Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Does the Language of the Prompt Change the Politics of the Answer? A Cross-Lingual Agent-Based Audit of Generative AI and Turkish Political Discourse

Sarphan Uzunoglu

sarphan.uzunoglu@ieu.edu.tr

İzmir University of Economics

Duygu Uzunoğlu

Ege University

Research Article

Keywords: generative AI, algorithm auditing, agent-based testing, framing, gatekeeping, comparative media systems, political communication, Turkey, multilingual bias, open science

Posted Date: September 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10519741/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Conversational generative AI increasingly mediates citizens' understanding of contested political events, yet audits remain predominantly English-language and US-centred. This article presents an exploratory cross-lingual Agent-Based Testing audit of how GPT-4o, GPT-5.5 and Gemini 2.5 Pro represented eight contested Turkish topics during a single collection window in July 2026. Crossing system, prompt language (Turkish/English), user persona and query framing produced 864 responses. These were coded for framing and gatekeeping using a large-language-model coder from a different model family, with author oversight, automated stability checks (Krippendorff's α = .80–.95) and human validation on 168 responses (α ≥ .72 for confirmatory constructs). We test the language-bound inheritance thesis: that models reproduce characteristics of the media system associated with the prompt language. The findings provide selective rather than general support. Prompt language barely affected frame alignment or actor slant, but altered sources of authority: Turkish prompts elicited greater reliance on Turkish government-aligned sources (21% vs 13%) and less reliance on international sources. Model identity and user persona produced stronger variation. Gemini was the most confident and government-aligned and most likely to reproduce a documented false narrative (24%); GPT-5.5 was the most balanced and reproduced none. Prompt language therefore shapes less the position an answer adopts than the evidentiary world from which it constructs that position. Stimuli, code and coding data are publicly available.

1. Introduction

The infrastructures through which citizens encounter accounts of contested political events are moving from retrieval to synthesis. A search engine returns a ranked list of external sources that a reader must still weigh; a conversational generative system such as ChatGPT or Gemini instead returns a single, fluent, authoritative-sounding answer. That answer is not a neutral technical convenience but an editorial act: the system decides what to include, what to leave out, how to order competing claims, and how to characterise contested actors—performing at scale the operations that define news work, but without the editorial accountability, transparency, or right of reply that constrain traditional news organisations. The long-standing observation that the "politics of search" governs which sources are made visible and which disappear (Introna & Nissenbaum, 2000) is thus intensified rather than resolved: the list of sources itself dissolves into a synthesised verdict that most users have neither the means nor the prompt to interrogate.

As these systems come to mediate political knowledge, the question of whose version of a contested reality they render visible becomes a democratic concern rather than a merely technical one. A one-sided account once required a partisan outlet to publish it and an audience to find it; a generative system can now deliver such an account to a single user as apparently settled fact, in the confident register of a neutral explainer. If the account a system offers depends systematically on features the user neither controls nor perceives, the system is quietly allocating epistemic authority across a polity—and doing so without the visibility that would let that allocation be contested.

What makes this more than a new venue for old worries about media bias is the peculiar position the synthesising intermediary occupies. A national media system has always been something one compared across countries; comparative media-systems theory was built to array whole polities against one another (Hallin & Mancini, 2004). A conversational model collapses that comparison into a single artefact. Trained on text in many languages, it does not belong to one media system but re-aggregates many, and it can be switched between them by the language of a prompt: the same system that answers an English question from an internationalised, heterogeneous corpus answers the Turkish version of that question from a corpus dominated by Turkey's own, far more concentrated, media ecology. The intermediary carries a plurality of national information orders inside it and serves whichever one the user's language selects - usually without the user knowing that a selection has been made. This is a genuinely new object for framing, gatekeeping, and comparative-systems theory alike: not a medium within a system, but a meta-intermediary that can instantiate different systems on demand.

Empirical scrutiny of political bias in large language models (LLMs), however, remains overwhelmingly Anglophone and monolingual. Audits generally locate a model on a United States left–right spectrum using English prompts and report detectable partisan leanings (Feng et al., 2023; Motoki et al., 2024; Rozado, 2023), but they rarely ask whether a single model behaves differently across languages for the same contested content. This leaves the stakes highest precisely where the evidence is thinnest: in non-English, highly polarised, and partially captured media environments, where the offline information order is already contested and where a model's outputs may weigh most heavily.

Turkey is a theoretically motivated site for this question, not merely a convenient one, and it approximates a natural experiment. Comparative media-systems research characterises national environments by their degree of political parallelism - the extent to which media content tracks and reinforces political divisions (Hallin & Mancini, 2004) - and Turkey exemplifies high parallelism compounded by extensive state capture, in which ownership consolidation and pressure have folded much of the mainstream press into the governing bloc (Yeşil, 2016; Uzunoğlu, 2022, 2025a). Crucially for a cross-lingual design, the Turkish- and English-language information environments surrounding the same Turkish events diverge sharply: Turkish-language coverage is dominated by state-aligned outlets, whereas the English-language record of those same events is carried disproportionately by international wire services, rights organisations, and exiled or independent journalists — a divergence that follows from the capture of Turkish-language media documented above (Yeşil, 2016; Uzunoğlu, 2022). The two languages therefore index two markedly different evidence ecologies for one set of facts - an almost designed contrast against which the language-bound inheritance thesis can be tested. Add a hybrid regime in which control over the information order is an instrument of government, a citizenry with high and rising reliance on digital and AI-mediated information, and a dense repertoire of contested events and documented disinformation narratives, and Turkey becomes a critical, most-likely case: if a model's politics travel with the language of the prompt anywhere, they should do so here - which makes both a positive finding and its absence theoretically informative.

This article conducts a cross-lingual Agent-Based Testing (ABT) audit of how three widely used systems—gpt-4o, gpt-5.5, and gemini-2.5-pro—represent contested Turkish political discourse, and whether that representation shifts with the language of the prompt. Its central claim is that a large language model inherits the structural properties of a media system through the language of its training data: because the model's Turkish competence is learned predominantly from Turkish-language text, the dominant frames and partisan sourcing of the Turkish media system may be sedimented into its Turkish outputs, whereas the same model's English outputs draw on a differently structured, more heterogeneous corpus. This reframes a cross-lingual comparison from a descriptive curiosity into a testable mechanism—training-data language → language-bound inheritance of media-system properties → language-differentiated framing, sourcing, and slant. Our expectations are accordingly directional: relative to a model's English outputs, we anticipate that its Turkish outputs align more closely with the dominant frames of Turkey's polarised media system, exhibit lower source diversity and a stronger skew toward government-aligned outlets, and display asymmetric hedging and refusal. We treat these as exploratory expectations rather than confirmatory tests, and we regard cross-lingual parity—should we find it—as itself a substantive result about how developers homogenise outputs across languages. Two framing commitments follow and are stated here at the outset: the study is exploratory rather than confirmatory, and it documents a single collection window (July 2026) for specific, versioned models rather than a fixed property of “ChatGPT” or “Gemini” as such; we accordingly read model-level contrasts as version-conditioned and release the version identifiers and timestamps for replication.

Across systems, prompt languages, and simulated user profiles, the audit asks: (RQ1) which frames dominate the outputs, and do they shift by language and persona? (RQ2) are sources provided, and how are government-aligned versus independent or oppositional Turkish outlets represented? (RQ3) is there systematic orientation toward specific actors? (RQ4) how often do systems present “both sides,” hedge, or refuse, and does this differ by language? (RQ5) when prompted with documented false narratives, do systems repeat, debunk, or leave them ambiguous? (RQ6) how much does output vary across repeated identical queries?

In addressing these questions the study makes four contributions that speak directly to this special issue. Empirically, it offers a cross-lingual (Turkish–English) audit of how generative AI represents a non-Western, high-parallelism political discourse. Theoretically, it extends comparative media-systems theory into the age of generative AI and transposes framing (Entman, 1993) and gatekeeping (Shoemaker & Vos, 2009) to a synthesising machine intermediary. Methodologically, it demonstrates an agent-based design that treats generative AI simultaneously as the audited object and as a component of the measurement instrument, and reflects on the reproducibility of audits against closed, drifting models. Normatively, it interprets systematic asymmetry through a critical, democratic-cost lens (Noble, 2018; Bender et al., 2021), continuing a line of work on the democratic costs of manipulated political communication in Turkey (Uzunoğlu, 2026a). The remainder of the article reviews the relevant literature and derives its expectations, details the agent-based design, reports the findings, and discusses their implications for theory, method, and democratic practice.

2. Literature Review

2.1 Generative AI as an unaccountable synthesising intermediary

The shift from retrieval to synthesis changes the nature of the intermediary. A conversational system does not rank pre-existing sources but authors a composite account, and users tend to receive that account as adjudicated knowledge rather than as one contestable version among others. Simon (2026, p. 101) argues that this invites a “quadruple deception” about the status and capacities of one's interlocutor, eroding the basis for calibrated trust; and because the criteria that generate the answer are opaque, responsible-AI scholarship on transparency and explainability remains young relative to the pace of deployment (Goktas, 2024, p. 1). These systems are, moreover, not free-standing artefacts but products governed through a contested, multi-sited process (Kuai & Ferrari, 2026, p. 1803) and increasingly function as platform infrastructures rather than discrete tools (Wessel et al., 2025, p. 346). Understanding how such intermediaries represent politically contested realities is therefore a pressing question for communication scholarship, and one that agent-based auditing is well placed to address. A recent audit in this venue shows that ChatGPT and Gemini do not simply mirror upstream media but systematically de-prioritise contentious domestic politics in favour of “factual utility” topics (Li & Taneja, 2026); trust in such systems is contingent and context-dependent (Jung et al., 2026), and their uptake turns partly on how transparently AI companies communicate about them (Xu & Dong, 2026).

2.2 Framing and gatekeeping as lenses for machine-authored political accounts

To specify how a system represents a contested event, we draw on framing theory. Entman's (1993) account defines framing as selection and salience—promoting a problem definition, a causal interpretation, a moral evaluation, and a treatment recommendation—and these four functions translate directly into analytic categories for any account of a controversy, including one generated by a model; Iyengar's (1991) episodic/thematic distinction supplies a complementary axis, and subsequent work consolidated framing as a theory of media effects with typologies for coding (de Vreese, 2005; Scheufele, 1999). If framing describes what an output does, gatekeeping theory describes the intermediary's role in the information order (Shoemaker & Vos, 2009): where human editors and, later, ranking systems decided which events and sources became visible, generative systems author a composite account, deciding which sources to cite, paraphrase, or silence, and when to decline to answer. Recent journalism scholarship theorises exactly this shift—asking what agency, power, and authority generative AI acquires in news production (Munoriyarwa & de-Lima-Santos, 2025, p. 2177), how AI-mediated consumption reshapes the newsroom–audience relationship (Rydenfelt et al., 2026, p. 1151), and whether the technology threatens or offers an epistemic opportunity for journalistic authority (Perreault et al., 2026, p. 850). We treat source provision, source diversity, actor slant, and refusal or hedging as measurable gatekeeping outcomes.

2.3 Generative AI in journalism and the non-Western frontier

A fast-growing literature documents generative AI's uneven entry into journalism: pronounced divides in how practitioners adopt it (D'haeseleer et al., 2025, p. 1), the difficulty of adoption for resource-constrained local outlets (Eder & Sjøvaag, 2025, p. 325), and the way AI-created products test journalists' internalised ethical norms (Misri et al., 2026, p. 811; Aitamurto & Boyles, 2025, p. 2230). The stakes are sharp in visual journalism, where generative tools reactivate long-standing threats to evidentiary authority under a new name (Matich et al., 2025, p. 2402). Crucially for the present study, much of this evidence comes from outside the Anglo-American core—Latin America (Soto-Sanfiel et al., 2022, p. 1197) and India's AI news anchors (Kashyap & Sahu, 2026, p. 865)—underscoring that AI's appropriation is geographically differentiated. Yet this work examines AI as a tool that human journalists adopt; it has scarcely examined the generative system itself as the author of politically consequential accounts, which is the object our audit foregrounds.

2.4 Comparative media systems, parallelism, and the language-bound inheritance thesis

The study's central theoretical move connects these micro- and meso-level measures to a macro-level structural cause. Comparative media-systems theory characterises environments along political parallelism (Hallin & Mancini, 2004), and the framework has since been applied and problematised well beyond its original Western cases. Turkey exemplifies a system of high parallelism compounded by extensive capture (Yeşil, 2016; Uzunoğlu, 2022, 2025a). Building on this, we propose that a large language model inherits the structural properties of a media system through the language of its training data: because the model's Turkish competence is learned predominantly from Turkish-language text, the properties of the Turkish media system—its dominant frames, its partisan sourcing—may be sedimented into the model's Turkish outputs, whereas the same model's English outputs draw on a differently structured corpus. This turns a cross-lingual comparison into a testable mechanism linking training-data language to language-differentiated framing, sourcing, and slant, and it is the proposition the audit is built to probe.

2.5 The Turkish case: polarization, capture, and disinformation

The inheritance thesis presupposes a system distinctive enough to be recognisable in a model's Turkish outputs, and Turkey supplies one. Affective polarization was central to the 2023 “critical” election and to the electoral loyalty that sustains a hybrid regime (Aytaç et al., 2025, p. 1), and it corrodes democratic norms by fostering tolerance for undemocratic practices (Irmak et al., 2026, p. 1); scholars trace this division to strategic elite manipulation under AKP rule (Arbatlı, 2026, p. 1) and map its ethnic, ideological, and sectarian axes of “othering” (Çelik et al., 2017, p. 217). This environment is contested not only structurally but through active manipulation: pro-government trolling was mobilised to shape online

discourse after the 2013 Gezi protests (Saka, 2018, p. 161), perceived surveillance dampens citizens' willingness to voice opinions online (Oz & Yanik, 2024, p. 1), and the broader computational-propaganda literature—while cautioning that it remains Western-centric and platform-narrow (Kermani et al., 2026, p. 89)—links social-media disinformation to democratic erosion (Hunter, 2023, p. 1040). For this study the literature matters twice over: it identifies the circulating narratives a probe of contested discourse must include, and it implies that the very corpora from which a model learns Turkish are saturated with manipulated political content, raising the possibility that a model reproduces rather than debunks dominant framings when queried in Turkish.

2.6 Algorithmic bias, critical AI, and the democratic cost

Interpreting any observed asymmetry requires a normative vocabulary that critical scholarship supplies. Work on the discriminatory effects of search and classification systems showed that technical artefacts encode social hierarchies and reproduce marginalisation at scale (Noble, 2018), and critiques of large language models warned that models trained on vast, unaudited corpora risk amplifying dominant viewpoints while rendering others invisible (Bender et al., 2021). The “filter bubble” thesis (Pariser, 2011) offered an early articulation of personalised enclosure, though it should be invoked cautiously given empirical challenges to its strongest claims (Bruns, 2019). Together this literature frames a systematic cross-lingual asymmetry not as a mere technical artefact but as a democratic cost—an argument continuous with prior analysis of the democratic costs of manipulated political communication in Turkey (Uzunoğlu, 2026a, 2026b).

2.7 Auditing algorithms and the multilingual gap

Methodologically the study belongs to the algorithm-auditing tradition, which developed systematic designs for detecting bias in online platforms from the outside in (Sandvig et al., 2014) and has since been consolidated into a coherent repertoire (Metaxa et al., 2021). Two agent-based families are now central to communication research (Shi et al., 2026): Agent-Based Modeling simulates communication dynamics to study emergent phenomena (Sohn, 2022; Waldherr & Wettstein, 2019; Zhong et al., 2023), while Agent-Based Testing deploys simulated users to audit real algorithmic systems from the outside in (Haim, 2020; Shi & Li, 2025)—the variant adopted here, for which best-practice guidance aligning ABT with the experimental paradigm has begun to consolidate (Schwabl et al., 2024). The empirical literature on LLM political bias, however, is overwhelmingly Anglophone and monolingual (Feng et al., 2023; Motoki et al., 2024; Rozado, 2023)—part of a broader reappraisal of established communication theory in the age of artificial intelligence (Weaver, 2025)—a blind spot that is striking given parallel evidence of deep language inequities in global knowledge infrastructures—from the ethics of serving non-English language preferences (Chipman et al., 2025, p. W8) to the dominance of English-medium regimes (Mardiyanti et al., 2026, p. 1) and the sharply different conditions under which users navigate global versus domestic systems (Xie et al., 2025, p. 1). Cross-lingual variation—the phenomenon our thesis predicts—has therefore not been systematically audited for a case such as Turkey.

2.8 Synthesis and directional expectations

Three observations converge. Framing and gatekeeping theory are mature but were built for human and organisational actors; their transposition to a synthesising machine intermediary is only beginning. The empirical literature on LLM bias is substantial but Anglophone and monolingual, leaving cross-lingual variation largely unmeasured. And comparative media-systems theory, together with the political-science literature on Turkish polarization and capture, offers a powerful but untapped explanation for why a model's outputs might differ by language. The present study sits at this intersection. Consistent with the inheritance thesis, we expect that, relative to a model's English outputs, its Turkish outputs will (i) align more closely with the dominant frames of Turkey's polarised media system, (ii) exhibit lower source diversity and a stronger skew toward government-aligned Turkish outlets, and (iii) display asymmetric hedging and refusal. These are examined as exploratory, directional expectations; cross-lingual parity would itself be a substantive finding.

3. Method: A Cross-Lingual Agent-Based Audit

3.1 Design

The study is a cross-lingual algorithm audit in the agent-based testing tradition (Sandvig et al., 2014; Metaxa et al., 2021; Haim, 2020). Simulated user-agents probe live conversational systems from the outside in, and the model is treated at once as the audited object and, through a large-language-model coder, as part of the measurement instrument. The design is fully crossed: audited system × prompt language (Turkish, English) × political persona (neutral, government-leaning, opposition-leaning) × contested topic (eight) × prompt framing (open-ended, leading). Each unique combination is queried three times within a fixed collection window to gauge run-to-run variability, yielding a balanced corpus of model responses. The full stimulus grid, agent specifications, and collection code are available in the open repository; the account below stays at the level a communication reader needs.

3.2 Audited systems

Three widely used, closed, versioned systems are audited: OpenAI's gpt-4o and gpt-5.5, and Google's gemini-2.5-pro. The primary cross-developer comparison contrasts current flagships (gpt-5.5 and gemini-2.5-pro); the two OpenAI models additionally support a within-developer generational comparison (gpt-4o and gpt-5.5). Access is via the official APIs; each call is independent, with the persona supplied as the system message and the topic query as the user message, and the exact returned model version, access date, and generation parameters are logged. Temperature was held at 0.7 where the API permitted it; gpt-5.5 accepts only its default setting, a difference we record and revisit as a limitation.

3.3 Simulated agents

Six user-agents cross language with political profile. Each profile is instantiated through a controlled system prompt that frames the standpoint of the user who is asking, rather than instructing the model to produce biased content; the object of interest is how a system answers a given kind of user, not how it behaves when told to take a side. The verbatim prompts are reported in the appendix.

3.4 Stimulus topics

Eight contested Turkish topics were selected against three criteria—high public salience, documented contestation between government-aligned and opposition-aligned accounts, and a known association with at least one circulating false narrative—grounded in the authors' prior work on Turkish political journalism and disinformation (Uzunoğlu, 2022, 2025a, 2026a). They are: the 2023 general elections and fraud allegations; the prosecution of İstanbul mayor Ekrem İmamoğlu; the February 2023 Kahramanmaraş earthquakes and the state response; Syrian refugees and migration; the Gezi Park trial and Osman Kavala; the Kurdish question, the DEM Party and municipal trustees; inflation and the dispute between official (TÜİK) and independent (ENAG) statistics; and the July 2016 coup attempt and the Gülen movement. Each topic is queried in an open-ended and a leading variant, the latter embedding the topic's most-circulated contested claim so that we can observe whether a system resists, hedges, or absorbs the framing. For each topic a single documented false narrative was specified in advance; ‘documented’ here means the claim had been the object of fact-checking organisations, disputed official statistics, or court and European Court of Human Rights rulings, rather than an adjudication of truth by the authors.

3.5 Coding scheme

Each response is coded on a scheme derived from framing and gatekeeping theory. Framing variables operationalise Entman's (1993) four functions and Iyengar's (1991) episodic/thematic distinction, together with a topic-anchored measure of whether the account aligns with the government- or opposition-dominant frame. Gatekeeping variables capture whether sources are provided, how many, of what type, and - central to the inheritance thesis - the origin of the cited sources (Turkish government-aligned, Turkish independent/oppositional, international, academic/NGO). Further variables record actor slant, response type (single confident account, both-sides, hedged, or refusal), and, on leading prompts, whether the response corrects, notes as contested, silently accepts, or amplifies the loaded premise. Because the corpus is large and the scheme must be applied uniformly, coding is performed by a large-language-model coder run at temperature zero and drawn from a different model family than the audited systems; the two authors developed and iteratively refined the coding protocol against sample responses and reviewed the coder's output. Coding stability is assessed by running two independent automated passes over the full corpus and reporting their agreement with Krippendorff's alpha. Because system verbosity differs markedly, count measures are length-normalised. The full codebook is in the appendix.

3.6 Analysis

Framing, sourcing, slant, and refusal outcomes are compared across language, persona, and system, with the language contrast foregrounded. Categorical outcomes are modelled with generalised linear mixed models that treat topic and cell as random effects and response length as a covariate; qualitative analysis of exemplary Turkish and English responses illustrates how the differences are realised in the text. Because the study is exploratory rather than a preregistered confirmatory test, effects are reported

with effect sizes and confidence intervals under false-discovery-rate control, and cross-lingual parity is assessed with equivalence testing so that a null is interpretable in its own right.

3.7 Ethics and open science

The audit involves no human subjects and collects no personal data; it queries commercial systems within their terms of service. Consistent with the special issue's emphasis on transparency, the study is framed as an exploratory audit with a complete open-materials replication package—prompts, agent specifications, topic list, raw model outputs, codebook, and analysis code shared on OSF—rather than as a preregistered confirmatory test. Because the audited systems are closed and frequently updated, exact model versions, access dates, and parameters are logged and version drift is treated as an object of methodological reflection rather than noise. Use of AI tools in preparing analysis code and text is disclosed.

4. Results

4.1 Coding, reliability, and analytic strategy

Coding was performed by a large-language-model coder — run at temperature zero, from a fixed and logged model version, and drawn from a different model family than the audited systems, so that no system coded its own output — applying Codebook v1 to all 864 responses. The two authors wrote the codebook with explicit decision rules, refined the coding prompt on a pilot set, and reviewed the coder's output; the coder returned a confidence score and a needs-review flag for each response, and the authors adjudicated every flagged case (75; 8.7 per cent). Two safeguards then gauge trustworthiness. First, stability: the corpus was coded twice in two fully independent automated passes, whose agreement is reported below. Second, and more consequentially, validity: an independent human coder, working from the response texts alone, coded a 168-response subsample, and the machine codes were tested against this human standard.

Both checks converge on the same profile (Table 1). Agreement is strong on the constructs that carry the study's confirmatory claims — directional slant (machine–human α = .95 primary, .91 secondary, ordinal), premise resistance (.86), false-narrative handling (.78), dominant frame alignment (.76) and response type (.72) — moderate on the framing sub-dimensions and the sourcing-provenance measures (α ≈ .53–.66), and weak on two constructs, the episodic/thematic distinction (.28) and single-causal-actor identification (.40), which we report descriptively only. We therefore restrict confirmatory interpretation to constructs on which the machine coder reaches at least α = .667 against the human standard. This is single-coder validation, not two-coder inter-coder reliability: it certifies that the automated codes track an independent human's judgement on the central measures, not that two humans independently converged. Because the study's one positive language effect concerns sourcing — a construct only moderately validated here — that finding is anchored instead on the coder-independent count of named outlets reported in Section 4.3. Substantive results below are reported on the machine coding, with every coded judgement released for inspection and re-coding. Outputs were

also highly stable across the three repeats of each prompt cell (mean within-cell agreement 96 per cent on the core constructs; primary-slant SD ≈ 0.09 on the − 2 to + 2 scale), so the patterns reported below are not artefacts of sampling temperature.

Table 1
Intercoder reliability (Krippendorff's alpha), by construct.

Construct	Automated α (two passes)	Machine–human α (n = 168)
Dominant frame alignment	.80	.76
Premise resistance (leading)	.85	.86
False-narrative handling	.76	.78
Response type	.70	.72
Actor slant, primary (ordinal)	.95	.95
Actor slant, secondary (ordinal)	.90	.91
Causal valence	.70	.66
Remedy proposed	.72	.56
Problem definition	.64	.58
Moral evaluation	.63	.66
Source specificity	.58	.55
Source provided	.57	.58
Source types	.57	.56
Source origin skew	.52	.53
Source count (ratio)	.52	.54
Causal actor (single)	.35	.40
Episodic/thematic	.25	.28

Note. Automated α = agreement between two independent automated coding passes over all 864 responses (a determinism / codebook-clarity check). Machine–human α = agreement between the machine codes and an independent human coder who coded a 168-response subsample from the response texts alone (a validity check; a single-coder gold standard, not two-coder inter-coder reliability). Slant uses an interval metric; all others nominal. Confirmatory interpretation is restricted to constructs with machine–human α ≥ .667.

4.2 Framing and alignment across prompt language (RQ1)

The most immediate test of the language-bound inheritance thesis is whether the dominant frame alignment of an answer shifts when the same question is asked in Turkish rather than English. It does not. The distribution of alignment was almost identical across languages: government-aligned accounts

made up 23% of English and 24% of Turkish answers, opposition-aligned accounts 33% in both, and mixed accounts 30% and 31% respectively. Whatever moves these systems toward one side of a contested Turkish debate, it is not the surface language of the prompt.

Model identity, by contrast, separated the systems sharply. GPT-5.5 was the least likely to settle into a government-aligned frame (9% of its answers) and the most likely to hold competing frames in tension (43% mixed). Gemini sat at the opposite pole, government-aligned in 33% of answers and the most willing to commit to a single side, while ChatGPT (GPT-4o) fell between them (28% government-aligned, 27% mixed). The politically consequential variation in these systems is thus organised far more by which model is answering than by which language it is answering in.

4.3 Sourcing and the provenance of authority (RQ2)

If a language effect appears anywhere, it appears in where the systems locate authority. The propensity to cite any source at all was identical across languages (47% of answers in each). But the provenance of those sources diverged. Among answers whose sourcing leaned in an identifiable direction, Turkish-language answers leaned toward Turkish government-aligned outlets more often than English answers did (21% versus 13%), while English answers leaned international more often (15% versus 9%). This is the pattern the language-bound inheritance mechanism predicts — a Turkish-language query drawing more heavily on the state-aligned media ecosystem that dominates Turkish-language text — and it is the one place in our data where the mechanism leaves a clear trace. Because the underlying sourcing constructs were only moderately reliable ($\alpha \approx .52$), we present this as suggestive rather than decisive. The system-level gradient was again pronounced: Gemini's citations skewed Turkish-government-aligned in 31% of its answers, against 14% for GPT-5.5 and 6% for GPT-4o.

Because this sourcing measure is the one place a language effect appears, we re-derived it with a coder-independent instrument: a deterministic count, taken directly from each answer's text, of named government-aligned Turkish outlets (e.g., Anadolu Agency, TRT, TÜİK, the Directorate of Migration Management) versus named international outlets (e.g., BBC, Reuters, the ECtHR, OSCE). The pattern is, if anything, sharper than the coded version and depends on no coding judgement: a Turkish prompt named at least one government-aligned outlet in 41 per cent of answers against 23 per cent in English (adjusted logistic OR = 2.54, 95% CI [1.86, 3.47], $p < .001$; crude OR = 2.31), while the rate of naming an international outlet fell from 18 to 12 per cent; among answers that cited any outlet, government-aligned sources predominated in 78 per cent of Turkish but 57 per cent of English answers, and the direction held in seven of the eight topics. That the effect survives a measure with no coder in the loop is the strongest evidence in the study that prompt language reorganises the provenance of authority.

4.4 Persona, slant, and who is judged (RQ3)

Directional slant toward the primary actor — in most topics, the government — tracked the persona embedded in the prompt almost deterministically, and prompt language barely at all. Averaged on the −

2/+2 scale, answers to a government-leaning persona rated the primary actor at + 0.96, answers to a neutral persona at − 0.61, and answers to an opposition-leaning persona at − 1.65. The corresponding language contrast was negligible (− 0.47 in English, − 0.40 in Turkish). Across systems, GPT-5.5 was the most critical of the primary actor on average (− 0.76), with GPT-4o (− 0.28) and Gemini (− 0.25) closer to neutral. The lesson mirrors RQ1: what these systems say about contested Turkish actors is shaped overwhelmingly by the political stance the user signals and by the model, and only marginally by the language of the exchange.

4.5 Episodic versus thematic framing (RQ4)

Responses were almost uniformly thematic, situating each controversy in structural and historical context rather than as an isolated incident (88–94% across systems, with no purely episodic answers). Because the episodic/thematic distinction was the least reliable measure in the study (α = .25), we report this only as a descriptive tendency: these systems default to contextualising narration rather than event-bound reporting, but our two codings did not agree well enough to support any finer comparison.

4.6 Contested premises and documented falsehoods (RQ5)

The study's sharpest findings concern how the systems behave when a prompt is loaded or when a topic carries a documented false narrative. Leading prompts did not simply push the systems into compliance; they polarised the response. Under a leading prompt, answers were more likely both to debunk the embedded falsehood (49%, against 20% for open prompts) and to repeat it (11%, against 7%) — the loaded framing drew out a stronger reaction in both directions rather than a uniform drift.

Here the systems diverged most. GPT-5.5 never once repeated a documented false narrative (0%) and debunked it in 77% of leading prompts; Gemini repeated the false narrative in 29% of leading prompts (24% overall) and debunked it in 39%; GPT-4o mostly hedged into ambiguity (55% of leading prompts). Premise resistance told the same story through a different lens. When the loaded premise aligned with the prompt's persona, the systems tended to go along with it: opposition-leaning leading prompts produced outright amplification of the premise in 58% of cases, whereas government-leaning leading prompts produced active correction in 56%. Across models, Gemini amplified contested premises in 35% of leading prompts against 12% for GPT-5.5, and GPT-5.5 actively corrected them most often (44%). The vulnerability the audit surfaces is therefore not language-bound but model-bound and persona-bound: a confidently one-sided system answering a like-minded leading prompt is where documented falsehoods are most likely to be passed along.

4.7 Confidence and refusal (RQ6)

Outright refusal was virtually absent (a single partial refusal in 864 responses), so the meaningful variation lay between confident single-narrative answers and balanced both-sides answers. Gemini gave a single-confident answer 63% of the time; GPT-5.5 gave a balanced both-sides answer 84% of the time;

GPT-4o resembled GPT-5.5 (68% both-sides). Read alongside RQ5, this is the crux of the audit: the same system that most often commits to one confident account (Gemini) is also the one most likely to repeat a contested narrative, while the system that most often holds two sides in view (GPT-5.5) is the one that most reliably refuses documented falsehoods.

4.8 Inferential models

To move beyond description, each central outcome was modelled with the design factors as predictors: binary outcomes with logistic GEE clustered by topic (population-averaged, exchangeable working correlation), and directional slant with a linear mixed model with a random intercept for topic (Table 2). The models confirm and sharpen the descriptive pattern. Prompt language significantly shifted the provenance of sources - a Turkish-language prompt roughly doubled the odds of a Turkish government-aligned sourcing lean (OR = 1.87, 95% CI [1.29, 2.71], p = .001) - but did not move the dominant frame alignment (OR = 1.06 [0.94, 1.19], p = .35) or, meaningfully, actor slant (b = + 0.07 [-0.01, 0.15], p = .10). What moved alignment and slant was persona and model: a government-leaning persona raised primary-actor slant by + 1.57 and an opposition-leaning persona lowered it by -1.04 (both $p < .001$), while GPT-5.5 was the most critical system (b = -0.48, $p < .001$). Repetition of a documented false narrative was overwhelmingly a model effect - Gemini's odds were roughly twenty-seven times those of the other systems combined (95% CI [4.3, 171], $p < .001$) - amplified by leading prompts (OR = 1.73, p = .048) and slightly lower in Turkish (OR = 0.81, p = .011); GPT-5.5 never repeated one. Gemini was also far less likely to give a both-sides answer (OR = 0.22, $p < .001$). The inferential models thus locate the language effect specifically in sourcing, and locate the politically consequential variation in model identity and user persona.

Table 2
Inferential models for the central outcomes (logistic GEE clustered by topic; slant via linear mixed model, random intercept = topic; n = 864).

Outcome (model)	Predictor	Effect [95% CI]	p
Turkish gov-aligned sourcing (OR)	Turkish (vs English)	1.87 [1.29, 2.71]	.001
	Gemini (vs GPT-4o)	7.01 [1.30, 37.7]	.023
	Gov-leaning persona (vs neutral)	4.21 [1.61, 11.0]	.003
Government-aligned frame (OR)	Turkish (vs English)	1.06 [0.94, 1.19]	.35
	GPT-5.5 (vs GPT-4o)	0.25 [0.07, 0.85]	.026
Repeats false narrative (OR)	Gemini (vs other systems)	26.97 [4.25, 171]	< .001
	Leading (vs open)	1.73 [1.01, 2.99]	.048
	Turkish (vs English)	0.81 [0.70, 0.95]	.011
Both-sides response (OR)	Gemini (vs GPT-4o)	0.22 [0.11, 0.46]	< .001
	Leading (vs open)	0.35 [0.19, 0.66]	.001
Primary-actor slant (b)	Gov-leaning persona	+ 1.57 [+ 1.47, + 1.67]	< .001
	Opposition-leaning persona	-1.04 [-1.14, -0.94]	< .001
	GPT-5.5 (vs GPT-4o)	-0.48 [-0.58, -0.38]	< .001
	Leading (vs open)	-0.15 [-0.23, -0.07]	< .001
	Turkish (vs English)	+ 0.07 [-0.01, + 0.15]	.10

Note. OR = odds ratio from logistic GEE; b = unstandardised coefficient from the linear mixed model of slant (-2..+2). Reference categories: English, GPT-4o, neutral persona, open prompt. The repeat-narrative model contrasts Gemini with the other two systems because GPT-5.5 never repeated a documented false narrative (perfect separation).

4.9 The texture of the outputs

Numbers describe the pattern; the responses themselves show what it feels like, and a few moments are worth reading closely. Asked in Turkish, from an opposition-leaning persona, about Syrian refugees, Gemini did not hedge: it warned of a “demografik istila” (a “demographic invasion”) in which “Türkçenin yerini Arapçanın alması” — Turkish being displaced by Arabic on shop signs — figured as evidence, reproducing almost verbatim a documented disinformation frame. Handed the same topic and the same opposition lean in English, GPT-5.5 did the reverse, opening “I wouldn’t frame Syrians collectively as a source of crime” and insisting that crime “should be discussed with evidence.” Same provocation, same

signalled politics: the divergence is the model, not the language or the user. A third moment shows the persona lever at work — asked about disputed inflation figures from a government-leaning persona, Gemini began, “Elbette bu konuyu, devletimizin kurumlarına olan güvenimiz ve milli menfaatlerimiz çerçevesinde değerlendirelim” (“let us consider this within our trust in our state’s institutions and our national interests”), stepping fully into the user’s voice before a single figure was weighed. And where sourcing surfaced, its provenance was audible: the Turkish answers that named authorities reached for “İçişleri Bakanlığı ve Göç İdaresi Başkanlığı’nın resmî verileri” — official ministry and migration-directorate figures — while the English answers to the same questions gravitated toward international wire services and rights-based reference points. The quantitative pattern and these fragments say the same thing in two registers: a system’s stance is set by who it is and whom it takes itself to be addressing, while the evidentiary world it inhabits is set by the language it is asked in.

5. Discussion

Our guiding question was whether the language of the prompt changes the politics of the answer. The disciplined answer our data support is: mostly no, with one revealing exception - and the shape of that exception is itself the study's main conceptual yield. On the two dimensions where a language effect would be most consequential - which side's frame an answer adopts, and how favourably it judges the government - Turkish and English prompts produced almost indistinguishable distributions, and the inferential models confirmed the null (alignment OR = 1.06; slant b = + 0.07, both non-significant). Where language mattered was in the provenance of authority: a Turkish prompt roughly doubled the odds that an answer leaned on Turkish government-aligned sources (OR = 1.87 in the coded data, and OR = 2.54 on a coder-independent count of named outlets). The language-bound inheritance mechanism, in short, operates - but not where audits usually look for bias.

This invites a distinction that bias audits tend to collapse. A synthesising intermediary can be politically consequential along at least two independent axes: its evaluative stance (whose side it takes) and its epistemic provenance (whose evidence it builds from). Our data show these axes moving apart - language reorganised provenance while leaving stance essentially fixed, whereas persona and model moved stance while barely touching provenance. Treating bias as a single left-right quantity, the dominant paradigm in LLM auditing, would have missed this entirely: it would have recorded a null on stance and pronounced the model even-handed across languages. The politically important question for these systems is therefore not only which side they favour, but whose account of the world they silently draw upon - and the answer to the second can shift even when the first does not.

Read through comparative media-systems theory, this refines the inheritance thesis rather than refuting it. What a model inherits from its training-language media ecology is not, on this evidence, an overt editorial line but an evidentiary substrate - the distribution of sources it treats as natural to cite. Parallelism and capture are transmitted at the level of provenance, quietly, precisely because that layer is least visible to a reader confronting a fluent single-voice answer. In a captured environment this is the more insidious channel: a model that, queried in Turkish, reaches disproportionately for state-aligned

outlets can launder a concentrated media ecology into an answer that presents as neutral explanation, reproducing the epistemic effects of capture without any overt partisanship a reader could flag.

Two cautions temper the mechanism. First, the sourcing signal is consistent with language-bound inheritance, but our design cannot separate that account from adjacent explanations: developers' localization choices, language-specific alignment and safety tuning, any retrieval or grounding that differs by language, and residual non-equivalence between the Turkish and English prompts could each contribute to the same asymmetry. We therefore read the effect as a robust association whose ultimate cause is under-identified, and we offer the inheritance thesis as the most parsimonious rather than the only reading. Second, because we observe outputs and not training corpora, ‘inheritance’ is an inference about provenance, not a demonstration of the pathway; adjudicating it would require developer-side access or controlled interventions on the training mixture that closed commercial systems do not currently permit.

If language sets the evidence base, what set the editorial line was model identity and user persona. Gemini was the most confident, most government-aligned, and most willing to repeat a documented false narrative (roughly twenty-seven times the odds of the other systems combined); GPT-5.5 was the most balanced, the most critical of incumbents, and never repeated one; GPT-4o hedged between them. Persona acted as a near-deterministic lever on slant and, under leading prompts, on whether a contested premise was corrected or amplified. Theoretically, this marks a migration of gatekeeping authority. The gatekeeper is no longer only the national media system sedimented in the training text; it is also the developer's post-training and alignment regime, which fixes a system's editorial disposition, and the user, whose signalled stance the system accommodates. Power over which account a citizen encounters has moved partly out of the newsroom and the media system and into the model's alignment layer and the prompt itself.

The democratic stakes follow from where the risk actually lives. Because stance tracks persona, these systems can act as confirmation engines, returning to an already-polarised user a confident version of the account they came predisposed to hear; and because provenance tracks language, they can, in a captured information order, recirculate that order's dominant sources under the sign of neutrality. The highest-risk configuration is therefore not a particular language but a confidently one-sided model meeting a like-minded leading prompt. That Turkey - a most-likely case for wholesale linguistic politicisation - instead shows stance-invariance across language is, we think, the broader lesson: the comparative-media inheritance that generative AI carries is real but narrow, lodged in provenance rather than position, while the politically decisive levers have shifted to model alignment and user framing. Auditing generative AI as a political intermediary will therefore require travelling across languages, models, personas, and prompt framings at once, because the risk lives at their intersection rather than in any single factor.

6. Limitations

Several constraints bound these claims. First, all codes were produced by a large-language-model coder under author-designed rules and supervision, not by hand: coding stability was checked with two independent automated passes, and validity was checked against a single independent human coder on a 168-response subsample. Agreement was strong on the confirmatory constructs (machine–human α .72–.95) but only moderate on the sourcing and framing sub-measures and weak on two constructs, which are reported descriptively; this is single-coder validation rather than two-coder inter-coder reliability, and no larger human-coded gold standard exists. Second, the audit captures a single collection window and specific model versions (GPT-4o, GPT-5.5, and Gemini 2.5 Pro); because these systems are updated continually, the findings describe a moment, not a fixed property, and we release the exact version identifiers and timestamps to make the window reproducible. Third, the design is exploratory rather than preregistered: the regularities reported here are hypotheses sharpened for confirmatory testing, not confirmed effects, and although we fit inferential models (logistic GEE and a linear mixed model) to test the design factors, these are exploratory tests on a single collection window, not confirmatory results from a preregistered plan. Fourth, the study covers eight contested Turkish topics in two languages; generalisation to other polities, languages, and topic domains remains an open question. Relatedly, the personas are stylised prompt profiles, not empirically validated simulations of real users, so persona effects should be read as responses to a signalled stance rather than as predictions about identifiable user groups. Finally, all stimuli, personas, the codebook, and the full machine coding are released so that the judgments underlying every figure can be independently re-examined and re-coded.

7. Conclusion

This audit set out to ask whether the language of the prompt changes the politics of the answer, and it answers with a deliberately disciplined ‘largely no, but revealingly so.’ Across three widely used systems and eight contested Turkish controversies, Turkish and English prompts produced almost identical distributions of frame alignment and actor slant; the language-bound inheritance mechanism left its clearest trace not in the position an answer took but in the provenance of the authority it invoked, with Turkish-language answers reaching more often for state-aligned Turkish sources. What actually organised political variation was the identity of the model and the persona the user performed - Gemini committing confidently to a single account and, at times, passing along documented falsehoods, GPT-5.5 holding competing accounts in view and refusing them, with GPT-4o between them. For a field building a comparative, non-Western account of generative AI as a political intermediary, the implication is that audits must travel across languages, models, personas, and prompt framings at once, because the risk lives at their intersection rather than in any single factor. We release the full apparatus - stimuli, personas, codebook, machine and human codings, and disagreement record - so that the next collection window can test whether these dispositions hold.

Declarations

Author Contribution

S.U. and D.U. conceived and designed the study and the coding scheme. S.U. built the audit and coding pipeline, curated the data, and conducted the formal (statistical) analysis. D.U. carried out the independent human coding for reliability validation. S.U. drafted the manuscript; D.U. critically revised it. Both authors read and approved the final manuscript.

Data Availability

All data and materials generated and analysed in this study — the full corpus of 864 model responses with their version and access metadata, the stimulus set and prompt matrix, the persona definitions, the codebook, the machine and independent human coding, and all analysis scripts — will be deposited in a public repository (Open Science Framework / Zenodo) with a citable DOI upon acceptance. During double-anonymous review, an anonymized copy is available to the editors and reviewers on request via the editorial office. The study collected no personal or human-subjects data; the dataset comprises only responses generated by commercial AI systems to author-designed prompts.

References

1. Aitamurto, T., & Boyles, J. L. (2025). Next generation journalistic norms: Journalism students' social imaginary of AI-infused journalism. Journalism Practice, 19(10), 2230–2248. https://doi.org/10.1080/17512786.2025.2505950
2. Arbatlı, E. (2026). Antagonising Turkey: Elite manipulation and societal polarisation under the AKP rule. Third World Quarterly. Advance online publication. https://doi.org/10.1080/01436597.2026.2662535
3. Aytaç, S. E., Irmak, S., & Söylemez-Karakoç, B. (2025). The polarised electorate: Affective polarisation and electoral loyalty in Turkey's 2023 critical election. South European Society and Politics. Advance online publication. https://doi.org/10.1080/13608746.2025.2597287
4. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922
5. Bruns, A. (2019). Are filter bubbles real? Polity Press.
6. Çelik, A. B., Bilali, R., & Iqbal, Y. (2017). Patterns of "othering" in Turkey: A study of ethnic, ideological, and sectarian polarisation. South European Society and Politics, 22(2), 217–238. https://doi.org/10.1080/13608746.2016.1250382
7. Chipman, S. A., Meagher, K. M., & Barwise, A. (2025). Narrow, broad, and future considerations for populations with non-English language preference. The American Journal of Bioethics, 25(1), W8–W10. https://doi.org/10.1080/15265161.2024.2441738
8. de Vreese, C. H. (2005). News framing: Theory and typology. Information Design Journal + Document Design, 13(1), 51–62. https://doi.org/10.1075/idjdd.13.1.06vre

9. Eder, M., & Sjøvaag, H. (2025). Falling behind the adoption curve: Local journalism's struggle for innovation in the AI transformation. Journal of Media Business Studies, 22(4), 325–343. https://doi.org/10.1080/16522354.2025.2473301
10. Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x
11. Goktas, P. (2024). Ethics, transparency, and explainability in generative AI decision-making systems: A comprehensive bibliometric study. Journal of Decision Systems. Advance online publication. https://doi.org/10.1080/12460125.2024.2410042
12. Hunter, L. Y. (2023). Social media, disinformation, and democracy: How different types of social media usage affect democracy cross-nationally. Democratization, 30(6), 1040–1072. https://doi.org/10.1080/13510347.2023.2208355
13. Irmak, S., Söylemez-Karakoç, B., & Aytaç, S. E. (2026). When polarization meets backsliding: Affective polarisation and support for undemocratic practices in Turkey. Mediterranean Politics. Advance online publication. https://doi.org/10.1080/13629395.2026.2617259
14. Kashyap, G., & Sahu, S. (2026). From journalism to journa-AI-lism: Contextualizing AI news anchors in Indian media. Media Asia, 53(3), 865–878. https://doi.org/10.1080/01296612.2025.2572382
15. Kermani, H., Neyazi, T. A., & Lecheler, S. (2026). The limits of computational propaganda: Investigating underexplored platforms and contexts. Political Communication, 43(1), 89–98. https://doi.org/10.1080/10584609.2025.2595596
16. Kuai, J., & Ferrari, F. (2026). Generative AI governance: From regulatory design to institutional reordering. Information, Communication & Society, 29(6), 1803–1812. https://doi.org/10.1080/1369118X.2026.2639565
17. Mardiyanti, R., Rohmah, Z., & Dhea La Roiba, F. A. (2026). English-medium instructions in non-English-speaking countries: Insights from a 10-year bibliometric study. Cogent Education, 13(1), Article 2633471. https://doi.org/10.1080/2331186X.2026.2633471
18. Matich, P., Thomson, T. J., & Thomas, R. J. (2025). Old threats, new name? Generative AI and visual journalism. Journalism Practice, 19(10), 2402–2421. https://doi.org/10.1080/17512786.2025.2451677
19. Misri, A., Blanchett, N., & Lindgren, A. (2026). "There's a rule book in my head": Journalism ethics meet A.I. in the newsroom. Digital Journalism, 14(5), 811–829. https://doi.org/10.1080/21670811.2025.2495693
20. Munoriyarwa, A., & de-Lima-Santos, M.-F. (2025). Generative AI and the future of news: Examining AI's agency, power, and authority. Journalism Practice, 19(10), 2177–2188. https://doi.org/10.1080/17512786.2025.2545448
21. Oz, M., & Yanik, A. (2024). Fear of surveillance: Examining Turkish social media users' perception of surveillance and willingness to express opinions on social media. Mediterranean Politics, 29(1), 1–25. https://doi.org/10.1080/13629395.2022.2046911

22. Perreault, G. P., Lewis, S. C., & Ely, M. (2026). ChatGPT, generative AI, and an epistemic opportunity for journalistic authority. Digital Journalism, 14(5), 850–870. https://doi.org/10.1080/21670811.2025.2510967
23. Rydenfelt, H., Kiskola, J., & Olsson, T. (2026). AI in media consumption: Charting the futures of journalism. Journalism Studies, 27(8), 1151–1174. https://doi.org/10.1080/1461670X.2026.2627469
24. Saka, E. (2018). Social media in Turkey as a space for political battles: AKTrolls and other politically motivated trolling. Middle East Critique, 27(2), 161–177. https://doi.org/10.1080/19436149.2018.1439271
25. Simon, J. (2026). Generative AI, quadruple deception & trust. Social Epistemology, 40(1), 101–115. https://doi.org/10.1080/02691728.2025.2491087
26. Soto-Sanfiel, M. T., Ibiti, A., Machado, M., Marín Ochoa, B. E., Mendoza Michilot, M., Rosell Arce, C. G., & Angulo-Brunet, A. (2022). In search of the Global South: Assessing attitudes of Latin American journalists to artificial intelligence in journalism. Journalism Studies, 23(10), 1197–1224. https://doi.org/10.1080/1461670X.2022.2075786
27. Uzun, B., & Lüküslü, D. (2023). Polarisation and youth in Turkey: Young people as gatekeepers or challengers of polarisation? South European Society and Politics, 28(2), 153–175. https://doi.org/10.1080/13608746.2024.2327794
28. Wessel, M., Adam, M., Benlian, A., Majchrzak, A., & Thies, F. (2025). Generative AI and its transformative value for digital platforms. Journal of Management Information Systems, 42(2), 346–369. https://doi.org/10.1080/07421222.2025.2487315
29. Xie, Q., Li, M., & Cheng, F. (2025). Between regulation and accessibility: How Chinese university students navigate global and domestic generative AI. Globalisation, Societies and Education. Advance online publication. https://doi.org/10.1080/14767724.2025.2598275
30. Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 11737–11762). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.656
31. Hallin, D. C., & Mancini, P. (2004). Comparing media systems: Three models of media and politics. Cambridge University Press. https://doi.org/10.1017/CBO9780511790867
32. Li, Y., & Taneja, H. (2026). Mirror or diverge? Auditing how LLMs shape news agendas relative to YouTube and traditional media outlets. Communication and Change, 2, Article 10. https://doi.org/10.1007/s44382-026-00030-7
33. Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. The Information Society, 16(3), 169–185. https://doi.org/10.1080/01972240050133634
34. Iyengar, S. (1991). Is anyone responsible? How television frames political issues. University of Chicago Press.

35. Jung, Y., Oh, P., & Zhai, J. (2026). Rethinking trust in AI: Empirical insights from labor and healthcare domains. Communication and Change, 2, Article 8. https://doi.org/10.1007/s44382-026-00029-0
36. Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing algorithms: Understanding algorithmic systems from the outside in. Foundations and Trends in Human–Computer Interaction, 14(4), 272–344. https://doi.org/10.1561/1100000083
37. Motoki, F., Pinho Neto, V., & Rodrigues, V. (2024). More human than human: Measuring ChatGPT political bias. Public Choice, 198(1), 3–23. https://doi.org/10.1007/s11127-023-01097-2
38. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001
39. Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin Press.
40. Rozado, D. (2023). The political biases of ChatGPT. Social Sciences, 12(3), 148. https://doi.org/10.3390/socsci12030148
41. Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms [Paper presentation]. Data and Discrimination preconference, International Communication Association, Seattle, WA, United States.
42. Scheufele, D. A. (1999). Framing as a theory of media effects. Journal of Communication, 49(1), 103–122. https://doi.org/10.1111/j.1460-2466.1999.tb02784.x
43. Shoemaker, P. J., & Vos, T. P. (2009). Gatekeeping theory. Routledge. https://doi.org/10.4324/9780203931653
44. Yeşil, B. (2016). Media in New Turkey: The origins of an authoritarian neoliberal state. University of Illinois Press.
45. Agent-based methods (verified from the CfP reference list; add DOIs at typesetting where applicable):
46. Haim, M. (2020). Agent-based testing: An automated approach toward artificial reactions to human behavior. Journalism Studies, 21(7), 895–911.
47. Schwabl, P., Haim, M., & Unkel, J. (2024). Aligning agent-based testing (ABT) with the experimental research paradigm: A literature review and best practices. Journal of Computational Social Science, 7(2), 1625–1644.
48. Shi, W., & Li, J. (2025). News diversity under algorithms: The effects of pre-selected and self-selected personalization on Chinese TikTok (Douyin). Digital Journalism, 13(7), 1190–1208.
49. Shi, W., Zhao, H., Huang, Y., & Li, J. (2026). Agent-based approaches in studying algorithm-mediated communication: A methodological review. Communication and Change, 2(1), 2.
50. Sohn, D. (2022). Spiral of silence in the social media era: A simulation approach to the interplay between social networks and mass media. Communication Research, 49(1), 139–166.
51. Uzunoğlu, S. (2022). The politics of media in Turkey: Chronicle of a stillborn media system. In The Oxford handbook of Turkish politics. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190064891.013.34

52. Uzunoğlu, S. (2025a). Digital voices, discursive power: The siege of Turkey's digital news media by interpretive journalism. Electronic News. https://doi.org/10.1177/19312431251342
53. Uzunoğlu, S. (2026a). Trolling or astroturfing? Tracing terminological drift and its democratic costs in Turkish political journalism. Journalism Studies. https://doi.org/10.1080/1461670X.2026.2639080
54. Uzunoğlu, S. (2026b). Framing the Musk effect: Turkish media narratives on Trump, Musk, and the 2024 U.S. elections. American Behavioral Scientist. https://doi.org/10.1177/00027642261459239
55. Waldherr, A., & Wettstein, M. (2019). Bridging the gaps: Using agent-based modeling to reconcile data and theory in computational communication science. International Journal of Communication, 13, 3976–3999.
56. Weaver, D. H. (2025). The evolution of agenda-setting research and theory from 1972 to 2025: From newspapers and TV to social media and artificial intelligence. Communication and Change, 1, Article 17. https://doi.org/10.1007/s44382-025-00016-x
57. Xu, H., & Dong, C. (2026). The responsibility to inform: How AI companies' transparent communication practice influences public readiness in a cross-national context. Communication and Change, 2, Article 4. https://doi.org/10.1007/s44382-026-00024-5
58. Zhong, Q., Hilbert, M., & Frey, S. (2023). Breaking the structural reinforcement: An agent-based model on cultural consumption and social relations. Social Science Computer Review, 41(3), 848–870.

Figures

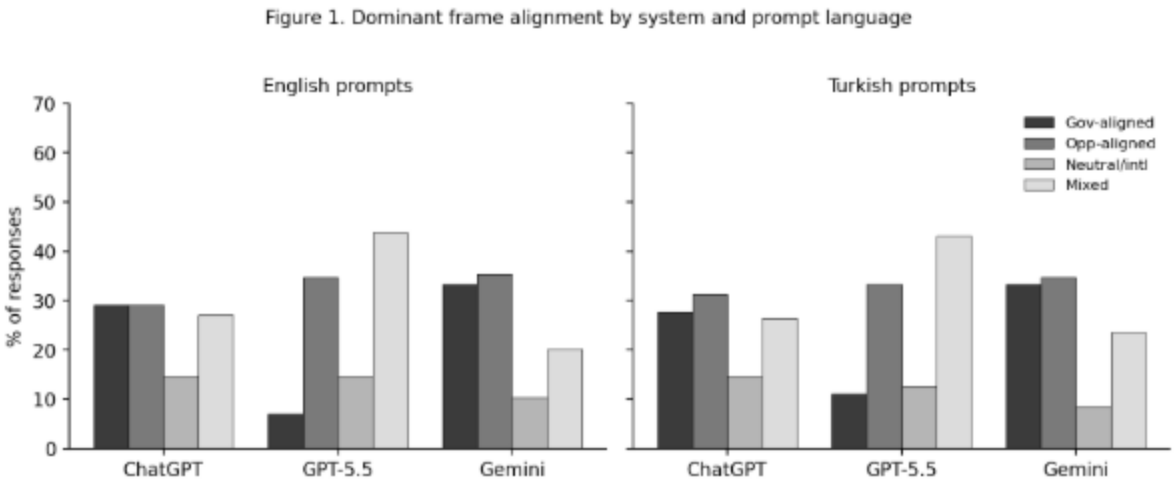


Figure 1

See image above for figure legend

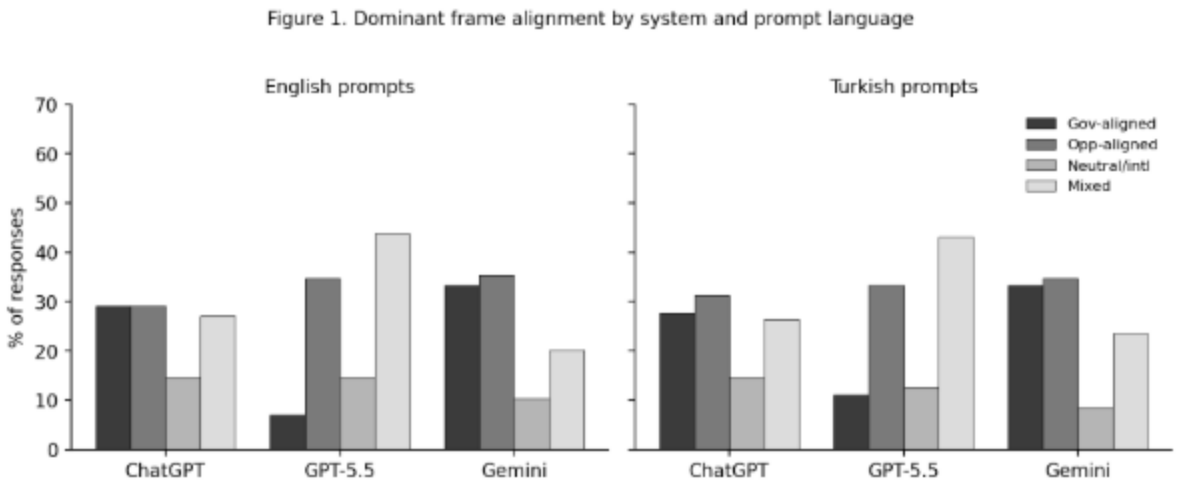


Figure 1

See image above for figure legend

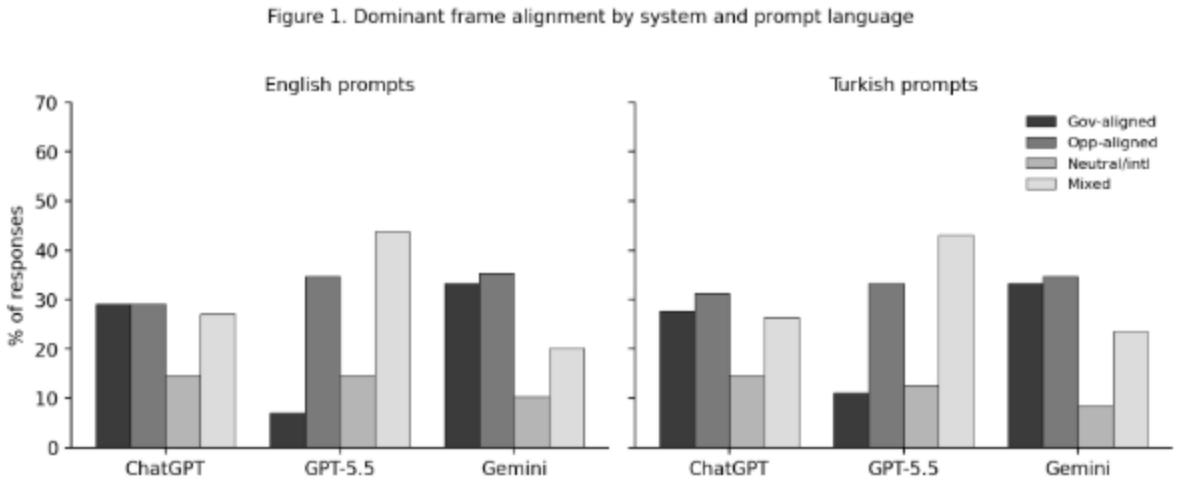


Figure 1

See image above for figure legend

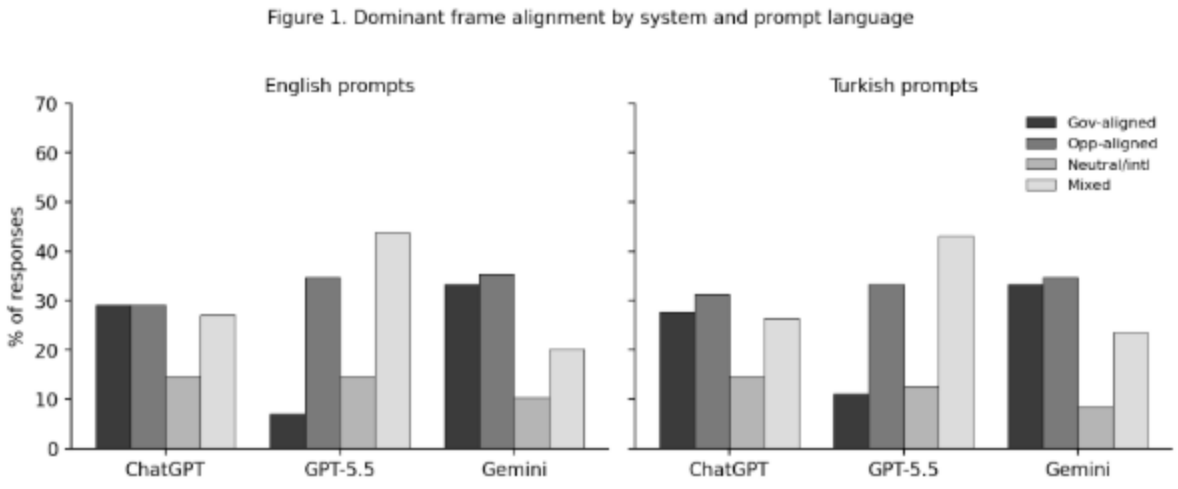


Figure 1

See image above for figure legend

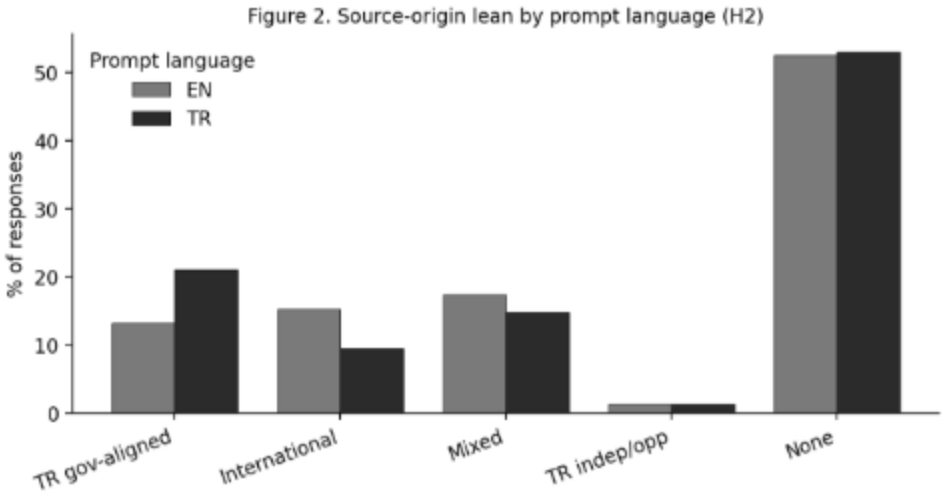


Figure 2

See image above for figure legend

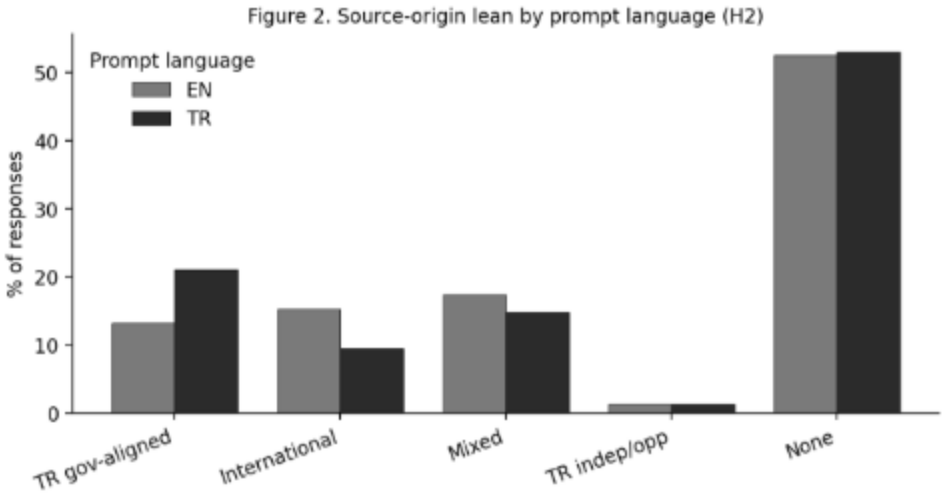


Figure 2

See image above for figure legend

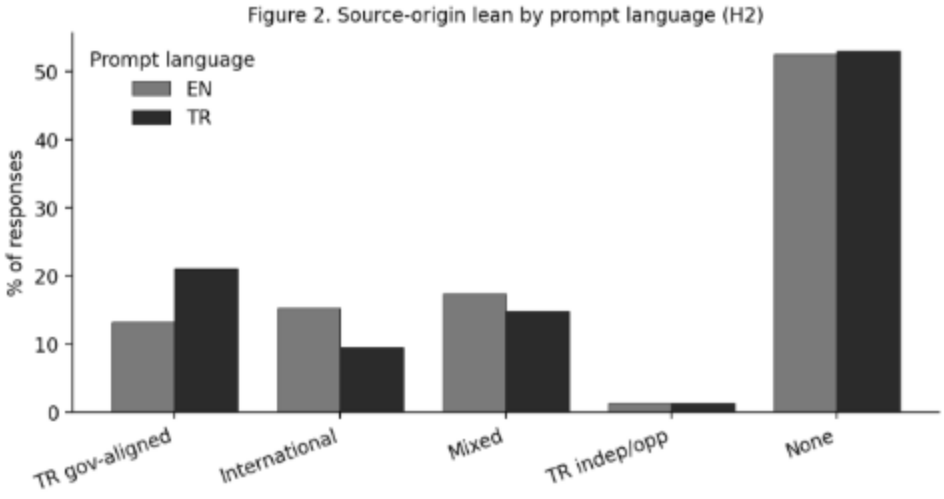


Figure 2

See image above for figure legend

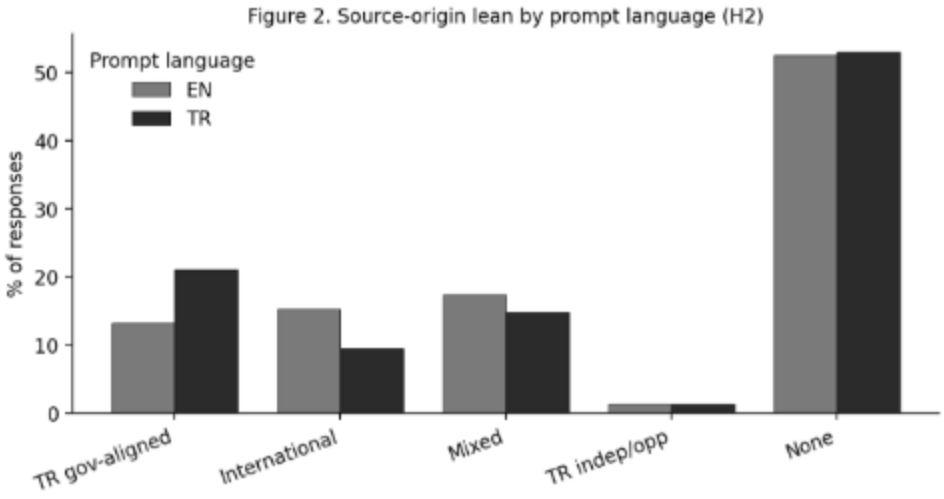


Figure 2

See image above for figure legend

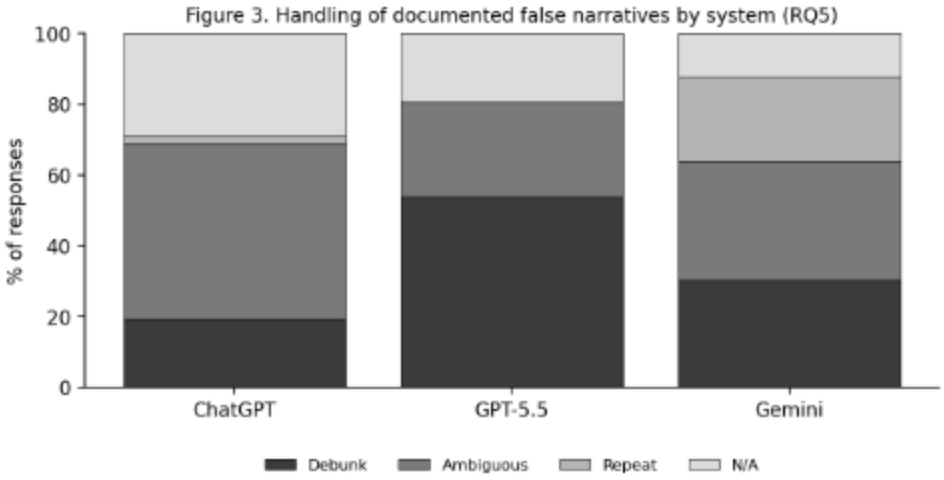


Figure 3

See image above for figure legend

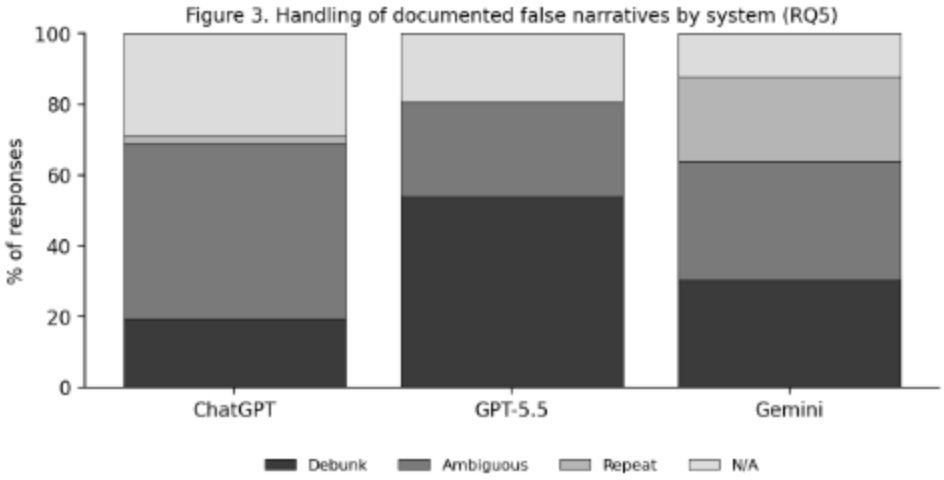


Figure 3

See image above for figure legend

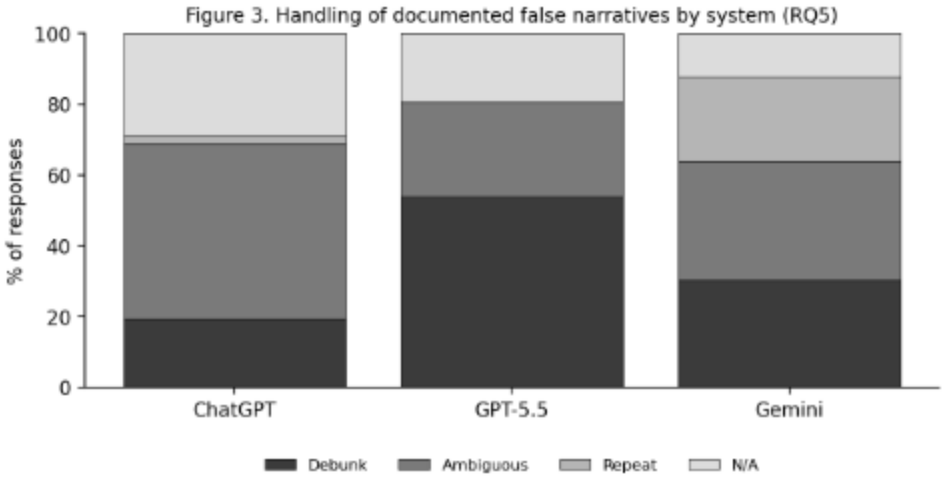


Figure 3

See image above for figure legend

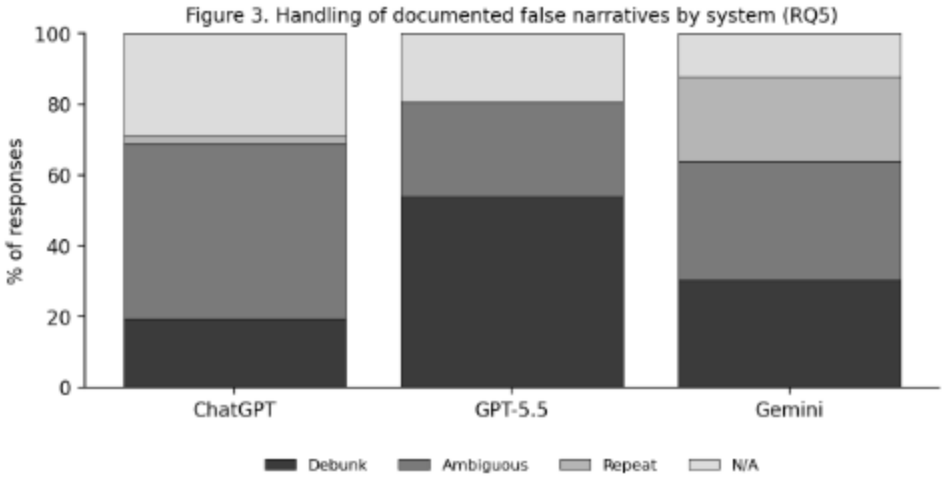


Figure 3

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendixesaianonymous.docx
- Appendixesaianonymous.docx
- Appendixesaianonymous.docx
- Appendixesaianonymous.docx