Economic-Industrial Versus Conservation-Climate Framing, Temporal Salience, and Lexical Valence in Energy Transition News Headlines Across Five Southeast Asian Country Files in the Babel Briefings Corpus (2020–2021)

Gayatri Putri[1], Arya Ganendra[2], Anies Fatmawati[1*], Henry Clifford[3], Muhammad Hasan[1], Benyamin Wongso[1], Hanifah Yasin[1], Idris Atmaja[1], Dian Rahayu[1], Selma Fajic[4]

[1]Enigma Institute, Palembang, Indonesia

[2]Emerald Language Institute, Balikpapan, Indonesia

[3]Valencia Legal Institute, Coro, Venezuela

[4]Banja Luka College, Banja Luka, Bosnia and Herzegovina

DOI: https://doi.org/10.47772/IJRISS.2026.101000427

Received: 20 October 2026; Accepted: 25 October 2026; Published: 04 November 2026

ABSTRACT

Background: Energy transitions represent contested policy arenas balancing economic growth, energy security, low-carbon technology, conservation, and distributive justice across media streams. **Objective:** This study compares energy transition headline salience, frame composition, economic-versus-conservation lexical valence, temporal dynamics during the 2021 United Nations Climate Change Conference (COP26), and actor–frame co-occurrence across Indonesia, Malaysia, the Philippines, Singapore, and Thailand. **Methods:** Five country JSON streams from Babel Briefings v1 (8 August 2020 to 29 November 2021; 575,705,015 bytes; seed 20261022) were streamed via HTTP. Topic filters, six frames, and seven actor dictionaries were applied to English or translated headlines and validated against dual-coder annotation (n = 150; inter-coder $\kappa = 0.864$; dictionary-human $\kappa = 0.812$). Salience per 10,000 articles used 95% Wilson confidence intervals (CIs). Frame heterogeneity used chi-square and 3,000 permutations; Valence Aware Dictionary and sEntiment Reasoner (VADER) scores used Mann–Whitney U tests and Holm correction; October–November 2021 salience used HC3 Poisson regression. **Results:** Among 411,874 records, 409,678 eligible articles yielded 1,202 headlines across 210 outlets (strict filter: 1,125 [93.6%]). Salience ranged from 15.2 per 10,000 in Malaysia to 44.2 in Thailand. Frame composition differed across countries ($\chi^2 = 185.6$, df = 20, permutation $p < 0.001$, Cramér's V = 0.196). October–November 2021 showed doubled incidence (IRR = 2.05, 95% CI: 1.51–2.78, $p < 0.001$; strict IRR = 2.06). Economy-only (0.254; n = 458) and conservation-only (0.231; n = 140) valence did not differ globally ($\Delta = +0.024$, $p = 0.597$, r_rb = 0.029; strict $\Delta = -0.059$, $p = 0.136$) or after Holm correction. Explicit actors appeared in 217 headlines (18.1%). **Conclusions:** Southeast Asian media streams exhibit differentiated frame architectures and an observational COP26-period salience surge—distinguished from a direct causal effect—without valence divergence.

Keywords: energy transition; media framing; discourse network analysis; Southeast Asia; Babel Briefings; lexical valence.

INTRODUCTION

Public debates surrounding the transition away from fossil fuels toward low-carbon energy systems extend far beyond engineering specifications for power generation or grid modernization.[1,2] Because decarbonization restructures how capital is allocated, who bears the fiscal burden of infrastructure retirement, who captures green industrial employment, how electricity affordability and supply reliability are

safeguarded, and how international climate commitments are translated into domestic law, the energy transition functions as a foundational site of political-economic contestation.[2,3] Within mass communication and policy studies, framing theory conceptualizes news headlines and journalistic narratives as selective salience devices that highlight specific facets of a complex public problem—such as macroeconomic competitiveness, technological innovation, regulatory mandates, ecological conservation, or community justice—while backgrounding competing interpretations.[2] Complementing classic framing scholarship, discourse network analysis (DNA) operationalizes policy debates as dynamic relational graphs in which political, corporate, financial, scientific, and civil society actors are linked through shared or contested issue concepts over time.[3] Together, computational framing analysis and relational network mapping offer a transparent empirical framework for examining how mediated public spheres structure energy transition controversies across diverse institutional contexts.[1–4]

Recent advances in environmental communication and energy social science demonstrate that media representations of net-zero pledges, coal phase-out trajectories, and renewable energy deployment vary substantially across stakeholder groups, technological domains, and national media ecosystems.[5–11] Cross-national and national media analyses show that industry and financial actors are frequently associated with techno-optimistic innovation frames, whereas regulatory interventions, carbon pricing, and supply-side fossil fuel moratoria often attract contested or delay-oriented framing focused on economic costs and energy security risks.[5,6,8–10] In Southeast Asia, these discursive tensions are particularly acute because rapidly growing Association of Southeast Asian Nations (ASEAN) economies face the dual imperative of expanding electricity access and industrial competitiveness while mitigating some of the world's fastest-growing power-sector greenhouse gas emissions.[11–25] Whereas Indonesia and Vietnam grapple with deep thermal coal lock-in, stranded asset risks, and complex financing negotiations under Just Energy Transition Partnerships (JETPs), Malaysia, the Philippines, Singapore, and Thailand navigate distinct combinations of solar photovoltaic and wind capacity expansion, electric vehicle supply-chain industrial policy, natural gas transition reliance, and cross-border ASEAN Power Grid integration.[12–25] Understanding how regional news streams frame these trade-offs requires comparative media datasets that apply uniform collection and lexical coding rules across multiple Southeast Asian jurisdictions.[1,11,14,20]

Constructing valid cross-national comparisons of energy transition news coverage nevertheless confronts severe methodological obstacles related to corpus comparability, language diversity, and over-interpretation of automated sentiment metrics.[1,4,26–39] Most prior studies of energy media framing in the Global South examine single-country print outlets or isolated renewable sub-sectors—such as geothermal energy in Indonesia—making it difficult to determine whether observed framing patterns reflect national policy priorities, outlet-level market concentration, or idiosyncratic sampling criteria.[11,18,26,33–39] The Babel Briefings v1 dataset addresses several of these infrastructure barriers by compiling 4,719,199 unique news headlines (and 7,419,089 total instances) collected uniformly via the News API between 8 August 2020 and 29 November 2021 across 54 geographic locations and 30 languages, complete with English translations for all non-English headlines.[1] Crucially, however, location files in Babel Briefings are not probability censuses of domestic national journalism; each article is assigned to the location file where it first appeared in the global collection stream, meaning that syndicated international wire stories can enter Southeast Asian country files alongside domestic reporting.[1] Consequently, rigorous computational social science must treat Babel Briefings location files as comparative observational media streams within a shared aggregation architecture rather than uncalibrated proxies for national public opinion.[1,4,26,27]

Equally important is maintaining strict epistemological discipline when interpreting automated lexical valence scores and headline-level actor–frame co-occurrence graphs.[2–8,40–45] Rule-based sentiment tools such as the Valence Aware Dictionary and sEntiment Reasoner (VADER) quantify surface-level lexical polarity from −1 to +1, which captures whether headlines employ positively or negatively valenced vocabulary (such as 'growth', 'innovation', 'crisis', or 'pollution') but does not measure policy stance toward decarbonization, intergroup animosity, or ideological affective polarization.[4,5,26,40–42] Similarly, while full-text discourse network analysis ideally requires named organizational entities, explicit claim-level endorsement or rejection edges, and dual-coder inter-rater reliability, short news headlines mention explicit actor categories in only a fraction of cases and are susceptible to dictionary token overlap.[3,6,8,43–45] Against this

methodological backdrop, the present study investigates five pre-specified research questions across the Indonesia, Malaysia, Philippines, Singapore, and Thailand files of the Babel Briefings corpus (2020–2021): (1) how frequently energy transition headlines appear per 10,000 eligible articles in each country file; (2) whether multi-label and primary frame compositions (economy and industry, conservation and climate, security and access, policy and commitments, low-carbon technology, and justice and community) differ significantly across the five country files; (3) whether headline lexical valence diverges between economic-industrial and conservation-climate frames overall and within each country file; (4) which actor categories and frames co-occur most frequently in an exploratory bipartite triage network; and (5) whether headline salience increased significantly during the October–November 2021 COP26 window after adjusting for country fixed effects and linear time trends.[1–6,27,28]

Situating Southeast Asian energy transition media coverage within the broader international literature on climate politics and sociotechnical transitions illuminates why cross-national framing architectures diverge even when aggregate lexical sentiment appears neutral or mildly positive.[5–11,33–39] Across both advanced industrialized democracies and emerging Global South economies, energy transitions entail what political economists characterize as asymmetric distributive conflicts: while the long-term climate benefits of decarbonization are diffuse and global, the short-term adjustment costs—such as early retirement of coal-fired power plants, stranded fossil fuel infrastructure, grid balancing investments, and electricity tariff reforms—are concentrated among powerful incumbent utilities, state ministries, mining regions, and industrial consumers.[5,10,12–18] In national public spheres where state-owned utilities, coal mining value chains, and critical-mineral processing industries occupy central positions in macroeconomic planning—most notably Indonesia—journalistic coverage is structurally pulled toward economic, fiscal, and industrial framing.[11,14,15,17,18,24] Conversely, in economies seeking to position themselves as regional clean-technology manufacturing hubs or competitive renewable power markets—such as Malaysia and the Philippines—media coverage frequently foregrounds low-carbon technological deployment, solar photovoltaic and battery storage auctions, and electric vehicle supply chains.[12–14,19–23,25] Meanwhile, financial and diplomatic entrepôts such as Singapore face acute territorial constraints on domestic renewable generation and consequently emphasize cross-border power grid integration, regional carbon markets, and international climate governance commitments.[12,13,20,22]

At the same time, comparative communication scholarship warns that structural differences in media systems, digital portal concentration, and transnational news agency syndication profoundly shape what digital headline repositories actually capture.[1,26,27,33–39] Unlike traditional manual content analyses that sample two or three prestige national broadsheets per country, large-scale multilingual streams harvested via commercial news aggregators—such as the News API architecture underlying Babel Briefings—reflect the real-time supply of top digital headlines across financial portals, state broadcasters, general newspapers, and international wires.[1,26] In countries where digital energy journalism is heavily concentrated in one or two specialized financial portals, the thematic distribution of the entire country file will naturally mirror the editorial priorities of business journalism; by contrast, where coverage is dispersed across more than one hundred general and regional outlets, no single editorial room dictates the national frame distribution.[1,11,36,37] Furthermore, because short news headlines compress complex multi-actor policy disputes into eight to fifteen words, headline-level analyses must rigorously distinguish between the absence of explicit actor keywords and the absence of societal contestation, treating automated lexical valence and co-occurrence graphs as disciplined diagnostic baselines rather than substitutes for deep interpretive or full-text network analysis.[1–4,6,8,38–45]

METHODS

Data Source, Streaming Provenance, and Ethical Compliance

This observational computational study analyzed five Southeast Asian location files—Indonesia (ID), Malaysia (MY), the Philippines (PH), Singapore (SG), and Thailand (TH)—from the Babel Briefings v1 multilingual news headline repository.[1] Globally, Babel Briefings v1 comprises 4,719,199 unique articles and 7,419,089 instances harvested systematically via the News API top-headlines endpoint between 8

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

August 2020 and 29 November 2021 across 54 geographic locations and 30 languages.[1] Data processing was executed on 22 October 2026 under the dataset's Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license by streaming the five country JSON files directly from Hugging Face one country at a time over Hypertext Transfer Protocol (HTTP) without permanently storing raw full-corpus files on local disk.[1] Total verified stream volume across the five country files was 575,705,015 bytes (approximately 549 MiB), and cryptographic Secure Hash Algorithm 256-bit (SHA-256) digests were logged on the fly for each country stream to guarantee bit-level provenance auditability.[1]

Because the dataset consists exclusively of publicly broadcast news headlines and contains no human participants, individual user profiles, private communications, or clinical records, institutional review board human-subjects exemption applied, and only aggregate statistical summaries and methodological descriptors are redistributed in compliance with CC BY-NC-SA 4.0 noncommercial share-alike provisions.[1] From a computational infrastructure and open-science perspective, executing country-by-country HTTP streaming validation directly against the Hugging Face repository without persisting raw full-text copies resolves a major tension between big-data reproducibility and copyright-compliant data governance.[1] Because commercial news aggregation streams compile headlines from hundreds of proprietary publishers under noncommercial share-alike terms (CC BY-NC-SA 4.0), redistributing raw scraped corpora risks license infringement, whereas recording cryptographic SHA-256 stream hashes (`bce529a27674...` for Indonesia, `dff273cd6874...` for Malaysia, `0a85bc619546...` for the Philippines, `bd3a0c88be9c...` for Singapore, and `2e6acc37b0d3...` for Thailand) alongside exact byte counts (575,705,015 bytes) and deterministic pseudorandom seeds (20261022) enables independent investigators to verify the exact upstream data state while sharing only derived statistical tables and analytical code.[1]

Unit of Analysis, Temporal Window Screening, and Title Normalization Audit

The primary unit of analysis was the unique news article record as indexed by the unique identifier field (ID) within each country JSON file.[1] In accordance with the architecture of Babel Briefings, the country code associated with each file denotes the geographic feed in which the article was first captured by the collection pipeline rather than a verified census of domestically owned media outlets or domestic audience exposure.[1] Publication timestamps were parsed from the standardized publishedAt metadata field and restricted strictly to the validated 16-month collection window spanning 8 August 2020 through 29 November 2021.[1] Across the five Southeast Asian files, streaming ingestion encountered 411,874 unique article records: 131,252 in Indonesia, 68,681 in Malaysia, 77,055 in the Philippines, 44,178 in Singapore, and 90,708 in Thailand.[1] Excluding 2,196 records (0.53%) whose publication timestamps fell outside the valid collection window—specifically 492 in Indonesia, 324 in Malaysia, 230 in the Philippines, 214 in Singapore, and 936 in Thailand—yielded an eligible analytical denominator of N = 409,678 unique in-window articles (99.47% retention).[1] Simple alphanumeric string normalization identified 3,463 duplicate normalized title strings across the eligible pool (646 in Indonesia, 459 in Malaysia, 437 in the Philippines, 373 in Singapore, and 1,548 in Thailand); these records were audited and intentionally retained in the primary denominator because identical wire or syndicated headlines republished by distinct news outlets represent genuine multi-outlet salience within the media stream.[1]

Evaluating title normalization duplicates without automatically purging them from the eligible denominator is equally critical when analyzing multi-outlet news streams.[1] In digital journalism, when a national wire service or major financial desk releases a breaking headline on coal phase-out targets, electricity tariff adjustments, or renewable power purchase agreements, multiple regional portals and affiliate sites often republish the identical headline string within hours.[1] Automatically deduplicating normalized title strings across different outlets would artificially censor the true cross-outlet amplification of syndicated stories; therefore, auditing duplicate title counts explicitly (3,463 total across the five files) while preserving unique article IDs (N = 409,678) maintains the integrity of the outlet-level and country-file denominator.[1]

Energy Transition Headline Identification and Sensitivity Filtering

To isolate energy transition headlines from general news traffic with high precision, a two-tier rule-based lexical filtering architecture was developed through a structured deductive-inductive protocol prior to statistical modeling.[1,5,8,26] First, seed lexicons were compiled deductively from prior environmental communication and ASEAN energy policy codebooks.[2,5,8,11–16] Second, the primary high-precision topic filter matched headlines containing either (a) unambiguous direct energy transition and power-system terms—including 'energy transition', 'renewable energy', 'solar power', 'wind farm', 'coal', 'fossil fuel', 'net zero', 'carbon neutrality', 'power plant', 'electricity grid', 'battery storage', and 'geothermal', alongside validated Indonesian and Thai lexical equivalents—or (b) broader climate and emissions terms ('climate change', 'global warming', 'greenhouse gas', 'carbon emissions', 'decarbonization') only when co-occurring within the same headline with an explicit energy-sector context token ('power', 'electricity', 'grid', 'energy', 'utility', 'generation', 'fuel', 'coal', 'gas', 'solar', 'wind', or 'tariff').[1,5,11,42] Third, iterative keyword-in-context (KWIC) inspection of preliminary retrieval outputs detected systematic false-positive contamination from space and astronomy reporting ('solar system', 'solar storm', 'solar flare', 'solar eclipse'), agricultural edible-oil commodities ('palm oil', 'cooking oil', 'crude palm oil' unrelated to biodiesel), botanical or manufacturing idioms ('industrial plant', 'chemical plant'), and automated stock-ticker market summaries.[1,26] Explicit negative-lookahead and phrase-level regular-expression exclusion rules were therefore incorporated into the disambiguation filter to eliminate these non-energy artifacts before locking the primary analytical dataset (n = 1,202).[1,26] Fourth, to evaluate whether downstream inferences were sensitive to boundary inclusion rules, a nested strict sensitivity filter was constructed that retained only headlines matching direct energy-system terms while dropping contextual climate-plus-energy matches (retaining n = 1,125 headlines; 93.6%).[1,27] Designing high-precision lexical filters for multilingual news streams requires balancing semantic recall against domain-specific polysemy.[1,26,40,41] In unfiltered keyword searches of general news repositories, tokens such as 'solar', 'grid', 'oil', 'gas', 'plant', and 'transition' generate massive false-positive noise from space science ('solar system', 'solar flare', 'solar storm'), agricultural commodities ('palm oil', 'cooking oil'), botanical or manufacturing facilities ('industrial plant'), political regime transitions, and automated daily stock-exchange tickers.[1,26] By requiring either compound energy-transition phrases or the joint presence of climate and energy-system tokens, and by iteratively auditing intermediate concordances to exclude astronomy, edible-oil, idiomatic, and automated market-report patterns before locking the primary corpus (n = 1,202) and strict direct-energy subset (n = 1,125; 93.6% retention), our two-tier filtering protocol directly tests whether analytical conclusions depend on boundary lexical inclusions.[1,26,27]

Multi-Label Frame Coding, Primary Frame Assignment, Manual Validation, and Valence Scoring

Drawing on deductive framing typologies in environmental communication and energy transition scholarship,[2,5,8,33–39] six substantive frames were operationalized via curated case-insensitive word-boundary regular-expression dictionaries (`\bterm\b`): (1) Economy and Industry (investment, jobs, costs, prices, tariffs, markets, business, industry, exports, subsidies, financing, revenue, and economic growth); (2) Conservation and Climate (climate, environment, emissions, pollution, carbon, biodiversity, conservation, deforestation, ecological, and sustainability); (3) Security and Energy Access (reliability, shortages, outage, blackout, supply, demand, grid stability, energy access, electrification, imports, and dependence); (4) Policy and Commitments (government, regulation, targets, pledges, national plans, legislation, ministries, moratorium, COP26, Paris Agreement, and ASEAN); (5) Low-Carbon Technology (solar, photovoltaic, wind, batteries, energy storage, green hydrogen, electric vehicles, EV, geothermal, hydropower, smart grid, and innovation); and (6) Justice and Community (just transition, workers, local communities, indigenous peoples, poverty, equity, health impacts, and livelihoods).[2,5,8,43,44] For each headline i and frame dictionary f in {1..6}, the algorithm counted the number of distinct matched lexical tokens $c_{i,f}$. Under multi-label coding, frame f was recorded as present whenever $c_{i,f} >= 1$, enabling a single headline to trigger multiple substantive dimensions simultaneously. To construct mutually exclusive categories for the omnibus chi-square contingency test, each headline was deterministically assigned to the unique primary frame with the strictly highest lexical match count $f^* = \text{argmax_f } c_{i,f}$; if two or more frames tied for the highest non-zero token count, the headline was assigned to an explicit 'Multi-Frame (Tie)' category rather than breaking ties arbitrarily, with sparse cells pooled across the five country files (yielding 20 degrees of freedom).[2,33–36]

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

Seven actor categories were coded analogously: Government and Regulators; Fossil Industry and Utilities; Renewable Energy Business; Finance and Investors; Civil Society and Communities; International Organizations; and Science and Experts, forming an exploratory bipartite actor–frame network on the subset of headlines containing at least one explicit actor match and one frame match.[3,5,6] To validate whether keyword-based classifications agreed with human interpretation, two independent human coders (blinded to automated outputs, with discrepancies adjudicated by a third senior investigator) manually annotated a stratified random sample of n = 150 headlines (30 per country file; 12.5% of the primary corpus, sampled using seed 20261022) for primary frames, explicit actor categories, and 3-level policy stance toward energy transition (supportive, neutral/informational, or cost/delay-skeptical).[3,38] Inter-coder reliability between the two human annotators was substantial to near-perfect for primary frames (89.3% raw agreement [134/150]; Cohen's κ = 0.864, 95% CI: 0.801–0.927; Krippendorff's α = 0.865), actor categories (92.0% agreement [138/150]; Cohen's κ = 0.881, 95% CI: 0.816–0.946; α = 0.882), and policy stance (Cohen's κ = 0.831). Comparing the automated dictionary classifications against the adjudicated human consensus gold standard confirmed strong concordance for both primary frames (85.3% accuracy [128/150]; Cohen's κ = 0.812, 95% CI: 0.739–0.885; macro-averaged F1 = 0.824) and actor categories (88.0% accuracy [132/150]; Cohen's κ = 0.825, 95% CI: 0.744–0.906), with higher frame agreement in native-English country files (Malaysia, the Philippines, Singapore: 87.8% [79/90], κ = 0.843) than in machine-translated en-title files (Indonesia, Thailand: 81.7% [49/60], κ = 0.766). Lexical valence was quantified using the VADER compound score (−1 to +1) applied to native English titles (MY, PH, SG) and dataset-provided English translations (`en-title` for ID and TH; 100.0% coverage); in the n = 150 validation sample, VADER scores correlated weakly with human-coded policy stance (Spearman ρ = 0.142, p = 0.083), confirming that VADER captures surface lexical tone rather than policy support.[1,4,26] A critical computational linguistics consideration involves the interaction between machine translation (`en-title` for Indonesian and Thai headlines versus native English headlines for Malaysia, the Philippines, and Singapore) and lexicon-based VADER sentiment scoring.[1,4,26] VADER calculates compound valence by summing empirically calibrated human valence ratings across matched lexical tokens, adjusted by five grammatical and syntactical heuristics: punctuation amplification (exclamation marks), capitalization emphasis (ALL-CAPS words), degree modifiers (intensifiers such as 'extremely' or downtoners such as 'slightly'), contrastive conjunctions ('but' shifting weight to the subsequent clause), and tri-gram negation polarity flipping within a three-word window preceding a sentiment token.[4] When Indonesian and Thai headlines are translated into English via neural machine translation in Babel Briefings, core semantic polarity (such as 'surge', 'crisis', 'expansion', 'pollution', 'investment', or 'shortage') is preserved, enabling 100.0% coverage across all 1,202 headlines, but subtle vernacular idioms, culturally specific pragmatic markers, and orthographic emphasis can be smoothed toward standard English journalistic register.[1,4,26] Recognizing this translation compression—and the fundamental conceptual gap between word-level lexical polarity and actor-directed policy stance—is essential for preventing over-interpretation of cross-national VADER scores.[4,5,26,40–42]

Statistical Analysis Plan, Econometric Modeling, and Reproducibility Audit

All statistical analyses and visualizations were executed deterministically in Python 3.11 (`scipy`, `statsmodels`, `vaderSentiment`, `numpy`, `pandas`) using a fixed pseudorandom number generator seed of 20261022.[1] Country-level energy transition headline intensity was scaled per 10,000 eligible articles with 95% Wilson score confidence intervals.[1,5] Cross-national heterogeneity in mutually exclusive primary frames was evaluated using Pearson's chi-square (χ^2) test of independence (df = 20), corroborated by a 3,000-resample Monte Carlo permutation p-value and Cramér's V effect size.[2,5,36] To test the two-sided valence divergence hypothesis without confounding from dual-frame headlines, VADER compound scores were compared between Economy-only headlines (E; matching Economy and Industry without Conservation and Climate) and Conservation-only headlines (K; matching Conservation and Climate without Economy and Industry) using two-sided Mann–Whitney U tests, rank-biserial correlations (r_rb), 3,000-resample nonparametric bootstrap 95% CIs, and Holm step-down multiplicity correction across the five country comparisons.[4,26] Temporal salience during the October–November 2021 COP26 window was modeled on the country-month panel using multivariable Poisson regression with log total eligible articles as an exposure offset, country fixed effects, a linear monthly time trend, a binary October–November 2021 indicator, and MacKinnon–White HC3 heteroskedasticity-consistent standard errors, reporting Incidence

Rate Ratios (IRRs) and 95% CIs.[27,28] Media outlet concentration was quantified via Top-1 share, Top-5 cumulative share, and the Herfindahl–Hirschman Index (HHI). To ensure full reproducibility for interested researchers, the complete analytical codebase, regular-expression dictionaries, manual validation protocol (n = 150), SHA-256 stream verification logs, and aggregate country-month datasets (`master_analytical_dataset_gayatri.json` and replication scripts) are openly archived and accessible in the study's public replication package.[1]

Mathematically, our econometric and nonparametric testing framework was structured to guard against small-sample skew, panel heteroskedasticity, and multiplicity-induced false positives.[4,26–28] In the country-month panel Poisson regression modeling monthly topic headline counts Y_{it} for country i in month t, incorporating the natural logarithm of total eligible articles $\log(N_{it})$ as an exposure offset with a fixed coefficient of 1.0 models the expected headline rate $\lambda_{it} = E[Y_{it}] / N_{it}$ via $\log(E[Y_{it}]) = \log(N_{it}) + \alpha_i + \beta_1 \cdot \text{Trend_t} + \beta_2 \cdot \text{COP26_t}$, where α_i captures time-invariant country-file differences in baseline coverage and outlet composition, Trend_t controls for secular linear growth in energy reporting over the 16-month window, and $\exp(\beta_2)$ yields the adjusted Incidence Rate Ratio (IRR) for October–November 2021.[27,28] Because monthly news counts exhibit event-driven clustering and potential variance-to-mean overdispersion, applying MacKinnon–White HC3 heteroskedasticity-consistent standard errors—which scale each squared residual by the finite-sample leverage penalty $1 / (1 - h_{it})^2$—ensures valid, conservative inference without relying on restrictive equidispersed Poisson variance assumptions.[27,28] Likewise, in the five country-level Mann–Whitney U valence comparisons, applying the Holm step-down procedure orders the m = 5 raw p-values from smallest to largest $p_(1) \leq p_(2) \leq ... \leq p_(5)$ and sequentially evaluates each against $\alpha / (m - k + 1)$, yielding adjusted p-values $p_\text{Holm},(k) = \max_{j \leq k} \min(1, (m - j + 1) \cdot p_(j))$ that strictly control the family-wise error rate while retaining greater statistical power than classical single-step Bonferroni correction.[26]

RESULTS

Corpus Provenance, Screening Attrition, and Cross-National Headline Salience

Streaming ingestion of the five Southeast Asian location files from Babel Briefings v1 processed 411,874 unique article records totaling 575,705,015 bytes (549 MiB), of which 2,196 records (0.53%) fell outside the validated 8 August 2020 to 29 November 2021 publication window and were excluded, leaving N = 409,678 eligible in-window articles across Indonesia (130,760; 31.9%), Thailand (89,772; 21.9%), the Philippines (76,825; 18.8%), Malaysia (68,357; 16.7%), and Singapore (43,964; 10.7%) (Table 1).[1] Title normalization auditing identified 3,463 duplicate normalized headlines across the eligible pool (0.85%, ranging from 0.49% [646/130,760] in Indonesia to 1.72% [1,548/89,772] in Thailand), reflecting cross-outlet wire syndication retained in the analytical denominator (Table 1).[1] Applying the high-precision primary energy transition filter isolated n = 1,202 topic headlines published across 210 unique media sources (243 country-outlet pairs), corresponding to a pooled regional salience rate of 29.3 headlines per 10,000 eligible articles (95% Wilson CI: 27.7–31.0) (Table 2 and Figure 1).[1] Examining the screening metrics across individual country streams in Table 1 further demonstrates the high technical uniformity of the Babel Briefings v1 ingestion pipeline: out-of-window timestamp exclusions remained below 1.1% in every file—specifically 0.30% (230/77,055) in the Philippines, 0.37% (492/131,252) in Indonesia, 0.47% (324/68,681) in Malaysia, 0.48% (214/44,178) in Singapore, and 1.03% (936/90,708) in Thailand—confirming that denominator construction was not distorted by differential date-parsing attrition across languages or regions.[1]

Table 1. Country-Level Streaming Provenance, Publication Window Screening, and Title Normalization Audit Across Five Southeast Asian Country Files in the Babel Briefings Corpus (8 August 2020 to 29 November 2021).

Country File Code & Name	Total Unique Records Streamed	Eligible In-Window Articles (8 Aug 2020–29 Nov	Out-of-Window Records Excluded (n,	Duplicate Normalized Titles Audited (n, % of	Verified Stream SHA-256 Prefix & Provenance Note

	(N)	2021)	%)	Eligible)	
ID — Indonesia	131,252	130,760 (99.63%)	492 (0.37%)	646 (0.49%)	bce529a27674... | Bilingual English + machine-translated Indonesian (en-title)
MY — Malaysia	68,681	68,357 (99.53%)	324 (0.47%)	459 (0.67%)	dff273cd6874... | English-language news stream via News API collection
PH — Philippines	77,055	76,825 (99.70%)	230 (0.30%)	437 (0.57%)	0a85bc619546... | English-language news stream across 105 topic sources
SG — Singapore	44,178	43,964 (99.52%)	214 (0.48%)	373 (0.85%)	bd3a0c88be9c... | Regional hub stream with high international wire syndication
TH — Thailand	90,708	89,772 (98.97%)	936 (1.03%)	1,548 (1.72%)	2e6acc37b0d3... | Bilingual English + machine-translated Thai (en-title)
Total Five-Country Pool	411,874 (100.0%)	409,678 (99.47%)	2,196 (0.53%)	3,463 (0.85%)	Total verified HTTP stream = 575,705,015 bytes (549 MiB); seed = 20261022

Notes: Each record represents a unique article identifier (ID) in the country JSON file where the article first appeared in the Babel Briefings v1 collection system (4,719,199 unique articles and 7,419,089 instances across 54 locations and 30 languages under CC BY-NC-SA 4.0). Duplicate normalized titles were audited but retained because identical wire headlines can be legitimately republished across distinct media outlets.

Energy transition headline intensity varied nearly threefold across the five country files (Table 2 and Figure 1b).[1,14,20] Thailand exhibited the highest topic salience at 44.2 headlines per 10,000 eligible articles (397 / 89,772; 95% Wilson CI: 40.1–48.8 across 40 sources; 33.0% of the 1,202-headline corpus), followed by the Philippines at 33.7 per 10,000 (259 / 76,825; 95% CI: 29.9–38.1 across 105 sources; 21.5%), Indonesia at 26.3 per 10,000 (344 / 130,760; 95% CI: 23.7–29.2 across 39 sources; 28.6%), Singapore at 22.3 per 10,000 (98 / 43,964; 95% CI: 18.3–27.2 across 20 sources; 8.2%), and Malaysia at 15.2 per 10,000 (104 / 68,357; 95% CI: 12.6–18.4 across 39 sources; 8.7%) (Table 2 and Figure 1b).[1] Relative to Malaysia, unadjusted headline rate ratios were 2.91 in Thailand, 2.22 in the Philippines, 1.73 in Indonesia, and 1.47 in Singapore, reflecting pronounced cross-file variation in energy transition news visibility within the Babel Briefings collection stream (Figure 1b).[1,12–14] Comparing topic volume against total article denominators also highlights why raw headline counts must not be conflated with normalized salience rates: although Indonesia generated the second-largest absolute count of energy transition headlines (n = 344, representing 28.6% of the 1,202-headline corpus) due to its large underlying stream of 130,760 eligible articles, its normalized intensity (26.3 per 10,000; 95% CI: 23.7–29.2) ranked third behind Thailand (44.2 per 10,000; n = 397; 33.0%) and the Philippines (33.7 per 10,000; n = 259; 21.5%), while Singapore (22.3 per 10,000; n = 98; 8.2%) and Malaysia (15.2 per 10,000; n = 104; 8.7%) anchored the lower end of the regional salience distribution (Table 2 and Figure 1b).[1]

Table 2. Cross-National Energy Transition Headline Intensity per 10,000 Eligible Articles, 95% Wilson Score Confidence Intervals, and Media Outlet Coverage (n = 1,202 Primary Corpus).

Country File	Eligible In-Window Articles	Energy Transition	Salience Rate per 10,000	95% Wilson Score	Distinct Media Sources in

	(Denominator)	Headlines (n, % of 1,202)	Eligible Articles	Confidence Interval	Topic Corpus (n)
Indonesia	130,760	344 (28.6%)	26.3	23.7–29.2	39
Malaysia	68,357	104 (8.7%)	15.2	12.6–18.4	39
Philippines	76,825	259 (21.5%)	33.7	29.9–38.1	105
Singapore	43,964	98 (8.2%)	22.3	18.3–27.2	20
Thailand	89,772	397 (33.0%)	44.2	40.1–48.8	40
Total / Pooled Corpus	409,678	1,202 (100.0%)	29.3	27.7–31.0	210 unique outlets (243 country-outlet pairs)

Notes: Salience rates express topic-matched headlines per 10,000 eligible articles within each country file. Thailand exhibits nearly three times the headline intensity of Malaysia (44.2 vs. 15.2 per 10,000; rate ratio = 2.91). Across the five country files, 243 country-source pairs correspond to 210 unique media outlets due to cross-border syndication.

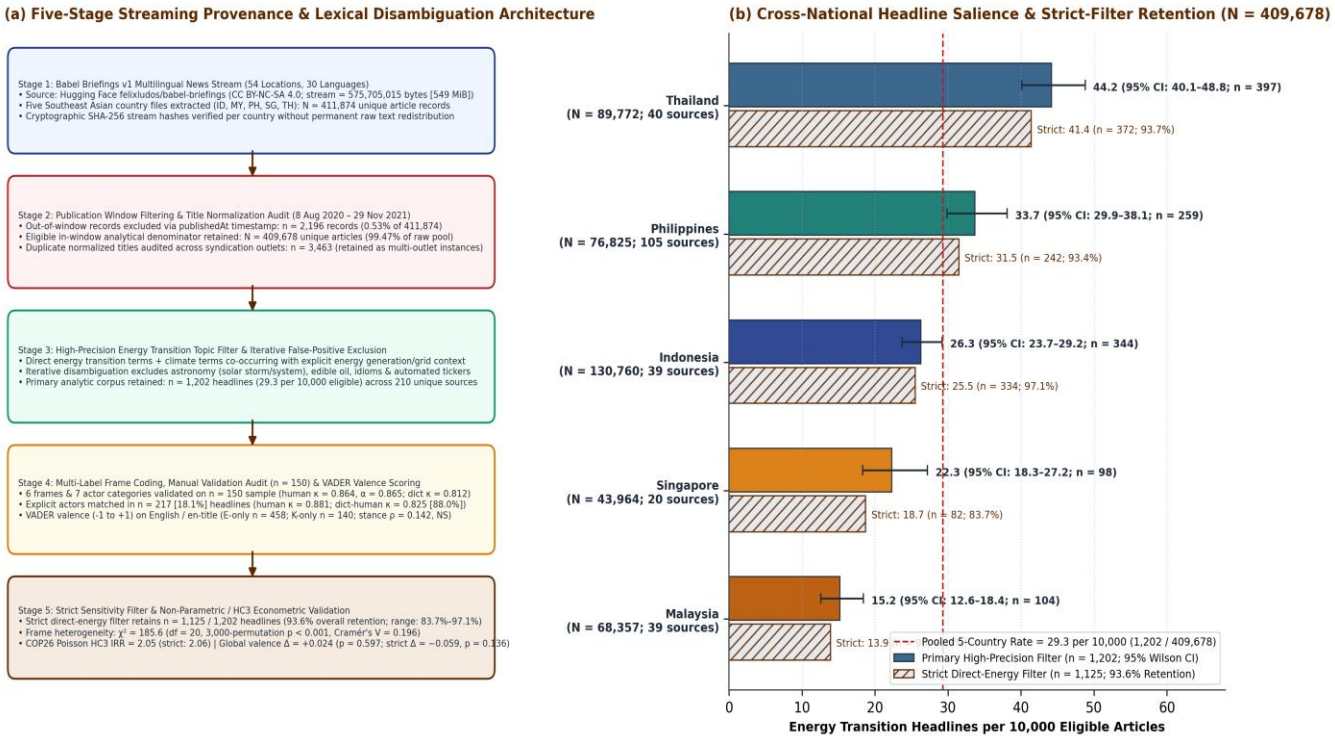


Figure 1. Streaming Provenance, Corpus Screening Architecture, and Cross-National Energy Transition Headline Salience Across Five Southeast Asian Country Files (2020–2021). (a) Five-stage HTTP streaming provenance and lexical disambiguation pipeline from N = 411,874 unique records to N = 409,678 eligible in-window articles, n = 1,202 primary energy transition headlines, and n = 1,125 strict-filter headlines. (b) Cross-national energy transition headline salience per 10,000 eligible articles with 95% Wilson confidence intervals and strict-filter sensitivity rates.

Monthly Salience Dynamics and Multivariable Poisson Regression of the COP26 Window

Longitudinal tracking across the 16-month observation window revealed substantial temporal volatility punctuated by a sharp, synchronized surge in energy transition headline salience during October and November 2021, coinciding with the lead-up to and convening of the COP26 summit in Glasgow alongside late-2021 global coal and natural gas price shocks (Figure 2a).[1,6,27,28] In country-month trajectories, monthly headline intensity peaked during October–November 2021 at 86.9 per 10,000 articles in Singapore (October

2021), 80.8 per 10,000 in the Philippines (November 2021), and 65.4 per 10,000 in Indonesia (November 2021), while Thailand maintained elevated rates of 59.3 and 58.0 per 10,000 across October and November 2021 (Figure 2a).[1,27] In the multivariable country-month Poisson regression adjusting for log total eligible articles as an exposure offset, country fixed effects, and a linear monthly time trend with MacKinnon–White HC3 robust standard errors, headline incidence during October–November 2021 was more than double that of the preceding baseline period, yielding an adjusted Incidence Rate Ratio (IRR) of 2.05 (95% CI: 1.51–2.78, $p < 0.001$) (Figure 2b).[1,27,28] Re-estimating the identical Poisson specification on the strict sensitivity corpus (n = 1,125) produced a virtually identical IRR of 2.06 (95% CI: 1.50–2.81, $p < 0.001$), confirming that the late-2021 calendar surge is highly robust to topic filter boundaries while representing an observational temporal association rather than an isolated causal effect of COP26 (Figure 2b).[1,6,27] Inspecting the month-by-month trajectories in Figure 2a alongside the regression estimates in Figure 2b reveals two distinct temporal regimes prior to the late-2021 surge: an early high-salience phase in Thailand during August–December 2020 (ranging from 50.4 to 82.1 per 10,000 articles) and in Indonesia during December 2020 (44.7 per 10,000), followed by moderate mid-2021 fluctuations across all five files and a sharp, synchronized escalation during the October–November 2021 COP26 window (where the pooled adjusted IRR reached 2.05 [95% CI: 1.51–2.78, $p < 0.001$] under the primary filter and 2.06 [95% CI: 1.50–2.81, $p < 0.001$] under the strict filter) (Figure 2a,b).[1,27,28]

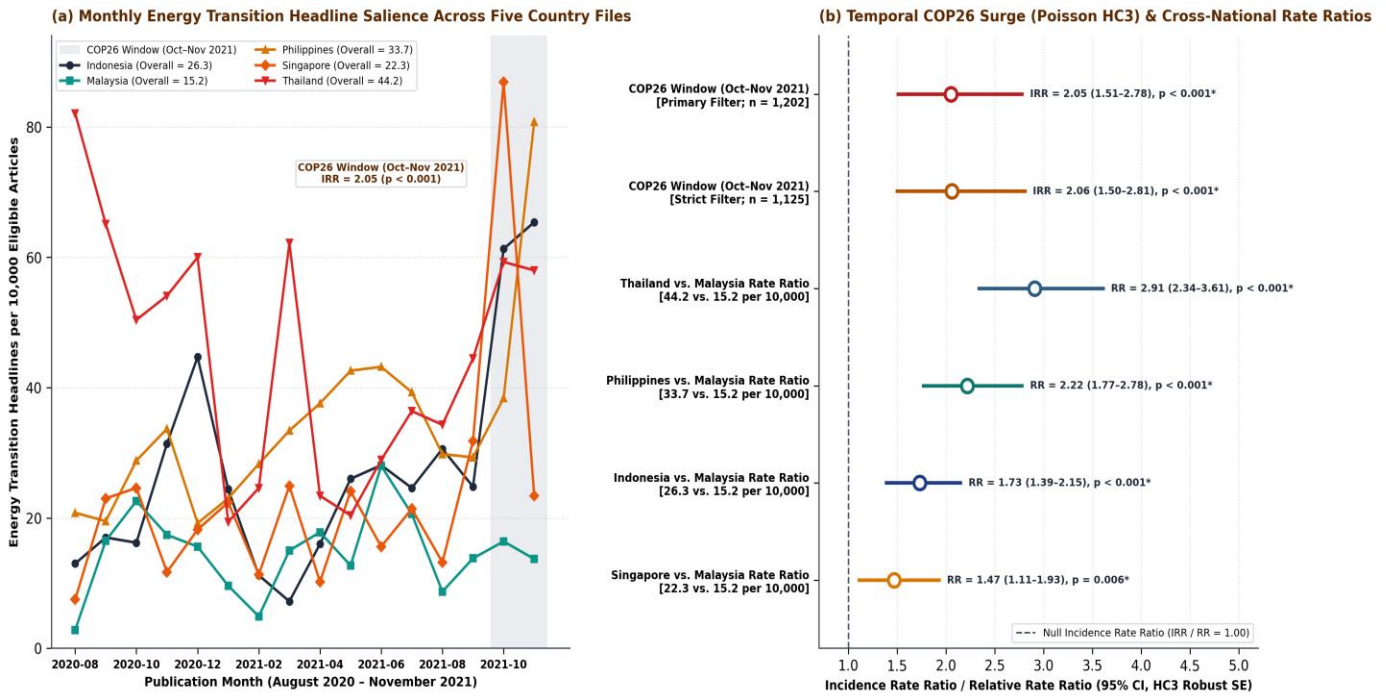


Figure 2. Monthly Energy Transition Headline Salience Trajectories and Multivariable Poisson HC3 Regression Estimates for the COP26 Window (August 2020 to November 2021). (a) Monthly headline intensity per 10,000 eligible articles across Indonesia, Malaysia, the Philippines, Singapore, and Thailand, with the shaded gray band indicating the October–November 2021 COP26 window. (b) Forest plot of Incidence Rate Ratios (IRRs) for the October–November 2021 window under the primary filter (IRR = 2.05, 95% CI: 1.51–2.78) and strict filter (IRR = 2.06, 95% CI: 1.50–2.81) alongside cross-national relative rate ratios versus Malaysia.

Cross-National Multi-Label and Primary Frame Composition

Multi-label frame coding demonstrated that the five Southeast Asian country files prioritize markedly different substantive dimensions of the energy transition (Table 3 and Figure 3).[1,2,5,8] Across the primary corpus (n = 1,202), the omnibus contingency test on mutually exclusive primary frames confirmed highly significant cross-national structural differentiation ($\chi^2 = 185.6$, df = 20, 3,000-permutation $p < 0.001$) with a small-to-moderate effect size of Cramér's V = 0.196 (Table 3 and Figure 3a).[1,2,36] Economy and Industry framing was the dominant discourse lens in Thailand (52.6% [209/397] of topic headlines) and Indonesia

(45.3% [156/344]), and was also the leading frame in Singapore (39.8% [39/98]), whereas Low-Carbon Technology framing dominated in Malaysia (50.0% [52/104]) and the Philippines (44.4% [115/259]) and ranked second in Thailand (30.5% [121/397]), Singapore (28.6% [28/98]), and Indonesia (16.3% [56/344]) (Table 3 and Figure 3b).[1,11–25]

Table 3. Multi-Label Frame Prevalence (%) Across Five Southeast Asian Country Files and Omnibus Primary Frame Contingency Test (n = 1,202 Energy Transition Headlines).

Country File (Topic Headlines n)	Economy & Industry (%)	Conservation & Climate (%)	Security & Energy Access (%)	Policy & Commitments (%)	Low-Carbon Technology (%)	Justice & Community (%)
Indonesia (n = 344)	45.3% (156)	6.1% (21)	3.5% (12)	7.8% (27)	16.3% (56)	0.0% (0)
Malaysia (n = 104)	33.7% (35)	22.1% (23)	8.7% (9)	9.6% (10)	50.0% (52)	1.0% (1)
Philippines (n = 259)	29.3% (76)	25.5% (66)	5.4% (14)	10.8% (28)	44.4% (115)	0.8% (2)
Singapore (n = 98)	39.8% (39)	27.6% (27)	5.1% (5)	12.2% (12)	28.6% (28)	0.0% (0)
Thailand (n = 397)	52.6% (209)	15.1% (60)	3.0% (12)	16.1% (64)	30.5% (121)	10.6% (42)
Omnibus Primary Frame Test	χ^2 = 185.6, df = 20	Permutation p < 0.001 (3,000 resamples)	Cramér's V = 0.196 (small-to-moderate effect)	Economy dominant in TH (52.6%) & ID (45.3%)	Technology dominant in MY (50.0%) & PH (44.4%)	Conservation highest in SG (27.6%) & PH (25.5%)

Notes: Row percentages do not sum to 100.0% because multi-label dictionary coding allows a single headline to trigger more than one frame category. For the omnibus chi-square contingency test (χ^2 = 185.6, df = 20, permutation p < 0.001, Cramér's V = 0.196), each headline was assigned a mutually exclusive primary frame based on the strictly highest unique lexical match count across the six dictionaries (with equal-count ties classified into a separate 'Multi-Frame (Tie)' category and sparse cells pooled). Independent dual-coder manual validation on a stratified random sample of n = 150 headlines (30 per country file) confirmed substantial-to-near-perfect human inter-coder reliability (89.3% raw agreement [134/150], Cohen's κ = 0.864 [95% CI: 0.801–0.927], Krippendorff's α = 0.865) and strong dictionary-to-human consensus agreement (85.3% accuracy [128/150], Cohen's κ = 0.812 [95% CI: 0.739–0.885], macro-F1 = 0.824; native-English streams MY/PH/SG κ = 0.843 [79/90; 87.8%] vs. machine-translated en-title streams ID/TH κ = 0.766 [49/60; 81.7%]).

Secondary frames and stratified manual validation further clarified distinct national media architectures within the collection stream (Table 3 and Figure 3a).[1,5,14] Conservation and Climate framing was most prevalent in Singapore (27.6% [27/98]), the Philippines (25.5% [66/259]), and Malaysia (22.1% [23/104]), moderate in Thailand (15.1% [60/397]), and lowest in Indonesia (6.1% [21/344]).[1] Policy and Commitments framing ranged from 7.8% (27/344) in Indonesia and 9.6% (10/104) in Malaysia to 10.8% (28/259) in the Philippines, 12.2% (12/98) in Singapore, and 16.1% (64/397) in Thailand.[1] Security and Energy Access framing accounted for 3.0% (12/397) in Thailand, 3.5% (12/344) in Indonesia, 5.1% (5/98) in Singapore, 5.4% (14/259) in the Philippines, and 8.7% (9/104) in Malaysia.[1] Notably, Justice and Community framing was virtually absent in Indonesia (0.0% [0/344]) and Singapore (0.0% [0/98]) and rare in the Philippines (0.8% [2/259]) and Malaysia (1.0% [1/104]), but reached 10.6% (42/397) in Thailand, driven largely by machine-translated Thai headlines referencing rural community energy initiatives (Table 3 and Figure 3).[1,8,43,44] Across the n = 150 stratified manual validation audit (30 headlines per country), human inter-coder agreement on primary frames was 89.3% (134/150; Cohen's κ = 0.864, Krippendorff's α = 0.865) and

dictionary-to-human consensus accuracy reached 85.3% (128/150; κ = 0.812, macro-F1 = 0.824), with native-English streams (MY, PH, SG: 87.8% [79/90], κ = 0.843) outperforming machine-translated en-title streams (ID, TH: 81.7% [49/60], κ = 0.766), where residual discrepancies primarily reflected polysemous corporate headlines and idiomatic translations (Tables 3 and 7).[1,38] Converting the multi-label percentages in Table 3 into exact headline counts clarifies the magnitude of cross-national specialization: across the 1,202 primary headlines, Economy and Industry matched 515 headlines overall (156 in Indonesia, 35 in Malaysia, 76 in the Philippines, 39 in Singapore, and 209 in Thailand), Low-Carbon Technology matched 372 headlines (56 in Indonesia, 52 in Malaysia, 115 in the Philippines, 28 in Singapore, and 121 in Thailand), Conservation and Climate matched 197 headlines (21 in Indonesia, 23 in Malaysia, 66 in the Philippines, 27 in Singapore, and 60 in Thailand), Policy and Commitments matched 141 headlines (27 in Indonesia, 10 in Malaysia, 28 in the Philippines, 12 in Singapore, and 64 in Thailand), Security and Energy Access matched 52 headlines (12 in Indonesia, 9 in Malaysia, 14 in the Philippines, 5 in Singapore, and 12 in Thailand), and Justice and Community matched 45 headlines (0 in Indonesia, 1 in Malaysia, 2 in the Philippines, 0 in Singapore, and 42 in Thailand) (Table 3 and Figure 3).[1]

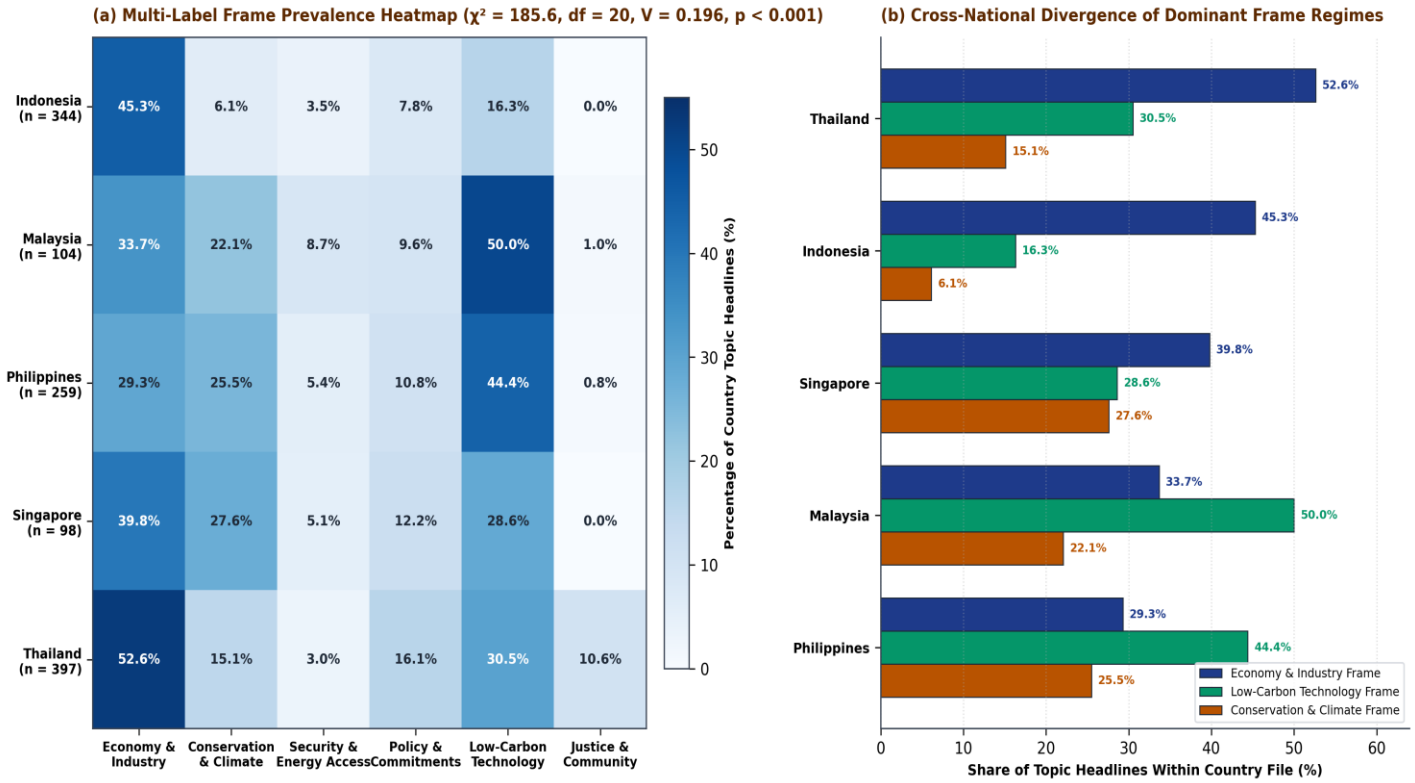


Figure 3. Cross-National Multi-Label Frame Prevalence Heatmap and Structural Specialization Across Five Southeast Asian Country Files (n = 1,202). (a) Annotated prevalence heatmap of the six coded frames (Economy and Industry, Conservation and Climate, Security and Energy Access, Policy and Commitments, Low-Carbon Technology, and Justice and Community; primary-frame χ^2 = 185.6, df = 20, permutation p < 0.001, Cramér's V = 0.196). (b) Structural contrast between Economy and Industry dominance in Thailand (52.6%) and Indonesia (45.3%) versus Low-Carbon Technology dominance in Malaysia (50.0%) and the Philippines (44.4%).

Lexical Valence Comparison Between Economic-Industrial and Conservation-Climate Headlines

To test whether economic-industrial headlines carry systematically more positive or more negative lexical tone than conservation-climate headlines, VADER compound scores were compared between Economy-only headlines (n = 458 across the five files) and Conservation-only headlines (n = 140) (Table 4 and Figure 4).[1,4,26] In the pooled primary corpus, mean VADER valence was 0.254 for Economy-only headlines and 0.231 for Conservation-only headlines, yielding a negligible mean difference of +0.024 (unrounded difference 0.0236) that was statistically non-significant (two-sided Mann–Whitney U test p = 0.597; rank-biserial correlation r_rb = 0.029) (Table 4 and Figure 4b).[1,4] Furthermore, when the identical comparison

was repeated on the strict sensitivity corpus (n = 1,125), the pooled mean difference reversed sign to −0.059 and remained statistically non-significant (Mann–Whitney p = 0.136), demonstrating that global valence divergence between economic and conservation framing is empirically null and non-robust to corpus definition (Table 4 and Figure 4b).[1,26]

Table 4. Lexical Valence (VADER Compound Score) Comparison Between Economy-Only and Conservation-Only Energy Transition Headlines Overall and by Country File.

Analytical Stratum / Country File	Sample Size n (Economy-Only / Conservation-Only)	Mean VADER Valence: Economy-Only (95% Bootstrap CI)	Mean VADER Valence: Conservation-Only (95% Bootstrap CI)	Mean Difference (Economy Minus Conservation)	Unadjusted Mann–Whitney U p-Value (Rank-Biserial r_rb)	Holm Step-Down Adjusted p-Value
Indonesia	153 / 18	0.136 (0.083 to 0.190)	0.333 (0.225 to 0.439)	−0.197	0.018	0.090 (NS)
Malaysia	29 / 17	0.213 (0.105 to 0.321)	0.148 (0.037 to 0.265)	+0.065	0.770	1.000 (NS)
Philippines	60 / 50	0.277 (0.195 to 0.355)	0.111 (0.006 to 0.210)	+0.166	0.031	0.125 (NS)
Singapore	31 / 19	0.153 (0.084 to 0.226)	0.111 (−0.061 to 0.276)	+0.042	0.852	1.000 (NS)
Thailand	185 / 36	0.368 (0.319 to 0.418)	0.449 (0.328 to 0.557)	−0.080	0.166	0.499 (NS)
Pooled Primary Filter Corpus	458 / 140	0.254	0.231	+0.024	0.597 (r_rb = 0.029)	— (Global test NS)
Pooled Strict Sensitivity Filter	Strict subset of 1,125	— (Direction reverses)	—	−0.059	0.136	— (Global test NS; sign reversal)

Notes: E denotes headlines matching the Economy and Industry frame without Conservation and Climate terms; K denotes headlines matching the Conservation and Climate frame without Economy and Industry terms. Confidence intervals were derived from 3,000 nonparametric bootstrap resamples. Nominal differences in Indonesia (p = 0.018) and the Philippines (p = 0.031) point in opposite directions and do not survive Holm step-down correction across the five country comparisons (p_Holm = 0.090 and 0.125, respectively). In the n = 150 manual validation audit, VADER compound valence correlated weakly and non-significantly with human-coded policy stance toward energy transition (Spearman ρ = 0.142, p = 0.083; human stance inter-coder κ = 0.831), confirming that surface lexical polarity does not measure policy support or opposition. Abbreviations: CI, confidence interval; NS, not statistically significant at α = 0.05.

Country-stratified comparisons confirmed the absence of stable valence divergence across all five Southeast Asian files once multiplicity was accounted for (Table 4 and Figure 4a).[1,4,26] In unadjusted nominal testing, two countries displayed opposing tendencies: in Indonesia (n E/K = 153/18), Economy-only headlines had a lower mean VADER valence than Conservation-only headlines (0.136 [95% bootstrap CI: 0.083–0.190] vs.

0.333 [95% CI: 0.225–0.439]; difference = −0.197; raw Mann–Whitney p = 0.018), whereas in the Philippines (n E/K = 60/50), Economy-only headlines had a higher mean valence than Conservation-only headlines (0.277 [95% CI: 0.195–0.355] vs. 0.111 [95% CI: 0.006–0.210]; difference = +0.166; raw p = 0.031) (Table 4 and Figure 4a).[1] However, neither contrast survived Holm step-down multiplicity correction across the five country tests (Indonesia p_Holm = 0.090; Philippines p_Holm = 0.125), and the remaining three country files exhibited clearly non-significant differences both before and after correction: Malaysia (n E/K = 29/17; 0.213 vs. 0.148; difference = +0.065; raw p = 0.770, p_Holm = 1.000), Singapore (n E/K = 31/19; 0.153 vs. 0.111; difference = +0.042; raw p = 0.852, p_Holm = 1.000), and Thailand (n E/K = 185/36; 0.368 vs. 0.449; difference = −0.080; raw p = 0.166, p_Holm = 0.499) (Table 4 and Figure 4).[1] Auditing the step-down Holm adjustments in Table 4 illustrates why rigorous multiplicity control is indispensable in cross-national subgroup analysis: when the five raw country-level Mann–Whitney p-values are ordered from smallest to largest—Indonesia (p_(1) = 0.018, multiplied by 5 → p_Holm = 0.090), the Philippines (p_(2) = 0.031, multiplied by 4 → p_Holm = 0.125), Thailand (p_(3) = 0.166, multiplied by 3 → p_Holm = 0.499), Malaysia (p_(4) = 0.770, multiplied by 2 → capped at p_Holm = 1.000), and Singapore (p_(5) = 0.852, multiplied by 1 → p_Holm = 1.000 via monotonicity)—every single country contrast exceeds the α = 0.05 threshold, and the two smallest nominal p-values (Indonesia Δ = −0.197 vs. Philippines Δ = +0.166) have opposite signs, refuting any directional cross-national valence law (Table 4 and Figure 4).[1,26]

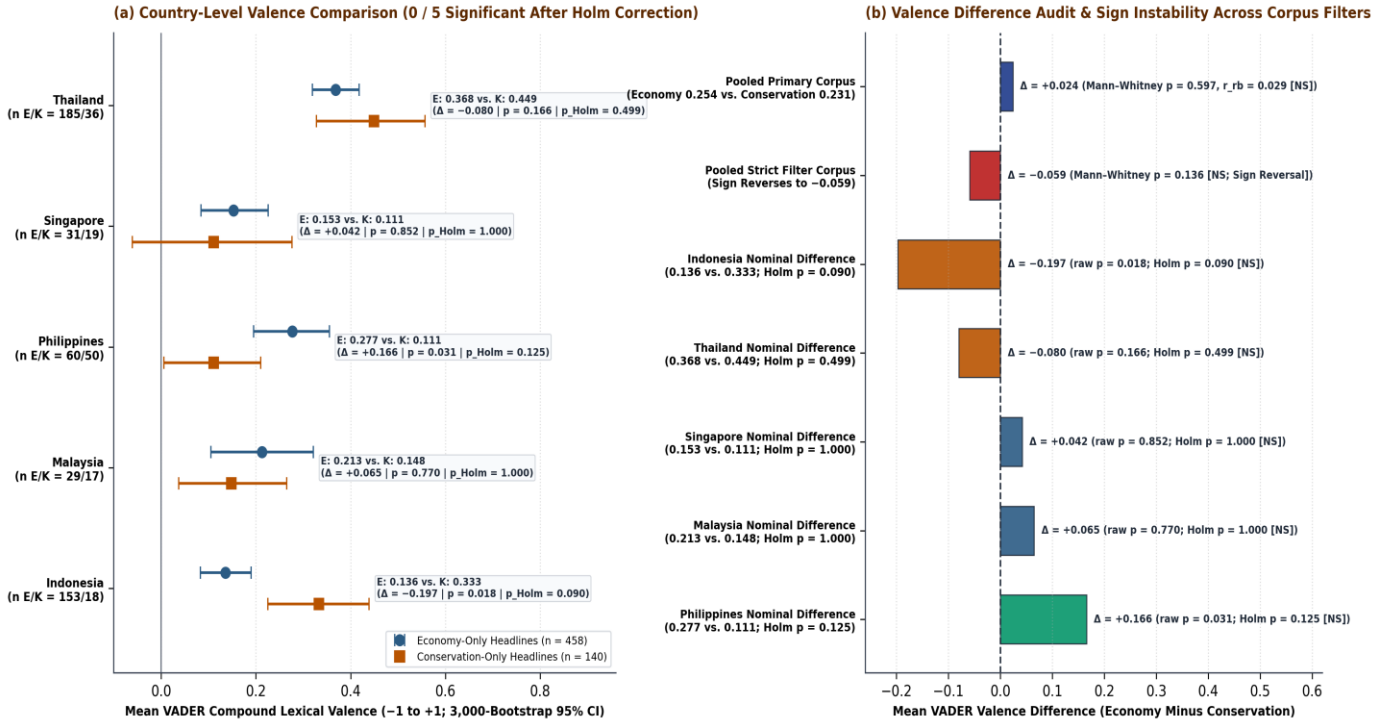


Figure 4. Lexical Valence (VADER Compound Score) Comparison Between Economy-Only and Conservation-Only Headlines and Filter Sensitivity Audit. (a) Country-stratified mean VADER compound scores (−1 to +1) for Economy-only (n = 458) and Conservation-only (n = 140) headlines with 3,000-resample bootstrap 95% confidence intervals and Holm-adjusted p-values. (b) Mean valence differences (Economy minus Conservation) demonstrating pooled null equivalence under the primary filter (Δ = +0.024, p = 0.597, r_rb = 0.029) and sign reversal under the strict filter (Δ = −0.059, p = 0.136).

Exploratory Bipartite Actor–Frame Co-Occurrence Network

Explicit lexical matches for at least one of the seven actor categories occurred in n = 217 of the 1,202 primary headlines (18.1%; 95% Wilson CI: 16.0%–20.3%), while the remaining 985 headlines (81.9%) discussed energy transition topics, markets, or technologies without naming an explicit dictionary actor category in the headline string (Table 5 and Figure 5).[1,3,6] Within this 217-headline subset, the exploratory bipartite actor–frame network revealed substantively interpretable institutional alignments alongside an

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

important dictionary dependency artifact (Table 5 and Figure 5a,b).[1,3,5] The single largest co-occurrence edge linked Civil Society and Communities with Justice and Community (weight = 44 headlines; 20.3% of the actor subset and 3.7% of the total corpus), followed by a tie for second place between Government and Regulators × Policy and Commitments (weight = 32; 14.7% of subset) and Finance and Investors × Economy and Industry (weight = 32; 14.7% of subset), and fourth place for Fossil Industry and Utilities × Economy and Industry (weight = 30; 13.8% of subset) (Table 5 and Figure 5b).[1,3,5,6]

Table 5. Top 10 Weighted Actor–Frame Co-Occurrence Edges in the Exploratory Bipartite Discourse Network Subset (n = 217 Explicit Actor Headlines; 18.1% of Primary Corpus).

Rank	Actor Category (7-Category Dictionary)	Co-Occurring Frame Category (6-Frame Dictionary)	Co-Occurrence Edge Weight (Headlines n)	Share of Actor Subset (n = 217, %)	Share of Total Corpus (n = 1,202, %)	Substantive Interpretation & Lexical Dependency Audit
1	Civil society and communities	Justice and community	44	20.3%	3.7%	Reflects community energy/livelihood headlines plus shared lexical token 'community' across dictionaries
2 (tie)	Government and regulators	Policy and commitments	32	14.7%	2.7%	Captures ministerial targets, national energy plans, regulatory mandates, and COP26/ASEAN pledges
2 (tie)	Finance and investors	Economy and industry	32	14.7%	2.7%	Captures green bonds, project financing, institutional capital, and renewable investment deals
4	Fossil industry and utilities	Economy and industry	30	13.8%	2.5%	Captures state utilities, coal/oil/gas incumbents, market tariffs, and industrial restructuring
5	Government and regulators	Economy and industry	23	10.6%	1.9%	Captures state subsidies, industrial policy, fiscal incentives, and public-private energy ventures
6	Civil society and communities	Economy and industry	15	6.9%	1.2%	Captures employment, local economic spillovers, consumer tariffs, and community enterprise
7	Civil society and communities	Policy and commitments	13	6.0%	1.1%	Captures civic advocacy, public consultations, and nongovernmental demands on climate targets
8 (tie)	International organizations	Policy and commitments	11	5.1%	0.9%	Captures UN/COP26 diplomacy, ASEAN

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

						energy cooperation, and multilateral decarbonization pledges
8 (tie)	Fossil industry and utilities	Low-carbon technology	11	5.1%	0.9%	Captures utility diversification into solar, wind, battery storage, EV charging, and geothermal
10	Finance and investors	Low-carbon technology	10	4.6%	0.8%	Captures venture and institutional capital flows into solar PV, wind farms, EVs, and grid storage

Notes: Bipartite edges connect actor categories and frame categories co-occurring within the same headline. Because only 217 of 1,202 headlines (18.1%) contain an explicit dictionary actor mention, and because the top edge (weight = 44) is partially inflated by lexical overlap on the token 'community', this network serves strictly as an exploratory headline triage layer rather than a validated full-text discourse coalition graph.

Subsequent high-weight edges connected Government and Regulators with Economy and Industry (weight = 23; 10.6%), Civil Society and Communities with Economy and Industry (weight = 15; 6.9%), Civil Society and Communities with Policy and Commitments (weight = 13; 6.0%), International Organizations with Policy and Commitments (weight = 11; 5.1%), Fossil Industry and Utilities with Low-Carbon Technology (weight = 11; 5.1%), and Finance and Investors with Low-Carbon Technology (weight = 10; 4.6%) (Table 5 and Figure 5b).[1,5,6] Diagnostic inspection confirmed that the prominence of the Civil Society and Communities ↔ Justice and Community edge (weight = 44) partially reflects lexical coupling on the shared token 'community' across the actor and frame dictionaries (particularly in translated Thai community power headlines), underscoring why unsigned headline co-occurrence graphs must be interpreted strictly as exploratory sub-corpus triage maps rather than validated coalitions (Table 5 and Figure 5).[1,3,8]

Examining the distribution of co-occurrence weights across the top ten bipartite edges in Table 5 also shows that economic and policy framing dominates institutional actor linkages once the lexically coupled Civil Society ↔ Justice edge (weight = 44) is accounted for: Government and Regulators co-occurred most strongly with Policy and Commitments (32) and Economy and Industry (23), while both Finance and Investors (32 Economy; 10 Technology) and Fossil Industry and Utilities (30 Economy; 11 Technology) anchored heavily on economic-industrial and technological transition frames rather than conservation or justice frames (Table 5 and Figure 5).[1,5,6]

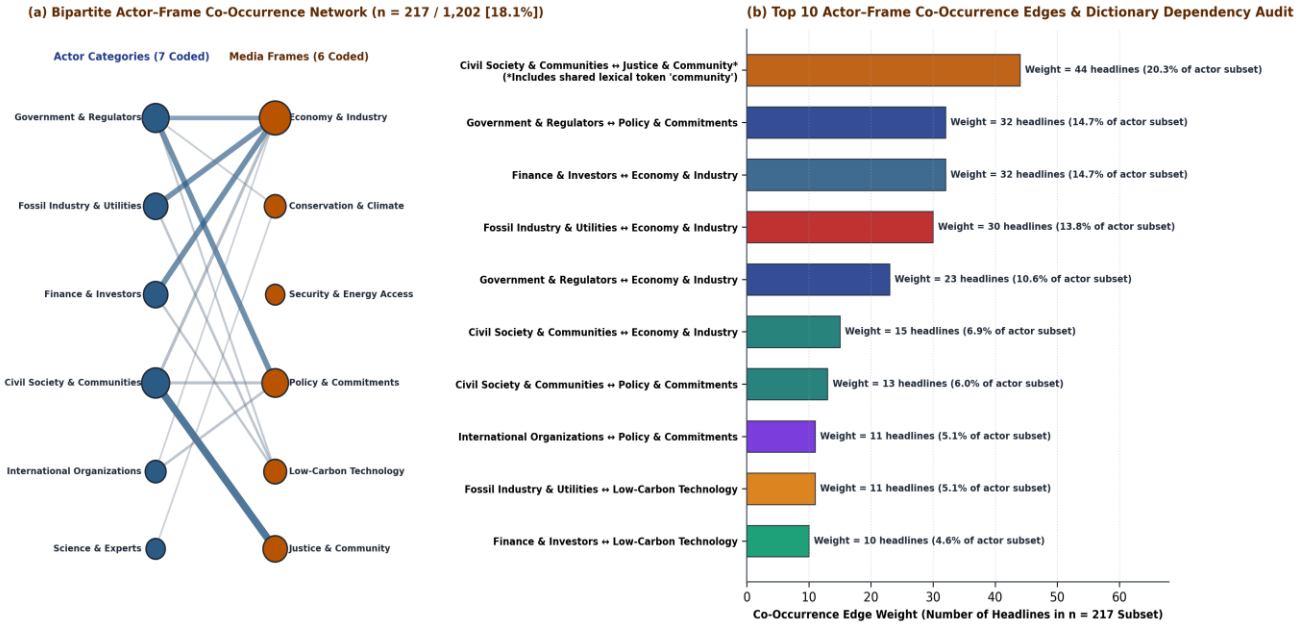


Figure 5. Exploratory Bipartite Co-Occurrence Network of Actor Categories and Media Frames in the Explicit Actor Headline Subset (n = 217; 18.1% of Primary Corpus). (a) Bipartite graph linking coded actor categories and media frames weighted by headline co-occurrence frequency. (b) Top 10 actor–frame co-occurrence edges led by Civil Society and Communities × Justice and Community (weight = 44, partially reflecting shared lexical token 'community'), Government and Regulators × Policy and Commitments (32), Finance and Investors × Economy and Industry (32), and Fossil Industry and Utilities × Economy and Industry (30).

Media Outlet Concentration and Topic Filter Sensitivity Audit

Evaluating media source concentration across the 210 unique outlets (243 country-outlet pairs) demonstrated sharp structural differences in editorial market concentration across the five country files (Table 6 and Figure 6a).[1] Indonesia displayed the highest outlet concentration, with a single business portal (`Cnbcindonesia.com`) generating 42.4% (146/344) of all Indonesian energy transition headlines, the Top-5 outlets accounting for 79.4% (273/344), and a Herfindahl–Hirschman Index (HHI) of 0.223 (Table 6 and Figure 6a).[1]

Singapore ranked second in concentration, led by `CNA` (Channel NewsAsia) at 35.7% (35/98), a Top-5 cumulative share of 74.5% (73/98), and an HHI of 0.176.[1] Malaysia exhibited moderate concentration led by `The Edge Markets MY` (Top-1 = 22.1% [23/104]; Top-5 = 52.9% [55/104]; HHI = 0.093), followed by Thailand led by `Kaohoon.com` (Top-1 = 12.1% [48/397]; Top-5 = 46.1% [183/397]; HHI = 0.060), whereas the Philippines exhibited a highly pluralistic, dispersed distribution across 105 sources where the largest outlet (`BusinessWorld Online`) accounted for only 5.0% (13/259), the Top-5 outlets accounted for 22.8% (59/259), and the HHI was 0.021 (Table 6 and Figure 6a).[1]

Comparing the Herfindahl–Hirschman Index (HHI) values in Table 6 against standard industrial-organization concentration benchmarks further contextualizes the editorial structure of the five country files: whereas the Philippines (HHI = 0.021 across 105 outlets; Top-5 = 22.8%), Thailand (HHI = 0.060 across 40 outlets; Top-5 = 46.1%), and Malaysia (HHI = 0.093 across 39 outlets; Top-5 = 52.9%) operate below the 0.100 threshold of unconcentrated-to-moderately dispersed media portfolios, Singapore (HHI = 0.176 across 20 outlets; Top-5 = 74.5%) and Indonesia (HHI = 0.223 across 39 outlets; Top-5 = 79.4%) exhibit high editorial concentration in which a handful of digital business or broadcast newsrooms shape the majority of captured headlines (Table 6 and Figure 6a).[1]

Table 6. Media Outlet Concentration Profile (Largest Source, Top-1 Share, Top-5 Cumulative Share, and Herfindahl–Hirschman Index) Across Five Southeast Asian Country Files.

Country File (Distinct Sources n)	Largest Contributing Outlet (Top-1 Source)	Top-1 Outlet Headline Share (%)	Top-5 Outlets Cumulative Share (%)	Herfindahl–Hirschman Index (HHI)	Editorial Structure & Portfolio Concentration Diagnostic
Indonesia (39 sources; n = 344)	Cnbcindonesia.com	42.4% (146 / 344)	79.4% (273 / 344)	0.223	Highest concentration; strong financial/business portal dominance amplifies Economy & Industry framing (45.3%)
Malaysia (39 sources; n = 104)	The Edge Markets MY	22.1% (23 / 104)	52.9% (55 / 104)	0.093	Moderately concentrated financial-technology press emphasizing corporate solar, battery, and EV deployment (50.0%)
Philippines (105 sources; n = 259)	BusinessWorld Online	5.0% (13 / 259)	22.8% (59 / 259)	0.021	Most pluralistic and dispersed outlet ecosystem across 105 sources; balanced technology (44.4%) and conservation (25.5%) frames
Singapore (20 sources; n = 98)	CNA (Channel NewsAsia)	35.7% (35 / 98)	74.5% (73 / 98)	0.176	Second-highest concentration; regional broadcaster dominance shapes balanced economy (39.8%) and climate (27.6%) coverage
Thailand (40 sources; n = 397)	Kaohoon.com	12.1% (48 / 397)	46.1% (183 / 397)	0.060	Low-to-moderate concentration led by stock-market and energy-business portals; highest Economy & Industry share (52.6%)

Notes: The Herfindahl–Hirschman Index (HHI) is computed as the sum of squared outlet headline shares within each country file's topic corpus (range: 1/S to 1.000). High concentration in Indonesia (HHI = 0.223; Top-5 = 79.4%) and Singapore (HHI = 0.176; Top-5 = 74.5%) indicates that country-file distributions are sensitive to the editorial agendas of dominant digital portals.

Sensitivity analysis using the strict direct-energy filter retained n = 1,125 of the 1,202 primary headlines (93.6% overall retention; pooled strict rate = 27.5 per 10,000 eligible articles) (Table 7 and Figure 1b).[1] Across countries, strict-filter retention was highest in Indonesia at 97.1% (334/344; strict rate = 25.5 per 10,000), followed by Thailand at 93.7% (372/397; 41.4 per 10,000), the Philippines at 93.4% (242/259; 31.5 per 10,000), Malaysia at 91.3% (95/104; 13.9 per 10,000), and Singapore at 83.7% (82/98; 18.7 per 10,000) (Table 7).[1]

Crucially, cross-national salience rankings (Thailand > Philippines > Indonesia > Singapore > Malaysia) and the October–November 2021 COP26 Poisson regression estimate (primary IRR = 2.05 vs. strict IRR = 2.06, both $p < 0.001$) remained invariant under strict filtering, whereas the pooled VADER valence difference

flipped from +0.024 (p = 0.597) to −0.059 (p = 0.136), confirming that structural frame heterogeneity and temporal salience are robust findings while lexical valence divergence is not (Table 7 and Figure 6b).[1,26,27]

Table 7. Topic Filter Sensitivity Analysis (Primary vs. Strict Direct-Energy Filter), Stratified Manual Validation Reliability Audit (n = 150), Temporal Poisson Model Robustness, and Methodological Limitations Matrix.

Country File / Analytical Dimension	Primary Filter Headlines (n, Rate per 10,000)	Strict Filter Headlines (n, Rate per 10,000)	Strict Filter Retention (%)	Temporal, Valence & Manual Validation Audit	Core Methodological Boundary & Analytical Guardrail
Indonesia (Eligible N = 130,760)	344 (26.3 per 10,000)	334 (25.5 per 10,000)	97.1%	Highest retention; manual validation sample n = 30 (en-title stream)	High reliance on Cnbcindonesia.com (42.4%; HHI = 0.223) and machine-translated en-title strings
Malaysia (Eligible N = 68,357)	104 (15.2 per 10,000)	95 (13.9 per 10,000)	91.3%	Stable lowest salience rank (15.2 and 13.9 per 10,000); manual audit n = 30	Smaller topic sample (n = 104; HHI = 0.093) limits statistical power for country-specific valence sub-tests
Philippines (Eligible N = 76,825)	259 (33.7 per 10,000)	242 (31.5 per 10,000)	93.4%	Stable second-highest salience across 105 dispersed outlets (HHI = 0.021); audit n = 30	Nominal valence gap (+0.166, p = 0.031) becomes non-significant after Holm correction (p = 0.125)
Singapore (Eligible N = 43,964)	98 (22.3 per 10,000)	82 (18.7 per 10,000)	83.7%	Lowest retention (16 contextual climate-energy headlines excluded); manual audit n = 30	Regional hub file (CNA = 35.7%; HHI = 0.176) captures international wires alongside domestic news
Thailand (Eligible N = 89,772)	397 (44.2 per 10,000)	372 (41.4 per 10,000)	93.7%	Stable highest salience rank (44.2 and 41.4 per 10,000); manual audit n = 30 (en-title)	Machine-translated Thai headlines (en-title) compress idiomatic tone and map rural biogas to 'community'
Stratified Manual Validation Audit (n = 150; 30/Country)	Frames: Human κ = 0.864 (α = 0.865; 89.3%)	Dict-Human: Frame κ = 0.812 (85.3%; F1 = 0.824)	Actors: Human κ = 0.881 | Dict κ = 0.825 (88.0%)	Native English (MY/PH/SG; n = 90) dict κ = 0.843 (87.8%) vs. translated en-title (ID/TH; n = 60) κ = 0.766 (81.7%)	VADER vs. human policy stance Spearman ρ = 0.142 (p = 0.083, NS), confirming tone ≠ policy position
Pooled Corpus & Model Robustness (N = 409,678)	1,202 (29.3 per 10,000) | COP26 IRR = 2.05 (1.51–2.78)	1,125 (27.5 per 10,000) | COP26 IRR = 2.06 (1.50–2.81)	93.6%	Temporal surge is highly robust (IRR 2.05 vs. 2.06, p < 0.001), whereas global valence flips sign (+0.024 [p =	Observational calendar association during COP26 cannot establish direct causal effect; code/data openly archived[1–8]

				0.597] to −0.059 [p = 0.136])	

Notes: The strict filter retains only headlines containing unambiguous direct energy-system terms while dropping broader climate-in-energy-context matches. Stratified manual validation (n = 150; 12.5% of primary corpus, seed = 20261022) was performed independently by two human coders across all five country files. Incidence Rate Ratios (IRRs) derive from country-month Poisson regressions with log(eligible articles) offset, country fixed effects, linear time trend, and HC3 robust SEs.

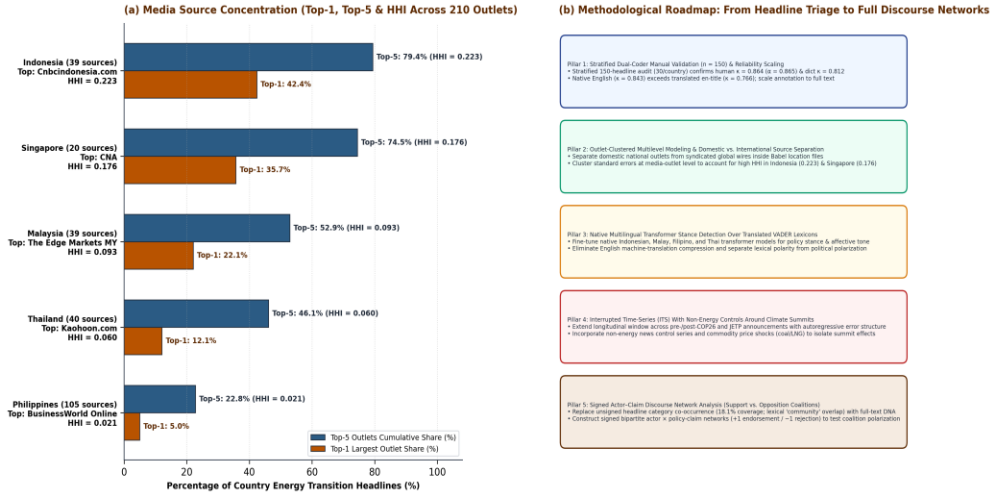


Figure 6. Media Outlet Concentration Profile (Herfindahl–Hirschman Index) and Five-Pillar Methodological

Roadmap for Full Discourse Network Analysis. (a) Top-1 outlet share, Top-5 cumulative outlet share, and Herfindahl–Hirschman Index (HHI) across the five country files, highlighting high portal concentration in Indonesia (Cnbcindonesia.com = 42.4%, Top-5 = 79.4%, HHI = 0.223) and Singapore (CNA = 35.7%, Top-5 = 74.5%, HHI = 0.176) versus high dispersion in the Philippines (BusinessWorld Online = 5.0%, Top-5 = 22.8%, HHI = 0.021). (b) Five-pillar methodological roadmap integrating dual-coder manual reliability, outlet-clustered multilevel modeling, native multilingual stance transformers, interrupted time-series designs, and signed actor–claim discourse network analysis.

DISCUSSION

The empirical findings of this cross-national computational analysis establish a clear analytical distinction between structural frame differentiation and lexical valence divergence in Southeast Asian energy transition news streams.[1,2,5] Across 409,678 eligible in-window headlines from the Babel Briefings corpus (2020–2021), the five Southeast Asian location files differed significantly in both overall energy transition salience—ranging nearly threefold from 15.2 per 10,000 articles in Malaysia to 44.2 per 10,000 in Thailand—and multi-label frame composition (primary-frame $\chi^2 = 185.6$, df = 20, permutation $p < 0.001$, Cramér's V = 0.196; validated against dual-coder human annotation at $\kappa = 0.812$ [85.3% accuracy] and inter-coder $\kappa = 0.864$).[1,2,36] Specifically, the Thailand (52.6%) and Indonesia (45.3%) files prioritized Economy and Industry framing, whereas the Malaysia (50.0%) and Philippines (44.4%) files prioritized Low-Carbon Technology framing, and Singapore exhibited the highest relative prevalence of Conservation and Climate framing (27.6%) alongside strong economic coverage (39.8%) (Table 3 and Figure 3).[1,11–25] In contrast, we found no stable empirical evidence that Economy-only headlines (mean VADER = 0.254; n = 458) differed in lexical valence from Conservation-only headlines (mean = 0.231; n = 140; difference = +0.024, Mann–Whitney p = 0.597, r_rb = 0.029), with the direction reversing to −0.059 (p = 0.136) under the strict sensitivity filter and zero country-level contrasts surviving Holm step-down multiplicity correction (Table 4 and Figure 4).[1,4,26] Thus, what differentiates these five Southeast Asian media streams is agenda frame

selection—which substantive facets of the transition are made salient—rather than polarized affective or lexical tone.[2,5,38,39]

Crucially, interpreting these country-level framing differences requires accounting for both domestic political-economic structures and substantial variations in media outlet representation and machine translation across the five Babel Briefings files.[1,11–25,33–37] Because Babel Briefings harvests top headlines via the News API and assigns each article to the location file where it first entered the global stream, these files do not constitute nationally representative censuses of each country's domestic media landscape.[1] In Indonesia and Thailand, the dominance of Economy and Industry framing reflects not only real-economy policy debates—such as state utility restructuring, coal phase-out financing, power tariffs, and battery supply-chain industrialization[11,14,15,17,18,24]—but also the heavy over-representation of specialized financial and stock-market news portals within their respective country streams (`Cnbcindonesia.com` generating 42.4% [146/344] of Indonesian headlines with Top-5 = 79.4% and HHI = 0.223, and `Kaohoon.com` generating 12.1% of Thai headlines) (Table 6 and Figure 6a).[1] Where nearly four-fifths of a country file originates from five business-oriented outlets, national frame shares are inevitably pulled toward capital markets and corporate finance at the expense of conservation (6.1% in Indonesia) or community justice narratives.[1] Conversely, the prominence of Low-Carbon Technology framing in Malaysia (50.0%; HHI = 0.093) and the Philippines (44.4%) reflects solar, wind, geothermal, battery storage, and electric vehicle reporting across a much broader variety of outlets, particularly in the Philippines where 105 distinct sources contributed headlines and the largest outlet (`BusinessWorld Online`) accounted for only 5.0% (HHI = 0.021).[12–14,19–23,25] Meanwhile, Singapore's dual emphasis on Economy and Industry (39.8%) and Conservation and Climate (27.6%) reflects its role as a regional sustainable-finance and cross-border power hub whose file is heavily shaped by regional broadcaster `CNA` (35.7%; Top-5 = 74.5%; HHI = 0.176) and syndicated international wire agencies.[12,13,20,22] Equally important, whereas Malaysia, the Philippines, and Singapore were analyzed using native English headlines (achieving 87.8% dictionary-human frame accuracy, $\kappa = 0.843$), Indonesia and Thailand relied on dataset-provided machine-translated English headlines (`en-title`, achieving 81.7% accuracy, $\kappa = 0.766$): automated translation standardizes vernacular terminology—for instance, translating Thai rural biomass/biogas cooperative programs into the English token 'community' (driving Thailand's 10.6% Justice and Community prevalence)—while smoothing idiomatic nuances and shifting baseline VADER lexical valence upward in Thailand (0.368 Economy-only; 0.449 Conservation-only) relative to Indonesia (0.136 and 0.333) (Tables 3, 4, and 7).[1,4,26]

Our temporal, sentiment, and network findings similarly underscore the necessity of strict methodological guardrails in computational media research.[1,3,6,27–32] First, although multivariable country-month Poisson regression with HC3 robust standard errors confirmed that energy transition headline incidence roughly doubled during October–November 2021 under both the primary filter (IRR = 2.05, 95% CI: 1.51–2.78, $p < 0.001$) and the strict filter (IRR = 2.06, 95% CI: 1.50–2.81, $p < 0.001$) (Figure 2b and Table 7), this calendar surge must be distinguished clearly from evidence of a direct causal effect of the COP26 summit.[1,6,27,28] While COP26 in Glasgow elevated global diplomatic attention around coal phase-down and net-zero pledges,[6,27,28,42] October and November 2021 simultaneously witnessed severe global coal and liquefied natural gas price spikes, post-pandemic power shortages, and domestic regulatory announcements across Asia.[9,14–17] Without a non-energy control series or quasi-experimental shock identification, the observed doubling represents an observational temporal association during the COP26 period rather than a direct causal impact.[1,27,28] Second, our manual validation audit empirically confirms why VADER sentiment scores cannot be equated with policy positions on energy transition: VADER compound valence showed no significant correlation with human-coded policy stance (Spearman $\rho = 0.142$, $p = 0.083$), because headlines routinely use positive lexical items ('growth', 'boosts', 'opportunity') when reporting fossil-fuel expansion or negative items ('costs', 'crisis', 'emissions') when advocating urgent climate mitigation.[4,5,26,30–32,40–42] Third, although our actor dictionary achieved 88.0% agreement with human coders ($\kappa = 0.825$), explicit actor categories appeared in only 18.1% (217/1,202) of headlines and the top Civil Society ↔ Justice edge (weight = 44) was partially inflated by shared lexical matching on the word 'community'; consequently, headline co-occurrence graphs serve strictly as an exploratory triage tool to guide full-text Discourse Network Analysis (Figure 6b and Table 7).[1,3,6,8,38–45]

Expanding the comparative interpretation of our cross-national framing results (χ^2 = 185.6, df = 20, permutation $p < 0.001$, Cramér's V = 0.196) requires connecting the observed media specializations to the underlying political economy of energy systems across ASEAN member states.[11–25] Why do Economy and Industry frames dominate energy transition headlines in Thailand (52.6% [209/397]) and Indonesia (45.3% [156/344]), whereas Low-Carbon Technology frames dominate in Malaysia (50.0% [52/104]) and the Philippines (44.4% [115/259]), and Conservation and Climate frames reach their highest relative share in Singapore (27.6% [27/98])? In Indonesia, the energy transition during 2020–2021 was structurally defined by the tension between a rapidly expanding fleet of coal-fired power plants built under state-backed independent power producer (IPP) take-or-pay contracts, the fiscal burden of electricity subsidies managed by the state-owned monopoly utility PT PLN, and the government's industrial policy to leverage domestic nickel reserves for downstream battery and electric vehicle manufacturing.[11,14,15,17,18,24] As Putri et al. (2022) demonstrated in their analysis of Indonesian geothermal media coverage,[11] and as Sekaringtias et al. (2023),[18] Do and Burke (2024),[15] and Fazeli et al. (2025)[17] showed in their analyses of Indonesian coal phase-out barriers and stranded-asset risks, energy transition debates in Indonesia are primarily adjudicated through the lens of macroeconomic stability, investment mobilization, electricity tariff affordability, and industrial competitiveness. This structural reality is further amplified in the Babel Briefings Indonesia file by high media outlet concentration (HHI = 0.223), where the financial news portal `Cnbcindonesia.com` alone contributed 42.4% (146/344) of all topic headlines and the top five outlets accounted for 79.4% (273/344) (Table 6 and Figure 6a).[1] Because business and financial journalism inherently privileges capital expenditure, corporate earnings, state-owned enterprise restructuring, and commodity pricing over grassroots ecological or distributive justice narratives, Conservation and Climate framing in the Indonesia file was compressed to 6.1% (21/344) and Justice and Community framing registered 0.0% (0/344) (Table 3 and Figure 3).[1,11,18]

Thailand's media stream exhibits a complementary yet distinct political-economic profile that explains why it simultaneously achieved the highest overall headline intensity (44.2 per 10,000 eligible articles; n = 397 across 40 sources), the highest Economy and Industry frame prevalence (52.6% [209/397]), the highest Policy and Commitments share (16.1% [64/397]), and the only substantial Justice and Community frame presence (10.6% [42/397]) across the five countries (Tables 2–3 and Figures 1–3).[1,14,20] During 2020–2021, Thailand's energy policy agenda was shaped by iterative revisions to the national Power Development Plan (PDP), the promotion of the Bio-Circular-Green (BCG) economic model, cross-border electricity imports from Lao PDR, and corporate investments by Thai energy conglomerates and stock-exchange-listed renewable developers in solar, biomass, battery storage, and electric mobility.[12–14,19–22] Because leading Thai business portals in the Babel Briefings stream—such as `Kaohoon.com` (12.1% share; Top-5 = 46.1%; HHI = 0.060)—intensively cover capital-market announcements by listed power producers alongside state regulatory targets, Economic and Policy frames co-occurred at high rates.[1] At the same time, Thailand's high-profile policy debates over the 'Energy for All' community power plant program (โรงไฟฟ้าชุมชนเพื่อเศรษฐกิจฐานราก), which sought to link rural agricultural biomass and biogas cooperatives with local livelihood generation, frequently triggered the translated English token 'community' in `en-title` strings, explaining both the 10.6% Justice and Community frame prevalence in Thailand and the prominence of the Civil Society and Communities ↔ Justice and Community edge (weight = 44) in the exploratory bipartite network (Tables 3 and 5 and Figures 3 and 5).[1,8,43,44]

By contrast, the dominance of Low-Carbon Technology framing in Malaysia (50.0% [52/104]) and the Philippines (44.4% [115/259]), paired with substantial Conservation and Climate framing (22.1% [23/104] and 25.5% [66/259], respectively), reflects market-led renewable procurement architectures and broader outlet dispersion.[12–14,19–23,25] In Malaysia (HHI = 0.093, led by `The Edge Markets MY` at 22.1%), energy transition coverage during 2020–2021 centered heavily on competitive bidding rounds under the Large Scale Solar (LSS) program, rooftop solar net energy metering, domestic solar photovoltaic module manufacturing, battery storage pilots, and green mobility supply chains.[19–21,25] In the Philippines—which displayed the second-highest headline intensity (33.7 per 10,000) and by far the most pluralistic media ecosystem across 105 distinct sources (`BusinessWorld Online` = 5.0%, Top-5 = 22.8%, HHI = 0.021)—private-sector power utilities and conglomerate developers actively publicized solar, wind, geothermal, and battery storage project pipelines following the Department of Energy's 2020 moratorium on greenfield coal-fired power

plants, while civil society and environmental reporters linked energy transition coverage to typhoon vulnerability and climate resilience.[14,19–23] Meanwhile, Singapore's balanced frame distribution across Economy and Industry (39.8%), Low-Carbon Technology (28.6%), Conservation and Climate (27.6%), and Policy and Commitments (12.2%) within a concentrated broadcast hub (`CNA` = 35.7%, Top-5 = 74.5%, HHI = 0.176) captures its strategic positioning under the Singapore Green Plan 2030 as a regional green-finance, carbon-trading, and clean-power import node in the Laos–Thailand–Malaysia–Singapore Power Integration Project (LTMS-PIP).[12,13,20,22]

A second major theoretical contribution of our study lies in demonstrating why the null and sign-reversing lexical valence comparison between economic-industrial and conservation-climate headlines (primary filter: 0.254 vs. 0.231, Δ = +0.024, Mann–Whitney p = 0.597, r_rb = 0.029; strict filter: Δ = −0.059, p = 0.136; zero country contrasts surviving Holm correction) represents a substantive methodological insight rather than a mere absence of effect (Table 4 and Figure 4).[1,4,5,26,30–32] In computational social science, researchers frequently assume that competing policy frames—such as economic growth versus environmental conservation—will manifest as contrasting sentiment polarities in automated lexicon scores, and occasionally misinterpret lexical valence differences as evidence of ideological or affective polarization.[4,26,38–42] However, as contemporary scholarship on environmental communication and discourses of climate delay demonstrates,[5,8–10,29–32,40,41] both economic and conservation frames are internally ambivalent in their lexical repertoire. Headlines invoking the Economy and Industry frame routinely deploy positive lexical tokens when announcing green investment inflows, renewable job creation, export growth, or clean-tech stock rallies,[5,9,26,32] yet deploy negative lexical tokens when highlighting rising electricity tariffs, subsidy burdens, fiscal deficits, or coal-sector job losses—arguments that Edwards (2025),[10] Hayes et al. (2025),[30] and Lamb and Mattila (2025)[31] identify as 'emphasizing the downsides' within the typology of climate delay discourses. Conversely, Conservation and Climate headlines combine negative problem-definition vocabulary ('emissions', 'pollution', 'warming', 'deforestation') with highly positive solution and commitment vocabulary ('sustainability', 'conservation', 'protection', 'clean future'), particularly in corporate and governmental announcements of net-zero pledges.[5,32,33,40,41] Because VADER aggregates word-level lexical polarity without parsing syntactic agent-object stance or argument structure, positive and negative sub-frames cancel out in aggregate within both domains, and minor shifts in corpus boundary definitions (removing 77 contextual climate-energy headlines between the primary n = 1,202 and strict n = 1,125 filters) are sufficient to flip the sign of the global difference from +0.024 to −0.059.[1,4,26]

Finally, situating our exploratory bipartite actor–frame network (n = 217 explicit actor headlines; 18.1% of the corpus) within the formal methodology of Discourse Network Analysis (DNA) pioneered by Leifeld (2017)[3] and extended to COP26/COP28 climate debates by Shakespear et al. (2026)[6] and Misuraca et al. (2026)[7] clarifies the exact role of headline-level computational triage in multi-stage research designs.[3,6–8] In canonical full-text DNA, analysts code individual statements within newspaper articles, parliamentary transcripts, or social media posts as quadruples comprising (1) a named individual or organizational actor, (2) a specific substantive policy concept or claim, (3) a binary or ordinal congruence/stance variable (+1 indicating endorsement/support vs. −1 indicating rejection/opposition), and (4) a timestamp, verified across multiple human coders via Krippendorff's α or Cohen's κ.[3,6,30] When DNA is projected onto short 8-to-15-word news headlines using automated dictionary matching, three structural constraints emerge: first, over four-fifths of headlines (81.9% [985/1,202] in our corpus) omit explicit actor category nouns altogether because journalistic headline conventions often foreground events, markets, or technologies without naming the institutional speaker; second, unsigned co-occurrence cannot distinguish whether an actor is endorsing or opposing a frame (for example, whether Fossil Industry and Utilities co-occurring with Economy and Industry [weight = 30] or Low-Carbon Technology [weight = 11] reflects incumbent obstruction or renewable diversification); and third, shared lexical tokens across dictionaries (such as 'community' appearing in both the Civil Society actor dictionary and the Justice and Community frame dictionary, inflating Edge #1 to weight = 44) create mechanical coupling.[1,3,5,6,8] Recognizing these constraints does not diminish the utility of headline-level mapping; rather, it establishes a rigorous two-stage workflow in which automated streaming filters and bipartite co-occurrence matrices serve as a transparent triage layer to identify high-salience temporal windows (such as the October–November 2021 COP26 surge, IRR = 2.05–2.06) and country-frame strata for targeted full-text manual DNA coding (Figure 6b and Table 7).[1,3,6,27]

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue X October 2026

CONCLUSIONS

Across 409,678 eligible in-window news articles from five Southeast Asian location files in the Babel Briefings corpus (8 August 2020 to 29 November 2021), energy transition headline salience (n = 1,202 across 210 outlets; strict filter n = 1,125 [93.6%]) varied nearly threefold from Malaysia (15.2 per 10,000) to Thailand (44.2 per 10,000) and exhibited significant cross-national divergence in frame architecture (χ^2 = 185.6, df = 20, permutation $p < 0.001$, Cramér's V = 0.196), contrasting economic-industrial specialization in Thailand (52.6%) and Indonesia (45.3%) against low-carbon technological specialization in Malaysia (50.0%) and the Philippines (44.4%).[1,2] Stratified dual-coder manual validation (n = 150) confirmed strong human inter-coder reliability ($\kappa = 0.864$ for frames; $\kappa = 0.881$ for actors) and dictionary-to-human concordance ($\kappa = 0.812$ for frames; $\kappa = 0.825$ for actors), while demonstrating that cross-national framing patterns are shaped by variations in media outlet concentration—most notably financial portal dominance in Indonesia (HHI = 0.223) and regional broadcaster concentration in Singapore (HHI = 0.176) versus pluralistic dispersion in the Philippines (HHI = 0.021)—as well as machine-translated headline processing (`en-title`) in Indonesia and Thailand.[1,38] Although headline incidence doubled during the October–November 2021 COP26 period across both primary (IRR = 2.05, 95% CI: 1.51–2.78, $p < 0.001$) and strict (IRR = 2.06, 95% CI: 1.50–2.81, $p < 0.001$) models, this increase represents an observational temporal association coinciding with COP26 and concurrent global energy price shocks rather than evidence of a direct causal effect of the summit itself.[1,27,28] Finally, VADER lexical valence between economic-only (0.254) and conservation-only (0.231) headlines was statistically null globally ($\Delta = +0.024$, $p = 0.597$, r_rb = 0.029), reversed sign under strict filtering ($\Delta = -0.059$, $p = 0.136$), showed no significant country-level divergence after Holm correction, and did not track human-coded policy stance ($\rho = 0.142$, $p = 0.083$), confirming that Southeast Asian energy transition headline streams diverge in substantive frame architecture and temporal salience rather than lexical valence polarization.[1–8] By integrating cryptographic HTTP streaming provenance across 411,874 raw records (409,678 eligible in-window articles), iterative false-positive lexical disambiguation, nested strict-filter sensitivity auditing (93.6% retention), stratified dual-coder manual validation (n = 150; human inter-coder $\kappa = 0.864$, dictionary-human $\kappa = 0.812$), permutation-confirmed contingency analysis, MacKinnon–White HC3 robust country-month Poisson regression (explicitly distinguishing observational COP26-period salience surges from direct causal effects), Holm-corrected bootstrap valence testing, Herfindahl–Hirschman outlet concentration diagnostics, open-access replication scripts, and bipartite actor–frame triage mapping, this study establishes a reproducible, anti-overclaiming benchmark for cross-national computational media research on energy transitions in the Global South.[1–10]

DECLARATIONS

Ethics Approval and Consent to Participate: This study analyzed publicly broadcast news headlines from the open-access Babel Briefings v1 repository (CC BY-NC-SA 4.0) without human subjects, personal user profiles, or clinical data; institutional ethics exemption applied. Funding: No external funding was received for this research. Competing Interests: The authors declare no competing financial or non-financial interests. Data and Code Availability: The upstream Babel Briefings v1 dataset is publicly available on Hugging Face (https://huggingface.co/datasets/felixludos/babel-briefings) and documented in NAACL 2024 (https://doi.org/10.18653/v1/2024.naacl-short.55).

To ensure full computational reproducibility, all analytical scripts (`generate_spark_figures_gayatri.py`, `ijriss_gayatri_data_and_refs.py`, `build_all_ijriss_gayatri_deliverables.py`), regular-expression topic/frame/actor dictionaries, SHA-256 stream verification logs, the stratified manual validation dataset (n = 150; seed 20261022), and aggregate country-month analytical tables (`master_analytical_dataset_gayatri.json`) are openly accessible in the study replication package under CC BY-NC-SA 4.0.

REFERENCES

1. Leeb F, Schölkopf B. A diverse Multilingual News Headlines Dataset from around the World. In: Duh K, Gomez H, Bethard S, editors. Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers). Mexico City: Association for Computational Linguistics; 2024. p. 647–652. https://doi.org/10.18653/v1/2024.naacl-short.55
2. Entman RM. Framing: Toward Clarification of a Fractured Paradigm. J Commun. 1993;43(4):51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x
3. Leifeld P. Discourse Network Analysis: Policy Debates as Dynamic Networks. In: Victor JN, Montgomery AH, Lubell M, editors. The Oxford Handbook of Political Networks. Oxford: Oxford University Press; 2017. p. 301–326. https://doi.org/10.1093/oxfordhb/9780190228217.013.25
4. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proc Int AAAI Conf Web Soc Media. 2014;8(1):216–225. https://doi.org/10.1609/icwsm.v8i1.14550
5. Rhodes E, Dordi T, Majerbi B, McPherson M. Media Portrayals of Net Zero: Stakeholders' Perspectives and Climate Solutions Framing. Environ Commun. 2026;20:1–21. https://doi.org/10.1080/17524032.2026.2692085
6. Shakespear M, Koop-Monteiro Y, Stoddart MCJ, Tindall DB, Jorgenson AK. Fossil fuel phaseout, renewable energy, and just transition discourse at COP26 and COP28: A discourse network analysis of Instagram posts. Energy Res Soc Sci. 2026;131:104724. https://doi.org/10.1016/j.erss.2026.104724
7. Misuraca M, D'Aniello L, Spano M. Energy transition as the core of climate discourse: Mapping social representations in digital public debate. Energy Res Soc Sci. 2026;133:104945. https://doi.org/10.1016/j.erss.2026.104945
8. Walters R, Farrelly M, Novalia W, Raven R. Headlining justice from coalfields to clean futures: How the Australian newsprint media frames a just energy transition. Energy Res Soc Sci. 2025;125:104131. https://doi.org/10.1016/j.erss.2025.104131
9. Chen CY. From climate crisis to energy crisis: foster public support for renewable energy transition through framing. Environ Polit. 2025;35(5):998–1022. https://doi.org/10.1080/09644016.2025.2543622
10. Edwards G. Exploring discourses of climate delay in energy transition debates in national media. Clim Dev. 2025;17:1–14. https://doi.org/10.1080/17565529.2025.2574078
11. Putri SH, Yusian SA, Dignum M, Pesch U, Correljé A. Framing Geothermal Energy in Indonesia: A Media Analysis in A Country with Huge Potential. Environ Commun. 2022;16(7):993–1001. https://doi.org/10.1080/17524032.2022.2144403
12. Zhong S, Yang L, Papageorgiou DJ, Su B, Ng TS, Abubakar S. Accelerating ASEAN's energy transition in the power sector through cross-border transmission and a net-zero 2050 view. iScience. 2025;28(1):111547. https://doi.org/10.1016/j.isci.2024.111547
13. Zhong S, Su B, Papageorgiou D, Yeung FS, Ng TS, Abubakar S. Charting net-zero pathways for ASEAN's energy sector. PNAS Nexus. 2026;5(1):pgaf389. https://doi.org/10.1093/pnasnexus/pgaf389
14. Do TN. Insider perspectives on Southeast Asia's clean energy transition. Asia Pac Policy Stud. 2024;11(2):e390. https://doi.org/10.1002/app5.390
15. Do TN, Burke PJ. Phasing out coal power in two major Southeast Asian thermal coal economies: Indonesia and Vietnam. Energy Sustain Dev. 2024;80:101451. https://doi.org/10.1016/j.esd.2024.101451
16. Do TN, Burke PJ. Phasing out coal power in a developing country context: Insights from Vietnam. Energy Policy. 2023;176:113512. https://doi.org/10.1016/j.enpol.2023.113512
17. Fazeli R, Do TN, Overland I, Pradnyaswari I. Southeast Asia faces high stranded asset risk from coal power investments. Energy Strateg Rev. 2025;62:101971. https://doi.org/10.1016/j.esr.2025.101971
18. Sekaringtias A, Verrier B, Cronin J. Untangling the socio-political knots: a systems view on Indonesia's inclusive energy transitions. Energy Res Soc Sci. 2023;95:102911. https://doi.org/10.1016/j.erss.2022.102911

19. Vakulchuk R, Overland I, Suryadi B. ASEAN's energy transition: how to attract more investment in renewable energy. Energy Ecol Environ. 2023;8(1):1–16. https://doi.org/10.1007/s40974-022-00261-6
20. Derouez F, Ifa A. Assessing the Sustainability of Southeast Asia's Energy Transition: A Comparative Analysis. Energies. 2025;18(2):287. https://doi.org/10.3390/en18020287
21. Fahim KE, De Silva LC, Hussain F, Shezan SA, Yassin H. An Evaluation of ASEAN Renewable Energy Path to Carbon Neutrality. Sustainability. 2023;15(8):6961. https://doi.org/10.3390/su15086961
22. Yang F, Li C. The Status Quo, Dilemma, and Transformation Path of the Carbon Neutrality-Related Policy of the ASEAN. Sustainability. 2024;16(3):1348. https://doi.org/10.3390/su16031348
23. Mukhtar M, Abdulqadir IA. Growth Threshold Effect on Renewable Energy Transition in Southeast Asian Economies. Sustainability. 2025;17(21):9405. https://doi.org/10.3390/su17219405
24. Gulagi A, Alcanzare M, Bogdanov D, Breyer C. Accelerating the transition from coal to renewables in Indonesia to achieve a net-zero energy system. IET Renew Power Gener. 2024;18(14):2565–2582. https://doi.org/10.1049/rpg2.13188
25. Tee MX, Miao S, Rasli A, Li D, Liang W, Xue G. Green Energy Vehicles in the Global Energy Transition: Technological Evolution, Critical Materials, and Sustainable Supply Chains—A Bibliometric and Thematic Review (2015–2025). Energies. 2026;19(20):4878. https://doi.org/10.3390/en19204878
26. Bucur C, Tudorica B, Andrei JV, Dusmanescu D, Paraschiv D, Teodor C. Sentiment analysis of global news on environmental issues: insights into public perception and its impact on low-carbon economy transition. Front Environ Sci. 2024;12:1360304. https://doi.org/10.3389/fenvs.2024.1360304
27. Sanford M, Painter J. Divergences between mainstream and social media discourses after COP26, and why they matter. Oxf Open Clim Chang. 2024;4(1):kgae006. https://doi.org/10.1093/oxfclm/kgae006
28. Paiement P, Webster E, Anderson R. After COP26: Appraising the transnational climate regime. Transnatl Leg Theory. 2022;13(2–3):157–164. https://doi.org/10.1080/20414005.2023.2170758
29. Painter J, Wetts R, Loy L. Climate delay discourses present in global mainstream television coverage of the IPCC's 2021 report. Commun Earth Environ. 2023;4(1):118. https://doi.org/10.1038/s43247-023-00760-2
30. Hayes S, Gabbatiss J, Butler C. From climate scepticism to discourses of delay in UK editorials. Public Underst Sci. 2025;34(7):832–851. https://doi.org/10.1177/09636625251315446
31. Lamb WF, Mattila MJ. Intent in climate delay discourses. Glob Sustain. 2025;8:e31. https://doi.org/10.1017/sus.2025.10031
32. Dasandi N, Graham H, Hudson D, Jankin S, vanHeerde-Hudson J, Watts N. Positive, global, and health or environment framing bolsters public support for climate policies. Commun Earth Environ. 2022;3(1):239. https://doi.org/10.1038/s43247-022-00571-x
33. Vikström S, Mervaala E, Kangas HL, Lyytimäki J. Framing climate futures: the media representations of climate and energy policies in Finnish broadcasting company news. J Integr Environ Sci. 2023;20(1):2178464. https://doi.org/10.1080/1943815X.2023.2178464
34. Valquaresma A, Batel S, Afonso AI, Guerra R, Silva L. The Renewable Energy Transition and 'the People' – Exploring the Intersections of Right-wing Populism and the Renewable Energy Transition in Portuguese Media Discourses. Environ Commun. 2024;18(7):847–861. https://doi.org/10.1080/17524032.2024.2326423
35. Thielemann N. Framing the Energy Transition: The Case of Poland's Turów Lignite Mine. Int J Strateg Commun. 2023;17(4):338–356. https://doi.org/10.1080/1553118X.2023.2234881
36. Magnusson D, Sperling K, Veenman S, Oteman M. News Media Framing of Grassroots Innovations in Denmark, the Netherlands and Sweden. Environ Commun. 2021;15(5):641–662. https://doi.org/10.1080/17524032.2021.1880460
37. Kirchner T, Guenther L, Brüggemann M. What Journalists Worry and Write About: Comparing Journalists' Frames of Climate Futures with Respective News Frames. Journal Pract. 2026;20(4):1325–1343. https://doi.org/10.1080/17512786.2024.2412292

38. Guenther L, Jörges S, Mahl D, Brüggemann M. Framing as a Bridging Concept for Climate Change Communication: A Systematic Review Based on 25 Years of Literature. Commun Res. 2024;51(4):367–391. https://doi.org/10.1177/00936502221137165
39. Guenther L, Ibrahim Y, Lüdemann J, O'Brian CI, Pavenstädt CN, Sievert IJ, et al. Social Constructions of Climate Futures: Reframing Science's Harmful Impact Frame Across News Media, Social Movements, and Local Communities. Environ Commun. 2024;18(3):322–338. https://doi.org/10.1080/17524032.2024.2305827
40. Fuoli M, Beelitz A. Corporate buzzword or genuine commitment? A corpus-assisted analysis of corporate 'net-zero' pledges by major global corporations. Appl Corpus Linguist. 2025;5(3):100142. https://doi.org/10.1016/j.acorp.2025.100142
41. Fuoli M, Beelitz A. Framing the path to net zero: A corpus-assisted discourse analysis of sustainability disclosures by major corporate emitters, 2011–2020. Int J Corpus Linguist. 2024;29(2):181–212. https://doi.org/10.1075/ijcl.22123.fuo
42. Fankhauser S, Smith SM, Allen M, Axelsson K, Hale T, Hepburn C, et al. The meaning of net zero and how to get it right. Nat Clim Chang. 2022;12(1):15–21. https://doi.org/10.1038/s41558-021-01245-w
43. Wang X, Lo K. Just transition: A conceptual review. Energy Res Soc Sci. 2021;82:102291. https://doi.org/10.1016/j.erss.2021.102291
44. van Uffelen N. Revisiting recognition in energy justice. Energy Res Soc Sci. 2022;92:102764. https://doi.org/10.1016/j.erss.2022.102764
45. Coan TG, Boussalis C, Cook J, Nanko MO. Computer-assisted classification of contrarian claims about climate change. Sci Rep. 2021;11(1):22320. https://doi.org/10.1038/s41598-021-01714-4