NBER WORKING PAPER SERIES

DIVERGENCE IN CLIMATE CHANGE COMMUNICATION: LLM-BASED EVIDENCE FROM THE IPCC AND THE PRESS

Sebastian Galiani
Franco Mettola La Giglia
Raul A. Sosa

Working Paper 35216
http://www.nber.org/papers/w35216

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2026

We used generative artificial intelligence tools (Claude by Anthropic) for coding assistance, literature verification, and editorial support. All analytical content and substantive conclusions are the authors' own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Divergence in Climate Change Communication: LLM-Based Evidence from the IPCC and the Press
Sebastian Galiani, Franco Mettola La Giglia, and Raul A. Sosa
NBER Working Paper No. 35216
May 2026
JEL No. D83, Q54

ABSTRACT

Public summaries of IPCC climate assessments lean toward the more severe end of the technical evidence. The pattern appears at two stages: the IPCC's lead authors and member governments produce the Summary for Policymakers (SPM) from the Technical Summary (TS), and newspapers then cover the SPM. We use LLMs to score about 114,000 matched claim pairs from all six Assessment Reports (1990 to 2023) and ten major US and UK outlets. Both stages systematically shift toward the more severe end of the source while staying inside the IPCC's accepted scientific ranges. The shift comes mainly from emphasizing higher-impact magnitudes within reported ranges, less from uncertainty compression, and almost none from selecting worst-case emissions scenarios. Left- and right-leaning outlets show similar patterns.

Sebastian Galiani
University of Maryland, College Park
Department of Economics
and NBER
sgaliani@umd.edu

Raul A. Sosa
Universidad de San Andrés
Department of Economics
rsosa@udesa.edu.ar

Franco Mettola La Giglia
Universidad de San Andrés
fmettolalagiglia@udesa.edu.ar

Divergence in Climate Change Communication: LLM-Based Evidence from the IPCC and the Press

By Sebastian Galiani, Franco Mettola La Giglia, and Raul A. Sosa*

Public summaries of IPCC climate assessments lean toward the more severe end of the technical evidence. The pattern appears at two stages: the IPCC's lead authors and member governments produce the Summary for Policymakers (SPM) from the Technical Summary (TS), and newspapers then cover the SPM. We use LLMs to score about 114,000 matched claim pairs from all six Assessment Reports (1990 to 2023) and ten major US and UK outlets. Both stages systematically shift toward the more severe end of the source while staying inside the IPCC's accepted scientific ranges. The shift comes mainly from emphasizing higher-impact magnitudes within reported ranges, less from uncertainty compression, and almost none from selecting worst-case emissions scenarios. Left- and right-leaning outlets show similar patterns.

Keywords: Climate change communication, IPCC, Scientific consensus, Large language models, Text-as-data
JEL: Q54, D83, L82

Climate policy is among the largest international resource-allocation negotiations of recent decades. Trillions of dollars per decade flow into mitigation and adaptation (Kreibiehl et al., 2022). Almost every national government has formal climate legislation (Eskander and Fankhauser, 2020), and United Nations treaties increasingly shape domestic policy (Falkner, 2016; Bodansky, 2016). The scientific foundation for this effort is the Intergovernmental Panel on Climate Change (IPCC). Established in 1988 by the World Meteorological Organization and the United Nations Environment Programme, the IPCC assesses published climate science and synthesizes the state of knowledge for policymakers in periodic Assessment Reports. These reports are cited at every major climate negotiation, in every national climate strategy document, and across the climate-science journalism that reaches the broader public (Vardy et al., 2017; Hulme and Mahony, 2010).

Among scientific advisory bodies, the IPCC is unusual. Its lead authors and member governments negotiate the wording of the public-facing summaries before release. Each Assessment Report comes from three Working Groups: WG1 covering physical science, WG2 covering impacts and adaptation, and WG3 covering

* Galiani: University of Maryland, email sgaliani@umd.edu. Mettola La Giglia and Sosa: Universidad de San Andrés. We used generative artificial intelligence tools (Claude by Anthropic) for coding assistance, literature verification, and editorial support. All analytical content and substantive conclusions are the authors' own.

mitigation. Within each Working Group, the lead authors produce a chapter-length scientific assessment, distill it into a Technical Summary (TS), and condense it further into a Summary for Policymakers (SPM). The SPM reaches policymakers, journalists, and the public. Before its release, representatives of all 195 IPCC member governments approve it line-by-line in plenary (Mach et al., 2016; De Pryck, 2021). No journal article, peer-reviewed survey, or national-academy report has anything like this.

Plenary approval gives the SPM political legitimacy and binds member states to its language. But it also makes the public-facing document a politically negotiated artifact rather than a neutral summary of the underlying science. Once the plenary closes, the SPM moves from negotiating room to newsroom. Journalists, editors, and columnists select among its claims, frame them, and filter them through editorial and ideological priors before findings reach voters and policymakers. At each step of this chain, we measure the directional reweighting between the summary and its immediate source.

To measure this at scale, we use large language models (LLMs) to extract structured records of scientific claims from each document in the chain. The corpus combines the IPCC Working Group documents across six Assessment Reports (1990 to 2023) with 116,045 newspaper articles from ten major US and UK outlets. We extract and link claims and then score about 1,000 TS-to-SPM pairs and 113,681 SPM-to-media pairs along three dimensions. *Severity shift* captures whether the summary picks the upper end of the source's range. *Uncertainty compression* captures whether hedges and qualifiers get dropped. *Scenario salience* captures whether worst-case emissions scenarios get foregrounded. We score each pair with three LLMs: OpenAI's GPT-5-mini, Anthropic's Claude Haiku 4.5, and Google's Gemini 2.5 Flash.

We document three findings. First, the TS-to-SPM stage reweights severity upward in every Assessment Report from 1990 to 2023. Second, the SPM-to-media stage shows the same upward reweighting across the four cycles we measure, AR3 (2001) through AR6 (2021 to 2023). Third, across the panel of ten outlets, left- and right-leaning groups show similar patterns. At both stages, the shift comes mainly from emphasizing higher-impact magnitudes within reported ranges, less from uncertainty compression, and almost none from selecting worst-case emissions scenarios.

The pattern we document is a reweighting within the IPCC's accepted scientific ranges. Every claim in our corpus traces back to an IPCC document, either drawn directly from the IPCC or made by a newspaper article covering an IPCC report, and the SPM and media summaries of these claims mostly stay within the IPCC's accepted scientific ranges. What we measure, then, is not falsehood, fabrication, or manipulation. Each new document in the chain makes the same directional choice when condensing its source, and that choice fits the rational institutional incentives at each stage. Lead authors write for a policy audience whose costs of under-preparation exceed its costs of over-preparation, so when the source admits a range, foregrounding the upper end is the conservative call. Every line of the SPM requires approval from 195 government delegations, and because a single objection blocks consensus, the wording converges on what the most risk-averse delegation will accept. News editors then select among the SPM's claims based on novelty, urgency, and salience, all of which favor the higher-impact end of any quantitative range.

This paper sits at the intersection of three literatures. The first treats expert assessment bodies as politically negotiated artifacts rather than neutral summaries of the underlying science. Foundational work locates them at a contested boundary between technical authority and political negotiation (Jasanoff, 1990). Hulme and Mahony (2010) surveys the extension of this argument to the IPCC, where compressing scientific disagreements into a single consensus document is itself a political act. Beck (2011) diagnoses this institutional pattern as a "linear model" of expertise that depoliticizes policy while politicizing the science. On the empirical side, O'Reilly, Oreskes and Oppenheimer (2012) provide ethnographic evidence that IPCC consensus on the West Antarctic ice sheet shifted between AR3 and AR4 through a mix of new observations and institutional factors including chapter restructuring and author turnover. De Pryck (2021) provides further ethnographic evidence from the AR5 Synthesis Report SPM that delegate negotiation reshapes wording and emphasis during the approval process. The closest quantitative precedent is Mach et al. (2016), who show that AR4 SPMs expand by 17 to 53 percent and AR5 SPMs by 20 to 35 percent during government review and approval, with plenary revisions emphasizing policy relevance and comprehensiveness. Their analysis operates at word-count and revision-purpose resolution across two cycles. From a different vantage point, Brysse et al. (2013) argue that IPCC physical-science projections err on the side of caution when checked against subsequently observed outcomes. Their benchmark is later observation; ours is the previous document in the chain.

The second literature is the political economy of media coverage, which models such coverage as a joint function of demand-side audience preferences and supply-side production incentives. The canonical method for measuring slant against an external reference text decomposes newspaper bias along these two dimensions (Gentzkow and Shapiro, 2010). Mullainathan and Shleifer (2005) formalize the demand side, showing that profit-maximizing outlets slant toward audience priors and that competition does not eliminate bias when readers segregate by ideology. Gentzkow, Shapiro and Stone (2015) synthesize the broader theoretical literature on both supply-driven and demand-driven bias. Slanted coverage causally moves political behavior (DellaVigna and Kaplan, 2007; Martin and Yurukoglu, 2017; DellaVigna and Gentzkow, 2010). In climate journalism, Boykoff and Boykoff (2004) documents that prestige US newspapers from 1988 to 2002 gave roughly equal attention to anthropogenic and naturalistic explanations of warming despite scientific consensus, generating "balance as bias." Stecula and Merkley (2019) extend the period through 2014 and document a steady decline in uncertainty framing in major US outlets. Cross-nationally, Painter (2013) characterizes coverage along risk and uncertainty frames. O'Neill et al. (2015) hand-code dominant frames in legacy and social-media coverage of the AR5 reports, the closest precedent for our second-stage analysis. Their object, however, is frame selection across outlets rather than severity reweighting of source claims.

The third literature is text-as-data and LLM-based measurement, the methodological lineage of our scoring. Two key reviews synthesize the dictionary, supervised, and topic-model approaches that underpin the field's foundational measurement papers (Grimmer and Stewart, 2013; Gentzkow, Kelly and Taddy, 2019). Applications include Baker, Bloom and Davis (2016) on policy uncertainty and Hassan et al. (2019) on firm-level political risk. Ash and Hansen (2023) updates the synthesis for the deep-learning era. The closest methodological precedent is Galiani et al. (2026), who apply LLM-based structured extraction to 27,464 economics journal articles, tracking efficiency- and equity-framing from 1950 to 2021. Validity work for LLM-based measurement comes from three sources. Gilardi, Alizadeh and Kubli (2023) demonstrate that frontier LLMs match or exceed crowd workers on text-annotation tasks. Argyle et al. (2023) introduce the notion of algorithmic fidelity for using LLMs in social-science research. Egami et al. (2023) develop design-based corrections for downstream regressions using LLM-extracted measures.

Our contribution is, to our knowledge, the first claim-level, cross-LLM-validated measurement of how IPCC findings are summarized for the public. The measurement spans the drafting of the SPM from the TS within the IPCC and the subsequent newspaper coverage of the SPM, across all six Assessment Reports from 1990 to 2023. It establishes that each step systematically emphasizes the more severe end of the source, with left- and right-leaning newspapers showing the same pattern.

The remainder of the paper is organized as follows. Section I describes the IPCC and media data. Section II presents the extraction and scoring methodology. Section III reports the results. Section IV concludes.

I. Data

We work with two text corpora: IPCC documents and newspaper articles. The IPCC corpus contains the official documents from all six Assessment Reports between 1990 and 2023, with three Working Groups per cycle, each contributing a TS source and an SPM. The media corpus contains 116,045 full-text newspaper articles from ten major US and UK outlets, covering IPCC release windows from AR3 (2001) through AR6 (2021 to 2023). Our unit of analysis is the *claim pair*. Each SPM claim is linked to its semantically nearest TS source; the same holds for each media claim and its nearest SPM source.

A. IPCC Documents

The IPCC documents we use form a layered structure within each Working Group of each Assessment Report. The chapter text contains the underlying scientific assessment with thousands of cited primary studies. The TS is the lead authors' Working Group-level synthesis, published as a separate document from AR4 (2007) onward; for AR1, AR2, and AR3, the synthesis role is filled by the full Working Group volume itself. The SPM is the policy-facing document, approved line-by-line by member governments during a plenary session lasting four to seven days.

We treat the TS as the lead authors' pre-plenary distillation of the underlying science and use it as the benchmark for SPM divergence. The chapter text contains too much heterogeneous material to serve as a single reference. The TS captures what the lead authors themselves were prepared to say before the plenary process began. We use the SPM as the benchmark for media coverage because journalists, policymakers, and press releases overwhelmingly cite the SPM at release time.

The corpus comprises eighteen Working Group by Assessment Report combinations across the six cycles. The claim-level extraction pipeline operates uniformly on the underlying scientific text, whether the AR4-onward TS document or the AR1-AR3 Working Group volume, and the assessment-to-summary separation is preserved by the published structure itself.

The IPCC documents support structured extraction. The SPM is not free-form narrative but an ordered list of numbered claims. From AR4 onward, each finding carries a hierarchical claim code such as `A.1.1` or `B.2` with bracketed cross-references to the supporting TS sections such as {`2.3, TS.2.2`}, so that each claim points back to its evidentiary basis. Calibrated uncertainty terms (*virtually certain*, *very likely*, *likely*) and confidence qualifiers (*very high confidence*, *high confidence*) likewise appear consistently from AR4 onward. We exploit both the cross-references (where present) and the calibrated language during extraction and linking. After linking and de-duplication, the corpus yields approximately 1,000 TS-to-SPM scored claim pairs.

B. Media Corpus

The media corpus comprises 116,045 full-text newspaper articles from ten outlets spanning the US and UK ideological spectrum. The corpus covers IPCC releases continuously across AR3 (2001) through AR6 (2021 to 2023). The five left-leaning outlets are The Guardian, the New York Times, the BBC, the Washington Post, and The Independent. The five right-leaning outlets are the Wall Street Journal, the Daily Mail, The Telegraph, Fox News, and the New York Post. Ideological classification draws on the slant measurement framework of Gentzkow and Shapiro (2010) for US outlets and on editorial endorsement records in recent UK general elections for British outlets. The ten outlets together account for the majority of legacy press IPCC coverage in their two countries during the AR6 window.

The Guardian Content API provides direct full-text access back to 1999. The New York Times Article Search API returns metadata back to 1851, and we recover full text for matched URLs from the Wayback Machine. We obtain Wall Street Journal coverage through ProQuest full-text exports under our institutional subscription. For the remaining seven outlets we use SerpAPI, a commercial search service that queries Google's index, to enumerate IPCC-related URLs and then fetch each article directly.[1] Across all channels, we match articles against IPCC-related search terms ("IPCC," "Intergovernmental Panel on Climate Change," "Climate Assessment Report," plus the specific report title for each cycle) and de-duplicate by URL canonicalization.

For each Assessment Report and each Working Group with a separate release date, we define a six-month window beginning on the official publication date. We collect every article in that window mentioning IPCC-related terms. AR1, AR2, and AR3 each consist of a single combined volume and contribute one window each. AR4, AR5, and AR6 each contribute four windows (one per Working Group plus one for the Synthesis Report), for fifteen release windows in total.

Two outlets dominate volume: The Guardian contributes 91,342 articles and the New York Times 7,020. Without intervention, The Guardian alone would account for 78 percent of the corpus, swamping cross-outlet comparisons. We therefore cap each outlet at 1,500 articles per Working Group per Assessment Report.[2] The cap binds only for The Guardian, in every Working Group by Assessment Report cell from AR3 onward. No other outlet exceeds 1,000 articles in any single cell. We extract individual scientific claims from each article (about 1.5 claims per article on average), and these extracted media claims are the unit we link and score. The per-cycle media-claim counts grow from 10,073 in AR3 to 27,444 in AR4, 30,137 in AR5, and 105,336 in AR6. The expansion reflects both the actual rise in coverage volume documented by Schmidt, Ivanova and Schäfer (2013) and the maturation of digital archives. Our scaled severity, uncertainty, and scenario measures stay insensitive to overall coverage volume because we compute each per claim pair rather than aggregating.

II. Methodology

We develop a three-step measurement pipeline. First, we extract structured claim records from each document and link each summary claim to its source. Second, we score each linked pair along three theoretically motivated dimensions. Third, we aggregate scored pairs into pooled estimates with bootstrap confidence intervals and audit measurement validity. Extraction uses GPT-5-mini. Scoring

[1] We do not access content behind paywalls. For outlets that require institutional subscriptions, we substitute database-mediated access where available.

[2] Appendix C reports five sample restrictions: three outlet-removal checks (without WSJ, without The Guardian, without both) and two equalizations (smallest-outlet cap, per-outlet $1/N$ pooling).

uses three LLMs: OpenAI's GPT-5-mini, Anthropic's Claude Haiku 4.5, and Google's Gemini 2.5 Flash. The per-pair mean is the headline. Their agreement is our measurement-error check.

A. *Claim Extraction and Linking*

For each IPCC document (TS or SPM), GPT-5-mini reads the full text and returns one structured record per scientific claim.[3] Each record encodes the claim's text, its assigned topic from an eleven-category taxonomy, its calibrated confidence level (one of the IPCC's five qualifiers from *very high* to *very low confidence*), any IPCC likelihood term (*virtually certain*, *very likely*, *likely* with associated probability ranges), and the emission scenario referenced when relevant. The IPCC defines a small number of standard emission scenarios spanning very low to very high greenhouse gas trajectories. We tag each claim with the specific scenario it cites or with no scenario at all. For newspaper articles, the schema mirrors the IPCC schema and adds three media-specific fields: the article's tone (alarming, neutral, reassuring, or skeptical), an indicator for whether the article reproduces the IPCC's calibrated language, and the set of editorial framing words that amplify or attenuate the underlying claim. The corpus yields approximately 20,000 TS records, 1,350 SPM records, and 174,000 media claim records (around 1.5 claims per article).

We link each summary claim to its semantically nearest source claim. TS-to-SPM linking uses two complementary signals. The first is the explicit cross-references the IPCC embeds in AR4 through AR6, where each SPM claim cites the underlying TS section in curly braces. The second is topic-category matching for AR1 through AR3, which lack formal cross-references. Topic matching assigns links at two confidence levels: an exact topic-category match receives linkage confidence 1.0, and a related topic match via a manually constructed topic-group mapping receives linkage confidence 0.7. SPM-to-media linking uses topic-category matching alone, which is deterministic and requires no model calls. We exclude articles tagged *general_climate*, which mention the IPCC without committing to a specific topic, to prevent spurious matches. After linking and de-duplication, the analyzable sample comprises approximately 1,000 TS-to-SPM claim pairs from all six Assessment Reports and 113,681 SPM-to-media claim pairs from AR3 through AR6.[4]

B. *Divergence Scoring*

For each linked pair we score divergence on a $\{-2, -1, 0, +1, +2\}$ scale, rescaled to $[-1, +1]$ by dividing by 2. On the rescaled scale, 0 represents faithful transmission, $+1$ a maximally upward semantic reweighting relative to the source, and -1 a maximally downward reweighting. The metric is signed by construction. A

[3]The full prompt suite for both extraction and scoring is in Appendix E.
[4]Appendix A reports the per-outlet, per-AR breakdown of these pairs.

pooled mean of +0.04 measures a directional asymmetry, not a magnitude. We use *amplification* as shorthand for the consistently positive direction we observe at both pipeline stages in every cycle we measure (all six Assessment Reports for TS-to-SPM; AR3 through AR6 for SPM-to-media). The term has a specific technical meaning: within-range reweighting, where both documents typically fall inside the IPCC's accepted scientific ranges. The broader interpretive framing appears in the introduction.

Each linked pair receives three dimension scores. *Severity shift* (denoted s) measures whether the summary emphasizes the more extreme end of the source's quantitative range or presents a more moderate framing. *Uncertainty compression* (denoted u) measures whether the summary drops calibrated qualifiers, removes numerical confidence intervals, or replaces hedged language with definitive language. Negative scores indicate the converse, added doubt or softened definitive language. *Scenario salience* (denoted c) measures whether the summary foregrounds a worst-case emissions scenario when the source covered the full set of scenarios with comparable weight. Negative scores indicate the converse, selective foregrounding of best-case scenarios. Table 1 reports one representative claim pair for each dimension at each pipeline stage.[5]

The composite divergence score is the unweighted mean of the three dimensions:

$$d_{ij} = \frac{s_{ij} + u_{ij} + c_{ij}}{3}, \tag{1}$$

where i indexes the source claim and j the summary. We score each linked pair independently with all three LLMs. The per-pair composite reported throughout Section III is the average of the three model scores.[6] Each model also returns a free-text justification for every score. We retain these justifications for spot auditing; they appear in the rightmost column of Table 1. For media articles, the scoring step batches each article's claims against all SPM claims sharing its topic, which keeps the number of model calls manageable across the 116,045-article corpus while preserving per-pair scoring fidelity.

C. Statistical Inference and Validity

We construct 95% confidence intervals via percentile bootstrap with 5,000 resamples. The null hypothesis at every step is faithful transmission, $H_0 : \bar{d} = 0$. For the TS-to-SPM analysis, the unit of observation is the SPM claim, pooled across Working Groups within each Assessment Report; the bootstrap resamples SPM claims directly. For the media analysis, the unit is the SPM-media claim pair, and we cluster the bootstrap at the article level because multiple claims from the same article share unobserved editorial context. The pooled mean weights

[5] Appendix D reports additional claim pairs spanning the full directional range (amplification, neutral, and attenuation) at both pipeline stages.

[6] Appendix B reports alternative ensemble methods (IVW mean, empirical-Bayes shrinkage, single-factor model) that yield essentially identical per-AR composites.

each pair equally, which means outlets that produce more IPCC-adjacent coverage contribute more weight to the headline composite.[7] The main results pool across Working Groups to maximize statistical power. The left-leaning and right-leaning outlet groups used in Section III.C are defined in Section I.B.

LLM-based measurement raises three substantive validity concerns. The first is hallucination: LLMs can fabricate content not present in the source text (Korinek, 2023). We attenuate this risk through cross-model averaging across three independently developed families and retain the per-pair reasoning fields each model returns for spot auditing. The second is single-model bias: a single model's training corpus or alignment objective could distort scoring systematically. Our three model families are independently trained on different pretraining corpora and post-training pipelines. Figure 4 shows that the cascade pattern appears in each model individually, with magnitude differences but consistent direction. The third is downstream inferential bias: Egami et al. (2023) show that naive use of LLM-extracted scores as outcome variables in regressions can produce biased coefficients and invalid confidence intervals. The descriptive results in Section III report unadjusted means with bootstrap CIs, so they are not subject to this concern. The central object is the pooled mean rather than a regression coefficient.

[7] Appendix C reports five sample restrictions that probe sensitivity to outlet weighting and to inclusion of high-volume outlets. All confirm that the cumulative cascade pattern survives.

TABLE 1—REPRESENTATIVE CLAIM PAIRS AT EACH PIPELINE STAGE, ONE PER DIVERGENCE DIMENSION.

Panel A. Technical Summary → Summary for Policymakers			
Dimension	**TS finding**	**SPM claim**	**Why it scores high**
Severity shift	Relatively small climate changes can cause large water resource problems in many areas, especially arid and semi-arid regions and those humid areas where demand or pollution has led to water scarcity.	Intensive assessment of water resources and water quality, especially in arid and semi-arid developing countries and their sensitivity to climate change.	Converts a descriptive statement about regional water-resource sensitivity into a prescriptive call for "intensive assessment," adding an urgency framing the TS evidence does not itself convey.
Scenario salience	An average rate of increase of global mean temperature during the next century of about 0.3°C per decade (with an uncertainty range of 0.2 to 0.5°C per decade) assuming IPCC Scenario A (Business-as-Usual). Lower rates apply under Scenarios B, C, and D.	The rate of change of global temperatures predicted for Business-as-Usual emissions will be greater than those which have occurred naturally on earth over the last 10,000 years, and the rise in sea level will be about three to six times faster than that seen over the last 100 years.	Reports only the Business-as-Usual rate as the headline finding. The TS gives equal weight to Scenarios B, C, and D, which yield slower rates; the SPM omits them, privileging the high-emissions trajectory.
Uncertainty compression	Natural processes and management regimes may reduce or increase the amount of carbon stored. The relative contributions of fossil fuel and land-use sources, and of ocean and land uptake, are subject to estimation uncertainty.	Emissions of CO_2 due to fossil fuel burning are *virtually certain* to be the dominant influence on the trends in atmospheric CO_2 concentration during the 21st century.	The SPM uses the IPCC's strongest calibrated qualifier (*virtually certain*, $> 99\%$) for fossil-fuel dominance. The TS hedges with "may reduce or increase" and acknowledges land-use and uptake uncertainties; the SPM compresses both.
Panel B. Summary for Policymakers → Media coverage			
Dimension	**SPM claim**	**Media excerpt**	**Why it scores high**
Severity shift	Knowledge of the global-mean warming and change in precipitation is of limited use in determining the impacts of climate change, for instance on agriculture. For this we need to know changes, regionally and seasonally.	The first point stems from a fear that environmental problems, especially those leading to food shortages, might ignite regional wars in which the U.S. might have to take sides.	Escalates a methodological caveat about needing regional impact data into the prospect of regional wars. The magnitude of the summary claim has no anchor in the source statement.
Scenario salience	Biodiversity and ecosystem services have limited capacity to adapt to increasing global warming levels, which will make climate resilient development progressively harder to achieve beyond 1.5°C warming (very high confidence).	Researchers at the University of Exeter modelled which coastlines might become more hospitable to the coral as waters warm, based on greenhouse gas emissions at the high end of current predictions.	The SPM discusses biodiversity limits across warming levels. The summary privileges a single high-emissions scenario as the baseline for its modelled outcome, without acknowledging the lower-warming alternatives.
Uncertainty compression	Changes in solar irradiance since 1750 are *estimated* to cause a radiative forcing of $+0.12$ [$+0.06$ to $+0.30$] W m^{-2}, which is less than half the *estimate* given in the TAR.	It is absolutely clear that the sun is nothing to do with the recent warming.	The SPM reports a quantified central estimate with an explicit $[+0.06, +0.30]$ uncertainty range and the qualifier *estimated*. The summary replaces the entire range with "absolutely clear," compressing both numerical bounds and linguistic hedging.

Note: Each row reports a claim pair scored by the three-LLM ensemble. Panel A claim pairs score +0.50 on the focal dimension. Panel B pairs score +1.00. The asymmetry reflects the paper's headline finding that TS-to-SPM divergence is smaller in magnitude than SPM-to-media divergence, not a sampling artifact.

III. Results

Section III.A establishes findings 1 and 2: positive amplification at both pipeline stages. Section III.B decomposes the composite into severity, uncertainty, and scenario channels, with severity dominant. Section III.C establishes finding 3: similar amplification across the ideological spectrum. Section III.D verifies the cascade reproduces across the three LLMs.

A. The Amplification Cascade

The TS-to-SPM stage shows positive amplification in every Assessment Report (Figure 1), ranging from +0.040 in AR1 and AR3 to +0.087 in AR4. Each 95% bootstrap CI excludes zero. The pattern is consistent with the institutional literature on IPCC consensus formation (Mach et al., 2016; De Pryck, 2021; Hulme and Mahony, 2010).

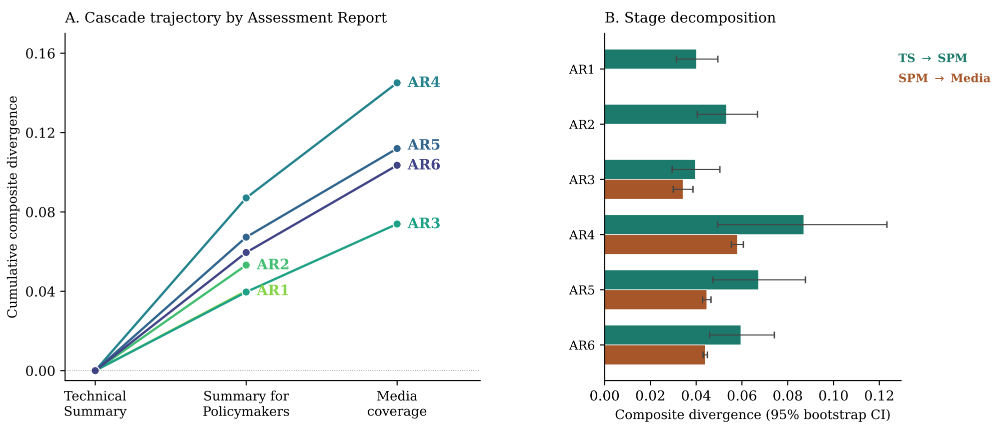

Figure 1. The amplification cascade, 1990 to 2023.

Note: Panel A. Cumulative composite divergence, with one polyline per Assessment Report tracing the trajectory from TS through SPM to media coverage. Panel B. Composite divergence per AR decomposed by stage, with 95% bootstrap CIs (TS → SPM in teal, SPM → Media in rust). The media stage covers AR3 through AR6 only. AR1 and AR2 polylines truncate at the SPM node in Panel A and lack rust bars in Panel B.

The SPM-to-media stage, measured from AR3 (2001) through AR6, also shows positive amplification in every cycle, ranging from +0.034 in AR3 to +0.058 in AR4, though uniformly smaller than the IPCC's internal amplification at the same cycles. The earliest cycle we observe (AR3, 2001) coincides with two contemporaneous developments: the weakening of the "balance as bias" norm in US prestige climate coverage (Boykoff and Boykoff, 2004, 2007) and a broader rise in coverage volume (Schmidt, Ivanova and Schäfer, 2013). The cumulative cascade pattern survives all five sample restrictions.[8]

B. Three Dimensions of Divergence

Severity shift is the primary channel of the directional reweighting at both stages (Figure 2). The SPM consistently selects the upper end of the TS's reported quantitative ranges, and media coverage reproduces and extends this selection. Both summaries mostly stay inside the source range. What shifts is which end gets foregrounded. The pattern is consistent with Painter (2013), who documents that "disaster" framings dominate international climate news relative to more hedged "uncertainty" framings.

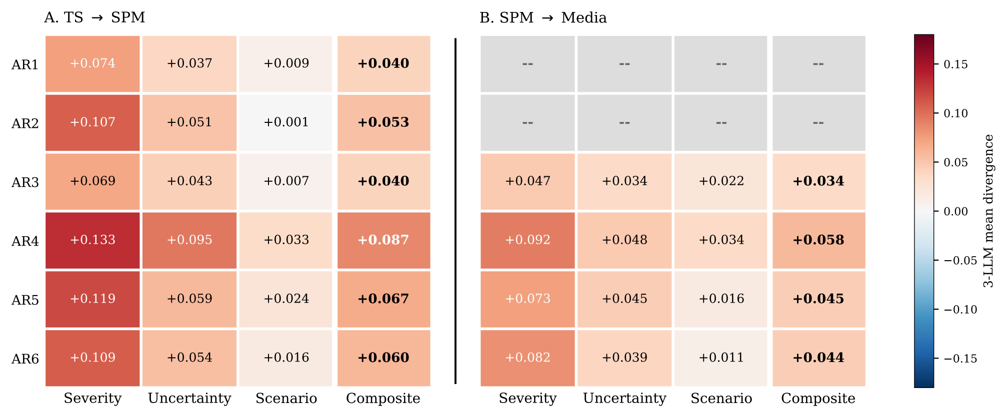

FIGURE 2. DIVERGENCE DECOMPOSED BY CHANNEL: SEVERITY, UNCERTAINTY, AND SCENARIO.

Note: Panel A. TS → SPM, AR1 through AR6. Panel B. SPM → Media, AR3 through AR6. Each cell reports the three-LLM mean divergence on $[-1, +1]$.

Uncertainty compression is a secondary channel, more prominent at the TS-to-SPM stage than in media. The pattern is consistent with Antilla (2005), who documents that early-2000s US newspaper coverage amplified uncertainty and skepticism frames rather than removing scientific hedges. Scenario salience is small everywhere: neither the IPCC editorial process nor media coverage systematically privileges worst-case scenarios over the underlying distribution. The within-range amplification we measure is therefore not driven by selective scenario citation.

C. Outlet Heterogeneity by Ideology

The dispersion within ideology groups exceeds the dispersion between them (Figure 3, Panel A). Most outlets cluster between +0.025 and +0.060. The

[8]Appendix C reports three restrictions that remove the highest-volume outlets and two that equalize outlet weight.

Guardian is the most stable amplifier. The Daily Mail is roughly in line with center-left outlets despite its right-leaning classification. Fox News pools to a near-zero composite, with substantial cross-cycle variance that the pooled estimate hides. The Wall Street Journal is the only outlet that consistently understates the SPM, and its largest negative score concentrates in AR3 (2001), the era Boykoff and Boykoff (2004) identifies as dominated by "balance as bias" coverage in the US prestige press.

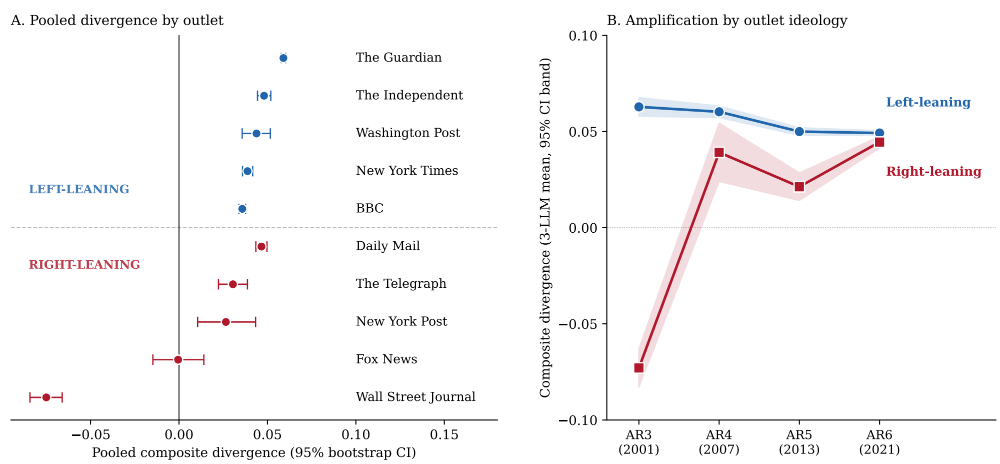

FIGURE 3. OUTLET-LEVEL HETEROGENEITY BY IDEOLOGY.

Note: Panel A. Pooled SPM-to-media composite divergence by outlet (sorted by descending mean within each ideology group), with 95% bootstrap CIs. Panel B. Three-LLM mean composite for left- and right-leaning outlet groups by AR (AR3 through AR6), with 95% bootstrap CI bands.

Across AR3 to AR6, left- and right-leaning outlets show similar amplification (Panel B). By AR6 the left-leaning group averages +0.046 and the right-leaning group +0.044, a gap of 0.002 well inside both bootstrap intervals. Neither group moves toward zero.

This cross-outlet similarity runs against demand-side predictions of persistent ideological niches (Mullainathan and Shleifer, 2005) and fits the supply-side reasoning of Gentzkow, Shapiro and Stone (2015). Section IV develops the theoretical interpretation in full, including the cross-country polarization context, the uncertainty-framing corroboration in US prestige press, and the distinction between cross-outlet coverage similarity and audience belief polarization.

D. Cross-Model Reproducibility

Every model detects the cascade at every cycle (Figure 4, Panel A). The cumulative shape rises through AR4 and plateaus at AR5 and AR6 in all three. Magnitudes differ by up to 0.06 in cumulative composite. Gemini scores systematically 0.04 to 0.06 below the ensemble mean, and GPT and Claude bracket the mean from above in most cycles. None is an outlier.

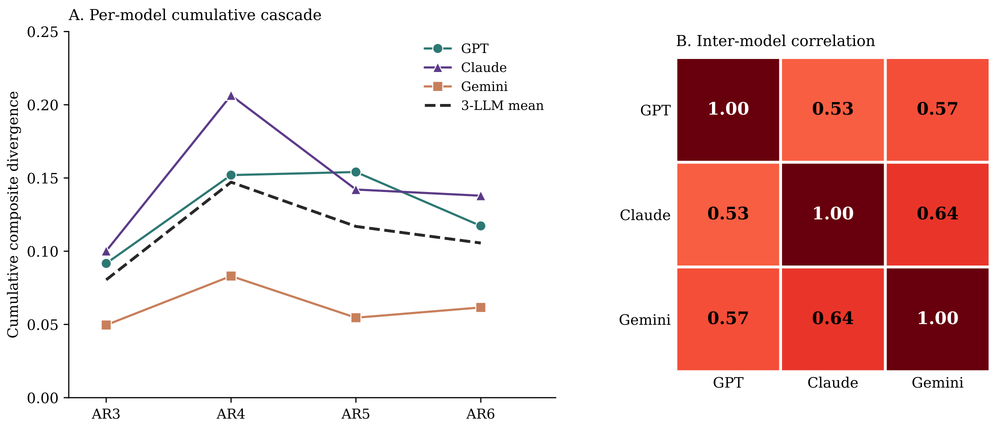

FIGURE 4. PER-LLM CUMULATIVE CASCADE AND INTER-MODEL CORRELATION.

Note: Panel A. Cumulative composite divergence (TS-to-SPM plus SPM-to-media) by AR (AR3 through AR6), plotted separately for each of the three LLMs alongside the three-LLM mean (dashed line). AR1 and AR2 are omitted because the media stage starts at AR3. Panel B. Pairwise Pearson correlations between LLM scores at the claim-pair level, pooled across the AR3 through AR6 media observations.

The pairwise inter-model correlations at the claim-pair level range from 0.54 to 0.65 (Panel B), indicating agreement on the direction and rough magnitude of divergence while leaving room for legitimate disagreement on borderline cases. Averaging across the three models reduces per-model idiosyncrasies, and the cascade pattern persists when each model is examined separately.

IV. Conclusion

We document a two-stage cascade of directional reweighting in IPCC climate communication: positive TS-to-SPM amplification in every cycle from AR1 (1990) through AR6 (2021 to 2023), positive SPM-to-media amplification in every cycle from AR3 (2001) onward, and similar amplification across left- and right-leaning outlets. Severity shift within accepted ranges carries the reweighting at both stages, and scenario salience is small at both, so the cascade is consistent with a small set of standard economic mechanisms rather than with selective citation of worst-case scenarios.

The TS-to-SPM stage follows from asymmetric loss in sender-receiver communication. Lead authors face a policy audience whose loss from underestimating climate risk exceeds its loss from overestimating, so the equilibrium signal leans toward the upper end of the calibrated range even when every actor reports honestly. Negotiated consensus across 195 government delegations compounds this effect: language must be acceptable to the most risk-averse delegation, and "we under-prepared" is much harder to defend in plenary than "we over-prepared." This consensus feature explains why TS-to-SPM amplification persists across all six cycles, including cycles that incorporated explicit institutional efforts to improve SPM accuracy.

The SPM-to-media stage follows from supply-side selection in news production (Gentzkow, Shapiro and Stone, 2015). Editors filter SPM claims through news-value criteria of novelty, urgency, and salience, all of which weight toward severity within the source range. The pattern's visibility from AR3 onward (Section III.A) coincides with the weakening of the "balance as bias" constraint in US prestige press: once that constraint loosened, supply-side selection on severity became measurable in coverage.

The similarity of amplification across left- and right-leaning outlets in our panel is direct evidence against the strong form of demand-side audience-niching theories (Mullainathan and Shleifer, 2005), which would predict large persistent ideological gaps in coverage of a high-stakes contested issue. Within-group dispersion exceeds between-group dispersion. Supply-side selection (Gentzkow, Shapiro and Stone, 2015) predicts exactly this pattern for issues that have reached broad scientific consensus and are no longer a profitable axis of differentiation. Independent corroboration comes from three angles. Stecula and Merkley (2019) document a steady decline in uncertainty framing across four major US outlets from 1988 to 2014. Boxell, Gentzkow and Shapiro (2017) and Boxell, Gentzkow and Shapiro (2024) situate the cross-outlet pattern in broader cross-country trends of moderate or declining polarization in news supply. Dechezleprêtre et al. (2025) document broad cross-country climate concern, including among conservative populations.

Polarization on the *belief* side remains real and growing (McCright and Dunlap, 2011), but our finding is on the *coverage* side. The two coexist: audience priors are not formed solely by news coverage, and Boxell, Gentzkow and Shapiro (2017) attribute most US polarization growth to demographic groups with the lowest internet and social media use.

The object we measure is directional reweighting relative to the immediate source, not the ex post accuracy of climate projections. To the extent that some high-severity scenarios later became viewed as less likely, that may reflect subsequent scientific updating rather than distortions introduced at the communication stage itself. The implication for readers of the IPCC chain is calibration rather than correction. A reader who treats the SPM as a neutral mid-range summary of the technical assessment receives a more severe-end-weighted version of that assessment. A reader who treats newspaper coverage as a neutral mid-range summary of the SPM receives a further severe-end-weighted version. Both summaries mostly stay within the scientifically accepted range, so the calibration adjustment is small and informational rather than corrective. IPCC outputs are anchored in peer-reviewed scientific evidence. The corrective is not to discount them but to read them with this within-range positioning in mind.

REFERENCES

Antilla, Liisa. 2005. "Climate of scepticism: US newspaper coverage of the science of climate change." *Global Environmental Change*, 15(4): 338–352.

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. "Out of One, Many: Using Language Models to Simulate Human Samples." *Political Analysis*, 31(3): 337–351.

Ash, Elliott, and Stephen Hansen. 2023. "Text Algorithms in Economics." *Annual Review of Economics*, 15: 659–688.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131(4): 1593–1636.

Beck, Silke. 2011. "Moving Beyond the Linear Model of Expertise? IPCC and the Test of Adaptation." *Regional Environmental Change*, 11(2): 297–306.

Bodansky, Daniel. 2016. "The Legal Character of the Paris Agreement." *Review of European, Comparative & International Environmental Law*, 25: 142–150.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Greater Internet Use Is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups." *Proceedings of the National Academy of Sciences*, 114(40): 10612–10617.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2024. "Cross-Country Trends in Affective Polarization." *Review of Economics and Statistics*, 106(2): 557–565.

Boykoff, Maxwell T., and Jules M. Boykoff. 2004. "Balance as bias: global warming and the US prestige press." *Global Environmental Change*, 14(2): 125–136.

Boykoff, Maxwell T., and Jules M. Boykoff. 2007. "Climate change and journalistic norms: A case-study of US mass-media coverage." *Geoforum*, 38(6): 1190–1204.

Brysse, Keynyn, Naomi Oreskes, Jessica O'Reilly, and Michael Oppenheimer. 2013. "Climate change prediction: Erring on the side of least drama?" *Global Environmental Change*, 23(1): 327–337.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2025. "Fighting Climate Change: International Attitudes Toward Climate Policies." *American Economic Review*, 115(4): 1258–1300.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics*, 122(3): 1187–1234.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics*, 2: 643–669.

De Pryck, Kari. 2021. "Intergovernmental Expert Consensus in the Making: The Case of the Summary for Policy Makers of the IPCC 2014 Synthesis Report." *Global Environmental Politics*, 21(1): 108–129.

Egami, Naoki, Musashi Hinck, Brandon M. Stewart, and Hanying Wei. 2023. "Using Imperfect Surrogates for Downstream Inference: Design-Based Supervised Learning for Social Science Applications of Large Language Models." Vol. 36.

Eskander, Shaikh M. S. U., and Sam Fankhauser. 2020. "Reduction in Greenhouse Gas Emissions from National Climate Legislation." *Nature Climate Change*, 10(8): 750–756.

Falkner, Robert. 2016. "The Paris Agreement and the New Logic of International Climate Politics." *International Affairs*, 92(5): 1107–1125.

Galiani, Sebastian, Ramiro H. Gálvez, Franco Mettola La Giglia, and Raul A. Sosa. 2026. "Measuring Efficiency and Equity Framing in Economics Research: LLM-Based Evidence from 1950 to 2021." National Bureau of Economic Research NBER Working Paper 34714.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica*, 78(1): 35–71.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature*, 57(3): 535–574.

Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone. 2015. "Media Bias in the Marketplace: Theory." In *Handbook of Media Economics*. Vol. 1B, , ed. Simon P. Anderson, Joel Waldfogel and David Strömberg, 623–645. North-Holland.

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences*, 120(30): e2305016120.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*, 21(3): 267–297.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. 2019. "Firm-Level Political Risk: Measurement and Effects." *Quarterly Journal of Economics*, 134(4): 2135–2202.

Hulme, Mike, and Martin Mahony. 2010. "Climate Change: What Do We Know About the IPCC?" *Progress in Physical Geography*, 34(5): 705–718.

Jasanoff, Sheila. 1990. *The Fifth Branch: Science Advisers as Policymakers.* Cambridge, MA:Harvard University Press.

Korinek, Anton. 2023. "Generative AI for Economic Research: Use Cases and Implications for Economists." *Journal of Economic Literature*, 61(4): 1281–1317.

Kreibiehl, Silvia, Tae Yong Jung, Stefano Battiston, Pablo Esteban Carvajal, Christa Clapp, Dipak Dasgupta, Nokuthula Dube, Raphaël Jachnik, Kanako Morita, Nahla Samargandi, and Mariama Williams. 2022. "Investment and Finance." *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, , ed. P. R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz and J. Malley, Chapter 15. Cambridge, UK and New York, NY, USA:Cambridge University Press.

Mach, Katharine J., Patrick T. Freeman, Michael D. Mastrandrea, and Christopher B. Field. 2016. "A Multistage Crucible of Revision and Approval Shapes IPCC Policymaker Summaries." *Science Advances*, 2(8): e1600421.

Martin, Gregory J., and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review*, 107(9): 2565–2599.

McCright, Aaron M., and Riley E. Dunlap. 2011. "The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010." *The Sociological Quarterly*, 52(2): 155–194.

Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The Market for News." *American Economic Review*, 95(4): 1031–1053.

O'Neill, Saffron, Hywel T. P. Williams, Tim Kurz, Bouke Wiersma, and Maxwell Boykoff. 2015. "Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report." *Nature Climate Change*, 5(4): 380–385.

O'Reilly, Jessica, Naomi Oreskes, and Michael Oppenheimer. 2012. "The Rapid Disintegration of Projections: The West Antarctic Ice Sheet and the Intergovernmental Panel on Climate Change." *Social Studies of Science*, 42(5): 709–731.

Painter, James. 2013. *Climate Change in the Media: Reporting Risk and Uncertainty.* London:I.B. Tauris.

Schmidt, Andreas, Ana Ivanova, and Mike S. Schäfer. 2013. “Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries.” *Global Environmental Change*, 23(5): 1233–1248.

Stecula, Dominik A., and Eric Merkley. 2019. “Framing Climate Change: Economics, Ideology, and Uncertainty in American News Media Content From 1988 to 2014.” *Frontiers in Communication*, 4: 6.

Vardy, Mark, Michael Oppenheimer, Navroz K. Dubash, Jessica O’Reilly, and Dale Jamieson. 2017. “The Intergovernmental Panel on Climate Change: Challenges and Opportunities.” *Annual Review of Environment and Resources*, 42: 55–75.

Descriptive Statistics

Table A1 reports the count of scored SPM-to-media claim pairs per outlet per Assessment Report for the four cycles used in the media-stage analysis (AR3 through AR6). The table makes two features of the corpus visible. First, the post-2001 corpus expands by roughly an order of magnitude per cycle, from 6,752 scored pairs in AR3 to 61,573 in AR6. Second, two outlets dominate volume after the 1,500-article cap is applied: The Guardian on the left and the Daily Mail on the right. The robustness checks reported in Appendix C address whether this volume asymmetry drives the headline pattern.

Table A1—Scored SPM-to-media claim pairs per outlet per Assessment Report (AR3 through AR6).

Outlet	**Ideology**	**AR3**	**AR4**	**AR5**	**AR6**	**Total**
The Guardian	Left	4,287	12,275	17,079	22,454	56,095
New York Times	Left	856	1,069	2,689	4,208	8,822
BBC	Left	64	397	4,096	15,842	20,399
Washington Post	Left	4	2	566	1,049	1,621
The Independent	Left	244	427	964	6,016	7,651
Wall Street Journal	Right	1,232	254	360	36	1,882
The Telegraph	Right	0	167	797	947	1,911
Daily Mail	Right	65	610	3,321	9,958	13,954
Fox News	Right	0	31	147	540	718
New York Post	Right	0	17	45	523	585
Total		6,752	15,249	30,064	61,573	113,638

Note: Each cell reports the number of scored SPM-to-media claim pairs surviving the linking and de-duplication steps in Section II. AR1 (1990) and AR2 (1995) are excluded from the media-stage analysis. The headline analysis caps each outlet's per-Working-Group per-Assessment-Report contribution at 1,500 articles to prevent any single outlet from swamping the pooled estimate. The cap binds only for The Guardian and binds in every cell from AR3 onward. A residual 43 pairs across AR4 (6), AR5 (2), and AR6 (35) lacked outlet metadata after the linking step and are omitted from this per-outlet breakdown. They are retained in the pooled composite divergence estimates reported in Section III (headline analyzable sample, 113,681 pairs).

Alternative Ensemble Methods

The headline composite reported in Section III is the unweighted arithmetic mean of the three LLMs' per-pair scores. We have computed three alternative ensemble methods on the same scored data and report them here as a robustness check.

The first alternative is the inverse-variance-weighted, or IVW, mean, where each model's per-pair score is weighted by the reciprocal of its within-AR variance on that dimension. Models that score with less spread carry more weight. The IVW estimate coincides with the simple mean when all three models share the same per-AR variance and is otherwise close to it.

The second alternative is empirical-Bayes shrinkage. Each per-pair simple-mean score is shrunk toward the AR-level grand mean by a factor that depends on the ratio of within-pair to between-pair variance. The shrinkage attenuates extreme per-pair scores but leaves AR-level pooled means almost unchanged.

The third alternative is a single-factor model that treats the three model scores as noisy measurements of a single latent divergence per pair, recovering the latent score via factor analysis. The single-factor model loads positively on all three indicators in every AR and reproduces the per-AR composite to within 0.01 of the simple mean in every cycle. Per-AR estimates under all four methods, namely simple mean, IVW, EB shrinkage, and single-factor, are reported in Table B1 below.

In all three alternatives, some pairs receive a score of zero or are dropped entirely. The reason is that not every claim pair receives a score from all three models. A small fraction of pairs return a parse error or an abstention from at least one model, leaving only one or two valid scores. The simple mean averages over the available models in those cases, so a pair with only one valid score uses that single score. The IVW and factor methods require all three models to have scored the pair, so single-model and two-model rows are dropped from those estimates. The shrinkage estimate uses the simple mean as input and inherits the simple-mean rule. Reported zeros for pairs in the IVW and factor columns therefore reflect dropped rows rather than computed null values.

Table B1—Per-AR composite divergence under four ensemble aggregation methods. 95% percentile bootstrap CIs in brackets.

Panel A. Technical Summary → Summary for Policymakers				
AR	**Simple mean**	**IVW**	**EB shrinkage**	**single-factor**
AR1	+0.040 [+0.031, +0.050]**	+0.013 [+0.008, +0.019]**	+0.040 [+0.037, +0.043]**	+0.040 [+0.031, +0.050]**
AR2	+0.053 [+0.041, +0.067]**	+0.000 [+0.000, +0.000]	+0.053 [+0.050, +0.057]**	+0.053 [+0.041, +0.067]**
AR3	+0.040 [+0.029, +0.051]**	+0.014 [+0.009, +0.020]**	+0.040 [+0.036, +0.044]**	+0.040 [+0.029, +0.051]**
AR4	+0.087 [+0.049, +0.124]**	+0.052 [+0.028, +0.079]**	+0.087 [+0.072, +0.102]**	+0.087 [+0.049, +0.124]**
AR5	+0.067 [+0.047, +0.087]**	+0.000 [+0.000, +0.000]	+0.067 [+0.061, +0.074]**	+0.067 [+0.047, +0.087]**
AR6	+0.060 [+0.046, +0.074]**	+0.036 [+0.025, +0.047]**	+0.060 [+0.054, +0.065]**	+0.060 [+0.046, +0.074]**
Panel B. Summary for Policymakers → Media coverage (AR3 through AR6)				
AR	**Simple mean**	**IVW**	**EB shrinkage**	**single-factor**
AR3	+0.034 [+0.030, +0.039]**	+0.035 [+0.030, +0.039]**	+0.034 [+0.031, +0.037]**	+0.034 [+0.030, +0.039]**
AR4	+0.058 [+0.055, +0.061]**	+0.058 [+0.055, +0.060]**	+0.058 [+0.056, +0.060]**	+0.059 [+0.056, +0.062]**
AR5	+0.045 [+0.043, +0.047]**	+0.045 [+0.043, +0.047]**	+0.045 [+0.043, +0.046]**	+0.045 [+0.044, +0.047]**
AR6	+0.044 [+0.043, +0.045]**	+0.044 [+0.043, +0.045]**	+0.044 [+0.043, +0.044]**	+0.044 [+0.043, +0.045]**

Note: **Methods.** *Simple mean* averages the three model scores per pair, then averages per AR. *IVW* weights each model score by the inverse of its within-AR variance on that dimension. Pairs with fewer than three valid model scores are dropped, which is why the IVW estimate equals zero in cells where all surviving rows attract the same model-specific weight. *EB shrinkage* shrinks each per-pair simple mean toward the AR-level grand mean by the ratio of within-pair to between-pair variance. *Single-factor* treats the three model scores as noisy measurements of a single latent divergence per pair, recovering the latent score via factor analysis and reporting the per-AR mean. The simple-mean column is the headline used in Section III. All three alternatives reproduce the cumulative cascade pattern within rounding.

Outlet-Balance Robustness

The pooled SPM-to-media composite reported in Section III weights each scored claim pair equally. Because the ten outlets in our panel produce very different volumes of IPCC-adjacent coverage, as Table A1 documents, pair-level weighting gives the high-volume outlets, especially The Guardian on the left and the Wall Street Journal in the early period and the Daily Mail later on the right, more weight in the pooled mean. We report five sample restrictions that probe whether the headline pattern depends on the volume-weighting. The first three restrictions remove the two highest-volume outlets, individually and jointly.

Without WSJ. Removes the right-leaning outlet that dominates the early-AR3 right-side sample while keeping The Guardian on the left.

Without The Guardian. Removes the highest-volume left-leaning outlet while keeping WSJ on the right.

Without WSJ and The Guardian. Pools the remaining eight outlets normally. Removes both highest-volume outlets simultaneously.

The last two restrictions equalize all ten outlets through subsampling or unweighted pooling.

Smallest-outlet cap. Within each Assessment Report and Working Group window, subsample each outlet's articles to the count of the smallest outlet in that window. Equalizes the number of articles each outlet contributes per window.

Per-outlet $1/N$ *pooling.* Compute each outlet's per-AR composite mean at the pair level, then take the unweighted mean of the outlet means. Each outlet receives $1/N$ weight regardless of article count, where N is the number of outlets with usable data in that AR.

We replicate three diagnostics from the main text under each restriction. Figure C1 replicates the cumulative cascade trajectory of Figure 1 Panel A. Figure C2 replicates the outlet-ideology panel of Figure 3 Panel B. Figure C3 replicates the per-LLM cumulative cascade of Figure 4 Panel A. In each figure, the upper row of three panels shows the outlet-removal restrictions and the lower row of two panels shows the equalization restrictions. Each panel reports cumulative composite divergence at the three pipeline nodes, namely the TS, the SPM, and media coverage, under one sample restriction.

The cascade pattern, the directional reweighting at the SPM-to-media stage from AR3 onward, and the cross-outlet similarity in amplification by AR6 all survive every restriction. The smallest-outlet cap strengthens the AR3-AR4 reweighting because it removes The Guardian's diluting effect on the pooled mean. The per-LLM diagnostic also reproduces the cascade shape under every restriction, confirming that the headline pattern is not an artifact of any single LLM interacting with any single high-volume outlet.

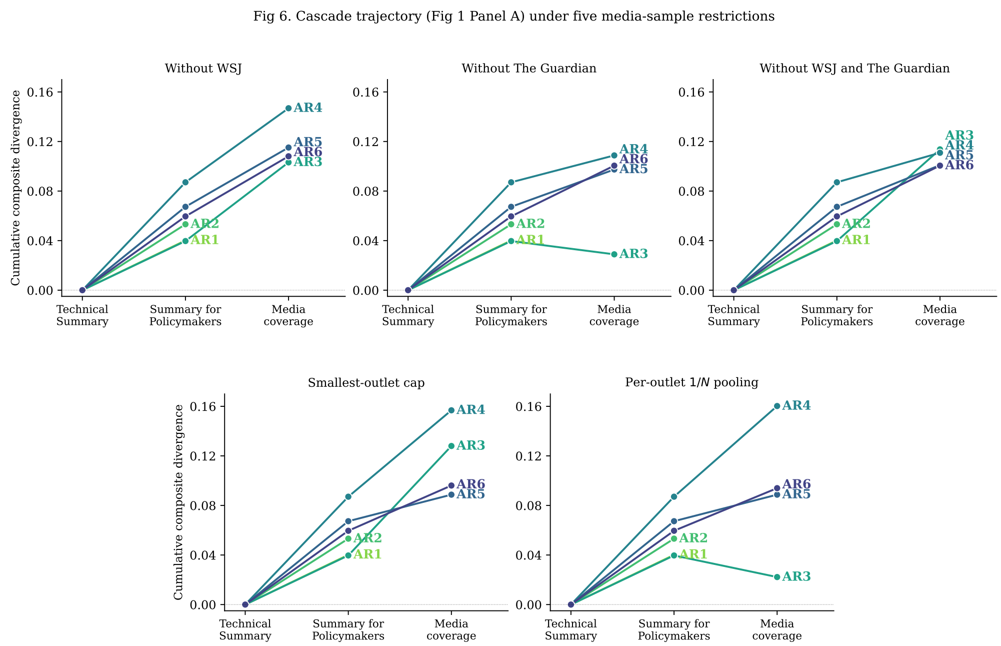

Figure C1. Cascade trajectory under five sample restrictions.

Note: One polyline per Assessment Report. AR1 and AR2 polylines truncate at the SPM node because the media stage covers AR3 through AR6 only. Upper row: outlet-removal restrictions. Lower row: equalization restrictions. The cumulative reweighting pattern from AR3 onward survives every restriction.

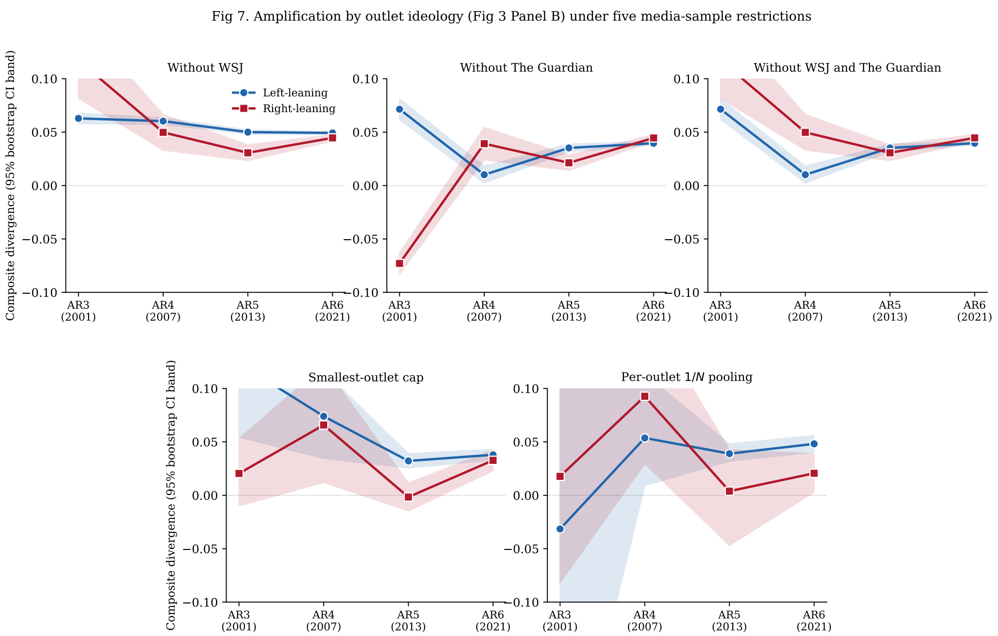

FIGURE C2. AMPLIFICATION BY OUTLET IDEOLOGY UNDER FIVE SAMPLE RESTRICTIONS.

Note: Mean SPM-to-media composite divergence by ideology group and Assessment Report (AR3 through AR6), with 95% bootstrap CI bands. Left- and right-leaning outlets sit at similar amplification levels under every restriction.

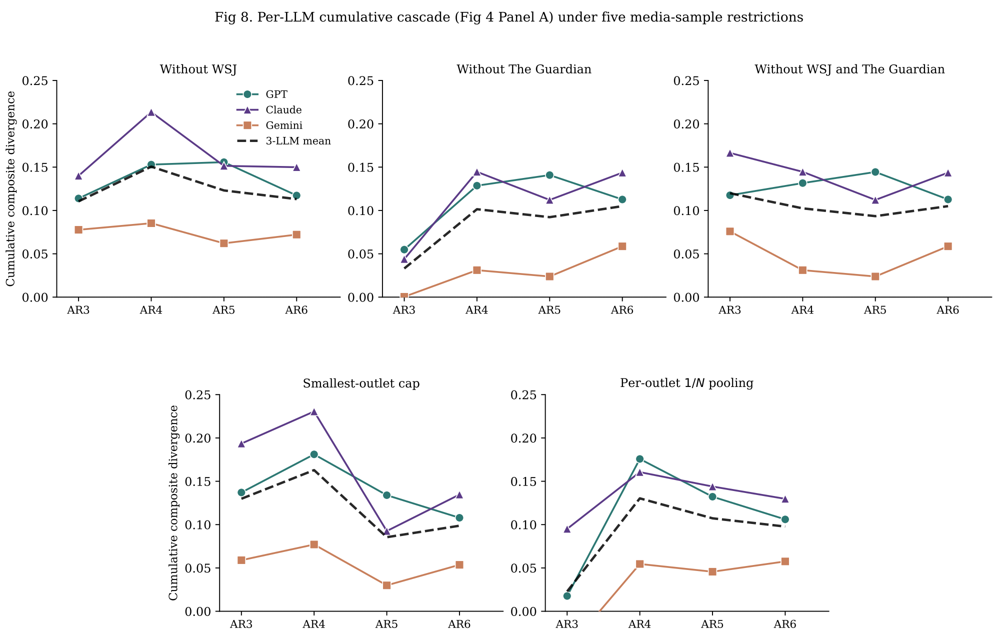

FIGURE C3. PER-LLM CUMULATIVE CASCADE UNDER FIVE SAMPLE RESTRICTIONS.

Note: Cumulative composite divergence (TS-to-SPM plus SPM-to-media) by Assessment Report, plotted separately for each LLM alongside the three-LLM mean. The per-model ordering (Gemini below the ensemble, GPT and Claude above) is preserved under every restriction.

Divergence Examples

This section lists representative claim pairs across the three divergence dimensions and the full directional range, namely amplification scoring +, neutral transmission scoring 0, and attenuation scoring −. Table 1 reports six pairs that are uniformly close to maximum amplification on the focal dimension. The pairs below are drawn from a wider score range and are intended to give the reader a feel for how the dimensions behave at borderline and negative values.

TS-to-SPM examples

Table D1 reports nine TS-to-SPM claim pairs spanning the three divergence dimensions and the three directional cases of amplification, neutral, and attenuation. The pairs are drawn from a mix of Assessment Reports and Working Groups.

SPM-to-media examples

Tables D2, D3, and D4 together report nine SPM-to-media claim pairs spanning the three divergence dimensions and the three directional cases, with one table per dimension and one row per directional case. The pairs are drawn from outlets across the ideological spectrum and from a mix of Assessment Reports.

Table D1—Representative TS-to-SPM claim pairs across the three divergence dimensions and the full directional range.

Dimension	Score	AR/WG & Topic	SPM excerpt
Severity shift	+0.50	AR1 WG1 radiative forcing	SPM (S1.1): "there is a natural greenhouse effect which already keeps the Earth warmer than it would otherwise be; emissions resulting from human activities are substantially increasing the atmospheric concentrations of the greenhouse gases."
Severity shift	+0.00	AR2 WG1 radiative forcing	SPM (S1.5): "Many greenhouse gases remain in the atmosphere for a long time (for CO_2 and N_2O, many decades to centuries), hence they affect radiative forcing on long time-scales."
Severity shift	−0.50	AR3 WG1 regional	SPM (S7.1): "Complex physically-based climate models are required to provide detailed estimates of feedbacks and of regional features. Such models cannot yet simulate all aspects of climate."
Scenario salience	+0.34	AR6 WG1 precipitation	SPM (B.3.4): "A projected southward shift and intensification of Southern Hemisphere summer mid-latitude storm tracks and associated precipitation is likely in the long term under high GHG emissions scenarios (SSP3-7.0, SSP5-8.5)."
Scenario salience	+0.00	AR4 WG3 mitigation	SPM (B.3): "Both bottom-up and top-down studies indicate that there is high agreement and much evidence of substantial economic potential for the mitigation of global GHG emissions over the coming decades, that could offset the projected growth of global emissions or reduce emissions below current levels."
Scenario salience	−0.56	AR6 WG1 extremes	SPM (B.2.3): "Some mid-latitude and semi-arid regions, and the South American Monsoon region, are projected to see the highest increase in the temperature of the hottest days, at about 1.5 to 2 times the rate of global warming."
Uncertainty compression	+0.66	AR6 WG1 radiative forcing	SPM (A.4.1): "Human-caused radiative forcing of 2.72 [1.96 to 3.48] W m^{-2} in 2019 relative to 1750 has warmed the climate system."
Uncertainty compression	+0.01	AR5 WG3 mitigation	SPM (SPM.4.2.3.6): "Technical and behavioural mitigation measures for all transport modes ... could reduce final energy demand in 2050 by around 40% below the baseline (*robust evidence, medium agreement*)."
Uncertainty compression	−0.50	AR4 WG3 mitigation	SPM (C.10.8): "By 2030, about 30% of the projected GHG emissions in the building sector can be avoided with net economic benefit, but their effect may be counteracted by growth in the sector."

Note: Each row reports a single TS-to-SPM claim pair and its per-pair composite score on the focal dimension. The rows span the full directional range, from amplification (+0.66) through neutral (+0.00) to attenuation (−0.56), complementing the amplification-only Panel A of Table 1.

Table D2—Representative SPM-to-media claim pairs: severity-shift dimension.

Score	Outlet & AR	Media excerpt	SPM context	Why it scores this way
+0.50	Daily Mail AR6	"Australia's Great Barrier Reef ... has suffered severe bleaching to an estimated 80 per cent of its corals since 2016."	"It is virtually certain that the global upper ocean has warmed since the 1970s. There is high confidence that oxygen levels have dropped in many upper ocean regions ..."	The 80% figure is presented with a strongly alarming framing and little hedging compared to the SPM's calibrated language.
+0.00	NYT AR3	"Unlike the simple simulations, the newer models include the circulation of the ocean currents."	"Understanding of climate processes and their incorporation in climate models have improved, including water vapour, sea-ice dynamics, and ocean heat transport."	A faithful, factual restatement of the SPM point that newer models include ocean circulation.
−0.50	BBC AR4	"Cliff erosion cannot be stopped, it is a natural process ... the speed of erosion and re-depositing of eroded material on the east coast has been relatively constant."	"For 1993 to 2003, the sum of the climate contributions is consistent within uncertainties with the total sea level rise that is directly observed."	The "relatively constant over thousands of years" framing downplays recent climate-driven sea-level changes the SPM documents.

Table D3—Representative SPM-to-media claim pairs: scenario-salience dimension.

Score	Outlet & AR	Media excerpt	SPM context	Why it scores this way
+0.50	Guardian AR5	"Focusing on the emissions scenario that the scientists say gives the greatest probability of limiting warming to two degrees, the rate of progress to decarbonise the global economy falls far short ..."	"A lower warming target, or a higher likelihood of remaining below a specific warming target, will require lower cumulative CO_2 emissions."	Highlights shortfall against the 2°C-consistent benchmark, slightly amplifying urgency by foregrounding that scenario.
+0.00	BBC AR4	"The tower's been sitting around 10 metres from the cliff edge ever since."	"Global average sea level rose at an average rate of 1.8 mm/yr."	A local factual observation about cliff proximity is neutral relative to the SPM statement on observed sea-level rise.
−0.50	NYT AR6	"... underscores the catastrophic consequences of global warming while also offering some hope that some coral reefs can be saved if humans move quickly to rein in greenhouse-gas emissions."	"Past GHG emissions since 1750 have committed the global ocean to future warming. Likely ocean warming ranges from 2–4 (SSP1-2.6) to 4–8 times (SSP5-8.5) the 1971–2018 change."	The "catastrophic but with hope" framing leans optimistic about mitigation benefits relative to SPM's commitment-to-warming language.

Table D4—Representative SPM-to-media claim pairs: uncertainty-compression dimension.

Score	**Outlet & AR**	**Media excerpt**	**SPM context**	**Why it scores this way**
+0.50	Daily Mail AR4	"The recent IPCC report predicted a half-metre rise in sea levels by 2100 . . . "	"Model-based projections of global average sea level rise at the end of the 21st century are shown in Table SPM.3. For each scenario, the midpoint of the range is within 10% of the model average."	A single "half-metre by 2100" figure is within the SPM range but drops the SPM's range/hedging, presenting one number as definitive.
+0.00	Daily Mail AR6	"He asked them to consider how much of the waste had been 'blown wider afield' . . . "	"It is virtually certain that the global upper ocean has warmed since the 1970s . . . "	The article is about litter blown into the sea, not a climate-science claim, so it neither amplifies nor downplays the SPM ocean findings.
−0.50	WSJ AR3	"No European nation, save Romania, has ratified the treaty and none is likely to."	"The SRES scenarios do not include additional climate initiatives, which means that no scenarios are included that explicitly assume implementation of the Kyoto Protocol . . . "	The near-total non-ratification claim minimizes global policy engagement and adds doubt relative to the SPM's neutral scenario framing.

Note: Each row of Tables D2, D3, and D4 reports a single SPM-to-media claim pair and the per-pair composite score on its focal dimension. Outlet labels follow the panel of ten outlets used throughout the paper. The rows span the full directional range, from amplification (+0.50) through neutral (0.00) to attenuation (−0.50), complementing the amplification-only Panel B of Table 1.

LLM Prompt Templates

The full prompt suite used in Section II.B is reproduced here. There are five distinct prompts. The first three are used for claim extraction, one each for the TS, the SPM, and newspaper articles. The remaining two are used for divergence scoring, one for the TS-to-SPM stage and one for the SPM-to-media stage. Each prompt is provided to the LLMs exactly as written below, with the document text inserted at the marked position.

Prompt 1: Technical Summary claim extraction

```
Extract ALL distinct scientific findings from this section of the IPCC
Technical Summary.

A "finding" is a statement that makes a factual or probabilistic claim about
the climate system, its changes, projections, or drivers. Each finding
should be a self-contained assertion (1-3 sentences).

SECTION: {section_id}
TEXT:
---
{section_text}
---

For EACH finding, return a JSON object with:
- "claim_id": Generate as "{section_id}.finding_{N}" where N starts at 1
- "claim_text": The exact text of the finding (may span 1-3 sentences)
- "confidence_level": One of ["very high confidence", "high confidence",
  "medium confidence", "low confidence", "very low confidence"] or null
- "likelihood_terms": Array of {"term": "...", "probability_range": "...",
  "applies_to": "..."}
- "scenario_references": Array of SSP scenarios mentioned. Empty if none.
- "quantitative_ranges": Array of {"metric": "...", "value": "...",
  "unit": "...", "range_low": number or null, "range_high": number or null,
  "range_type": "..."}
- "topic_category": One of: "temperature", "precipitation", "sea_level",
  "ice_cryosphere", "ocean", "extremes", "carbon_cycle", "radiative_forcing",
  "water_cycle", "regional", "mitigation"
- "chapter_references": Array of underlying chapter references cited in/near
  this finding
- "is_summary_block": true if this comes from a bold summary statement at the
  start of a section, false otherwise

Return a JSON object: {"findings": [array of finding objects]}
```

```
Return ONLY the JSON, no other text.
```

Prompt 2: Summary for Policymakers claim extraction

```
Extract structured information from this IPCC SPM claim.

CLAIM TEXT:
---
{claim_text}
---

Return a JSON object with EXACTLY these fields:
- "claim_id": The claim identifier (e.g., "A.1.1")
- "claim_text": The full text of the finding, stripped of footnote markers
  like [6] [7]
- "confidence_level": One of ["very high confidence", "high confidence",
  "medium confidence", "low confidence", "very low confidence"] or null if
  none stated
- "likelihood_terms": Array of objects, each: {"term": "...",
  "probability_range": "...", "applies_to": "brief description of what the
  term qualifies"}. Use exact IPCC terms: virtually certain, extremely
  likely, very likely, likely, more likely than not, about as likely as not,
  unlikely, very unlikely, extremely unlikely, exceptionally unlikely.
- "scenario_references": Array of SSP scenarios mentioned (e.g.,
  ["SSP1-1.9", "SSP5-8.5"]). Empty array if none.
- "quantitative_ranges": Array of objects, each: {"metric": "what is
  measured", "value": "central or best estimate as string", "unit": "unit of
  measurement", "range_low": number or null, "range_high": number or null,
  "range_type": "very_likely" or "likely" or "best_estimate" or null}
- "topic_category": Primary topic, one of: "temperature", "precipitation",
  "sea_level", "ice_cryosphere", "ocean", "extremes", "carbon_cycle",
  "radiative_forcing", "water_cycle", "regional", "mitigation"
- "chapter_references": Array of chapter/TS section references from the curly
  braces (e.g., ["2.3", "5.2", "TS.2.2", "Box TS.5"])

Return ONLY the JSON object, no other text.
```

Prompt 3: Newspaper-article claim extraction

```
Extract ALL distinct climate science claims from this news article.

A "claim" is any statement the article makes about climate change, its
impacts, projections, or policy responses that relates to physical science
findings (temperature, sea level, ice, extremes, etc.).
```

```
Ignore: author opinions, political commentary, quotes about policy
preferences. Focus on: scientific assertions, data cited, projections
mentioned, characterizations of IPCC findings.

ARTICLE SOURCE: {source}
HEADLINE: {headline}
DATE: {pub_date}
TEXT:
---
{article_text}
---

For EACH claim, return a JSON object with:
- "claim_id": Generate as "MEDIA_{source_short}_{article_idx}.claim_{N}"
  where N starts at 1
- "claim_text": The exact text making the claim (1-3 sentences)
- "tone": One of ["alarming", "neutral", "reassuring", "skeptical"]
- "uses_ipcc_language": true if it uses calibrated IPCC terms (likely, very
  likely, etc.), false otherwise
- "specificity": One of ["quantitative", "qualitative", "vague"]
- "topic_category": One of: "temperature", "precipitation", "sea_level",
  "ice_cryosphere", "ocean", "extremes", "carbon_cycle", "radiative_forcing",
  "water_cycle", "regional", "mitigation", "general_climate"
- "framing_words": Array of emotive/framing words used (e.g., "catastrophic",
  "devastating", "hoax", "uncertain")

Return a JSON object: {"claims": [array of claim objects]}
Return ONLY the JSON, no other text.
```

Prompt 4: Technical Summary-to-SPM divergence scoring

```
Compare these two related climate findings and assess whether the SPM version
presents a more alarming framing than the Technical Summary version.

SPM CLAIM ({spm_id}):
---
{spm_text}
---

TECHNICAL SUMMARY FINDING(S):
---
{ts_text}
---
```

```
Consider:
1. Does the SPM emphasize the more extreme end of reported ranges?
2. Does the SPM omit caveats, conditions, or lower-impact findings present
   in the TS?
3. Does the SPM use stronger qualitative language (e.g., "dramatic" vs
   "notable")?

Rate on this scale:
-2: SPM significantly downplays severity vs TS
-1: SPM somewhat less alarming than TS
 0: Equivalent framing (faithful summary)
+1: SPM somewhat more alarming than TS
+2: SPM significantly amplifies severity vs TS

Return JSON: {"severity_score": <integer -2 to +2>, "reasoning": "<2-3
sentences explaining your assessment>"}
Return ONLY the JSON.
```

Prompt 5: SPM-to-media divergence scoring (batch)

```
You will compare a batch of news media claims against IPCC SPM claims on the
same topic. For each media claim, assess whether it amplifies or downplays
the SPM finding.

TOPIC: {topic}

IPCC SPM CLAIMS (the scientific benchmark):
{spm_block}

MEDIA CLAIMS TO SCORE:
{media_block}

For EACH media claim, score on THREE dimensions:

1. severity_score: integer from -2 to +2
   -2 = media significantly downplays/dismisses vs SPM
   -1 = media somewhat less alarming than SPM
    0 = equivalent framing (faithful reporting)
   +1 = media somewhat more alarming than SPM
   +2 = media significantly amplifies alarm vs SPM

2. scenario_salience: integer from -2 to +2
   -2 = media cherry-picks best-case scenarios only
```

```
   -1 = media slightly favors optimistic projections
    0 = balanced or no scenario references
   +1 = media favors worst-case scenarios (e.g., high-emission SSP5-8.5)
   +2 = media ONLY mentions catastrophic/worst-case projections, ignoring
        moderate scenarios

3. uncertainty_score: integer from -2 to +2
   -2 = media adds doubt not in SPM ("scientists aren't sure")
   -1 = media somewhat less certain than SPM
    0 = equivalent certainty
   +1 = media drops some hedging from SPM
   +2 = media presents as absolute certainty what SPM hedges

Return JSON: {"scores": [
  {"media_claim_id": "...", "spm_claim_id": "best matching SPM claim ID",
   "severity_score": <int>, "scenario_salience": <int>,
   "uncertainty_score": <int>, "reasoning": "<1 sentence>"}
]}

IMPORTANT: For each media claim, pick the SINGLE best-matching SPM claim and
score against that one.
Return ONLY the JSON.
```