based illustrations rather than predictive certainties.

Third, while this study touches on the role of generative AI in shaping future publishing dynamics, it does not empirically examine how AI is currently used by researchers across different contexts, nor does it model the potential disruptive effects of platform-specific LLMs on global knowledge hierarchies. Further qualitative and ethnographic research is needed to understand how AI tools are being adopted, adapted, or resisted in everyday academic practices.

Finally, our analysis privileges English-language and internationally indexed data sources, which may underrepresent the contributions of non-Anglophone scholars and alternative knowledge systems. Future research should explore bibliodiversity, epistemic justice and linguistic inclusion in greater depth, especially as open access infrastructures and regional platforms continue to evolve.

Advancing this line of inquiry will require not only methodological refinement but also critical engagement with the politics of knowledge production. Cross-disciplinary collaborations, multimodal data integration and participatory approaches with underrepresented academic communities will be essential to move beyond description toward transformation.

References

- Agarwal, D., Naaman, M. & Vashistha, A. (2025). AI Suggestions Homogenize Writing toward Western Styles and Diminish Cultural Nuances. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). Online: <https://doi.org/10.1145/3706598.3713564>
- Ahmed, N. & Wahed, M. (2020). The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research. *arXiv*. Online: <https://doi.org/10.48550/arXiv.2010.15581>
- Al-Marzouqi, A. H. & Arabi, A. A. (2024). A Comparative Analysis of the Performance of Leading Countries in Conducting Artificial Intelligence Research. *Human Behavior and Emerging Technologies, 1*. Online: <https://doi.org/10.1155/2024/1689353>
- Bourdieu, P. (2009). *Language and Symbolic Power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Polity Press.
- Bourdieu, P. (with Wacquant, L.) (1998). *The State Nobility. Elite Schools in the Field of Power*. Stanford University Press.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis. Forecasting and Control*. Holden-Day.
- Brainard, J. (2024). Open for Business: Authors Are Increasingly Paying to Publish Their Papers Open Access. But Is It Fair or Sustainable? *Science, 385*(6708), 486–489. Online: <https://doi.org/10.1126/science.ads1013>
- Burnham, J. F. (2006). Scopus Database: A Review. *Biomedical Digital Libraries, 3*. Online: <https://doi.org/10.1186/1742-5581-3-1>
- Chen, Y. (2025). The Accuracy and Biases of AI-Based Internet Censorship in China. *Journal of Research in Social Science and Humanities, 4*(2), 27–36. Online: <https://doi.org/10.56397/JRSSH.2025.02.05>

- Chinese Academy of Sciences & Cyberspace Administration of China (2025). Measures for the Labeling of AI-generated and Synthesized Content [方案]. Ministry of Science and Technology.
- Cicero, T. (2025). Forecasting the Scientific Production Volumes of G7 and BRICS Countries in a Comparative Analysis. *Publications*, 13(1). Online: <https://doi.org/10.3390/publications13010006>
- De Sousa Santos, B. (2016). *Epistemologies of the South. Justice against Epistemicide*. Routledge.
- Demeter, M. (2019). The World-Systemic Dynamics of Knowledge Production: The Distribution of Transnational Academic Capital in the Social Sciences. *Journal of World-Systems Research*, 25(1). Online: <https://doi.org/10.5195/jwsr.2019.887>
- Demeter, M. (2020). *Academic Knowledge Production and the Global South. Questioning Inequality and Under-representation*. Springer. Online: <https://doi.org/10.1007/978-3-030-52701-3>
- Demeter, M. & Goyanes M. (2021). A World-Systemic Analysis of Knowledge Production in International Communication and Media Studies: The Epistemic Hierarchy of Research Approaches. *The Journal of International Communication*, 27(1), 38–58. Online: <https://doi.org/10.1080/13216597.2020.1817121>
- Demeter, M., Goyanes, M., Navarro, F., Mihalik, J. & Mellado, C. (2022). Rethinking De-Westernization in Communication Studies: The Ibero-American Movement in International Publishing. *International Journal of Communication*, 16.
- Demeter, M. & Istratii, R. (2020). Scrutinising what Open Access Journals Mean for Global Inequalities. *Publishing Research Quarterly*, 36(4), 505–522. Online: <https://doi.org/10.1007/s12109-020-09771-9>
- Demeter, M., Jele, A. & Major, Z. B. (2021). The International Development of Open Access Publishing: A Comparative Empirical Analysis over Seven World Regions and Nine Academic Disciplines. *Publishing Research Quarterly*, 37(3), 364–383. Online: <https://doi.org/10.1007/s12109-021-09814-9>
- Fehér, K. & Demeter, M. (2025). Generative Knowledge Production Pipeline Driven by Academic Influencers. *arXiv*. Online: <https://doi.org/10.48550/arXiv.2505.24681>
- Filetti, S., Fenza, G. & Gallo, A. (2024). Research Design and Writing of Scholarly Articles: New Artificial Intelligence Tools Available for Researchers. *Endocrine*, 85(3), 1104–1116. Online: <https://doi.org/10.1007/s12020-024-03977-z>
- Galtung, J. (1971). A Structural Theory of Imperialism. *Journal of Peace Research*, 8(2), 81–117.
- Galtung, J. (1980). 'A Structural Theory of Imperialism' – Ten Years Later. *Millennium*, 9(3), 181–196. Online: <https://doi.org/10.1177/03058298800090030201>
- Gao, Y., Wu, H., Shu, R., Dong, H., Xu, F., Chen, R. R., Yan, Y., Wen, Q., Hu, X., Wang, K., Wu, J., Li, Q., Xiong, H. & Huang, X. (2025). OneForecast: A Universal Framework for Global and Regional Weather Forecasting. *arXiv*. Online: <https://arxiv.org/abs/2502.00338v2>
- Gerhards, J., Silke, H. & Carlson, S. (2017). *Social Class and Transnational Human Capital. How Middle and Upper Class Parents Prepare Their Children for Globalization*. Routledge – CRC Press. Online: <https://doi.org/10.4324/9781315313733>
- Global Times (2023, December 24). Four Chinese Generative AI Models Pass Official Assessment, Establishing Compliance Directory for Industry. *Global Times*. Global Times. Online: https://www.globaltimes.cn/page/202312/1304207.shtml?utm_source=chatgpt.com
- Greussing, E., Guenther, L., Baram-Tsabar, A., Dabran-Zivan, S., Jonas, E., Klein-Avraham, I., Taddicken, M., Agergaard, T., Beets, B., Brossard, D., Chakraborty, A., Fage-Butler, A., Huang, C.-J., Kankaria, S., Lo, Y.-Y., Middleton, L., Nielsen, K. H., Riedlinger, M. & Song, H. (2025). Exploring Temporal and Cross-National Patterns: The Use of Generative AI in Science-Related Information Retrieval across Seven Countries. *Journal of Science Communication*, 24(2). Online: <https://doi.org/10.22323/2.24020205>

- Gupta, T. & Pruthi, D. (2025). All That Glitters is Not Novel: Plagiarism in AI Generated Research. *arXiv*. Online: <https://doi.org/10.48550/arXiv.2502.16487>
- He, L. (2023, July 14). China Takes Major Step in Regulating Generative AI Services Like ChatGPT. *CNN Business*. Online: <https://edition.cnn.com/2023/07/14/tech/china-ai-regulation-intl-hnk>
- Henke, J. (2025). The New Normal: The increasing Adoption of Generative AI in University Communication. *Journal of Science Communication*, 24(2). Online: <https://doi.org/10.22323/2.24020207>
- Heuritsch, J. (2024). Towards a Democratic University: A Call for Reflexive Evaluation and a Participative Culture. *arXiv*. Online: <https://doi.org/10.48550/arXiv.2112.08963>
- Hwang, S. I., Lim, J. S., Lee, R. W., Matsui, Y., Iguchi, T., Hiraki, T. & Ahn, H. (2023). Is ChatGPT a “Fire of Prometheus” for Non-Native English-Speaking Researchers in Academic Writing? *Korean Journal of Radiology*, 24(10), 952–959. Online: <https://doi.org/10.3348/kjr.2023.0773>
- Jahn, N. (2025). How Open Are Hybrid Journals Included in Transformative Agreements? *Quantitative Science Studies*, 6, 242–262. Online: https://doi.org/10.1162/qss_a_00348
- Kaiwei, Z. & Wenxing, Z. (2024, December 2). Generative AI Product User Base in China Reaches 230 mln. *People’s Daily Online*. Online: https://en.people.cn/n3/2024/12/02/c90000-20248924.html?utm_source=chatgpt.com
- Lauer, K., Gladun, E. & Laudani, P. (2025, March 18). Springer Nature Forecasts 2025 Revenue Boost Amid Open Access Growth. *Reuters*. Online: <https://www.reuters.com/markets/europe/springer-nature-forecasts-2025-revenue-boost-amid-open-access-growth-2025-03-18/>
- Lee, J., Le, T., Chen, J. & Lee, D. (2023). Do Language Models Plagiarize? *Proceedings of the ACM Web Conference 2023*, (pp. 3637–3647). Online: <https://doi.org/10.1145/3543507.3583199>
- Li, M., Xie, Q., Enkhtur, A., Meng, S., Chen, L., Yamamoto, B. A., Cheng, F. & Murakami, M. (2025). A Framework for Developing University Policies on Generative AI Governance: A Cross-national Comparative Study. *arXiv*. Online: <https://arxiv.org/abs/2504.02636v1>
- Liao, Z. & Zhang, C. (2024). Generative AI Makes for Better Scientific Writing – But Beware the Pitfalls. *Nature*, 631(8021), 505–505. Online: <https://doi.org/10.1038/d41586-024-02319-3>
- Liu, J. (2020). China’s Data Localization. *Chinese Journal of Communication*, 13(1), 84–103. Online: <https://doi.org/10.1080/17544750.2019.1649289>
- Lund, B. D. & Naheem, K. T. (2024). Can ChatGPT Be an Author? A Study of Artificial Intelligence Authorship Policies in Top Academic Journals. *Learned Publishing*, 37(1), 13–21. Online: <https://doi.org/10.1002/leap.1582>
- Lund, B., Lamba, M. & Oh, S. H. (2024). The Impact of AI on Academic Research and Publishing. *arXiv*. Online: <https://doi.org/10.48550/arXiv.2406.06009>
- McMorrow, R. & Hu, T. (2024, July 17). China Deploys Censors to Create Socialist AI. *Financial Times*. Online: <https://www.ft.com/content/10975044-f194-4513-857b-e17491d2a9e9>
- Mollema, W. J. T. (2024). Decolonial AI as Disenclosure. *Open Journal of Social Sciences*, 12(2), 574–603. Online: <https://doi.org/10.4236/jss.2024.122032>
- Moran, R. (2024, July 19). 49 Facts for 49 Decades. *Cambridge University Press & Assessment*. Online: <https://www.cambridge.org/news-and-insights/49-facts-for-49-decades>
- Mwambari, D., Ali, F. A. & Barak, C. (2022). The Impact of Open Access on Knowledge Production, Consumption and Dissemination in Kenya’s Higher Education System. *Third World Quarterly*, 43(6), 1408–1424. Online: <https://doi.org/10.1080/01436597.2022.2056010>
- OECD (2024). Fostering an Inclusive Digital Transformation as AI Spreads among Firms. *OECD Policy Briefs*, 8. Online: <https://doi.org/10.1787/5876200c-en>

- Oladokun, B. D., Enakrire, R. T., Emmanuel, A. K., Ajani, Y. A. & Adetayo, A. J. (2025). Hallucination in Scientific Writing: Exploring Evidence from ChatGPT Versions 3.5 and 4o in Responses to Selected Questions in Librarianship. *Journal of Web Librarianship*, 19(1), 62–92. Online: <https://doi.org/10.1080/19322909.2025.2482093>
- Pooley, J. (2020, February 21). Read-and-Publish Open Access Deals Are Heightening Global Inequalities in Access to Publication. *Impact of Social Sciences. LSE Impact Blog*. Online: <https://blogs.lse.ac.uk/impactofsocialsciences/2020/02/21/read-and-publish-open-access-deals-are-heightening-global-inequalities-in-access-to-publication/>
- Qu, Y. & Wang, J. (2024). Performance and Biases of Large Language Models in Public Opinion Simulation. *Humanities and Social Sciences Communications*, 11(1), 1–13. Online: <https://doi.org/10.1057/s41599-024-03609-x>
- Salesforce (2025). Top Generative AI Statistics for 2025. *Salesforce*. Online: <https://www.salesforce.com/news/stories/generative-ai-statistics/>
- Silver, A. (2025, April 25). China's Data Protection Rules Prompt Pause from Major European Research Funders. *Reuters*. Online: <https://www.reuters.com/sustainability/society-equity/chinas-data-protection-rules-prompt-pause-major-european-research-funders-2025-04-25/>
- Silverman, J. A., Ali, S. A., Rybak, A., van Goudoever, J. B., Leleiko, N. S. & JPGN Editorial Board. (2023). Generative AI: Potential and Pitfalls in Academic Publishing. *Journal of Pediatric Gastroenterology and Nutrition*, 77(5), 573–574. Online: <https://doi.org/10.1097/MPG.0000000000003931>
- Singla, A. & Sukharevsky, A. (2024). *The State of AI in Early 2024*. McKinsey Global Survey. Online: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024?utm_source=chatgpt.com
- Smith, A. C., Merz, L., Borden, J. B., Gulick, C. K., Kshirsagar, A. R. & Bruna, E. M. (2021). Assessing the Effect of Article Processing Charges on the Geographic Diversity of Authors Using Elsevier's "Mirror Journal" System. *Quantitative Science Studies*, 2(4), 1123–1143. Online: https://doi.org/10.1162/qss_a_00157
- STM (2024). *Uptake of Open Access*. STM Association. Online: <https://stm-assoc.org/oa-dashboard/oa-dashboard-2024/uptake-of-open-access/>
- Stokel-Walker, C. (2024, July 1). AI Chatbots Have Thoroughly Infiltrated Scientific Publishing. *Scientific American*. Online: <https://www.scientificamerican.com/article/chatbots-have-thoroughly-infiltrated-scientific-publishing/>
- Thomas, S. P. (2023). Grappling with the Implications of ChatGPT for Researchers, Clinicians, and Educators. *Issues in Mental Health Nursing*, 44(3), 141–142. Online: <https://doi.org/10.1080/01612840.2023.2180982>
- Thorp, H. H. (2023). ChatGPT Is Fun, but Not an Author. *Science*, 379(6630), 313–313. Online: <https://doi.org/10.1126/science.adg7879>
- Van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R. & Bockting, C. L. (2023). ChatGPT: Five Priorities for Research. *Nature*, 614(7947), 224–226. Online: <https://doi.org/10.1038/d41586-023-00288-7>
- Visser, M., van Eck, N. J. & Waltman, L. (2020, May 21). Large-scale Comparison of Bibliographic Data Sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *arXiv*. Online: <https://doi.org/10.48550/arXiv.2005.10732>
- Wallerstein, I. (1990). Culture as the Ideological Battleground of the Modern World-System. *Theory, Culture & Society*, 7(2–3), 31–55. Online: <https://doi.org/10.1177/026327690007002003>
- Wallerstein, I. (2004). *World-Systems Analysis. An Introduction*. Duke University Press. Online: <https://doi.org/10.2307/j.ctv11smzx1>

- Wang, S. P. R. & Yu, A. (2025, June 11). Introduction to the Measures for the Labeling of AI-Generated and Synthesized Content and Review of Industry Practices. *Lexology*. Online: <https://www.lexology.com/library/detail.aspx?g=2f09c52e-cc40-4769-b762-b42a5e0b2c92>
- Webb, M. (2025, May 28). Generative AI Adoption Nearly Universal, Challenges Persist. *Consulting Magazine*. Online: <https://www.consultingmag.com/2025/05/28/survey-generative-ai-adoption-nearly-universal-challenges-persist/>
- White & Case LLP (2025, May 29). AI Watch: Global Regulatory Tracker – China. *White & Case LLP*. Online: <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china>
- Wu, H. & Zha, Q. (2018). A New Typology for Analyzing the Direction of Movement in Higher Education Internationalization. *Journal of Studies in International Education*, 22(3), 259–277. Online: <https://doi.org/10.1177/1028315318762582>
- Yang, J., Cheng, W. X., Zhang, P., Wu, G., Sheng, S. T., Yang, J., Zhao, S., Hu, Q., Ji, W. & Shi, Q. (2025). Conformational Ensembles for Protein Structure Prediction. *Scientific Reports*, 15(1). Online: <https://doi.org/10.1038/s41598-024-84066-z>
- Yin, S., Huang, S., Xue, P., Xu, Z., Lian, Z., Ye, C., Ma, S., Liu, M., Hu, Y., Lu, P. & Li, C. (2025). Generative Artificial Intelligence (GAI) Usage Guidelines for Scholarly Publishing: A Cross-Sectional Study of Medical Journals. *BMC Medicine*, 23(1). Online: <https://doi.org/10.1186/s12916-025-03899-1>

This page intentionally left blank.

From Tools to Teammates: Reimagining Moral Agency and Human–Robot Partnerships in Journalism Practice

Wenxi Wen,^{*} Ashraf Sadat Ahadzadeh,^{**} Kam-Fong Lee^{***}

* Department of Journalism, School of Communication, Xiamen University Malaysia, MALAYSIA

E-mail: wenwenximy[at]163.com

** Department of Journalism, School of Communication, Xiamen University Malaysia, MALAYSIA

E-mail: ashrafsadat.ahadzadeh[at]xmu.edu.my; ahadzadeh1980[at]gmail.com

*** Faculty of Education, Languages, Psychology and Music, SEGi University, MALAYSIA

E-mail: Mtvchris1002[at]gmail.com

Advances in robot design are reshaping human-dominated fields such as journalism. The integration of autonomous, adaptive systems into society increasingly blurs the boundaries between humans and machines. In this evolving landscape, journalism, rooted in human judgment and ethics, now faces unique challenges and opportunities. China, a global leader in robotic innovations, has emerged as a pioneer in robot journalism, further underscoring the importance of understanding human–robot dynamics. Amid these developments, the concept of Artificial Moral Agents has intensified debates about whether machines can possess moral agency, and how they might coexist with humans as collaborators. Central questions arise: How do humans perceive entities that resemble humans but are not alive? Can such artificial beings be ascribed moral agency and integrated into professional domains as colleagues? This study collected data through semi-structured interviews with 15 Chinese students of journalism aspiring to join the profession with an interest in the intersection between technology and journalism. Analysis of their insights yielded six themes: 1. anthropomorphic appearance; 2. human-like cognitive and emotional abilities; 3. the decision-making capacity of robot journalists; 4. robot journalists as incompetent moral agents; 5. erosion of news authenticity and diversity; and

Address for correspondence: Ashraf Sadat Ahadzadeh, e-mail: ashrafsadat.ahadzadeh[at]xmu.edu.my; ahadzadeh1980[at]gmail.com

Article received on 25 December 2025. Article accepted on 13 April 2026.

Conflict of interest: The authors declare no conflicts of interest.

Funding: This study was funded by Xiamen University Malaysia Research Grant (XMUMRF/2022-C9/IART/0014).

6. whether the future of human–robot collaboration is as partners or rivals. This research highlights the need for philosophers and ethicists to revisit the definition of ‘moral agency’ and to develop frameworks to address the ambiguous moral status of artificial agents. By bridging theoretical discourse with empirical insights, this study contributes to ongoing conversations about the integration of artificial entities into journalism and society.

Keywords: robot journalist, moral agency, human–robot cooperation, autonomy, responsible gap

Introduction

Moral agency has long been central to journalism and a valued trait in journalists. Defined as “the capacity of a moral agent to provide messages to audiences”, it carries ethical responsibility and accountability for actions or inaction (Matilainen 2024, p. 235). With data, algorithms, and automation entering journalism, and humanoid robots moving from sci-fi into reality, moral agency is increasingly a question (Mays et al., 2025). In this context, humanoid robots in journalism refer to artificial entities designed to perform roles and tasks that are traditionally carried out by human journalists. These entities may appear in various forms, ranging from software-based systems that automatically generate news stories from collected data to AI-powered anchors and robotic presenters that deliver news directly to audiences (Ayu Karunianingsih et al., 2025; Broussard et al., 2019). Recent developments have demonstrated this shift, for example in the introduction of AI news anchors that can imitate the appearance, voice and movements of human journalists, and report news on digital platforms (Kim & Kim, 2021; Wu et al., 2024). In this paper, *a journalist* is defined as a media professional who gathers, verifies, interprets, produces, and disseminates news and information to the public through various media platforms because while television presenters and reporters do deliver journalistic content on screen, they represent only one role within the broader field of journalism and should not be treated as if interchangeable with the profession of journalism itself. In this context, *robot journalists* refer to AI-driven systems capable of performing tasks traditionally associated with journalism, such as generating news stories, analysing data, or presenting news content through embodied forms such as AI-powered news anchors or robotic television presenters. As robots begin to mirror humans in external and some internal traits, they take on traditionally human roles and hint at future collaboration as equals (Henschel et al., 2021). As the human–machine boundary blurs, key questions arise about whether or not robot journalists can truly embody moral agency; if ethical responsibility has to be redefined; and of how this human–robot partnership will unfold in practice.

These inquiries are pressing given the human tendency to assign decision-making power to artificial agents. In journalism, journalists have a ‘relatively rosy view’ of robot journalism (Kim & Kim, 2018), emphasising potential and risks. Their ambivalent stance may be reinforced by the public’s positive response to news being delivered by AI-generated anchors and a belief that AI and traditional journalism should synergistically enhance the

future of news production (Sun et al., 2024). The equivalent credibility of human-written and automated news suggests readers' acceptance of AI-generated content (Henestrosa et al., 2023; Wang & Huang, 2024), could also prompt journalists to embrace their robot counterparts. Moreover, the notion that the 'flawed human subjectivity' of traditional journalism should be replaced by algorithms perceived as neutral and indispensable could facilitate the acceptance of robot journalism (Carlson, 2018).

The human tendency to anthropomorphise technological artefacts – the attribution of thoughts and feelings to entities without evolutionary ties to humanity (Epley et al., 2007; Roesler et al., 2024) – underscores a key rationale for exploring Artificial Moral Agents (AMAs) (Moor, 2006). These systems are studied to assess whether artificial entities may be regarded as moral agents (MA) capable of ethical decision-making (Graff, 2024). Proponents frame AMAs as autonomous, self-improving and self-regulating (Formosa & Ryan, 2020), yet this challenges the anthropocentric ethos of journalism rooted in human judgment and lived experience (Nguyen & Hekman, 2022). Perceptions of moral agency are tied to the way individuals view themselves and others as morally accountable (Pasupathi & Wainryb, 2010). Rejection of 'algorithmic objectivity' and the valuing of human originality and judgment destabilises the conceptual scaffolding for AMAs (van Dalen, 2026). While much research critiques AI ethics in journalism (Jones & Jones, 2025; Porlezza, 2024), little focuses on the extent to which audiences perceive robot journalism as MA in journalistic roles.

In the realm of traditional journalism, China advanced journalism by introducing during the Fifth World Internet Conference the world's first comprehensive AI news anchor with a human-like appearance and reporting skills (Hua, 2018). This commitment led Xinhua to introduce an upgraded AI news anchor in 2019, named 'Xin Xiaomeng', who subsequently covered major events, such as the China Import and Export Fair and the Spring Festival Gala, as a representative of Xinhua News (Zhang, 2019). These initiatives have garnered widespread approval from both the public and the media (Zhong et al., 2025), advancing China's digital infrastructure goals as outlined in the 18th National Congress (National Development and Reform Commission of the People's Republic of China, 2023). Concurrently, the National Professional Committee on AI Governance released its Ethical Code for New Generation Artificial Intelligence in 2023 (Ministry of Science and Technology of the People's Republic of China, 2021), advocating for ethical AI practices, including privacy protection, security, accountability, and transparency in directing the responsible development and deployment of AI-powered systems (Gao & Yu, 2024). Thus the study of how moral agency is both theoretically and practically negotiated within journalism alongside human–robot journalist collaboration in contexts such as China, where rapid technological advancements aim to emulate human-like cognition (e.g. the world's first AI-powered robot news anchor) is essential.

The current study

The present study targeted Chinese students of journalism, prospective journalists with experience of using robots or AI technologies. Modern, Chinese, Generation Z students have an open-minded attitude towards AI technology tools and actively seek engagement

with them (Jing, 2024), and some have even formed virtual relationships with Chatbot companions as romantic partners (Hua, 2024). Furthermore, the younger generation faces growing pressure to take responsibility for their professional and technological development, juggling human labour and technology (Lyons et al., 2015). This study established three research objectives to guide the investigation.

The first objective was to explore how Chinese students of journalism perceive robot journalists, including their imagined appearance, capabilities, and roles within the news production process. As robotic journalism evolves from simple automated writing systems to more sophisticated artificial agents, understanding how the journalists of the future conceptualise these technologies becomes increasingly important.

The second objective is to investigate how Chinese students of journalism interpret the concept of moral agency in relation to robot journalists. Moral agency generally refers to the capacity of an entity to understand moral principles, make ethical decisions, and be held responsible for its actions (Formosa & Ryan, 2020). Exploring students' views on whether robot journalists could possess such capacities offers insight into how emerging professionals understand the ethical boundaries between human and artificial actors in journalism.

The third objective was to explore how Chinese students of journalism envision the future relationship between human and robot journalists within their professional context. As robotic technologies become more advanced and autonomous, they may shift from being technical tools to collaborative partners in the newsroom, prompting reconsideration of how responsibilities, tasks, and authority are distributed between humans and machines.

In light of these considerations, this study employs semi-structured interviews to address the following research questions:

RQ1: What perceptions do Chinese students of journalism have of robot journalists?

RQ2: How do Chinese students of journalism conceptualise the integration of moral agency into robot journalists, and how might this conceptualisation navigate the complex moral dilemmas inherent in journalistic practice?

RQ3: How do Chinese students of journalism expect the interaction between human and robot journalists to develop in journalistic contexts?

The journey of robot: From inception to innovation

Throughout human history, there have been numerous narratives and experiments exploring artificially created, living and intelligent human-like beings. These creations reflect human efforts to unravel the mysteries of life and intelligence, often drawing on metaphors from the dominant technologies of their eras (Horáková & Kelemen, 2009). In ancient times, legends about robots were prevalent across different civilisations. One of the earliest documented projects for a programmable humanoid robot was conceived by Leonardo da Vinci around 1495. His drawing outlined plans for a mechanical knight capable of standing, waving its arms, and moving its head and jaw (Gasparetto, 2016). Centuries later, the conceptual evolution of robots continued with the coining of the term 'robot', widely attributed to Karel

Čapek (1920) who introduced it in his play *Rossum's Universal Robots*. Derived from the Czech word *robot*, referring to compulsory labour, Čapek used the term to represent symbolised artificial beings created to serve humans, establishing robots in the popular imagination as autonomous entities designed to perform human tasks (Horáková & Kelemen, 2009).

Prior to the 18th century, theoretical concepts of robots far outpaced practical applications due to limitations in mechanical technology (Bogue, 2020). Early robots relied on mechanical power and were often novelties or toys until the Industrial Revolution redirected focus toward industrial needs, marking the end of 'amateur' robots (Bayram & İnce, 2017). The 1950s marked a new era with advancements in information science and technology, notably with the debut of Unimate, the first programmable robot (Moran, 2007). This innovation enabled robots to execute tasks programmed by humans and to replace humans in hazardous environments with commensurate economic advantages (Bayram & İnce, 2017). By 1972, Shake emerged as the first intelligent robot, pioneering the integration of intelligent systems and robots. Throughout the 20th century, the development of robots remained connected to progress in mechanics, computer science and AI (Wang et al., 2018). Scientists envisioned machines capable not only of mechanically replicating human movements but also of autonomous decision-making based on incomplete information and predefined criteria (Doncieux et al., 2022).

The intellectual roots of AI extend deep into human history, as ancient myths across various civilisations featured autonomous machines and artificial beings (Natale & Ballatore, 2020). However, the mathematical and philosophical foundations necessary for modern AI only emerged in the 19th and early 20th centuries. A key figure in this transformation was the British mathematician, Alan Turing. In 1936, Turing proposed the Turing Machine, a theoretical computing device demonstrating that symbolic manipulation could be executed through algorithmic procedures (Haenlein & Kaplan, 2019). This model not only laid the groundwork for modern computers but also suggested that machines might eventually perform tasks associated with human intelligence (Jiang et al., 2022).

Building on this theoretical foundation, Turing later proposed the Turing Test. He argued that if a machine could engage in conversation so effectively that a human could not distinguish it from another person, it should be considered intelligent (Haenlein & Kaplan, 2019). This proposal was revolutionary because it shifted the debate from metaphysical speculation about the soul to observable behaviours and measurable performance (Hoffmann, 2022). Rather than asking if machines could truly think, Turing focused on whether they could convincingly imitate intelligent behaviour (Danziger, 2022). This pragmatic perspective became one of the earliest operational definitions of AI and has continuously influenced contemporary debates on machine cognition (Hoffmann, 2022).

The 1950s and 1960s saw the rapid maturation of these theories into functional research. Scientists began designing programs to tackle tasks previously considered uniquely human. A milestone was the development of the General Problem Solver by Allen Newell, Herbert Simon, and Cliff Shaw, which attempted to replicate human problem-solving strategies using symbolic logic (Haenlein & Kaplan, 2019). Similarly, Joseph Weizenbaum created ELIZA, which was one of the earliest natural language processing programs capable of simulating conversations with human users (Haenlein & Kaplan, 2019). Research on imitating human learning mechanisms also made progress. The development of the neural learning model

inspired the early exploration of artificial neural networks. In 1957, Rosenblatt proposed the Perceptron Model, which became one of the earliest machine learning algorithms capable of recognising patterns (Jiang et al., 2022). These studies laid the theoretical foundation for the subsequent development of machine learning and deep learning.

In recent years, AI, machine learning, and deep learning began a radical advance in mechanical independence. These technologies utilise calculations that enable robots to memorise data, updating their experiences, capability, and flexibility in directing complex errands and circumstances (Licardo et al., 2024). Unlike ordinary rule based humanoid robots for specific tasks, AI-powered robots can learn from experience, adapt to unusual circumstances, and make choices based on their environment and past behaviour (Licardo et al., 2024). This integration locks in robots with updated capabilities, driving their application over businesses such as healthcare, instruction, family assignments, advantage divisions and amusement. Scientists have made efforts to design robot agents able to perform errands that require cognitive capacities such as organising, decision-making, and learning related to human capabilities (Licardo et al., 2024). In showing disdain toward the fact that turning robots into moral agents still poses challenges, the questions of whether and to what extent robots can reflect people in acting as moral agents has captured the imaginations of philosophers and scientists (Cervantes et al., 2020). The moral status of robots competent enough to work autonomously in real-world circumstances for extended periods has been around for a long time (Brožek & Janik, 2019; Gordon & Gunkel, 2022). Can an intelligent robot be held morally responsible for its actions? Even if robots have not yet achieved this level of sophistication, what does the near future hold? Will they become moral agents? These questions beckon us toward exploring them from a philosophical angle, seeking a deeper understanding of what it means for a robot to be a moral agent and how this could reshape human interactions and societal norms.

Moral agency, journalism and robots

Journalists have been regarded as moral agents (MAs) because of their adherence to moral intentions and their navigation of moral dilemmas through editorial judgment in the news production process (Matilainen, 2024). Autonomy is a prerequisite for moral agency, as it grants journalists the freedom and the free choice to make independent editorial decisions while balancing external pressures, such as political and economic influences (Sjøvaag, 2013). While traditional norms emphasise detachment and objectivity, journalists can struggle to separate personal emotions from moral decisions (Matilainen, 2024). Some scholars argue that emotions should be recognised as a crucial component of journalistic ethics, as they can facilitate more human-centred reporting that reconciles professional obligations with moral responsibility (Stupart, 2021). Rather than perceiving emotions as a threat to objectivity, sentiments of responsibility, empathy, and ethical concern can guide journalists in making morally sound decisions (Stupart, 2021).

Decades of effort in expanding moral agency from humans to non-humans led to the term Artificial Moral Agents (AMAs) (Moor, 2006), reinforcing the belief that machines can make ethical decisions and act morally (Formosa & Ryan, 2020; Martinho et al., 2021).

Moor (2006) classifies AMAs as: 1. *implicit MAs* following pre-defined rules without reasoning; 2. *explicit MAs* analysing situations and justifying decisions; and 3. *full MAs* possessing consciousness, autonomy and intentionality. Wallach et al. (2020) also distinguish between *top-down AMAs* (rule-based) and *bottom-up AMAs* (learning-based). Top-down AMAs are too rigid, while bottom-up AMAs risk misgeneralising because they lack the human's 'form of life'. While AMAs lack full moral sensitivity, they may still function in constrained public morality domains where subtle moral distinctions matter less (Graff, 2024).

The extant literature also identifies reasons for extending moral agency beyond humans: 1. *functional equivalence*: if an artificial entity, e.g. robot, is capable of moral reasoning and decision-making akin to humans, it may be regarded as morally agentic (Gudmunsen, 2024); 2. *social and legal recognition*: the legal framework along with social involvement in dealing with it determines their moral agency status (Gordon & Pasvenskiene, 2021); 3. *attribution of responsibility*: the growing autonomous skills and unmonitored operation of robots are increasingly seen as bearers of responsibility, at least in part (Coeckelbergh, 2020); and 4. *anthropomorphism and human perception*: people tend to attribute robots with human-like characteristics such as mind and emotion (Epley et al., 2007). These perspectives align with the functionalist view that asserts moral agency is determined by an entity's observable behaviour and functional capacities rather than its intrinsic nature (Behdadi & Munthe, 2020). In contrast, the ontological view holds that moral agency is fundamentally tied to an entity's intrinsic properties (Véliz, 2021) and robots cannot be MA due to their lack of intrinsic qualities like consciousness (Søvik, 2022).

Despite the ongoing debate on moral agency from functionalist and ontological perspectives, the increasing integration of AI into various fields, including journalism, has intensified discussions about both the advantages and ethical limitations of AI systems. Proponents of the functionalist view argue that moral subjects should not be limited to entities with inherent traits such as consciousness or biological experience (Dung, 2025). If an artificial system can perform tasks similar to moral reasoning, such as evaluating results, applying ethical rules, or generating decisions in accordance with a normative framework, then it can be regarded as functionally equivalent to a moral agent (Behdadi & Munthe, 2020; Gudmunsen, 2024). Observable behaviour and decision-making capacity rather than internal mental states determine whether an entity is able to participate in morally relevant actions; and while this argument reflects the growing sophistication of artificial systems, the ontological critique maintains that functional imitation alone does not constitute genuine moral agency (Véliz, 2021; Søvik, 2022). This tension becomes particularly visible in journalism, where technological efficiency increasingly intersects with ethical responsibility (Molla & Ahsan, 2025).

In practical terms, AI technologies have demonstrated significant instrumental advantages in journalistic practice because AI systems can quickly analyse large amounts of data, identify patterns from complex information environments, and provide support to journalists through automatic fact-checking, background research and trend analysis (Broussard et al., 2019; Kim & Kim, 2021). Moreover, algorithmic tools are capable of producing template-based news stories in data-driven domains such as financial reporting, sports coverage and weather forecasting, where datasets can be translated into standardised narratives (Leppänen et al., 2017).

However, despite these practical benefits, significant limitations remain when evaluating AI systems as potential moral agents. One fundamental limitation concerns the cognitive dimension of moral agency, particularly the absence of intentionality and genuine moral understanding. From an ontological perspective, agency requires internal mental states, such as beliefs, desires and intentions, which guide actions in meaningful ways (Swanepoel & Corks, 2024). AI systems, by contrast, generate outputs through statistical correlations and algorithmic optimisation rather than through intentional reasoning. Although their decisions may resemble human actions, they lack the internal states necessary for genuine moral agency (Sebastián, 2021). Human journalists, in contrast, interpret the ethical significance of their reporting through reflective judgment and contextual awareness (Mellado & Gajardo, 2026). For instance, decisions regarding whether to publish sensitive information often involve balancing competing values, such as the public's right to know and an individual's right to privacy. These dilemmas require normative reasoning grounded in professional ethics and lived experience, which are capacities that cannot be reduced to computational processes (Cervantes et al., 2020).

Another limitation concerns the emotional dimension of ethical decision-making. Journalism often involves reporting on human suffering, injustice and traumatic events, situations that demand empathy and emotional sensitivity. Emotions such as compassion, concern, and responsibility can guide journalists in producing ethically responsible coverage and engaging respectfully with vulnerable individuals (Stupart, 2021). AI, however, lacks genuine emotional experience and embodied perception. While AI systems can analyse sentiment or generate language that mimics empathy, these capabilities represent algorithmic simulations rather than authentic emotional understanding.

The debate extends beyond theoretical frameworks to practical considerations about responsibility and accountability. Moral agents are typically expected to bear responsibility for the consequences of their actions (Santoni de Sio & Mecacci, 2021). In journalism, this responsibility is embedded within professional norms, editorial oversight and public accountability mechanisms (Sun et al., 2024). Journalists who publish harmful or inaccurate information may face reputational damage, professional sanctions, or demands for correction and apology. When AI systems contribute to journalistic production, however, responsibility becomes more diffuse (Molla & Ahsan, 2025). Such a scenario challenges existing frameworks of moral and legal responsibility (Coeckelbergh, 2020). Scholars have warned that attributing agency to AI systems may create 'responsibility gaps', in which neither humans nor machines can be clearly held accountable for harmful outcomes (Santoni de Sio & Mecacci, 2021).

Some scholars posit the concept of *distributed morality*, which recognises that responsibility in socio-technical systems may be shared among human actors and technological agents rather than attributed to a single entity (Floridi, 2016; 2025). Others adopt a more expansive view of agency. List (2021), for example, compares AI with forms of *group agency*, such as in corporations, arguing that both raise similar questions about responsibility, legal status and moral regulation. Similarly, Laukyte (2017) suggests that artificial agents might eventually be recognised as legal actors if they meet certain conditions associated with agency, including rationality and social interaction. Moreover, some scholars argue that AI agency should be understood as existing to degrees rather than as a binary distinction between agents and non-agents (Symons & Abumusab, 2024).

Nevertheless, even these more flexible perspectives acknowledge that artificial systems currently lack many of the intrinsic characteristics associated with human moral agency (Formosa, 2021; Laitinen & Sahlgren, 2021). As Véliz (2021) argues, moral agency is deeply embedded in human social practices, cultural norms and lived experiences. Without such existential grounding, AI decision-making may remain a form of symbolic computation rather than genuine ethical judgment.

Collaboration between journalists and robots

AI-powered robot journalists now influence not only what content users see but also how media is shaped based on audience reactions (Henestrosa et al., 2023; Sun et al., 2024). This shift has fuelled debates about whether machines can truly have agency in journalism, bringing up ethical concerns such as the decontextualisation of news, lack of transparency, and the challenge of dividing responsibility between human journalists and AI systems (Dörr & Hollnbuchner, 2017). The human-machine operational collaboration represents an acceptable solution that combines oversight of computer system automation by humans who lead decision-making processes and align with journalistic goals (Rydenfelt, 2022). In this model, AI assists in workflows but follows human-led moral guidance. This view indicates that while machines show moral agency in partnerships, human involvement remains more dominant. Journalists must supervise AI systems because they must not relinquish their moral responsibility to AI-driven journalists (Rydenfelt, 2022).

Method

This study – which is rooted in the interpretive paradigm that assumes multiple realities and prioritises contextual meaning-making – employed an exploratory qualitative design using basic interpretive qualitative methodology (Krauss, 2005; Merriam, 2002). Basic qualitative research was derived philosophically from constructionism, phenomenology, and symbolic interaction and is used to understand how people interpret, construct and attribute meaning to their experiences. The overall purpose is to understand how people make sense of their lives and their experiences (Merriam, 2009). This approach offered a structured framework to understand how journalism students – prospective journalists – interpret their experiences with robot journalists, their evolving professional landscape, and the key issues such as: 1. perceptions of robot journalists; 2. the potential to ascribe moral agency (including decision-making capacity) to these technologies; and 3. the dynamics of prospective human-robot collegial collaboration.

We employed purposive sampling (criterion sampling) a common method in qualitative research (Nyimbili & Nyimbili, 2024), to select 15 participants who met the following criteria: 1. enrolment onto a journalism programme (undergraduate/post graduate) in China; 2. passionately aspire to a career in journalism; 3. have an express interest in the role of technology in journalism's future; 4. prior experience of engagement with robots or other AI technologies. Sample size justification was guided by data saturation, a principle

recommended by scholars such as Braun and Clarke (2021). Reviewing empirical studies on the sample sizes required to achieve saturation in qualitative research, Hennink and Kaiser (2022) discovered that saturation is often typically reached with between 9–17 interviews or 4–8 focus group discussions. In light of these findings, the sample size of 15 in this study has been deemed appropriate to ensure theoretical saturation.

A set of preliminary screening questions (including items related to knowledge of the current state of the news industry and prior experience with robots or AI) was designed to verify that potential participants met the study criteria. To reach potential participants, Xiaohongshu (also known as the ‘Little Red Book’), one of the most popular platforms among young people in China, was used. Those potential participants who met the criteria received interview invitations and consent forms via WeChat. Upon accepting the invitation, semi-structured interviews were conducted as the next phase of the research.

With regard to ethical considerations, participants were provided with clear information about the research aim, the data collection methods, and the potential benefits and risks before their involvement. Participants’ rights and freedoms were respected, including their ability to withdraw from the study at any stage, the assurance of anonymity to encourage honest responses, and adherence to privacy and confidentiality during data collection, analysis and reporting. These were all taken into account as ethical measures. Data collection was conducted from May to August 2024 via the Tencent Meeting conferencing tool. The interview session duration ranged from 28 to 51 minutes during which participants received 20 RMB (an equivalent of 3 USD) as a token of appreciation.

The data was collected in Chinese by a Chinese researcher who was fluent in both Chinese and English. Key insights from the interviews were translated into English before back-translation was performed to ensure optimal accuracy. The researcher used bilingual Chinese–English speakers to verify translation accuracy. All data were stored securely through password-protected devices, accessible only to authorised team members. Both physical and electronic records were securely archived for five years, with digital encryption to ensure confidentiality.

This study employed reflexive thematic analysis (RTA) as the analysis method. RTA functions as a methodologically accessible theoretical framework that facilitates researchers in identifying and analysing patterns or themes within a dataset (Braun & Clarke, 2019). Researcher’s interpretations of patterns of meaning spread across the dataset are represented by codes. The method involves analysing the intersection of the dataset with theoretical assumptions of the study, and the analytical skills of the researcher (Braun & Clarke, 2019). The strengths of RTA lie in its accessibility, transparency and flexibility, which permits researchers to develop themes throughout the analysis process (Byrne, 2022). The adaptability and flexibility of RTA match the objectives of this study. An inductive approach was employed for the data analysis in this study, which involves coding data without requiring any predefined frameworks or adherence to the researcher’s preconceptions (Naeem et al., 2023). The data emerges naturally from its source as an organic process without necessitating strict adherence to participant questions (Naeem et al., 2023). The study applies inductive analysis effectively as researchers move from specific observations to broader generalisations (Byrne, 2022) to understand the public’s perception of robot journalists.

Findings

This study conducted semi-structured interviews with 15 Chinese students of journalism who aspired to becoming journalists and expressed an interest in the integration of technology and journalism. Some of them had previous internship experiences at newspapers. Table 1 presents details from the interviewees’ profiles. The research questions focused on three main research areas that investigate the moral awareness and ethical issues of Chinese students of journalism and their expectations of robot journalists. Transcript analyses uncovered six main patterns that generated more complete knowledge about how living journalists differ from robot journalists.

*Table 1:
Interviewees’ profile*

Interviewee	Nickname	Age	Gender	Major	Degree (Undergraduate (UG) / Postgraduate (PG))	University	Internship Experience
1	Xinhao	23	Female	Media & Communication	PG	Hankuk University of Foreign Studies	Newspaper outlet
2	Yiling	22	Female	Journalism & Communication	UG	Shenzhen University	Newspaper outlet
3	Ning	23	Male	Journalism & Communication	PG	Jiangxi University of Finance and Economics	None
4	Miaomiao	22	Female	Journalism	UG	Ningbo University	None
5	Aurelia	22	Female	Journalism	UG	Ningbo Tech University	None
6	Dong	24	Male	Journalism	PG	Anhui University	Newspaper outlet
7	Haiyang	23	Male	Journalism	PG	Henan Normal University	Newspaper outlet
8	Wanmei	22	Female	Journalism	UG	Anhui Xinhua University	Newspaper outlet
9	Qinyuan	22	Female	Media & Communication	PG	University of Nottingham Ningbo China	None
10	Ruixue	23	Male	Journalism	PG	Anhui University	None

Interviewee	Nickname	Age	Gender	Major	Degree (Undergraduate (UG) / Postgraduate (PG))	University	Internship Experience
11	Maomao	24	Female	Journalism & Communication	PG	East China University of Political Science and Law	Newspaper outlet
12	Xiaoman	23	Female	Journalism	PG	Shandong University	In newspaper
13	Congcong	22	Female	Journalism & Communication	PG	Central South University	In newspaper
14	Sue	24	Female	Journalism	PG	Beijing Normal University at Zhuhai	In PR
15	Zhijia	22	Male	Journalism	PG	Shanxi University	None

Source: Compiled by the authors.

RQ1: What perceptions do Chinese students of journalism have of robot journalists?

In exploring the perceptions of Chinese students of journalism regarding robot journalists, the findings reveal two primary themes: anthropomorphic appearance and human-like capabilities. Each of them is accompanied by multiple sub-themes, as illustrated in Table 2.

*Table 2:
Theme, sub-themes and the initial codes
for the perception of journalism students on robotic journalist*

Theme	Sub-theme	Initial Code
Anthropomorphic appearance		'human-like'; 'a gentle smile'; 'an elegant female figure'
Human-like abilities	Cognitive ability Emotional ability	'as sharp as an eagle's'; 'as sensitive as a hunting dog's'; 'display sadness'

Source: Compiled by the authors.

Participants generally expressed expectations that align with findings commonly observed in earlier studies: that robots should resemble humans in form, as presented in Theme 1 below.

Theme 1: Anthropomorphic appearance

The theme emphasising the anthropomorphic appearance of robot journalists reflects the beliefs of Chinese students of journalism that these robots should exhibit human-like attributes. Most of the interviewees expressed the idea that robot journalists should appear 'human-like', featuring nonverbal cues such as facial expressions and gestures, like 'a gentle smile'. They also accentuated the importance of attire, suggesting either 'an elegant female figure' or 'a male figure in a suit and tie'. However, some interviewees found overly human-like designs 'creepy' or 'uncomfortable'.

However, for robot journalists, given the distinctive characteristics of this profession, participants placed particular emphasis on human-like cognitive and emotional capabilities, thereby offering new insights that enrich our findings.

Theme 2: Human-like abilities

The findings indicated that robot journalists exhibit human-like capabilities essential for journalism. These capabilities encompass cognitive and emotional traits.

Cognitive ability

Cognitive abilities entail a logical reasoning mechanism. Our interviewees stated that robot journalists should be able to employ some mechanism of logic and reasoning, as this would enable them to respond promptly to their surroundings. Such capabilities would allow them to identify relevant news stories while gathering accurate information. Robot journalists must be able to recognise social cues and latent information. Our respondents also emphasised that perspective-taking emerges as an important capability for robot journalists, like communication and smooth interaction, highlighting the necessity of understanding multiple dimensions of aptitude of journalism when interpreting individuals' intentions and actions.

Emotional ability

Sensory abilities were also identified as essential attributes for robot journalists. This imperative was conveyed through expressions such as having eyes 'as sharp as an eagle's' and a nose 'as sensitive as a hunting dog's', capturing a comprehensive view on robots as an event and effective narrator. Emotional sensitivity and understanding were also highlighted as prerequisites, as evidenced by their expressions, such as showing a 'tight face' when feeling 'nervous' or 'stressed' while delivering a talk. All interviewees concurred that robot journalists are expected to demonstrate a certain degree of empathy, such as the capacity to 'display

sadness' when encountering 'bereaved families weeping in sorrow', to 'sense' the pain and express understanding and being able to 'feel [others'] anger'. Some interviewees described synthetic empathy generated through programming as 'fake empathy', as opposed to genuine understanding that derived from the heartfelt feelings. Other interviewees were even more cynical, describing this programmed empathy as akin to 'mathematical' problem-solving, where emotions are mechanically imitated without inner authenticity.

RQ2: How do Chinese students of journalism conceptualise the integration of moral agency into robot journalists, and how might this conceptualisation navigate the complex moral dilemmas inherent in journalistic practice?

When analysing the perceptions of Chinese students of journalism concerning the moral agency of robot journalists, the primary themes indicate that robot journalists possess decision-making capabilities but are fundamentally flawed as moral agents (MAs). The main themes and sub-themes, along with their initial codes, are presented in Table 3.

*Table 3:
Theme, sub-themes and initial codes for perceptions
of Chinese students of journalism on moral agency on robotic journalists*

Theme	Sub-theme	Initial code
Decision-making capacity of robotic journalists	Predetermined logic vs. machine learning in decision-making	'complex'; 'unforeseeable'; 'flexibility'
	Utilitarianism in robotic decision-making	'so narrow-minded'; 'free of human bias'; 'references from its database'
	Complexity of human morality	'nerds'; 'facts accurately'; 'lack the warmth of human care'
Robotic journalists as incompetent moral agents	Tool-like nature of robotic journalists	'manipulated'; 'free of will'; 'puppets'
	Accountability dilemma for robotic journalists	'meaningless'; 'malicious intentions'; 'scapegoats'
	Possible solutions to accountability dilemma	'one-to-one accountability'; 'proves false or harmful'; 'facilitate subsequent accountability'

Source: Compiled by the authors.

Participants' perceptions of the moral agency of robot journalists generally align with the prevailing understanding of AI systems as data-driven, tool-like entities. However, beyond these general perceptions, this study further reveals additional multi-faceted dimensions to how people understand moral decision-making and accountability, particularly within the journalistic context in which moral dilemmas frequently arise and human judgment is highly valued. The findings highlight the deeper complexity involved in understanding

and managing moral responsibility in the context of robotic journalism. In other words, the findings reveal that, within specific professional contexts, people's understanding of the moral responsibility of AI and AI-powered entities is moving beyond traditional tool-oriented frameworks, giving rise to more diverse interpretations.

Theme 1: *Decision-making capacity of robot journalists*

The results indicated that robot journalists do have decision-making abilities. Nevertheless, they possess predetermined logic and have experienced machine learning in their decision-making processes. It was also discovered that the robot journalist utilised utilitarianism in decision-making as a result of the complexity of human morality.

Predetermined logic vs. machine learning in decision-making

Interviewees generally attributed the decision-making abilities of robot journalists to data-driven, programmed settings and machine learning; which was evident in verbatim quotes as highlighted by Sue: "As long as we give them the standards and principles on how to make proper news judgments, they can handle procedural decisions like choosing newsworthy stories effectively;" and Wanmei stated: "Like a journalism student, it learns and memorises all the rules and knowledge, but in a way that is much faster than a human, inputting data and setting up programmes." However, the downsides of these pre-set decision-making methods raise a red flag to some of the interviewees, as some situations are 'complex' and 'unforeseeable', and the lack of references in the previous training model, combined with a lack of 'flexibility', makes it difficult to make 'appropriate judgments'. On a more optimistic and lenient note, robot journalists were evaluated as capable of learning on their own, helping them make more sound decisions through 'advanced data analysis and calculations', as expressed by Miaomiao and by Xinhao: "...catching the patterns, and then working out better solutions through computation." One interviewee, Ning stated: "With a lot of practice and 'reflection' – if this word can be applied to robots – they become smarter and learn to make better decisions."

Utilitarianism in robot decision-making

Some interviewees argued that robot journalists maintain ethical standards better than human journalists because they follow programmed principles while having no capacity for emotion. These mind-sets appeared concertedly in excerpts such as robots "aren't tempted by anything and they aren't scared of threats", "free of personal values or thoughts of their own", "massive size of the data help robots not be 'so narrow-minded'", and "free of human bias". Relying on their strong computational abilities and vast access to data, robot journalists

demonstrate objective reporting. They should be able to discover beneficial solutions as suggested by Congcong: “When we try to solve a moral issue, we think about the good and bad sides. Robots can do this more accurately because they make the best choice to maximise benefits.” Since they get “references from its database” “should surely be able to make more forward-looking judgments”.

Complexity of human morality

Several interviewees raised concerns about the ethical judgment of robot journalists. They argued that ethical decision-making cannot simply be reduced to computation due to the intricacy of ethical dilemmas. Interviewees highlighted that the ‘best’ solution calculated by the robot may not necessarily be equivalent to that of a human. As highlighted by Zhihua: “If it were a robot journalist, I guess it would keep reporting the truth.” Another interviewee believed that robot journalists are unable to make good decisions. Besides, Dong indicated they were like ‘nerds’, one who “learns mechanically, rigidly adhering to principles without regard for practical results”. They report ‘facts accurately’ but could ‘lack the warmth of human care’, lacking moral considerations. Some interviewees stand at the midpoint on the moral decision-making spectrum, adopting a ‘so-so’ perspective and relativism that relies on social norms to establish moral values. Given this, robot journalists would fall short when they maintain absolute standards as an inflexible approach in news reporting.

Theme 2:

Robot journalists as incompetent moral agents

It was also discovered that robot journalists demonstrate signs of incompetence as MAs when functioning as tools while facing problems related to accountability. Nevertheless, insightful solutions were recommended for overcoming these constraints.

Tool-like nature of robot journalists

Our interviewees generally held the view that robot journalists do not possess autonomy and are controlled by humans. This stance was explicitly reflected in discourses including the terms ‘manipulated’, free ‘will’, ‘puppets’, and free of ‘spontaneous desires’. Interviewees believed that robot journalists are ‘instruction’ dependent and lack ‘self-awareness’. Other interviewees also stated that granting robots the right to act independently is impractical. Another interviewee, Aurelia, emphasised that the “existence of a ‘black box’ for their actions makes it difficult to assign specific responsibilities” while Ning added that “tools are not responsible for anything”. They all argued that treating robot journalists as independent agents would complicate the assignment of responsibility and create an enigma, given the inherent incompatibility between independence and irresponsibility.

Accountability dilemma for robot journalists

Since most of the interviewees perceived robot journalists as mere tools, they concurred that such entities cannot be held accountable in the same manner as humans. This stance is reflected in the interviewees' statements, such as the assertion that there is "no sense in punishing the robot" or that it is 'meaningless' to punish them, as they are incapable of feeling 'any pain or guilt' or feeling 'good'. Moreover, interviewees noted that robots lack 'malicious intentions' and are 'not driven by greed or selfishness to act harmfully'. One interviewee believed that if robot journalists were held responsible for their actions and reports, human journalists might use them as 'scapegoats'. Conversely, other interviewees held a diametrically opposing view, arguing that if robot journalists were not held accountable for their actions, humans could be unfairly blamed. This view implies that future robot journalists may easily deceive humans, as the authenticity of robot-generated content might be dubious. For instance, Maomao stated: "A human editor might unknowingly publish a fabricated photo. In such cases, it's unreasonable to hold the editor responsible if the content proves to be false or harmful." One interviewee, Aurelia, highlighted the need for legal frameworks to address robots' potential for unpredictable and uncontrollable actions as their behaviour could "move beyond human control". She further emphasised: "Such highly intelligent robot journalists may seem a long way off at this point in time, the system and the law must always be one step ahead."

Possible solutions to the accountability dilemma

Our interviewees offered some solutions to address the issue of accountability for robot journalists. Several interviewees contended that implementing a responsibility-traceability system would be beneficial, encompassing designers of robot journalists, programmers, producers, and the news organisations employing them. Alternatively, other interviewees stated that the role of news organisations in upholding accountability should be strengthened, considering that they are the entities directly using robot journalists. One interviewee recommended appointing an experienced human journalist or editor to oversee robot journalists, concurring Qinyuan, another interviewee's analogy, who mentioned: "It's like an old master leading his apprentice, with one-to-one teaching and one-to-one accountability. When the apprentice goes wrong, the master is naturally held accountable." Another interviewee, Wanmei also emphasised the need for transparency. She added that "this would not only guarantee the public's right to know the source of the news, but also demonstrate a responsible attitude and facilitate subsequent accountability".

RQ3: How do Chinese students of journalism expect the interaction between human journalists and robot journalists to develop in journalistic contexts?

The study participants expressed a variety of viewpoints regarding the extent to which Chinese students of journalism expect human journalists to interact with robot journalists. The interview analysis revealed two primary themes: 1. erosion of news authenticity and

diversity; and 2. partners or rivals? The future of human–robot collaboration. Each theme includes two corresponding sub-themes, as illustrated in Table 4.

Table 4:
Theme, sub-themes and the initial code for the expectation of Chinese students of journalism

Theme	Sub-theme	Initial code
Erosion of news authenticity and diversity	Homogenisation of news and loss of diverse perspectives	‘repetitive formulas’; ‘loss of diverse voices’; ‘only copy the views of others’
	Absence of human touch in news	‘miss the nonverbal cues’; ‘flat’; ‘cold’;
Partners or rivals? The future of human–robot collaboration	Robotic journalist as helpful sidekicks	‘ensuring the timeliness of news’; ‘free of errors’; ‘detecting and correcting errors’
	Robotic journalists as emerging rivals	‘do the work of three human journalists’; ‘he’ll replace me’; ‘inevitable careless mistakes’
	Human-led cooperative relationship	‘a mascot or an attention-grabbing gimmick’; ‘lubricant’; ‘grasp the depth of human nature’

Source: Compiled by the authors.

Across the data, participants expressed relatively consistent views regarding both the capabilities and limitations of robot journalists, which are findings that align with prior research on human–machine collaboration. Notably, their perspectives extended beyond technical considerations to address broader issues related to industry structure, audience engagement, and the evolving role of human journalists, thereby illuminating the complex and evolving relationship between humans and AI-powered systems in journalism.

Theme 1: ***Erosion of news authenticity and diversity***

Students’ expectations on the extent of interaction between human journalists and robot journalist were consistent. It was reported that the authenticity and diversity of the news produced by robot journalists is compromised. The two main sub-themes derived from the analysis encompass: homogenisation of news and loss of diverse perspectives and the absence of the human touch in news.

Homogenisation of news and loss of diverse perspectives

Our interviewees also raised the concern that robot journalists might lead to homogenised news content, stifling diverse viewpoints. This was reflected in critiques by Dong: “It would be boring if all news were written by robot journalists using repetitive formulas”, alongside

phrases like “loss of diverse voices”, and that “robot journalists only copy the views of others”. Smaller outlets, unable to adopt robot journalism due to financial constraints, risk losing competitiveness, being phased out, and contributing to the homogenisation of voices in journalism. For example, one interviewee, Aurelia, explained that large newspapers owned by wealthy capitalists or backed by the state can afford to purchase robot journalists. These outlets, she argued, would amplify the perspectives of powerful groups, stating: “The voices they promote will reflect the thoughts of these classes.” She added that if newspapers are controlled by authorities, press freedom would be compromised, making it difficult for the public to hear diverse voices.

Absence of the human touch in news

While our interviewees generally agreed that robot journalists can uphold professional values, they also raised the concern that these systems might struggle to fully grasp the human elements of journalism. For example, Haiyang expressed the thought that “they may miss the nonverbal cues that human journalists can often catch during interviews”. One interviewee, Qinyuan, noted that news composed by robot journalists might be perceived as ‘flat’ or ‘cold’, lacking the emotional appeal necessary to “move people or inspire deep reflection”. Another interviewee, Ning, argued that “even though robot journalists can learn this through imitation and training, after all, it’s not what they do best”.

Theme 2:

Partners or rivals? The future of human–robot collaboration

Another main theme is the consideration of robots being either partners or rivals. The future of human–robot journalistic collaboration remains contradictory. The relationship functions both in support of each other at times yet can become competitive on other occasions while humans maintain authority over decisions.

Robot journalists as helpful sidekicks

Many interviewees argued that robot journalists may significantly enhance productivity in the news industry by outperforming humans in both efficiency and accuracy. This notion was manifested in practices “ensuring the timeliness of news”, while counterbalancing the “inevitable careless mistakes” made by human journalists. Unlike humans, robot systems are “free of errors” and excel at “detecting and correcting errors”. They assist human journalists by quickly processing and integrating large amounts of information, lightening their workload and enabling them to focus on in-depth investigative reporting.

Robot journalists as emerging rivals

Robot journalists are increasingly perceived as formidable competitors, sparking concerns among professionals about the potential displacement of human journalists. These arguments were reflected in statements such as “one robot journalist can do the work of three human journalists at the same time”, (Yiling) and “if I don’t work hard, he’ll replace me with a robot journalist” (Haiyang). To ensure survivability in the age of robots, interviewees emphasised the imperative for education in journalism to evolve. This evolution should focus on teaching collaboration with robot systems, enabling journalists to leverage the synergistic strengths of human creativity and machine efficiency. Another interviewee, Ruixue, argued that “learning and familiarising themselves [human journalists] more with new technologies and thinking about how to work better with robots”, is critical for effective human–robot collaboration. The interviewees argued that the amalgamation of human intuition with machine precision would achieve a formidable synergy.

Human-led cooperative relationship

Some interviewees advocated for a clear division of tasks between human journalists and robot systems. Robot journalists can efficiently handle technical tasks while humans focus on translating complex information into accessible narratives for audiences. For example, Zhihua stated: “We could equip robot journalists with scene reconstruction capabilities to simulate the scene of news events.” Robot journalists might function more like ‘a mascot or an attention-grabbing gimmick’ while human journalists would act as a ‘lubricant’ between robot systems and audiences. Our interviewees highlighted the inability of robot journalists to ‘grasp the depth of human nature’ in interactions and relationships, asserting that journalism fundamentally revolves around connecting with people, rather than merely processing data or deploying technology.

Discussion

The interviewees preferred robot journalists with human-like appearances, reflecting schema congruence (familiarity boosting acceptance) (Lou et al., 2022) and anthropomorphism (attributing human traits to nonhumans) (Epley et al., 2007). However, the phenomenon of the Uncanny Valley Theory positing that overly human-like robots cause discomfort (Mori, 2012) was also evident in participants’ accounts. Anthropomorphism extends beyond physical traits or behaviours to include attributing human mental states to nonhumans (Dubois-Sage et al., 2023). This heightened anthropomorphism, echoing Theory of Mind (ToM, i.e. the ability to infer mental states), was also reflected in interviewees’ positioning (Lopez-Soto et al., 2024). They assigned cognitive and emotional abilities to robot journalists, deeming these essential for effective journalism. Interviewees also perceived human-like robots as increasing trust and relatability, reinforcing existing evidence that anthropomorphism

enhances predictability, trustworthiness (Roesler et al., 2024), and collaboration with nonhuman agents (Zhao & Malle, 2022).

Given empathy's critical role in journalism – its absence hinders work in emotionally charged scenarios – and its salience in human–robot trust (Pelau et al., 2021), interviewees deemed empathy essential for robot journalists but highlighted its artificiality. This mirrors weak anthropomorphism, where humans treat agents as if they possess human traits, despite knowing otherwise (Epley et al., 2007). Critics call artificial empathy deceptive, as it relies on predefined rules, not consciousness or authentic emotion (Mlonyeni, 2025; Umbrello & Natale, 2024).

Two predominant themes emerged from interviewees' accounts regarding robot journalists' decision-making capabilities: *programme-based abilities* and *experience-based abilities*. The former is principle-based and algorithm-centric, enabling robot journalists to manage both straightforward news decisions and complex ethical judgments. Conversely, the latter is established based on robot journalists' ability to learn from human journalists' experiences, resembling collective intelligence. The programme-driven approach exemplifies the top-down methodology for designing AMAs, wherein AI systems are endowed with ethical frameworks rooted in moral philosophy (Wallach et al., 2020). This paradigm also echoes the codifiability thesis, which posits that ethics can be condensed into universal moral rules applicable regardless of individual moral competence (Véliz, 2021). The bottom-up approach is characterised by decision-making rooted in internal practices and the autonomous redefinition of decision-making skills. Robot journalists utilise adaptive mechanisms to render ethically sound judgments (Wallach et al., 2020). Robot journalists' ethical decision-making as it manifested in interviewee discourses is also reminiscent of Moor's (2006) explicit ethical agent framework, in which agents are designed to recognise morally salient information in specific contexts and act in accordance with externally derived ethical principles.

Robot journalism's resistance to bias and aid from diverse databases were also shared by the interviewees. This suggests the assertion of 'machine objectivism', where machines could outperform humans as MA by avoiding irrationality and self-interest (Martinho et al., 2021). Scholars outlined AI as Kantian agents applying universal ethics (via programmed principles), not subjective motives (Brožek & Janik, 2019). Interviewees similarly saw robot journalists making principled decisions using pre-set moral frameworks, ensuring consistent, unbiased outcomes free from human fallibility. Another cluster of interviewees endorsed utilitarian ethical principles, supporting Grau's (2006) argument that machines can act as superior utilitarian MA, maximising overall utility. Interviewees' perception of robot journalists as emotionless, objective entities reflects this utilitarian stance. However, robot journalists might rigidly follow abstract ethical rules, risking unintended harm. This highlights a key flaw in the top-down AI approach, that is, overly abstract principles struggle to address specific situations effectively (Wallach et al., 2020). For example, utilitarianism's goal to "maximise the greatest good" offers no guidance on who to include in calculations or how to quantify 'utility' (Brožek & Janik, 2019). Interviewees noted that an AI calculated 'best' outcome might conflict with human values or societal appropriateness, casting doubt on its capacity for ethical decision-making. This aligns with the existing argument that AMAs cannot replace humans in resolving ambiguous moral dilemmas, as ethical decision-making involves complex cultural, social, and personal considerations that are not purely algorithmic (Martinho et al., 2021).

Similar concerns were raised about the bottom-up approach. Interviewees noted that there is risk of acquiring biases or unethical tendencies from their training data or exposure to human practices. Interviewees also expressed concerns about biases embedded in algorithms that guide the robot journalists. Therefore, they proposed solutions to mitigate bias including pre-processing techniques during data selection, in-processing adjustments during model training, and post-processing refinements in application (Mehrabani et al., 2021).

Interviewees also described robot journalists as devices that reproduce human moral decision-making and exert moral influence, thus conveying moral agency. This manifests the *functional equivalence* view, moral agency becomes equivalent to or functions as an MA (Søvik, 2022). Robot journalists could qualify as MAs by employing top-down and bottom-up ethical reasoning systems. Customary use of human moral criteria still falls short to qualify robots as agents who demonstrate authentic moral agency. An essential problem with moral practices stems from an inability to translate them into formal instructions (Purves et al., 2015). Machines frequently prove able to make sound moral choices yet they struggle with unfamiliar scenarios or unprogrammed exceptions, rendering morality beyond their reach (Véliz, 2021). As algorithms assume more roles in public and private sectors, humans may increasingly shift blame to automated systems when errors occur. If algorithms become targets of praise or blame, accountability for those who design, deploy, or oversee them could erode (Véliz, 2021).

According to the ontological view of moral agency, attributes such as autonomy and consciousness are required for true moral responsibility (Cervantes et al., 2020; Formosa, 2021; Mosakas, 2021); however, our interviewees described robot journalists as lacking autonomy. As in previous studies, the interviewees argued that machines cannot be true MA because in the absence of morality they act as tools, hence they cannot bear moral responsibility (Véliz, 2021). Some scholars contest the necessity of autonomy for moral responsibility, contrasting with interviewees' emphasis on autonomy as central to moral agency. Others propose sentience – subjective experiences of pleasure, pain and awareness of consequences – as the true basis of moral agency (Mosakas, 2021; Véliz, 2021). This is corresponded with interviewees' critiques of robot journalists as entities devoid of emotional understanding or subjective experience. For instance, robot journalists mechanically prioritise programmed 'values' as numerical rankings (Janoušek, 2024). Hence, they cannot engage moral action while bearing responsibility according to Confucian theories that both directly oppose the ideas of fairness, as well as objectivity (Chuang, 2025). The interviewees noted that punishing robot journalists is futile, as they cannot feel guilt or pain; and while humans may anthropomorphise machines (e.g. blaming inanimate objects), they inherently recognise that these entities lack true accountability.

Some interviewees, echoing scholars, suggested robot journalists could evolve into MA as AI advances are achieved via a 'master algorithm' (Bertoncini & Serafim, 2023). They speculated such systems might surpass human cognition, developing independent 'minds' worthy of accountability. However, critics counter with parallels to psychopaths – while highly intelligent, their inability to feel shame or grasp morality negates moral agency (Mosakas, 2021). This aligns with the orthogonality thesis, which warns that even supremely intelligent AI could lack inherent concern for human well-being (Véliz, 2021).

The situation demonstrates why Confucian values regarding ethical choices would be missing from robot reporters.

Interviewees' comparison of robot journalists to 'puppets' matches the instrumentalist theory position (Gunkel, 2020), implying human responsibility as the only factor. However, critics argue that if developers face liability for AI's unintended actions, they may limit advanced, autonomous systems to avoid culpability (Gunkel, 2020). This tension grows as robot journalists increasingly demonstrate independent news and ethical judgments, blurring the line between tool and agent. A key concern is the unpredictability of AI systems with learning capabilities, which complicates assigning human responsibility – a dilemma termed the 'responsibility gap' (Hindriks & Veluwenkamp, 2023). However, hybrid approaches still struggle to distinguish between accountable agents and mere tools. To bridge this gap, interviewees proposed a responsibility traceability system involving designers, programmers and news organisations, aligned with the principle of 'meaningful human control' (Santoni de Sio & van den Hoven, 2018) that requires two conditions: tracking (monitoring AI decisions) and tracing (mapping accountability across stakeholders) (Santoni de Sio & van den Hoven, 2018).

The interviewees downplayed robot journalists' capabilities and assert human dominance in journalistic areas requiring sensitivity and discernment, such as investigative reporting, empathetic storytelling and ethical decision-making. This aligns with prior research showing journalists often dismiss automation as competition by strategically emphasising their superior skills (e.g. emotional intelligence, creativity and news judgment) (Rydenfelt, 2022; van Dalen, 2026). Such arguments reinforce a human-centric view of journalism, which may reflect speciesism, a human cognitive prejudice against other species (Pauketat & Anthis, 2022).

Conclusion

The increasing integration of new technological advancements – specifically robots – in journalism raises concerns about how these advancements, once viewed as merely submissive to human journalists, are now becoming more pronounced in the field. This shift presents new opportunities that could significantly impact the landscape of journalism. The current study, pioneering among limited existing research, puts forward inquiries into the relevance of moral agency in robot journalists, its influence on the decision-making process in journalism, and how this perceived conceptualisation of moral agency flows within the collaboration between human journalists and robot journalists. Although diverse perspectives exist with regard to AI-powered journalism robots, a prevailing view is that these entities fall short of being full moral agents. They are better described as explicit moral agents – guided by pre-programmed systems – capable of making seemingly sensible moral decisions, yet lacking consciousness and emotional sensitivity. While there is an argument that robots should possess human-like attributes – including cognitive and sensory abilities – to effectively function as journalists, the 'uncanny valley' phenomenon emerges as a significant pitfall. This effect may hinder their acceptance and effectiveness in human-facing roles such as journalism. These findings could practically contribute to both software and hardware developers by

informing the design of robot features that reflect anthropomorphic qualities while avoiding the ‘uncanny valley effect’, which undermines trust and communication effectiveness.

Given that the term AMAs is still not fully developed in practical and operational terms within the field of journalism, traditional expectations – such as empathising with the subjects involved in a story – remain essential. This highlights that AMAs still have a long way to go before becoming concrete and reliable in journalistic practice, and before their theoretical axioms can translate into practical application. Such limitations support the reservations and critical stances toward incorporating machines into journalism, despite the appeal of a functionalist perspective. AI-powered robots still pose a significant accountability challenge in journalism. This calls for urgent measures from Chinese policymakers to establish clear governance frameworks in newsrooms, ensuring human oversight and clearly defined accountability systems. Ultimately, primary agency and leadership should remain with human beings, without burdening them with the uncertainties arising from robot journalists.

Although this study yields contributions, its limitations, which could serve as a point of departure for future research, should be acknowledged. While the targeted subjects were students of journalism with prior exposure to robots and AI, their limited real-world journalism experience may have influenced their responses. Experienced or traditional journalists might hold different perspectives on robot journalism and the concept of AMAs, particularly as practiced in real journalistic settings; and their interactions with robots could also differ significantly from those of journalism students. This study employed a basic interpretative design with semi-structured interviews as the primary data collection method. However, to better capture the complexities of human–robot relationships – particularly those shaped by mutual, unmediated interaction and relational dialectics – an ethnographic observational approach could address limitations inherent in interview-based methods and current research designs.

Acknowledgment

We extend our sincere gratitude to all the participants who contributed their time and insights to this study. Their willingness to share their experiences and perspectives during the interviews made this research possible.

References

- Ayu Karunianingsih, D., Sarifah, S., Korina Relawati, D., Dwi Setyorini, M. & Utomo, A. (2025). Use of Artificial Intelligence Technology AI News Presenter in the News Production Process at TVOne.ai. *COMMICAST*, 6(2), 348–363. Online: <https://doi.org/10.12928/commicast.v6i2.14323>
- Bayram, B. & İnce, G. (2017). Advances in Robotics in the Era of Industry 4.0. In A. Ustundag & E. Cevickan (Eds.), *Industry 4.0: Managing the Digital Transformation* (pp. 187–200). Springer. Online: https://doi.org/10.1007/978-3-319-57870-5_11
- Behdadi, D. & Munthe, C. (2020). A Normative Approach to Artificial Moral Agency. *Minds and Machines*, 30, 195–218. Online: <https://doi.org/10.1007/s11023-020-09525-8>

- Bertoncini, A. L. C. & Serafim, M. C. (2023). Ethical Content in Artificial Intelligence Systems: A Demand Explained in Three Critical Points. *Frontiers in Psychology*, 14. Online: <https://doi.org/10.3389/fpsyg.2023.1074787>
- Bogue, R. (2020). Robots in a Contagious World. *Industrial Robot*, 47(5), 637–642. Online: <https://doi.org/10.1108/IR-05-2020-0101>
- Braun, V. & Clarke, V. (2019). Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. Online: <https://doi.org/10.1080/2159676x.2019.1628806>
- Braun, V. & Clarke, V. (2021). To Saturate or Not to Saturate? Questioning Data Saturation as a Useful Concept for Thematic Analysis and Sample-Size Rationales. *Qualitative Research in Sport, Exercise and Health*, 13(2), 201–216. Online: <https://doi.org/10.1080/2159676x.2019.1704846>
- Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M. & Chuan, C.-H. (2019). Artificial Intelligence and Journalism. *Journalism & Mass Communication Quarterly*, 96(3), 673–695. Online: <https://doi.org/10.1177/1077699019859901>
- Brożek, B. & Janik, B. (2019). Can Artificial Intelligences Be Moral Agents? *New Ideas in Psychology*, 54, 101–106. Online: <https://doi.org/10.1016/j.newideapsych.2018.12.002>
- Byrne, D. A. (2022). A Worked Example of Braun and Clarke’s Approach to Reflexive Thematic Analysis. *Quality & Quantity*, 56, 1391–1412. Online: <https://doi.org/10.1007/s11135-021-01182-y>
- Carlson, M. (2018). Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism. *New Media & Society*, 20(5), 1755–1772. Online: <https://doi.org/10.1177/1461444817706684>
- Cervantes, J., López, S., Rodríguez, L., Cervantes, S., Cervantes, F. & Ramos, F. (2020). Artificial Moral Agents: A Survey of the Current Status. *Science and Engineering Ethics*, 26(2), 501–532. Online: <https://doi.org/10.1007/s11948-019-00151-x>
- Chuang, S. (2025). A Philosophical Look at Confucian Humanity in the Age of Robots. *International Journal of Organizational Analysis*, 33(5), 1324–1334. Online: <https://doi.org/10.1108/IJOA-03-2024-4380>
- Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 26(4), 2051–2068. Online: <https://doi.org/10.1007/s11948-019-00146-8>
- Danziger, S. (2022). Intelligence as a Social Concept: A Socio-Technological Interpretation of the Turing Test. *Philosophy & Technology*, 35(3). Online: <https://doi.org/10.1007/s13347-022-00561-z>
- Domingos, P. (2015). *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake our World*. Basic Books.
- Doncieux, S., Chatila, R., Straube, S. & Kirchner, F. (2022). Human-centered AI and Robotics. *AI Perspectives*, 4(1). Online: <https://doi.org/10.1186/s42467-021-00014-x>
- Dörr, K. N. & Hollnbuchner, K. (2017). Ethical Challenges of Algorithmic Journalism. *Digital Journalism*, 5(4), 404–419. Online: <https://doi.org/10.1080/21670811.2016.1167612>
- Dubois-Sage, M., Jacquet, B., Jamet, F. & Baratgin, J. (2023). We Do Not Anthropomorphize a Robot Based only on its Cover: Context Matters Too! *Applied Sciences*, 13. Online: <https://doi.org/10.3390/app13158743>
- Dung, L. (2025). Understanding Artificial Agency. *The Philosophical Quarterly*, 75(2), 450–472. Online: <https://doi.org/10.1093/pq/pqae010>
- Epley, N., Waytz, A. & Cacioppo, J. T. (2007). On Seeing Human: A Three-factor Theory of Anthropomorphism. *Psychological Review*, 114(4), 864–886. Online: <https://doi.org/10.1037/0033-295X.114.4.864>
- Floridi, L. (2016). Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083). Online: <https://doi.org/10.1098/rsta.2016.0112>

- Floridi, L. (2025). AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. *Philosophy & Technology*, 38(1). Online: <https://doi.org/10.1007/s13347-025-00858-9>
- Formosa, P. (2021). Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. *Minds and Machines*, 31(4), 595–616. Online: <https://doi.org/10.1007/s11023-021-09579-2>
- Formosa, P. & Ryan, M. (2020). Making Moral Machines: Why We Need Artificial Moral Agents. *AI & Society*, 36, 839–851. Online: <https://doi.org/10.1007/s00146-020-01089-6>
- Gao, D. & Yu, D. (2024). Challenges and Cracks: Ethical Issues in the Development of Artificial Intelligence. *Science, Technology and Society*, 29(3), 415–434. Online: <https://doi.org/10.1177/09717218241246372>
- Gasparotto, A. (2016). Robots in History: Legends and Prototypes from Ancient Times to the Industrial Revolution. In C. López-Cajún & M. Ceccarelli (Eds.), *Explorations in the History of Machines and Mechanisms. Proceedings of the Fifth IFToMM Symposium on the History of Machines and Mechanisms* (pp.39–49). Springer.
- Gordon, J.-S. & Gunkel, D. J. (2022). Moral Status and Intelligent Robots. *The Southern Journal of Philosophy*, 60(1), 88–117. Online: <https://doi.org/10.1111/sjp.12450>
- Gordon, J.-S. & Pasvenskiene, A. (2021). Human Rights for Robots? A Literature Review. *AI and Ethics*, 1(3–4), 579–591. Online: <https://doi.org/10.1007/s43681-021-00050-7>
- Graff, J. (2024). Moral Sensitivity and the Limits of Artificial Moral Agents. *Ethics and Information Technology*, 26(1). Online: <https://doi.org/10.1007/s10676-024-09755-9>
- Grau, C. (2006). There Is No “I” in “Robot”: Robots and Utilitarianism. *IEEE Intelligent Systems*, 21(4), 52–55. Online: <https://doi.org/10.1109/MIS.2006.81>
- Gudmunsen, Z. (2024). The Moral Decision Machine: A Challenge for Artificial Moral Agency Based on Moral Deference. *AI and Ethics*, 5, 1033–1045. Online: <https://doi.org/10.1007/s43681-024-00444-3>
- Gunkel, D. J. (2020). Mind the Gap: Responsible Robotics and the Problem of Responsibility. *Ethics and Information Technology*, 22(4), 307–320. Online: <https://doi.org/10.1007/s10676-017-9428-2>
- Haenlein, M. & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5–14. Online: <https://doi.org/10.1177/0008125619864925>
- Henestrosa, A. L., Greving, H. & Kimmerle, J. (2023). Automated Journalism: The Effects of AI Authorship and Evaluative Information on the Perception of a Science Journalism Article. *Computers in Human Behavior*, 138. Online: <https://doi.org/10.1016/j.chb.2022.107445>
- Hennink, M. & Kaiser, B. N. (2022). Sample Sizes for Saturation in Qualitative Research: A Systematic Review of Empirical Tests. *Social Science & Medicine*, 292. Online: <https://doi.org/10.1016/j.socscimed.2021.114523>
- Henschel, A., Laban, G. & Cross, E. S. (2021). What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You. *Current Robotics Reports*, 2(1), 9–19. Online: <https://doi.org/10.1007/s43154-020-00035-0>
- Hindriks, F. & Veluwenkamp, H. (2023). The Risks of Autonomous Machines: From Responsibility Gaps to Control Gaps. *Synthese*, 201(1). Online: <https://doi.org/10.1007/s11229-022-04001-5>
- Hoffmann, C. H. (2022). Is AI Intelligent? An Assessment of Artificial Intelligence, 70 Years after Turing. *Technology in Society*, 68. Online: <https://doi.org/10.1016/j.techsoc.2022.101893>
- Horáková, J. & Kelemen, J. (2009). Artificial Living Beings and Robots: One Root, Variety of Influences. *Artificial Life and Robotics*, 13(2), 555–560. Online: <https://doi.org/10.1007/s10015-008-0502-z>

- Hua, X. (2018, June 15). Xinhua Upgrades AI-based News Production System. *Xinhua*. Online: http://www.xinhuanet.com/english/2018-06/15/c_137256200.htm
- Hua, X. (2024, November 4). Feature: China's AI-savvy Younger Generation Creates a New Wave of Cultural Memes. *Xinhua*. Online: <https://english.news.cn/20241104/621d370da3ef40d08cad850884ca30bb/c.html>
- Janoušek, H. (2024). Brentano and Husserl on Hume's Moral Philosophy. *Journal of the British Society for Phenomenology*, 55(2), 133–148. Online: <https://doi.org/10.1080/00071773.2024.2308281>
- Jiang, Y., Li, X., Luo, H., Yin, S. & Kaynak, O. (2022). Quo Vadis Artificial Intelligence? *Discover Artificial Intelligence*, 2(1). Online: <https://doi.org/10.1007/s44163-022-00022-8>
- Jing, S. (2024, May 30). Report Reveals Youth's Views on AIGC. *China Daily*. Online: <https://www.chinadaily.com.cn/a/202405/30/WS66586a99a31082fc043ca12d.html>
- Jones, B. & Jones, R. (2025). Action Research at the BBC: Interrogating Artificial Intelligence with Journalists to Generate Actionable Insights for the Newsroom. *Journalism*, 26(8), 1708–1725. Online: <https://doi.org/10.1177/14648849251317150>
- Kim, D. & Kim, S. (2018). Newspaper Journalists' Attitudes towards Robot Journalism. *Telematics and Informatics*, 35(2), 340–357. Online: <https://doi.org/10.1016/j.tele.2017.12.009>
- Kim, D. & Kim, S. (2021). A Model for User Acceptance of Robot Journalism: Influence of Positive Disconfirmation and Uncertainty Avoidance. *Technological Forecasting and Social Change*, 163. Online: <https://doi.org/10.1016/j.techfore.2020.120448>
- Krauss, S. E. (2005). Research Paradigms and Meaning Making: A Primer. *The Qualitative Report*, 10(4), 758–770. Online: <https://doi.org/10.46743/2160-3715/2005.1831>
- Laitinen, A. & Sahlgrén, O. (2021). AI Systems and Respect for Human Autonomy. *Frontiers in Artificial Intelligence*, 4. Online: <https://doi.org/10.3389/frai.2021.705164>
- Laukyte, M. (2017). Artificial Agents among Us: Should We Recognize Them as Agents Proper? *Ethics and Information Technology*, 19(1), 1–17. Online: <https://doi.org/10.1007/s10676-016-9411-3>
- Leppänen, L., Munezero, M., Granroth-Wilding, M. & Toivonen, H. (2017). Data-driven News Generation for Automated Journalism. In J. M. Alonso, A. Bugarin & E. Reiter (Eds.), *Proceedings of the 10th International Conference on Natural Language Generation* (pp. 188–197). Online: <https://doi.org/10.18653/v1/W17-3528>
- Licardo, J. T., Domjan, M. & Orehovački, T. (2024). Intelligent Robotics – A Systematic Review of Emerging Technologies and Trends. *Electronics*, 13(3). Online: <https://doi.org/10.3390/electronics13030542>
- List, C. (2021). Group Agency and Artificial Intelligence. *Philosophy & Technology*, 34(4), 1213–1242. Online: <https://doi.org/10.1007/s13347-021-00454-7>
- Lopez-Soto, T., Garcia-Lopez, A. & Salguero-Lamillar, F. J. (2024). The Theory of Mind under Scrutiny: Psychopathology, Neuroscience, Philosophy of Mind and Artificial Intelligence. In T. Lopez-Soto, A. Garcia-Lopez & F. J. Salguero-Lamillar (Eds.), *The Theory of Mind under Scrutiny*. Springer. Online: <https://doi.org/10.1007/978-3-031-46742-4>
- Lou, C., Kang, H. & Tse, C. H. (2022). Bots vs. Humans: How Schema Congruity, Contingency-based Interactivity, and Sympathy Influence Consumer Perceptions and Patronage Intentions. *International Journal of Advertising*, 41(4), 655–684. Online: <https://doi.org/10.1080/02650487.2021.1951510>
- Lyons, S., Urick, M., Kuron, L. & Schweitzer, L. (2015). Generational Differences in the Workplace: There Is Complexity beyond the Stereotypes. *Industrial and Organizational Psychology*, 8(3), 346–356. Online: <http://doi:10.1017/iop.2015.48>

- Martinho, A., Poulsen, A., Kroesen, M. & Chorus, C. (2021). Perspectives about Artificial Moral Agents. *AI and Ethics*, 1(4), 477–490. Online: <https://doi.org/10.1007/s43681-021-00055-2>
- Matilainen, T. (2024). A Communicational Ontology Inspired by Peter Singer. *Journal of Media Ethics*, 39(4), 230–243. Online: <https://doi.org/10.1080/23736992.2024.2418047>
- Mays, K. K., Cummings, J. J. & Katz, J. E. (2025). The Robot Rights and Responsibilities Scale: Development and Validation of a Metric for Understanding Perceptions of Robots' Rights and Responsibilities. *International Journal of Human-Computer Interaction*, 41(5), 3413–3430. Online: <https://doi.org/10.1080/10447318.2024.2338332>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 1–35. Online: <https://doi.org/10.1145/3457607>
- Mellado, C. & Gajardo, C. (2026). The Importance of the Human Touch in Journalism: Journalistic Values from an Audience Perspective. *Journalism Practice*, 20(3), 747–764. Online: <https://doi.org/10.1080/17512786.2024.2387668>
- Merriam, S. B. (2002). *Qualitative Research in Practice. Examples for Discussion and Analysis*. Jossey-Bass.
- Merriam, S. B. (2009). *Qualitative Research. A Guide to Design and Implementation*. Jossey-Bass.
- Ministry of Science and Technology of the People's Republic of China. (2021, September 26). 新一代人工智能伦理规范 [New Generation of Artificial Intelligence Ethics]. Online: https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html
- Mlonyeni, P. M. T. (2025). Personal AI, Deception, and the Problem of Emotional Bubbles. *AI & Society*, 40, 1927–1938. Online: <https://doi.org/10.1007/s00146-024-01958-4>
- Molla, M. A. M. & Ahsan, M. M. (2025). Artificial Intelligence and Journalism: A Systematic Bibliometric and Thematic Analysis of Global Research. *Computers in Human Behavior Reports*, 20. Online: <https://doi.org/10.1016/j.chbr.2025.100830>
- Moor, J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), 18–21. Online: <https://doi.org/10.1109/mis.2006.80>
- Moran, M. E. (2007). Evolution of Robotic Arms. *Journal of Robotic Surgery*, 1(2), 103–111. Online: <https://doi.org/10.1007/s11701-006-0002-x>
- Mori, M. (2012). *The Uncanny Valley: The Original Essay by Masahiro Mori*. Trans. K. F. MacDorman & N. Kageki. *IEEE Spectrum*, 6(1), 1–6. (Original work 1970.)
- Mosakas, K. (2021). On the Moral Status of Social Robots: Considering the Consciousness Criterion. *AI & Society*, 36, 429–443. Online: <https://doi.org/10.1007/s00146-020-01002-1>
- Naeem, M., Ozuem, W., Howell, K. & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, 22. Online: <https://doi.org/10.1177/16094069231205789>
- Natale, S. & Ballatore, A. (2020). Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence. *Convergence: The International Journal of Research into New Media Technologies*, 26(1), 3–18. Online: <https://doi.org/10.1177/1354856517715164>
- National Development and Reform Commission of the People's Republic of China (2023, December 6). Stand at the Forefront of the Times and Seize the Opportunities for Development. *NDRC gov*. Online: https://www.ndrc.gov.cn/wxdw/ztlz/NEW_srxgjcjppjjsx/jjsxyjqk/sxzj/202401/t20240116_1363486.html
- Nguyen, D. & Hekman, E. (2022). The News Framing of Artificial Intelligence: A Critical Exploration of How Media Discourses Make Sense of Automation. *AI & Society*, 39(2), 437–451. Online: <https://doi.org/10.1007/s00146-022-01511-1>

- Nyimbili, F. & Nyimbili, L. (2024). Types of Purposive Sampling Techniques with Their Examples and Application in Qualitative Research Studies. *British Journal of Multidisciplinary and Advanced Studies*, 5(1), 90–99. Online: <https://doi.org/10.37745/bjmas.2022.0419>
- Pauketat, J. V. & Anthis, J. R. (2022). Predicting the Moral Consideration of Artificial Intelligences. *Computers in Human Behavior*, 136. Online: <https://doi.org/10.1016/j.chb.2022.107372>
- Pasupathi, M. & Wainryb, C. (2010). Developing Moral Agency through Narrative. *Human Development*, 53(2), 55–80. Online: <https://doi.org/10.1159/000288208>
- Pelau, C., Dabija, D. & Ene, I. (2021). What Makes an AI Device Human-like? The Role of Interaction Quality, Empathy and Perceived Psychological Anthropomorphic Characteristics in the Acceptance of Artificial Intelligence in the Service Industry. *Computers in Human Behavior*, 122. Online: <https://doi.org/10.1016/j.chb.2021.106855>
- Porlezza, C. (2024). The Datafication of Digital Journalism: A History of Everlasting Challenges between Ethical Issues and Regulation. *Journalism*, 25(5), 1167–1185. Online: <https://doi.org/10.1177/14648849231190232>
- Purves, D., Jenkins, R. & Strawser, B. J. (2015). Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice*, 18, 851–872. Online: <https://doi.org/10.1007/s10677-015-9563-y>
- Roesler, E., Vollmann, M., Manzey, D. & Onnasch, L. (2024). The Dynamics of Human–Robot Trust Attitude and Behavior – Exploring the Effects of Anthropomorphism and Type of Failure. *Computers in Human Behavior*, 150. Online: <https://doi.org/10.1016/j.chb.2023.108008>
- Rydenfelt, H. (2022). Transforming Media Agency? Approaches to Automation in Finnish Legacy Media. *New Media & Society*, 24(12), 2598–2613. Online: <https://doi.org/10.1177/1461444821998705>
- Santoni de Sio, F. & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them. *Philosophy & Technology*, 34(4), 1057–1084. Online: <https://doi.org/10.1007/s13347-021-00450-x>
- Santoni de Sio, F. & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robots and AI*, 5. Online: <https://doi.org/10.3389/frobt.2018.00015>
- Schmitt, B. (2020). Speciesism: An Obstacle to AI and Robot Adoption. *Marketing Letters*, 31, 3–6. Online: <https://doi.org/10.1007/s11002-019-09499-3>
- Sebastián, M. Á. (2021). First-person Representations and Responsible Agency in AI. *Synthese*, 199(3), 7061–7079. Online: <https://doi.org/10.1007/s11229-021-03105-8>
- Sjøvaag, H. (2013). Journalistic Autonomy: Between Structure, Agency and Institution. *Nordicom Review*, 34(Special Issue), 155–166. Online: <https://doi.org/10.2478/nor-2013-0111>
- Søvik, A. O. (2022). How a Non-Conscious Robot Could Be an Agent with Capacity for Morally Responsible Behaviour. *AI and Ethics*, 2(4), 789–800. Online: <https://doi.org/10.1007/s43681-022-00140-0>
- Stupart, R. (2021). Feeling Responsible: Emotion and Practical Ethics in Conflict Journalism. *Media, War & Conflict*, 14(3), 268–281. Online: <https://doi.org/10.1177/17506352211013461>
- Sun, M., Hu, W. & Wu, Y. (2024). Public Perceptions and Attitudes towards the Application of Artificial Intelligence in Journalism: From a China-based Survey. *Journalism Practice*, 18(3), 548–570. Online: <https://doi.org/10.1080/17512786.2022.2055621>
- Swanepoel, D. & Corks, D. (2024). Artificial Intelligence and Agency: Tie-breaking in AI Decision-making. *Science and Engineering Ethics*, 30(2). Online: <https://doi.org/10.1007/s11948-024-00476-2>
- Symons, J. & Abumusab, S. (2024). Social Agency for Artifacts: Chatbots and the Ethics of Artificial Intelligence. *Digital Society*, 3(1). Online: <https://doi.org/10.1007/s44206-023-00086-8>

- Umbrello, S. & Natale, S. (2024). Reframing Deception for Human-centered AI. *International Journal of Social Robotics*, 16(11), 2223–2241. Online: <https://doi.org/10.1007/s12369-024-01184-4>
- Van Dalen, A. (2026). Revisiting the Algorithms behind the Headlines. How Journalists Respond to Professional Competition of Generative AI. *Journalism Practice*, 20(3), 897–914. Online: <https://doi.org/10.1080/17512786.2024.2389209>
- Véliz, C. (2021). Moral Zombies: Why Algorithms Are Not Moral Agents. *AI & Society*, 36(2), 487–497. Online: <https://doi.org/10.1007/s00146-021-01189-x>
- Wallach, W., Allen, C. & Smit, I. (2020). Machine Morality: Bottom-up and Top-down Approaches for Modelling Human Moral Faculties. In W. Wallach & P. Asaro (Eds.), *Machine Ethics and Robot Ethics* (pp. 249–266). Routledge. Online: <https://doi.org/10.4324/9781003074991>
- Wang, S. & Huang, G. (2024). The Impact of Machine Authorship on News Audience Perceptions: A Meta-analysis of Experimental Studies. *Communication Research*, 51(7), 815–842. Online: <https://doi.org/10.1177/00936502241229794>
- Wang, T.-M., Tao, Y. & Liu, H. (2018). Current Researches and Future Development Trend of Intelligent Robot: A Review. *International Journal of Automation and Computing*, 15(5), 525–546. Online: <https://doi.org/10.1007/s11633-018-1115-1>
- Wu, H., Chen, Z., Huang, Y. & Tu, H. (2024). Research on the Uncanny Valley Effect in Artificial Intelligence News Anchors. *Multimedia Tools and Applications*, 83, 62581–62606. Online: <https://doi.org/10.1007/s11042-023-18073-z>
- Zhang, Z. (2019, March 11). Will Female AI News Anchors Replace Humans on the Job?: China Daily columnist. *The Straits Times*. Online: <https://www.straitstimes.com/asia/east-asia/will-female-ai-news-anchors-replace-humans-on-the-job-china-daily-columnist>
- Zhao, X. & Malle, B. F. (2022). Spontaneous Perspective Taking toward Robots: The Unique Impact of Humanlike Appearance. *Cognition*, 224. Online: <https://doi.org/10.1016/j.cognition.2022.105076>
- Zhong, D., Wu, F., Huang, Z. & Chen, Y. (2025). Navigating the Human-digital Nexus: Understanding Consumer Intentions with AI Anchors in Live Commerce. *Behaviour & Information Technology*, 44(16), 4096–4111. Online: <https://doi.org/10.1080/0144929X.2025.2469655>

Perception of Information and Disinformation on Social Media

Daily Access and Age of Adolescents and Young People as Predictive Factors

Beatriz Catalina-García,^{*} Antonio García-Jiménez,^{**}
Manuel Montes-Vozmediano^{***}

* Universidad Rey Juan Carlos. Despacho 033. Edificio departamental I. Campus de Fuenlabrada, SPAIN
E-mail: beatriz.catalina[at]urjc.es

** Universidad Rey Juan Carlos. Despacho 242. Edificio departamental I. Campus de Fuenlabrada, SPAIN
E-mail: antonio.garcia[at]urjc.es

*** Universidad Rey Juan Carlos. Despacho 136. Edificio departamental I. Campus de Fuenlabrada, SPAIN
E-mail: manueljavier.montes[at]urjc.es

Social media currently plays a key role for adolescents and young people in accessing information. The integration of these platforms into their media diet leads them to build their own digital architecture to stay informed and to avoid disinformation. Based on academic literature and a survey (n = 1,800), this study analyses the perceptions of Spanish young people and adolescents (aged 14 to 24) regarding various characteristics of social media in relation to information and disinformation. Our findings reveal that daily access to social media is a good predictor regarding propositions related to information, whereas age is generally a slightly better indicator for those related to disinformation. These results provide deeper insight into the components that influence adolescents and young people, helping to better understand how they shape their own media repertoires.

Keywords: adolescents, young people, social media, information, disinformation

Address for correspondence: Beatriz Catalina-García, e-mail: beatriz.catalina[at]urjc.es

Article received on 23 December 2025. Article accepted on 23 February 2026.

Conflict of interest: The authors declare no conflicts of interest.

Funding: This work is part of the R&D&I project 'Media repertoires and practices in adolescence and youth: uses, cyber wellbeing and digital vulnerabilities in social networks' (PID2022-138281NB-C21) has been funded by the Ministry of Science, Innovation and Universities – Spain (MCIN/AEI/10.13039/501100011033) and supported by 'ERDF a way of building Europe'.

Introduction

Social media platforms are a basic part of daily life for young people and adolescents, who use them not only for entertainment (Harff & Schmuck, 2025), but also as channels for information sources, which for them translates into a form of social connection (Anderson, 2026). This is a general trend among young Europeans aged 16 to 24, who place more importance on social media than television as a source of news (European Parliament, 2025). They are drawn to a variety of more engaging formats to them than traditional media (Chou et al., 2023), and appreciate the opportunity to personalise their own content (Anter & Kümpel, 2025). Although they do not always actively seek out information (Hendrickx, 2025), they are convinced of the importance of staying informed in order to take on the role of citizens through the cultural value that news provides (Tamboer et al., 2022).

The high level of digital exposure to news also means that social media has become the most common medium used by young people and adolescents (Farias-Battle et al., 2024). However, this leads to the indiscriminate and widespread consumption of unverified news (Selnes, 2024), a situation made worse by their overconfidence – especially among adolescents – in their ability to detect disinformation (Zozaya-Durazo et al., 2024). This study examines how Spanish young people and adolescents perceive certain features that, according to academic literature, characterise information and disinformation. The findings will help to further understand the key factors shaping how this part of the population accesses information and, in turn, help clarify how they configure their media repertoires.

Status of the issue

Relationships with social media

The constant access to mobile phones and social media, linked to the need for social interaction and the search for entertainment, as well as the way digital behaviour is adapted to each platform, are some of the aspects that help explain the reality in which adolescents and young people live (García-Jiménez et al., 2020). At earlier ages, there is also a link with the risks posed by social media, such as cyberbullying, false information, certain addictions, and various aspects related to personal safety. It is similarly associated with opportunities in fields such as education and communication between individuals (Osepashvili, 2025).

Social media are presented as spaces where young people can share experiences, and where the content they consider relevant for keeping informed and for entertainment is increasingly found. In this respect, perceptions of the platforms depend on age, gender and family social class (López-de-Ayala et al., 2020). X (formerly Twitter) is a space for critical analysis of information and of social movements, especially climate change. In this space, adolescents defend their values and ideology. Meanwhile, Instagram provides a showcase for fashion and beauty, where brands promote an idealised and desirable lifestyle, and YouTube is geared towards entertainment (Lozano-Blasco et al., 2023).

Younger users display varying levels of trust depending on the type of content they access. The study by Neira Placer et al. (2025) shows that they trust content more on technical than on political platforms. Age has also been identified as a predictor of which platform they use: TikTok is more popular among younger users, while X is more common among older users. Similarly, perceptions of social media reliability are linked to age, with critical awareness increasing over time. Gender differences are also observed in the case of Instagram.

The detachment of news from traditional media is changing how citizens conceptualise it and how they distinguish it from other cultural objects, such as entertainment or advertising, particularly among younger generations. Currently, there are differences between what they recognise cognitively and affectively as news (Swart & Broersma, 2024). In addition, the decline of traditional media and news websites, compared with online aggregators, social networks, and video platforms, is evident. Public figures and influencers have already become significant actors in the public agenda. Fragmentation across social media platforms is increasing, including on WhatsApp (Newman et al., 2025).

It is in this context that the distinction between what constitutes journalism and what does not is becoming increasingly 'blurred' (Wunderlich et al., 2022), coinciding with the growing relevance of non-journalistic sources. Adolescents aged 15 to 17, for instance, perceive influencers as suitable media actors in staying well informed. This perception differs among young adults (18–25 years) and older adults (40–53 years), also in relation to the functions associated with journalistic and non-journalistic sources. Similarly, there are divergences in the level of trust placed in journalistic activity and in the various verification strategies applied, such that young people view journalistic work as reliable for staying informed and for cross-checking information, whereas not-strictly journalistic sources are associated with fulfilling social needs. Age continues to emerge as a key factor.

Young people tend to consume news incidentally, that is, through social media while using their phones. In a sense, they 'encounter' the news rather than actively seek it out, which ultimately makes such content indistinguishable from other types of content, and differentiates it from news consumption in other media contexts (Boczkowski et al., 2017). Nevertheless, some young people still actively seek out information on social media, often combining that with traditional media. Our research also draws on the theory known as 'the news finds me'. Indeed, this segment of the population considers itself well-informed just through the information they receive without actively searching for it, which can accordingly have personal and social consequences (Goyanes et al., 2023).

In Spain, under-24s obtain their news via social media, specifically: 28% on Instagram and 25% on TikTok (Sierra et al., 2025). In addition to high daily exposure to social media with its centrality as a news-consumption channel, there is generally little interest in verifying information, and it generates low levels of trust. Access is incidental, news is read sporadically, and it is often shared with contacts (Farias-Batlle et al., 2024). In this context, ideological stance, education and age all influence the credibility that adolescents and young people attribute to current news encountered on social media and its sources. However, according to Ceballos-del-Cid et al. (2025), while interest in news is increasing among these age groups, trust in social media remains lower than in traditional media (Neira Placer et al., 2025; Cureses, 2023),

likely due to the influence of parental opinions in certain age ranges, which emphasise the lower reliability of social media compared with conventional outlets (Smahel et al., 2020). At the same time, as young people get older, their ability to recognise certain mechanisms and biases in the media environment, such as clickbait, increases, which directly undermines credibility (Cureses, 2023; Pastor Ruiz et al., 2019).

Relationships with disinformation

Social media platforms present an environment in which the dissemination of misleading or false information is particularly easy (Bode & Vraga, 2018). This research also highlights the ease with which content goes viral, specifically the ability of influencers to spread information (Ouvrein et al., 2021) that may be incorrect, taking into account Kelman's (1974) theory of social influence. In fact, according to the findings of the work by Mulcahy et al. (2025), a greater virality of disinformation posts leads to a reduced perception of deceit, which in turn encourages the intention to share them. This, in its turn, could contribute to increased social radicalisation (Moskalenko & Romanova, 2022). Diaz Ruiz (2025) notes that disinformation is, in fact, an expected outcome rather than an anomaly within the economic logic of social media platforms.

Children aged 12 to 15 are already aware of the existence of false information. They are able to associate it with rumours, as well as posts and videos on social media. They also recognise that fake news can be generated and shared by individuals for ideological or economic purposes, linked to deception and intimidation (Vartiainen et al., 2023). Nevertheless, in some cases, exposure to disinformation prompts adolescents to consult traditional media to verify potentially false information (Selnes, 2024). Despite this, they constitute a vulnerable group because of the amount of time they spend on social media and the content they consume (Herrero-Diz et al., 2023).

Overall, susceptibility to both deliberate and inadvertent disinformation decreases throughout secondary education (Siani et al., 2024). The same trend applies to other educational attainment (Gómez-Calderón et al., 2023), with higher levels of education being associated with a greater ability to recognise false information and to cross-check and verify news. Other factors, such as urban/rural environments and political orientation, also appear to play a role. Likewise, perceptions of fake news on social media are not independent of context. As Shin et al. (2022) note, in the case of X, the assessment of whether a post is credible depends on whether it is received through a close contact and whether it has a high number of 'likes' and comments, compared with retweets from celebrities without these indicators. The evaluation of these attributes, as well as individual interactions, emerges as an important factor to consider.

In general terms, young people are aware that there is more disinformation on social media than in the traditional media (Galarza-Molina, 2023) and they understand that not all platforms carry such content to the same degree; for example, it would be more difficult to find disinformation among TikTok content than Facebook content. Other researchers refer to what is termed adolescents' 'invulnerability to disinformation'. In the study by

Papapicco et al. (2022), it was noted that adolescents are aware of the existence of fake news but find it harder to recognise or recall it. In fact, they tend to overestimate their ability to detect false information and often feel the need to share content with their network (Zozaya-Durazo et al., 2024). This connects with what Martínez-Costa et al. (2023) call the “perception that no one can deceive me”. This cognitive bias, which is more pronounced at higher levels of education, is based on the perception that they are capable of detecting false information and are more immune to it than others, even if they have not internalised certain valid criteria for assessing news quality (Herrero-Diz et al., 2021). Likewise, young people perceive older adults as being more likely to believe false information, a condition known as the third-person effect (Galarza-Molina, 2023; Martínez-Costa et al., 2023).

Regarding the sharing of information, adolescents are more likely to share news on WhatsApp if it relates to their personal interests, regardless of its veracity. They also share content based on the appearance of journalistic interest (Herrero-Diz et al., 2020). Furthermore, the tendency to share content without prior analysis has been linked to moral disengagement (Herrero-Diz et al., 2023). In addition, the study by Paciello et al. (2023) highlights that self-efficacy beliefs about dealing with disinformation are connected both with perceptions of the validation of journalistic content and with the avoidance of sharing information of uncertain reliability. There are also links with the capacity for reflection and self-efficacy in managing negative emotions. Some approaches emphasise the potential of sharing fake news to raise awareness among contacts about certain content (Valencia-Arias et al., 2023).

Finally, Trninić et al. (2022) highlight the need for education in the realm of social and digital media, given the complexity of the media environment that adolescents and young people face. Such an educational approach should also take into account communication and interpersonal skills when addressing disinformation and false information, not merely technical knowledge (Vissenberg et al., 2023).

Objective and research questions

This study aims to understand the perceptions of minors and young people aged 14 to 24 residing in Spain regarding four propositions related, on the one hand, to access to news via social media, and on the other, to the possible relationship between the viral spread of disinformation and these digital platforms. The propositions established to assess the level of agreement among the participants are:

1. Social media is an appropriate space for accessing news.
2. Consulting social media is sufficient to know what is happening in their environment and around the world.
3. The greatest spread of rumours and disinformation on the internet occurs via social media.
4. Continuous exposure to information on social media helps distinguish fake from real news.

Based on age and frequency of social media use as possible predictors, three research questions are proposed: the first relates to information (propositions 1 and 2) and the following two relate to the spread of disinformation (proposition 3) and its detection (proposition 4):

RQ1: Do minors and young people perceive social media as an appropriate channel for accessing news and staying informed?

RQ2: Do they believe that social media constitutes the main channel for the spread of disinformation?

RQ3: Do minors and young people believe that frequent social media use helps detect disinformation?

Methodology

This study adopts a dual approach, combining descriptive analysis with regression analysis, using four dependent variables corresponding to the four established propositions, and two independent variables: age and daily time spent on social media.

The quantitative questionnaire technique was applied to a sample drawn from the population of residents in Spain aged 14 to 24. To ensure representativeness, sex, age, and geographic region were taken into account. The final sample comprised 1,800 participants who completed a web-assisted interview via online panels, with fieldwork carried out between 28 October and 27 November 2024. The study was reviewed and approved by the Research Ethics Committee of Rey Juan Carlos University. The data were analysed using SPSS Statistics v.29.02.0.

To gauge young people's perceptions and level of agreement with the established propositions, a Likert scale from 1 to 5 was used, corresponding to the items 'Totally disagree', 'Somewhat agree', 'Moderately agree', 'Mostly agree', and 'Completely agree'. A sixth option, 'Don't know/No answer', was included and treated as missing data in the analysis. Regarding the independent variables, age was categorised into three groups: 14 to 16 years, 17 to 20 years, and 21 to 24 years. Daily social media use was divided into the categories 'Never', 'Less than 1 hour', '1 to 3 hours', '3 to 5 hours', and 'More than 5 hours'. However, for the regression analysis, these items were adjusted to avoid statistical instability, as only five participants (0.3% of minors and young people) reported not accessing social media at all. For this adjustment, the 'Never' category was merged with 'Less than 1 hour'.

In the first phase, a descriptive analysis was conducted to determine the frequency and distribution of the four propositions (dependent variables), using percentages, measures of central tendency, and standard deviation. Subsequently, the contingency associated with age and social media access time, established as predictors (independent variables), was calculated. To this end, the chi-squared (χ^2) coefficient was applied to determine whether significant differences exist and if a linear association can be detected within these differences.

In the second phase, an ordinal regression analysis was carried out, suitable for dependent variables measured on a Likert scale. In this phase, the goodness of fit for the two established predictors was analysed using the model's chi-squared significance and the Cox and Snell and Nagelkerke R-squared coefficients. Where results were statistically

significant, the Odds Ratio – Exp(B) and the regression coefficient B were calculated to determine the strength and direction of the association between variables. For this purpose, the highest-range items were taken as the reference groups for the predictors: the 21–24 age group for age and the ‘more than 5 hours’ group for daily social media access time.

Results

Frequency and distribution of the dependent variables (propositions)

Regarding the perception that ‘Social media is an appropriate space for accessing news’ (Proposition 1), a high level of agreement was recorded among young people, with a mean of 3.82 and a mode of 4, corresponding to the ‘Mostly agree’ item. The standard deviation ($\sigma = 1.162$) suggests a moderate level of dispersion, ranging roughly between 0.8 and 1.5. Therefore, there is some variability, but with a tendency towards higher levels of agreement. This trend is reflected in the 83.1% of young people who show the highest levels of agreement (25.6% ‘Moderately agree’, 30.3% ‘Mostly agree’, and 27.2% ‘Totally agree’). The remaining percentages do not exceed 10 points.

There is moderate consensus regarding young people’s perception that ‘Consulting social media is sufficient to know what is happening in their environment and around the world’ (Proposition 2). The mean obtained is $\bar{x} = 3.45$ and both the median and the mode are 3, values that indicate that most responses fall under ‘Moderately agree’, although with a general trend toward ‘Mostly agree’. In this case, the standard deviation is slightly higher, $\sigma = 1.237$, but remains at a moderate level, still far from complete dispersion.

In percentage terms, there is no clear linear trend. Although more than half fall within the categories of ‘Moderately agree’ (28.6%) and ‘Mostly agree’ (27.6%), the intermediate percentages are split evenly, with 16.9% in both ‘Somewhat agree’ and ‘Totally agree’. It is also notable that only 6% of young people categorically reject this proposition.

As for the variables relating disinformation to social media, significant results are also observed. First, regarding the idea that ‘The greatest spread of rumours and disinformation on the internet occurs via social media’ (Proposition 3), the measures of central tendency fall around the value corresponding to ‘Mostly agree’ ($\bar{x} = 4.05$; Mo = 4; Med = 4). The standard deviation is the lowest found and, although it still falls within the moderate range, it is the value showing the greatest concentration of results at the higher end ($\sigma = 1.023$).

This strong concentration in the most favourable categories is confirmed by the 67.6% of responses positioned as ‘Totally agree’ and ‘Mostly agree’. As support for the proposition decreases, the percentages also diminish progressively, from 21.5% in ‘Moderately agree’ to 1.2% of young people who reject it.

Secondly, regarding the idea that ‘Continuous exposure to information on social media helps distinguish fake from real news’ (Proposition 4), the mean value ($\bar{x} = 3.54$) lies almost exactly halfway between the items ‘Somewhat agree’ and ‘Mostly agree’. However, both the mode and the median coincide at 3, and the standard deviation shows the highest figure among the dependent variables established. The value $\sigma = 1.208$ indicates that, although still

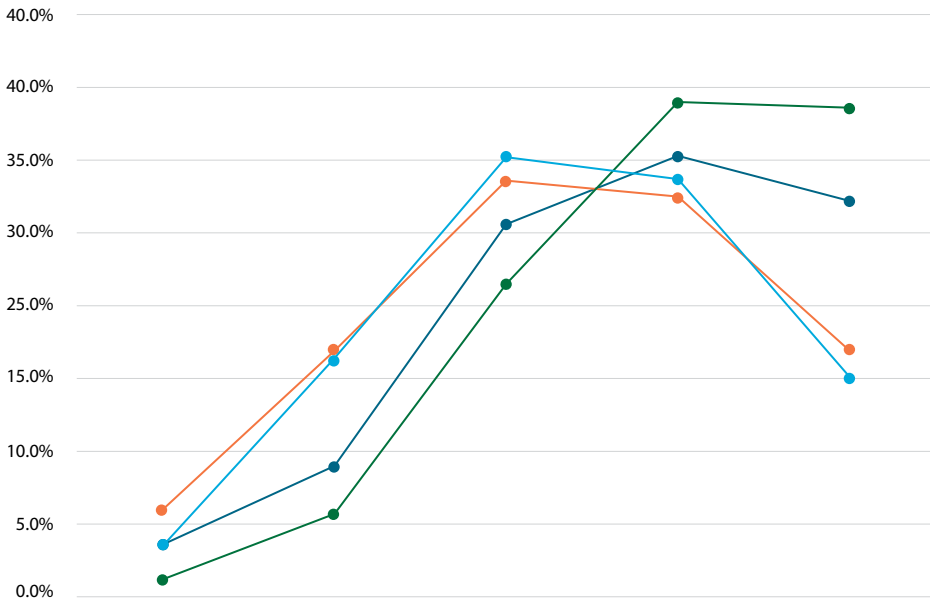


Figure 1:
Percentage distribution of the dependent variables established
Source: Compiled by the authors.

within a moderate range, there is notable dispersion, which is confirmed by the percentages recorded for each item.

In this respect, there is a slightly higher tendency among young people to be moderately in agreement (30.2%). However, the next highest level is only one and a half points lower. The remaining categories are distributed in very similar percentages between somewhat (16.4%) and totally agree (15.2%). Only 3.5% do not support this statement. Figure 1 and Table 1 show the summary of the frequency distribution values of the dependent variables and the statistical indexes obtained.

Table 1:
Central tendency measures and standard deviation values for the dependent variables

	Social media is an appropriate space for accessing news	Consulting social media is sufficient to know what is happening in their environment and around the world	The greatest spread of rumours and disinformation on the internet occurs via social media	Continuous exposure to information on social media helps distinguish fake from real news
Mean (\bar{X})	3.82	3.45	4.05	3.54
Median (Med)	4.00	3.00	4.00	3.00
Mode (Mo)	4	3	4	3
Standard deviation (σ)	1.162	1.237	1.023	1.208

Source: Compiled by the authors.

Contingency based on age and daily frequency of access to social media

In the two dependent variables relating to how social media are rated as a channel for accessing current affairs information, age shows statistically significant differences according to Pearson’s chi-squared test ($\chi^2 < 0.05$), with a linear trend that rises as the age group increases.

Looking closely at Proposition 1, that ‘Social media is an appropriate space for accessing news,’ a clearly significant linear association is found, with $p = 0.001$. This is reflected in the categories of mostly and totally agree. In the first of these, responses are distributed as follows: 28% among those aged 14 to 16, 33.3% in the intermediate group, and 38.7% among those aged 21 to 24. In the fully supportive category, the respective values are 20.3% in the youngest group, 37.6% in the intermediate group, and 42.1% among the oldest. No linear trend appears in the rest of the agreement scale, where the highest percentages occur among those aged 17 to 20.

Regarding the statement ‘Consulting social media is sufficient to know what is happening in their environment and around the world’ (Proposition 2), a linear association also emerges, though with a slightly more moderate p-value of 0.006. Except for the extreme categories of this proposition, all other categories show higher percentages as age increases. However, the differences between those aged 17–20 and 21–24 are only a few tenths of a percentage point across all agreement levels. These minimal variations confirm the moderate nature of the linear association. It is also relevant that the oldest respondents show the strongest support for this proposition, with 38.8% in the Mostly agree category and 39.1% expressing full support, although the latter is slightly lower than that recorded among those aged 17 to 20 (39.4%). By contrast, the 14 to 16 age group records more responses in the categories expressing the greatest disagreement with the proposition: 7.7% say they do not agree at all (compared with 5.7% and 5.0% in the other two groups) and 18.6% are in the slightly agree category, in contrast with 15.7% in the intermediate group and 16.9% among the oldest.

Significant differences are observed by age, with a chi-squared value below 0.05, in both propositions that relate disinformation to social media. However, no linear association is found ($p = 0.634$) regarding young people's perception that 'The greatest spread of rumours and disinformation on the internet occurs via social media'. All age groups clearly support this statement, especially those aged 14 to 16. In this group, 70.2% are mostly or totally in agreement, compared with 64.5% in the intermediate group and 69% among the oldest. The group showing the least support is the 17 to 20 age range, but in none of the groups do the categories of somewhat or not at all in agreement reach a combined 10%.

A moderate linear association ($p = 0.020$) does emerge in Proposition 4, that 'Continuous exposure to information on social media helps distinguish fake from real news', with support increasing as age rises. According to the percentages recorded, 16.1% of young people aged 21 to 24 choose the totally agree category, compared with 15.4% in the intermediate group and 13.2% among those aged 14 to 16. At the opposite end, the trend reverses, with greater rejection among the youngest: 4.8% in this group, compared with 3.4% and 2.6% in the older groups.

Table 2 shows the statistical significance and the linear-by-linear association of the contingencies between the dependent variables established and age.

On the other hand, there are statistically significant differences ($\chi^2 < 0.05$) in the two propositions that link information on current affairs with the frequency of access to social networks, and a linear association is also observed in both cases with $p = 0.001$. Regarding the first one (Social media is an appropriate space for accessing news), the level of support expressed by young people is positively related to their frequency of access to social networks, such that those who most frequently show full agreement are those who connect for more than five hours, with 38.4%. As the frequency of access decreases, this percentage also drops, reaching 19.8% among the young people who spend less than an hour on these platforms.

The linear association is also evident when examining the statement Consulting social media is sufficient to know what is happening in their environment and around the world. In this case, the trend is mainly found in the values for Totally disagree, which decrease as time spent on social media increases – from 8.7% among those who spend less than an hour, compared with 2.5% among those who connect for more than five hours. In contrast, those with the lowest frequency of access show a Mostly agree level of 23.4% and a Totally agree level of 16.8%, figures well below those found among the group with the highest frequency of access, which display respective percentages of 30% and 22.6%.

For the propositions that link social networks with disinformation, the chi-squared tests show different results. For the statement that 'The greatest spread of rumours and disinformation on the internet occurs via social media', Pearson's value does not indicate statistically significant differences. In contrast, such differences do appear in 'Continuous exposure to information on social media helps distinguish fake from real news' ($\chi^2 = 0.001$), with a linear association of $p = 0.001$. Consequently, greater access corresponds to greater agreement. This is reflected in the Totally agree item, which is selected by 20.9% of the young people who connect for more than five hours, a percentage that decreases as frequency drops, reaching 11.5% among those who access for less than an hour (Table 3).

*Table 2:
Chi-squared tests for the dependent variables by age*

	Pearson's Chi-squared			Linear-by-linear association		
	Value	DF	Asymptotic significance	Value	DF	Asymptotic significance
Social media is an appropriate space for accessing news	42.144	10	0.001	10.262	1	0.001
Consulting social media is sufficient to know what is happening in their environment and around the world	24.419	10	0.007	7.626	1	0.003
The greatest spread of rumours and disinformation on the internet occurs via social media	38.798	10	0.001	0.226	1	0.634
Continuous exposure to information on social media helps distinguish fake from real news	36.674	10	0.001	5.429	1	0.020

Source: Compiled by the authors.

*Table 3:
Summary of the chi-squared values obtained for the established propositions when cross-tabulated with daily time spent on social networks*

	Pearson's Chi-squared			Linear-by-linear association		
	Value	DF	Asymptotic significance	Value	DF	Asymptotic significance
Social media is an appropriate space for accessing news	61.936	20	0.001	24.741	1	0.001
Consulting social media is sufficient to know what is happening in their environment and around the world	47.746	20	0.001	13.963	1	0.001
The greatest spread of rumours and disinformation on the internet occurs via social media	22.232	20	0.328	0.733	1	0.392
Continuous exposure to information on social media helps distinguish fake from real news	55.765	20	0.001	20.645	1	0.001

Source: Compiled by the authors.

Age and frequency of access as predictive factors: Regression analysis

Based on the ordinal logistic regression models, it is observed that agreement with the statement ‘Social media is an appropriate space for accessing news’ (Proposition 1) is mainly conditioned by the frequency of access to networks, which shows a good fit ($\chi^2 = 40.540$; $p < 0.001$). Although age also shows statistically significant differences ($\chi^2 = 6.506$), the fit is poor ($p = 0.039$), so this link should be interpreted with caution. In addition, the low Cox and Snell and Nagelkerke Pseudo R-Squared indices (0.004 in both cases) show that age has a very weak association and minimally explains the fluctuations of the dependent variable. In contrast, daily access is more relevant to the level of agreement that social media is an appropriate space for accessing news, according to the Cox and Snell (0.025) and Nagelkerke (0.027) indices.

The regression coefficients for age indicate that the 14–16 group is less likely to agree than the older group, with a negative value ($B = -0.291$), an Odds Ratio of 0.75 and significance at $p = 0.012$. In the case of young people aged 17 to 20, no significant differences are observed compared with the 21–24 group ($p = 0.332$). Frequency of access explains variability more effectively, with a higher probability of agreement with proposition 1 among those who use networks for more than five hours a day, although the p value (0.073) is not significant for those who access between 3 and 5 hours. The probability of agreement decreases as time spent on these platforms decreases, with increasingly negative B values and decreasing $\text{Exp}(B)$ indices, data that indicate a clear relationship between frequency of use and degree of agreement (Table 4).

Table 4:
Overall statistics and group comparison for Proposition 1

Predictor	χ^2 model	p model	R^2_{CS}	R^2_N	Groups compared	B	Exp (B)	p-value	Significance
Age	6.506	0.039	0.004	0.004	14–16 vs. 21–24	-0.291	0.75	0.012	Yes
					17–20 vs. 21–24	-0.105	0.9	0.322	No
Daily access	40.540	< 0.001	0.025	0.027	Less than 1 hour vs. more than 5 hours	-0.891	0.41	< 0.001	Yes
					1 and 3 hours vs. more than 5 hours	-0.675	0.50	< 0.001	Yes
					3–5 hours vs. more than 5 hours	-0.257	0.77	0.073	No

Source: Compiled by the authors.

Age again proves to be a poorer predictor ($\chi^2 = 7.084$; $p = 0.029$) than frequency of access ($\chi^2 = 25.538$; $p < 0.001$) for the statement ‘Consulting social media is sufficient to know what is happening in their environment and around the world’ (Proposition 2). In both cases, a good fit is evident, with Pseudo R-Squared values explaining the variance based on both factors, mainly frequency of use, which reaches an index of 0.016 for Cox and Snell and 0.017 for the Nagelkerke coefficient. For age, the indices are lower ($R^2_{CS} = 0.004$ and $R^2_N = 0.005$).

Specifically for age, the 14–16 group [$B = 0.271$; $p = 0.019$; $\text{Exp}(B) = 0.763$] shows a lower probability of agreeing with this statement compared with the 21–24 age group. Those aged 17–20 do not show significant differences compared with the older group ($p = 0.970$). On the other hand, there is no linear regression for daily access to social media. While it is generally observed that lower usage corresponds with a lower probability of agreement, the lowest Odds Ratio is found in the group accessing networks between 1 and 3 hours [$\text{Exp}(B) = 0.57$]; this result, together with a negative B coefficient, indicates a 43% lower probability for this group to agree with Proposition 2. The only group that does not show statistically significant differences compared with the highest frequency group is the one accessing networks between 3 and 5 hours (Table 5).

For the statements linking disinformation to social media, in the first of them (The greatest spread of rumours and disinformation on the internet occurs via social media – Proposition 3), daily access time is not a significant predictor, with a chi-squared value greater than 0.05 ($p = 0.404$). Age is significant, although extremely moderately ($\chi^2 = 6.322$; $p = 0.042$), with a poor fit (goodness-of-fit tests with $p < 0.001$), and its effect on the level of agreement with the analysed statement is weak, as reflected in the Cox and Snell and

*Table 5:
Overall statistics and group comparison for Proposition 2*

Predictor	χ^2 model	p model	R^2_{CS}	R^2_N	Groups compared	B	Exp (B)	p-value	Significance
Age	7.084	0.029	0.004	0.005	14–16 vs. 21–24	-0.271	0.76	0.019	Yes
					17–20 vs. 21–24	0.004	1	0.970	No
Daily access	25.538	< 0.001	0.016	0.017	Less than 1 hour vs. more than 5 hours	-0.518	0.60	0.004	Yes
					1–3 hours vs. more than 5 hours	-0.569	0.56	< 0.001	Yes
					3–5 hours vs. more than 5 hours	-0.169	0.85	0.232	No

Source: Compiled by the authors.

Table 6:
Overall statistics and age comparison for Proposition 3

Predictor	χ^2 model	P model	R ² _{CS}	R ² _N	Groups compared	B	Exp (B)	p-value	Significance
Age	6.322	0.042	0.004	0.004	14–16 vs. 21–24	0.137	1.14	0.244	No
					17–20 vs. 21–24	-0.159	0.85	0.136	No
Daily access time	2.918	0.404	0.002	0.002					

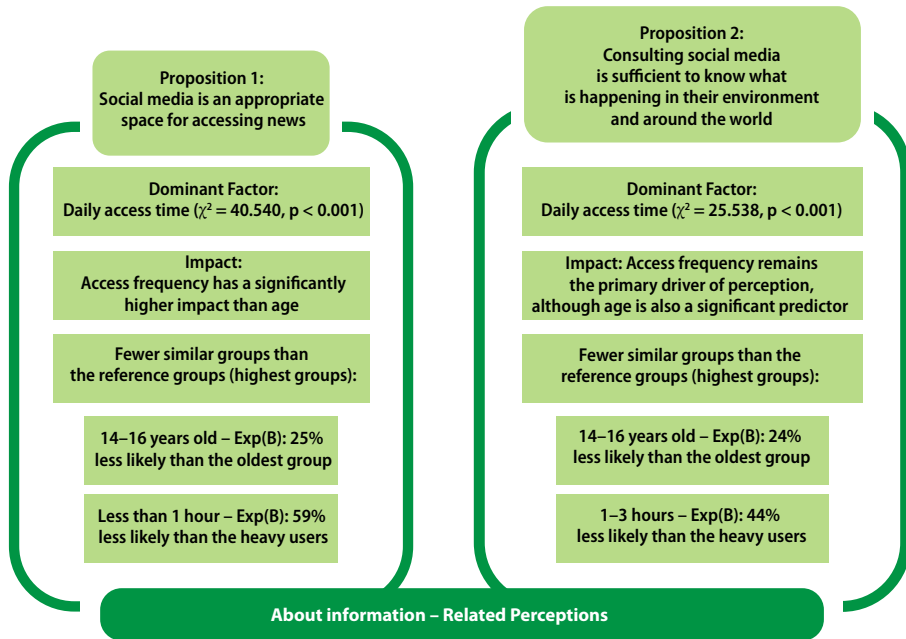
Source: Compiled by the authors.

Nagelkerke coefficients (0.004 in both cases). When both predictors are considered together, they also provide no significant value, exceeding the 0.005 threshold ($p = 0.058$).

The regression indices for age reinforce the weak effect it has on support for the statement that ‘The greatest spread of rumours and disinformation on the internet occurs via social media’ (Table 6). Although significance appeared in the overall model, none of the age groups are significant when using the 21–24 group as a reference. For the 14–16 group, the Odds Ratio is $\text{Exp}(B) = 1.14$, indicating a 14% higher probability of agreeing than the older group, but the values of $B = 0.137$ and $p = 0.244$ prevent this from being considered statistically significant. A similar pattern is observed for the 17–20 group. Table 6 details the overall coefficients but does not include a group comparison for daily access time, as the chi-squared value in the model does not show significant differences.

Regarding the second statement on the relationship between disinformation and social networks (Continuous exposure to information on social media helps distinguish fake from real news – Proposition 4), age is a good predictor of the level of agreement with this statement ($\chi^2 = 6.112$; $p = 0.047$), with a good model fit ($p = 0.114$ in the goodness-of-fit test), although the pseudo R-squared tests show low variability, with identical Cox and Snell and Nagelkerke values of 0.004. Daily access time to social networks is a stronger predictor than age ($\chi^2 = 30.484$; $p < 0.001$), although the model fit is not optimal ($p = 0.015$). Nevertheless, the Cox and Snell (0.020) and Nagelkerke (0.021) values indicate a somewhat stronger predictive capability than age.

The regression analysis shows that the 14–16 age group differs significantly from the 21–24 group ($p = 0.014$), with a 25% lower probability of agreeing with the statement [$B = -0.288$; $\text{Exp}(B) = 0.75$]. No significant differences are apparent in the 17–20 age group. Regarding daily access to social media, the 3–5 hours group is the only one that does not show any statistical differences compared with the highest-use group. In the other groups, significant values are recorded with $p < 0.001$. In these groups, a lower probability of agreement is observed as daily access time decreases. For young people accessing social media for less than 1 hour, there is a 48% lower probability of supporting Proposition 4 than those accessing more than five hours [$B = -0.654$; $\text{Exp}(B) = 0.52$]. This percentage decreases to 44% [$B = -0.583$; $\text{Exp}(B) = 0.56$] for those spending slightly more time on networks, between 1 and 3 hours daily (Table 7).



OUTLINE: INFLUENCING FACTORS ON PERCEPTIONS

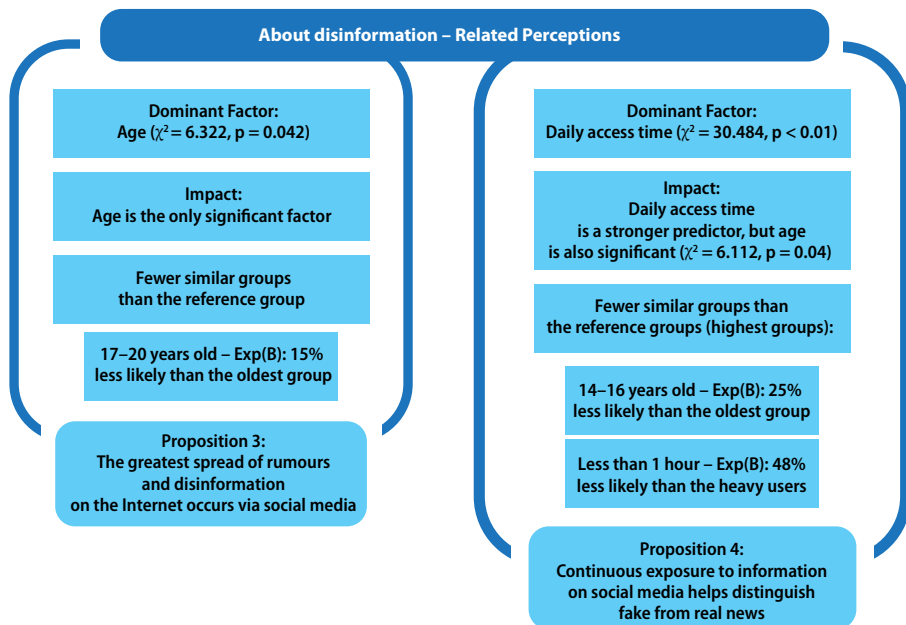


Figure 2: Outline: Influencing factors on perceptions

Source: Compiled by the authors.

*Table 7:
Overall statistics and group comparison for Proposition 4*

Predictor	χ^2 model	P model	R^2_{CS}	R^2_N	Groups compared	B	Exp (B)	p-value	Significance
Age	6.112	0.047	0.004	0.004	14–16 vs. 21–24	-0.288	0.75	0.014	Yes
					17–20 vs. 21–24	-0.095	0.91	0.371	No
Daily access time	30.484	< 0.001	0.020	0.021	Less than 1 hour vs. more than 5 hours	-0.654	0.52	< 0.001	Yes
					1–3 hours vs. more than 5 hours	-0.583	0.56	< 0.001	Yes
					3–5 hours vs. more than 5 hours	-0.144	0.87	0.319	No

Source: Compiled by the authors.

Conclusions and discussion

The main contribution of this research relates to the study of age and frequency of access to social media as potential predictors in terms of information consumption, the spread of disinformation, and the ability to detect it.

Regarding the first research question – whether minors and young people perceive social media as an appropriate channel for accessing news – it can be stated that eight out of ten respondents fall within the three highest levels of agreement. This indicates a widespread acceptance of social media as a preferred information channel among young people. Although the recorded values suggest overall acceptance, there is a slight tendency towards the higher levels of agreement, which could be interpreted as growing trust in social networks as a primary source of information. This finding aligns with other studies on this issue (Newman et al., 2025; Sierra et al., 2025; Swart & Broersma, 2024). While the perception of social media as a suitable vehicle for consulting news achieves majority consensus, it was moderated if consumption was limited exclusively to this channel. In this regard, the contribution of our study lies in identifying through regression models that frequency of access to social media is a stronger predictor than age in determining whether adolescents and young people consider social networks an appropriate place to keep them informed.

Furthermore, regarding whether young people feel informed based on their consumption through social media, the fact that only 6% categorically reject this proposition indicates that negative perceptions about the informative capacity of social networks are in the minority. In this regard, age emerges as a less consistent predictor than frequency of access.

In relation to the second research question – whether young people associate the viralisation of disinformation with social networks – the high concentration of responses at the higher levels of agreement and the low dispersion demonstrate a consensus among the

young people surveyed. This allows us to conclude that, for adolescents and young adults, social media are the main channel for the dissemination of false information. The results indicate that more than two-thirds of respondents fall between 'Mostly agree' and 'Totally agree', revealing a widespread perception among young people of the prominent role of social media in spreading disinformation.

However, the results regarding the ability to discern between false and real information through continued access to social media show greater dispersion. This disparity indicates that, while some young people believe that frequent use of social networks can improve their information judgment, others do not share this view. Likewise, the percentage distribution of responses shows significant fragmentation, with similar percentages across the different levels of agreement and disagreement. We therefore find that those who are more predisposed to accessing social media – platforms that facilitate the spread of misinformation (Bode & Vraga, 2018) – tend to rely excessively on their own judgement to distinguish between truthful and untruthful content. These results are consistent with the idea of cognitive bias described by Kruger and Dunning (1999), according to which people with little skill in a given area tend to overestimate their own competence. This type of bias must be taken into account first and foremost in studies on perceptions of disinformation and, at the same time, in media literacy programmes that may assume adolescents and young people have skills that are not demonstrated in everyday life.

In this regard, access to social media cannot, in any respect, be associated with greater digital literacy. This research highlights the importance of fostering digital competencies that enable young people not only to identify false content but also to understand the mechanisms that make it viral, thereby promoting a more conscious and responsible use of digital platforms. Frequent access to information does not, by itself, guarantee greater competence; rather, it requires specific skills to evaluate sources, identify biases, or cross-check data, and incidental consumption impairs the ability to discern between different types of content (Boczkowski et al., 2017).

Since young people perceive themselves as properly informed solely through the information they receive, a false sense of informational competence can take hold. The high level of agreement among those who most strongly identify with the idea that social media is an ideal space for accessing news (Proposition 1) aligns with the findings of Campbell and Hawkins (2025) regarding the common perception among social media users of the 'news finds me' effect. The intensity of this perception is likewise associated with a greater tendency to believe false information, as users feel they are inured to such untruthful content. According to our findings, those who have greater access to social media are also more likely to agree that continued exposure to information on these platforms helps them distinguish between false and real content (Proposition 4).

This line of reasoning is concerning given the high level of exposure to information that young people experience on social media. At this point, it is important to distinguish between the quantity of exposure and the quality of cognitive processing, particularly among young populations who are still consolidating their critical skills. Regarding predictors that could anticipate the association of disinformation viralisation with social media, the results rule out daily access time but do confirm age, albeit in a very limited way.

Concerning the third research question – whether age and time spent on social media are good predictors of young people linking these platforms with information and/or disinformation – it is observed that frequency of access constitutes a more significant predictive factor than age regarding the level of agreement on the suitability of these platforms for accessing news. As daily social network usage decreases, so too does the probability of regarding these environments as appropriate for staying informed. Nevertheless, age remains a relevant factor in perceiving social networks as an informational channel, as agreement percentages regarding this proposition increase progressively from younger to older participants.

These findings suggest that maturity may be associated with a higher valuation of social media as an information source, possibly due to a greater capacity for critical selection of these platforms within their news routines. Considering the results obtained, in addition to age, gender, ideology, and educational level are presented as determining factors that influence the perception of fake news. Thus, older age and higher educational level would be associated with a greater ability to recognise and verify disinformation. When young people reach adulthood, their perception of disinformation appears to be more firmly established, and factors other than age – such as cultural differences or the frequency of social media use – begin to shape their perceptions.

In any case, our results indicate that age remains a determining factor in the perception of disinformation, and that Media and Information Literacy (MIL) interventions adapted to each age group are necessary. In this regard, Suárez-Perdomo et al. (2025) propose cognitive-intellectual techniques aimed at adolescents, focusing on searching for and evaluating sources, as well as axiological techniques for identifying biases in media content. For university students, Turpo-Gebera et al. (2025) argue that Media and Information Literacy must necessarily be linked to civic education that reinforces critical engagement.

Furthermore, the perception of adolescents and young people regarding the viral spread of rumours and disinformation on social media shows clear differences across age groups, but without an evident linear association. The uniformity in responses is reflected in the recognition of the problem, which is high across all age groups, indicating a widespread awareness of informational risks in digital environments. This finding suggests that, regardless of age, there is a shared understanding of the role of social media in the dissemination of false content, which supports the relevance of promoting educational interventions focused on media literacy from an early age.

Finally, based on our results, the preference for social media as an information source challenges traditional media and educational actors to foster critical skills in news literacy. It is necessary to reflect on the role that both traditional media and digital platforms should play in the critical formation of young people regarding the information they consume, since the quality, veracity, and diversity of the content circulating in these environments will, in turn, shape the public opinion of this more easily influenced population segment.

This study offers robust results based on a broad and representative sample of young people's perceptions of information and disinformation in Spain. However, its restriction to a single country constitutes a limitation, as it is difficult to extrapolate our findings to other regions of the world when taking into account data reported in other studies and international reports. Indeed, when comparing our results with those recorded in the European Parliament's

Eurobarometer (2025), we observe that young Spaniards follow a pattern similar to that of most of their counterparts across the continent. Nonetheless, they diverge from the trends observed among young people in neighbouring countries (France, Italy and Portugal), who show a slightly stronger preference for television as a source of information

In Latin America, a region with close historical and cultural ties to Spain, the trend is very similar, as reported in the *Digital News Report* (Newman et al., 2025). According to this report, young people primarily access information through social media and content creators (influencers), relegating traditional media (television, radio and print press) to a secondary position. However, in line with our findings, they are largely aware of the disinformation circulating on social media. Latin Americans, however, express greater concern about not knowing how to detect disinformation and, overall, are more worried about this issue than Europeans. In summary, the trends observed in other countries within Spain's geographical or historical-cultural sphere reveal both similarities and differences. Therefore, as a continuation of this study, we suggest establishing a transnational comparison addressing the premises analysed in this research. Such an approach would provide valuable insights into the behaviour and information ecosystem developed by young people from a broader perspective.

References

- Anderson, K. J. (2026). What Audiences Do with News: A Broader Definition of News Consumption. *Media International Australia*, 198(1), 228–244. Online: <https://doi.org/10.1177/1329878X241270608>
- Anter, L. & Kümpel, A. S. (2025). Young Adults' Information Needs, Use, and Understanding in the Context of Instagram: A Multi-Method Study. *Digital Journalism*, 13(6), 1112–1130. Online: <https://doi.org/10.1080/21670811.2023.2211635>
- Boczkowski, P. J., Mitchelstein, E. & Matassi, M. (2017). *Incidental News: How Young People Consume News on Social Media*. Proceedings of the 50th Hawaii International Conference on System Sciences, (pp. 1785–1792). Online: <https://doi.org/10.24251/HICSS.2017.217>
- Bode, L. & Vraga, E. K. (2018). See Something, Say Something: Correction of Global Health Misinformation on Social Media. *Health Communication*, 33(9), 1131–1140. Online: <https://doi.org/10.1080/10410236.2017.1331312>
- Campbell, S. W. & Hawkins, I. (2025). Social (Media) Psychology of the 'News-Finds-Me' Perception: Habits, Mindsets, and Beliefs. *Journal of Computer-Mediated Communication*, 30(5). Online: <https://doi.org/10.1093/jcmc/zmaf010>
- Ceballos-del-Cid, Y., Gómez-Calderón, B. & Córdoba-Cabús, A. (2025). Social Media and News Consumption among Young Spaniards: A Diachronic Analysis (2021–2023). *Revista Mediterránea de Comunicación/Mediterranean Journal of Communication*, 16(1). Online: <https://www.doi.org/10.14198/MEDCOM.28010>
- Chou, Y.-L., Lee, H.-J., Tsai, J., Liang, E.-C. & Chang, Y.-J. (2023). *I Like Their Autonomy and Closeness to Me. Uncovering the Perceived Appeal of Social-Media Influencers*. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–19. Online: <https://doi.org/10.1145/3544548.3580898>
- Cureses, M. M. (2023). Boomers versus Millennials y Generación Z: Alfabetización digital y redes sociales. *Ámbitos. Revista Internacional de Comunicación*, (61), 116–131. Online: <https://doi.org/10.12795/Ambitos.2023.i61.07>

- Díaz Ruiz, C. (2025). Disinformation on Digital Media Platforms: A Market-Shaping Approach. *New Media & Society*, 27(4), 2188–2211. Online: <https://doi.org/10.1177/14614448231207644>
- European Parliament (2025). *Eurobarometer. Youth Survey 2024*. Online: <https://europa.eu/eurobarometer/api/deliverable/download/file?deliverableId=96862>
- Fariás-Batlle, P., Córdoba-Cabús, A. & Gómez-Calderón, B. (2024). Jóvenes y redes sociales: hábitos de consumo informativo y credibilidad de las noticias. *Comunicar*, 32(78), 155–165. Online: <https://doi.org/10.58262/V32I78.13>
- Galarza-Molina, R. (2023). Youth in the Face of Disinformation: A Qualitative Exploration of Mexican College Students' Attitudes, Motivations, and Abilities around False News. *Communication & Society*, 36(2), 97–113. Online: <https://doi.org/10.15581/003.36.2.97-113>
- García-Jiménez, A., López-de-Ayala López, M. C. & Montes-Vozmediano, M. (2020). Características y percepciones sobre el uso de las plataformas de redes sociales y dispositivos tecnológicos por parte de los adolescentes. *Zer*, 25(48), 269–286. Online: <https://doi.org/10.1387/zer.21556>
- Gómez-Calderón, B., Córdoba-Cabús, A., & López-Martín, A. (2023). Las fake news y su percepción por parte de los jóvenes españoles: el influjo de los factores sociodemográficos. *Doxa Comunicación*, 36, 19–42. Online: <https://doi.org/10.31921/doxacom.n36a1741>
- Goyanes, M., Ardèvol-Abreu, A. & Gil de Zúñiga, H. (2023). Antecedents of News Avoidance: Competing Effects of Political Interest, News Overload, Trust in News Media, and “News Finds Me” Perception. *Digital Journalism*, 11(1), 1–18. Online: <https://doi.org/10.1080/21670811.2021.1990097>
- Harff, D. & Schmuck, D. (2025). Who Relies on Social Media Influencers for Political Information? A Cross-Country Study Among Youth. *The International Journal of Press/Politics*, 30(3), 841–864. Online: <https://doi.org/10.1177/19401612241234898>
- Hendrickx, J. (2025). ‘Normal News is Boring’: How Young Adults Encounter and Experience News on Instagram and TikTok. *New Media & Society*, 27(10), 5736–5754. Online: <https://doi.org/10.1177/14614448241255955>
- Herrero-Diz, P., Conde-Jiménez, J. & Reyes de Cózar, S. (2020). Teens' Motivations to Spread Fake News on WhatsApp. *Social Media + Society*, 6(3). Online: <https://doi.org/10.1177/2056305120942879>
- Herrero-Diz, P., Conde-Jiménez, J. & Reyes-de-Cózar, S. (2021). Spanish Adolescents and Fake News: Level of Awareness and Credibility of Information. *Culture and Education*, 33(1), 1–27. Online: <https://doi.org/10.1080/11356405.2020.1859739>
- Herrero-Diz, P., Sánchez-Martín, M., Aguilar, P. & Muñoz-Velázquez, J. A. (2023). La vulnerabilidad de los adolescentes frente a la desinformación. *Revista Española de Pedagogía/ Spanish Journal of Pedagogy*, 81(285), 317–336. Online: <https://www.jstor.org/stable/48729279>
- Kelman, H. C. (1974). Further Thoughts on the Processes of Compliance, Identification, and Internalization. In J. T. Tedeschi (Ed.), *Social Power and Political Influence* (pp. 125–171). Routledge.
- Kruger, J. & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. Online: <https://psycnet.apa.org/doi/10.1037/0022-3514.77.6.1121>
- López-de-Ayala, M. C., Vizcaíno-Laorga, R. & Montes-Vozmediano, M. (2020). Social Network Practices and Attitudes of Young People: The Influence of Sex, Age, and Social Class. *Profesional de la Información*, 29(6). Online: <https://doi.org/10.3145/epi.2020.nov.04>
- Lozano-Blasco, R., Mira-Aladrén, M. & Gil-Lamata, M. (2023). Redes sociales y su influencia en los jóvenes y niños: Análisis en Instagram, Twitter y YouTube. *Comunicar*, 31(74), 125–137. Online: <https://doi.org/10.3916/C74-2023-10>

- Martínez-Costa, M. P., López-Pan, F., Buslón, N. & Salaverría, R. (2023). Nobody-Fools-Me Perception: Influence of Age and Education on Overconfidence about Spotting Disinformation. *Journalism Practice*, 17(10), 2084–2102. Online: <https://doi.org/10.1080/17512786.2022.2135128>
- Moskalenko, S. & Romanova, E. (2022). Deadly Disinformation: Viral Conspiracy Theories as a Radicalization Mechanism. *The Journal of Intelligence, Conflict, and Warfare*, 5(2), 129–153. Online: <https://doi.org/10.21810/jicw.v5i2.5032>
- Mulcahy, R., Barnes, R., de Villiers Scheepers, R., Kay, S. & List, E. (2025). Going Viral: Sharing of Misinformation by Social Media Influencers. *Australasian Marketing Journal*, 33(3), 296–309. Online: <https://doi.org/10.1177/14413582241273987>
- Neira Placer, P., Zozaya, L., Sádaba, C. & Feijoo, B. (2025). Percepciones de adolescentes españoles sobre la fiabilidad de la información en redes sociales. *Revista OBETS*, 20(1), 119–140. Online: <https://doi.org/10.14198/obets.27600>
- Newman, N., Ross Arguedas, A., Robertson, C. T., Nielsen, R. K. & Fletcher, R. (2025). *Digital News Report 2025*. Reuters Institute for the Study of Journalism. Online: <https://doi.org/10.60625/risj-8qqf-jt36>
- Osepashvili, D. (2025). The Risks and Opportunities of Use of Social Media: Youth Perceptions. *Studies in Media and Communication*, 13(1), 452–461. Online: <https://doi.org/10.11114/smc.v13i1.7495>
- Ouvrein, G., Pabian, S., Giles, D., Hudders, L. & De Backer, C. (2021). The Web of Influencers. A Marketing-Audience Classification of (Potential) Social Media Influencers. *Journal of Marketing Management*, 37(13–14), 1313–1342. Online: <https://doi.org/10.1080/0267257X.2021.1912142>
- Paciello, M., Corbelli, G. & D’Errico, F. (2023). The Role of Self-Efficacy Beliefs in Dealing with Misinformation among Adolescents. *Frontiers in Psychology*, 14. Online: <https://doi.org/10.3389/fpsyg.2023.1155280>
- Papapicco, C., Lamanna, I. & D’Errico, F. (2022). Adolescents’ Vulnerability to Fake News and to Racial Hoaxes: A Qualitative Analysis on Italian Sample. *Multimodal Technologies and Interaction*, 6(3). Online: <https://doi.org/10.3390/mti6030020>
- Pastor Ruiz, Y., Martín Nieto, R. & Montes Vozmediano, M. (2019). Patrones de uso, control parental y acceso a la información de los adolescentes en la red. *Estudios sobre el Mensaje Periodístico*, 25(2), 995–1012. Online: <https://doi.org/10.5209/esmp.64821>
- Selnes, F. N. (2024). Fake News on Social Media: Understanding Teens’ (Dis) engagement with News. *Media, Culture & Society*, 46(2), 376–392. Online: <https://doi.org/10.1177/01634437231198447>
- Shin, I., Wang, L. & Lu, Y. T. (2022). Twitter and Endorsed (Fake) News: The Influence of Endorsement by Strong Ties, Celebrities, and a User Majority on Credibility of Fake News During the Covid-19 Pandemic. *International Journal of Communication*, 16, 2573–2595. Online: https://pure.mpg.de/rest/items/item_3383320/component/file_3383321/content
- Siani, A., Joseph, M. & Dacin, C. (2024). Susceptibility to Scientific Misinformation and Perception of News Source Reliability in Secondary School Students. *Discover Education*, 3(1), 93. Online: <https://doi.org/10.1007/s44217-024-00194-8>
- Sierra, A., Labiano-Juanguarcía, R., Novoa Jaso, M. F. & Vara Miguel, A. (2025). *Digital News Report España 2025*. Universidad de Navarra. Online: <https://doi.org/10.15581/019.2025>
- Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S. & Hasebrink, U. (2020). *EU Kids Online 2020: Survey Results from 19 Countries*. EU Kids Online. Online: <https://doi.org/10.21953/lse.47fdeqj01of0>
- Suárez-Perdomo, A., Garcés-Delgado, Y. & Arvelo-Rosales, C. N. (2025). Systematic Review on Adolescents’ Construction of Media Literacy in Educational Settings. *Review of Education*, 13(1). Online: <https://doi.org/10.1002/rev3.70048>

- Swart, J. & Broersma, M. (2024). What Feels Like News? Young People's Perceptions of News on Instagram. *Journalism*, 25(8), 1620–1637. Online: <https://doi.org/10.1177/14648849231212737>
- Tamboer, S. L., Kleemans, M. & Daalmans, S. (2022). 'We are a Neecew Generation': Early Adolescents' Views on News and News Literacy. *Journalism*, 23(4), 806–822. Online: <https://doi.org/10.1177/1464884920924527>
- Trninić, D., Kuprešanić Vukelić, A. & Bokan, J. (2022). Perception of "Fake News" and Potentially Manipulative Content in Digital Media – A Generational Approach. *Societies*, 12(1). Online: <https://doi.org/10.3390/soc12010003>
- Turpo-Gebera, O., Rosales-Márquez, C., Gutiérrez-Aguilar, O. & Rivera-Mansilla, E. (2025). Media and Information Literacy and Citizenship Education in University Students. *Revista Latina de Comunicación Social*, (83), 1–23. Online: <https://doi.org/10.4185/RLCS-2025-2447>
- Valencia-Arias, A., Arango-Botero, D. M., Cardona-Acevedo, S., Paredes Delgado, S. S. & Gallegos, A. (2023). Understanding the Spread of Fake News: An Approach from the Perspective of Young People. *Informatics*, 10(2). Online: <https://doi.org/10.3390/informatics10020038>
- Vartiainen, H., Kahila, J., Tedre, M., Sointu, E. & Valtonen, T. (2023). More than Fabricated News Reports: Children's Perspectives and Experiences of Fake News. *Journal of Media Literacy Education*, 15(2), 17–30. Online: <https://doi.org/10.23860/JMLE-2023-15-2-2>
- Vissenberg, J., De Coninck, D., Mascheroni, G., Joris, W. & d'Haenens, L. (2023). Digital Skills and Digital Knowledge as Buffers Against Online Mis/Disinformation? Findings from a Survey Study among Young People in Europe. *Social Media + Society*, 9(4). Online: <https://doi.org/10.1177/20563051231207859>
- Wunderlich, L., Hölig, S. & Hasebrink, U. (2022). Does Journalism Still Matter? The Role of Journalistic and Non-Journalistic Sources in Young People's News Related Practices. *The International Journal of Press/Politics*, 27(3), 569–588. Online: <https://doi.org/10.1177/194016122211072547>
- Zozaya-Durazo, L. D., Sádaba-Chalezquer, C. & Feijoo-Fernández, B. (2024). "Fake or Not, I'm Sharing It": Teen Perception about Disinformation in Social Networks. *Young Consumers*, 25(4), 425–438. Online: <https://doi.org/10.1108/YC-06-2022-1552>

Hijacked Victimhood: The Political Communication Strategy of Vice President Sara Duterte at Congress Hearings

Audrey Buenavista Morallo*

* College of Education, University of the Philippines Diliman, Quezon City, REPUBLIC OF THE PHILIPPINES
E-mail: abmorallo@outlook.up.edu.ph

This article examines the way in which hijacked victimhood was used to evade accountability and preserve the power of Vice President Sara Duterte-Carpio at three congressional hearings. To achieve its goals, the study used the concept of indexicals to identify linguistic forms that signal a repurposing of the traditional concept of victimhood. The study found five discursive devices of victimhood employed to demonise critics, deflect their criticism, and elicit public sympathy and support. Through these discursive strategies, the vice president avoided critical scrutiny of her budget during the hearings, which could have caused her legal difficulties. She launched personal attacks against her opponents through her answers, a tactic that appears in populist rhetoric. The results highlight the need for critical vigilance and analysis of political communication in the Philippines to ensure that concepts traditionally associated with campaigns for justice for actual victims of harm are not reconfigured to serve partisan purposes.

Keywords: hijacked victimhood, strategic victimhood, Philippine politics, indexicals

Article received on 12 March 2025. Article accepted on 1 December 2025.

Conflict of interest: The author declares no conflicts of interest.

Funding: The author received no financial support for the research, authorship, or publication of this article.

Introduction

Over the years, considerable attention has been paid to the concept of victimhood in politics and political discourse (e.g. Al-Ghazzi, 2021; Armaly & Enders, 2022; Banet-Weiser, 2021; Hronešová & Kreiss, 2024; Neville-Shepard, M. & Neville-Shepard, R., 2022). This article contributes to the literature on victimhood in contemporary political discourse by investigating how victimhood was repurposed and hijacked in the Philippines during the 2024 budget deliberations for the Office of the Vice President (OVP) in order to avoid accountability and preserve the dominant group's grip on power (Hronešová & Kreiss, 2024). Specifically, this paper attempts to demonstrate that the language used by Vice President Sara Duterte-Carpio during congressional hearings on the budget for her office was an *indexical* of this use of the concept of *strategic victimhood* (i.e. hijacked victimhood) (Hronešová & Kreiss, 2024) and of the contestation, preservation and expansion of the dominant group's hold on power in the Philippines. Through her language, Vice President Duterte-Carpio encoded *tactical vulnerability* (Chouliaraki, 2021) and employed victimhood as a powerful discursive device in her rhetoric (Chouliaraki & Banet-Weiser, 2021). Investigating victimhood in this context is important because, while the literature has extensively discussed strategic and hijacked victimhood, little is known about how these operate in Southeast Asian populist democracies such as the Philippines, or indeed about how strategic victimhood is encoded in institutional discourse, such as in congressional testimony.

At the centre of this project lies the concept of victimhood, which serves as a framework for affective communication, turning public discussions into a battleground over differing assertions of suffering and the quest for their acknowledgment (Chouliaraki, 2021). Recognition refers to the successful procurement of the moral values associated with the vulnerable (Chouliaraki, 2021, p. 17). Chouliaraki further argues that victimhood is a 'dispersed and mutable' communication framework that uses claims of traumatisation and injury to show the self as the sufferer and to invite sympathy from others. Claims of victimhood are made in a symbolic field where there are contending powers and abilities for rights violations (Hronešová & Kreiss, 2024).

The concept of victimhood is also central to politics and contemporary political messages (Armaly & Enders, 2022, p. 1605). Today, many points surrounding political debates centre on who victims are (Chouliaraki & Banet-Weiser, 2021), on who or what the source of that victimisation is, and on who can address the injuries incurred through victimhood. The feeling of being a victim in politics is related to attitudes toward the establishment, political efficacy, personality traits, and support for particular candidates, among other factors (Armaly & Enders, 2022). However, this does not mean that the use of victimhood is new (Zembylas, 2021). It has been a dominant discourse in politics and the history of trauma, pain and resentment in the 20th century (Chouliaraki, 2021).

In one way or another, politicians have capitalised on victimhood, especially in rhetoric, in making their cases to their electorates (Armaly & Enders, 2022). Armaly and Enders noted that some political actors, for example, portrayed ordinary citizens as victims of specific and abstract policies and circumstances such as high tax rates, exorbitant healthcare costs, and unfair taxation. In the process of this portrayal, these politicians would also present themselves as being in the best position to address those problems. Viewed from one perspective,

the preeminent place of victimhood in politics and political discourse is not necessarily a bad idea, as any electorate would want politicians to address its needs and develop the conditions for a happy and satisfying life (Armaly & Enders, 2022). What is problematic, however, is the extent to which politicians are willing to exploit victimhood, severely threatening a civilised democratic system (Armaly & Enders, 2022), and one controversial form requiring serious consideration and investigation is that of hijacked victimhood.

Hijacked victimhood is a form of strategic victimhood that is commonly employed by interest groups to claim victimhood status in power and legitimacy struggles, and, as Hronešová & Kreiss (2024) go on to explain, strategic victimhood can be a politically valuable construction of harm for the achievement of equity and justice. However, they also recognised that this tool could be co-opted and hijacked, inverting moral relations and perverting an understanding of harm to defend those in powerful positions (Hronešová & Kreiss, 2024). The term hijacked victimhood – a co-optation of strategic victimhood – is used by dominant groups to show themselves as imperilled and in need of protection at the expense of those in precarious situations (Neville-Shepard, M. & Neville-Shepard, R., 2022). Thus, hijacked victimhood is a way of subverting the moral power usually accrued by victims by undermining moral hierarchies (Hronešová & Kreiss, 2024). This hijacking and the co-optation of victimhood are evident in the rhetoric of populist leaders such as U.S. President Donald J. Trump and Hungarian Prime Minister Viktor Orbán to mobilise support and justify aggressive actions (Hronešová & Kreiss, 2024; Neville-Shepard, M. & Neville-Shepard, R., 2022; Patrona, 2025).

Hijacked victimhood, particularly in political discourse, can thus help us to understand the language of populists such as Vice President Duterte-Carpio's father, former Philippine President Rodrigo R. Duterte. Although a precise definition of populism still evades experts (Curato, 2016), some see populism as an ideology that splits society into two homogeneous and opposing groups – pure people and a corrupt elite – asserting that politics should reflect the will of the people (Mudde, 2004). For others, populism is better defined by its social base, which is mostly composed of those who are frustrated by their diminishing space in society (Lipset, 1960, as cited in Curato, 2016; Moffitt, 2015, as cited in Curato, 2016). Former Philippine President Rodrigo Duterte employed an illiberal populist law and order message when he campaigned for and won the Philippine presidency in May 2016. Vice President Duterte-Carpio's ascent to the second highest position in the Philippines in 2022 can be seen as riding on the back of the populism of former President Duterte, who was still popular with the Filipino electorate at the time of the election. In her own right, Vice President Duterte has generally dovetailed with the populist law and order message of her father and cultivated a tough politician persona. Iglesias (2022) argued that the elder Duterte played a significant role in the victory of his daughter and her running mate, Philippine President Ferdinand Marcos, Jr., by creating an environment of violence and impunity.

The concept of victimhood can help in investigating the way, in her testimony before Congress, Vice President Duterte-Carpio – a member of one of the country's elite political families – indexed hijacked victimhood to portray herself as being under attack by legislators whom she accused of politicising the budget hearing of her office. Such analysis became all the more important following Duterte-Carpio's impeachment at the Philippine House of Representatives in early 2025 on allegations of misuse of public funds and of threatening

to have Marcos, Jr., the First Lady and the Speaker of the House assassinated should she be killed (Elemia & Wee, 2025). That trial did not come to fruition owing to the Supreme Court decision to declare the process unconstitutional (Marcelo, 2025). Her portrayal of herself as being under attack from members of Congress could be considered a form of ‘tactical vulnerability’ – a form of suffering claimed by individuals for personal gain or benefit (Chouliaraki, 2021). Through all these deployed strategies, she was able to flip the victim–victimiser dichotomy, an action that reverses the moral order and mobilises real or perceived security threats (Hronešová & Kreiss, 2024). By encoding tactical vulnerability and turning the victim–victimiser dichotomy on its head, she might have been able to preserve, expand and protect the budget of her office, which was already substantially more significant than that of her predecessor (Panti, 2024). Furthermore, she hoped to evade scrutiny of her office’s budget utilisation, especially the way it had spent around ₱125 million in confidential funds over 11 days in 2022 (Crisostomo, 2023). All of these had the goal of protecting and preserving her own and her family’s power. To better understand how Vice President Duterte-Carpio’s language indexed hijacked victimhood in her congressional testimony to protect and preserve power, this study answered the following questions:

1. Which indexicals of hijacked victimhood did Vice President Sara Duterte-Carpio use in her testimony before Congress during her office’s budget hearings to portray herself as a victim?
2. How did the reversal of the victimiser–victim dichotomy allow Vice President Sara Duterte-Carpio to avoid scrutiny of her actions and contest, preserve and expand her power?

Understanding how political figures hijack victimhood is crucial because such rhetorical strategies reshape the way in which accountability, justice and power are perceived in democratic societies. When powerful actors claim victimhood, they blur the moral distinction between genuine victims and political elites (Chouliaraki & Banet-Weiser, 2021; Hronešová & Kreiss, 2024), undermining public empathy for those who suffer real harm and eroding trust in democratic institutions. These hijacked narratives transform legitimate scrutiny into perceived persecution, weakening mechanisms of democratic oversight and normalising impunity while pushing genuine issues of injustice into the background. In the Philippine context, where populist leaders already command intense affective loyalty, such discursive practices can silence criticism and redirect public attention away from accountability issues such as corruption and mass murder, as in the case of the Duterte Government’s war on drugs (Iglesias, 2022). Recognising these rhetorical manoeuvres is therefore essential for citizens, journalists, and watchdog organisations to critically evaluate political communication, resist the co-optation of moral discourse, and defend victimhood as a framework for justice rather than for self-preservation by an elite (Chouliaraki & Banet-Weiser, 2021).

The rise of populism in the Philippines

From one perspective, Duterte’s rise to power on the back of populism is part of a worldwide trend of politicians using bombastic rhetoric and diplomacy to paint themselves as true

representatives of the people against corruption. Railing against the economic and global elites, a new breed of populist leaders speedily and simultaneously emerged in different parts of the globe by echoing the public's concern over the cost of globalisation (McCoy, 2017). Once in office, these populist leaders take on a moral tone and claim to lead in the name of and following the will of the people and do not recognise the legitimacy of their opposition (McCoy, 2017). This trend can be observed in Europe (e.g. the U.K. Independence Party, France's Marine LePen, Hungary's Viktor Orbán and Turkey's Recep Erdogan), in the USA in President Donald J. Trump, in Asia in India's Narendra Modi, Thailand's Thaksin Shinawatra and the Philippines' Rodrigo R. Duterte. However, despite being part of and from this lineage of populist politicians, McCoy argues that Duterte was unique because of his violent social policies and disdain for the world order. This leads to the point that Philippine populism is an intersection of a rising global tide and a meandering local movement whose origins could be traced back to the time of former Philippine President Manuel Quezon.

Although populism has been framed as a global reaction to widespread economic and social conditions, its uncritical application in the Philippines does not completely explain the rise of the movement on this archipelago (McCoy, 2017). The strategy of many former Philippine leaders, including Duterte and the former dictator Ferdinand Marcos, Sr., the father of the current chief executive, could be explained by a balancing act between political patronage and performative violence, which allowed them to promise law and order if given power (McCoy, 2017). According to McCoy, both present and past Filipino strongmen used performative violence to project domestic strength and success on the diplomatic front to show the local electorate international influence. The country's gravitation toward populist leaders offers better analytical tools for studying similar shifting trends overseas, as two key ingredients have been identified: local performative violence and skilled diplomacy (McCoy, 2017).

What is relevant in this study is the rise of former President Duterte to national power from his power base in the southern Philippine city of Davao. His ascendancy represented a significant shift in Philippine electoral politics, considering that his predecessor – former President Benigno Aquino III – was relatively popular when his term ended in June 2016. Duterte's presidency could be considered an inflection point in Philippine political history as he put in place an illiberal law and order regime (Thompson, 2016b). The most egregious example of this was Duterte's 'war on drugs', which claimed the lives of somewhere between 6,000 and 30,000 individuals (Iglesias, 2022), many of whom were drug users and street peddlers.

Duterte's populist appeal hinges on the dichotomy between virtuous law-abiding citizens and the dregs of society (Curato, 2016). This type of populism is termed *penal populism* (Pratt, 2007) and is a response to the public's clamour for political institutions to take a tougher stance against crime, leading to stronger punitive measures for control (Curato, 2016). Curato further explains that Duterte was able to leverage his appeal based on his supposed ability to surmount the slow wheels of the country's justice system and take action decisively. He was able to mount this challenge to the country's liberal order through a confluence of factors, primarily centred on the dysfunction in this dominant political order, such as the weak influence of once influential groups (e.g. the Catholic Church, non-governmental organisations), besmirching of the good governance (*Daang Matuwid*

or the Straight Path) narrative of the previous government, and the weaknesses of critical institutions (Thompson, 2016a). Duterte swiftly consolidated his power and established a new regime focused on his penal populism, characterised by human rights violations and impunity in the bloody crackdown against illegal drugs with no effective political opposition to speak of (Thompson, 2016a).

Until the end of this term, Duterte enjoyed ‘excellent’ net satisfaction ratings across geographical areas (Social Weather Stations, 2022). In fact, SWS noted that Duterte’s satisfaction rating in June 2022 was 10 points higher than the polling firm’s preceding reading in April 2022. This enduring popularity remained despite a series of controversies that beset his administration and eroded his appeal (Teehankee, 2022). The political capital that Duterte had at the time of the 2022 elections was further demonstrated when it helped propel the Marcos, Jr. and Duterte-Carpio ticket to an absolute majority win, the first since the end of the Marcos, Sr. dictatorship (Iglesias, 2022). Duterte’s political capital elevated his family to national prominence status and fuelled them in a successful run for the vice presidency, evidence of the transferability of his appeal. It was this transferred political capital that Duterte-Carpio was riding on and trying to protect as she faced severe political headwinds that she alleged were related to her possible 2028 presidential election candidacy (Lalu, 2024). However, the current satisfaction and approval ratings data demonstrate that this did not work, with her numbers, which had been the highest among government officials, sliding below those of Marcos, Jr. and congressional leaders (Mateo, 2024) for a certain period.

Victimhood in politics

Although victimhood is a crucial element of political and public opinion, it is often overlooked (Armaly & Enders, 2022). According to Armaly and Enders, victimhood is related to anti-establishment views, personality traits, and support for specific politicians. It has become a central element in today’s political battles, as debates today centre on the question of who the victims are (Chouliaraki & Banet-Weiser, 2021). Previously, the discourse on victimhood had centred on claims of injury on behalf of the powerless (e.g. victims of war or emotional trauma); this has radically changed in the 21st century with worldwide gyrations such as the 2008 financial crisis and the rise of social media platforms that have allowed everyone to stake a claim to victimhood (Chouliaraki & Banet-Weiser, 2021). Specifically, victimhood today, once the righteous expression of the vulnerable, is seen and claimed everywhere and ‘politically weaponized’ (Chouliaraki & Banet-Weiser, 2021, p. 4); for example, perceiving oneself as a victim is everyday political fare in the United States (Armaly & Enders, 2022) and an important signal of identity position (Horwitz, 2018).

Hijacked victimhood is a form of politicisation in which narratives of victimhood are used to depict the dominant group as being in danger and in need of protection – from the narrative’s creator in particular – suggesting that they are current or future victims facing threats, often from others (Hronešová & Kreiss, 2024). As a subset of strategic victimhood, it refers to the practice of political communication to seek justice and equity (Hronešová & Kreiss, 2024). Such appeals can then be mobilised for electoral and policy gains, as they often dovetail with the groups that politicians and their parties represent

(Lerner & O’Loughlin, 2023). Although hijacked victimhood traditionally manifests in the usurpation of a marginalised group’s experiences, delegitimation of that group, presentation of vulnerable groups as threats, and portrayal of the opposition as enablers of dangerous elements (Hronešová & Kreiss, 2024), this paper posits that it also covers attempts to portray ‘tactical vulnerability’, which is suffering claimed to the advantage of particular individuals or groups (Chouliaraki, 2021). As inherent in the structure of victimhood, anyone who successfully mobilises grand victimhood narratives and stakes a claim to suffering, regardless of their social position, will be able to claim victimhood status and, consequently, the moral value attached to it (Chouliaraki, 2021). Chouliaraki explains that this ability to mobilise grand narratives for victimhood is especially acute in today’s post-2008 crisis and platformed world, where pain is detached from the victims of neoliberalism and attached to those benefiting from current structural relations of power (Chouliaraki, 2021), and is amplified and monetised on and through social media (van Dijck et al., 2018).

Informed by the scholarship on how powerful and vested individuals and groups have deployed strategic vulnerability – as opposed to systemic vulnerability arising from conditions that perpetuate violence and suffering – and hijacked the concept of victimhood, this study examines how Vice President Duterte-Carpio strategically deployed victimhood in her congressional testimonies to preserve and protect her political power amid questions on her office’s budget and its utilisation. This paper argues that Duterte-Carpio indexed hijacked victimhood in her responses at congressional hearings to avoid scrutiny of her actions and contest, preserve and expand her political power in preparation for a possible presidential run in 2028. By investigating her *indexicals of victimhood*, this article is intended to shed light on how politicians have usurped victimhood and its associated moral value for their personal and political gain to the detriment of genuine victims of systemic and political violence in the country such as the casualties of former President Duterte’s war on drugs. This article is also intended to demonstrate another way of deploying victimhood – through a combination of victimhood indexicals and aggression – in keeping with the Duterte family’s penal populism (Pratt, 2007).

Investigation of the way in which Vice President Duterte-Carpio indexed hijacked victimhood in her selected congressional testimonies is important for two reasons. While previous studies have examined the strategic use of victimhood in populist rhetoric, these inquiries have primarily centred on Western contexts (Armaly & Enders, 2022; Hronešová, 2024; Neville-Shepard, M. & Neville-Shepard, R., 2022; Patrona, 2025) and prominent leaders such as Donald Trump and Viktor Orbán (Hronešová & Kreiss, 2024; Chouliaraki, 2021). Limited empirical analysis exists on how hijacked victimhood manifests in Southeast Asian political communication, particularly in the Philippines, where populist discourse can intersect with entrenched political dynasties and localised moral hierarchies. Moreover, existing research has tended to focus on speeches and campaign materials (e.g. Hronešová & Kreiss, 2024; Patrona, 2025) rather than on institutional forums such as congressional hearings, which provide a different arena for the enactment of victimhood. This study fills these gaps by examining Vice President Sara Duterte-Carpio’s congressional testimonies to show how hijacked victimhood operates linguistically as a strategy for the avoidance of accountability and the preservation of power within Philippine populism. In doing so,

it extends the conceptual framework of strategic victimhood to a new geopolitical and discursive context.

Methods

This study used a qualitative design and examined the transcripts of three hearings at the Philippine Senate and House of Representatives: 1. 20 August 2024, Hearing of the Senate Committee on Finance on the Proposed Budget of the Office of the Vice President; 2. 27 August 2024, Hearing of the House Committee on Appropriations on the Proposed Budget of the Office of the Vice President; and 3. 18 September 2024, Hearing of the House Good Governance and Public Accountability Committee on the Budget Use of the Office of the Vice President. The analysis of these transcripts involved two phases. In the first phase, incidents of hijacked victimisation were examined. These critical incidents were signalled by the Vice President's recourse to language that evoked persecution, unfair treatment, victimisation, vulnerability and willingness to suffer, among other affective investments she made to evince victimhood and vulnerability. Essentially, this phase investigated the discursive construction of tactical vulnerability and victimhood. QualCoder 3.5 was utilised in the coding process to examine these incidents. QualCoder is open-source, free software for the analysis of qualitative data, which allows the coding of text, images and videos, the writing of notes and memos, and the generation of tree-like hierarchical categorisation of data (QualCoder, 2019).

To establish the link between her language and the recourse to victimhood and vulnerability, the concept of indexicality was used to understand how this was reflected in language. Indexicality refers to the linking of semiotic forms, such as language, to context-specific social meanings (Bucholtz, 2015). An index is a linguistic (or semiotic) form that depends on the meaning of the social environment (Bucholtz & Hall, 2005). The second phase of the data analysis involved categorising the indexicals found in the first phase to ensure that the categories were conceptually distinguishable from each other and did not overlap, or if they did, their definitions were combined.

Results and discussion

This section presents the indexicals of hijacked victimhood that Vice President Sara Duterte-Carpio employed in three of her testimonies before the Philippine House of Representatives and Senate, which conducted hearings on her office's proposed budget and the use of previous capital outlays. Based on the results of the analysis of the congressional-hearing transcripts, this paper argues that Duterte-Carpio relied on aggressiveness and passiveness to hijack the concept of victimhood (Hronešová & Kreiss, 2024) to serve her own political ends. To do this, she specifically employed sharp answers to inquiries, a veneer of deference to Congress's decision, claims of political harassment, an expression of willingness to answer allegations in the proper venue, and projecting herself as a responsible, transparent, and kind leader, among others. The analysis of the congressional transcripts revealed how the language of

Duterte-Carpio employed both aggressive and non-aggressive techniques to defend her budget and avoid scrutiny of her utilisation of capital outlays in the past two years. Table 1 summarises the five main indexicals of hijacked victimhood found in this investigation. This list excludes three themes that were recorded only once or twice: use of house rules, defence against criticism, and recall of previous favour given.

*Table 1:
Summary of Duterte-Carpio’s indexicals of hijacked victimhood*

Strategy	Number	Examples
Sharp answers in testimonies	52	Madam Chair, I do not understand <i>bakit</i> [why] a person convicted of child abuse is still sitting in a seat of the House of Representatives.
Veneer of deference to congressional decision	42	Madam Chair, we only submit the budget, and we leave it to the discretion of all the members of Congress to decide on the budget. We can only propose a budget for the Office of the Vice President.
Claims of political harassment	21	What we are witnessing now is no ordinary legislative inquiry. This exercise is a well-funded and coordinated political attack.
Willingness to face allegations in the proper venue	18	If there are audit findings, we shall gladly respond to them before the Commission on Audit (COA), and if there are legitimate cases to be filed, then we shall gladly respond to them before the appropriate courts.
Projection of responsibility and transparency as a leader	9	It is then my lawful duty to give aid to every person who reaches out to the OVP seeking relief. I must do justice to any man, and I intend to do so.

Source: Compiled by the author.

The following sections provide a succinct discussion of the identified indexicals using representative samples from congressional transcripts. These indexicals were identified by examining the patterns within and across transcripts.

Sharp answers in testimonies

The most preponderant indexical found in Duterte-Carpio’s congressional testimonies was her deployment of sharp and aggressive answers during the hearings. In many instances, this was in response to the critical questions asked and comments provided during the hearings. In some episodes where this was found, she asserted that the panels conducting the hearings should allow her ‘snide’ remarks in the same way they allowed those of some legislators. These sharp answers included personal attacks, outright refusal to answer questions, sarcastic comments and engaging legislators in back-and-forth verbal tussles. The questions that she refused to answer were primarily related to her office’s rapid utilisation of confidential funds in previous years.

Extract 1:

Madam Chair, I do not understand bakit [why] a person convicted of child abuse is still sitting in a seat of the House of Representatives (27 August 2024, House Appropriations Committee Hearing).

Extract 2:

Madam Chair, the invitation letter that we received from the House of Representatives: 'Dear Vice President Duterte, the House Committee on Appropriations will hold briefing hearings on the budget proposals of various departments, agencies, and corporations relative to the fiscal year 2025 National expenditure program as a basis of the fiscal year 2025 general appropriations bill.' Nasaan sa presentation ko kanina – ang haba – ang confidential funds? [Where in my earlier presentation – which was very long – were confidential funds?] (27 August 2024, House Appropriations Committee Hearing).

Extract 3:

We would like to thank all the honorable members of the Senate for doing your work here in the Senate of the Philippines, except—joke lang, Ma'am [just a joke, Ma'am] (20 August 2024, Senate Appropriations Committee Hearing).

In Extract 1, Duterte-Carpio questioned why ACT Teachers Party-list Rep. France Castro was still holding a congressional seat despite her child abuse conviction. This statement presumably referred to the guilty verdict handed down against Castro and several others for allegedly endangering the lives of Lumad children (Galvez, 2024). Prior to this, Castro criticised Duterte-Carpio for not providing direct answers to questions about the notice of disallowance issued by the COA over her office's use of confidential funds. Instead of giving direct answers, Duterte-Carpio launched a personal attack on Castro, who is appealing against the lower court's verdict. Extract 2, meanwhile, shows Duterte-Carpio's refusal to answer questions on her confidential funds use asked during a budget scrutiny hearing. Finally, Extract 3 demonstrates Duterte-Carpio by jokingly excluding a senator from her message of thanks, presumably Senator Risa Hontiveros who had a sharp exchange with the vice president over the latter's budget request for the printing of a children's book she had written.

Duterte-Carpio's sharp and confrontational remarks exemplify what Cervi et al. (2023) identify as a populist personalisation rhetoric where opponents are vilified or mocked to focus on their personal lives and besmirch their public persona. This rhetorical aggressiveness is not just a question of temperament; it functions as an indexical sign of strength and weakness, allowing the speaker to frame herself simultaneously as strong and victimised. From the lens of strategic victimhood (Hronešová & Kreiss, 2024), such responses invert the traditional moral hierarchy by portraying the powerful as besieged by unfair critics. The sharpness of tone thus encodes both populist antagonism, the people vs. elite, law-abiding vs. law-breaking citizens divide (Curato, 2016; Mudde, 2004), and tactical vulnerability (Chouliaraki, 2021), signalling that the vice president is being unjustly targeted by a hostile establishment. In this sense, her sarcasm, ridicule and personal attacks become semiotic tools that fuse aggression and victimhood, a hallmark of hijacked victimhood narratives.

This interplay of aggression and victimhood typifies the communicative logic of populism, where emotional performance replaces rational accountability. The next theme, ‘Veneer of Deference to Congressional Decision’, reveals the contrasting strategy through which Duterte-Carpio tempers this aggression with an apparent submission to authority.

Veneer of deference to congressional decision

In contrast to the overt aggressiveness encoded in the first theme, this second indexical encoded passiveness and the vice president’s willingness to submit herself to Congress. Specifically, in this indexical, Duterte-Carpio communicated her willingness to abide by Congress’s decision regarding her submitted budget. This stance portrayed her as helpless before the powerful body and as a law-abiding individual.

Extract 4:

Hence, just as we have done last year, we again leave the 2025 OVP budget to the pleasure of Congressman Martin Romualdez. I am not asking for any special treatment, nor am I asking you to uphold any tradition. There is no disrespect. All I am saying is that you have the complete freedom to do whatever you wish with the OVP budget. If you feel that all the documentary submissions are not enough, then by all means, huwag kayong magbigay ng budget [do not give us any budget] (18 September 2024 Hearing, House Committee on Good Government and Public Accountability).

This statement showed the vice president’s apparent willingness to accept Congress’s decision on the budget for her office. What was notable here was the specific mention of House Speaker Martin Romualdez, whom she accused of actually controlling the country’s national budget (Peralta-Malonzo, 2024). By naming the House Speaker, Duterte-Carpio subtly attacked and signalled that Romualdez was behind the critical questioning of her outlay utilisation. This subtle offensive was hidden beneath a veneer of deference to Congress’s decision regarding her proposed budget. The line: “...you have the complete freedom to do whatever you wish with the OVP budget”, demonstrated this. This investigation views this apparent deference as just a veneer because of the context in which it was uttered. Prior to this, the vice president had called the hearing a “well-funded and coordinated political attack”, signalling the third indexical of victimhood in the investigation, political harassment. The next statement also claimed that the hearings were meant to make a “case for impeachment”, again tying everything to political harassment.

Duterte-Carpio’s apparent humility before Congress can be read as a strategic performance of tactical vulnerability (Chouliaraki, 2021), a rhetorical move where powerful actors portray themselves as weak or submissive to elicit sympathy and moral legitimacy. Her language of surrender (e.g. “we leave the budget to your pleasure”) indexes compliance while simultaneously undermining institutional authority, suggesting that the process is politically biased and therefore meaningless. This ambivalence typifies hijacked victimhood (Hronešová & Kreiss, 2024), where moral hierarchies are inverted and scrutiny from others is reframed as persecution. At the same time, such stylised passiveness resonates with populist

performance, which often oscillates between aggression (Cervi et al., 2023) and humility to project authenticity (Curato, 2016). By cloaking defiance in politeness, Duterte-Carpio positions herself as the wronged yet dutiful servant of the people, an indexical sign of strategic victimhood that appeals to audiences who reject elite authority (McCoy, 2017). In doing so, she claims moral high ground and turns accountability into a test of loyalty rather than legality.

This oscillation between deference and defiance sets the stage for the next indexical, claims of political harassment, where Duterte-Carpio explicitly transforms institutional scrutiny into evidence of persecution.

Claims of political harassment

The third indexical in the data was the vice president's claim that what she was experiencing was due to political harassment. Under this theme, the vice president viewed the scrutiny of her budget as politically motivated and simply political harassment. The vice president purported that there was a coordinated and orchestrated campaign against her to politically damage her in the 2028 presidential elections, where she is one of the frontrunners (Hufana, 2024).

Extract 5:

Madam Chair, this is an example of politicising the budget hearing through the questions of a senator. Ang problema niya kasi, nakalagay iyong pangalan ko doon sa libro. At iyong libro na iyan, ibibigay namin doon sa mga bata. At iyong mga bata na iyan, may mga magulang na boboto. At iyong pangalan ko ay nagkalat doon sa kung saan man ibibigay iyong libro [Her problem is my name is written on the book. Copies of the book will be given to children whose parents are voters. My name will spread among those who will receive copies] (20 August 2024, Senate Appropriations Committee Hearing).

In this extract, Duterte-Carpio painted the questions raised by Senator Risa Hontiveros on the budget allocation for a children's book that she had written as an example of budget deliberation politicisation. Instead of explaining and justifying the requested allocation, the vice president went on the offensive in claiming that the senator was politicising the budget hearings of her office through her critical questions. It is important to note that many panel members supported the vice president's budget proposal during the hearing and did not scrutinise the proposed outlay or how the office had utilised its previous allocations. In the other data, the vice president claimed that the legislators asking questions about her budget were following a 'script' and that the hearings were a prelude to a possible impeachment.

Duterte-Carpio's framing of congressional inquiry as political harassment exemplifies what Hronešová and Kreiss (2024) describe as hijacked victimhood, where moral hierarchies are reversed and those in power portray themselves as being in danger and in need of protection (Neville-Shepard, M. & Neville-Shepard, R., 2022). By claiming that critical questioning constitutes a form of politicisation, she re-indexes scrutiny as injustice, transforming an exercise of democratic oversight into evidence of her own oppression. This rhetorical inversion performs what Chouliaraki (2021) calls tactical vulnerability, which in this case

is an exhibition of weakness calculated to accrue moral capital in the face of a legitimate congressional oversight hearing. Within the populist logic of antagonism, such moves divide the field into moral binaries of the people against the elite (Curato, 2016; Mudde, 2004). Through her narrative, the vice president positions herself as the authentic voice of the people besieged by a hostile establishment that is out to damage her politically, thereby activating affective solidarity from supporters who perceive her as a victim of the elite. This dual move, vilifying institutions while mobilising empathy, illustrates the affective politics of contemporary victimhood (Chouliaraki & Banet-Weiser, 2021), where emotional appeal supersedes factual accountability.

The claim of political harassment thus functions as a bridge between overt aggression and performative compliance, paving the way for the next indexical, willingness to face allegations at the proper venue, where the vice president further consolidates her image as a law-abiding victim rather than a subject of accountability.

Willingness to face allegations at the proper venue

Through this indexical, the vice president expressed readiness to answer allegations at proper venues such as the COA and the courts. In the data, these allegations primarily centred on her use of confidential funds during the previous two years. Her refusal to answer questions on the issue in Congress was based on a stated belief that questions on her budget use should be answered in the other forums; she believed Congress should just evaluate her proposed outlay for the following year as presented. One notable observation in the lines that encoded this indexical was their formulaic nature, where the vice president simply stated that they had received the COA's report and would respond to that and cooperate in the probe. In these lines, she also assured the public of their cooperation. This was demonstrated by Extract 6.

Extract 6:

On the issue of confidential funds, we have received audit actions from COA regarding the confidential funds of the Office of the Vice President. We have submitted our responses, and we assure the public of our full cooperation with the ongoing and unfinished audit (27 August 2024, House Appropriations Committee Hearing).

Duterte-Carpio's repeated invocation of the 'proper legal forum' argument is another manifestation of strategic victimhood (Hronešová & Kreiss, 2024), where appeals to institutional procedure serve as discursive shields against moral and political accountability. By affirming her readiness to face charges before the courts, she projects legal compliance while simultaneously delegitimising the congressional inquiry as partisan or improper. This manoeuvre functions as a classic blame avoidance behaviour (Hinterleitner & Sager, 2017), reframing accountability as a question of technical procedure rather than moral responsibility. From the standpoint of indexicality (Bucholtz & Hall, 2005), her formulaic statements signal both transparency and persecution. This indexes the stance of a compliant official wronged by unfair scrutiny. This rhetorical performance resonates with populist narratives of moral purity, where there exists a dichotomy between virtuous law-abiding citizens and

the decadent law-breaking scum of society (Curato, 2016). Ultimately, Duterte-Carpio's insistence on legalism operates as a linguistic strategy that transforms accountability into virtue, thereby consolidating her image as a victimised yet law-abiding servant of the people.

This rhetorical balance between compliance and persecution is further reinforced in the next indexical, projection of responsibility and transparency as leader, where Duterte-Carpio expands her legal defensiveness into a broader narrative of benevolent public service.

Projection of responsibility and transparency as leader

Another non-aggressive indexical used was the vice president's projection of herself as a responsible and kind-hearted leader who was just doing her part in helping the Filipino people. This theme also covered her transparency in her actions and readiness to protect people around her and her staff from the political attacks against her, as demonstrated by Extract 7.

Extract 7:

On the matter of OVP projects being similar to the mandate of other government offices, there is certainly no agency that has the monopoly of government services. No office of the government should be so helpless as to deny assistance to a person in distress. After all, the duty of government is to accord our countrymen with faithful public service (27 August 2024, House Appropriations Committee Hearing).

In this extract, the vice president was justifying the budget earmarks for several of her office's projects, which seemed to overlap with those of other government agencies. There were suggestions for the funds to be transferred to concerned government agencies and for the OVP to direct requests for assistance to those offices. Carpio-Duterte explained that no government agency had a monopoly over public service. In other parts of the data, she claimed that she had sworn an oath to do justice to every man, and her projects were just part of the fulfilment of that oath.

Duterte-Carpio's presentation as a compassionate and dutiful leader exemplifies how hijacked victimhood operates through moral appropriation. By foregrounding her benevolence and commitment to 'faithful public service', she reframes political accountability as moral virtue and insulates herself from critique. This aligns with Hronešová and Kreiss's (2024) observation that powerful actors invert moral hierarchies by adopting the ethical language of the marginalised to defend privilege. Her discourse also indexes tactical vulnerability (Chouliaraki, 2021), which is the strategic invocation of care and service to evoke sympathy and legitimacy while deflecting accusations of wrongdoing. Within the populist moral framework (Curato, 2016; Mudde, 2004), this projection reinforces the binary of a selfless leader serving the 'ordinary people' against self-interested elites. Linguistically, these statements function as indexicals of moral authority (Bucholtz & Hall, 2005), signalling integrity and care for other people to authenticate her public persona. In effect, her appeals to responsibility and transparency do not simply describe governance; they perform it, transforming ethical language into a communicative shield that sustains her image as a persecuted yet virtuous leader.

Collectively, these indexicals – ranging from aggression to humility – reveal how Duterte-Carpio linguistically constructs a dual identity as both a victim and a moral exemplar, enabling her to preserve political capital while deflecting institutional scrutiny.

Conclusion

Considering the cluster of indexicals identified in Duterte-Carpio's congressional testimonies, her hijacking of victimhood can be characterised as a combination of aggressive and non-aggressive strategies. These indexicals, from sharp aggression to moral self-presentation, reveal a coherent rhetorical pattern that aligns with the logic of populism and hijacked victimhood. These linguistic performances illustrate how power speaks the language of vulnerability to legitimise itself. The combination of confrontational and deferential strategies demonstrates the possibility of affective duality in populist discourse: the ability to be in power and yet claim victimhood status (Hronešová & Kreiss, 2024). In this sense, Duterte-Carpio's congressional rhetoric exemplifies how political elites mobilise victimhood as a discursive device (Chouliaraki & Banet-Weiser, 2021) to maintain legitimacy in times of legal crisis. This allows her to claim victimhood status strategically and, at the same time, demonise her opponents, similar to the strategies used by populist leaders like Trump and Orbán (Hronešová & Kreiss, 2024). Duterte-Carpio's strategies enable her to assert victimhood, thereby establishing moral high ground and deflecting guilt (Hronešová, 2024). Additionally, the vice president weaponised her hijacked victimhood status to attack her opponents. This resembles the weaponisation of victimhood to justify aggressive actions, such as war in other countries (Hronešová, 2024).

On the one hand, the vice president's first component of hijacked victimhood, composed of a veneer of congressional deference, claims of political harassment, willingness to answer allegations, and projection of responsibility and transparency, allows her to claim the moral high ground (Chouliaraki, 2021; Hronešová, 2024) and deflect scrutiny of possible wrongdoing. The claimants to hijacked victimhood rewrite moral hierarchies, use legitimate victims' moral power, and adopt harm to defend their current political standing and power (Hronešová & Kreiss, 2024). Duterte-Carpio portrays herself as a law-abiding, compassionate public servant simply doing her job, who is subjected to political harassment at these hearings because of her popularity for the 2028 presidential elections (Hufana, 2024). However, despite this persecution, she continues to serve the public who voted her into office and is willing to face these allegations at proper venues. On the other hand, the second component of Duterte-Carpio's hijacked victimhood gives her the opportunity to demonise opponents through aggressive and pointed responses. This is characteristic of populist rhetoric, which is rich in character assassination, personal attacks and sharp language, and is also observed in other countries (Cervi et al., 2023). Cervi et al. explained that personalisation in politics is a critical element of populism and allows focus on opponents' personal lives and qualities to undermine their reputations. This also affords amplification on social media platforms to strengthen populist communication (Cervi et al., 2023), one of the key characteristics of contemporary notions of victimhood (Chouliaraki & Banet-Weiser, 2021). Taken together, the five indexicals demonstrate a discursive choreography of power that moves fluidly between

aggression and humility. This interplay shows how indexicality (Bucholtz & Hall, 2005) operates in populist victimhood rhetoric to produce recognisable moral cues that audiences can identify with. By drawing from the semiotics of sincerity and service, Duterte-Carpio's discourse constructs tactical vulnerability (Chouliaraki, 2021), which allows the powerful to embody both authority and fragility at once.

By hijacking victimhood, Duterte-Carpio avoided having to detail and justify her office's use of confidential funds by refusing to answer congressional questions while claiming to be ready to answer inquiries on that issue at the proper forum and supposedly deferring to congressional wisdom on the current budget. The 'proper legal forum' argument is part of blame avoidance behaviour (BAB), which covers all actions to protect the integrity of officeholders amid possible threats from events (Hinterleitner & Sager, 2017). The public's expectations are duly managed, and accountability is established by demonstrating that the matter is being taken with the requisite seriousness and addressed through appropriate legal channels (Hinterleitner & Sager, 2017). Using the 'proper legal forum' argument, the vice president avoids legal exposure and risks that might stem from her statements before Congress. Additionally, by portraying herself as a mere public servant striving to 'do justice to every man' while experiencing political oppression, she was able to paint herself as a victim who needed the sympathy and support of the public. These moves are tactics under hijacked victimhood, which the political elite, including populists, use to usurp victimhood's moral power in their pursuit or preservation of power (Hronešová & Kreiss, 2024). They use potential threats to maintain or gain political standing (Hronešová & Kreiss, 2024). In the case of Duterte-Carpio, she caricatures herself as a responsible and compassionate official whom legislators are politically harassing because of the threat she poses to the ruling coalition. Through her deployment of hijacked and aggressive victimhood, she hopes to protect herself legally and, at the same time, preserve her popularity, which will fuel her possible 2028 presidential run.

Previous literature on victimhood has shown that politicians who use victimhood narratives use a variety of indexicals in their political communication. For example, some politicians distinguish between in-groups and out-groups through their 'we vs. they' rhetoric in their speeches (Hronešová & Kreiss, 2024; Matos & Miller, 2023; Muhammad Jasim & Mustafa, 2021). Meanwhile, others promote their victimisation narratives through urgent temporal constructions and the infusion of legitimacy and urgency into their rhetoric, which allows them to urge immediate action (Al-Ghazzi, 2021). The current study contributes to this discourse by identifying the indexicals of hijacked victimhood in the Philippines. This paper demonstrates that Duterte-Carpio primarily deployed five indexicals to hijack victimhood in order to protect and preserve herself and her political power. This paper argues that these indexicals were intended to mobilise support and protection for the vice president and delegitimise congressional inquiries into her budget utilisation by painting them as having been politically motivated. She evidently hoped that, by doing this, she could protect herself from the possible legal repercussions of her actions (e.g. impeachment) and preserve her and her family's political power in time for the next presidential polls. However, this strategy did not seem to pay off, as her trust and approval ratings continued to slide at that time (Mendoza, 2024).

The insights of this study are essential to better understand political movements and the rhetoric employed by politicians. Specifically, the study shows how victimisation is utilised in Philippine political communication to achieve local politicians' goals and ends. This study shows how the vice president, who belongs to one of the most powerful political families in the country, exploited victimhood to preserve and protect her political power. The results of this study allow observers and the public to be vigilant against attempts to avoid scrutiny of officials' actions involving public interest and welfare. Furthermore, by shedding light on strategic vulnerability and hijacked victimhood, this study obliquely highlights the need to remember the real victims of injustice and violence in the country, such as those of Duterte-Carpio's father's war on drugs.

These findings extend the conceptual work of Hronešová and Kreiss (2024) by demonstrating that hijacked victimhood is not only an ideological strategy but also a linguistic and performative one. The Philippine context shows that such strategies are localised through familiar cultural values (e.g. humility and service) that make moral inversion even more persuasive to the public. This exhibits that hijacked victimhood adapts to local moral repertoires and can therefore thrive in democracies where affective performance carries more weight than institutional accountability. The results thus broaden our understanding of the way in which strategic victimhood travels across contexts, from Western populisms to Southeast Asian political communication.

Beyond its academic contribution, this study has implications for civic education and democratic participation. When political victimhood is hijacked by the powerful, it risks dulling the public's moral sensitivity through the exploitation of the concept of victimhood (Hronešová, 2024). Such discourse reconfigures citizens from active agents of accountability into spectators of conflict, thereby weakening public vigilance. Civil society organisations, educators and the media can counter this trend by promoting critical media literacy and discourse awareness that expose how emotional appeals and inverted victimhood narratives serve partisan interests (Chouliaraki & Banet-Weiser, 2021). Encouraging citizens to interrogate political communication and re-focus on the experiences of genuine victims of structural and political violence (e.g. war on drugs) is essential to restore moral clarity and sustain democratic accountability in the Philippines.

Further research is needed on strategic and hijacked victimisation in the Philippines. First, this study only investigated hijacked victimhood in Duterte-Carpio's three congressional testimonies. A richer understanding can be developed if expanded to include all of her testimonies following her resignation from the cabinet, pronouncements, statements, and actions. This might demonstrate how hijacked victimisation is encoded textually and through other semiotic resources. Second, it is also interesting how former officials of the Duterte Government (e.g. former presidential spokesperson Harry Roque) utilise victimhood narratives now that the tables have turned and they are the ones alleging political persecution. Finally, research can also be conducted on the role of strategic victimhood (Hronešová & Kreiss, 2024) in successful advocacies and campaigns for people in peripheries worldwide.

In conclusion, while victimhood has been helpful in redressing harms done in the past, this study and previous literature have demonstrated that politicians around the world have repurposed this to serve political and personal ends (Al-Ghazzi, 2021; Banet-Weiser, 2021; Chouliaraki, 2021; Hronešová, 2024; Hronešová & Kreiss, 2024; Zembylas,

2021). The Philippines is no exception to this trend, with no less than its vice president employing hijacked victimhood indexicals in her congressional testimonies. As more research is conducted on this topic, a critical examination of victimhood claims is warranted to ensure that it serves the ends of justice for those who truly experienced harm and trauma rather than the vested interests of political actors. This study underscores that the struggle over victimhood is also a struggle for moral authority. As populist leaders increasingly rely on affective communication to project authenticity and grievance, the ability to decode such rhetoric becomes all the more crucial. Future scholarship should continue to map how these indexicals of victimhood circulate in digital and institutional arenas, shaping public perception and political legitimacy. At the same time, civic education and critical journalism must reclaim victimhood as a moral vocabulary of justice rather than a rhetorical tool of evasion.

Acknowledgments

The author extends his deepest gratitude to the Philippine Senate and the Philippine House of Representatives for sharing a copy of the official transcripts of the concerned congressional hearings. Any errors or omissions remain the sole responsibility of the author.

References

- Al-Ghazzi, O. (2021). We Will Be Great Again: Historical Victimhood in Populist Discourse. *European Journal of Cultural Studies*, 24(1), 45–59. Online: <https://doi.org/10.1177/1367549420985851>
- Armaly, M. T. & Enders, A. M. (2022). ‘Why Me?’ The Role of Perceived Victimhood in American Politics. *Political Behavior*, 44(4), 1583–1609. Online: <https://doi.org/10.1007/s11109-020-09662-x>
- Banet-Weiser, S. (2021). ‘Ruined’ Lives: Mediated White Male Victimhood. *European Journal of Cultural Studies*, 24(1), 60–80. Online: <https://doi.org/10.1177/1367549420985840>
- Bucholtz, M. (2015). The Elements of Style. In D. N. Djenar, A. Mahboob & K. Cruickshank (Eds.), *Language and Identity across Modes of Communication* (pp. 27–60). De Gruyter Mouton.
- Bucholtz, M. & Hall, K. (2005). Identity and Interaction: A Sociocultural Linguistic Approach. *Discourse Studies*, 7(4–5), 585–614. Online: <https://doi.org/10.1177/1461445605054407>
- Cervi, L., Tejedor, S. & Villar, M. G. (2023). Twitting Against the Enemy: Populist Radical Right Parties Discourse Against the (Political) “Other.” *Politics and Governance*, 11(2). Online: <https://doi.org/10.17645/pag.v11i2.6391>
- Chouliaraki, L. (2021). Victimhood: The Affective Politics of Vulnerability. *European Journal of Cultural Studies*, 24(1), 10–27. Online: <https://doi.org/10.1177/1367549420979316>
- Chouliaraki, L. & Banet-Weiser, S. (2021). Introduction to Special Issue: The Logic of Victimhood. *European Journal of Cultural Studies*, 24(1), 3–9. Online: <https://doi.org/10.1177/1367549420985846>
- Crisostomo, S. (2023, September 26). OVP Spent P125 Million Confidential Funds in 11 Days. *The Philippine Star*. Online: <https://www.philstar.com/headlines/2023/09/26/2299068/ovp-spent-p125-million-confidential-funds-11-days>

- Curato, N. (2016). Politics of Anxiety, Politics of Hope: Penal Populism and Duterte's Rise to Power. *Journal of Current Southeast Asian Affairs*, 35(3), 91–109. Online: <https://doi.org/10.1177/186810341603500305>
- Elemia, C. & Wee, S.-L. (2025, February 5). Philippines' Vice President Sara Duterte Is Impeached, Deepening Political Turmoil. *The New York Times*. Online: <https://www.nytimes.com/2025/02/05/world/asia/philippines-vice-president-duterte-impeached.html>
- Galvez, D. (2024, July 15). Satur, Castro Convicted for Endangering Lumad Children. *The Philippine Star*. Online: <https://www.philstar.com/headlines/2024/07/16/2370510/satur-castro-convicted-endangering-lumad-children>
- Hinterleitner, M. & Sager, F. (2017). Anticipatory and Reactive Forms of Blame Avoidance: Of Foxes and Lions. *European Political Science Review*, 9(4), 587–606. Online: <https://doi.org/10.1017/S1755773916000126>
- Horwitz, R. B. (2018). Politics as Victimhood, Victimhood as Politics. *Journal of Policy History*, 30(3), 552–574. Online: <https://doi.org/10.1017/s0898030618000209>
- Hronešová, J. B. (2024). The Uses of Victimhood as a Hegemonic Meta-Narrative in Eastern Europe. *Journal of Contemporary European Studies*, 32(2), 442–458. Online: <https://doi.org/10.1080/14782804.2022.2110456>
- Hronešová, J. B. & Kreiss, D. (2024). Strategically Hijacking Victimhood: A Political Communication Strategy in the Discourse of Viktor Orbán and Donald Trump. *Perspectives on Politics*, 22(3), 717–735. Online: <https://doi.org/10.1017/S1537592724000239>
- Hufana, C. M. (2024, November 18). VP Duterte, Senator Tulfo Are Filipinos' Most Preferred Presidential Candidates in 2028 – Survey. *Business World*. Online: <https://www.bworldonline.com/the-nation/2024/11/18/635750/vp-duterte-senator-tulfo-are-filipinos-most-preferred-presidential-candidates-in-2028-survey/>
- Iglesias, S. (2022). Violence and Impunity: Democratic Backsliding in the Philippines and the 2022 Elections. *Pacific Affairs*, 95(3), 575–593. Online: <https://doi.org/10.5509/2022953575>
- Lalu, G. P. (2024, September 18). New House Probe Wants to Tell People “Don't Vote for Sara in 2028” – VP. *Inquirer.net*. Online: <https://newsinfo.inquirer.net/1984440/duterte-new-house-probe-wants-to-tell-people-dont-vote-for-sara-in-2028>
- Lerner, A. B. & O'Loughlin, B. (2023). Strategic Ontologies: Narrative and Meso-Level Theorizing in International Politics. *International Studies Quarterly*, 67(3). Online: <https://doi.org/10.1093/isq/squad058>
- Marcelo, E. (2025, July 26). VP Sara's Impeachment Unconstitutional – SC. *The Philippine Star*. Online: <https://www.philstar.com/headlines/2025/07/26/2460776/vp-saras-impeachment-unconstitutional-sc>
- Mateo, J. (2024, November 3). Sara Ratings Now Lower Than Those of Chiz, Romualdez. *The Philippine Star*. Online: <https://www.philstar.com/headlines/2024/11/03/2397244/sara-ratings-now-lower-those-chiz-romualdez>
- Matos, Y. & Miller, J. L. (2023). The Politics of Pronouns: How Trump Framed the Ingroup in the 2016 Presidential Election. *Politics, Groups, and Identities*, 11(3), 507–525. Online: <https://doi.org/10.1080/1565503.2021.2007964>
- Mendoza, R. (2024, October 31). Duterte Ratings Dip Again. *The Manila Times*. Online: <https://www.manilatimes.net/2024/11/01/news/national/duterte-ratings-dip-again/1994944>
- McCoy, A. W. (2017). Global Populism: A Lineage of Filipino Strongmen from Quezon to Marcos and Duterte. *Kasarinlan: Philippine Journal of Third World Studies*, 32(1–2), 7–54.
- Mudde, C. (2004). The Populist Zeitgeist. *Government and Opposition*, 39(4), 541–563. Online: <https://doi.org/10.1111/j.1477-7053.2004.00135.x>

- Muhammad Jasim, R. & Mustafa, S. (2021). English Personal Pronouns as a Manipulation Strategy in Political Discourse: A Critical Discourse Analysis. *Journal of the College of Languages*, (44), 1–20. Online: <https://doi.org/10.36586/jcl.2.2021.0.44.0001>
- Panti, L. (2024, August 28). VP Sara: Comparing 2025 OVP Budget with Robredo's Time Is Absurd. *GMA News Online*. Online: <https://www.gmanetwork.com/news/topstories/nation/918491/vp-sara-comparing-ovp-budget-robredo-time-absurd/story/>
- Patrona, M. (2025). From Victim to Avenger: Trump's Performance of Strategic Victimhood and the Waging of Global Trade War. *Journalism and Media*, 6(3). Online: <https://doi.org/10.3390/journal-media6030134>
- Peralta-Malonzo, T. A. (2024, September 10). Sara Duterte: Romualdez, Co Control Budget Allocations. *Sun Star*. Online: <https://www.sunstar.com.ph/manila/sara-duterte-romualdez-co-control-budget-allocations>
- Pratt, J. (2007). *Penal Populism*. Routledge.
- Neville-Shepard, M. & Neville-Shepard, R. (2022). Outfitting the Conservative Civil Rights Movement: Rehearsed White Victimhood and the MAGA Hat. *Rhetoric and Public Affairs*, 25(4), 35–63. Online: <https://doi.org/10.14321/rhetpublaffa.25.4.0035>
- QualCoder (2019, August 13). *What is QualCoder?* Online: <https://qualcoder.wordpress.com/reports/>
- Social Weather Stations (2022, September 23). Second Quarter 2022 Social Weather Survey: Pres. Rodrigo Duterte's final net satisfaction rating at +81. *Social Weather Stations*. Online: <https://www.sws.org.ph/swsmain/artcldispage/?artcsyscode=ART-20220923101814>
- Teehanke, J. C. (2022). The Philippines in 2021. *Asian Survey*, 62(1), 126–136. Online: <https://doi.org/10.1525/as.2022.62.1.12>
- Thompson, M. R. (2016a). Bloodied Democracy: Duterte and the Death of Liberal Reformism in the Philippines. *Journal of Current Southeast Asian Affairs*, 35(3), 39–68. Online: <https://doi.org/10.1177/186810341603500303>
- Thompson, M. R. (2016b). The Early Duterte Presidency in the Philippines. *Journal of Current Southeast Asian Affairs*, 35(3), 3–14. Online: <https://doi.org/10.1177/186810341603500301>
- Van Dijck, J., Poell, T. & De Waal, M. (2018). *The Platform Society. Public Values in a Connective World*. Oxford University Press. Online: <https://doi.org/10.1093/oso/9780190889760.001.0001>
- Zembylas, M. (2021). Interrogating the Affective Politics of White Victimhood and Resentment in Times of Demagoguery: The Risks for Civics Education. *Studies in Philosophy and Education*, 40(6), 579–594. Online: <https://doi.org/10.1007/s11217-021-09777-6>

Political Communication on Telegram

Emotional Framing and Strategic Messaging in Europe

Rubén Rivas-de-Roca,^{*}  Concha Pérez-Curiel^{**} 

* Universidade de Santiago de Compostela, SPAIN
E-mail: rubenrafael.rivasderoca[at]usc.es

** Universidad de Sevilla, SPAIN
E-mail: cperez1[at]us.es

In a technological age, many political actors provide orientations on digital platforms, especially during periods of crisis. This research aims to explore political messaging and the perceptions of public opinion in Europe within the period of the 2022 energy crisis that resulted from the war in Ukraine. First, the evolution of the perception of the citizenry is described, analysing trust in public institutions through a secondary analysis on data from the Eurobarometer. Then, Telegram messages posted over a trimester in 2022 (n = 929) by some of the governing parties and the main populist opposition parties in the four most populous EU countries (Germany, France, Italy, and Spain) were subjected to content analysis. Results show how trust in national institutions has been eroded compared to the EU. The communication on Telegram was quite emotional, applying both positive and negative feelings, but the agenda of populists was much more fragmented than those parties in government.

Keywords: political communication, public opinion, populism, emotions, Europe, Telegram

Address for correspondence: Rubén Rivas-de-Roca, e-mail: rubenrafael.rivasderoca[at]usc.es.

Article received on 6 November 2025. Article accepted on 20 March 2026.

Conflict of interest: The authors declare no conflicts of interest.

Funding: The authors received no financial support for the research, authorship, or publication of this article.

Introduction

Crises are moments of strong transformation that imply a change in social priorities. In the summer of 2022, the energy-related crisis that resulted from the war in Ukraine caused European countries to adopt complex saving measures. In times of massive use of social media for political purposes (Stier et al., 2018), political actors tend to explain their positions on social media or mobile instant messaging services such as Telegram (Alonso-Muñoz et al., 2022). As part of digital political communication, mobile instant messaging services represent a new opportunity to connect political actors with citizens. Specifically, these digital platforms are a channel to spread messages more directly (Newman et al., 2019).

Political leaders and organisations try to provide media orientations (Jacobs & Wonneberger, 2019). These actors have adapted their communication practices to the digital media logic (Bimber, 2014), incorporating elements of negative campaigns that are commonplace on platforms such as WhatsApp and Telegram (Tirado-García, 2023). Emotional messages linked to criticism achieve more impact, but they also generate disaffection (Patterson, 1993), particularly in moments of political polarisation shaped by social media (Tucker et al., 2018).

Moreover, the rise of polarisation overlaps with the consolidation of populist rhetoric and a loss of trust in public institutions (Jungherr et al., 2019). According to a recent study (Verboord et al., 2025), trust is an essential foundation for democracies that is negatively associated with social media use. With regard to the war in Ukraine, its effects have especially touched Europe, and the European Union (EU) has been targeted as the objective of populist movements (Caiani & Guerra, 2017). Although the EU has traditionally been an issue in European politics, right-wing populists have developed social media strategies widely to give prominence to their Eurosceptic claims (Carral et al., 2023).

In light of the above, the objective of this research is to examine the political messages on Telegram and the perceptions of public opinion during a crisis that affected the economy and well-being. Indeed, its impact on the European Public Sphere (EPS) makes shedding light on how Telegram is used for political purposes in different EU countries relevant. Previous literature has analysed the communication style of populist and non-populist candidates, highlighting the negativity of the first (Nai, 2021), but there is scant research on the emotional nature of political parties (Widmann, 2021). As mobile instant messaging services enable direct communication, we aim to understand the messages posted from these organisations, ranging from parties in government to populist opposition groups.

Theoretical framework

The political use of emotions on social media

Social media have become essential tools for political actors in order to share messages without filters or limitations (Bossetta, 2018; Dimitrova & Matthes, 2018). This dynamic is not only relevant during periods of the electoral campaigns, but also in times of conflict. The outcomes of communication on social networks are highly recognised, generating echo

chambers that reinforce existing beliefs. However, the media environment and political interest of the population influence the effects of these eco chambers (Dubois & Blank, 2018).

In an age of digital capitalism, social media platforms are an essential part of society (Fuchs, 2017), one shaping a new relationship between politicians and the public. Besides that, people are incidentally exposed to news on these platforms, especially the youth (Fletcher & Nielsen, 2018). This also explains why social media are a tool for political power (Shirky, 2011) that tends toward personalisation (McAllister, 2007), understood as the prominence of leaders. The massive use of television made the personality of leaders a central element in the assessment of their performance, triggering the so-called 'pop politics' (Mazzoleni & Sfarдини, 2009) that is now reinforced by digital platforms.

In this context, social media allow messages to be adapted to the needs of the audience in terms of interests or demographics (Kreiss et al., 2018). Mostly, it is electoral goals that are pursued; which is why previous studies have focused on electoral campaigns (Stier et al., 2018). Despite the relevance of social media for all political actors, there seems to be a strong connection between these platforms and populism. According to Gerbaudo (2018), they provide a massive channel for the development of the typical strategies of populism, such as claiming to be the voice of the people.

Likewise, populist parties on social media address ideologically thematic issues that provoke controversy, impacting on the agenda of news media organisations, including fact-checkers (Rivas-de-Roca et al., 2024). These digital tools serve to create a populist communication style based on criticism of the establishment to the point of antagonising the elite. Its rhetoric seeks to build a collective identity by applying a strategic use of colloquial language and a rebellious narrative (Block & Negrine, 2017).

Emotions are also distinguished from populist communication. The persuasiveness of populist messages is largely fuelled by the emotionalised attribution of blame to elites (Hameleers et al., 2016). They portray a society that does not work well, and ask for political change in order to prevent social collapse. These assignments based on negative emotions have a stronger effect on citizens (Tirado-García, 2023), which overlaps with the rise in political polarisation.

Beyond traditional social media, the consolidation of mobile instant messaging apps is also altering political communication. These services allow users to send messages and reach mobile devices directly without opening the application or searching for a specific account. Thus, there is a tighter environment where people may feel more confident about sharing and discussing political content (Valenzuela et al., 2021). Prior scholarship has described how these apps facilitate political conversations in closed environments (Vermeer et al., 2021) as messages reach the devices of each user through private chats without others accessing them.

Telegram as a political tool

In recent years, Telegram is one of the platforms that has grown the most. According to Marcos-García et al. (2023), there are two reasons for its increased importance in the political sphere: 1. the possibility of establishing a more emotional connection with citizens

in contrast to classic social media; and 2. the need for political actors to seek alternatives to private communication in the face of public overexposure.

Literature has shed light on the prominence of Telegram within disintermediated communication. Many people discuss politics on this platform, which is shaping a new form of political discussion (Valeriani & Vaccari, 2018). In this sense, it is mainly elections that have been studied, evidencing the way in which Telegram is used as mobilisation tool that enables closeness to candidates and parties (Sierra et al., 2022). Electoral information is delivered while taking advantage of the private environment on Telegram to develop a personal agenda and criticise opponents. Particularly, this app works well for radical mobilisation as it has direct message retransmission, which is useful in the sharing of populist appeals (Jost & Dogruel, 2023).

The rise of Telegram should be considered within the framework of platformisation, which is a concept that refers to how digital platforms (Google, Meta, X, Telegram, TikTok, etc.) are becoming essential in understanding how people communicate. The penetration of digital platforms into daily life brings a huge transformation in cultural practices (Poell et al., 2019). This change affects access to information, but with significant country differences (Nielsen & Fletcher, 2023), necessitating comparative studies. Furthermore, information disorders have been magnified by the proliferation of these mobile instant messaging services, where disinformation circulates in an inconspicuous way.

This kind of platform emotions are conceptualised as a central mediating mechanism between populist rhetoric and institutional trust in periods of crisis. Building on the literature concerning the emotionalisation of politics and affective publics (Papacharissi, 2016), emotions are understood not as secondary reactions but as constitutive forces in political meaning-making. At the same time, emotionalised populist communication is not platform-neutral but deeply conditioned by the affordances of digital infrastructures. Drawing on affordance theory (Ronzhyn et al., 2023), we argue that platforms also shape the forms of interaction through which emotions are reinforced or normalised, in addition to the visibility and circulation of emotional content.

Telegram's affordances – such as its large broadcast channels, limited content moderation and semi-private communication spaces – facilitate the sustained dissemination of emotionally charged narratives with minimal contestation (Rogers, 2020). Unlike algorithmically curated platforms such as Facebook or X, where emotional content competes for attention within heterogeneous networks, Telegram enables the consolidation of affective publics characterised by ideological homogeneity and high emotional resonance. This social media ecology amplifies populist frames and fosters affective alignment between communicators and audiences (Kiess, 2025).

Within this conceptual framework, Telegram emerges as a distinctive communicative infrastructure for crisis communication and populist mobilisation. Its affordances support a mode of emotional governance in which institutional trust is eroded not primarily through factual counterclaims but through repeated emotional cues that reframe crises as the moral and political failures of institutions. By enabling persistent, emotionally synchronised communication between populist actors and their audiences, Telegram blurs the boundary between information, mobilisation and affective bonding. This capacity to stabilise

emotionalised distrust over time reshapes how crises are interpreted and how institutional authority is evaluated.

Bearing these facts in mind, this research aims to explore the political messages on Telegram and the perceptions of public opinion during the 2022 energy-related crisis, a period (triggered by the war) that impacted on the economy and well-being. As Europe is probably the place most affected by the saving measures against the economic crisis, our study focuses on the four most populous countries in the EU: Germany, France, Italy and Spain. In accordance with what is presented, the following research questions were posed:

RQ1: How has trust in public institutions at the EU level evolved during the period of the energy crisis?

RQ2: What communicative purposes (functions) and thematic priorities characterise political messages that shape the agenda on Telegram?

RQ3: What emotions prevail in party-political Telegram messages?

Methodology

This study develops a triangulation between research techniques, mixing secondary data analysis and quantitative content analysis. The goal was to understand the way emotions work in the messages of political parties (RQ2 and RQ3) and the evolution of public opinion during the energy crisis (RQ1). To answer RQ1, the evolution of the citizens' perception was described, analysing trust in public institutions. On this matter, a secondary analysis of data was applied to four waves of the Eurobarometer (EB 96, 97, 98 and 99), specifically:

1. The Standard Eurobarometer 96 (winter 2021–2022). This survey was conducted between 18 January and 14 February 2022. The final sample of responses was 26,696 people from 27 countries, using a multi-stage, random and stratified sample to analyse 1,001 variables. The Standard Eurobarometer was also carried out in 12 other countries or territories such as the Balkans or the United Kingdom, but they were not included in our study as we consider the EU a space especially touched by the consequences of the war in Ukraine.
2. The Standard Eurobarometer 97 (summer 2022). This study was conducted between 17 June and 17 July 2022. In some countries, face-to-face interviewing was mixed with online interviews. The sample reached 26,468 individuals from EU27, applying a stratified system. This wave was the first Eurobarometer survey since the invasion of the Ukraine.
3. The Standard Eurobarometer 98 (winter 2022–2023). This survey was conducted between 12 January and 6 February 2023. Interviews included 26,468 people (in person or virtually) from all 27 Member States, based on a stratified sample.
4. The Standard Eurobarometer 99 (spring 2023). This survey was carried out between 31 May and 25 June 2023. A total of 26,425 interviews were conducted across the whole of the EU.

The Eurobarometer waves had a sample error of $\pm 1.4\%$ for a 95% confidence level, targeting citizens from the age of 15 years. These surveys are held by Kantar Public; meanwhile, the European Commission's Directorate-General for Communication (DG COMM), through the Strategy, Corporate Communication Actions and Eurobarometer Unit, is responsible for its design. For our study, two variables were extracted:

- Trust in public institutions. Three categories, referring to trust in the European Union, the national government, or the national parliament.
- Opinion regarding the future of the EU. Three categories were included by the survey: optimism, pessimism and do not know (DN).

The comparison of Eurobarometer variables with data retrieved from other methods has been already applied with a qualitative approach (Rivas-de-Roca & García-Gordillo, 2022), but here we used a quantitative content analysis. This method (Krippendorff, 2004) allows the content of messages on social networks to be examined in a systematic way. Our research design addressed Telegram as a rising mobile instant messaging service, analysing the content for three months (15 July – 15 October 2022). This period was selected because it was the moment at which energy saving measures began to be adopted.

The sample was composed of all the messages published by some of the parties in government and the main populist opposition parties at that moment, located in the four most populous EU countries: Germany, France, Italy and Spain. Parties in government (ruling) are here defined as political parties that, at the time of analysis in 2022, held executive power either by leading the government or by participating formally in the governing coalition at the national level. These parties bear direct responsibility for policymaking and implementation, and their communication could therefore be shaped by institutional accountability and governance constraints. Opposition parties, by contrast, are those represented in parliament but not participating in the executive. They usually position themselves as critics of the incumbent government's decisions.

Populism is conceptualised as a transversal political characteristic rather than a fixed party type. It refers to an ideological frame that constructs politics as a moral struggle between a virtuous 'people' and a corrupt 'elite', following an anti-establishment rhetoric (Boulianne et al., 2020). Social media boosts this dichotomous boundary where the leader is part of the in-group (Bracciale et al., 2021). Populism is also defined as a particular communication style for political purposes (Jagers & Walgrave, 2007); hence, it can be articulated by both governing and opposition parties, depending on context and strategy. However, in the specific period analysed in this study, populist discourse was predominantly observed among the selected opposition parties.

Data were collected from the official Telegram channels of selected political parties over the trimester in question. Four trained researchers systematically monitored these channels and manually retrieved all messages published during the observation window. To ensure consistency and completeness, researchers followed a shared protocol, recording message content and documenting metadata such as date and time of publication. All collected messages were subsequently translated into English to enable standardised analysis across researchers. Translations were reviewed for accuracy to preserve the original meaning and the tone of the messages.

The accounts collected included the following eight Telegram media profiles:

- Germany: Bündnis 90/Die Grünen (ruling party, part of the government coalition), AfD – Alternative für Deutschland (opposition party).
- France: Renaissance (ruling party), RN-Rassemblement National (opposition party).
- Italy: PD – Partito Democratico (ruling party), Fratelli d'Italia (opposition party, later in government since October 2022).
- Spain: PSOE – Partido Socialista Obrero Español (ruling party), Vox (opposition party).

The dataset consists of content drawn from the official Telegram profiles of these political parties. Specifically, the data were extracted manually by accessing each party's official Telegram channel and copying the complete posts, including text and multimedia contents. To address linguistic diversity, profile texts published in languages other than English were translated into English using DeepL. This tool was selected due to its strong performance in translating political and institutional language. Following automated translation, all texts were manually reviewed to verify semantic accuracy and to ensure consistency across the sample.

Based on a comparative approach between nations that may be helpful in illustrating different reasons for the political use of Telegram (Valeriani & Vaccari, 2018), our study analyses the social messages posted by the above parties at the interstate level in the context of the European Union. To answer RQ2 and 3, we first created a list of 15 exclusive categories to delve into the distribution of the thematic agenda. Beyond that, the digital message functions of political parties were studied through the analysis protocol defined by López-Meri et al. (2017), adapting the strategies identified to the logic of Telegram.

Regarding emotions, we combined the basic psychological emotions identified by the literature (Palmero Cantero et al., 2002) with other emotions that show the positive and negative feelings towards political actors (López-López et al., 2020). Since previous studies had found that the emotional charge of digital discourses correlates not only to the level engagement (Rivera Otero et al., 2021) but also to the electoral result, we developed a specific analysis list on emotions (Table 1), arguing that this concept refers to feelings triggered by actions or circumstances.

A manual content analysis was applied to the messages on Telegram during the aforementioned trimester ($n = 929$), whereas the data were processed with IBM SPSS Statistics, Version 28. The message functions and the thematic agenda were coded by a single researcher as they are nominal categories and have little interpretation. Both the collection of the sample and the coding of emotions were carried out manually by a team of four coders, during which time a pre-test was conducted on 10% of the sample (93 units).

The intercoder agreement reached high levels. The coefficients per category ranged between 0.88 and 0.97, all remaining above the recommended threshold of 0.80 according to Krippendorff's alpha values (Neuendorf, 2002). The detailed Krippendorff alphas for each individual variable are fear ($\alpha = 0.95$), joy ($\alpha = 0.91$), anger ($\alpha = 0.93$), sadness ($\alpha = 0.97$), disgust ($\alpha = 0.88$), blame ($\alpha = 0.94$), shame ($\alpha = 0.90$), pride ($\alpha = 0.96$), empathy ($\alpha = 0.92$), gratitude ($\alpha = 0.94$), hope ($\alpha = 0.93$) and other ($\alpha = 0.89$).

Table 1:
Categories of emotions on Telegram

Items (emotions)	Description
Fear	Messages that share a strong concern for any scenario from the past, present or future.
Joy	Messages that contain emotions that spark joy.
Anger	Contents based on disapproval or annoyance at political behaviour or a specific event.
Sadness	Messages related to a strong disappointment.
Disgust	Contents that disseminate rejection (contempt) towards specific attitudes or political facts.
Blame	Aggressive messages seeking responsibility for certain events.
Shame	Contents contrary to pride as they express the will to escape from a political situation.
Pride	Messages that acknowledge the importance of the achievement of a political actor.
Empathy	Contents that show understanding in the face of specific situations or events.
Gratitude	It occurs when messages share the feeling of gratitude for a specific event or person.
Hope	Contents on trust in the political future.
Other	Unclassifiable in the previous categories.

Source: Compiled by the authors.

In this sense, emotions were coded by assigning each message to one of the 15 mutually exclusive categories predefined in the codebook, based on the dominant emotional cue expressed in the text. These categories derived from established literature were previously used by the research group, which might explain the results for intercoder agreement. The items aim to examine how emotional framing contributes to the distribution of the thematic agenda.

Our analysis adapts to digital logic by dealing with many variables during the time frame in question. It should be noted that the number of posts on Telegram differs by country (125 in Germany, 90 in France, 405 in Italy and 309 in Spain), but the results are sometimes presented as aggregated data here to make comparisons. To avoid the problem of cherry-picking, and also considering that the field of emotions on Telegram is little explored, two independent scholars were invited to monitor the research and discuss its findings.

Results

Support for public institutions

At the beginning of this energy-related crisis, the data shows that the EU was the main source of trust (Table 2), which status remained almost unchanged between 2021 and 2023 according to the Eurobarometer. Conversely, trust in national institutions (government and parliament) decreased by 3%. These authorities were responsible for managing some of the economic measures brought in to deal with the crisis, possibly weakening their support in the waves analysed (from EB 96 to EB 99). However, other major events or factors could also have determined changes in the level of trust. What can be stated on the basis of empirical evidence is that the EU did not experience any decline in public support during the years of the energy crisis.

Table 2:
Trust in public institutions (%)

	EB 96 (2021–2022)	EB 97 (2022)	EB 98 (2022–2023)	EB 99 (2023)	Change after war
European Union	47	49	47	47	0
National government	35	34	32	32	–3
National parliament	36	34	33	33	–3

Source: Compiled by the authors.

Following the above, the Eurobarometer also indicates a little growth in optimism about the future for the EU (Table 3). The variation was greater when the war in Ukraine started (EB 97, 2022), with a 3% rise in optimistic views accompanied by a similar decrease in pessimism. This change later stabilised at 1% (EB 99, 2023); thus, opinion on the EU was not apparently negatively affected, especially at the end of 2022, and in 2023 when the problem became chronic.

Table 3:
Opinion on the future of the EU (%)

	EB 96 (2021–2022)	EB 97 (2022)	EB 98 (2022–2023)	EB 99 (2023)	Change after war
Optimistic	62	65	62	63	+1
Pessimistic	35	32	35	34	–1
Do not know	3	3	3	3	0

Source: Compiled by the authors.

Also noteworthy is the fact that trust in public institutions is the variable with the highest degree of change on the Eurobarometer. This is relevant because the level of trust could influence the commitment of the public to the saving measures implemented. Nevertheless, the combination of these categories reveals that citizens' perceptions of the political institutions in Europe, including the EU, have improved little during the energy crisis. These exploratory findings serve to contextualise the content analysis on the political communication on Telegram.

Strategies and thematic agenda on Telegram

The existing literature presents message functions as political management strategies (López-Meri et al., 2017). Our results show that the use of message functions on Telegram varies significantly among European parties (Table 4), but the organisation of political events was a key mission for both the ruling (50% in Bündnis 90/Die Grünen, 40% Renaissance) and populist opposition political parties (23.3% in RN, 23.5% in Vox) during the crisis.

Table 4:
Message functions by political party (%)

	DGrün (DE)	AfD (DE)	Renaissance (FR)	RN (FR)	PD (IT)	Fratelli d'Italia (IT)	PSOE (ES)	Vox (ES)	Total
	A	B	C	D	E	F	G	H	
Organisation of political events	50.0 BEFG	2.8	40.0 BEGF	23.3 BEF	7.8	3.3	4.4	23.5 BEF	13.8
Electoral programme / promises	22.2	12.1	13.3	8.3	26.6 GH	13.0	2.2	10.2	15.6
Criticism of the opponent	–	33.6	–	21.7	24.5	18.7	40.0	23.5	23.8
Ideology (values)	5.6	2.8	10.0	–	6.0	8.9	4.4	4.2	5.2
Political achievements	–	5.6	20.0 H	8.3	6.0	8.1	37.8 BDEFH	4.5	7.9
Reference to media	–	6.5 E	–	–	0.4	10.6 E	–	4.5 E	3.6
Appeal to mobilisation	16.7	6.5	6.7	13.3 E	3.5	22.0 BEH	–	4.5	7.4
Forwarding other messages	–	22.4 D	6.7	1.7	–	–	11.1	20.5 D	9.3
Others	5.9	7.5	3.3	23.3 H	25.2 BH	15.4 H	–	4.5	13.6

Note: Data with a significance level of .05 (in bold), based on two-tailed tests for the column proportion (Bonferroni correction).

Source: Compiled by the authors.

Criticism of opponents was also quite apparent throughout the sample (23.8% as average level), with only the Bündnis 90/Die Grünen and Renaissance parties not employing it as a strategy.

Moreover, political achievements are relevant for ruling parties (20% in Renaissance and 37.8% in PSOE); meanwhile, populist parties resort to appeals to mobilise (13.3% in RN and 22% in Fratelli d'Italia) and forwarding other messages (22.4% in AfD and 20.5% in Vox) as the main strategies. Therefore, populists seek not only to mobilise the public, but also to sharing with it other relevant messages. These messages include contents spread through meetings or legacy media.

With regard to thematic agendas, interesting empirical evidence is provided (Table 5). First, there is some relatively fragmented agenda in most populist parties (AfD, Fratelli d'Italia, Vox), dealing with many different topics compared to thematic concentration in ruling organisations (Bündnis 90/Die Grünen, Renaissance, PSOE).

*Table 5:
Thematic agenda by political party (%)*

	DGrün (DE)	AfD (DE)	Renaissance (FR)	RN (FR)	PD (IT)	Fratelli d'Italia (IT)	PSOE (ES)	Vox (ES)	Total
	A	B	C	D	E	F	G	H	
Elections and campaign	44.4 GH	14.0	10.0	25.0 H	20.9 H	34.1 BGH	4.4	7.2	17.5
Economy/ industry	5.6	29.9 DEFH	10.0	8.3	14.9	11.4	15.6	8.3	13.6
SMEs	–	0.9	–	–	–	–	–	0.8	0.3
Immigration	–	6.5 E	–	1.7	0.7	5.7 E	–	3.8	2.9
Covid-19 and healthcare	5.6	1.9	–	–	2.5	–	–	0.4	1.2
Social policy	11.1	11.2	16.7	8.3	13.1	10.6	35.6 BDEFH	8.7	12.2
National security	5.6	0.9	6.7	11.7 BEFH	1.4	0.8	–	2.7	2.5
Foreign relations	–	2.8	–	3.3	5.7 H	0.8	2.2	0.4	2.6
Facilities	–	2.8	–	–	4.6	0.8	–	1.1	2.2
Environment	16.7 F	11.2 F	–	–	5.7	1.6	–	15.5 EF	8.0
Science and technology	–	0.9	–	–	1.1	–	–	–	0.4
Personal issues	11.1	2.8	36.7 BEGH	26.7 BEGH	1.8	5.7	6.6	8.7 E	7.5
Political competition	–	3.7	6.7 E	10.0 E	0.4	2.4	24.4 BEFH	7.2 E	5.0
Relation with the media	–	9.3	3.3	–	–	–	–	5.7	2.8
Others	–	0.9	10.0	5.0	27.3 BD	26.0 BD	11.1	29.5 BD	21.4

Note: Data with a significance level of .05 (in bold), based on two-tailed tests for the column proportion (Bonferroni correction).

Source: Compiled by the authors.

The thematic concentration is greater among the parties in government. Thus, 44.4% of Bündnis 90/Die Grünen messages published on Telegram were about elections and campaign, being its most popular issue, followed by the environment (16.7%), personal issues (11%) and social policy (11%). The preponderance of personal issues in terms of the cabinet of ministers is also detected with Renaissance (Figure 1), which devotes 36.7% of its messages



Figure 1:
Example of a message
on personal issues (Renaissance, FR), applying pride as an emotion
Source: Renaissance party Telegram channel (t.me/PartiRenaissance).

to personal issues, compared against social policy (16.7%), elections and campaign (10%) or economy/industry (10%). For its part, PSOE addresses social policy (35.6%) and political competition (24.4%) as significant categories.

Likewise, huge differences are found in the issue preference, but the importance of elections and campaigns in several countries (17.5% of the sample) is noteworthy. Besides that, many topics work as priorities for only specific parties such as economy/industry (29.9% in AfD), social policy (35.6% in PSOE), national security (11.7% in RN), personal issues (36.7% in Renaissance with Macron, 26.7% in RN with Le Pen) or political competition (24.4% in PSOE).

Nevertheless, some political messages on Telegram allude to more than one topic, which was coded properly; for instance, the right-wing populist AfD links economy/industry to migration (Figure 2), adding a feeling of anger that is now examined in the analysis of emotions.



Figure 2:
Example of message concerning the issues
of economy/industry and migration (AfD, DE), accompanied by anger as emotion
Source: AfD party Telegram channel in the Bundestag (t.me/afdfraktionimbundestag).

Appeal to emotions in the political discourse on Telegram

Once again, categories are much more concentrated for ruling parties (Figure 3), ‘pride’ standing out for Renaissance (60%). Although emotions represent a classic element of the rhetoric of populist political actors, appeals to emotions are massively present for all the parties selected. This finding is aligned with previous literature on leaders (Alonso-Muñoz & Casero-Ripollés, 2023), but the way our study illustrates this beyond populism is remarkable since only 21.3% of the analysed messages on Telegram do not use emotions.

Positive and negative feelings are mixed as shown by Bündnis 90/Die Grünen, which focuses on sadness (22.2%), happiness (22.2%) and fear (11.1%). The latter is also common for the German populist party analysed (15% in AfD). For its part, happiness is important for many different parties (22.2% in Die Grünen, 16.8% in AfD, 16.7% in Renaissance, 20% in RN, 21.3% in PD, and 27.6% in Fratelli d’Italia), but at the same time some of those populist organisations use anger (22.4% in AfD, 23.3% in RN).

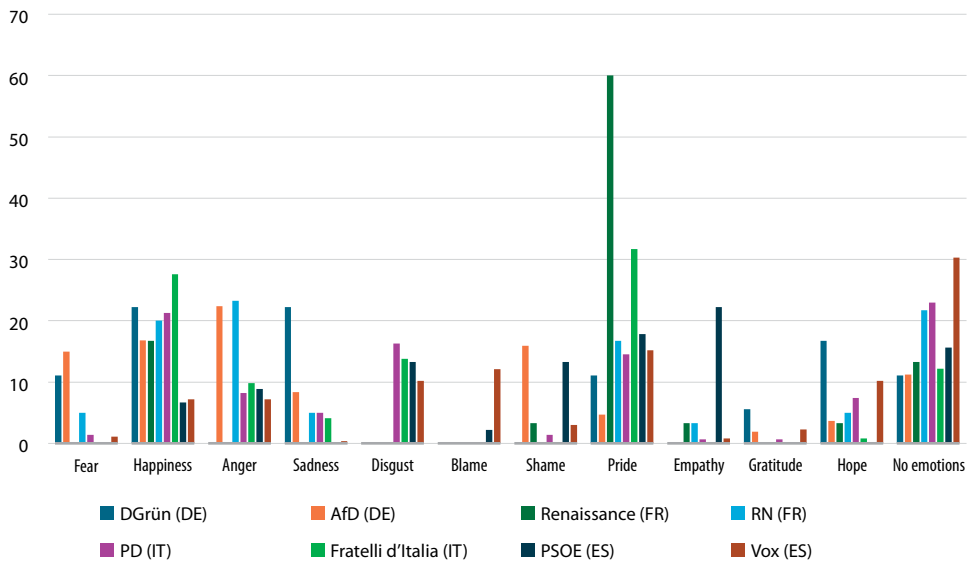


Figure 3:
Emotions on Telegram (%)
Source: Compiled by the authors.

With regard to positive feelings, pride and empathy are also found. Pride is the most frequent emotion (17.5%), reaching high levels for political organisations across Europe such as Renaissance (60%), Fratelli d'Italia (31.7%), PSOE (17.8%), RN (16.7%), Vox (15.2%) and PD (14.5%). The use of pride is significant for Renaissance and Fratelli d'Italia. Conversely, empathy is only detected for PSOE (22.2%), connecting this emotion to topics such as social policy.

The relationship between emotions and message functions reveals the purpose for which the appeal to emotions is used (Table 6). Indeed, an absence of emotions is essential for the organisation of political events (43%), references to media (36.4%), forwarding other messages (24.4%) and others (51.6%). Thus, some political issues are communicatively shaped on Telegram without emotions.

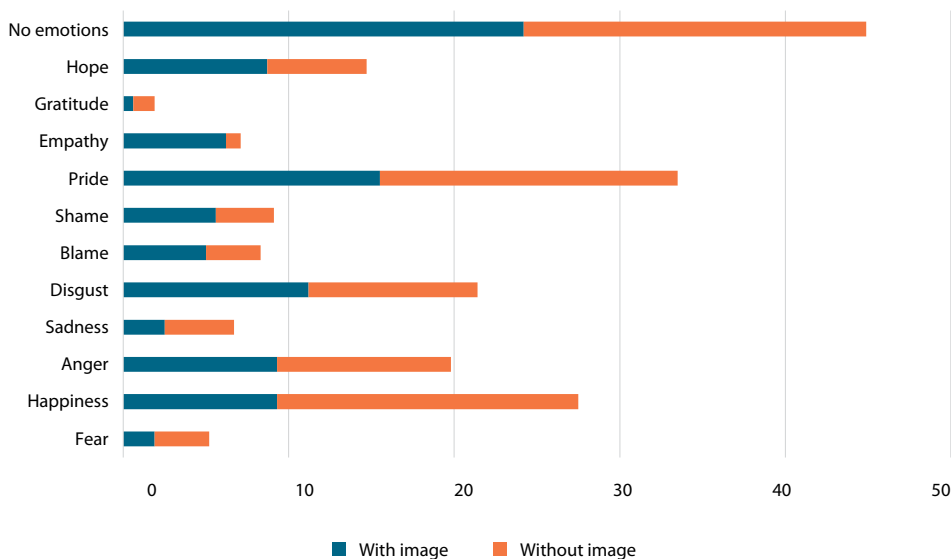
Despite the above, most emotions connect with message functions. Happiness is linked to electoral programme/promises (27.6%), disgust to criticism of an opponent (36.2%), and pride to ideology (52.1%), political achievements (54.8%) and appeals to mobilise (33.3%). The feeling of pride serves a diversity of objectives that range as functions employed by both ruling and populist parties.

As for the extent to which the variable of 'other' message functions do not convey emotions (51.6%), future research is needed to reflect upon why additional functions are less likely to convey feelings. On this matter, a classic element in conveying emotions is the employment of images. According to the results (Figure 4), images are not as important as might have been expected. Only messages involving disgust (11.2%), blame (5%), shame (5.6%), empathy (6.2%), and hope (8.7%) tended to have images.

*Table 6:
Emotions by message functions (%)*

	Organise political events	Electoral programme / promises	Criticism of the opponent	Ideology (values)	Political achievements	Reference to media	Appeal to mobilisation	Forwarding other messages	Other
Fear	3.9	0.7	4.5	–	–	3.0	2.9	3.5	4.8
Happiness	27.3	27.6	0.5	18.8	23.3	27.3	21.7	15.1	12.7
Anger	0.8	3.4	25.8	10.4	–	9.1	8.7	11.6	7.1
Sadness	1.6	2.1	3.2	4.2	1.4	3.0	2.9	2.3	12.7
Disgust	0.8	2.1	36.2	2.1	–	18.2	–	3.5	1.6
Blame	–	–	13.1	2.1	–	–	–	2.3	0.8
Shame	–	1.4	10.0	4.2	–	–	–	11.6	–
Pride	15.6	26.2	0.9	52.1	54.8	–	33.3	10.5	4.8
Empathy	0.8	0.7	–	–	13.7	–	1.4	2.3	1.6
Gratitude	–	2.1	–	2.1	–	–	1.4	5.8	0.8
Hope	6.3	22.1	–	–	2.7	3.0	13.0	7.0	1.6
No emotions	43.0	11.7	5.9	4.2	4.1	36.4	14.5	24.4	51.6

Source: Compiled by the authors.



*Figure 4:
Emotions by type of message (%)*
Source: Compiled by the authors.

PSOE

NUEVO PAQUETE DE MEDIDAS DEL GOBIERNO

<https://cadenaser.com/nacional/2022/10/13/guia-practica-para-entender-como-afectara-el-nuevo-plan-de-ahorro-energetico-cadena-ser/>

cadena SER

Guía práctica para entender cómo afectará el nuevo plan de ahorro energético

El presidente del Gobierno, Pedro Sánchez, ha anunciado un nuevo paquete de medidas de ahorro energético que se aproba...



Figure 5:

Example of social policy with image (PSOE, ES), applying empathy as emotion

Source: PSOE party Telegram channel (t.me/canalPSOE)

For their part, images seem to be avoided for fear (3.3%), happiness (18.2%), anger (10.5%), sadness (4.2%), pride (18%) and gratitude (1.3%). It should be noted that pride or happiness were some of the most frequent categories. They have been widely employed throughout to disseminate sympathy currents about leaders and parties (López-López et al., 2020).

Also, in considering positive feelings, empathy is one in which images are preferred (Figure 5). This was particularly detected for the social democratic party PSOE in Spain. However, political parties do not usually employ images to support appeals to emotions. By contrast, have a higher percentage of images appear in messages without emotions (24.2%), which reveals that there is no clear relationship between both variables on Telegram.

Discussion and conclusions

This article aimed to examine the political messages on Telegram and some of the perceptions of public opinion during the energy crisis in Europe. Telegram has been mostly studied to evaluate mobilisation strategies (Jost & Dogruel, 2023). However, scant attention has been paid to the employment of emotion. Accordingly, our study provides three inter-related contributions on European political communication on Telegram for that period.

First, we illustrate how trust in national institutions has been eroded compared to the EU, which was the main source of trust during the period of the energy crisis. Nevertheless, this setback was slight and opinions on the future of Europe were not affected (RQ1).

Besides that, the evolution could have been impacted by additional factors. Overall, changes over time were modest and did not indicate any major shift in public attitudes. This relative stability suggests a degree of resilience in citizens' perceptions despite the challenging context.

With regard to RQ2, we offer insightful findings on the distribution of message functions and thematic agenda. Political actors mostly use Telegram for the organisation of events and criticism of their opponents, which are classic functions of political action. The results on agendas are different as this is relatively fragmented for most populist parties; meanwhile, organisations in government address fewer topics. Elections/programme as a topic is a common finding within the sample, but ruling parties are more likely to focus on political achievements, and populist parties do it when appealing for mobilisation and forwarding other messages.

Furthermore, many topics are priorities for only specific parties, revealing strong differences between parties and ideologies. For instance, the thematic focus of Telegram messages varies noticeably across parties and national contexts. In France, communication by Renaissance under Macron is more strongly oriented toward personal issues, whereas the RN led by Le Pen places comparatively less emphasis on this dimension. In Spain, by contrast, messages from the PSOE are more frequently centred on political competition. At the same time, several Telegram posts address multiple topics simultaneously and were coded accordingly. A clear example can be found in the communication of German right-wing populist AfD, which often connects economic and industrial concerns with migration, while simultaneously introducing an explicit tone of anger that is further examined in the analysis of emotions.

Our third contribution furthers our understanding of the working of emotions. The presence of emotions can be clearly and explicitly observed throughout the sample (RQ3), but they are not always accompanied by images. The parties in government tend to use less variety in appeals to this type of feeling. Sadness, happiness and fear are key emotions, together with pride, which means combining positive and negative emotions. These categories overlap with message functions, depending on the objective pursued. For instance, happiness is linked to electoral programme/promises or disgust to criticism of the opponent.

Our findings show an emotional narrative in the parties, but with relevant differences. The results are in line with prior scholarship that has extensively described how positive and negative emotions are conducted together to reach the same objective (Alonso-Muñoz & Casero-Ripollés, 2023). On this matter, the accounts from some countries (Germany and France) have minimal activity, whereas others produce a large number of messages, linked to political events.

Beyond that, trust in public institutions has diminished a little. Nevertheless, it should be taken into account that some authors consider Eurobarometer to be propagandistic and barely reliable (Höpner & Jurczyk, 2015). Parties in government have responded to this national distrust by focusing on political achievements, together with emotions such as pride. Likewise, negative emotions are more present in the fragmented agenda of populist parties, using appeals to mobilisation and forwarding other messages as the main message functions.

Finally, we acknowledge some limitations in this study. The most prolific political parties on Telegram are overrepresented, especially Partito Democratico in Italy and Vox in Spain. Future research could expand the scope of this work by applying more extended time

frames on larger samples, including additional parties, and also dealing with the potential conversation between politicians and citizens through these mobile instant messaging services. Scholars face the challenge of providing empirical findings on the role of mobile services such as Telegram in political polarisation (Tucker et al., 2018), considering incidental exposure to these messages on social media (Fletcher & Nielsen, 2018).

In conclusion, the analysis of messages on Telegram contributes to budding literature on the use of mobile instant messaging services for political activities. Social media was traditionally applied as a tool of opposition by populists, blaming the elite (van Kessel & Castelein, 2016). Populist communication strategically mobilises emotions such as anger by constructing antagonistic narratives that contrast a morally pure ‘people’ against corrupt or unresponsive elites.

Although our article also shows a greater negativity in populism, we found out how positive and negative emotions were combined to mitigate the effects of a crisis. Emotions thus function as a discursive hinge linking populist rhetoric to broader dynamics of institutional trust. The agenda of populists in opposition showed greater plurality than the parties in government, but all of them use emotions widely to share their proposals, which may have implications in the shaping of public debate.

References

- Alonso-Muñoz, L. & Casero-Ripollés, A. (2023). The Appeal to Emotions in the Discourse of Populist Political Actors from Spain, Italy, France and the United Kingdom on Twitter. *Frontiers in Communication*, 8. Online: <https://doi.org/10.3389/fcomm.2023.1159847>
- Alonso-Muñoz, L., Tirado-García, A. & Casero-Ripollés, A. (2022). Telegram in Campaign: The Use of Mobile Instant Messaging Services in Electoral Political Communication. *Communication & Society*, 35(4), 71–88. Online: <https://doi.org/10.15581/003.35.4.71-88>
- Bimber, B. (2014). Digital Media in the Obama Campaigns of 2008 and 2012: Adaptation to the Personalized Political Communication Environment. *Journal of Information Technology & Politics*, 11(2), 130–150. Online: <https://doi.org/10.1080/19331681.2014.895691>
- Block, E. & Negrine, R. (2017). The Populist Communication Style: Toward a Critical Framework. *International Journal of Communication*, 11(1), 178–197.
- Bossetta, M. (2018). The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism & Mass Communication Quarterly*, 95(2), 471–496. Online: <https://doi.org/10.1177/1077699018763307>
- Boulianne, S., Koc-Michalska, K. & Bimber, B. (2020). Right-wing Populism, Social Media and Echo Chambers in Western Democracies. *New Media & Society*, 22(4), 683–699. Online: <https://doi.org/10.1177/1461444819893983>
- Bracciale, R., Andretta, M. & Martella, A. (2021). Does Populism Go Viral? How Italian Leaders Engage Citizens through Social Media. *Information, Communication & Society*, 24(10), 1477–1494. Online: <https://doi.org/10.1080/1369118X.2021.1874472>
- Caiani, M. & Guerra, S. (Eds.). (2017). *Euro-scepticism, Democracy and the Media. Communicating Europe, Contesting Europe*. London: Palgrave Macmillan.

- Carral, U., Tuñón, J. & Elías, C. (2023). Populism, Cyberdemocracy and Disinformation: Analysis of the Social Media Strategies of the French Extreme Right in the 2014 and 2019 European Elections. *Humanities and Social Sciences Communications*, 10(23), 1–12. Online: <https://doi.org/10.1057/s41599-023-01507-2>
- Dimitrova, D. V. & Matthes, J. (2018). Social Media in Political Campaigning around the World: Theoretical and Methodological Challenges. *Journalism & Mass Communication Quarterly*, 95(2), 333–342. Online: <https://doi.org/10.1177/1077699018770437>
- Dubois, E. & Blank, G. (2018). The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media. *Information, Communication & Society*, 21(5), 729–745. Online: <https://doi.org/10.1080/1369118X.2018.1428656>
- Fletcher, R. & Nielsen, R. K. (2018). Are People Incidentally Exposed to News on Social Media? A Comparative Analysis. *New Media & Society*, 20(7), 2450–2468. Online: <https://doi.org/10.1177/1461444817724170>
- Fuchs, C. (2017). *Social Media. A Critical Introduction*. London: Sage.
- Gerbaudo, P. (2018). Social Media and Populism: An Elective Affinity? *Media, Culture & Society*, 40(5), 745–753. Online: <https://doi.org/10.1177/0163443718772192>
- Hameleers, M., Bos, L. & de Vreese, C. H. (2016). ‘They Did It’: The Effects of Emotionalized Blame Attribution in Populist Communication. *Communication Research*, 4(6), 870–900. Online: <https://doi.org/10.1177/0093650216644026>
- Höpner, M. & Jurczyk, B. (2015). How the Eurobarometer Blurs the Line between Research and Propaganda. *MPIfG Discussion Paper*, (15/6). Cologne: Max Planck Institute for the Study of Societies.
- Jacobs, S. H. J. & Wonneberger, A. (2019). Dealing with Increasing Complexity: Media Orientations of Communication Managers in Public Sector Organizations. *International Journal of Communication*, 13, 918–937.
- Jagers, J. & Walgrave, S. (2007). Populism as Political Communication Style: An Empirical Study of Political Parties’ Discourse in Belgium. *European Journal of Political Research*, 46(3), 319–345. Online: <https://doi.org/10.1111/j.1475-6765.2006.00690.x>
- Jost, P. & Dogruel, L. (2023). Radical Mobilization in Times of Crisis: Use and Effects of Appeals and Populist Communication Features in Telegram Channels. *Social Media + Society*, 9(3), 1–12. Online: <https://doi.org/10.1177/20563051231186372>
- Jungherr, A., Schroeder, R. & Stier, S. (2019). Digital Media and the Surge of Political Outsiders: Explaining the Success of Political Challengers in the United States, Germany, and China. *Social Media + Society*, 5(3), 1–12. Online: <https://doi.org/10.1177/2056305119875439>
- Kiess, J. (2025). Euroscepticism and Local Far-Right Mobilization via Telegram in Light of the Fundamental Transformation of the Public Sphere. *Political Studies Review*, 23(2), 635–642. Online: <https://doi.org/10.1177/14789299231190731>
- Kreiss, D., Lawrence, R. G. & McGregor, S. C. (2018). In Their Own Words: Political Practitioner Accounts of Candidates, Audiences, Affordances, Genres, and Timing in Strategic Social Media Use. *Political Communication*, 35(1), 8–31. Online: <https://doi.org/10.1080/10584609.2017.1334727>
- Krippendorff, K. (2004). *Content Analysis. An Introduction to Its Methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- López-López, P. C., Oñate, P. & Rocha, Á. (2020). Social Media Mining, Debate and Feelings: Digital Public Opinion’s Reaction in Five Presidential Elections in Latin America. *Cluster Computing*, 23, 1875–1886. Online: <https://doi.org/10.1007/s10586-020-03072-8>

- López-Meri, A., Marcos-García, S. & Casero-Ripollés, A. (2017). What Do Politicians Do on Twitter? Functions and Communication Strategies in the Spanish Electoral Campaign of 2016. *Profesional de la información*, 26(5), 795–804. Online: <https://doi.org/10.3145/epi.2017.sep.02>
- Marcos-García, S., Tirado-García, A. & Casero-Ripollés, A. (2023). Respread, Mobilize and Fragment the Agenda: The Secondary Role of Telegram in the Spanish Electoral Campaign of November 2019. *Revista de Comunicación*, 22(1), 185–206. Online: <https://doi.org/10.26441/RC22.1-2023-2993>
- Mazzoleni, G. & Sfondini, A. (2009). *Politica pop. Da 'Porta a Porta' a 'L'isola dei famosi'*. Bologna: Collata Contemporanea Il Mulino.
- McAllister, I. (2007). The Personalization of Politics. In R. Dalton & H. D. Klingemann (Eds.), *The Oxford Handbook of Political Behaviour* (pp. 571–588). Oxford – New York: Oxford University Press.
- Nai, A. (2021). Fear and Loathing in Populist Campaigns? Comparing the Communication Style of Populists and Non-populists in Elections Worldwide. *Journal of Political Marketing*, 20(2), 219–250. Online: <https://doi.org/10.1080/15377857.2018.1491439>
- Neuendorf, K. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.
- Newman, N., Fletcher, R., Kalogeropoulos, A. & Nielsen, R. K. (2019). *Reuters Institute Digital News Report 2019*. Oxford: Reuters Institute for the Study of Journalism.
- Nielsen, R. K. & Fletcher, R. (2023). Comparing the Platformization of News Media Systems: A Cross-country Analysis. *European Journal of Communication*, 38(5), 484–499. Online: <https://doi.org/10.1177/02673231231189043>
- Palmero Cantero, F., Fernández-Abascal, E., Martínez-Sánchez, F. & Chóliz Montañes, M. (Eds.) (2002). *Psicología de la motivación y las emociones*. Madrid: McGraw Hill.
- Papacharissi, Z. (2016). Affective Publics and Structures of Storytelling: Sentiment, Events and Mediality. *Information, Communication & Society*, 19(3), 307–324. Online: <https://doi.org/10.1080/1369118X.2015.1109697>
- Patterson, T. E. (1993). *Out of Order*. New York: Alfred A. Knopf.
- Poell, T., Nieborg, D. & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4), 1–13. Online: <https://doi.org/10.14763/2019.4.1425>
- Rivas-de-Roca, R. & García-Gordillo, M. (2022). Brussels Will Pay for Everything. The Shaping of the European Public Sphere before NextGenerationEU. *Profesional de la información*, 31(3). Online: <https://doi.org/10.3145/epi.2022.may.01>
- Rivas-de-Roca, R., Pérez-Curiel, C. & Casero-Ripollés, A. (2024). Effects of Populism: The Agenda of Fact-checking Agencies to Counter European Right-wing Populist Parties. *European Journal of Communication*, 39(2), 105–121. Online: <https://doi.org/10.1177/02673231231197032>
- Rivera Otero, J. M., Lagares Diez, N., Jaráiz Gulías, E. & López-López, P. C. (2021). Emociones y engagement en los mensajes digitales de los candidatos a las elecciones generales de 2019. *Cultura, Lenguaje y Representación*, 26(2), 229–245. Online: <https://doi.org/10.6035/clr.5844>
- Rogers, R. (2020). Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media. *European Journal of Communication*, 35(3), 213–229. Online: <https://doi.org/10.1177/0267323120922066>
- Ronzhyn, A., Cardenal, A. S. & Batlle Rubio, A. (2023). Defining Affordances in Social Media Research: A Literature Review. *New Media & Society*, 25(11), 3165–3188. Online: <https://doi.org/10.1177/14614448221135187>
- Shirky, C. (2011). The Political Power of Social Media: Technology, the Public Sphere, and Political Change. *Foreign Affairs*, 90(1), 28–41.

- Sierra, A., González-Tosat, C. & Rodríguez-Virgili, J. (2022). El uso de Telegram por los partidos políticos españoles en las elecciones generales de 2019. *Observatorio (OBS*)*, 16(1), 138–159. Online: <https://doi.org/10.15847/obsOBS16120221959>
- Stier, S., Bleier, A., Lietz, H. & Strohmaier, M. (2018). Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter. *Political Communication*, 35(1), 50–74. Online: <https://doi.org/10.1080/10584609.2017.1334728>
- Tirado-García, A. (2023). The Negative Campaign on Telegram: The Political Use of Criticism during the 2021 Community of Madrid Elections. *Social Sciences*, 12(2). Online: <https://doi.org/10.3390/socsci12020093>
- Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Nyhan, B. (2018). *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Social Science Research Network. Online: <https://doi.org/10.2139/ssrn.3144139>
- Valenzuela, S., Bachmann, I. & Bargsted, M. (2021). The Personal Is the Political? What Do WhatsApp Users Share and How It Matters for News Knowledge, Polarization and Participation in Chile. *Digital Journalism*, 9(2), 155–175. Online: <https://doi.org/10.1080/21670811.2019.1693904>
- Valeriani, A. & Vaccari, C. (2018). Political Talk on Mobile Instant Messaging Services: A Comparative Analysis of Germany, Italy, and the UK. *Information, Communication & Society*, 21(11), 1715–1731. Online: <https://doi.org/10.1080/1369118X.2017.1350730>
- Van Kessel, S. & Castelein, R. (2016). Shifting the Blame. Populist Politicians' Use of Twitter as a Tool of Opposition. *Journal of Contemporary European Research*, 12(2), 594–614. Online: <https://doi.org/10.30950/jcer.v12i2.709>
- Verboord, M., Janssen, S., Kristensen, N. N. & Marquart, F. (2025). Institutional Trust and Media Use in Times of Cultural Backlash: A Cross-National Study in Nine European Countries. *The International Journal of Press/Politics*, 30(3), 752–774. Online: <https://doi.org/10.1177/19401612231187568>
- Vermeer, S. A. M., Kruikemeier, S., Trilling, D. & de Vreese, C. H. (2021). WhatsApp with Politics?!: Examining the Effects of Interpersonal Political Discussion in Instant Messaging Apps. *The International Journal of Press/Politics*, 26(2), 410–437. Online: <https://doi.org/10.1177/1940161220925020>
- Widmann, T. (2021). How Emotional Are Populists Really? Factors Explaining Emotional Appeals in the Communication of Political Parties. *Political Psychology*, 42(1), 163–181. Online: <https://doi.org/10.1111/pops.12693>

This page intentionally left blank.

Data Pollution as Resistance

A Commentary on Partisan Responses to Government-Dependent Pollsters before the 2026 Hungarian Parliamentary Elections

János József Tóth*^{ORCID}

* Ludovika University of Public Service, Office for Science Strategy; Ludovika University of Public Service, Department of Social Communication; HUNGARY
E-mail: toth.janos.jozsef[at]uni-nke.hu

Prior to the 2026 Hungarian parliamentary elections, pro-government and government-independent pollsters published, quite consistently, hugely different figures on the levels of support for the main government and the main opposition parties. In the wake of the elections, it became apparent that rather than the ruling Fidesz party finishing narrowly ahead, as predicted, government-independent pollsters' numbers were much closer to the supermajority the opposition party achieved on the 12 April voting. This commentary reflects on the insincerity of the partisan cheerleader influencing surveys conducted by government-dependent pollsters as reported on Hungarian-language Reddit forums, and consider it a possible contributing factor in their forecasts being less accurate than those of government-independent ones.

Keywords: election forecasting, response substitution, expressive responding, partisanship, survey quality

Article received on 27 April 2026. Article accepted on 4 May 2026.

Conflict of interest: János József Tóth is a member of the Hungarian Sociological Association and the Editor-in-Chief of KOME.

Funding: The author received no financial support for the research, authorship, or publication of this article.

2026 Hungarian parliamentary election polls in retrospect

At the 2026 Hungarian parliamentary elections, the Tisza party won a two-thirds supermajority, thus Hungary joined the recent process of weakening in the right-wing populism of the Visegrád countries, driven by mix of economic frustration, high inflation rates and anti-corruption sentiments. The landslide victory of Péter Magyar, himself a former lower-tier beneficiary of the illiberal regime of the Fidesz party that governed the country between 2010 and 2026, is especially interesting from the perspective of communication studies because it occurred in the midst of/despite an information autocracy (Guriev & Treisman, 2019, 2020; Sik & Krekó, 2025) exercised by the leader of Fidesz and Prime Minister Viktor Orbán. Relying on the volume of government propaganda, indirect censorship, foreign (mostly Russian) interference, in addition to the results of pre-election opinion polling, the former ruling party and its supporters could have made a reasonable bet on the results of the election being under their control, the general sentiment being that Fidesz would win although, in contrast to most of their previous 16 years of governance, they would not get a supermajority on their own. When it turned out that the reality was far from what Fidesz voters, and perhaps even the party leadership expected, a discussion began about what might have caused the differences between expectations and perceptions, a discussion which is still in its early stages and would, most likely, translate to a variety of research projects in the field of political sciences, media, communication and journalism studies.

Early consensus seems to indicate that, among other effects, trust in the polling results of government-dependent pollsters and distrust in government-independent one's might have contributed to the false perception of the volume of Fidesz's support among their voters, as the figures these two published were consistently different over the past year, with government-dependent pollsters reporting a stable Fidesz lead over the Tisza party. Ten days prior to the elections, an exhaustive comparison of polling numbers measured by independent and pro-government Hungarian pollsters for the 2026 Hungarian parliamentary elections was published by *Átlátszó.hu*, a watchdog NGO and a collective of investigative journalists (Zubor & Szabó, 2026). Their data showed that between January 2025 and 2026, government-independent pollsters: 21KK, the Compass Institute, IDEA, Medián, Republikon, Publicus and Závecz Research measured support among voters for Fidesz at between 32% and 40%, and Tisza support between 39% and 53% across the year. Government-dependent pollsters like the Institute of the Twenty-first Century, the Center for Fundamental Rights, Magyar Társadalomkutató LLC, McLaughlin & Associates, Nézőpont Institute and Real-PR 93 LLC measured a higher range (between 43% and 51%) for Fidesz support and a lower range (between 35% and 42%) for Tisza support.

Naturally, government-owned or affiliated pollsters in information autocracies do not often conduct polls for the sake of simply measuring but rather to influence public opinion and voting behaviour, and their sampling and question design might therefore be less solidly based on scientific criteria than on the need to conform to and support a government they have strong financial ties with. It seems that there is strong support among the country's research community for this line of thinking being applicable to the Hungarian situation, as signalled by the position of the Hungarian Sociological Association published on the 20th of April. In their communiqué (2026), the Association challenged several government-dependent

polling institutes by name to publish their database and the precise methods upon which they based their research, voicing the suspicion of incompetence or data manipulation, and citing concerns of eroding public trust in polling and ongoing politically driven attacks against pollsters and academics.

Although we think that these claims merit transparency and further investigation, we would like to draw some attention to the possibility that these differences in reported numbers may also be linked to external factors, resulting in contaminated data being used to generate the results in question. This commentary offers insights on possible data contamination affecting only government-dependent pollsters due to a specific type of expressive survey response so far less observed in the Hungarian context; fuelled by partisan activism and social media effects on Hungarian-language Reddit channels, especially on r/hungary.

Hungarian-language Reddit and anti-government partisanship

The use of social media to convey political messages is a phenomenon almost as old as the platforms themselves. Political communication research (and practice) is generally quick to keep up with new platforms and formats, but since there are significant differences in popularity and reach between social media platforms across different regions and age groups, some platforms are more often sampled and better covered in political communication research focusing on specific countries. In Hungary, such research is mostly focused on Facebook, Instagram, YouTube, and (more recently) TikTok, while the role of Reddit has so far eluded research interest, despite it being slowly discovered by Hungarian political actors and partisan activism in recent years. I expect this focus in coverage to adjust in the years ahead.

Reddit is unique among social networking platforms in the sense that it provides users with the actual ability to create self-governing forums ('subreddits') while letting members make use of the publicity of such forums without forcing them to surrender their anonymity. Users here engage in complex public social interactions, which makes Reddit an ideal setting for both politicians and activists to engage communities and generate feedback on events of public interest. Some U.S. politicians have been using Reddit for AMAs¹ since the early 2010s; the most notable early adopter being Barack Obama from 2012, his example was followed by not only active politicians but political operatives and congressional staff members as well. In Hungary, besides relying on traditional media outlets, the majority of which were under governmental influence and closely coordinated, pro-government actors and politicians have a history of being most active on Facebook, where audience characteristics and Meta's advertising system provide optimal environment for government-financed propaganda. Reddit – which is popular among younger urban users

¹ Abbreviation for "Ask Me Anything"; posts under which users can submit questions as comments directly to the Original Poster (OP). These questions are usually ranked by the community through public votes which the OP answers in order of popularity during a designated timeframe. The more popular and visible the question, the better the chances of it being answered.

with higher digital literacy – however, remained mostly a platform for government-critical discussions. When it came to campaigning and endorsement, Hungarian subreddits are almost exclusively used by government-critical actors. High-profile politicians like Péter Márki-Zay (leader of the united opposition for the 2022 national election), the newly elected Prime Minister, Péter Magyar, and the new Minister of the Prime Minister's Office, Bálint Ruff, held multiple AMAs in the last few years on the biggest Hungarian-language subreddit; r/hungary. This sub has 5–600k visitors weekly, actively moderates political advertising and linking to certain types of media associated with governmental propaganda, and generally has a government-critical sentiment through its user-generated content. Its political impact seems to be meaningful in the sense that members of the new Tisza Government seem to follow and react to feedback received through this community. Beside direct engagement through AMAs; for example when rumours began to spread about the person of the future Minister of Education; the community of r/hungary vocally expressed their discontent about Rita Rubovszky, an education professional with a conservative Christian background. This feedback was addressed by Péter Magyar in one of the first press conferences he held after it became certain that his party had won the elections, and he asked the public not to lynch rumoured candidates – or anyone else – verbally, not even on Reddit (ATV Magyarország, 2026). It is also worth mentioning that after a few days, Tisza announced its new Minister of Education in the person of Judit Lannert.

Besides politician AMAs and feedback to events of political significance, r/hungary was also used by various individual accounts in the past years to report on occasions when a pollster had contacted them by telephone. Based on the general content of such threads – the majority of them are still accessible and can be found with a simple keyword search – these posts and comments discussed the questions the pollsters had asked, whether they were pro-government or not, and the professionalism of the person or AI assistant asking the questions.

Many of the aforementioned posts and comments were focused on presenting evidence that the polling entity was associated with the government, and offering opinions on this characteristic. The evidence offered was both direct (when the interviewer introduced themselves as being associated with a known pro-government pollster) or indirect (based on the type of questions they asked; as certain loaded questions, topics, style and language use are connected to well-known government narratives), and evaluations of perceived pro-government-ness or government dependency was generally negative. Discussions under posts mentioning pro-government pollsters and their questionnaires also contained reports from individual users promoting the idea of how one should contaminate pro-government pollsters' data to ensure the results they got would be unusable without them knowing that would actually be the case.

The first step was to ensure that they did not accidentally pollute government-independent polling data. If the polling entity was not clearly identifiable from the introduction or they did not introduce themselves at all, users wanted to obtain that information anyway through directly asking the interviewer. It seems that, at least in some cases, interviewers did not want to disclose this information:

“They cut me off when I asked which polling company they were with”, one user reported (drnst, 2025). This behaviour can be explained either by legal or ethical issues the interviewer was aware of (e.g. they had not obtained the interviewee's data in compliance with GDPR

or other applicable laws) or by the presupposition that the respondent's answers would be biased by knowing the political affiliation of the pollster; rendering them unusable.

Second, when a government-dependent pollster was identified, motivated individuals tried to ensure that they would actually be selected for inclusion and not be filtered out by initial screening questions.

"I've already got it offpat: I am a 50-year-old with elementary school education living in a village and hesitating between Fidesz and the Good-for-Nothings [“mihasznák”; a reference to Our Hungary; a far-right Hungarian nationalist party]. I always tell them that.” (Adventurous-Ice3325, 2025).

Finally, once one was 'in', various individuals pushed other members of the community to give false answers to pollute the database and weaken the government party's ability to accurately assess their situation.

If you receive a call from Társadalomkutató (or any other unknown pollster working for Fidesz) you should tell them the opposite [of what your opinion is]. Let them believe everything is in order (gd42, 2024).

In my opinion, it is worth saying in these cases that you are a Fidesz voter. This will introduce a false input into their survey, and their strategy [drawn from the results] will be wrong (austurist, 2026).

[I told them that] I used to be a Fidesz supporter, but now I'd definitely vote for the Tisza Party. As for [question of] the room, I believe Péter Magyar when he says he didn't use drugs; in fact, my sympathy for him has grown, and my attitude toward Fidesz has become more negative. Meanwhile, I'm worse off than I was four years ago. Let Tóni [Antal Rogán, at that time Minister of the Prime Minister's Office] see that even Fidesz voters withdraw support (I was never a Fidesz voter) (erer1000, 2026).

This type of data pollution is difficult to clean (because countermeasures should have been built into the survey design and data cleaning methods beforehand) and, if introduced systematically, would potentially lead to untrustworthy results. Redditors also reported another opportunity for data pollution when the interviewer's apparent political biases were discussed. For example, when one respondent evaluated Viktor Orbán's actions consistently with the worst possible values, the interviewer started asking question like “are you sure about that, young man?” and proffered comments like “I wouldn't bet on that” or “interesting; that's not how most people see it” (OwlOnly4079, 2026). Others reported how interviewers were biased towards the very questions they had to ask:

[The interviewer] was cool, for example they did not even read out the question if I agreed with Viktor Orbán on many things and that he should continue his work. Instead they just marked the opposite end of the scale (GrassNo7577, 2026).

I had the same yesterday. Just like you did, the interviewer could hardly hold back their laughter (ElectricalMacaroon96, 2026).

This would raise questions not only about the methods based on which government-dependent pollsters had gathered their data but also about the recruitment of their interviewers, the training they received, and the quality control measures they implemented. A politically biased interviewer, or those with a ‘conspiratorial’ attitude could have encouraged partisan responding or may have contributed directly to the contamination of datasets with misleading information.

Self-reported information in Reddit posts and comments thus provide sporadic evidence on a type of expressive response as yet unexplored in Hungarian political communication. Beside the general characteristics of expressive responding, i.e. a sharp difference between misinformed answers and answers given with the aim of expressing the respondent’s identity commitments, such behaviour would point toward insincere partisan cheerleading (Peterson & Iyengar, 2021), where users intentionally distort their answers to signal support for their party. This support, in the Hungarian case, seems to have indirect parts since support to the preferred side was provided by hindering the disliked side’s ability to compile a database appropriate, after analysis, to identifying voter sentiment, and thus to draw conclusions for campaign strategies and to allocate resources to implement said strategies. In other words, these acts went beyond merely signalling support for the preferred side: they were intended to damage the opposing one.

Naturally, a verdict on the overlap between self-reported respondent behaviour on Reddit and the actual activities undertaken, as well as the extent, influence and other characteristics of Reddit discussions to respondent behaviour in the government-affiliated polling of the 2026 Hungarian parliamentary election cannot be reached without future studies analysing Reddit data and transparent, publicly accessible databases and the methods of said pollsters. This commentary serves as a mere reminder that discrepancies between pollsters or group of pollsters with different political affiliations can be affected by the strength, commitment, and level of organisation of partisan actions targeting pollsters with a specific (and apparent) political affiliation.

References

- Adventurous-Ice3325 (2025, September 2). *En mar betanultam, hogy egy 50 éves vidéki kistelepülésen élő* [Comment on the online forum post Engem is felhívott a közvéleménykutató]. Reddit. Online: <https://www.reddit.com/r/hungary/comments/1n6jnnp/comment/nc0xx6q/>
- ATV Magyarország (2026, April 17). *Magyar Péter először a parlamentben – sajtótájékoztató* [Video]. YouTube. Online: <https://www.youtube.com/watch?v=tgBL5EBxQho>
- austurist (2026, February 19). Szerintem ezekben az esetekben érdemes azt mondani, hogy Fideszes vagy [Comment on the online forum post Tegnap este sikerült felb@sznom egy “közvélemény kutatót”]. Reddit. Online: <https://www.reddit.com/r/hungary/comments/1r8sy1k/comment/o67hi0z/>
- drnst (2025, September 2). *Nálam akkor bontották, amikor megkérdeztem, melyik piackutató cég* [Comment on the online forum post Engem is felhívott a közvéleménykutató]. Reddit. Online: <https://www.reddit.com/r/hungary/comments/1n6jnnp/comment/nc1orn1/>

- ElectricalMacaroon96 (2026, March 04). *Ugyanezek voltak nálam is tegnap. Szintén alig bírta ki röhögés nélkül a kérdező* [Comment on the online forum post *Fideszes közvéleménykutató: Magyar ellenes szlogeneket kellett véleményeznem*]. Reddit. Online: <https://www.reddit.com/r/hungary/comments/1n6jnpr/comment/nc1orn1/>
- erer1000 (2026, February 19). Engem aki felhívott közvélemény kutató, azzal a kérdéseket összefoglalva [Comment on the online forum post *Tegnap este sikerült felb@sznom egy "közvélemény kutatót"*]. Reddit. Online: <https://www.reddit.com/r/hungary/comments/1r8sy1k/comment/o67mttu/>
- gd42 (2024, December 04). *Az ellenkezőjét kell mondani, ha a társadalomkutató (vagy egyéb ismeretlen, fidesznek dolgozó közvéleménykutató hív) [Comment on the online forum post *előbb hívtak közvéleménykutatás kapcsán – telefonos mérés*].* Reddit. Online: <https://www.reddit.com/r/hungary/comments/1h6cnby/comment/m0cl6lz/>
- GrassNo7577 (2026, March 04). *Nem nagyon. De jó fej volt és pl félsem olvasta a kérdést, hogy több mindenben egyetérték-e* [Comment on the online forum post *Fideszes közvéleménykutató: Magyar ellenes szlogeneket kellett véleményeznem*]. Reddit. Online: <https://www.reddit.com/r/hungary/comments/1n6jnpr/comment/nc1orn1/>
- Guriev, S. & Treisman, D. (2020). A Theory of Informational Autocracy. *Journal of Public Economics*, 186. Online: <https://doi.org/10.1016/j.jpubeco.2020.104158>
- Guriev, S. & Treisman, D. (2019). Informational Autocrats. *Journal of Economic Perspectives*, 33(4), 100–127. Online: <https://doi.org/10.1257/jep.33.4.100>
- Hungarian Sociological Association (2026) A közvélemény-kutatás nem vélemény, hanem szakma [Polling Is not Opinion, It Is a Profession]. Online: <https://szociologia.hu/allasfoglalasok/a-kozvelemeny-kutatas-nem-velemeny-hanem-szakma>
- OwlOnly4079 (2026, February 19). *Tegnap este sikerült felb@sznom egy "közvélemény kutatót"* [Online forum post]. Reddit. Online: https://www.reddit.com/r/hungary/comments/1r8sy1k/tegnap_este_siker%C3%BClt_felbsznom_egy_k%C3%B6zvelem%C3%A9nykut%C3%A9ny/
- Peterson, E. & Iyengar, S. (2021). Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading? *American Journal of Political Science*, 65(1), 133–147. Online: <https://doi.org/10.1111/ajps.12535>
- Sik, E. & Krekó, P. (2025). Hungary as an Ideological Informational Autocracy (IA) and the Moral Panic Button (MPB) as its Basic Institution. *Central and Eastern European Migration Review*, 14(2), 97–113. Online: <https://doi.org/10.54667/ceemr.2025.21>
- Zubor, Z. & Szabó, K. (2026, April 02). A Massive Gap Is Developing between Polling Results by Independent and Pro-Government Pollsters. *Átlátszó*. Online: <https://english.atlatszo.hu/2026/02/04/a-massive-gap-is-developing-between-polling-results-by-independent-and-pro-government-pollsters/>

Contents

MÁRTON DEMETER, FERENC PETRUSKA, GERGŐ HÁLÓ, ERIK PALUSEK: <i>PARALLEL EMPIRES OF KNOWLEDGE</i>	3
WENXI WEN, ASHRAF SADAT AHADZADEH, KAM-FONG LEE: <i>FROM TOOLS TO TEAMMATES: REIMAGINING MORAL AGENCY AND HUMAN-ROBOT PARTNERSHIPS IN JOURNALISM PRACTICE</i>	43
BEATRIZ CATALINA-GARCÍA, ANTONIO GARCÍA-JIMÉNEZ, MANUEL MONTES-VOZMEDIANO: <i>PERCEPTION OF INFORMATION AND DISINFORMATION ON SOCIAL MEDIA</i>	73
AUDREY BUENAVISTA MORALLO: <i>HIJACKED VICTIMHOOD: THE POLITICAL COMMUNICATION STRATEGY OF VICE PRESIDENT SARA DUTERTE AT CONGRESS HEARINGS</i>	95
RUBÉN RIVAS-DE-ROCA, CONCHA PÉREZ-CURIEL: <i>POLITICAL COMMUNICATION ON TELEGRAM</i>	115
JÁNOS JÓZSEF TÓTH: <i>DATA POLLUTION AS RESISTANCE</i>	137