ISSN (онлайн): 2663-8959 | ISSN (друк): 2308-6912
УДК 339.9:004.85:502.131.1(477)
DOI: https://doi.org/10.17721/apmv.2026.167.1.210-220

UKRAINE'S GREEN RECOVERY IN THE INTERNATIONAL MEDIA: THE POSSIBILITIES AND LIMITS OF FINBERT-BASED SENTIMENT ANALYSIS

ЗЕЛЕНА ВІДБУДОВА УКРАЇНИ В МІЖНАРОДНИХ МЕДІА: МОЖЛИВОСТІ ТА ОБМЕЖЕННЯ АНАЛІЗУ ТОНАЛЬНОСТІ НА ОСНОВІ FINBERT

Vasyl Namoniuk
PhD in Economics, Associate Professor, Head of the Chair of International Finance, Educational and Scientific Institute of International Relations, Taras Shevchenko National University of Kyiv,
e-mail: vasyl.namoniuk@knu.ua
ORCID ID: https://orcid.org/0000-0002-9454-6658

Cheslava Namoniuk
PhD in Political Science, Associate Professor of the Department of International Organisations and Diplomatic Service, Taras Shevchenko National University of Kyiv.
e-mail: cheslava.namoniuk@knu.ua
ORCID ID: https://orcid.org/0009-0002-5134-253X

Василь Намонюк
кандидат економічних наук, доцент, завідувач кафедри міжнародних фінансів Навчально-наукового інституту міжнародних відносин Київського національного університету імені Тараса Шевченка,
e-mail: vasyl.namoniuk@knu.ua;
ORCID ID: https://orcid.org/0000-0002-9454-6658

Чеслава Намонюк
Кандидат політичних наук, доцент кафедри міжнародних організацій і дипломатичної служби Навчально-наукового Інституту міжнародних відносин Київського національного університету імені Тараса Шевченка.
e-mail: cheslava.namoniuk@knu.ua
ORCID ID: https://orcid.org/0009-0002-5134-253X

Abstract. This paper examines the tone of international media coverage of Ukraine's green recovery and asks how far that measurement depends on choices made in constructing it. The corpus consists of 456 English-language headlines retrieved from Google News RSS across eight queries between November 2019 and August 2026, classified with FinBERT (ProsusAI/finbert). The overall tone is moderately positive: 35.31% of headlines are positive, 57.02% neutral and 7.68% negative, giving a sentiment index of +0.276 (95% CI [+0.222, +0.331]). Two dimensions of genuine variation are identified. Institutional publishers are three times more likely than media outlets to produce a positive headline (+0.581 against +0.245; Fisher exact p = 0.00062). Tone also differs sharply by retrieval frame, from +0.545 for reconstruction-and-energy coverage to exactly 0.000 for climate finance, where positive and negative headlines are in perfect balance (chi-square = 58.84; $p < 0.0001$). The apparent recovery of tone over the course of the war, by contrast, does not survive scrutiny: the aggregate rise from +0.016 in 2022 to +0.363 in 2025-2026 is driven by a shift in which queries return content in which period, and disappears when the retrieval frame is held constant (p = 0.121). Two further demonstrations point the same way. Removing the publisher name that Google News appends to every headline changes 14.69% of classifications and moves the index from +0.217 to +0.276; and classifying topic by keyword match rather than by retrieval query turns a significant effect into a non-significant one. Applying financial language models to reconstruction discourse therefore requires full reporting of preprocessing, explicit control for retrieval composition and validation against a hand-coded subsample.

Keywords: green recovery, Ukraine, sentiment analysis, FinBERT, sustainable finance, media discourse, natural language processing, post-war reconstruction.

Анотація. У статті досліджено тональність міжнародного медійного висвітлення зеленої відбудови України та з'ясовано, наскільки результат такого вимірювання залежить від рішень, ухвалених у процесі його побудови. Корпус становлять 456 англомовних заголовків, зібраних

https://doi.org/10.17721/apmv.2026.167.1.210-220
із Google News RSS за вісьмома запитами з листопада 2019 року до серпня 2026 року та класифікованих моделлю FinBERT (ProsusAI/finbert). Загальний тон висвітлення є помірно позитивним: 35,31% заголовків позитивні, 57,02% нейтральні і 7,68% негативні, індекс сентименту становить +0,276 (95% ДІ [+0,222; +0,331]). Виявлено два виміри справжньої варіації. Інституційні джерела втричі частіше за медіа продукують позитивні заголовки (+0,581 проти +0,245; точний тест Фішера $p = 0,00062$). Тональність також різко різниться залежно від пошукової рамки — від +0,545 для висвітлення відбудови та енергетики до рівно 0,000 для кліматичного фінансування, де позитивні й негативні заголовки перебувають у цілковитій рівновазі ($\chi^2 = 58,84$; $p < 0,0001$). Натомість позірне поліпшення тону впродовж війни перевірки не витримує: агрегатне зростання з +0,016 у 2022 році до +0,363 у 2025–2026 роках зумовлене зміною того, які запити повертають матеріал у який період, і зникає, коли пошукову рамку тримати сталою ($p = 0,121$). Ще дві перевірки вказують у той самий бік. Вилучення назви видання, яку Google News додає до кожного заголовка, змінює 14,69% класифікацій і зсуває індекс з +0,217 до +0,276; а класифікація тем за ключовими словами замість пошукового запиту перетворює значущий ефект на незначущий. Застосування фінансових мовних моделей до дискурсу відбудови потребує повного опису передобробки, явного контролю за композицією вибірки та валідації на вручну закодованій підвибірці.

Ключові слова: зелена відбудова, Україна, аналіз тональності, FinBERT, стале фінансування, медійний дискурс, обробка природної мови, повоєнна реконструкція.

Introduction. The scale of Ukraine's reconstruction task is now measured in hundreds of billions of dollars, and the terms on which that money arrives are increasingly conditional. The Lugano Declaration of 2022, the successive Ukraine Recovery Conferences and, most consequentially, the European Union's Ukraine Facility have tied financial support to sustainability criteria, so that green recovery has moved from aspiration to something closer to a disbursement condition. The architecture assembled to deliver it — blended public and private instruments, donor coordination mechanisms, conditionality frameworks — has been mapped in some detail (Becker et al., 2022; Namoniuk, 2026), and what began as a normative preference among donors has become an operating constraint on how reconstruction is designed, procured and reported.

The information environment in which that agenda unfolds has attracted far less attention, although it bears directly on the financing question. Media coverage shapes how sovereign and corporate risk is priced, how donor parliaments justify continued expenditure to their electorates, and how private capital reads the plausibility of a green rebuild conducted under missile attack. The evidence that press tone carries information beyond fundamentals is long established (Tetlock, 2007), and in the climate domain specifically, indices built from newspaper text have been shown to track asset prices closely enough to support hedging strategies (Engle et al., 2020). Favourable coverage tends to compress perceived risk premia and sustain political commitment; sceptical or hostile coverage does the reverse. For a reconstruction effort dependent on sustained external financing over a decade or more, the tone of international reporting is not merely descriptive of conditions but partly constitutive of them.

To the best of our knowledge, no study has systematically measured how the green recovery narrative is framed in the international press. The relevant literature divides along an unhelpful line. Financial text analysis has concentrated on corporate disclosure in developed markets, while media research on the war and on post-crisis recovery has proceeded largely through framing analysis, whether qualitative or computational, without quantifying valence (Ptaszek et al., 2024; Ibrahim et al., 2025). Neither tradition has addressed a discourse that sits at their intersection, one in which war reporting, climate policy and financial-instrument coverage are interleaved within the same corpus and often within the same sentence.

Literature review. The analysis of financial text has moved through three broad phases. Dictionary-based approaches dominated the first, following Loughran and McDonald (2011), whose demonstration that general-purpose sentiment word lists systematically misclassify financial language established the case for domain-specific resources; roughly three-quarters of the words flagged as negative in the Harvard dictionary are not negative in a financial context. Their subsequent survey (Loughran & McDonald, 2016) documents the breadth of adoption but also the method's

ceiling: word lists cannot resolve negation, cannot weight by syntactic position, and treat every occurrence of a term identically regardless of what surrounds it. Machine-learning classifiers addressed some of this at the cost of requiring labelled training data, with Malo et al. (2014) contributing the Financial PhraseBank, a set of expert-annotated sentences from financial news labelled from the standpoint of a retail investor, which remains the standard benchmark in the field.

The transformer architecture introduced by Devlin et al. (2019) altered the terms of the problem by making contextual representation learnable at scale, and domain-adapted variants followed quickly. Araci (2019) produced the model distributed as ProsusAI/finbert, further pre-training BERT on the Reuters TRC2 financial corpus before fine-tuning on the Financial PhraseBank, and reporting classification accuracy of approximately 0.86, a substantial improvement over both dictionary methods and generic BERT. A parallel model developed by Huang et al. (2023) took a comparable approach with a different pre-training corpus, and has since been extended by its authors into task-specific releases for ESG categorisation and forward-looking statement detection. These are distinct models with distinct training regimes, frequently conflated in the applied literature under the single label FinBERT; the distinction matters for replication and is observed explicitly here. The same logic of domain adaptation has been applied beyond finance proper: ClimateBERT (Webersinke et al., 2021) was pre-trained on climate-related text and used to show that corporate climate disclosures are systematically selective, with firms reporting cheap-to-disclose items while omitting the commitments that would bind them (Bingler et al., 2022). That result is instructive here in a second sense: a domain-adapted classifier proved able to detect a gap between declared ambition and concrete undertaking, which is close to the distinction this study finds between reconstruction framing and instrument-level reporting.

Applications of sentiment measurement in finance and economics are by now extensive, beginning with Tetlock (2007), who established that pessimism in financial press coverage predicts downward pressure on market prices with subsequent reversal. The literature has since extended to corporate disclosure events, where the textual tone of earnings conference calls carries information incremental to the reported figures (Price et al., 2012); to central bank communication, where even the vocal and textual characteristics of policy announcements move markets (Gorodnichenko et al., 2023); and to social media, where intraday investor sentiment extracted from message volume predicts return patterns (Renault, 2017). A related strand examines media sentiment in relation to sustainable finance. Green bonds trade at a small but persistent yield discount attributable to pro-environmental investor preferences (Zerbib, 2019), a premium whose formation and measurement have been analysed in the Ukrainian and tokenised-instrument context by Namoniuk and Matei (2025), and corporate green bond issuance is associated with improved environmental performance and positive market reaction (Flammer, 2021). What unites these applications is that they operate within the domain for which the models were built: corporate actors, financial instruments, market outcomes.

Reconstruction discourse does not fit that description, and here the literature thins considerably. Research on media coverage of conflict and post-crisis recovery has proceeded largely within the framing tradition inaugurated by Entman (1993), concerned with how selection and salience construct problem definitions rather than with quantifying valence. Recent computational work on the Russia–Ukraine war has advanced that tradition methodologically — text mining of Russian and Ukrainian agency coverage has identified a rigid technical frame on one side and an evolving moralising frame on the other (Ptaszek et al., 2024). Unsupervised modelling of two years of international coverage has traced shifts in thematic content and in the portrayal of principal actors (Ibrahim et al., 2025) — yet both identify frames rather than measure tone, and neither addresses reconstruction financing. Work specifically on Ukraine's reconstruction has concentrated on instruments and institutional mechanisms (Becker et al., 2022; Namoniuk, 2026) rather than on the surrounding discourse, and studies of the country's post-war energy trajectory have modelled scenarios rather than their reception (Lukash & Namoniuk, 2024). The gap is therefore twofold. Substantively, the tone of international coverage of Ukraine's green recovery has not been measured. Methodologically, the behaviour of financial language models on reconstruction text, a register in which the vocabulary of finance appears but its semantics do not reliably hold, has not been examined, despite the growing tendency to deploy such models wherever financial vocabulary is present.

https://doi.org/10.17721/apmv.2026.167.1.210-220

The aim of this paper is to establish the tone of English-language media coverage of Ukraine's green recovery and to determine how far that measurement depends on the choices made in constructing it. The first task is descriptive: what is the net tone of the corpus, and along which dimensions does it genuinely vary? The second is diagnostic: whether the patterns that emerge survive statistical testing, whether they survive the removal of confounds introduced by retrieval itself, and what systematic errors arise when a model trained on corporate financial text is turned on reconstruction reporting.

Data and methodology. The corpus consists of English-language news headlines concerning Ukraine's green recovery, retrieved from the Google News RSS endpoint. Eight queries were used to span the discourse: Ukraine reconstruction energy, Ukraine “energy transition”, Ukraine “renewable energy”, Ukraine “green recovery”, Ukraine “climate finance”, Ukraine “green bonds”, Ukraine “green reconstruction” and Ukraine “sustainable reconstruction”. Retrieval returned 456 records, each carrying a headline, a publication timestamp, the originating query and a resolved link; headline strings and links are unique throughout, so no item was returned by more than one query and each carries an unambiguous retrieval frame. The earliest item dates from 7 November 2019 and the latest from 20 August 2026. The temporal distribution is uneven but supports period comparison: 15 headlines fall before the full-scale invasion (2019–2021), 62 in 2022, 87 across 2023–2024 and 292 in 2025–2026. The concentration in recent years reflects how RSS retrieval works, since the endpoint returns what is currently indexed rather than a complete archive, and the earliest years should be read as a thinning trace rather than a representative sample. This property of the feed turns out to matter for more than sample size, as the results show.

Google News returns each headline with the publishing outlet appended after a hyphen, in the form Headline – Publisher. This is a delivery convention of the feed rather than part of the headline as published, and it was removed before classification by splitting at the final hyphen-space sequence; the publisher was retained as a variable in its own right, yielding 235 distinct outlets across the 456 records. The step is treated as substantive rather than as housekeeping, because leaving the publisher attached changes the classification of roughly one headline in seven. Both versions were therefore scored and the comparison is reported below. Apostrophes and quotation marks were normalised to their ASCII equivalents and HTML entities decoded; no further normalisation was applied, since casing, punctuation and named entities carry information the model was trained to use.

Sentiment was assigned with FinBERT in the ProsusAI release (Araci, 2019), run through the Hugging Face Transformers library at default settings. The model applies the BERT-base architecture of twelve encoder layers, twelve attention heads and 768 hidden units, further pre-trained on the Reuters TRC2 financial corpus and fine-tuned on the Financial PhraseBank (Malo et al., 2014), where it reports accuracy of approximately 0.86. A second and distinct model circulates under the same name (Huang et al., 2023), trained on a different corpus; the release used here is specified so that the analysis can be reproduced. For each headline the model returns a probability distribution over three classes. The arg-max was taken as the assigned label and the full distribution retained, since the probability mass on the runner-up class proves informative when examining misclassification. No headline approached the 512-token input limit, so truncation was never triggered.

Four grouping variables are used. The retrieval query is the frame under which each headline was found, and is the closest thing the data contain to an externally defined topic, since it reflects how the discourse was searched rather than how an individual headline happens to be worded. As an alternative and, in applied work, more common operationalisation, headlines were also assigned to four keyword categories by string matching on the headline itself: green recovery, energy transition, climate finance and a residual category, the last of which absorbs 343 of 456 records. Publishers were sorted into institutional sources, corporate sources and media outlets by matching on the outlet name; the institutional group comprises United Nations bodies, international financial institutions, government agencies and development organisations, and the corporate group comprises company-operated domains such as those of DTEK and ORLEN. Finally, records were grouped into four periods keyed to the war and the reconstruction-financing timetable: pre-invasion (2019–2021), the invasion year (2022), the Ukraine Recovery Conference and Ukraine Facility period (2023–2024) and the current period (2025–2026). Net tone is summarised by a sentiment index defined as the positive share minus the negative share, bounded by −1 and +1. Arithmetic differences between index

values are not treated as self-evidently meaningful. Independence between each grouping variable and sentiment class was tested by Pearson's chi-square, with effect size reported as Cramér's V, and pairwise contrasts assessed by Fisher's exact test where cell counts are sparse. Interval estimates for class proportions use the Wilson method; intervals for the index, a difference of two dependent proportions with no convenient closed form, were obtained by non-parametric bootstrap with 10,000 resamples. All tests are two-sided at the 5% level.

Main results of the research. Across the 456 headlines, 260 (57.02%) were classified neutral, 161 (35.31%) positive and 35 (7.68%) negative, giving a sentiment index of +0.276 with a bootstrap 95% confidence interval of [+0.222, +0.331]. Positive coverage exceeds negative by a ratio of roughly four and a half to one, and neither pole is a fragile estimate: the Wilson interval places the positive share within [0.311, 0.398] and the negative share within [0.056, 0.105]. Set against the period the corpus covers, that asymmetry is the notable feature. These seven years include a full-scale invasion, the systematic destruction of generating and transmission capacity, repeated winter blackouts and the suspension of payments on state-linked green bonds. Coverage of the green recovery agenda nonetheless carries four and a half positive headlines for every negative one, which is not what the underlying events alone would predict. The dominance of the neutral class calls for less comment, since headline conventions favour declarative construction and a model calibrated on financial reporting assigns most of that to neutral; the informative quantity is the balance between the two evaluative poles.

Two dimensions of genuine variation lie beneath that average. The first is who is speaking. Institutional publishers return an index of +0.581 across 43 records, against +0.245 across 404 media outlets, with corporate publishers higher still at +0.667 on nine records. The three-way association is significant (chi-square = 18.44, df = 4, p = 0.00101), and the contrast in positive share between institutional and journalistic sources survives an exact test with corporate publishers excluded (Fisher exact p = 0.000323, odds ratio 3.26). Institutional sources are, in round terms, three times more likely to produce a positive headline than media outlets covering the same subject. The consequence is methodological as much as substantive. Google News aggregates press releases from the United Nations Development Programme, UNECE, UNIDO and the EBRD alongside reporting from Reuters, Bloomberg and Ukrainian outlets, and presents them in a single undifferentiated feed. Any corpus drawn from such a source is a blend of promotional and journalistic material, and its overall tone depends partly on the mix: institutional sources make up 9.4% of records here, and a different query set or retrieval date would produce a different proportion and hence a different headline figure, without anything having changed in the discourse itself.

The second dimension is how the discourse is searched. Table 1 reports the sentiment index by retrieval query.

The association is strong and highly significant (chi-square = 58.84, df = 14, p < 0.0001; Cramér's V = 0.254), and the extremes are far apart. Coverage retrieved by reconstruction energy runs at +0.545 with barely any negative content, while coverage retrieved by climate finance is exactly balanced, positive and negative shares standing at an identical 20.41%. The contrast between these two in positive share gives a Fisher exact p of 0.000047 and an odds ratio of 4.88. The gradient has an interpretable shape.

Table 1

Sentiment index by retrieval query

Query	n	Negative	Positive	Index	95% CI
Ukraine reconstruction energy	99	1.01%	55.56%	+0.545	[+0.444, +0.646]
Ukraine “sustainable reconstruction”	15	0.00%	46.67%	+0.467	[+0.200, +0.733]
Ukraine “renewable energy”	79	6.33%	43.04%	+0.367	[+0.228, +0.494]
Ukraine “green recovery”	49	2.04%	36.73%	+0.347	[+0.204, +0.490]
Ukraine “green reconstruction”	23	0.00%	21.74%	+0.217	[+0.044, +0.391]
Ukraine “green bonds”	48	12.50%	29.17%	+0.167	[−0.021, +0.333]

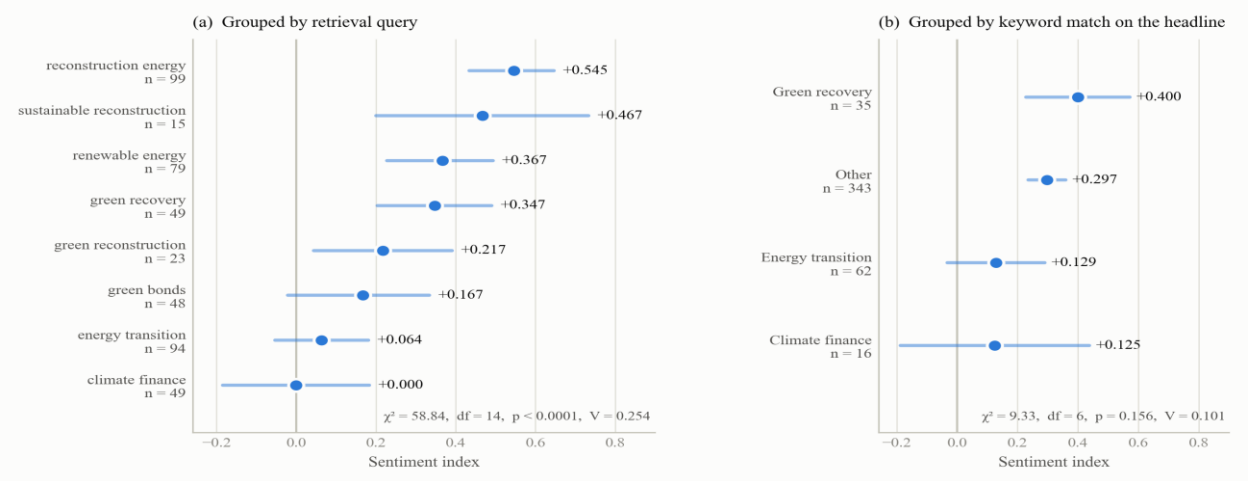
Ukraine “energy transition”	94	12.77%	19.15%	+0.064	[−0.053, +0.181]
Ukraine “climate finance”	49	20.41%	20.41%	0.000	[−0.184, +0.184]

Source: calculated by the authors on the basis of the FinBERT-classified corpus.

Queries framed around rebuilding and physical capacity attract the language of commitment: funding announced, partnerships signed, capacity installed. Queries framed around financial instruments and policy processes attract the language of execution, where the news is delay, default, dispute and shortfall. The same underlying agenda produces markedly different tone depending on which end of it is being reported, which suggests that confidence in the vision runs considerably ahead of confidence in the machinery meant to deliver it.

Figure 1

Sentiment index by retrieval query and by keyword match



Source: calculated by the authors on the basis of the FinBERT-classified corpus.

That finding, however, is available only under one of the two ways of dividing the corpus by topic, and the comparison between them is instructive. Figure 1 sets the retrieval-query grouping beside the keyword grouping applied to the identical 456 headlines. Sorted by keyword match, the four categories run from +0.400 for green recovery through +0.297 for the residual category and +0.129 for energy transition to +0.125 for climate finance. The ordering looks similar and a reader given only those numbers would draw much the same conclusion, but the association is not significant: chi-square = 9.33, df = 6, p = 0.156, with Cramér's V of 0.101, and restricting to positive against non-positive gives chi-square = 3.08, df = 3, p = 0.380. This follows from how the rule operates. Headlines rarely contain the phrase by which they were found; the item “Norway and UNDP strengthen partnership to accelerate Ukraine's energy recovery” was returned by a green recovery query but contains none of the trigger phrases and therefore lands in the residual category. The keyword rule consigns 75.2% of the corpus to that category and in doing so pools the sharply divergent sub-corpora that the query variable keeps apart. Post-hoc keyword classification of short texts is common in applied sentiment work and is usually presented as a neutral operationalisation of topic. It is not: here it destroys an effect that is present in the data, significant at $p < 0.0001$ and of moderate size, and replaces it with a non-significant ordering that reads plausibly and would in all likelihood have been reported as substantive.

An apparent change in tone over the course of the war does not survive the removal of a confound introduced by retrieval itself. Table 2 reports sentiment by period, together with the share of each period's headlines that came from queries scoring above the corpus mean.

Table 2

Sentiment and corpus composition by period

Period	n	Neg.	Neut.	Pos.	Index	95% CI	Above-mean q.
Pre-invasion (2019–2021)	15	0.00%	53.33%	46.67%	+0.467	[+0.200, +0.733]	0.0%
Invasion year (2022)	62	17.74%	62.90%	19.35%	+0.016	[−0.129, +0.161]	4.8%
URC / Facility (2023–2024)	87	11.49%	63.22%	25.29%	+0.138	[+0.011, +0.264]	31.0%
Current (2025–2026)	292	4.79%	54.11%	41.10%	+0.363	[+0.298, +0.428]	72.6%

Source: calculated by the authors on the basis of the FinBERT-classified corpus. The final column reports the share of each period's headlines returned by queries whose own index exceeds the corpus mean of +0.276.

Taken at face value the trajectory is dramatic and statistically significant (chi-square = 25.84, df = 6, p = 0.0002; Cramér's V = 0.168). In 2022 the index stands at +0.016, indistinguishable from zero, with a negative share of 17.74%; by 2025–2026 it has risen to +0.363 with the negative share down to 4.79%. The obvious reading is that coverage brightened as the financing architecture consolidated, and the temptation is to conclude that the framing of reconstruction changed while conditions on the ground did not.

The final column of Table 2 shows why that conclusion cannot be drawn. The retrieval frame did not hold still across the period. Two of the four highest-scoring queries, reconstruction energy at +0.545 and renewable energy at +0.367, return no material at all before 2025 and together account for 61% of the current period's records. The 2022 sub-corpus, by contrast, is dominated by energy transition and climate finance, the two lowest-scoring frames. The share of each period drawn from above-mean queries runs 0%, 4.8%, 31.0% and 72.6%, and the association between period and query is overwhelming (chi-square = 227.30, df = 21, $p < 0.000001$). Later periods are not merely later; they are composed of different searches.

Figure 2 separates the two effects. Panel (a) shows the compositional shift. Panel (b) sets the aggregate index against the index computed on the three queries present in every period, which hold the retrieval frame constant across 191 records. On that constant frame the series runs +0.467, +0.019, 0.000 and +0.083, and the association with period is not significant (chi-square = 10.09, df = 6, p = 0.121). The upward movement that the aggregate displays is therefore largely compositional. Holding the frame fixed leaves a slight and statistically unsupported drift, not a recovery.

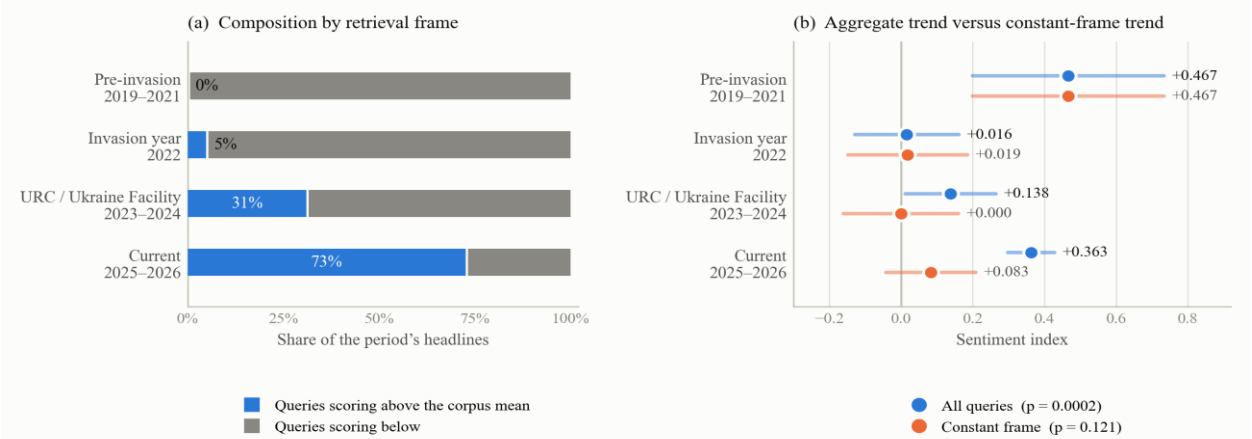
One alternative confound can be dismissed. Because institutional sources are markedly more positive and their share also varied across periods (chi-square = 14.68, df = 3, p = 0.00212), the trajectory might have been a change in the promotional share of the feed. It is not: restricted to media outlets alone the period association remains significant (chi-square = 25.78, df = 6, p = 0.00024). Source mix is not what drives the aggregate pattern; retrieval frame is. The substantive conclusion is correspondingly narrow. This corpus cannot support a claim that international coverage of Ukraine's green recovery grew more favourable as the war continued. What it shows is that the searches which return material about Ukraine's green recovery have themselves shifted towards frames that attract favourable coverage.

The measurement is sensitive to a preprocessing decision that published applications do not ordinarily report. Scoring the same 456 records with the publisher name left attached, as a pipeline built directly on the feed would do, produces 122 positive (26.75%), 311 neutral (68.20%) and 23 negative (5.04%), for an index of +0.217.

Figure 2

Corpus composition by period and the aggregate trend against the constant-frame trend

https://doi.org/10.17721/apmv.2026.167.1.210-220

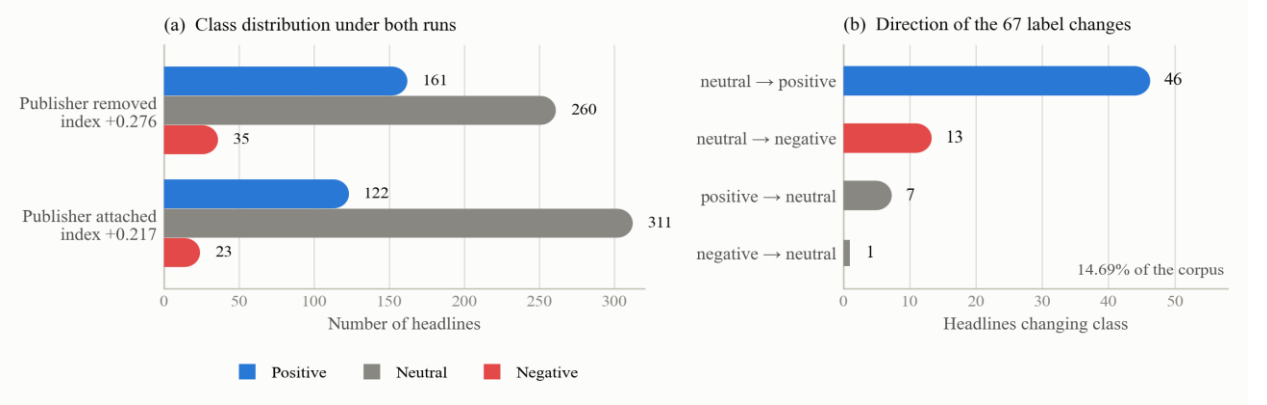


Source: calculated by the authors on the basis of the FinBERT-classified corpus. The constant frame comprises the three queries returning material in every period (n = 191).

The two runs agree on 85.31% of labels; 67 headlines, or 14.69% of the corpus, change class depending on whether the outlet name is present. Figure 3 reports both the class distributions and the direction of the changes, which are not symmetrical: 46 run from neutral to positive, 13 from neutral to negative, 7 from positive to neutral and 1 from negative to neutral. Attaching the publisher pulls the corpus towards neutrality, converting evaluative headlines into non-evaluative ones far more often than the reverse. The mechanism is visible in individual cases. “DTEK Renewables wins the Climate Bonds Initiative's Green Bond Pioneer Award” is classified positive once the outlet is removed but neutral when the string continues “– dtek.com”; “Global sustainable bond issuance takes Q1 hit on Ukraine crisis” is negative alone and neutral once “– Reuters” is appended. The added tokens are institutional names carrying no valence, and in a text of a dozen words they dilute the evaluative signal. The gap between +0.217 and +0.276 is of the same order as the substantive differences reported above, which means an unreported preprocessing choice can be as consequential as the phenomenon under study.

The classifications themselves also err systematically rather than randomly. Three mechanisms account for most of it. The model treats rising quantities as favourable, which holds in the corpus it was trained on and inverts here: “Ukraine reconstruction estimate jumps 12% to $588 billion, World Bank says” is classified positive at 0.938, “Chart of the week: Ukraine's reconstruction needs rise 12% to $588 billion” positive at 0.947, and “Energy costs rise to 75% in Ukraine's greenhouse production” positive at 0.613, though all three report deterioration. It fails on negation, classifying “Russia's invasion fails to prevent progress in Ukraine's energy sector” as negative at 0.835 by resolving on the salient negative tokens. And it responds to isolated lexical markers detached from context, flagging headlines containing blackout, power cut or shortage as negative irrespective of the surrounding proposition, which is precisely the failure for which dictionary methods were criticised (Loughran & McDonald, 2011), reappearing in a contextual model operated outside its training domain.

Figure 3
Effect of publisher-name removal on classification



Source: calculated by the authors on the basis of two FinBERT runs over the same 456 headlines.

These errors do not cancel, and they are not evenly distributed. Thirty headlines, 6.6% of the corpus, contain the cost-and-damage vocabulary on which the first mechanism operates, and they are classified positive at 46.7% against 34.5% for the remainder. Their share rises from 0% in 2019–2021 and 1.6% in 2022 to 9.2% in 2025–2026 (chi-square test of independence, $p = 0.023$), and varies by frame from 14.1% under reconstruction energy and 12.5% under green bonds down to 2.1% under energy transition and 2.0% under green recovery ($p = 0.006$). The bias therefore concentrates in the most recent period and in the highest-scoring frames, which is to say in exactly the cells that drive the aggregate figure upward. The corpus index of +0.276 should be read as an upper bound rather than as an unbiased estimate, and the comparative results cannot be assumed to be insulated from the level effect in the way that a uniformly distributed error would allow. Quantifying the bias requires manual coding of a stratified subsample against which the classifier can be scored; that is identified here as the necessary next step rather than claimed as performed.

Conclusions. International coverage of Ukraine's green recovery is, on this measure, supportive: an index of +0.276 with four and a half positive headlines for every negative one, sustained across a period that includes invasion, systematic attacks on the energy system and a debt restructuring. The tone is not uniform, and the two dimensions along which it varies are informative. Institutional publishers are three times more likely than journalists to produce a positive headline, so that the apparent favourability of any aggregator-drawn corpus depends partly on how much promotional material the feed happens to contain. More consequentially, tone tracks the framing of the subject: coverage of reconstruction and physical energy capacity runs at +0.545, while coverage of climate finance sits at exactly 0.000, with positive and negative headlines in perfect balance, and green bonds and energy transition are only marginally better. Confidence in the vision runs well ahead of confidence in the instruments meant to realise it, and that is where a policy reader should expect the aggregate to turn first.

The second conclusion is methodological and, in our assessment, the more transferable of the two. Three demonstrations in this study show the same thing from different angles. Removing the publisher name that the feed appends to every headline changes 14.69% of classifications and moves the corpus index by 0.059, a shift comparable to the substantive differences the study set out to measure. Classifying topic by keyword match on the headline rather than by retrieval query consigns three-quarters of the corpus to a residual category and converts an effect significant at $p < 0.0001$ into a non-significant ordering that nevertheless reads plausibly. And an apparent recovery of tone across the war, significant at $p = 0.0002$ and available for exactly the kind of narrative interpretation that such a result invites, dissolves once the retrieval frame is held constant ($p = 0.121$), because the searches that return material about Ukraine's green recovery have themselves migrated towards frames that attract favourable coverage. In each case the corpus, the model and the metric are identical, and the finding depends entirely on a decision that published applications do not ordinarily report.

These findings do not discredit such tools; they specify what their use requires. Preprocessing must be reported in full, because the operations that appear to be housekeeping are not. Any disaggregated comparison must be tested rather than read off a ranking of point estimates. Where the corpus is assembled by retrieval, the composition of that retrieval must be treated as a variable and

controlled for, since a change in what the search returns is indistinguishable, in the aggregate, from a change in what is being said. And the classifier's output must be validated against a hand-coded subsample, particularly where the vocabulary of the target domain overlaps the training domain while its semantics do not, as the reading of a rise in reconstruction cost estimates as good news illustrates. The availability of pre-trained financial language models has outpaced the scrutiny applied to their output.

The limits of the study bound these claims. It works from headlines rather than full texts and so measures framing rather than argument. It draws on a single aggregator whose retrieval favours recent material, which leaves the pre-2022 sample too thin to characterise and which, as the temporal analysis shows, introduces composition effects that must be modelled rather than assumed away. It covers English only, describing how the story reads to an international audience and saying nothing of Ukrainian, German or Polish reporting. The source-type classification rests on matching outlet names and would benefit from hand-coding. Each of these is addressable by applying the same pipeline to full-text, multi-source data with source region, outlet type and topic as designed strata rather than as variables recovered after the fact, provided the inferential discipline argued for here is carried across with it.

Funding: *no external funding.*

Conflict of Interest: *the authors declare no conflict of interest regarding the publication of this article.*

Data Availability Statement: *not applicable*

References:

Araci, D. (2019). *FinBERT: Financial sentiment analysis with pre-trained language models* (arXiv:1908.10063). arXiv. https://arxiv.org/abs/1908.10063

Becker, T., Eichengreen, B., Gorodnichenko, Y., Guriev, S., Johnson, S., Mylovanov, T., Rogoff, K., & Weder di Mauro, B. (2022). *A blueprint for the reconstruction of Ukraine*. CEPR Press.

Bingler, J. A., Kraus, M., Leippold, M., & Webersinke, N. (2022). Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Research Letters, 47*, 102776. https://doi.org/10.1016/j.frl.2022.102776

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies, 33*(3), 1184–1216. https://doi.org/10.1093/rfs/hhz072

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Flammer, C. (2021). Corporate green bonds. *Journal of Financial Economics, 142*(2), 499–516. https://doi.org/10.1016/j.jfineco.2021.01.010

Gorodnichenko, Y., Pham, T., & Talavera, O. (2023). The voice of monetary policy. *American Economic Review, 113*(2), 548–584. https://doi.org/10.1257/aer.20220129

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research, 40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Ibrahim, M., Wang, B., Xu, M., & Xu, H. (2025). A multidimensional analysis of media framing in the Russia-Ukraine war. *Journal of Computational Social Science, 8*(2), 34. https://doi.org/10.1007/s42001-025-00363-1

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research, 54*(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

Lukash, O., & Namoniuk, V. (2024). Post-war development energy scenarios for Ukraine. In *Positive tipping points towards sustainability* (pp. 101–125). Springer Climate. https://doi.org/10.1007/978-3-031-50762-5_6

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology, 65*(4), 782–796. https://doi.org/10.1002/asi.23062

Namoniuk, V. Ye. (2026). The architecture of green financing for the post-war reconstruction of Ukraine: Instruments and institutional mechanisms. *Investytsiyi: Praktyka ta Dosvid*, (4), 144–149. https://doi.org/10.32702/2306-6814.2026.4.144 [In Ukrainian]

Namoniuk, V., & Matei, V. (2025). Tokenised green bonds in the architecture of sustainable finance and the formation of the greenium. *Actual Problems of International Relations*, (165), 172–181. https://doi.org/10.17721/apmv.2025.165.1.172-181

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance, 36*(4), 992–1011. https://doi.org/10.1016/j.jbankfin.2011.10.013

Ptaszek, G., Yuskiv, B., & Khomych, S. (2024). War on frames: Text mining of conflict in Russian and Ukrainian news agency coverage on Telegram during the Russian invasion of Ukraine in 2022. *Media, War & Conflict, 17*(1), 41–61. https://doi.org/10.1177/17506352231166327

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance, 84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2021). *ClimateBERT: A pretrained language model for climate-related text* (arXiv:2110.12010). arXiv. https://arxiv.org/abs/2110.12010

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance, 98*, 39–60. https://doi.org/10.1016/j.jbankfin.2018.10.012

Received: 05.05.26 / Revised: 24.05.26 / Accepted: 03.06.26 / Published: 30.06.26