Measuring Media Bias Toward Reform Prosecutors: A Multi-Method NLP Analysis of Bay Area News Coverage, 2019–2024

Dvir Yogev

Criminal Law & Justice Center, UC Berkeley

Abstract

Media coverage is the primary channel through which the public evaluates prosecutorial policy, yet whether the press covers reform-oriented prosecutors differently from their traditional counterparts remains untested at scale. I analyze 136,313 Bay Area news articles (2019–2024) using a multi-method NLP pipeline that separately measures emotional tone, evaluative stance, thematic emphasis, narrative framing, and source ecology across five district attorneys spanning the progressive–traditional spectrum. Media bias toward prosecutors is not a unitary phenomenon: its dimensions diverge sharply. Emotional tone surrounding prosecutor mentions is statistically equivalent across ideological groups, yet evaluative stance classification reveals substantially more critical framing of progressive prosecutors, an effect confirmed by within-county comparisons and robust to regression controls. When critical themes were explicitly attributed to a named prosecutor, the differential increases to the study's largest effect. Progressive prosecutors receive accountability framing at nearly twice the rate of traditional counterparts, who are more often covered through human-interest narratives. LLM-based structural extraction reveals a corresponding asymmetry in source ecology: progressive coverage draws disproportionately on advocacy, expert, and politician voices, while traditional coverage relies

more heavily on the prosecutor's own statements. These differentials are event-driven rather than reflecting a stable editorial disposition. The press covers reform prosecutors differently in kind, not in degree: through systematic differences in evaluative framing, sourcing, and narrative structure that are invisible to tonal analysis alone.

1 Introduction

The progressive prosecution movement represents one of the most significant developments in American criminal justice over the past decade. Beginning with the election of reform-minded district attorneys in major cities (including Larry Krasner in Philadelphia (2017), Kim Foxx in Chicago (2016), and Chesa Boudin in San Francisco (2019)) a new generation of prosecutors has sought to reduce mass incarceration, curtail cash bail, decline to prosecute low-level offenses, and hold law enforcement accountable (Bellin, 2020; Sklansky, 2017). These prosecutors have faced intense political opposition, including recall campaigns, legislative challenges, and sustained media scrutiny (Bazelon, 2019). Goldstein (2024) characterizes this resistance as "toplash": elite-driven opposition from wealthy donors, police unions, and political establishments that is distinct from grassroots backlash and specifically targets prosecutors' reform agendas through media campaigns and institutional pressure.

Media coverage plays a crucial role in shaping public attitudes toward criminal justice reform. How prosecutors are portrayed in the press, whether through sympathetic human-interest narratives or critical accountability frames, can influence public support for reform policies and, ultimately, electoral outcomes (Iyengar, 1991; Entman, 1993). More broadly, research in political economy and criminal justice communication suggests that media exposure can shape democratic accountability, policy judgments, and legitimacy perceptions toward legal institutions (Prat and Strömberg, 2007; Graziano, 2019; Intravia et al., 2018). The recall of San Francisco District Attorney Chesa Boudin in June 2022, widely attributed to perceptions of rising crime and prosecutorial leniency, underscores the political salience of media framing in this domain (Yogev, 2026).

Despite the importance of media coverage in the politics of prosecution, there has been little systematic analysis of whether progressive prosecutors receive measurably different treatment from the press compared to their traditional counterparts. Existing scholarship on media bias in criminal justice has focused primarily on race and crime (Dixon and Linz, 2000), sentencing coverage (Surette, 2015), and outlet-level ideological slant or article-level bias measurement rather

than actor-specific evaluative framing (Groseclose and Milyo, 2005; Gentzkow et al., 2016; Budak et al., 2016; Spinde et al., 2023). The prosecutor-specific question, whether the press covers reform-oriented officeholders more negatively and if so through what mechanisms, remains largely unexamined. The most systematic prior effort, a pilot study of prosecutor coverage across four states, found that the median incumbent prosecutor appeared in only 24 articles during an entire election year, and 89 percent of that coverage contained no information about prosecutors' decisions or policies (Hessick and Thornburg, 2023).

This study addresses that gap. I analyze 136,313 news articles from 21 Bay Area publications over a five-year period (2019–2024), comparing coverage of five district attorneys who span the progressive–traditional spectrum. The San Francisco Bay Area provides an unusually strong research design: multiple counties with sequential tenures of progressive and traditional prosecutors in the same jurisdiction allow within-county comparisons that control for local media markets, crime conditions, and editorial cultures.

I contribute to the literature in three ways. First, I develop a multi-method NLP pipeline that goes beyond simple sentiment analysis to measure aspect-based sentiment (tone specifically toward the prosecutor, not the article generally), zero-shot stance classification (whether the article's framing criticizes or defends the prosecutor), and media frame detection (the narrative lens through which prosecutors are presented). Second, I provide the first large-scale empirical test of differential media treatment of prosecutors by ideological orientation. Third, I demonstrate that framing differences, specifically the disproportionate use of accountability and reform frames for progressive prosecutors, may be more consequential than aggregate tone differences. In doing so, the paper follows recent work that treats media bias as multidimensional and text-based measurement as layered rather than unitary (Spinde et al., 2023; Gentzkow et al., 2019).

2 Literature Review

2.1 Media Framing and Criminal Justice

The study of media framing in criminal justice draws on Entman's (1993) definition of framing as selecting and emphasizing particular aspects of perceived reality to promote specific interpretations. In criminal justice contexts, framing choices—whether to emphasize individual responsibility versus systemic factors, crime trends versus policy outcomes, victims versus defendants—shape public understanding of complex policy questions (Iyengar, 1991; Gross, 2008). Tversky and Kahneman (1981) foundational work on framing effects demonstrates that logically equivalent information presented in different frames produces systematically different judgments, a finding repeatedly confirmed in criminal justice settings (Slovic et al., 2002).

Research on crime news specifically has documented a persistent "mean world" effect (Gerbner et al., 2002): local news overrepresents violent crime relative to its actual incidence, contributing to public fear and support for punitive policies. Dixon and Linz (2000) demonstrate that racial minorities are overrepresented as perpetrators and underrepresented as victims in local television news, while Gilliam and Iyengar (2000) show that implicit racial cues in crime coverage activate punitive policy preferences among white viewers. The asymmetry extends to victim portrayal: white victims are significantly more likely than Black victims to receive sympathetic coverage featuring personal photographs and family context (Global Strategy Group and Equal Justice Initiative, 2021). These media-constructed racial associations drive support for punitive policies independently of actual crime rates (Ghandnoosh, 2014), a dynamic directly relevant to reform prosecutors who frame their agendas around addressing racial disparities in prosecution. This body of work establishes that media coverage of criminal justice is neither neutral nor proportional to underlying reality.

Crime news is also shaped by recurring sourcing asymmetries. Police are often the primary definers of crime events in news coverage, giving law enforcement unusual agenda-setting power over

which facts, suspects, and causal interpretations enter early reporting (Chermak, 1995). Even when journalists retain some autonomy, the police-media relationship remains asymmetric because police communication offices control access, timing, and official confirmation in fast-moving crime stories (Mawby, 2010). Recent evidence extends this logic from source access to language structure itself: news coverage of police killings more often adopts obfuscatory constructions that diffuse responsibility, patterns that appear to trace back in part to police press narratives rather than to overt ideological editorializing (Moreno-Medina et al., 2024).

These patterns reflect structural features of the media industry rather than individual editorial choices. Beale (2006) argues that commercial pressures determine the news media's contemporary treatment of crime: sensationalized coverage is cost-effective and audience-attracting, creating persistent overrepresentation of violent crime. Market incentives also generate structural dependencies on law enforcement as readily available, authoritative sources, a reliance that shapes the narrative framework within which prosecutors are evaluated (Beale, 2006; Surette, 2015). These media dynamics have measurable downstream consequences: Ash and Poyker (2024) provide causal evidence that exposure to conservative crime coverage increases sentencing harshness among elected judges, while Romer et al. (2003) confirm that local television news viewing increases fear of crime independent of actual crime rates.

These asymmetries persist in a changed media environment. Social media allows police departments to bypass traditional gatekeepers through direct, curated communication while also creating a contested arena in which citizens can circulate counter-footage and counter-narratives (Cheng, 2021; Colbran, 2020; Walsh and O'Connor, 2019). The result is not the disappearance of institutional power but a more complex communication ecology in which official actors retain major agenda-setting advantages even as their narratives can be challenged more rapidly.

2.2 Prosecutorial Politics and Public Perception

The rise of the progressive prosecution movement has generated a growing scholarly literature on prosecutorial discretion, electoral accountability, and political backlash (Pfaff, 2017; Sklansky, 2017; Bellin, 2020). Wright (2009) provides a framework for understanding prosecutorial elections as mechanisms of democratic accountability, while Davis (2007) documents the broad discretionary powers that make the prosecutor's office a site of significant policy variation. Bazelon (2019) chronicles the political dynamics surrounding progressive prosecutors, noting that opponents have mobilized law enforcement unions, victims' rights organizations, and media campaigns to challenge reform agendas. Police unions occupy a dual position in this dynamic: they serve as primary media sources for crime stories while simultaneously funding opposition campaigns against reform prosecutors and lobbying for legislation limiting prosecutorial discretion (Goldstein, 2024). This structural asymmetry means that the most frequently quoted institutional actors in crime coverage are also the most organized opponents of prosecutorial reform. Recent reviews document the scope of this opposition, including legislative challenges across multiple states and coordinated recall campaigns (Mitchell and Petersen, 2025).

However, empirical analysis of media coverage of individual prosecutors remains scarce. Despite prosecutors' central role in criminal justice, they are structurally under-covered: a UNC pilot study found that prosecutors receive minimal media attention, with coverage rarely mentioning their elected status even during election years (Hessick and Thornburg, 2023). This coverage deficit is stark relative to law enforcement; in the Bay Area corpus analyzed here, 82 percent of crime articles mention police compared to 15 percent that mention the district attorney. The scarce attention that prosecutors do receive is thus particularly high-stakes, as it constitutes the primary channel through which the public evaluates prosecutorial policy. Most studies of media bias in politics focus on candidate coverage in electoral contexts (D'Alessio and Allen, 2000; Puglisi and Snyder, 2011) or ideological slant at the outlet level (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010). The prosecutor-specific question is distinct because prosecutors are both political figures

and quasi-judicial actors, operating in a policy domain, criminal justice, where media coverage has documented and substantial effects on public attitudes.

Although this adjacent literature concerns policing rather than prosecution, it provides a useful theoretical bridge for why prosecutor coverage should matter. Reviews and survey-based studies show that media type shapes evaluations of criminal justice institutions, with internet and social media often associated with more skeptical legitimacy judgments than traditional formats, though these effects are heterogeneous and can be outweighed by direct experience (Graziano, 2019; Gauthier and Graziano, 2018; Graziano and Gauthier, 2018; Intravia et al., 2018; Dowler, 2002). If media exposure helps structure public judgments about police effectiveness and legitimacy, there is strong reason to expect similar dynamics in evaluations of elected prosecutors, whose policy choices are even less directly observable to most voters.

2.3 Computational Text Analysis in Political Communication

Recent advances in natural language processing have expanded the toolkit available for media bias measurement and, more broadly, for turning large-scale text corpora into inferential social-science data (Gentzkow et al., 2019). Transformer-based models, including BERT (Devlin et al., 2019) and its derivatives, have achieved state-of-the-art performance on sentiment analysis, stance detection, and text classification tasks (Liu et al., 2019). Zero-shot classification approaches, which leverage models trained on natural language inference to classify text without task-specific training data, are particularly valuable in political communication research where labeled training corpora are scarce (Yin et al., 2019).

Applications of NLP to media bias detection include Hamborg et al. (2019), who develop a framework for identifying framing bias through word choice and emphasis, and Fan et al. (2019), who use BERT-based models to detect informational and lexical bias in news articles. Card et al. (2015) apply latent frame analysis to immigration coverage, demonstrating that computational methods can reliably detect framing patterns at scale. Recent work further argues that media bias is a

multi-level construct that cannot be reduced to a single indicator and that LLM-based approaches increasingly outperform lexicon methods, crowd annotation, and older supervised models on political text tasks (Spinde et al., 2023; Gilardi et al., 2023; Birkenmaier and Lechner, 2025; Griswold et al., 2025). This study builds on these approaches by integrating multiple NLP measurements (sentiment analysis, stance detection, thematic emphasis, frame classification, and structural extraction) to capture distinct but related dimensions of media treatment rather than treating any single text score as a sufficient proxy for bias.

3 Data and Methods

The empirical strategy proceeds from sample construction to complementary measurement layers and then to comparative inference. I first construct a prosecutor-attributed corpus from the full news archive through relevance filtering and attribution/disambiguation. I then estimate four directional prosecutor-focused scores (Methods A–D) on a common signed scale and combine them into a composite index for a parsimonious overall test. Framing is estimated separately because it is a categorical narrative-lens outcome rather than a signed intensity measure. Prosecutor-attributed theme detection is also separate because it measures explicit prosecutor–theme linkage rather than local thematic salience alone. Finally, LLM structural extraction captures article architecture (who is quoted, what claims are made, and how causality is asserted), which complements score-based NLP by measuring content structure directly. Inference combines group and paired comparisons, controlled regressions, and segmented interrupted time-series models with quarterly heterogeneity summaries.

3.1 Data

The analysis draws on a corpus of 136,313 news articles published between January 2019 and December 2024, collected from LexisNexis Academic across 21 Bay Area publications. The corpus includes major regional outlets (*San Francisco Chronicle*, sfgate.com, *East Bay Times*, *Mercury*

News), television station web content (kron4.com, ktvu.com, nbcbayarea.com, cbsnews.com), and digital-native local news sources (berkeleyside.com, Palo Alto Online). The temporal range spans the full tenure of Chesa Boudin as San Francisco District Attorney (2020–2022), the subsequent tenure of his replacement Brooke Jenkins (2022–present), the transition from Nancy O'Malley to Pamela Price in Alameda County (2023), and the ongoing tenure of Steve Wagstaffe in San Mateo County.

3.2 Article Filtering and Prosecutor Attribution

Not all articles in the corpus concern crime and criminal justice. I apply a relevance filter requiring at least two crime/justice terms (prosecutor, arrest, sentencing, felony, etc.), which retains 107,713 articles; a zero-shot classifier (BART-MNLI; Lewis et al. 2020) serves as a secondary relevance screen for additional validation checks. Prosecutor attribution then uses regex-based name matching across known variants (e.g., "Boudin," "Chesa Boudin," "DA Boudin") with contextual disambiguation for "Price" (counted as Pamela Price only within 50 words of terms such as "DA," "prosecutor," "Alameda," or "Pamela"). Articles are assigned a primary prosecutor by mention frequency, with headline mentions weighted at $3\times$. A further disambiguation step requires first-name confirmation for ambiguous pre-tenure surname matches (e.g., "Jenkins" before Brooke Jenkins took office in July 2022; "Price" before Pamela Price took office in January 2023), reducing false positives from other individuals or non-name uses of "price." When no named prosecutor appears but the article refers to "district attorney"/"prosecutor," attribution falls back to county-by-date officeholder assignment; these cases are flagged and evaluated in sensitivity analyses. This process yields $n = 12{,}953$ articles attributed to five prosecutors (Table 1). Importantly, the relevance filter uses generic crime and justice vocabulary ("prosecutor," "arrest," "felony") that is orthogonal to the outcome-specific theme dictionaries applied later ("soft-on-crime," "recall campaign," "public-safety-failure"); the filtering step therefore does not select articles on the basis of the thematic variables subsequently measured.

Table 1: Prosecutor sample characteristics.

Prosecutor	County	Ideology	Tenure	*n*
Chesa Boudin	San Francisco	Progressive	2020–2022	5,773
Brooke Jenkins	San Francisco	Traditional	2022–present	3,207
Nancy O'Malley	Alameda	Traditional	2009–2023	1,497
Pamela Price	Alameda	Progressive	2023–2024	828
Steve Wagstaffe	San Mateo	Traditional	2010–present	1,648
Progressive total				6,601
Traditional total				6,352

3.3 Multi-Method Bias Detection

I estimate four prosecutor-focused article-level scores (Methods A–D) and combine them into a weighted composite used for the paper's primary directional comparisons. These methods are grouped because each produces a signed directional measure on a common $[-1, +1]$ scale while still targeting conceptually distinct dimensions of media bias (Spinde et al., 2023).

Method A: Aspect-based sentiment analysis (weight: 0.35). For each prosecutor mention in an article, I extract a three-sentence context window centered on the mention and compute sentiment polarity using a fine-tuned RoBERTa model (cardiffnlp/twitter-roberta-base-sentiment-latest; Loureiro et al. 2022). The article-level score averages across mention windows, measuring sentiment *toward the prosecutor* rather than overall article tone—a critical distinction in crime reporting, where subject matter is inherently negative.

Method B: Zero-shot stance classification (weight: 0.30). For paragraphs containing prosecutor mentions, I classify stance using BART-MNLI (Lewis et al., 2020) with three labels: criticism, defense/support, and neutral reporting about the prosecutor. The article-level score is the balance of supportive minus critical paragraph probabilities.

Method C: Enhanced keyword analysis (weight: 0.20). I apply curated theme dictionaries (crime-rising, soft-on-crime, recall, case dismissal, releasing criminals, public safety failure, victim neglect, police conflict, and office dysfunction) within three-sentence windows around pros-

ecutor mentions. A negation check in the preceding words reduces false positives. These dictionaries operationalize the anti-prosecutor framing patterns identified in prior qualitative work on prosecutorial media coverage (Bazelon, 2019; Goldstein, 2024).

Method D: Document-level sentiment baseline (weight: 0.15). Whole-article sentiment serves as a control, capturing general negativity in crime reporting that is not necessarily prosecutor-specific.

Composite construction (A–D only). The composite score is the weighted average of Methods A–D (0.35, 0.30, 0.20, 0.15) on a common $[-1, +1]$ direction where more negative values indicate more negative treatment. If one or more method scores are missing, weights are renormalized over available methods. The weights follow a precision–generality hierarchy: the two prosecutor-targeted measures (aspect sentiment, 0.35; stance, 0.30) are prioritized over the two text-level measures (keywords, 0.20; document sentiment, 0.15). This scheme is empirically conservative: it assigns the highest weight to the method with the weakest effect (aspect sentiment, $d = 0.037$) and lower weights to the methods with larger effects (stance $d = -0.341$; keywords $d = -0.218$), mechanically attenuating the composite. Six alternative weighting specifications (equal, inverse, effect-proportional, and three leave-one-out variants) confirm the composite effect is negative and significant under every scheme, with d ranging from -0.08 (dropping stance) to -0.39 (retaining only evaluative methods); the current weights produce one of the most conservative estimates. The composite provides a parsimonious overall test of directional differential treatment, while per-method analyses identify which dimensions drive that aggregate result in line with text-as-data approaches that separate measurement construction from downstream inference (Gentzkow et al., 2019).

3.4 Media Framing Analysis

Framing is estimated separately from the A–D composite because it measures narrative *type* (which frame is used), not directional evaluative intensity on a single signed scale. For each attributed

article, I estimate `frame_*` probabilities and a `dominant_frame` from prosecutor-centered text windows (five-sentence windows around mentions) using BART-MNLI in multi-label mode, with keyword-based fallback scoring when no valid mention window is available or model inference fails.

1. **Accountability**: The prosecutor is held responsible for outcomes (e.g., crime rates, case decisions).
2. **Conflict**: Coverage emphasizes disagreement (prosecutor vs. police, victims, or community).
3. **Consequences**: Focus on policy outcomes and their effects.
4. **Human interest**: Individual stories of victims or defendants as narrative vehicles.
5. **Reform/ideology**: The prosecutor is characterized through an ideological or reformist lens.

Method B and framing both use BART-MNLI but operationalize different constructs: stance asks whether language criticizes or supports a prosecutor, whereas framing asks which narrative lens structures the article. Pairwise correlations between article-level stance scores and all five frame probabilities range from $r = -0.10$ (human interest) to $r = -0.21$ (accountability), far below the $r > 0.8$ threshold for methodological redundancy; the modest negative values are theoretically expected (critical stance co-occurs with accountability framing) and confirm convergent validity rather than collinearity.

3.5 Prosecutor-Attributed Theme Detection

Prosecutor-attributed theme detection is estimated separately from both the A–D composite and frame analysis because it targets explicit prosecutor-linked critical narratives rather than general evaluative tone or frame category prevalence. Nine themes are tracked: crime-rising, releasing criminals, victim neglect, police conflict, office dysfunction, soft-on-crime, case dismissal, recall, and public safety failure. To complement Method C's proximity-based theme signal, the attribution algorithm requires explicit prosecutor-theme linkage via four detection methods: (1) context-aware

dictionary patterns requiring both a prosecutor reference and a theme keyword within the same expression (e.g., "crime rising because of the DA" rather than merely "crime" near "prosecutor"), (2) relationship-based regex patterns capturing causal structures linking prosecutors to outcomes, (3) targeted criticism patterns, and (4) sentence-level co-occurrence checks. Method scores (0–25 each) are combined with weights 0.30, 0.35, 0.15, and 0.20, then adjusted by method-agreement rules (single-method caps at 15; 1.1× multiplier for two-method agreement, 1.25× for three or more). The resulting `ta_*` variables are used in dedicated theme-attribution analyses.

3.6 LLM-Based Structural Content Extraction

The preceding NLP analyses (Methods A–D, framing, and prosecutor-attributed theme detection) quantify evaluative, framing, and attributional signal in text. They are well suited for comparative inference at scale but do not directly enumerate full article structure (who is quoted, what claims are made, and how causal attribution is framed). To capture that layer, I add a complementary LLM structural extraction approach using Google's **langextract** library (Gemini 2.5 Flash), applied to all attributed articles in the analysis sample ($n = 12{,}953$, with $n = 6{,}601$ progressive and $n = 6{,}352$ traditional). Recent evidence suggests that prompted frontier models can perform text-measurement and annotation tasks at or above crowd-worker and older supervised baselines, making them useful for structured extraction when paired with auditability constraints (Gilardi et al., 2023; Birkenmaier and Lechner, 2025).

Langextract is extractive rather than classificatory: outputs are schema-constrained instances linked to exact source spans. The extraction schema covers source attribution, claims against the prosecutor, causal claims, policy actions, and explicit comparisons. Model behavior is stabilized through fixed prompting and few-shot calibration examples, yielding auditable structural measures that can be aligned with score-based NLP patterns.

The NLP methods characterize how coverage evaluates and frames prosecutors, while langextract identifies the structural content through which that coverage is built. When both layers converge,

Table 2: Complementary measurement layers: NLP methods vs. LLM structural extraction.

Dimension	NLP scoring/detection (A–D + framing + theme attribution)	LLM extraction (langextract)	Complementary role
Construct	Evaluative, framing, and attribution signal	Source and claim structure	"How" coverage evaluates vs. "what" coverage contains
Unit	Article/paragraph scores and attribution flags	Span-level extracted instances	Cross-layer triangulation
Output	Continuous indices plus validated indicators	Discrete counts and rates	Mechanism mapping for score differences
Strength	High comparability and statistical power	Source-grounded, auditable outputs	Convergence strengthens interpretation
Limitation	Can obscure specific mechanisms	Depends on schema and calibration choices	Divergence is diagnostically informative

meaning when the excess of prosecutor-attributed harm claims (Section 4.8) aligns with the excess of loaded negative language detected in the bias-indicator pilot (Appendix B), the evidence is stronger than either method alone. When they diverge, as when the overall bias-indicator result is null (Appendix B) despite significant stance and keyword effects (Section 4), the divergence reveals that the differential operates through aggregate framing mechanisms rather than discrete journalistic violations. Main-text structural results appear in Section 4.8; full extraction protocol details remain in Appendices A–B.

3.7 Statistical Approach

I employ six analytical strategies: (1) group-level comparison (Welch's t-test, Mann–Whitney U, Cohen's d, Cliff's delta, bootstrap 95% CIs); (2) same-county paired comparisons (Boudin vs. Jenkins in San Francisco; Price vs. O'Malley in Alameda); (3) OLS regression of bias score on prosecutor ideology, controlling for county, year, and article length, with cluster-robust standard errors by publication; (4) framing differential analysis (chi-square tests on dominant frame fre-

quencies by prosecutor type); (5) segmented interrupted time-series models on monthly county-level outcomes around prosecutor transitions, estimating immediate level shifts (`post`) and post-transition slope changes (`time_after`) with HAC-robust inference, complemented by quarterly heterogeneity summaries; and (6) TOST equivalence testing (Lakens, 2017) with an equivalence bound of $d = 0.2$. Analyses (1), (2), (3), and (5) are reported for the composite and for each A–D component, while (4) is framing-specific. Prosecutor-attributed theme detection uses its own effect-size and prevalence tests, reported in the dedicated theme-attribution subsection.

The $d = 0.2$ bound corresponds to Cohen's definition of a "small" effect and is standard in equivalence testing applications. Crucially, in a large-N study ($n > 13{,}000$ articles) trivially small effects achieve statistical significance, so the TOST serves as a meaningful practical check: an equivalence result ($p = .007$) indicates the composite effect is not only small by conventional standards but falls within a range generally considered substantively negligible. The per-method analyses are critical: by running each test separately for stance, sentiment, keywords, and document tone, I can identify *which dimensions* of coverage differ by ideology rather than relying on a composite that may average opposing signals to zero. Throughout, p-values below .001 are reported as $p < .001$; exponential notation (e.g., $p < 10^{-127}$) is used selectively where the extreme magnitude is itself informative.

4 Results

4.1 Bias Is Not One Thing: Per-Method Decomposition

The central finding of this study is that "media bias" toward prosecutors is not a single phenomenon. The four bias detection methods, each measuring a distinct dimension of coverage, produce sharply divergent results (Figure 1):

- **Stance classification (B)** produces the largest effect ($d = -0.341$, $p < .001$): articles about progressive prosecutors contain significantly more critical and fewer defensive paragraphs.

This method captures evaluative framing: whether coverage *judges* the prosecutor favorably or unfavorably.

- **Keyword analysis (C)** shows a small-to-moderate effect ($d = -0.218$, $p < .001$): anti-prosecutor themes (soft-on-crime, crime-rising, recall) appear more frequently near mentions of progressive prosecutors. This method captures thematic emphasis.
- **Aspect-based sentiment (A)** shows a negligible effect ($d = 0.037$, $p = .058$): the emotional tone surrounding prosecutor mentions is essentially equivalent across groups. This method captures emotional valence.
- **Document-level sentiment (D)** shows a small reversed effect ($d = 0.047$, $p = .008$): articles about traditional prosecutors are slightly *more* negative in overall tone, reflecting the inherent negativity of crime reporting rather than prosecutor-specific bias.

The divergence between stance ($d = -0.341$) and sentiment ($d = 0.037$) is the study's key conceptual finding: media coverage of progressive prosecutors maintains a neutral emotional tone while adopting a critical evaluative stance. The press covers progressive prosecutors in the same emotional register as their traditional counterparts but evaluates their performance more critically. A diagnostic visualization of this tone–evaluation split is reported in Appendix C.

This pattern has a direct methodological implication: studies relying solely on sentiment analysis would conclude that media treatment of prosecutors is equivalent. The differential emerges only when measurement captures evaluative framing, a dimension invisible to emotional tone detection.

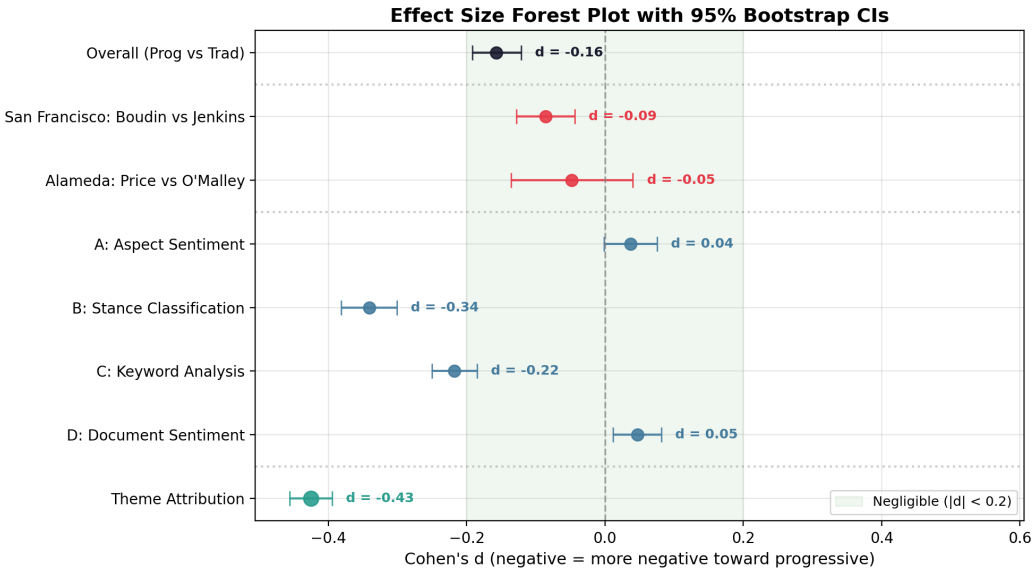

Figure 1: Effect size forest plot with 95% bootstrap confidence intervals. The green band marks the negligible effect zone ($|d| < 0.2$). Results are grouped by analysis type: overall, paired-county, per-method, and theme attribution. Theme attribution ($d = -0.43$) is the study's largest effect.

4.2 Same-County Paired Comparisons by Method

The within-county design provides the strongest quasi-experimental identification. Per-method analysis reveals *which dimensions* of coverage differ within each jurisdiction:

San Francisco (Boudin vs. Jenkins). Stance classification detects significant differential treatment ($d = -0.207$, $p < .001$) and keyword analysis confirms it ($d = -0.116$, $p < .001$). Aspect sentiment shows no significant difference ($d = 0.033$, $p = .180$), and document sentiment is reversed ($d = 0.058$, $p = .009$). The bias is evaluative, not tonal.

Alameda County (Price vs. O'Malley). On a composite measure, this pair shows a non-significant difference ($d = -0.048$, $p = .295$), an apparent null. However, decomposition by method reveals that the null is an artifact of aggregation: stance classification ($d = -0.275$, $p < .001$) and keyword analysis ($d = -0.373$, $p < .001$) both detect substantial differential treatment, while sentiment ($d = 0.082$, $p = .104$) and document tone ($d = 0.124$, $p = .005$) pull in the *opposite* direction. The composite averages these opposing signals into a near-zero result, masking real differential

treatment that operates through evaluative framing.

Table 3: Per-method Cohen's d for same-county paired comparisons. Significant results ($p < .05$) in bold.

	San Francisco		Alameda	
Method	d	p	d	p
B: Stance	$\mathbf{-0.207}$	$< .001$	$\mathbf{-0.275}$	$< .001$
C: Keywords	$\mathbf{-0.116}$	$< .001$	$\mathbf{-0.373}$	$< .001$
A: Aspect sentiment	0.033	.180	0.082	.104
D: Document sentiment	$\mathbf{0.058}$	.009	$\mathbf{0.124}$	.005
Composite	$\mathbf{-0.086}$	$< .001$	-0.048	.295

Figure 2 displays per-prosecutor mean composite scores with 95% confidence intervals. Figure 3 presents the paired county comparisons.

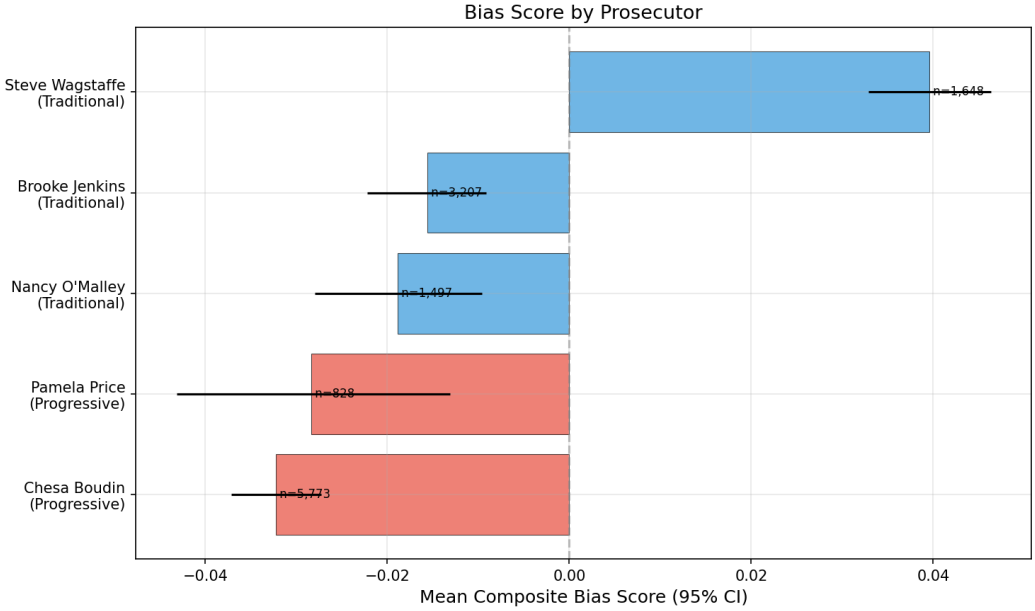

Figure 2: Mean aggregate tone-evaluation index by prosecutor with 95% confidence intervals. Red bars indicate progressive prosecutors; blue bars indicate traditional prosecutors.

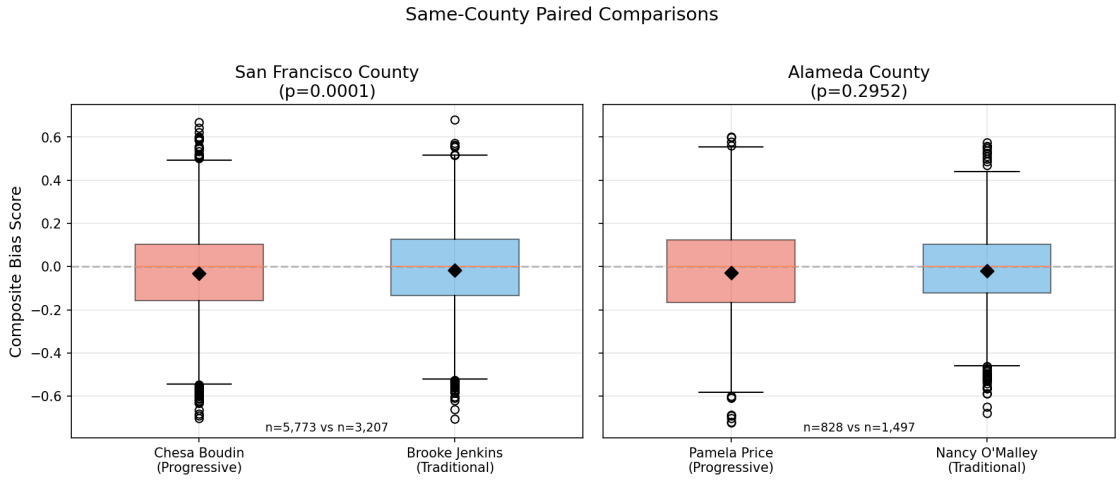

Figure 3: Same-county paired comparisons: San Francisco (Boudin vs. Jenkins) and Alameda (Price vs. O'Malley). Diamond markers indicate group means.

4.3 Regression Analysis by Method

To test whether each bias dimension survives controls for county, year, article length, and publication clustering, I estimate separate OLS regressions for each method (Figure 4; full coefficient table in Appendix C).

Stance classification and keyword analysis, the evaluative and thematic dimensions, both show significant progressive coefficients that survive all controls. Aspect sentiment and document sentiment do not. The composite regression falls to non-significance ($p = .089$) precisely because it averages the strong evaluative signal with the null tonal signal. Running regressions per method resolves this: evaluative bias is robust to controls; tonal bias does not exist.

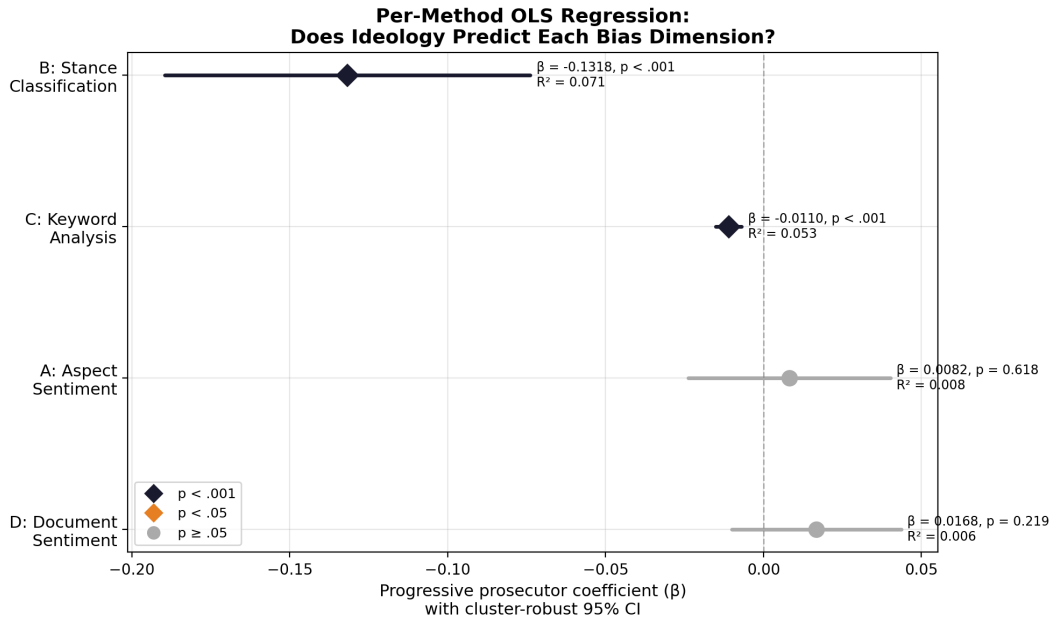

Figure 4: Per-method OLS regression: progressive prosecutor coefficient with 95% cluster-robust CIs. Stance and keyword methods survive controls; sentiment methods do not.

4.4 Temporal Dynamics by Method

Per-method temporal analysis combines segmented interrupted time-series modeling with quarterly decomposition to test whether the evaluative bias identified above is stable or event-driven (Figure 5).

Segmented interrupted time series. Monthly segmented models estimate transition-date level changes and post-transition slope changes for each outcome (Appendix G, Table 12, Figure 16). In San Francisco, the composite shows no immediate level break but a positive post-transition slope, consistent with gradual movement toward less negative coverage after the Boudin-to-Jenkins transition. In Alameda, the composite and stance outcomes show immediate negative level shifts at the O'Malley-to-Price transition with little compensating slope change, consistent with a step increase in critical evaluative coverage. Across transitions, evaluative measures (especially stance, then keywords) carry the temporal signal, whereas sentiment measures remain weak and inconsistent.

Quarterly heterogeneity. Figure 5 decomposes the temporal patterns of each method across 21

quarters. Stance classification ($SD = 0.32$, range $[-0.66, +0.30]$) and keyword analysis ($SD = 0.35$, range $[-0.94, +0.44]$) show dramatic quarterly variation, with the deepest effects coinciding with the Boudin recall (2021Q4–2022Q2) and the Price recall movement (2023Q2–Q3). Aspect sentiment ($SD = 0.28$, range $[-0.30, +0.75]$) and document sentiment ($SD = 0.22$, range $[-0.15, +0.69]$) show less variation and no systematic pattern tied to political events. The monthly time series (Figure 6) overlays a coverage-volume-weighted three-month rolling average alongside the raw monthly means; the weighted and unweighted trends are nearly identical, confirming that the temporal patterns are not artifacts of high-variance low-coverage months.

This per-method temporal decomposition resolves a key ambiguity in the aggregate analysis. The composite quarterly d (range $[-0.39, +0.84]$, $SD = 0.37$) appears to show dramatic cyclical variation, but this aggregate masks the distinct temporal signatures of its component methods. Evaluative bias (stance, keywords) is event-driven, surging during recall campaigns and subsiding between them. Tonal bias (sentiment) is absent throughout. The aggregate temporal variation is driven entirely by the evaluative dimensions.

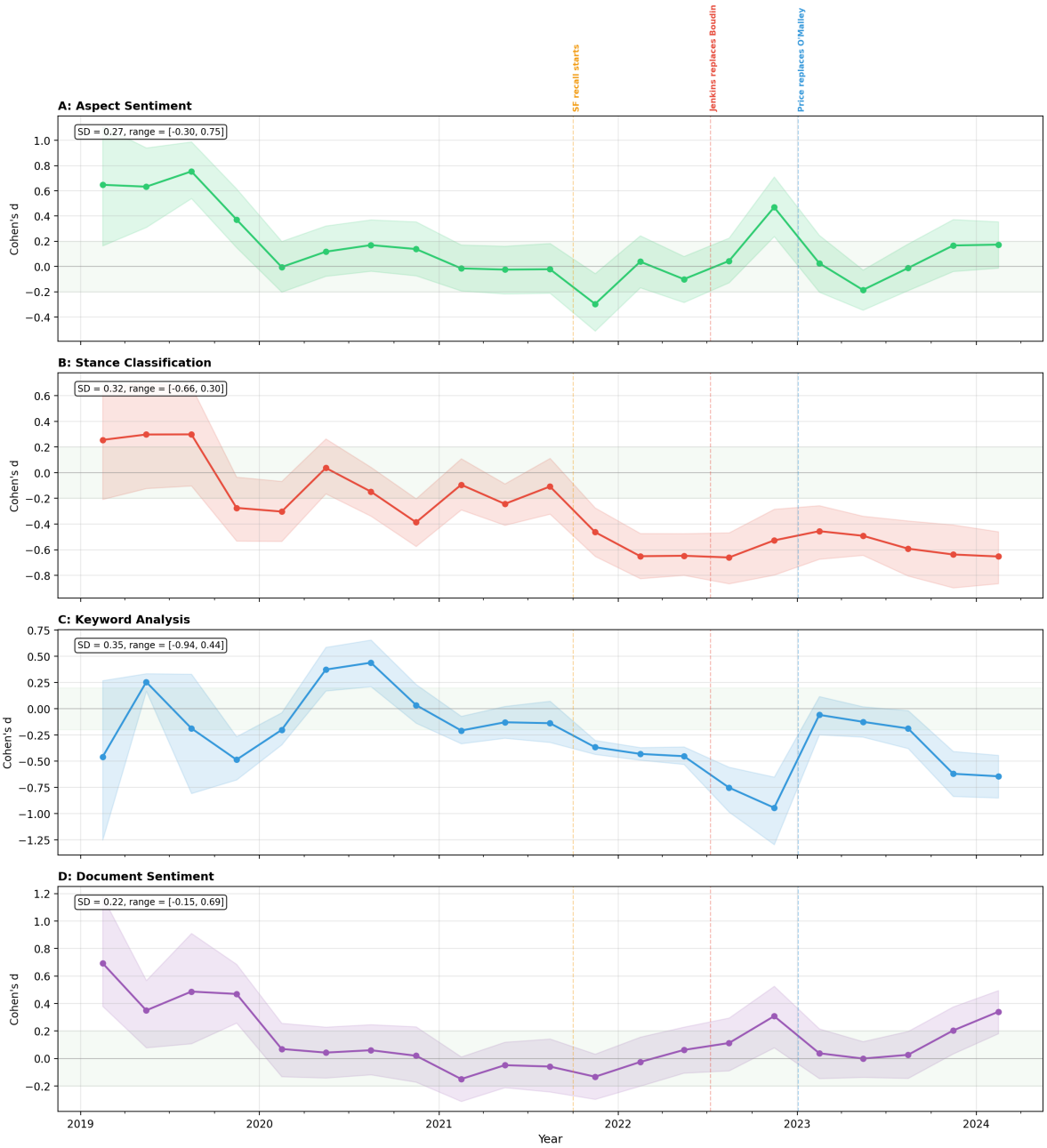

Figure 5: Per-method quarterly Cohen's d (progressive vs. traditional) with 95% bootstrap CIs. The green band marks the negligible zone ($|d| < 0.2$). Dashed lines mark prosecutor transitions. Stance and keyword methods show event-driven variation; sentiment methods remain flat throughout.

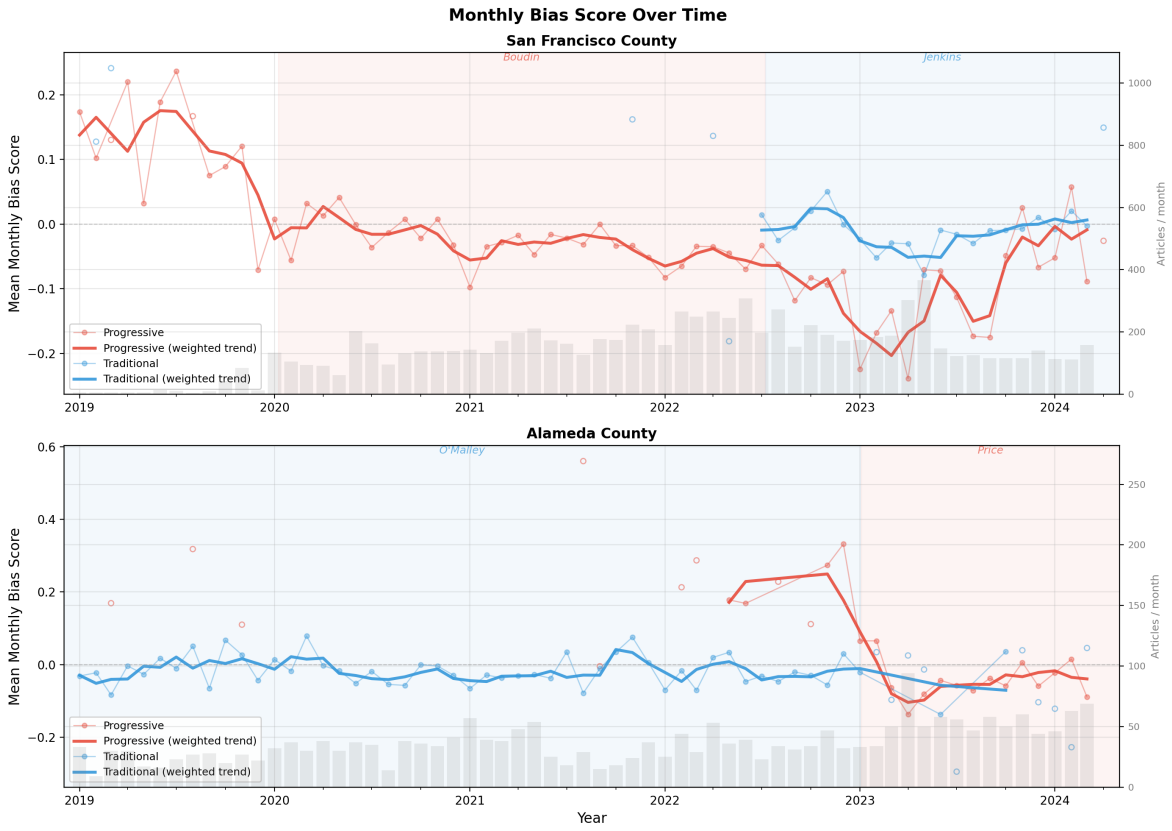

Figure 6: Monthly mean composite scores for San Francisco (top) and Alameda County (bottom), 2019–2024. Faint dots and lines show raw monthly means; thick lines show coverage-volume-weighted three-month rolling averages. Open circles mark months with fewer than five articles. Shaded bands indicate prosecutor tenures; red = progressive, blue = traditional.

4.5 Composite Summary

The preceding A–D analyses constitute the study's primary findings and demonstrate that evaluative and tonal dimensions of coverage diverge sharply. The composite below is reported immediately afterward because it is constructed only from Methods A–D and serves as a summary measure for overall comparison and time-series visualization, not as the principal test of differential treatment.

When the four methods are combined into a weighted composite (weights: 0.35 sentiment, 0.30 stance, 0.20 keywords, 0.15 document), the overall group comparison yields $d = -0.157$ ($p < .001$,

95% bootstrap CI $[-0.036, -0.023]$), with a TOST equivalence test confirming the effect falls within conventional equivalence bounds ($p = .007$ at $d = 0.2$; see Section 3.7 for bound justification). As reported in Section 3.3, the composite effect is negative and significant under all six alternative weighting specifications tested (d ranging from -0.08 to -0.39); the current weights produce one of the most conservative estimates. This composite serves as a proxy for the aggregate reader experience: a reader does not consume "stance" and "sentiment" as separate channels but is exposed to both the emotional tone and the evaluative framing of an article simultaneously. The composite approximates this net impression and provides a single dependent variable for the time series and overall comparison (Figure 7). However, the per-method results above demonstrate that this aggregate figure obscures the distinct mechanisms at play: it averages a large evaluative effect with a null tonal effect, producing a small overall number that understates the stance differential and overstates the sentiment contribution.

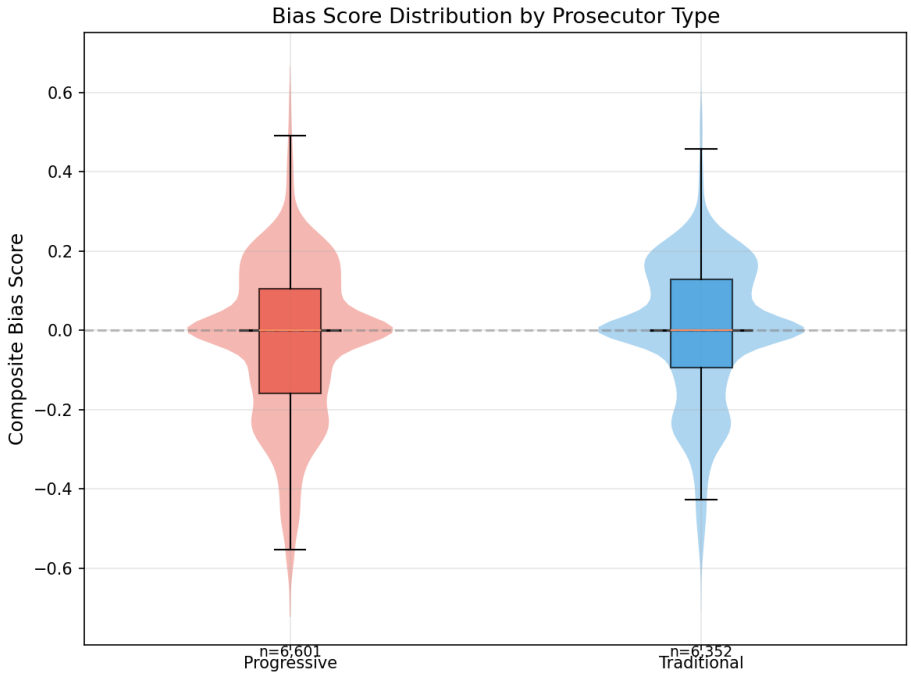

Figure 7: Aggregate tone-evaluation index distributions by prosecutor ideology. Violin plots show the full distribution; inner box plots show the interquartile range. Negative scores indicate more negative coverage.

4.6 Framing Analysis

The distribution of dominant media frames differs significantly between progressive and traditional prosecutors, $\chi^2(4) = 443.75$, $p < .001$, Cramér's $V = 0.225$ (see Figure 8):

Table 4: Dominant media frame distribution by prosecutor ideology.

Frame	Progressive	Traditional
Accountability	39.7%	22.9%
Conflict	30.9%	36.5%
Human interest	21.4%	35.6%
Reform	4.3%	1.3%
Consequences	3.7%	3.7%

Progressive prosecutors are nearly twice as likely to be covered through an accountability frame (39.7% vs. 22.9%, $d = 0.37$, $p < .001$) and more than three times more likely through a reform frame (4.3% vs. 1.3%, $d = 0.19$, $p < .001$). Conversely, traditional prosecutors receive substantially more human-interest framing (35.6% vs. 21.4%, $d = -0.32$, $p < .001$): stories organized around individual victims, defendants, or community members rather than evaluations of prosecutorial performance. Traditional prosecutors also receive slightly more conflict framing (36.5% vs. 30.9%, $d = -0.12$, $p < .001$), while consequences framing is equally distributed ($d = -0.01$, $p = .713$). Effect sizes here reflect per-frame dominant-frame assignment rates, matching the proportions in Table 4.

These framing patterns are consistent with the stance classification finding: the media evaluates progressive prosecutors through accountability and ideological lenses while presenting traditional prosecutors through less evaluative, story-driven narratives. That combination of accountability framing with tonal equivalence is consistent with evidence that media bias often operates through attributional and semantic choices, not just overt negativity (Moreno-Medina et al., 2024).

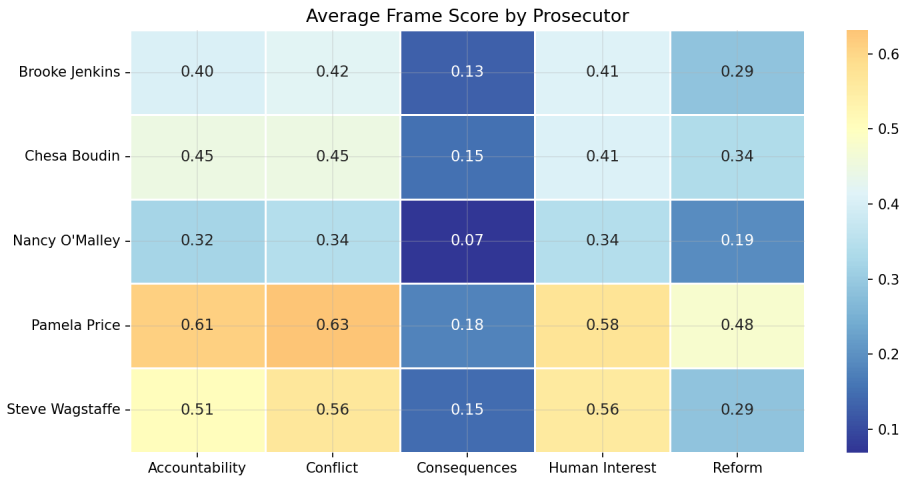

Figure 8: Heatmap of average frame scores by prosecutor. Warmer colors indicate higher frame prevalence. Progressive prosecutors (Boudin, Price) score higher on accountability and reform; traditional prosecutors (Jenkins, O'Malley, Wagstaffe) score higher on human interest. Note: values represent average model probability scores, whereas Table 4 reports dominant frame assignments (the frame with the highest probability per paragraph).

4.7 Theme Prevalence and Prosecutor-Attributed Themes

This subsection reports two complementary theme analyses. *Part A (Method C descriptive prevalence):* Figure 9 presents keyword prevalence from the Method C dictionaries, measured as local term density within three-sentence windows around prosecutor mentions. By this measure, recall-related language dominates progressive prosecutor coverage (9.8% vs. 2.7% for traditional), driven overwhelmingly by Boudin's 2022 recall campaign. After recall, the most prevalent anti-prosecutor themes for progressive coverage are soft-on-crime (1.6% vs. 0.7%), crime-rising (1.5% vs. 0.3%), and public-safety-failure (1.2% vs. 0.3%). The "releasing criminals" theme is the only category where traditional prosecutors receive higher prevalence (5.5% vs. 4.1%), likely reflecting routine bail and parole coverage rather than ideologically targeted framing.

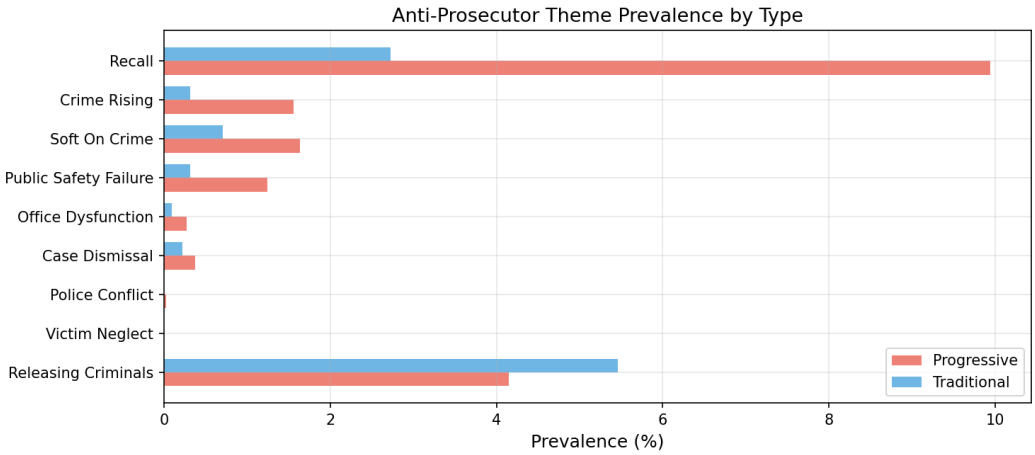

Figure 9: Anti-prosecutor theme prevalence by ideology. Themes are sorted by the difference between progressive and traditional prevalence rates.

Part B (prosecutor-attributed theme model): The multi-method prosecutor-attributed theme analysis requires themes to be explicitly linked to prosecutors through compound regex patterns rather than simple keyword proximity. This stricter attribution measure provides the strongest evidence of differential thematic coverage. Progressive prosecutors receive significantly higher theme attribution scores ($M = 1.80$, $SD = 3.07$) than traditional prosecutors ($M = 0.74$, $SD = 1.71$), $t(10{,}765) = 24.43$, $p < 10^{-127}$, $d = -0.43$ (signed to match the composite convention: negative indicates more anti-prosecutor coverage), 95% bootstrap CI for the mean difference $[0.98, 1.15]$. This medium effect size is the study's largest, exceeding the stance classification effect ($d = -0.341$) and confirming that the evaluative differential operates through the explicit attribution of critical narratives to specific prosecutors.

Per-theme analysis (Figure 10) uses a different metric: the percentage of articles in which *any* of the four attribution methods flags the theme, producing higher prevalence rates than the keyword-proximity measure above. By this article-level flag, recall-related themes appear 4.02 times more often in progressive prosecutor coverage (19.5% vs. 4.8%, $\chi^2 = 641.9$, $p < 10^{-141}$, Cramér's $V = 0.22$). Crime-rising themes appear 1.72 times more often (21.0% vs. 12.2%, $\chi^2 = 178.7$, $p < 10^{-40}$), and soft-on-crime framing is 3.74 times more prevalent ($\chi^2 = 68.5$, $p < 10^{-16}$). The

paired county theme comparisons confirm (reported in the standard progressive-minus-traditional direction, where positive d indicates higher theme attribution for the progressive prosecutor): in San Francisco, Boudin scores higher than Jenkins ($d = 0.28$, $p < .001$), and in Alameda, Price scores higher than O'Malley ($d = 0.58$, $p < .001$), the study's largest within-county effect.

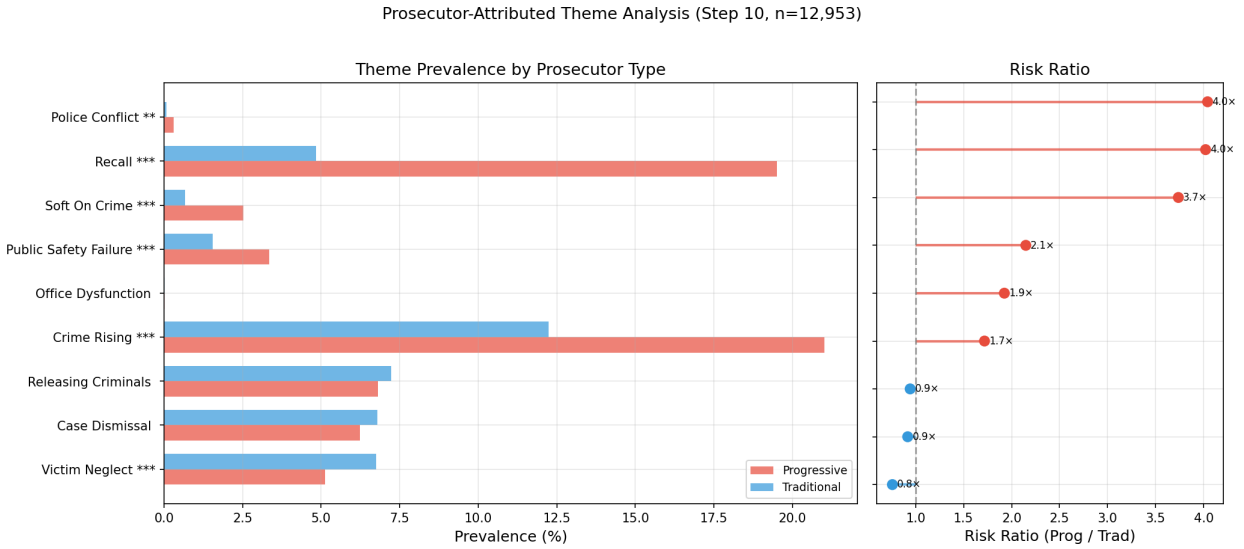

Figure 10: Prosecutor-attributed theme prevalence by ideology (left) and risk ratios (right). Significance: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. Themes sorted by risk ratio.

Multi-method validation shows that 41.4% of articles with any detected theme trigger two or more independent detection methods, confirming convergent evidence rather than single-method artifacts (Appendix C).

4.8 Structural Source Ecology and Claims (Full-Corpus LLM Extraction)

After establishing evaluative differentials with NLP scoring, this section identifies the structural content patterns that co-occur with those differentials. Source types differ strongly by prosecutor ideology ($\chi^2(9) = 1720.3$, $p \approx 0$), with progressive prosecutor coverage drawing disproportionately on advocacy voices (2.3×), journalist commentary (1.5×), politicians (1.5×), and experts (1.4×) relative to traditional coverage.

Progressive prosecutors attract both more critical and more supportive source attributions per arti-

Table 5: Source stance per article by prosecutor ideology (full extraction, $n = 12{,}953$).

Source Stance	Prog./art.	Trad./art.	d	Ratio
Critical	1.70	1.28	0.14	1.3×
Supportive	1.42	0.95	0.22	1.5×
Neutral	3.16	3.84	−0.20	0.8×

cle, while traditional prosecutors receive more neutral sourcing. The largest absolute source-rate difference remains prosecutor self-quotation (2.08 per article for progressives vs. 2.56 for traditional), indicating that traditional coverage is more often anchored in the officeholder's own framing while progressive coverage draws more external contestation.

Causal attributions show the same directional asymmetry: "prosecutor caused harm" appears at 0.41 per progressive article versus 0.24 per traditional article (1.7×), while "prosecutor helped" also appears slightly more often for progressives (0.15 vs. 0.13, 1.2×). For claim content, policy and performance criticism show the largest differentials (policy: 0.48 vs. 0.31 per article, 1.6×; performance: 0.45 vs. 0.31, 1.5×), while character and competence claims are comparatively symmetric (1.1× and 1.1×).

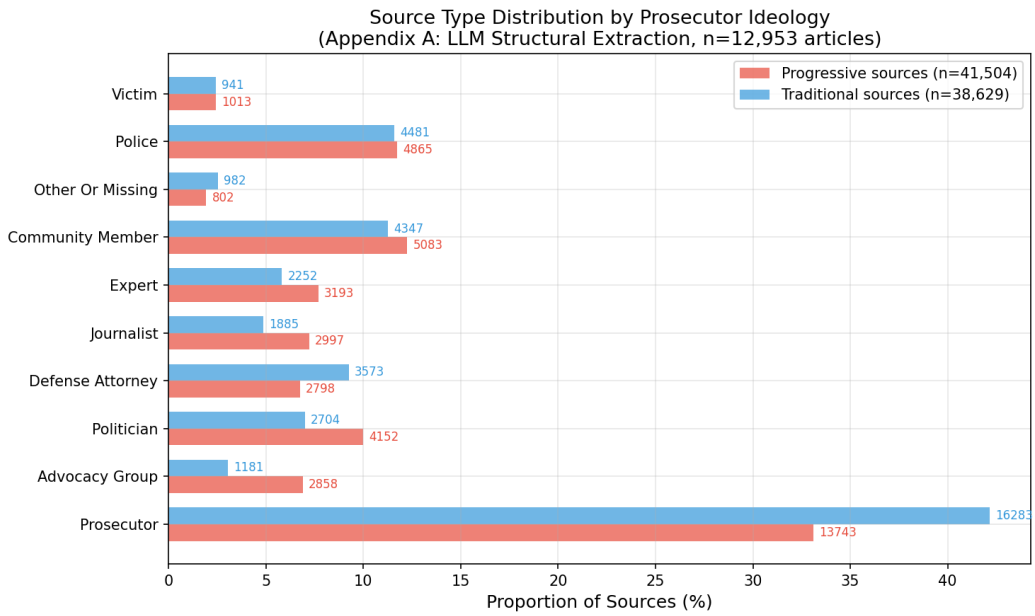

Figure 11: Source type distribution by prosecutor ideology from LLM structural extraction ($n =$ 12,953 articles). Bars show proportions of total sources within each group; annotations show raw counts. The largest relative differentials are in advocacy groups (2.3×), journalist commentary (1.5×), politicians (1.5×), and expert sources (1.4×).

These structural patterns are consistent with the evaluative-framing findings but remain descriptive rather than causal. In that sense, the source-ecology results extend long-standing research on police-media asymmetries into the prosecutor domain: coverage appears shaped not only by what is said about prosecutors, but by which institutional actors are repeatedly positioned to define the story in the first place (Chermak, 1995; Mawby, 2010).

Robustness checks point in the same direction: excluding Brooke Jenkins from the traditional baseline and excluding fallback-attributed articles both increase effect magnitudes, indicating attenuation rather than artifact in the full-sample estimates (Appendices E and F).

5 Discussion

This study provides the first large-scale, multi-method analysis of media coverage of prosecutors by ideological orientation. Three principal findings emerge.

First, media bias toward prosecutors is multi-dimensional, and the dimensions diverge. The study's central contribution is demonstrating that what is commonly treated as a unitary construct, "media bias," decomposes into at least two independent dimensions: evaluative framing and emotional tone. Stance classification detects a medium effect ($d = -0.341$) that survives regression controls ($\beta = -0.132$, $p < .001$); sentiment analysis detects no effect ($d = 0.037$, $p = .058$; $\beta = 0.008$, $p = .618$ after controls). This divergence is not a methodological artifact but a substantive finding: the media covers progressive prosecutors in the same emotional register as their traditional counterparts but evaluates their performance more critically. This pattern aligns with Entman's (1993) observation that framing operates through selection and salience rather than through overt evaluative language, and with prior scholarship documenting how structural accumulation of story selection, source amplification, contextual juxtaposition, and semantic form creates critical impressions without lexical negativity (Beale, 2006; Surette, 2015; Spinde et al., 2023; Moreno-Medina et al., 2024). The prosecutor-attributed theme analysis provides the most direct evidence of this mechanism: when themes must be explicitly linked to prosecutors through compound linguistic patterns, the effect size increases to $d = -0.43$, confirming that the evaluative differential operates through the *attribution* of critical narratives rather than through diffuse tonal differences. The full-corpus structural extraction converges with this result: progressive prosecutor coverage is organized around more externally contestatory source ecologies and more harm-attribution content, while traditional coverage relies more heavily on neutral self-quotation. Together, the score-based and extraction-based layers support a complementary interpretation of differential coverage while remaining descriptive rather than causal.

The Alameda County comparison illustrates why decomposition matters. The composite paired comparison yields a non-significant $d = -0.048$ ($p = .295$), an apparent null. However, per-method analysis reveals that stance ($d = -0.275$, $p < .001$) and keywords ($d = -0.373$, $p < .001$) detect substantial differential treatment while sentiment pulls in the opposite direction ($d = 0.082$, $d = 0.124$). The composite averages these opposing signals to near-zero. A single-index approach would miss the differential treatment entirely.

Second, evaluative bias is event-driven rather than stable. The per-method temporal analysis (Figure 5) resolves a key ambiguity. Segmented ITS estimates around transition dates (Appendix G) reinforce the same pattern: limited transition effects in San Francisco but clear immediate evaluative shifts in Alameda, concentrated in stance and keyword outcomes rather than sentiment. Stance classification and keyword analysis show dramatic quarterly variation ($SD = 0.32$ and 0.35, respectively), with the deepest effects coinciding with the Boudin recall campaign (2021Q4–2022Q2) and the Price recall movement (2023Q2–Q3). Aspect sentiment and document sentiment show less variation ($SD = 0.28$ and 0.22) and no systematic pattern tied to political events. The aggregate temporal heterogeneity (composite range $d = -0.39$ to $+0.84$, $SD = 0.37$) is driven entirely by the evaluative dimensions; tonal bias is absent throughout. This finding suggests that recall campaigns and politically charged events amplify evaluative scrutiny of progressive prosecutors, rather than a stable editorial disposition against them, a pattern that is more consistent with politically responsive media incentives than with fixed outlet ideology (Gentzkow et al., 2016; Prat and Strömberg, 2007). Accordingly, the term "bias" as used throughout this paper denotes differential evaluative responsiveness to political events, not a stable editorial preference: the finding is not that newsrooms maintain a constant anti-progressive disposition, but that politically charged events trigger evaluative scrutiny of reform prosecutors that has no tonal counterpart in coverage of traditional prosecutors facing comparable circumstances.

Third, framing differences are more pronounced and potentially more consequential than any tonal measure. The Cramér's V of 0.225 for dominant frame distributions, a medium effect, indicates that progressive and traditional prosecutors are covered through substantively different narrative lenses. Progressive prosecutors are held to an accountability standard (40% of articles vs. 23%) and viewed through an ideological lens (4% vs. 1%), while traditional prosecutors receive more human-interest framing (36% vs. 21%). These framing choices may matter more than any aggregate tone measure: they determine whether voters receive performance evaluations or human-interest stories about structurally similar officeholders.

5.1 Limitations

Several limitations warrant caution. First, this is an observational study: differences in coverage may reflect genuine differences in prosecutorial conduct, public controversy, or newsworthy events rather than ideological bias per se. Boudin's recall, Price's contentious relationship with law enforcement, and the unique political dynamics of Bay Area criminal justice all represent confounds that within-county comparisons can mitigate but not eliminate. Second, the NLP methods employed, while state-of-the-art, have known limitations: zero-shot classifiers may exhibit systematic biases, including the potential confusion of quoted source stance with authorial stance in attribution-heavy journalism (e.g., classifying "critics say Boudin is soft on crime" as the article's evaluative position rather than a reported viewpoint); sentiment models trained on social media text may transfer imperfectly to news prose, and keyword dictionaries inevitably embed researcher assumptions about what constitutes negative framing. Third, the study examines a single metropolitan area; whether these patterns generalize to other progressive prosecutors (e.g., Krasner in Philadelphia, Foxx in Chicago, Gascón in Los Angeles) is an empirical question. Fourth, news production is highly redundant: a single press conference or recall milestone often generates near-duplicate stories across outlets and wire services, inflating the effective sample size. Although the cluster-robust standard errors (clustered by publication) partially address within-outlet dependence, they do not account for story-level clustering across publications covering the same event. The extreme p-values reported throughout (e.g., $p < 10^{-127}$) should accordingly be interpreted as strong directional evidence rather than precise probability statements; the effect sizes, which are less sensitive to sample size inflation, are the substantively meaningful quantities.

5.2 Implications and Future Directions

These findings have implications for scholarship on prosecutorial politics, media bias, and computational text analysis. For prosecutorial politics, the results suggest that reform prosecutors face a structurally different media environment: one organized around accountability and ideological evaluation rather than the human-interest narratives that characterize coverage of traditional pros-

ecutors. This asymmetry may contribute to the political vulnerability of progressive prosecutors, who face constant performance evaluation in the press while their traditional counterparts benefit from more diffuse, story-driven coverage.

For media bias research, the divergence between sentiment (null) and stance (large effect) highlights the importance of multi-dimensional measurement. Studies relying solely on sentiment analysis may underestimate or miss entirely the evaluative dimensions of media coverage that operate through framing rather than emotional tone. More broadly, this paper extends media-bias research from outlet slant and article-level ideological scaling toward actor-specific evaluative framing: even when hard-news outlets appear relatively similar in aggregate, they can still distribute accountability, criticism, and narrative attention unevenly across officeholders (Groseclose and Milyo, 2005; Budak et al., 2016; Spinde et al., 2023).

These findings also raise concerns about democratic accountability in criminal justice. Media coverage is the primary conduit through which the public learns about prosecutorial policy (Wright, 2009), and the results indicate that this conduit applies systematically different evaluative standards depending on a prosecutor's ideological orientation. Adjacent criminal-justice research suggests that media exposure can shape institutional legitimacy judgments, implying that repeated differences in prosecutor framing may alter how voters evaluate prosecutorial reform even when they encounter little direct evidence about office performance (Graziano, 2019; Gauthier and Graziano, 2018; Intravia et al., 2018). When progressive prosecutors are disproportionately covered through accountability frames while traditional prosecutors receive human-interest coverage, voters receive different types of information about structurally similar officeholders, developing calibrated judgments about a reform prosecutor's policy outcomes while remaining uninformed about a traditional prosecutor's performance. This asymmetry is compounded by the "toplash" dynamics Goldstein (2024) describes, in which institutional actors leverage media relationships to amplify opposition to reform. If small, persistent shifts in evaluative framing accumulate over thousands of articles, they may contribute to the political vulnerability of reform prosecutors through a mechanism that

is invisible in any individual article but consequential in aggregate. Taken together, these results demonstrate that the press covers reform prosecutors differently in kind rather than in degree, not through uniform hostility, but through systematic differences in evaluative framing, sourcing, and narrative structure that are invisible to tonal analysis alone.

Future work should expand the corpus to include other reform prosecutors nationally, develop human-coded validation samples to benchmark NLP performance, and pursue causal identification strategies, such as regression discontinuity designs around close prosecutorial elections, to distinguish ideological bias from coverage differences attributable to newsworthy events.

References

Ash, E. and Poyker, M. (2024). Conservative news media and criminal justice: Evidence from exposure to the Fox News Channel. *The Economic Journal*, 134(660):1331–1355.

Bazelon, E. (2019). *Charged: The New Movement to Transform American Prosecution and End Mass Incarceration*. Random House, New York.

Beale, S. S. (2006). The news media's influence on criminal justice policy: How market-driven news promotes punitiveness. *William & Mary Law Review*, 48(2):397–481.

Bellin, J. (2020). Theories of prosecution. *California Law Review*, 108:1203–1253.

Birkenmaier, L. and Lechner, C. (2025). Measuring politicians' public personality traits using computational text analysis: A multimethod feasibility study for agency and communion. *Political Analysis*.

Budak, C., Goel, S., and Rao, J. M. (2016). Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271.

Card, D., Gross, A., Smith, N. A., and Tsvetkov, Y. (2015). The media frames corpus: Annotations of frames across issues. In *Proceedings of ACL-IJCNLP 2015*.

Cheng, T. (2021). Social media, socialization, and pursuing legitimation of police violence. *Criminology*, 59(3):391–418.

Chermak, S. (1995). Image control: How police affect the presentation of crime news. *American Journal of Police*, 14(2):21–44.

Colbran, M. P. (2020). Policing, social media and the new media landscape: Can the police and the traditional media ever successfully bypass each other? *Policing and Society*, 30(3):295–309.

D'Alessio, D. and Allen, M. (2000). Media bias in presidential elections: A meta-analysis. *Journal of Communication*, 50(4):133–156.

Davis, A. J. (2007). *Arbitrary Justice: The Power of the American Prosecutor*. Oxford University Press.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*.

Dixon, T. L. and Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication*, 50(2):131–154.

Dowler, K. (2002). Media influence on citizen attitudes toward police effectiveness. *Policing and Society*, 12(3):227–238.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

Fan, L., White, M., Sharma, E., Su, R., Choubey, P. K., Huang, R., and Wang, L. (2019). In plain sight: Media bias through the lens of factual reporting. In *Proceedings of EMNLP 2019*.

Gauthier, J. F. and Graziano, L. M. (2018). News media consumption and attitudes about police: In search of theoretical orientation and advancement. *Journal of Crime and Justice*, 41(5):504–520.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M., Shapiro, J. M., and Stone, D. F. (2016). Media bias in the marketplace: Theory. In Anderson, S. P., Waldfogel, J., and Strömberg, D., editors, *Handbook of Media Economics*, volume 1, pages 623–645. Elsevier, Amsterdam.

Gerbner, G., Gross, L., Morgan, M., Signorielli, N., and Shanahan, J. (2002). Growing up with television: Cultivation processes. In Bryant, J. and Zillmann, D., editors, *Media Effects: Advances in Theory and Research*, pages 43–67. Lawrence Erlbaum Associates.

Ghandnoosh, N. (2014). Race and punishment: Racial perceptions of crime and support for punitive policies. Technical report, The Sentencing Project, Washington, DC.

Gilardi, F., Alizadeh, M., and Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120:e2305016120.

Gilliam, F. D. and Iyengar, S. (2000). Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science*, 44(3):560–573.

Global Strategy Group and Equal Justice Initiative (2021). Innocent until proven guilty? A look at media coverage of criminal defendants in the U.S. Technical report, Global Strategy Group and Equal Justice Initiative.

Goldstein, R. (2024). Toplash: Progressive prosecutors under attack from above. *American Criminal Law Review*, 61:1157–1203.

Graziano, L. M. (2019). News media and perceptions of police: A state-of-the-art review. *Policing: An International Journal*, 42(2):209–225.

Graziano, L. M. and Gauthier, J. F. (2018). Media consumption and perceptions of police legitimacy. *Policing: An International Journal*, 41(5):593–607.

Griswold, M., Robbins, M. W., and Pollard, M. S. (2025). Stay tuned: Improving sentiment analysis and stance detection using large language models. *Political Analysis*.

Groseclose, T. and Milyo, J. (2005). A measure of media bias. *Quarterly Journal of Economics*, 120(4):1191–1237.

Gross, K. (2008). Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, 29(2):169–198.

Hamborg, F., Donnay, K., and Gipp, B. (2019). Automated identification of media bias in news articles: An interdisciplinary literature review. *International Journal on Digital Libraries*, 20:391–415.

Hessick, C. B. and Thornburg, R. (2023). Media coverage of prosecutors and their elections: Results of a pilot study. Technical report, Prosecutors and Politics Project, University of North Carolina School of Law.

Intravia, J., Wolff, K. T., and Piquero, A. R. (2018). Investigating the effects of media consumption on attitudes toward police legitimacy. *Deviant Behavior*, 39(8):963–980.

Iyengar, S. (1991). *Is Anyone Responsible? How Television Frames Political Issues*. University of Chicago Press.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4):355–362.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of ACL 2020*.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Loureiro, D., Barbieri, F., Neves, L., Anke, L. E., and Camacho-Collados, J. (2022). TimeLMs: Diachronic language models from Twitter. In *Proceedings of ACL 2022 Demo Track*.

Mawby, R. C. (2010). Police corporate communications, crime reporting and the shaping of policing news. *Policing and Society*, 20(1):124–139.

Mitchell, O. and Petersen, N. (2025). The rise of progressive prosecutors in the United States: Politics, prospects, and perils. *Annual Review of Criminology*, 8:459–484.

Moreno-Medina, J., Ouss, A., Bayer, P., and Ba, B. (2024). Officer-involved: The media language of police killings. Working paper, February 2024.

Pfaff, J. F. (2017). *Locked In: The True Causes of Mass Incarceration and How to Achieve Real Reform*. Basic Books.

Prat, A. and Strömberg, D. (2007). The political economy of mass media. *Annual Review of Political Science*, 10:103–126.

Puglisi, R. and Snyder, J. M. (2011). Newspaper coverage of political scandals. *Journal of Politics*, 73(3):931–950.

Romer, D., Jamieson, K. H., and Aday, S. (2003). Television news and the cultivation of fear of crime. *Journal of Communication*, 53(1):88–104.

Sklansky, D. A. (2017). The progressive prosecutor's handbook. *UC Davis Law Review Online*, 50:25–42.

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2002). The affect heuristic. In Gilovich, T., Griffin, D., and Kahneman, D., editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 397–420. Cambridge University Press.

Spinde, T., Hinterreiter, S., Haak, F., Ruas, T., Giese, H., Meuschke, N., and Gipp, B. (2023). The media bias taxonomy: A systematic literature review on the forms and automated detection of media bias. *arXiv preprint arXiv:2312.16148*.

Surette, R. (2015). *Media, Crime, and Criminal Justice: Images, Realities, and Policies*. Cengage Learning, Stamford, 5th edition.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.

Walsh, J. P. and O'Connor, C. (2019). Social media and policing: A review of recent research. *Sociology Compass*, 13:e12648.

Wright, R. F. (2009). How prosecutor elections fail us. *Ohio State Journal of Criminal Law*, 6:581–610.

Yin, W., Hay, J., and Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *Proceedings of EMNLP-IJCNLP 2019*.

Yogev, D. (2026). Holding justice accountable: Intensive vs. extensive margins in prosecutor elections. *Public Opinion Quarterly*, 89(4):1087–1123.

Appendix A LLM-Based Structural Content Extraction

The multi-method analyses reported in Section 4 measure evaluative direction, narrative framing, and prosecutor-attributed themes but do not directly enumerate the structural building blocks of articles: who is quoted, what claims are made, and with what evidence. Understanding these structural features is essential for explaining *why* evaluative coverage differs across prosecutor types, yet manual content coding at the scale of $n = 12{,}953$ articles is prohibitively expensive. This appendix reports extraction protocol details and extended tables for the full-corpus LLM-based structured extraction.

Method

I use Google's langextract library, which leverages Gemini 2.5 Flash to extract structured information from text with source grounding: every extraction is anchored to a specific span in the original article. The model was prompted with a detailed extraction schema and three diverse few-shot examples (one critical of a progressive prosecutor, one neutral about a traditional prosecutor, one with mixed coverage) to calibrate extraction behavior across ideological contexts. I applied the extraction pipeline to the full corpus of 12,953 prosecutor-attributed articles (6,601 progressive, 6,352 traditional), covering all five prosecutors: Boudin ($n = 5{,}773$), Jenkins ($n = 3{,}207$), O'Malley ($n = 1{,}497$), Price ($n = 828$), and Wagstaffe ($n = 1{,}648$).

Extraction prompt

The following prompt was passed to the model for every article. It defines five extraction classes, each with structured attributes that constrain the model's output to a fixed schema:

```
Extract the following types of information from this news article about a
district attorney / prosecutor.  Use exact text spans from the article.
```

1. `claim_against_prosecutor`: Specific accusations, criticisms, or negative claims made about the prosecutor's performance, policies, or character.

Attributes: `claim_type` (performance | policy | character | competence), `specificity` (vague | specific | quantified), `evidence_cited` (none | anecdotal | statistical | official_report).

2. `source_attribution`: Identify who is speaking, quoted, or cited as a source.
   Attributes: `source_type` (police | victim | defense_attorney | prosecutor | politician | community_member | journalist | expert | advocacy_group), `stance_toward_prosecutor` (critical | supportive | neutral).

3. `causal_claim`: Claims that the prosecutor's actions caused or contributed to some outcome.
   Attributes: `effect` (crime_increase | public_safety_decline | case_outcome | community_impact | positive_outcome), `causal_strength` (explicit | implied | speculative), `direction` (prosecutor_caused_harm | prosecutor_helped | ambiguous).

4. `policy_action`: Concrete policies, decisions, or actions attributed to the prosecutor.
   Attributes: `action_type` (declined_to_prosecute | reduced_charges | new_policy | fired_staff | reversed_predecessor | enhanced_prosecution | other), `domain` (drugs | property_crime | violent_crime | bail | sentencing | staffing | juvenile | general), `framing` (positive | negative | neutral).

5. `comparison`: Explicit comparisons between the current prosecutor and a predecessor, another prosecutor, or a general standard.
   Attributes: `compared_to` (predecessor | other_prosecutor | general_standard), `dimension` (toughness | case_outcomes | policy | competence | ideology), `who_favored` (current | predecessor | neither).

`Only extract items that are clearly present. Do not fabricate or infer. Extract text spans in order of appearance. Do not paraphrase.`

Few-shot examples

Three examples were provided with each API call to demonstrate the expected extraction format. Each example pairs a synthetic article excerpt with the complete set of extractions the model should produce. The examples were designed to span the ideological and tonal range of the corpus.

Example 1: Critical coverage of a progressive prosecutor.

> *"San Francisco's District Attorney Chesa Boudin faced renewed criticism this week after police data showed a 15% increase in car break-ins since he took office. Police Officers Association president Tony Montoya said officers are demoralized because suspects they arrest are quickly released without charges. 'Why bother making arrests when the DA won't prosecute?' Montoya asked. Boudin defended his record, saying his office focuses on serious violent crime rather than low-level offenses. Recall organizers said the crime statistics prove Boudin's progressive policies have failed San Francisco residents."*

Class	Extracted Text	Key Attributes
claim_against	"a 15% increase in car break-ins since he took office"	performance, quantified, statistical
source_attrib.	"Police Officers Association president Tony Montoya said officers are demoralized"	police, critical
source_attrib.	"Boudin defended his record…"	prosecutor, supportive
causal_claim	"the crime statistics prove Boudin's progressive policies have failed"	crime_increase, explicit, harm
policy_action	"suspects they arrest are quickly released without charges"	declined_to_prosecute, negative
source_attrib.	"Recall organizers said…"	advocacy_group, critical

Example 2: Neutral coverage of a traditional prosecutor.

"Brooke Jenkins, who replaced Boudin after the recall, announced a new initiative targeting retail theft rings in the downtown area. Jenkins said her office would file felony charges against organized shoplifting groups, reversing Boudin's practice of treating most retail theft as misdemeanors. Supervisor Matt Dorsey praised the move. Defense attorney Niki Solis warned that harsher penalties would disproportionately affect low-income communities without reducing crime."

Class	Extracted Text	Key Attributes
policy_action	"file felony charges against organized shoplifting groups"	enhanced_prosecution, property_crime, positive
comparison	"reversing Boudin's practice of treating most retail theft as misdemeanors"	predecessor, toughness, current
source_attrib.	"Supervisor Matt Dorsey praised the move"	politician, supportive
source_attrib.	"Defense attorney Niki Solis warned…"	defense_attorney, critical

Example 3: Mixed coverage of a progressive prosecutor.

"Alameda County DA Pamela Price faces a recall effort barely a year into her term. Critics say Price has been too lenient, pointing to cases where violent offenders received reduced sentences. The Save Alameda for Everyone committee cited three homicide cases where Price's office offered plea deals. Price responded that her office has actually increased the conviction rate for violent felonies compared to her predecessor Nancy O'Malley."

Class	Extracted Text	Key Attributes
claim_against	"Price has been too lenient"	policy, vague, none
claim_against	"violent offenders received reduced sentences"	performance, specific, anecdotal
source_attrib.	"The Save Alameda for Everyone committee cited three homicide cases"	advocacy_group, critical
policy_action	"Price's office offered plea deals"	reduced_charges, violent_crime, negative
comparison	"increased the conviction rate... compared to her predecessor"	predecessor, case_outcomes, current
source_attrib.	"Price responded that her office has actually increased..."	prosecutor, supportive

The extraction schema defines five content categories: (1) *claims against the prosecutor*, classified by type (performance, policy, character, competence) and evidence quality; (2) *source attributions*, identifying the speaker's role and stance toward the prosecutor; (3) *causal claims* linking the prosecutor's actions to outcomes; (4) *policy actions* attributed to the prosecutor; and (5) *comparisons* between the current prosecutor and predecessors or standards. The pipeline produced 135,178 total extraction instances across 12,953 articles (mean = 10.4 per article).

Results

Source type distribution. The types of sources quoted differ significantly between prosecutor groups ($\chi^2 = 1720.3$, $p \approx 0$, $df = 9$). Main-text summaries are in Section 4.8; this table provides the full source-type breakdown.

Coverage of progressive prosecutors features proportionally more advocacy group voices (2.3×), journalist commentary (1.5×), politician quotes (1.5×), and expert sources (1.4× per article). No-

Table 6: Source type distribution by prosecutor ideology. Proportions are within-group; rates are per attributed article (progressive $n = 6{,}601$; traditional $n = 6{,}352$).

Source Type	Prog. %	Trad. %	Prog./art.	Trad./art.	Ratio
Police	11.7%	11.6%	0.74	0.71	1.0×
Victim	2.4%	2.4%	0.15	0.15	1.0×
Defense attorney	6.7%	9.2%	0.42	0.56	0.8×
Prosecutor (self)	33.1%	42.2%	2.08	2.56	0.8×
Politician	10.0%	7.0%	0.63	0.43	**1.5**×
Community member	12.2%	11.3%	0.77	0.68	1.1×
Expert	7.7%	5.8%	0.48	0.35	**1.4**×
Journalist	7.2%	4.9%	0.45	0.30	1.5×
Advocacy group	6.9%	3.1%	0.43	0.19	**2.3**×
Other/missing type	1.9%	2.5%	0.12	0.15	0.8×

tably, traditional prosecutors receive more defense-attorney (0.8×) and self-quote (0.8×) sources per article, while police, victim, and community sources appear at roughly equal rates. These sourcing differentials suggest that progressive prosecutor coverage draws on a broader and more politically diverse source ecology, while traditional prosecutor coverage relies more heavily on the prosecutor's own statements and legal actors. This is consistent with the accountability and reform framing patterns documented in Section 4, and with the "toplash" dynamics described by Goldstein (2024): the elevated presence of advocacy, politician, and expert sources around progressive prosecutors is precisely the source ecology one would expect when institutional actors mobilize media relationships to contest reform agendas.

Source stance details are reported in the main text (Table 5); briefly, progressive prosecutor coverage has higher critical and supportive source rates, while traditional coverage has higher neutral source rates, largely driven by self-quotes (2.56 vs. 2.08 per article).

Causal claims. Progressive prosecutors face 1.7× as many "prosecutor caused harm" attributions per article (0.41 vs. 0.24). "Prosecutor helped" attributions are also 1.2× higher per article for progressives (0.15 vs. 0.13). The asymmetric harm ratio identifies a specific structural mechanism through which progressive prosecutors receive more negative coverage: causal blame is attributed more frequently regardless of evidence quality.

Claim types. Per article, progressive prosecutors face more claims across all categories, but the differentials are sharply asymmetric by type. Policy (0.48 vs. 0.31, 1.6×) and performance (0.45 vs. 0.31, 1.5×) criticisms show the largest gaps, while character (0.42 vs. 0.36, 1.1×) and competence (0.31 vs. 0.29, 1.1×) claims are relatively symmetric across ideology.

Per-prosecutor extraction density. Pamela Price generates the most extraction-dense coverage (13.3 per article, $n = 828$), followed by Brooke Jenkins (10.8, $n = 3{,}207$), Chesa Boudin (10.7, $n = 5{,}773$), Nancy O'Malley (10.0, $n = 1{,}497$), and Steve Wagstaffe (7.9, $n = 1{,}648$).[1]

Limitations

The extraction covers the full corpus of prosecutor-attributed articles ($n = 6{,}601$ progressive, $n = 6{,}352$ traditional), eliminating sampling bias as a concern. However, the LLM-based extraction itself introduces measurement uncertainty: Gemini 2.5 Flash may systematically over- or under-extract certain categories, and extraction quality likely varies with article length and complexity. Without a human-coded gold standard, extraction accuracy cannot be validated directly. The source type asymmetry, particularly the advocacy-group disparity (2.3×), suggests that tone differences detected in the main analysis may partly reflect differential journalistic sourcing practices rather than editorial bias per se.

Appendix B LLM-Based Bias Indicator Extraction (Pilot Study)

The structural extraction in Appendix A characterizes *what* articles contain; this appendix reports a complementary pilot that asks *how* the coverage is biased. A well-sourced, evidence-rich article can be legitimately critical of a prosecutor without being biased, while a superficially neutral article can exhibit bias through ungrounded assertions, source imbalance, loaded language, or omission

[1] One might assume that progressive-prosecutor articles are simply richer or more detailed, producing more extractable content. The density data do not support this: a traditional prosecutor (Jenkins, 10.8) produces slightly *more* extractable content per article than the most-covered progressive (Boudin, 10.7), and the three middle prosecutors cluster between 10.0 and 10.8 regardless of ideology. The sourcing and claim differentials reported above reflect differences in *what kind* of content articles contain, not in how much.

of relevant context.

Method

Using the same langextract + Gemini 2.5 Flash infrastructure, I designed a schema targeting four categories of bias indicator: (1) *ungrounded negative claims*; (2) *source prominence imbalance*; (3) *loaded language*; and (4) *missing context*. A key methodological challenge is preventing the model from conflating legitimate criticism with bias, addressed through three calibration examples demonstrating that criticism with evidence is not bias. The extraction was applied to a balanced stratified sample of 200 articles (100 progressive, 100 traditional).

Extraction prompt

The following prompt was passed to the model for every article. Unlike the structural extraction in Appendix A, this schema explicitly distinguishes legitimate criticism from bias through calibration guidance:

```
Analyze this news article about a district attorney / prosecutor for BIAS
INDICATORS -- patterns of coverage that reveal systematic slant rather than
fair reporting.  Use exact text spans from the article.

Only extract genuine bias indicators.  Legitimate critical reporting with
evidence is NOT bias.  Balanced articles may have zero extractions.
```

1. `ungrounded_negative_claim`: A negative assertion about the prosecutor that lacks adequate supporting evidence within the article.
   Attributes: `claim_content` (performance | policy | character | competence), `evidence_quality` (none | anonymous_source | single_anecdote | stats_without_baseline | adequately_sourced), `systemic_blame` (true | false).

2. `source_prominence_imbalance`: A source given disproportionate prominence relative to other viewpoints, creating an imbalanced picture.

Attributes: `source_stance` (critical | supportive), `prominence_mechanism` (placement_lede | placement_closing | extended_quote | sole_named_source | no_counterbalance | headline_framing), `counterbalance_present` (true | false).

3. `loaded_language`: Emotionally charged, non-neutral word choices that go beyond factual reporting.
   Attributes: `language_type` (pejorative_label | scare_quotes | presuppositional_verb | hyperbole | ideological_framing | dehumanizing), `target` (prosecutor | prosecutor_policy | prosecutor_supporters | other), `position` (headline_or_lede | body | closing).

4. `missing_context`: Important context that is conspicuously absent, making the coverage misleading.
   Attributes: `what_is_missing` (trend_context | comparison_baseline | policy_rationale | systemic_factors | legal_constraints | prosecutor_response | alternative_explanation), `claim_it_supports` (free text).

CALIBRATION GUIDANCE:

- A well-sourced article that is critical of a prosecutor is NOT biased if the criticism is grounded in evidence and the prosecutor's perspective is included.
- An article that is favorable to a prosecutor IS biased if it cherry-picks achievements, omits failures, or relies on uncritical source selection.
- Apply the same standard regardless of which prosecutor is covered.

Only extract items that are clearly present. Do not fabricate or infer. Extract text spans in order of appearance. Do not paraphrase.

Few-shot examples

Three examples were provided to calibrate the model's threshold for flagging bias. Critically, Example 2 demonstrates that a well-sourced critical article should produce *minimal* bias indicators, preventing the model from conflating criticism with bias.

Example 1: Heavily biased anti-progressive article.

> *"San Francisco's failed experiment: How Chesa Boudin let criminals run wild. Since the radical progressive prosecutor took office, residents say the city has become unrecognizable. Car break-ins are rampant, shoplifters brazenly clear shelves with impunity, and violent offenders walk free hours after arrest. 'He cares more about criminals than victims,' said retired officer Jim Walsh, who spent 30 years on the force. Walsh described case after case of suspects released without charges. A business owner on Market Street, who asked not to be named for fear of retaliation, said Boudin's policies have destroyed the neighborhood. Critics point to a 12% rise in property crime, though experts note similar trends across major cities regardless of prosecutor ideology. Boudin's office did not respond to requests for comment."*

Class	Extracted Text	Key Attributes
loaded_lang.	"San Francisco's failed experiment"	presuppositional_verb, policy, headline
loaded_lang.	"let criminals run wild"	hyperbole, prosecutor, headline
loaded_lang.	"the radical progressive prosecutor"	pejorative_label, prosecutor, body
ungrounded	"shoplifters brazenly clear shelves with impunity, and violent offenders walk free hours after arrest"	performance, none, systemic_blame
source_imbal.	"retired officer Jim Walsh…case after case of suspects released"	critical, extended_quote, no counterbalance
ungrounded	"A business owner…asked not to be named…Boudin's policies have destroyed the neighborhood"	policy, anonymous_source
missing_ctx.	"Critics point to a 12% rise in property crime"	comparison_baseline
missing_ctx.	"Boudin's office did not respond to requests for comment"	prosecutor_response

Example 2: Well-constructed critical article (minimal bias).

"Alameda County DA Pamela Price's first year in office has drawn sharp scrutiny. An East Bay Times analysis of court records found that Price's office offered plea deals in 73% of violent felony cases, compared to 58% under predecessor Nancy O'Malley. The analysis examined 412 cases filed between January and September 2023. Victims' rights advocate Maria Chen said the plea rates concern families who expected tougher prosecution. Price defended the approach in a press conference, saying her office prioritizes cases with the strongest evidence and that conviction rates on cases taken to trial have actually increased to 89%. Criminal

justice professor David Lang at UC Berkeley noted that higher plea rates do not necessarily indicate leniency, as they may reflect more realistic case assessment. The recall campaign against Price has seized on the plea statistics."

Class	Extracted Text	Key Attributes
missing_ctx.	"The recall campaign against Price has seized on the plea statistics"	systemic_factors (plea rates attributable to case composition, not ideology alone)

This example is critical for calibration: the article contains sharp criticism grounded in court records, includes the prosecutor's response and an expert perspective, and produces only one extraction (a minor missing-context flag). This teaches the model that evidence-based critical reporting is *not* bias.

Example 3: Pro-traditional puff piece.

"New DA Brooke Jenkins is restoring order to San Francisco. In her first six months, the tough-on-crime prosecutor has brought a sense of accountability that residents desperately needed. Jenkins announced felony charges against a prolific shoplifter, drawing praise from the business community. 'Finally, someone who takes our concerns seriously,' said Union Square Alliance director Tom Richards, who described Jenkins as a breath of fresh air after the chaos of the Boudin era. Police Chief Bill Scott lauded the improved cooperation between his department and the DA's office. Crime data for the period is not yet available from official sources."

Class	Extracted Text	Key Attributes
loaded_lang.	"restoring order to San Francisco"	presuppositional_verb, prosecutor, headline
loaded_lang.	"the tough-on-crime prosecutor has brought... accountability that residents desperately needed"	ideological_framing, prosecutor, body
source_imbal.	"Tom Richards... a breath of fresh air after the chaos of the Boudin era"	supportive, extended_quote, no counterbalance
source_imbal.	"Police Chief Bill Scott lauded the improved cooperation"	supportive, no_counterbalance
missing_ctx.	"Crime data for the period is not yet available"	trend_context (effectiveness claimed without evidence)

This example demonstrates that the schema applies symmetrically: a favorable article about a traditional prosecutor triggers loaded-language, source-imbalance, and missing-context indicators just as readily as a negative article about a progressive prosecutor.

Results

Primary test. Progressive prosecutors receive a slightly more negative mean bias score ($M = -0.143$) than traditional prosecutors ($M = -0.111$), but the difference is not statistically significant: $t(198) = -0.83$, $p = .406$, $d = -0.118$, bootstrap 95% CI $[-0.105, +0.043]$. The null primary result contrasts with the significant $d = -0.157$ from the main pipeline (Section 4).

Per-indicator decomposition. Individual bias indicators show small, mixed effects (see Figure 12):

Loaded negative language is the only indicator favoring the progressive bias hypothesis ($d = +0.14$): progressive prosecutors receive more emotionally charged negative word choices. Para-

Table 7: Bias indicator means by prosecutor ideology (Appendix B pilot, $n = 200$).

Indicator	Progressive M	Traditional M	d
Ungrounded claims	0.50	0.74	−0.12
Ungrounded (severe)	0.36	0.62	−0.14
Systemic blame	0.23	0.23	0.00
Source imbalance (critical)	0.64	0.67	−0.02
Source imbalance (supportive)	0.10	0.16	−0.16
Loaded language (total)	1.45	1.19	+0.10
Loaded language (negative)	0.55	0.38	**+0.14**
Loaded language (headline)	0.24	0.25	−0.02
Missing context	0.77	0.95	−0.12

doxically, ungrounded claims are *higher* for traditional prosecutors ($d = -0.12$; the severe subcategory specifically: $d = -0.14$). The distribution of *prominence mechanisms* differs significantly across groups ($\chi^2 = 14.58$, $p = .012$, $df = 5$): traditional prosecutor coverage relies more on sole named sources (25 vs. 7 instances), while progressive prosecutor coverage exhibits more "no counterbalance" patterns (16 vs. 10).

Language type decomposition. Disaggregating loaded language by type reveals a qualitative distinction in *how* bias manifests. Progressive prosecutor coverage exhibits more presuppositional verbs (34 vs. 24 instances) and ideological framing (41 vs. 24 instances): language that embeds assumptions about the prosecutor's politics or takes contested claims as given. Traditional prosecutor coverage shows more pejorative labels (35 vs. 27 instances). The *mechanism* of bias thus differs: progressive prosecutors are framed through an ideological lens where their reform identity is constantly invoked and assumed; bias against traditional prosecutors, when it occurs, is expressed more bluntly. Evidence quality further distinguishes the groups: articles about traditional prosecutors more often contain ungrounded negative claims with *no* supporting evidence at all (53 vs. 29 instances).

Per-prosecutor decomposition. The most striking pattern in the pilot data is that Brooke Jenkins, the traditional San Francisco prosecutor, receives the worst-quality coverage in the sample by standard journalistic metrics: 0.93 ungrounded claims per article (vs. Boudin's 0.47), 0.84 severe

ungrounded claims (vs. 0.35), and 1.12 missing context instances (vs. 0.74). By contrast, Nancy O'Malley receives the best-quality coverage of any prosecutor (0.26 ungrounded claims/article, 1.6 total indicators/article). This intra-traditional heterogeneity means that the aggregate null result for the traditional group conceals substantial within-group variation. The Jenkins finding is consistent with the politically charged nature of her appointment: she replaced a highly controversial progressive predecessor and was herself a focal point of the Boudin recall narrative, rather than evidence of systematic pro-progressive favoritism.

Convergent validity. The bias score correlates moderately with the aggregate tone-evaluation index from the main pipeline: Pearson $r = 0.257$, $p = .0002$; Spearman $\rho = 0.237$, $p = .0007$. The two approaches capture overlapping but distinct constructs: the NLP pipeline measures aggregate tone and evaluative stance, while the LLM extraction targets specific journalistic violations.

Per-prosecutor ranking. Pamela Price receives the most negative bias score (-0.158), followed by Chesa Boudin (-0.140), Brooke Jenkins (-0.118), Steve Wagstaffe (-0.104), and Nancy O'Malley (-0.100).

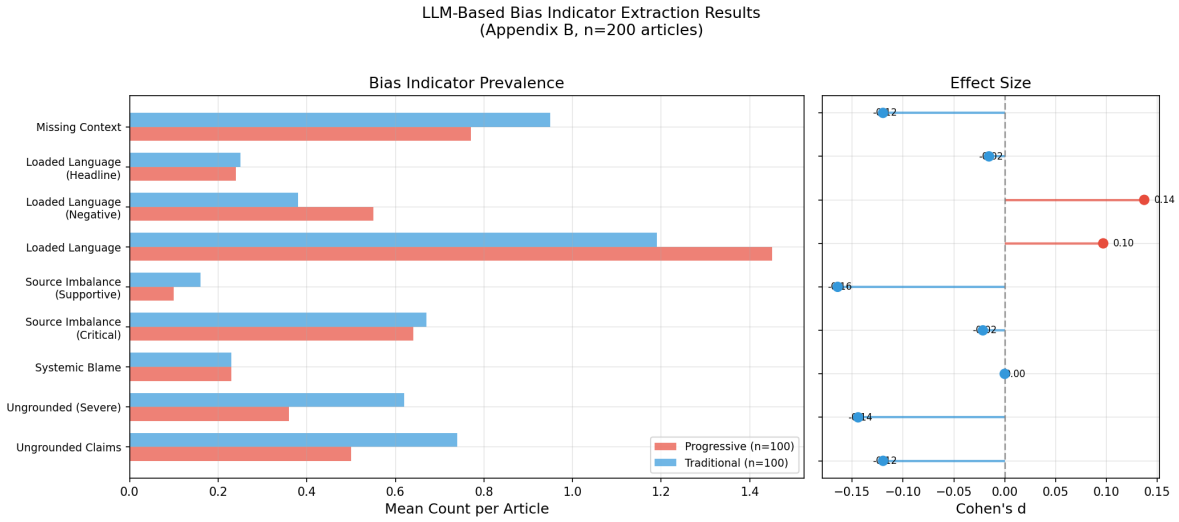

Figure 12: LLM-based bias indicator extraction results (Appendix B pilot, $n = 200$). Left panel: mean indicator counts by ideology. Right panel: Cohen's d effect sizes for each indicator. The only indicator favoring the progressive bias hypothesis is loaded negative language ($d = +0.14$); ungrounded claims are paradoxically higher for traditional prosecutors.

Discussion

The null primary result ($d = -0.118$, $p = .406$) contrasts with the significant $d = -0.157$ from the main pipeline (Section 4). The interpretation is that media bias toward progressive prosecutors operates through tone and framing, the aggregate evaluative register of coverage, rather than through discrete, identifiable journalistic violations (ungrounded claims, source imbalance, missing context).

The loaded language finding ($d = +0.14$) is particularly important here: it provides the mechanistic bridge between what the stance classifier detects as a "vibe" and what is observable in the text as a specific lexical practice. The main pipeline's stance effect operates through how articles frame prosecutors evaluatively; the loaded language result identifies the textual mechanism: negative word choices carrying evaluative loading beyond their denotative content, presuppositional verbs that embed assumptions about performance, and ideological labels that invoke the prosecutor's reform identity as a frame for criticism. Progressive prosecutors are not simply described more harshly; they are described through language that *presupposes* rather than asserts the critical evaluation. This is exactly the kind of subtle framing effect that aggregate tone measures (ungrounded claims, missing context) would not capture, because the individual sentences may be technically accurate. Loaded language is therefore not a minor corroborating finding but the observable textual trace of the stance mechanism: the mechanism by which evaluative framing enters coverage without constituting a discrete journalistic violation. The moderate convergent validity ($r = .26$) between the two pipelines is consistent with this view: they overlap because loaded language is tone-adjacent, and they diverge because the rest of the bias indicator schema targets a different domain.

The per-prosecutor and language-type decompositions further reinforce this. Progressive prosecutor bias operates through ideological framing and presuppositional language, subtle mechanisms that do not register as discrete violations. Traditional prosecutors, particularly during politically contested transitions (Jenkins replacing Boudin), are not immune from low-quality coverage; the

aggregate null result conceals substantial within-group heterogeneity. These findings reinforce the main paper's argument that bias detection requires multi-dimensional measurement: an approach sensitive only to explicit violations would miss the evaluative framing that drives the main effects.

This remains a pilot study ($n = 200$). The Alameda comparison lacks statistical power ($n = 15$ and $n = 23$). A full-corpus extraction is planned to achieve adequate power and enable publication-level subgroup analyses.

Appendix C Supplementary Diagnostic Figures and Tables

This appendix consolidates diagnostic visuals and supporting tables moved from the main text for readability.

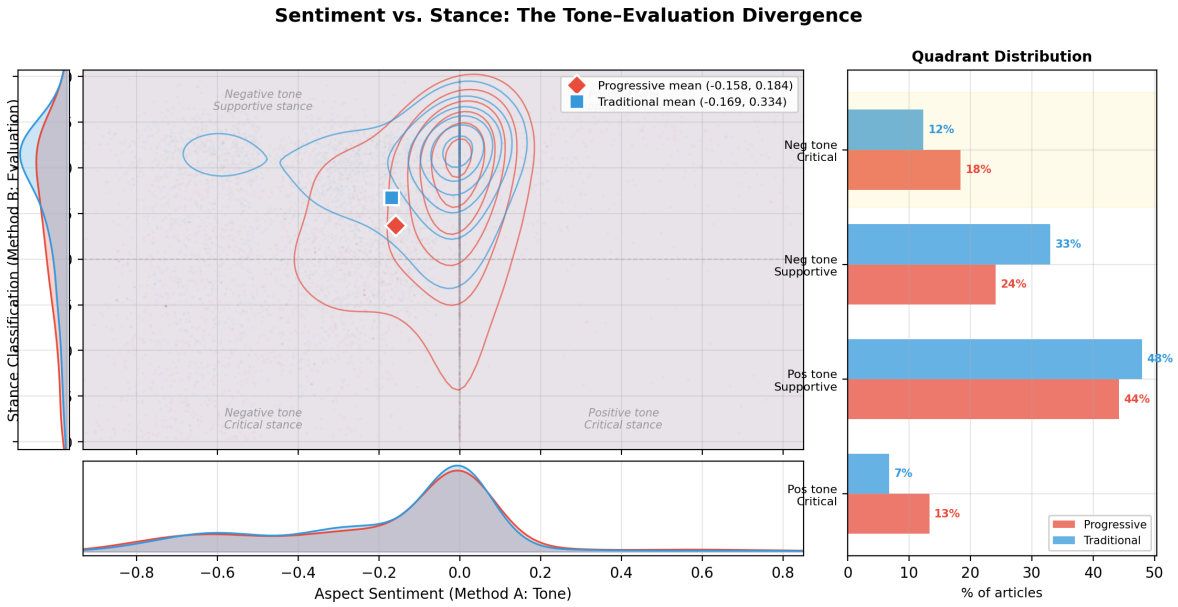

Figure 13: Tone–evaluation divergence at the article level. Central panel: KDE contour plot of Method A (aspect sentiment) vs. Method B (stance classification) scores for progressive (red) and traditional (blue) prosecutors. The horizontal separation in stance contrasts with the overlapping sentiment distributions, confirming that media coverage differs in evaluative framing rather than emotional tone.

Table 8: Per-method OLS regressions: progressive prosecutor coefficient with cluster-robust SEs.

Method	β	SE	p	95% CI	R^2
B: Stance	$\mathbf{-0.132}$	0.030	$< .001$	$[-0.190, -0.074]$	.071
C: Keywords	$\mathbf{-0.011}$	0.002	$< .001$	$[-0.015, -0.007]$	.053
A: Aspect sentiment	0.008	0.016	.618	$[-0.024, 0.040]$	.008
D: Document sentiment	0.017	0.014	.219	$[-0.010, 0.044]$	.006
Composite	-0.022	0.013	.089	$[-0.047, 0.003]$	.022

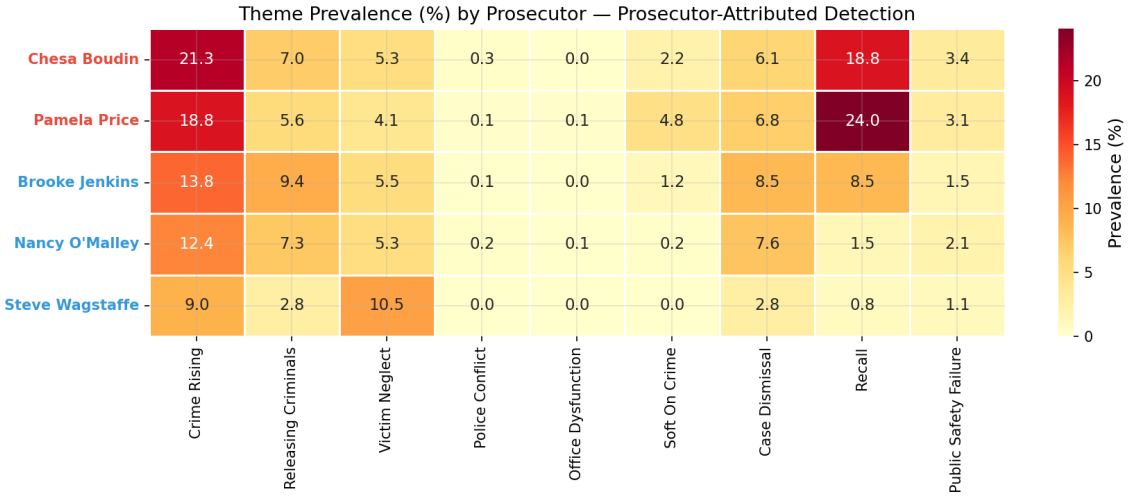

Figure 14: Theme prevalence (%) by individual prosecutor. Rows are colored by ideology (red = progressive, blue = traditional). Darker cells indicate higher prevalence.

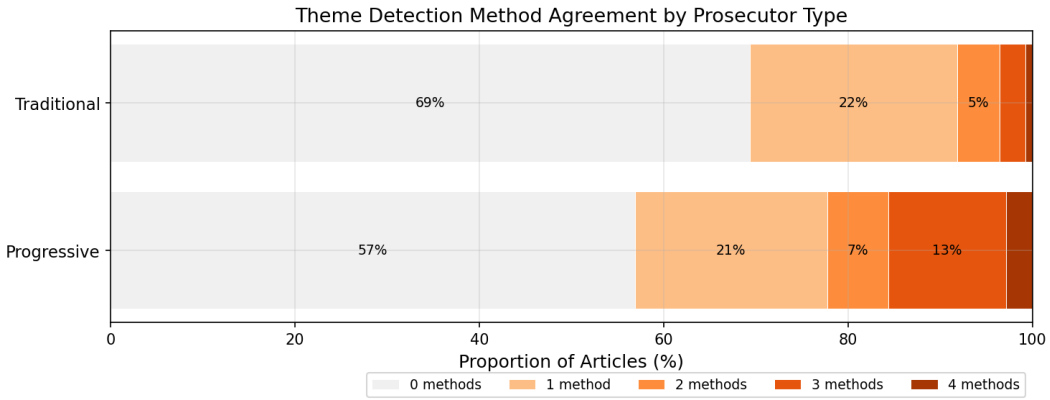

Figure 15: Distribution of detection method agreement by prosecutor type. Segments show the proportion of articles detected by 0–4 independent methods.

Appendix D Illustrative Examples of Identified Bias by Method

This appendix presents concrete examples from the corpus to illustrate what each method detects. Examples are drawn from articles at the extremes of each method's score distribution and are intended to make the abstract measurement pipeline tangible.

Method A: Aspect-Based Sentiment Analysis

Method A extracts three-sentence context windows around prosecutor mentions. The following window scored $A = -0.932$ (*San Francisco Chronicle*, about Pamela Price):

> "Don't blame Alameda DA Price for crime. The pandemic changed everything, and we failed low-opportunity communities, especially our youth; crime in our cities has climbed in some areas and it is frightening."

Although the article's argument is *defensive* of Price, the RoBERTa model responds to the emotional valence of surrounding language, including "crime," "climbed," and "frightening," rather than the argumentative intent. By contrast, a reader letter produced $A = +0.955$: "I am so thankful for Alameda County District Attorney Pamela Price. She has taken on an enormous job and she is doing it well."

Method B: Zero-Shot Stance Classification

A paragraph from kron4.com received a critical stance score of $B = -0.996$:

> "[The letter] cites failed leadership, the movement to defund the police and the district attorney's unwillingness to charge and prosecute people who commit life threatening crimes."

BART-MNLI classifies this as criticizing the prosecutor's handling of crime with near-certainty. The divergence between Methods A and B ($d = 0.037$ for sentiment vs. $d = -0.341$ for stance)

is visible in individual examples: an article can score near zero on sentiment while scoring highly critical on stance.

Method C: Enhanced Keyword Analysis

An article about Nancy O'Malley (berkeleyside.org) triggered three themes simultaneously: **soft-on-crime** ("very **light sentences** or no sentences at all"), **releasing criminals** ("Boudin has **released** a person accused of a crime"), and **recall** ("announced a **recall campaign** against O'Malley"). The keywords capture thematic framing rather than sentiment: "light sentences" is emotionally neutral but thematically codes the soft-on-crime narrative.

Method D: Document-Level Sentiment

Because crime reporting is inherently negative, Method D reveals a paradoxical pattern: articles about traditional prosecutors are slightly *more* negative in overall tone ($d = +0.047$, $p = .008$). This underscores why prosecutor-specific methods are necessary.

Media Framing Analysis

The framing analysis classifies paragraphs (including standalone headlines) into five narrative frames. Illustrative examples include:

Table 9: Illustrative examples of each dominant media frame (Appendix C).

Frame	Example Headline	Prosecutor	Score
Accountability	"SF DA Boudin Blamed by Some for Release of Parolee…"	Boudin	1.00
Conflict	"Jenkins blames Boudin in decision to drop case…"	Jenkins	1.00
Human interest	"Family of woman killed in S.F. New Year's hit and run…"	Boudin	1.00
Reform	"As new D.A., Jenkins vows to be tougher than Boudin"	Jenkins	1.00
Consequences	"Price's office lags on providing charging data"	Price	0.99

Progressive prosecutors are nearly twice as likely to receive accountability framing (39.7% vs. 22.9%) and more than three times more likely to receive reform framing (4.3% vs. 1.3%), while

traditional prosecutors receive substantially more human-interest coverage (35.6% vs. 21.4%; see also Figure 8).

Prosecutor-Attributed Theme Detection

The theme attribution method ($d = -0.43$) requires anti-prosecutor themes to be explicitly linked to a named prosecutor through compound linguistic patterns, distinguishing it from the proximity-based keyword method (Method C). An article titled "Boudin Blunders: SF DA's downfall leading up to recall" (*kron4.com*, 2022-06-08) received the corpus's highest theme attribution score (21.3), with four independent detection methods identifying themes including crime-rising, soft-on-crime, case-dismissal, recall, and public-safety-failure. By contrast, a routine crime article about a man accused of hitting a woman with a car (*San Mateo Daily Journal*, 2021-09-24) attributed to Wagstaffe received a theme score of 0.0, illustrating the method's specificity in not flagging ordinary crime coverage merely because a prosecutor is mentioned.

LLM-Based Bias Indicator Detection

A *San Francisco Chronicle* article about Tenderloin drug enforcement produced 19 bias indicators, including **loaded language** ("drug pushers" as pejorative label; "a stunning 70% increase" as hyperbole), **ungrounded claims** ("arrested and released with few real consequences" without case data), and **missing context** ("Boudin and Public Defender Mano Raju declined to comment"). Critically, many articles produce zero indicators: evidence-based critical reporting is not flagged as biased. The method also detects bias *favoring* prosecutors, supporting the finding that bias is not unidirectional.

Appendix E Sensitivity to Traditional Baseline Composition

Brooke Jenkins accounts for more than half of the traditional baseline ($n = 3{,}207$ of $n = 6{,}352$ traditional articles). Because this can attenuate pooled traditional estimates, I recomputed group

comparisons after excluding Jenkins and retaining all progressive articles ($n = 6{,}601$ progressive; $n = 3{,}145$ traditional).

Table 10: Per-method sensitivity to excluding Brooke Jenkins from the traditional baseline.

Outcome	Full sample (d, p)	Excluding Jenkins (d, p)
Composite	$d = -0.157$, $p = 4.05 \times 10^{-19}$	$d = -0.230$, $p = 1.25 \times 10^{-29}$
Stance	$d = -0.341$, $p = 1.08 \times 10^{-64}$	$d = -0.466$, $p = 1.05 \times 10^{-89}$
Keywords	$d = -0.218$, $p = 1.91 \times 10^{-35}$	$d = -0.313$, $p = 2.89 \times 10^{-71}$
Aspect sentiment	$d = 0.037$, $p = 0.058$	$d = 0.027$, $p = 0.249$
Document sentiment	$d = 0.047$, $p = 0.008$	$d = 0.024$, $p = 0.253$

Excluding Jenkins strengthens the composite, stance, and keyword differentials materially, while sentiment metrics remain small and comparatively weak. This indicates that full-sample estimates are conservative with respect to evaluative dimensions.

For the complementary robustness check that removes fallback-attributed articles instead of a specific prosecutor baseline component, see Appendix F.

Appendix F Sensitivity Analysis Excluding Fallback Attributions

In the attribution pipeline, fallback assignments occur when an article contains generic DA references (e.g., "the district attorney") but no named prosecutor mention; in these cases, the article is assigned by publication county and article date. This mechanism expands coverage but can introduce measurement noise when outlets report cross-county stories.

Using the article-level analysis sample from Section 4 ($n = 12{,}953$), fallback assignment accounts for 6,630 articles (51.2%). Excluding these fallback-attributed cases leaves 6,323 directly at-

tributed articles (named-mention based), with $n = 3{,}341$ progressive and $n = 2{,}982$ traditional.

Table 11: Composite group comparison: full sample vs. excluding fallback attributions.

Sample	n_{Prog}	M_{Prog}	n_{Trad}	M_{Trad}	d	p_{Welch}
Full attributed sample	6,601	-0.0317	6,352	-0.0020	-0.157	4.05×10^{-19}
Exclude fallback-assigned	3,341	-0.0387	2,982	$+0.0241$	-0.310	3.86×10^{-35}

The no-fallback estimate is approximately double the full-sample effect size in magnitude. The bootstrap mean difference (progressive minus traditional) is -0.0628 with 95% CI $[-0.0728,\ -0.0527]$, confirming that the progressive group remains significantly more negative after removing all fallback assignments. Substantively, this indicates that fallback attribution is conservative in this corpus: it attenuates the ideology differential rather than driving it.

Appendix G Segmented Interrupted Time-Series Robustness

The main text reports a concise segmented ITS summary for temporal dynamics by method. This appendix provides the full model specification, coefficient table, and plots for those temporal analyses:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 \text{Post}_t + \beta_3 \text{TimeAfter}_t + \varepsilon_t$$

where β_2 captures an immediate level shift at transition and β_3 captures post-transition slope change. Models are weighted by monthly article counts and use Newey–West (HAC) robust standard errors.

Table 12: Segmented ITS results (monthly, weighted, HAC-robust).

Transition	Outcome	Level shift β_2	Slope change β_3	12-month net effect
SF (Boudin $\rightarrow$ Jenkins)	Composite	$+0.024$ ($p = 0.248$)	$+0.0033$ ($p = 0.044$)	$+0.063$ ($p = 0.001$)
SF (Boudin $\rightarrow$ Jenkins)	Stance	$+0.055$ ($p = 0.311$)	$+0.0122$ ($p = 0.002$)	$+0.201$ ($p < .001$)
SF (Boudin $\rightarrow$ Jenkins)	Keywords	$+0.006$ ($p = 0.270$)	$+0.0018$ ($p < .001$)	$+0.028$ ($p = 0.002$)
Alameda (O'Malley $\rightarrow$ Price)	Composite	-0.089 ($p < .001$)	$+0.0020$ ($p = 0.505$)	-0.065 ($p = 0.025$)
Alameda (O'Malley $\rightarrow$ Price)	Stance	-0.291 ($p < .001$)	$+0.0005$ ($p = 0.946$)	-0.285 ($p < .001$)
Alameda (O'Malley $\rightarrow$ Price)	Keywords	$+0.005$ ($p = 0.079$)	-0.0023 ($p < .001$)	-0.023 ($p < .001$)

Figure 16 visualizes the same segmented models for the three most policy-relevant outcomes (composite, stance, keywords), plotting observed monthly means and fitted segmented trends around each transition date.

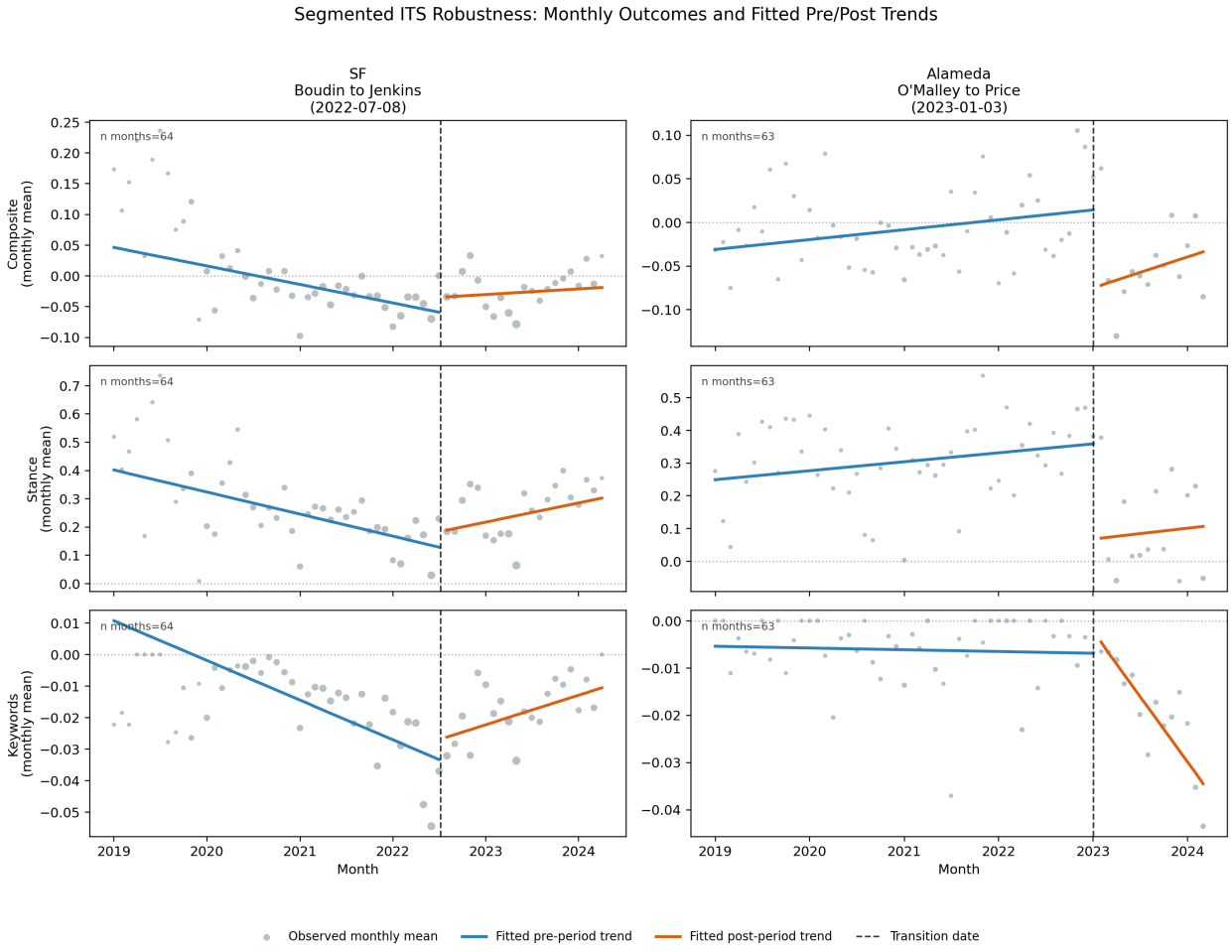

Figure 16: Segmented ITS robustness plots. Points are monthly outcome means (marker size proportional to monthly article count). Colored lines are fitted segmented trends before and after transition; dashed vertical lines mark transition dates.

These segmented models sharpen interpretation by separating immediate transition shifts from post-transition trend changes. In San Francisco, the composite does not show an abrupt level break but does show a modest positive post-transition slope, consistent with gradual movement toward less negative coverage. In Alameda, the composite shows a clear immediate negative level shift at the Price transition with no significant slope change, consistent with a step increase in negative coverage. Across both transitions, stance remains the strongest and most consistent evaluative signal. This robustness analysis improves temporal identification but should still be interpreted descriptively: without an external control series, it does not establish strict causal effects.