Balancing Engagement and Polarization: Multi-Objective Alignment of News Content Using LLMs

Mengjie (Magie) Cheng[+*] Elie Ofek[+] Hema Yoganarasimhan[#]

[+]Harvard Business School
[#]University of Washington

Abstract

We study the problem of how media firms can use LLMs to generate news content that aligns with multiple objectives – making content more engaging while maintaining a preferred level of polarization/slant to keep it consistent with the firm's editorial policy. First, using a dataset of news articles from *The New York Times*, we show that more engaging content is also more polarizing. Further, naively employing LLMs (with prompts or vanilla DPO-based approaches) to make content more engaging can also increase polarization. This has an important policy implication – using LLMs in news media without building in controls for limiting slant can exacerbate media polarization. We present a constructive solution to this problem based on the Multi-Objective Direct Preference Optimization (MODPO) algorithm, a newly proposed approach that integrates Direct Preference Optimization with multi-objective optimization techniques. We build on open-source LLMs and develop a new language model that simultaneously makes content engaging while maintaining a preferred editorial stance and establishing its effectiveness. Our model achieves this by strategically focusing on content strategies strongly associated with polarization, but not with engagement. Our approach and findings apply to other settings where firms seek to use LLMs for content creation to achieve multiple objectives, such as advertising and social media.

Keywords: LLMs, DPO, Content Creation, News, Digital Marketing, Media, Polarization

[*]Please address all correspondence to: macheng@hbs.edu, eofek@hbs.edu, and hemay@uw.edu.

1 Introduction

1.1 LLMs and Generative AI in News Media

Recent advancements in Large Language Models (LLMs) have transformed the landscape of content creation, enabling the scalable and automated generation of content across multiple industries.[1] In particular, the use of generative AI is reshaping the news media and journalism industry. Prominent media companies such as the *Associated Press*, *Forbes*, *The New York Times*, *The Washington Post*, and *ProPublica*, have all adopted AI to help refine/generate articles and automatically report on crimes, financial markets, politics, sporting events, and foreign affairs (Marconi 2020). For example, in September 2020, *The Guardian* published an article entirely written by GPT-3 (The Guardian 2020). More recently, since January 2023, *CNET* has been using automation technology to generate news articles and financial advice, accompanied by a disclaimer: "This article was generated using automation technology, and thoroughly edited and fact-checked by an editor on our editorial staff" (Futurism 2023b). Other publishers, such as *BuzzFeed* and *Men's Journal*, have adopted similar practices.[2]

The appeal of generative AI in journalism is twofold. First, LLMs can rapidly generate large volumes of content at a lower cost than humans, i.e., they are highly efficient and cheap. This efficiency presumably frees journalists to focus on more complex tasks, such as investigative reporting and in-depth analysis (Dörr 2016). Second, generative AI can be used to enhance content quality and boost reader engagement. Editors are increasingly using LLMs to refine rough drafts into polished and compelling articles that resonate with readers and improve the article's visibility in the competitive media landscape (Nishal and Diakopoulos 2024).

1.2 Challenges and Research Agenda

In principle, using LLMs to generate or refine content to make it more appealing to users is straightforward – an editor or journalist can simply provide a basic draft of a news article to the LLM and prompt it to rewrite the article to make it more engaging. However, this naive approach overlooks a fundamental challenge associated with modifying any language-based content (or, more generally, any unstructured data) – changing the language to improve its performance on one objective (e.g., engagement) can inadvertently affect its performance on other dimensions such as the editorial stance of the media firm. In other words, while an LLM can modify text through the choice of words, emotion, tone, emphasis on specific details, and selective omission to increase users' interest or engagement with the content, these changes can also change the slant, polarization, and editorial stance of the article in unexpected ways. As a result, modifying news content to make it more engaging while maintaining the preferred editorial stance is a non-trivial problem.

In this paper, we study the problem of how media firms can use LLMs to align news content on multiple, often conflicting, objectives. Specifically, we focus on how firms can make news content more engaging while controlling the level of polarization/slant to keep it consistent with the firm's editorial policy. In the

[1] Platforms such as https://perma.cc/4VBZ-H5JC and copy.ai demonstrate the ability of generative AI to create content in diverse settings, including novels, essays, headlines, podcast notes, resumes, recipes, computer code, company statements, websites, advertisements, and more.

[2] *BuzzFeed* utilizes an AI assistant, "Buzzy, our creative AI assistant," and notes that articles are "collaboratively written" by its human staff and Buzzy (Futurism 2023a). *Men's Journal* explains its article generation approach as "a curation of expert advice from Men's Fitness, using deep-learning tools for retrieval combined with OpenAI's LLMs for various stages of the workflow" (The Wall Street Journal 2023).

first part of our paper, we document the relationship between engagement and polarization – establishing that these two language objectives tend to move together, i.e., changing one affects the other. Hence, simple prompt-based approaches designed to make content more engaging can also make it more polarizing. In the second part of the study, we present a framework for multi-objective alignment of LLMs in media settings. We build on open-source LLMs and develop a new language model policy that simultaneously makes content more engaging while maintaining a preferred editorial stance (i.e., controlling the extent of polarization), and establish its effectiveness. Finally, in the third part of the paper, we examine how our new language model designed to accommodate multiple objectives differs in content creation strategies compared to simpler language models fine-tuned for a single objective (i.e., maximizing engagement).

1.3 Our Approach and Findings

We use a large dataset consisting of 13,508 news articles from *The New York Times*, published between January 1, 2019, and May 30, 2021, for our main study. These articles span a wide range of topics, including health, business, politics, and the Covid-19 pandemic. For each article, we use an LDA analysis to obtain its topic distribution and employ the "GPT-4o-mini" model to assess both engagement and polarization levels.

First, we find that articles that are more engaging also tend to be more polarizing, even after controlling for the topics covered by the article. This suggests that certain aspects of language, such as structure, flow, and tone, simultaneously influence how engaging the content is as well as how polarizing it is. To further explore the interplay between engagement and polarization, we conduct an exercise that mimics common editorial practice of LLMs: we provide a baseline version of each article to an off-the-shelf LLM (i.e., "GPT-4o-mini") and prompt it to rewrite it in a more engaging fashion. We then reassess both engagement and polarization levels for the recreated articles. We find that these LLM-generated versions have higher engagement scores compared to the original versions. However, this improvement in engagement comes with an unintended consequence – greater polarization. On average, the recreated corpus is significantly more polarizing than the original one.

If prompting LLMs to rewrite articles to increase engagement inadvertently increases polarization, a natural question is whether prompting LLMs to rewrite articles to reduce engagement will lower polarization? Therefore, we conduct another exercise, where we provide the original versions of the articles and prompt "GPT-4o-mini" to generate less engaging versions of these articles. As expected, the new corpus of articles are less engaging than the original one. However, we again observe an unintended effect on the polarization scores – the new corpus is also less polarizing (i.e., more neutral) than the original corpus. Furthermore, the magnitude of this unintended effect on polarization is positively correlated with the degree of change in engagement scores: articles that experience greater increases (or decreases) in engagement also exhibit proportional increases (or decreases) in polarization. In sum, these observations suggest that naively adjusting the engagement levels of news content using prompt-based approaches on off-the-shelf LLMs can inadvertently change an article's overall polarization/slant. This, in turn, can potentially make the content inconsistent with the preferred editorial stance of the media firm.

In the second part of the paper, we propose a solution to address this problem. We develop a new language model policy that simultaneously makes content more engaging while aiming to maintain a preferred editorial stance (or controlling polarization). Our framework builds on open-source LLMs (e.g., Llama and Alpaca)

and leverages the Multi-Objective Direct Preference Optimization (MODPO) algorithm (MODPO; Zhou et al. 2024), a novel approach that combines Direct Preference Optimization (DPO; Rafailov et al. 2024) with multi-objective optimization techniques (Van Moffaert and Nowé 2014; Li et al. 2020) for aligning the LLM to satisfy multiple, potentially conflicting objectives.

MODPO first employs a linear scalarization strategy to transform a multi-objective problem into a single-objective one. This can be achieved by assigning weights to each objective and combining them into a weighted sum, allowing firms to pre-specify the weight of each objective. By iterating over different weight vectors, we can approximate the Pareto front of optimal language models (LMs), thereby offering firms a spectrum of solutions tailored to their strategic priorities. Once the objective function is defined, MODPO follows four key phases: Phase 0 – Model Pre-training, Phase 1 – Supervised Fine-Tuning, Phase 2 – Marginal Reward Modeling, and Phase 3 – Language Modeling. The process begins with model pre-training, a foundational stage where the model learns general linguistic patterns, knowledge, and reasoning capabilities from a vast and diverse dataset. We take the pre-training stage (Phase 0) as given and use a state-of-the-art open-source LLM – Meta's Llama-3-8b (Meta 2024) – as our base model. Next, in the Supervised Fine-Tuning (SFT) phase (Phase 1), we enhance the pre-trained model's ability to follow instructions for a given task using demonstration data. In our main analysis, this task involves generating a more engaging version of an article based on a given base version. To achieve this, we construct a demonstration dataset by pairing the original content with its more engaging counterpart, resulting in an *engaging SFT model*. However, in our setting, since we need to account for the agent's latent preference for both engagement and preferred editorial stance (i.e., polarization), we introduce the marginal reward modeling phase (Phase 2), where we learn a reward function for the polarization objective based upon a comparison dataset that captures agents' polarization preferences. Lastly, in the language modeling phase, we incorporate the learned reward model from the marginal reward modeling step as an adjustment term into the single-objective DPO loss function. This gives us a new language model, denoted as the *multi-objective DPO model*, designed to generate content that is more engaging without being excessively polarizing (i.e., adheres to the preferred editorial stance of the media firm).

To ensure a robust evaluation, we split our demonstration and comparison data into training (80%), validation (10%), and test (10%) sets. The test set is used for evaluation through Gemini, Claude3, and human raters. As a baseline for comparison, we also develop an *engaging DPO model*, which is designed to generate content optimized solely for engagement by applying the standard DPO approach to maximize the engagement objective without considering polarization. We then recreate all the articles in our test dataset using both the *multi-objective DPO model* and the *engaging DPO model*, evaluate the recreated versions of each article on engagement and polarization, and compare their performance to the original versions.

We find that the *multi-objective DPO model* effectively increases engagement without significantly increasing polarization, i.e., it adheres closely to the original editorial stance of the firm. In contrast, the *engaging DPO model* generates articles that are more engaging, but also more polarizing. This suggests that aligning an LLM to a single objective (i.e., enhance engagement) using DPO can lead to unintended consequences similar to those observed in naive prompt-based approaches. Overall, our results suggest that media firms can effectively leverage LLMs to generate content that is engaging/appealing to users while

minimizing unintended negative consequences such as higher polarization (i.e., a substantial shift in the firm's editorial stance). However, achieving this balance requires a thoughtful approach that combines open-source LLMs with multi-objective alignment methods to ensure firms can develop a language policy that balances their multiple objectives.

We also examine two possible levers to modify the extent to which the *multi-objective DPO model* balances the two objectives – (1) the weight vector, which represents the relative importance of engagement and polarization in phase 3, and (2) the baseline task used for the SFT model in phase 1. As the relative weight on the polarization objective increases, the language model policy changes – it starts producing content that is both less polarizing (and less engaging). Similarly, when changing the reference policy from the *engaging SFT model* to a *less polarizing SFT model*, it shifts the final language model towards a policy that is inherently less polarizing. Thus, if a media firm wants to adopt a more neutral stance, it can do so by increasing the weight on the polarization objective or using this objective directly for supervised fine-tuning. However, given the inherent tension between these two objectives, reducing polarization/ideological slant inevitably comes at the cost of some engagement. Thus, depending on the extent to which a firm cares about adhering to a preferred editorial stance vs. increasing engagement, firms can strategically use these levers to calibrate the language model.

In the third part of the paper, we examine how the *multi-objective DPO model* enhances engagement while controlling polarization. We focus on 12 theory-driven content strategies and find that our language model achieves this balance by strategically adjusting specific content strategies that significantly impact polarization but have minimal impact on engagement, such as *provocative language* and *balanced perspective*. In the process of aligning multiple conflicting objectives, it seems the language model learns to identify and adjust content structure and language that can effectively balance both objectives simultaneously.

Finally, we present an extensive set of robustness checks and extensions to validate our approach and findings. These include: (1) using an alternative open-source LLM (Alpaca) as the base model, (2) ruling out hallucination as a significant concern, (3) validating results with a real-world news dataset, and (4) extending our main analysis where we also use data from *Allsides* to train a different *multi-objective DPO model* that prioritizes left-leaning perspectives while also accounting for right-leaning perspectives. Across all these robustness checks and extensions, we find consistent support for our findings and approach.

1.4 Contributions

Our paper makes several key contributions to the literature. First, from a substantive perspective, we establish a positive relationship between engagement and polarization – the choice of words, tone, emphasis (or omission) on specific details that make content more engaging also tend to make it more polarizing. Furthermore, we show that naively using prompt-based approaches to boost engagement using LLMs can inadvertently intensify polarization and make the recreated content deviate from the firm's editorial priorities. Second, from a methodological perspective, we provide a constructive solution to this problem. We leverage the recently developed Multi-Objective Direct Preference Optimization (MODPO) algorithm and demonstrate how it can be applied to news media settings to effectively enhance engagement while maintaining a preferred level of slant. Our approach is built on open-source LLMs, relatively light on data and computational resources, and can flexibly accommodate multiple objectives and varying preferences over these objectives. Third, from

a managerial perspective, our work provides a practical framework for media firms seeking to create content that is both engaging and aligned with their editorial priorities. The framework is not only applicable to news media but also extends to a variety of business contexts where firms use LLMs for content creation to achieve multiple objectives, such as advertising, social media, and e-commerce platforms. Fourth, we contribute to the understanding of how *multi-objective DPO model* balances multiple objectives. We demonstrate that MODPO articles usually leverage content strategies that strongly influence one objective but have a more limited impact on the other objective.

Finally, from a public policy perspective, our research has significant implications for the impact of generative AI and LLMs on media polarization. A growing body of literature shows that media polarization has been rising over the past few decades and that consumer trust in the media industry is declining (Gentzkow and Shapiro 2008; Gerber et al. 2009; Skelley and Fuong 2022). Our research suggests that increasing the use of generative AI without proper controls to limit polarization can further exacerbate these concerns. Nevertheless, our work also offers a viable solution – thoughtful use of LLMs with careful alignment can help policymakers, news aggregators, and media firms limit polarization while enhancing engagement.

2 Related Literature

First, our work contributes to the growing body of research on the role of LLMs in marketing applications. Early studies on LLMs in marketing primarily focused on simulating human behavior and examining their potential in market research (Li et al. 2024; Gui and Toubia 2023; Brand et al. 2023). More recently, researchers have begun to examine the potential of LLMs for content creation and experimentation. Ye et al. (2024) introduce LOLA, an LLM-Assisted Online Learning Algorithm framework that integrates LLMs with adaptive experimentation to optimize content delivery. Angelopoulos et al. (2024) fine-tune a language model to generate new email creatives for marketing communication and evaluate their performance through a large-scale field experiment. Ellickson et al. (2024) combine the contextual embeddings of LLMs with customer demographics and prior treatment outcomes to model heterogeneous treatments of content creatives. We add to this literature by introducing a multi-objective LLM alignment approach that balances content engagement and polarization/editorial slant in news media settings.

Second, our work also contributes to the literature on how language shapes consumer behavior and market outcomes. Much of this research has focused on understanding which aspects of language drive engagement and content diffusion (Toubia and Netzer 2017; Toubia et al. 2021; Netzer et al. 2019; Berger et al. 2021, 2023; Berger and Packard 2018). Recent studies, such as Banerjee and Urminsky (2024) and Batista and Ross (2024), have examined hundreds of cognitive, motivational, affective, and grammatical cues to understand their impact on engagement with news headlines. While some linguistic constructs, such as emotional intensity (Andersen and Blackburn 2004), consistently enhance engagement, others yield mixed findings. For example, Scacco and Muddiman (2020) find that news articles featuring question headlines reduce engagement intentions, while Lai and Farbrot (2014) report the opposite effect. Building on this literature, our work not only investigates how language affects engagement with news articles but also examines a critically underexplored outcome: polarization. Additionally, we analyze the relative importance of various content features and explore how our algorithm leverages these features to balance the dual objectives of fostering engagement and managing polarization to ensure alignment with the firms' editorial stance.

Finally, our research relates to the literature on media bias and polarization. Although a large body of work has documented the existence of polarization in media text and political speech using Natural Language Processing (NLP) techniques (Groseclose and Milyo 2005; Gentzkow and Shapiro 2010; Martin and Yurukoglu 2017; Gentzkow et al. 2019), our study is the first to document that language shapes both engagement and polarization, and that polarization can be an unintended consequence of media firms' desire to offer more engaging content. We further show that naively using LLMs to increase engagement can also make news content more polarizing and contribute to the overall polarization level of the news landscape. Finally, we provide a constructive solution based on multi-objective alignment that can be used to manage polarization while boosting engagement using LLMs.

3 Setting and Data

Our data consists of 13508 popular articles from *The New York Times* (NYTimes) published between January 1, 2019, and May 30, 2021. The dataset was initially assembled by Yoganarasimhan and Iakovetskaia (2024) to study users' sharing patterns of news articles in emails and on Facebook. This corpus serves as an excellent setting for our research since it covers a wide variety of news articles over a sufficiently long period of time. We provide a brief description of the data here and refer interested readers to the earlier study for details of the data collection process.

For each article in the data, we have information on its metadata (which includes the headline, publication date, abstract, and section) as well as the full text of the article. The articles in the dataset span a variety of topics, and for each article, we have the distribution of topics based on an LDA analysis. Figure A1 in Web Appendix §A shows the 40 most common topics across all articles in the corpus, in decreasing order of popularity. We see that the most prevalent topics in our corpus are *Family*, *Politics*, and *Emotions and Feelings*, while the least prevalent are *Judaism*, *Pets and Animals*, and *Horse Racing and Farms*.

Next, we obtain the engagement and polarization scores for each of these articles using "GPT-4o mini", a LLM developed by OpenAI, known for its strong performance across diverse language tasks (OpenAI 2024). We ask the model to score each article on a scale of 1 to 10 on both polarization and engagement, with higher values indicating the article is more engaging or polarizing. To ensure output consistency, we set the model's "temperature" parameter to zero and score each article on both dimensions once.[3] For details of the prompts used, please refer to Web Appendix §B. The engagement score captures the extent to which an article attracts the readers' attention and keeps them interested, while the polarization score captures the extent to which an article is considered to be politically divisive.[4] Table 1 shows summary statistics for the engagement and polarization scores of the articles in our corpus.

We expect the scores obtained to be realistic and sufficiently accurate representations of human user ratings for three reasons. First, LLM models have been shown to possess remarkable zero-shot learning proficiency, which allows them to understand and respond to queries they were not explicitly trained for. This adaptability

[3]The GPT model is stochastic and, therefore, may generate different responses for the same prompt over different iterations. Setting the temperature to zero ensures stable and consistent responses.

[4]All the articles in our corpus are left-leaning to neutral given the source. So when we ask GPT for a polarization score, we are effectively asking it to score the article on the extent of leftward slant. We also consider an alternative scoring approach, where we define polarization as the extent of extreme views/left-leaning bias in the article and repeat the scoring procedure. The results are largely the same as those shown here.

Table 1: Summary statistics of the engagement and polarization scores of the original corpus.

	Engagement	Polarization
Mean	7.68	5.54
Std	0.82	2.37
Median	8.00	6.00
N	13508.00	13508.00

enables them to handle structured questions on possibly unfamiliar topics, such as scoring news articles on engagement and polarization (Xie et al. 2021). Second, prior research supports the reliability of LLMs in the context of news content. For example, the GPT-3 technical report (Brown et al. 2020) demonstrates that GPT-3 can generate news articles that are often indistinguishable from authentic human-written content, as indicated by human evaluations. These capabilities have further improved with the development of more recent LLM models. Third, LLM ratings have been shown to align well with human ratings in this specific dataset; Yoganarasimhan and Iakovetskaia (2024) compare "GPT-3.5-turbo" ratings and user ratings for polarization scores on this dataset and show that GPT ratings correspond well with those of humans, on average. Finally, to further validate the reliability of the scores in our case, we also collect human ratings for both engagement and polarization on a random subset of approximately 1,000 articles in the data; see Web Appendix §E.3. We find a strong correlation between user ratings and GPT-4o mini's ratings, with correlation coefficients of 0.72 for polarization and 0.65 for engagement.[5]

4 Relationship between Engagement and Polarization

We now examine the relationship between an article's engagement and polarization scores. First, we regress the polarization score of an article on its engagement score and present the results in column (1) of Table 2. The results suggest a strong and significant positive relationship – more engaging articles tend to be more polarizing (β = 1.092, p $<$ 0.001). However, this positive relationship may simply reflect the fact that certain topics are both more engaging and more polarizing. Therefore, in column (2) of Table 2, we present the regression results after controlling for the topic distribution of each article (obtained from the LDA analysis discussed earlier). We see that the positive relationship between engagement and polarization remains robust.[6] This suggests that certain aspects of language may simultaneously influence content's engagement and polarization levels.

To further explore the interplay between engagement and polarization and how they can co-evolve, we consider the perspective of an editor whose goal is to make news content more engaging. As discussed in §1, an increasingly common approach to this problem is to use LLMs to enhance engagement by providing them with baseline news content and then prompting them to create a "more engaging" version of the article. To examine whether such simple prompt-based approaches can work, we use "GPT-4o-mini" to create a more engaging version for each article in the corpus. Details of the prompts used are shown in Web Appendix

[5]Prior research has shown that even among human raters, the agreement rate is only around 70% in most annotation tasks (Ouyang et al. 2022; Liu et al. 2020). Thus, the agreement between GPT ratings and human ratings is similar to the agreement between human annotators.

[6]The positive relationship between polarization and engagement is also found in human ratings (β = 0.728, $p < 0.001$. Similarly, human ratings continue to show a significant positive relationship between polarization and engagement even when controlling for an article's topics (β = 0.724, $p < 0.001$). See Web Appendix E.3 for details.

Table 2: Relationship between Polarization and Engagement in Original Data

	(1) Polarization		(2) Polarization	
Engagement	1.092***	(0.0227)	1.070***	(0.0177)
Constant	-2.852***	(0.175)	-0.508	(0.386)
Topics	No		Yes	
N	13508		13508	
R^2	0.1427		0.5725	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Summary Statistics of Engagement and Polarization Scores of More and Less Engaging Versions

	Engagement Scores		Polarization Scores	
	More engaging version	Less engaging version	More engaging version	Less engaging version
Mean	8.99***	3.10***	5.89***	3.09***
Std	0.11	0.33	2.46	0.95
Median	9.00	3.00	7.00	3.00
N	13508	13938	13508	13508

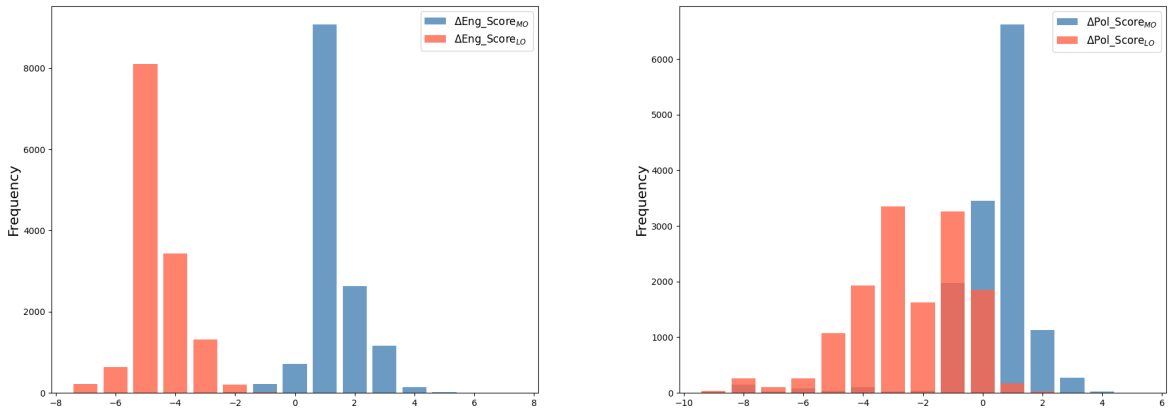
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

§B. Next, we obtain the engagement and polarization scores for this new "more-engaging" version of each article using the LLM-based scoring approach discussed earlier. The results are summarized in the left panel of Table 3. As expected, these LLM generated versions of the articles are rated as more engaging, with the average engagement score increasing to 8.99, compared to 7.68 in the original dataset. However, this improvement in engagement comes with an unintended consequence – greater polarization. We see that the average polarization score of the "more-engaging" articles increases to 5.89, up from 5.53 in the original dataset (Table 1). Two-sample t-tests indicate that both differences are significant. Essentially, the LLM-enhanced versions have moved away from the editorial stance in the original corpus and become more left-polarized. Thus, while this approach can make content more engaging, it does so at the cost of increased polarization, drifting from the original editorial stance.

If increasing engagement also increases polarization, a natural question follows: does reducing engagement lead to lower polarization? To explore this, we prompt the LLM to generate a less engaging version of each article. The details of these prompts are provided in Web Appendix §B. We then evaluate the engagement and polarization levels of these "less-engaging" versions, and present the results in the right panel of Table 3. As expected, the engagement scores for the less engaging versions are significantly lower than those of the original articles, with a mean of 3.09 compared to 7.68 for the original articles (see Table 1). Further, we see that polarization also decreases, with an average polarization score of 3.09, compared to 5.53 in the original corpus. Two-sample t-tests confirm that these differences are statistically significant as well. To illustrate, we provide an example from *The New York Times*, published in May 2020, which reports on "coronavirus parties" as a potential means of gaining immunity. In Web Appendix §C, we include the original article alongside both a "more-engaging" and a "less-engaging" version generated using the "GPT-4o-mini" model. This example shows how changing engagement levels often involves strategic choices in language, emotion, tone, and

emphasis (or omission) on specific details, which, in turn, influence polarization.

So far, we have summarized how the engagement and polarization scores shift at the corpus level. We now examine these changes at the article level. Recall that we now have three versions of each article – (1) the original version, (2) the more engaging version, and (3) the less engaging version. To quantify how engagement and polarization scores of the LLM-enhanced articles change relative to the original, we define four measures: (1) ΔEng_Score$_{MO}$ – the difference in engagement scores between the more engaging version and the original version, (2) ΔEng_Score$_{LO}$ – the difference in engagement scores between the less engaging version and the original version, (3) ΔPol_Score$_{MO}$ – the difference in polarization scores between the more engaging version and the original version, and (4) ΔPol_Score$_{LO}$ – the difference in polarization scores between the less engaging version and the original version. Figure 1 illustrates the distributions of these four difference measures. We see that the distributions of ΔEng_Score$_{MO}$ and ΔPol_Score$_{MO}$ are right-skewed, whereas the distributions of ΔEng_Score$_{LO}$ and ΔPol_Score$_{LO}$ are left-skewed; together these findings confirm the results from the previous analysis.



ΔEng_Score ΔPol_Score

Figure 1: Engagement/Polarization Difference between LLM-Modified Versions and Original Articles

Next, we examine whether the *degree* to which an article's engagement increases or decreases after it is modified to be more (or less) engaging affects the *extent* to which it becomes more (or less) polarizing. Table 4 presents the results of two regressions: (1) the regression of ΔEng_Score$_{MO}$ on ΔPol_Score$_{MO}$, and (2) the regression of ΔEng_Score$_{LO}$ on ΔPol_Score$_{LO}$. In both cases, we find a direct positive correlation between engagement change and polarization change – articles that experience larger increases (or decreases) in engagement also exhibit proportional increases (or decreases) in polarization. These results suggest that when LLMs are used to adjust an article's engagement level, they simultaneously induce a proportional shift in its polarization level.

In sum, our analysis in this section highlights three key takeaways. First, we find a positive correlation between an article's engagement score and its polarization score, even after controlling for its topics. This suggests that the choice of words, emphasis (or omission) on specific details, and tone elements that make

Table 4: Difference in Polarization Scores = f(Difference in Engagement Scores)

	(1) ΔPol_Score$_{MO}$	(2) ΔPol_Score$_{LO}$
ΔEng_Score$_{MO}$	0.559***	
	(0.0161)	
ΔEng_Score$_{LO}$		0.748***
		(0.0157)
Constant	0.00638	-0.326
	(0.321)	(0.343)
Topics	Yes	Yes
N	13508	13508
R^2	0.1612	0.3907

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

content more engaging may also make it more polarizing. Second, we demonstrate that modern LLMs can effectively enhance engagement, offering media firms a powerful tool to increase the appeal of their content. However, this comes with an unintended consequence: naively optimizing news content for engagement using LLMs can amplify polarization. This presents a trade-off for news editors and outlets – while efforts to boost engagement may attract more readers, they may also inadvertently intensify polarization and make content inconsistent with the firm's editorial stance. Finally, our findings have important policy implications. The widespread use of LLMs in news media, without safeguards to limit ideological slant, can exacerbate media polarization. Recognizing this challenge, in the next section, we examine how news outlets and editors can leverage LLMs to enhance engagement while controlling for unintended consequences, such as increased polarization.

5 Problem Formulation

Consider a media firm that has dual goals – (1) make content more engaging, and (2) maintain a specific editorial stance (e.g., a preferred degree of left/right-leaning slant).

- **Enhancing engagement:** The goal of increasing news content engagement is directly tied to revenue generation. Consumers who find content engaging are more likely to spend additional time on the platform, return in the future, and interact with more advertisements. As a result, high news content engagement is likely to translate to greater readership (eyeballs), increased advertising revenue, and potentially greater subscription revenues (Iizuka et al. 2021; SmartNews 2024).
- **Maintaining a preferred editorial stance:** A media firm's commitment to a specific editorial stance can stem from three motivations. First, this editorial stance may reflect the ideological preferences of its owners, journalists, or editors, using the news outlet as a vehicle to promote their viewpoints (Baron 2006). For instance, in February 2025, Jeff Bezos directed *The Washington Post*'s opinion section to focus exclusively on "personal liberties and free markets," explicitly stating that "viewpoints opposing those pillars will be left to be published by others" (Bezos 2025). Similarly, *Fox News* has long upheld a conservative editorial stance, shaped by its founder, Rupert Murdoch, with coverage frequently aligning

with right-leaning perspectives on issues such as gun rights, immigration, and economic policy (News 2024). Second, firms may wish to maintain a preferred level of left/right leaning slant to align with the preferences of their readership because consumers tend to favor news that conforms to their existing beliefs (Mullainathan and Shleifer 2005). Finally, firms may adopt an editorial stance consistent with their readership due to reputational concerns (Gentzkow and Shapiro 2006).[7]

The firm's goal is to develop a content generation policy that effectively balances these dual objectives – enhancing engagement while maintaining a desired editorial stance. While our discussion focuses on these two objectives, the framework and solution can extend to any context, where a firm aims to achieve multiple objectives in its content creation process.[8]

To formalize the firm's problem, we start by defining a few relevant concepts. We first introduce the concept of a Language Model (LM) policy π, which governs how content is generated.

Definition 1. *The LM policy $\pi(y|x)$ represents a machine learning model that assigns probabilities to sequences of words y, given an input sequence x.*

The policy π represents how a language model generates content for a specific task. For example, if the task is question-answering, x is the input question, y denotes the generated answer, and π denotes the LM policy $\pi(y|x)$ used to generate y based upon input x. Similarly, in a summarization task, x denotes the original content, and y represents the generated summary. In our setting, the task is to take original news content (x) and generate a modified version (y) that aligns with a set of K objectives. To formalize each objective k, we introduce an objective-specific reward function, r_k: $X \times Y \rightarrow \mathbb{R}$, that captures how well the response y aligns with objective k, given input x. Building on this, we can further define the optimal language model for a specific objective k as:

Definition 2. *The optimal language model $\pi^*_{r_k}$ for objective k is the policy that maximizes the expected reward for that objective such that $\pi^*_{r_k} = \underset{\pi}{argmax} E_\pi [r_k(x, y)]$.*

The definitions above provide the foundation for developing an optimal language model $\pi^*_{r_k}$ for a single objective k. In the single-objective case, optimizing the policy is straightforward: responses can be rank-ordered based on the reward function $r_k(x, y)$, allowing for the development of an optimal policy. For example, if the firm's sole objective is to increase engagement ($k = 1$), it can develop a policy $\pi^*_{r_1}$ that maximizes the expected reward for engagement given input x (i.e., $E(r_1(x, y))$.

However, in a multi-objective setting, the firm's goal is to develop an optimal language model π^* that aligns with multiple objectives (K). This complicates the development of the optimal policy π^* because there

[7] Gentzkow and Shapiro (2006) show that firms may slant their news content to be consistent with that of their readership because consumers often perceive news content as higher quality when it conforms to their prior expectations. Thus, even when neither the firm nor the consumers have an inherent preference for slanted news, the firm's goal can devolve into reporting slanted news that is consistent with its readership's view of the world. See Gentzkow and Shapiro (2008) for a broader discussion of these issues.

[8] While a simplistic approach to solve this problem is to use prompt engineering on an off-the-shelf LLM such as GPT or Gemini, prompt-based approaches tend to have several problems – poor performance compared to fine-tuned LLMs, particularly for complex or domain-specific tasks (Zhao et al. 2021; Wei et al. 2022; Ye et al. 2024), inability to tune/quantify how the LLM should trade-offs between different objectives, inability to incorporate multiple objectives effectively, scalability, cost, and data and privacy concerns. Therefore, we focus on solutions that employ open-source LLMs on the firm's own historical data and are explicitly calibrated on the firm's preferred balance of objectives using a weight vector.

can be inherent trade-offs across objectives. For example, if the firm also has a secondary objective ($k = 2$), where the goal is to reduce polarization, a dilemma arises. As discussed in §4, a policy that maximizes engagement does not necessarily reduce polarization; in fact, it may increase polarization. One solution is to develop a new policy $\pi^*_{r_2}$ that minimizes the expected reward for polarization given input x (i.e., $E(r_2(x, y))$. This approach, however, presents two key challenges. First, the two policies are likely to be misaligned. Specifically, if $\pi^*_{r_1}$ outperforms $\pi^*_{r_2}$ on engagement ($E_{\pi^*_{r_1}}(r_1(x, y)) > E_{\pi^*_{r_2}}(r_1(x, y))$) but underperforms on polarization control ($E_{\pi^*_{r_1}}(r_2(x, y)) < E_{\pi^*_{r_2}}(r_2(x, y))$), it is unclear which language model is preferable, as each outperforms the other on one of the objectives. Second, this produces two separate policies that accommodate different objectives, whereas, in practice, the firm needs one optimal policy that effectively balances both objectives – maximizing engagement while controlling polarization. A natural solution to this problem is to leverage the concept of Pareto optimality, where the Pareto frontier of the policy space represents the set of policies that are non-dominated by any other policy. Formally, we define our goal as:

Definition 3. *Consider a firm that seeks to achieve K objectives simultaneously. The firm's goal is to find the Pareto front Π_f – a set of policies that are Pareto optimal with respect to each objective k, characterized by the reward function r_k. That is, for each $\pi \in \Pi_f$, there exists no other policy π' such that $E_{\pi'}(r_k) > E_{\pi}(r_k)$ for every objective k.*

A common approach to solving multi-objective optimization problems is linear scalarization (Van Moffaert et al. 2013). Given a non-negative weight vector, $\mathbf{w} = [w_1, w_2, \ldots, w_K]$, where $\sum_{k=1}^{K} w_k = 1$, and a corresponding reward function vector, $\mathbf{r}(x, y) = [r_1(x, y), r_2(x, y), \ldots, r_K(x, y)]$, we can convert the firm's multiple objectives into a single objective using a weighted sum. The weight w_k captures the relative importance of the k^{th} objective for the firm. See Rafieian et al. (2024) and Wang et al. (2024) for recent marketing papers that also use this scalarization approach for multi-objective optimization. In our setting, a firm that cares a lot about engagement and not too much about the editorial stance/polarization may assign a lower weight to polarization, and vice-versa. The firm's goal can be formally defined as follows.

Definition 4. ***Firm's Problem:*** *Given a weight vector ($\mathbf{w} = [w_1, w_2, ...w_K]$), where $\sum_{k=1}^{K} w_k = 1$, and a reward function vector ($\mathbf{r}(x, y) = [r_1(x, y), r_2(x, y), ..., r_K(x, y)]$), the firm's goal is to identify the Pareto-optimal language model (LM) $\pi^*_{\mathbf{w^T r}}$ by optimizing the following objective function:*

$$\pi^*_{\mathbf{w^T r}} = \underset{\pi}{argmax} E_{\pi}\left[\mathbf{w^T r}(x, y)\right]. \tag{1}$$

In summary, we formulate the firm's language model alignment as a multi-objective optimization problem, where each objective has a pre-specified weight that can be chosen by the firm based on its strategic priorities. By iterating over different weight vectors, we can approximate the Pareto front of the optimal language models (LM), denoted as $\{\pi_{\mathbf{w^T r^*}} | \mathbf{w} \in \Omega\}$, thereby providing firms with a spectrum of solutions tailored to their editorial and business objectives.

6 Model

We now provide an overview of the solution concept. In §6.1, we start with a baseline case, where the goal is to optimize a single objective. We introduce the basic notions of reward modeling, Reinforcement Learning

from Human or AI Feedback (RLHF or RLAIF), and Direct Preference Optimization (DPO) algorithms for aligning with a single objective. Building on these ideas, §6.2 extends the discussion to multi-objective settings, where we introduce the Multi-Objective Direct Preference Optimization (MODPO) algorithm as a natural extension for handling multiple competing objectives.

6.1 Baseline Case with a Single Objective

Consider a simple setting where we only have one objective that we want to maximize, i.e., $K = 1$. Then, the firm's objective function simplifies to:

$$\pi_r^* = \underset{\pi}{argmax} E_\pi \left[r(x, y) \right]. \tag{2}$$

In §6.1.1, we describe how this problem can be solved using a reinforcement learning algorithm. Then, in §6.1.2, we discuss some drawbacks of this approach and present the simpler algorithm, DPO.

6.1.1 Reinforcement Learning from Human/AI Feedback

RLHF/RLAIF is a machine learning algorithm that incorporates feedback from humans/AI agents, often in the form of rankings or comparisons, to perform tasks more aligned with the agents' revealed preferences. We provide a brief outline of the RLHF/RLAIF pipeline here and refer readers to prior works for detailed discussions; see Bai et al. (2022a); Ouyang et al. (2022); Lee et al. (2024); Bai et al. (2022b). The pipeline for this algorithm comprises three key stages: (1) supervised fine-tuning (SFT), (2) preference sampling and reward modeling, and (3) reinforcement learning (RL) optimization. We describe each of these stages below.

Supervised Fine-Tuning (SFT): The first step of the process starts with adapting a pre-trained, general-purpose Language Model (LM) to a specific downstream task through supervised learning on demonstration data (D). Formally, we define the demonstration dataset as $D = \left\{(x, y)^{(i)}\right\}_{i=1}^{N}$, where x represents the input, y represents the ideal output, and N denotes the number of observations in the dataset D. This step enhances the model's ability to follow instructions for a given task, thereby laying the foundation for subsequent RL training (Chu et al. 2025). In our context, SFT can enhance the model's ability to produce content that performs a specific task, such as generating modified news articles based on the original article. This process results in a fine-tuned LM, denoted as π_{sft}.

Reward Modeling (RM): In this phase, we use comparison data (C) that captures the preferences of downstream users of the LM to develop a reward model that reflects their preferences. For example, given a prompt x, the comparison data would consist of two responses, y_w and y_l, reflecting the agent's (human or AI) rating or relative preference ranking for y_w and y_l. We denote this preference as $y_w \succ y_l | x$, where y_w and y_l represent the more preferred and less preferred responses, respectively.[9] These preferences in the comparison data are assumed to be governed by a latent reward model $r^*(x, y)$. Assuming preferences follow the Bradley-Terry (BT) model (Bradley and Terry 1952), the preference distribution p^* can be expressed as:

$$p^*(y_w \succ y_l | x; C) = \frac{exp(r^*(x, y_w))}{exp(r^*(x, y_w)) + exp(r^*(x, y_l))} = \sigma(r^*(x, y_w) - r^*(x, y_l)), \tag{3}$$

[9]The subscript w in y_w denotes the *w*inning response and the subscript l in y_l denotes the *l*osing response, respectively.

where σ denotes the sigmoid function.[10] Given a dataset of comparisons $C = \left\{(x, y_w, y_l)^{(i)}\right\}_{i=1}^{N}$ sampled from p^*, we can define a reward model $r_\phi(x, y)$ and estimate its parameters via maximum likelihood to approximate the latent reward model $r^*(x, y)$. This optimization can be framed as a binary classification problem, where the parameters are updated to minimize the loss function, $L_R(r_\phi; C)$, as follows:

$$L_R(r_\phi; C) = -E_{r_\phi}[log\left(\sigma(r_\phi(x, y_w) - r_\phi(x, y_l))\right)]. \tag{4}$$

RL optimization: In the final phase, we use the learned reward function ($r_\phi(x, y)$) to further optimize the language model using reinforcement learning techniques. Specifically, suppose we have a language model (LM) policy parameterized by θ, defined π_θ. We can then align π_θ to the learned reward function $r_\phi(x, y)$ by solving the following optimization problem:

$$\pi^*_{r_\phi} = \underset{\pi_\theta}{argmax} E_{\pi_\theta}\left[r_\phi(x, y) - \beta log \frac{\pi_\theta(y|x)}{\pi_{ref}(y|x)}\right]. \tag{5}$$

The first term in the above optimization problem directly corresponds to the firm's problem defined in Equation (2). The second term represents a Kullback–Leibler (KL) constraint, which serves as a penalty for moving too far from a reference policy. The parameter β controls the degree of deviation from the reference policy π_{ref}, which is usually the initial supervised fine-tuned (SFT) model π_{sft} (from the SFT step discussed earlier). This penalty plays a critical role in fine-tuning LLMs, as it prevents the model from deviating too far from the distribution on which the reward model is valid while preserving generative diversity. Additionally, it mitigates the risk of reward over-optimization, which can degrade the quality of the generated content (Gao et al. 2023). Finally, we note that the above optimization problem is typically solved using reinforcement learning algorithms, such as Proximal Policy Optimization (PPO) (Schulman et al. 2017).

6.1.2 Direct Preference Optimization (DPO)

While RLHF/RLAIF has shown promise in producing high-quality models, it is not without drawbacks. In particular, the third step (reinforcement learning) can be unstable and computationally intensive due to frequent sampling from the policy during training, which entails significant computational costs. Further, RLHF/RLAIF is an *online* algorithm, which means that, at each step of the optimization in Equation (5), the algorithm generates new text based on the updated policy, which then needs to be annotated by humans/AI to get preference ratings. This can be very time-consuming and costly.

Recent research on LLM alignment proposes an alternative approach that is simpler to implement and can perform LLM alignment directly on the policy without the explicit Reward Modeling (RM) step in §6.1.1 using only *offline* data, eliminating the need to continuously generate new text and obtain labels or preferences during training. This approach, proposed by Rafailov et al. (2024), is known as Direct Preference Optimization (DPO). Instead of relying on explicit reward modeling and reinforcement learning, DPO derives a theoretical mapping between the true reward function r^* and the optimal policy, $\pi^*_{r^*}$. This mapping allows the preference loss to be defined directly as a function of the optimal policy, enabling an analytical solution to

[10] Mathematically, the Bradley-Terry (BT) model shares structural and functional similarities with the logit model (Ben-Akiva et al. 1997), which is widely used in economics and marketing to study consumer behavior.

Equation (5), as follows:[11]

$$r^*(x,y) = \beta log \frac{\pi^*_{r^*}(y|x)}{\pi_{ref}(y|x)} + \beta log Z(x). \tag{6}$$

Here $Z(x)$ is the partition function, defined as: $\sum_y \pi_{ref}(y|x) exp(\frac{1}{\beta} r^*(x,y))$.[12] By substituting this reparameterization of $r^*(x,y)$ given in Equation (6) into the preference model in Equation (3), the partition function $Z(x)$ cancels out. This allows us to express the agent's preference probability purely in terms of the optimal policy $\pi^*_{r^*}$ and the reference policy π_{ref}. Thus, the optimal policy $\pi^*_{r^*}$ under the Bradley-Terry model satisfies the following preference model:

$$p^*(y_w \succ y_l | x; C) = \frac{1}{1 + exp\left(\beta log \frac{\pi^*_{r^*}(y_w|x)}{\pi_{ref}(y_w|x)} - \beta log \frac{\pi^*_{r^*}(y_l|x)}{\pi_{ref}(y_l|x)}\right)}. \tag{7}$$

This reformulation bypasses the need for explicit reward modeling by directly connecting human preference probabilities to the optimal policy $\pi^*_{r^*}$. Using this, we can define the maximum likelihood objective for a parameterized policy π_θ. The resulting objective, known as the DPO loss, can be defined as follows:

$$L_{DPO}(\pi_\theta; \pi_{ref}, C) = -E_{\pi_\theta}\left[log\sigma\left(\beta log \frac{\pi_\theta(y_w|x)}{\pi_{ref}(y_w|x)} - \beta log \frac{\pi_\theta(y_l|x)}{\pi_{ref}(y_l|x)}\right)\right]. \tag{8}$$

We can minimize this loss function for the language model policy π_θ by directly using maximum likelihood, without relying on reinforcement learning. In summary, DPO consists of two steps:

- Step 1: Supervised Fine-Tuning (SFT), where we fine-tune the pre-trained language model on the demonstration data to obtain the SFT policy π_{sft}, which serves as the reference policy.
- Step 2: Directly optimize the language model policy π_θ using maximum likelihood estimation on the comparison dataset, based on the loss function shown in Equation (8).

In sum, DPO avoids both the explicit reward modeling in Equation (4) (Step 2 in RLHF) and the computationally intensive reinforcement learning process in Equation (5) (Step 3 in RLHF). As such, it is a more stable, efficient, and practical alternative to LLM alignment.

6.2 Multi-Objective Direct Preference Optimization (MODPO)

While the DPO approach described above works well, it is designed for single-objective settings. However, in our context, we have two objectives – (1) increasing engagement and (2) controlling polarization. To that end, we first adopt the optimization function for the multi-objective setting (Equation (1)) by including a KL constraint similar to the formulation in Equation (5). With this adjustment, our objective function becomes:

$$\pi^*_{\mathbf{w^T r}} = \underset{\pi}{argmax} E_\pi \left[\mathbf{w^T r}(x,y) - \beta log \frac{\pi(y|x)}{\pi_{ref}(y|x)}\right]. \tag{9}$$

[11] This is done by replacing the true reward function r^* in place of the estimated reward model r_ϕ in Equation (5), taking the first-order condition with respect to π_θ, and then normalizing by the partition function $Z(x)$.

[12] The partition function is a mathematical construct that computes the probability of an agent's possible trajectories. It serves as a normalization constant, ensuring that the sum of probabilities across all states sums to one (Rahme and Adams 2019).

To solve the multi-objective alignment problem described above, we adopt the recently proposed Multi-Objective Direct Preference Optimization (MODPO) algorithm (Zhou et al. 2024). This novel algorithm combines ideas from Direct Preference Optimization (DPO; Rafailov et al. 2024) with multi-objective optimization techniques (Van Moffaert and Nowé 2014; Li et al. 2020) to align LLMs on multiple objectives. This method has mainly been used in LLM safety research in computer science, where the goal is to train LLMs to provide answers that are both helpful and harmless in response to sensitive prompts such as, 'how to conduct suicide?' or 'how to make a bomb' (Dai et al. 2024; Ji et al. 2023a; Liu et al. 2024b). Our work extends these principles to the media domain where firms seek to increase engagement while maintaining a preferred editorial stance. We provide a brief summary of this approach below.

Similar to the theoretical mapping of DPO for a single objective as shown in Equation (6), MODPO builds on the theoretical relationship between the weighted ground-truth collective reward models $\mathbf{w^T r}^*$ across multiple objectives and the optimal LM $\pi_{\mathbf{w^T r}^*}$ as follows:

$$\mathbf{w^T r}^*(x,y) = \beta log \frac{\pi^*_{\mathbf{w^T r}^*}(y|x)}{\pi_{ref}(y|x)} + \beta log Z_m(x), \tag{10}$$

where $Z_m(x) = \sum_y \pi_{ref}(y|x) exp(\frac{1}{\beta}\mathbf{w^T r}^*(x,y))$ is the partition function for the MODPO optimization problem. Let $r^*_k(x,y)$ denote the true reward function for the k^{th} objective. Then, as before, we can assume that the agent's preferences for each objective k follow the Bradley-Terry (BT) model. Consequently, the preference distribution p^*_k for two responses y^k_w and y^k_l to the same input x for each objective k is given by:

$$p^*_k(y^k_w \succ y^k_l | x; C_k) = \frac{exp(r^*_k(x, y^k_w))}{exp(r^*_k(x, y^k_w)) + exp(r^*_k(x, y^k_l))} = \sigma(r^*_k(x, y^k_w) - r^*_k(x, y^k_l)). \tag{11}$$

In the above equation, C_k denotes the comparison data for the k^{th} objective. We can then rearrange Equation (10) to parameterize r^*_k as a function of $\pi^*_{\mathbf{w^T r}^*}$, $\mathbf{r}^*_{-\mathbf{k}}$, π_{ref}, and $Z_m(x)$ as follows:[13]

$$r^*_k(x,y) = \frac{1}{w_k}\left(\beta log \frac{\pi^*_{\mathbf{w^T r}^*}(y|x)}{\pi_{ref}(y|x)} + \beta log Z_m(x) - \mathbf{w^T_{-k} r^*_{-k}}(x,y)\right). \tag{12}$$

Substituting the parametrization for $r^*_k(x,y)$ above into the preference model in Equation (11) causes the partition function $Z_m(x)$ to cancel out, and we get:

$$p^*(y^k_w \succ y^k_l | x; C_k) = \tag{13}$$
$$\sigma\left(\frac{1}{w_k}\left(\beta \log \frac{\pi^*_{\mathbf{w}^T\mathbf{r}^*}(y^k_w|x)}{\pi_{\text{ref}}(y^k_w|x)} - \beta \log \frac{\pi^*_{\mathbf{w}^T\mathbf{r}^*}(y^k_l|x)}{\pi_{\text{ref}}(y^k_l|x)} - \mathbf{w}^T_{-k}\left(\mathbf{r}^*_{-k}(x,y^k_w) - \mathbf{r}^*_{-k}(x,y^k_l)\right)\right)\right).$$

The equation above captures the preference distribution for the k^{th} objective. Notice that it is similar in spirit to Equation (7) in the single-objective DPO case, with one key difference – an additional term, $\mathbf{w}^T_{-k}\left(\mathbf{r}^*_{-\mathbf{k}}(x,y^k_w) - \mathbf{r}^*_{-\mathbf{k}}(x,y^k_l)\right)$, which represents the adjustment added to capture the marginal impact of the other objectives on the k^{th} objective. This added complexity means that, unlike in the single-objective

[13] r^*_k represents element k of vector $\mathbf{r}$; $\mathbf{r}^*_{-\mathbf{k}}$ represents all elements of vector $\mathbf{r}$ except for element k.

DPO case, we cannot bypass the explicit reward modeling step. However, based upon Theorem 1 from (Rafailov et al. 2024), we can parameterize the reward model implicitly. For all objectives except k, we define $K-1$ reward models, $r_{\phi,i}(x,y)$, each approximating the latent reward function r_i^* as follows:

$$r_{\phi,i}(x,y) = \beta log \frac{\pi_{\phi,i}(y|x)}{\pi_{ref}(y|x)}. \tag{14}$$

Note that this parameterization assumes that we have already obtained $\pi_{\phi,i}(y|x)$ for each objective through the single-objective DPO process illustrated in §6.1.2. We then construct the reward function vector for all objectives except the k^{th} objective, and denote it as $\mathbf{r}_{\phi,-\mathbf{k}}(x,y)$. Next, we substitute $\mathbf{r}^*_{-\mathbf{k}}(x,y)$ with the pre-estimated values $\mathbf{r}_{\phi,-\mathbf{k}}(x,y)$ and formulate the objective for the target policy $\pi_{\theta_\mathbf{w}}$ by training on comparison dataset C_k, analogous to Equation (8):

$$L_{MODPO}(\pi_{\theta_\mathbf{w}}; \mathbf{r}_{\phi,-\mathbf{k}}, \pi_{ref}, C_k) = \tag{15}$$
$$- E_{\pi_\theta}\left[log\sigma \left(\frac{1}{w_k} \left(\beta log \frac{\pi_{\theta_w}(y_w^k|x)}{\pi_{ref}(y_w^k|x)} - \beta log \frac{\pi_{\theta_w}(y_l^k|x)}{\pi_{ref}(y_l^k|x)} - \underbrace{\mathbf{w}^\mathbf{T}_{-\mathbf{k}}(\mathbf{r}_{\phi,-\mathbf{k}}(x,y_w^k) - \mathbf{r}_{\phi,-\mathbf{k}}(x,y_l^k))}_{margin, m_\phi(x,y_w^k,y_l^k)} \right) \right) \right].$$

This formulation ensures that the LM $\pi_{\theta_\mathbf{w}}$ is optimized for a specific weight vector $\mathbf{w}$.[14] Intuitively, L_{MODPO} is similar in spirit to the loss function of the single-objective DPO (shown in Equation (8)), with two main differences: (1) additional weights w_k to capture the weighted objective function, and (2) a margin term to ensure the LM is guided by multiple objectives.

In summary, MODPO consists of three main phases:

- Phase 1: Similar to RLHF and DPO, the process begins with Supervised Fine-Tuning (SFT), where we fine-tune the pre-trained language model on the demonstration dataset to obtain the SFT policy π_{sft}, which serves as the reference policy for subsequent steps.
- Phase 2: Given a set of comparison datasets, denoted as $\{C_1, C_2, ..., C_K\}$, where each dataset $C_k = \left\{(x, y_w^k, y_l^k)^{(i)}\right\}_{i=1}^N$ is sampled from p_k^* for a specific objective k, we proceed with reward modeling. Specifically, for each objective k, we perform reward modeling for all objectives except for k using $\mathbf{C}_{-\mathbf{k}}$, which comprises the comparison datasets for all the remaining objectives other than k. This results in a reward function vector $\mathbf{r}_{\phi,-\mathbf{k}}(x,y)$.
- Phase 3: Building on the reference policy π_{ref} from Phase 1 and the reward models $\mathbf{r}_{\phi,-\mathbf{k}}(x,y)$ from step 2, we can optimize the language model policy $\pi_{\theta_\mathbf{w}}$ based on the loss function shown in Equation (15) for a given weight vector $\mathbf{w}$.

MODPO offers many advantages for our setting. First, from a managerial perspective, it is capable of accommodating diverse and potentially conflicting business objectives. By adopting a linear scalarization strategy (Van Moffaert et al. 2013), we are able to transform a multi-objective problem into a single-objective one by assigning weights to each objective and combining them into a weighted sum. MODPO allows

[14] By selecting different comparison datasets C_k in this loss function, we can learn the corresponding $\mathbf{r}_{\phi,-\mathbf{k}}$, all of which should yield $\pi_{\theta_\mathbf{w}}$ for a given weight vector $\mathbf{w}$. The proof for this can be found in Appendix A2 of Zhou et al. (2024).

firms and managers to specify how much they care about different objectives and develop LLM policies precisely aligned with those weights. Second, MODPO is both practically stable and computationally efficient, making it feasible for real-world implementation. Similar to DPO, it does not rely on reinforcement learning algorithms for LLM alignment. Instead, it employs a novel parameterization of the reward model, enabling it to learn optimal policies in closed form. Under this approach, the optimization problem collapses to simple classification loss, and eliminates the need for sampling from the language model during fine-tuning and extensive hyperparameter tuning. Finally, MODPO is theoretically guaranteed to produce the exact optimal solution as multi-objective RLHF[15], and has been empirically demonstrated to match or surpass existing methods in safety alignment and long-form question answering (Zhou et al. 2024).[16]

6.3 Implementation

Figure 2 illustrates the four phases of MODPO: model pre-training, supervised fine-tuning, marginal reward modeling, and language modeling. Each phase builds upon the previous one, enabling the model to progressively align with multiple business objectives. We describe each in detail below.

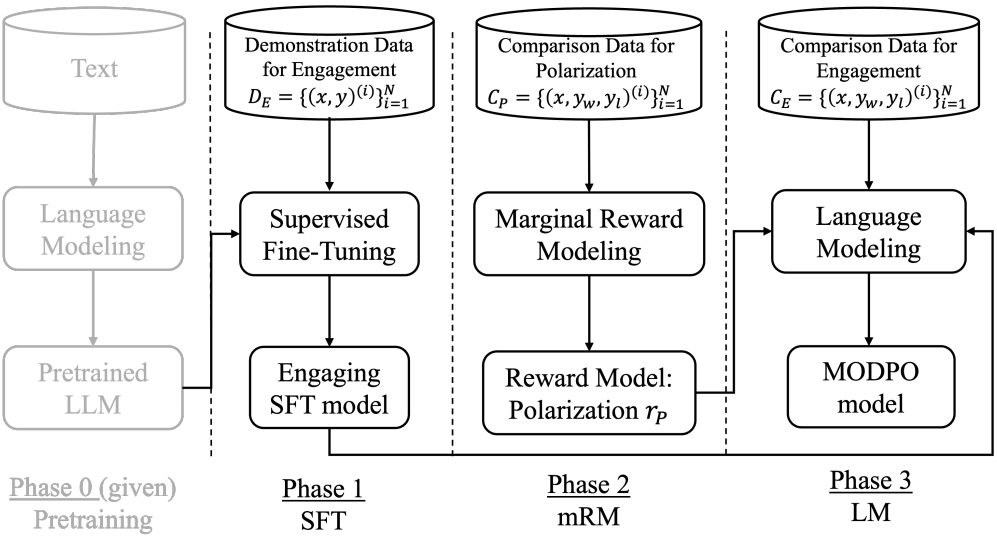


Figure 2: MODPO Overview

6.3.1 Phase 0: Pre-Trained LLM

Phase 0 consists of model pre-training, a foundational stage where the model learns general linguistic patterns, knowledge, and reasoning capabilities from a vast and diverse dataset. In other words, this is the

[15]Multi-objective RLHF (Rame et al. 2024; Williams 2024) extends RLHF to optimize multiple objectives by incorporating human input to guide the learning process. Like RLHF, it involves both reward modeling and reinforcement learning. However, we do not discuss multi-objective RLHF in detail, since MODPO offers greater stability and efficiency than RLHF-style approaches.

[16]Additionally, since this approach leverages open-source LLMs and can be hosted directly on the firm's servers, it is privacy compliant. In contrast, naive prompt-based approaches require the firm to provide their data to GPT/Gemini models, which can then be stored and used as training data by OpenAI/Google, a practice that conflicts with many firms' values and data protection standards. For example, *The New York Times* recently sued OpenAI and Microsoft for the unauthorized use of its articles to train their models (Harvard Law Review 2024).

stage where a general-purpose LLM is built. The standard approach in LLM alignment is to take a pre-trained LLM as given, since training LLMs from scratch requires millions of dollars in resources (Zvornicanin 2024). Following this practice, we adopt a state-of-the-art open-source LLM – Meta's Llama-3-8b (Meta 2024) – as our base model, taking its pre-training as given in Phase 0. To ensure that the substantive results are not idiosyncratic to the choice of the base LLM, in §7.4.1, we present robustness checks with alternative pre-trained open-source LLMs, such as Alpaca.

6.3.2 Phase 1: Supervised Fine-Tuning (SFT)

In Phase 1, we adapt the pre-trained model to perform specific tasks or operate effectively within targeted domains using demonstration data. In our setting, this task involves generating a new version of an article from a base version.

Demonstration Data: To fine-tune the pre-trained model, we use a demonstration dataset, denoted as $D_E = \left\{(x, y)^{(i)}\right\}_{i=1}^{N}$, where the original content serves as the input (x), and the more engaging response, described in §4, serves as the preferred output (y). Essentially, this gives us a new language model that generates more engaging versions of the original articles. Firms can also use alternative demonstration datasets for supervised fine-tuning, depending on their strategic goals.[17] For instance, if a firm's primary objective is to limit polarization/bias, it could instead construct a demonstration data D_P where the ideal output (y) consists of a less polarized version of the original article; we consider this extension in §7.2.2.

Supervised Fine-Tuning (SFT): Building on Llama-3-8b, we use the demonstration data D_E to fine-tune and train a model optimized for generating engaging content, referred to as the *engaging SFT model*. To fine-tune the model efficiently, we employ Low-Rank Adaptation (LoRA), a widely used Parameter-Efficient Fine-Tuning (PEFT) technique (Dettmers et al. 2024; Ye et al. 2024; Angelopoulos et al. 2024), originally developed by Hu et al. (2021). Please see Ye et al. (2024) for a recent application of LoRA in the marketing literature. LoRA enables practical fine-tuning of large-scale LLMs on medium-scale hardware under the assumption that updates during model adaptation exhibit an intrinsic low-rank property. Specifically, instead of modifying all model parameters, LoRA introduces a set of low-rank trainable matrices into each layer of the Transformer model, significantly reducing both memory usage and computational time. This significantly reduces the number of trainable parameters. Please see Web Appendix §D for additional details on the LoRA fine-tuning technique.

6.3.3 Phase 2: Marginal Reward Modeling (mRM)

Comparison Data for Polarization Objective, C_P: For reward modeling in Phase 2, we leverage comparison data to capture the agent's preferences for polarization. Formally, we define this dataset as $C_P = \left\{(x, y_w, y_l)^{(i)}\right\}_{i=1}^{N}$, where each observation consists of an original content (x) and two modified versions (y_w and y_l) evaluated by the agent in terms of polarization. Recall that our goal is to avoid excessive left-slant/left-polarization and ensure that the final output remains close to the original version of the article, which aligns with *The New York Times*'s current editorial stance.[18] As such, in our setting, y_w corresponds to

[17] Note that because of the KL-divergence term added in Equation (9), the final policy produced by the MODPO model will not deviate too far from the baseline/reference policy generated by in this phase. As such, it is critical to pick a demonstration dataset that is aligned with the main objectives of interest.

[18] In our main analysis, we assume that the original article already reflects the preferred editorial stance of the news outlet. Our goal is to enhance engagement while maintaining the polarization level of the original article. In §8.1, we relax this assumption and

the less polarizing version, while y_l represents the more polarizing version; both versions are obtained from the two versions of the articles generated in §4.

Marginal Reward Modeling: In our setting, we need to account for the firm's preference for both engagement ($k = 1$) and polarization ($k = 2$). To learn the optimal policy $\pi_{\theta_\mathbf{w}}$ based on Equation (15), we need to optimize on one objective k and model rewards for the remaining $K - 1$ objectives. In theory, we can choose any of the k objectives to optimize Equation (15), and all of them should yield the same optimal language model for a given weight $\mathbf{w}$, as discussed earlier. In practice, we choose polarization ($k = 2$) as the objective for which we learn the reward function, while engagement serves as the objective on which we will learn the language policy based on Equation (15). To achieve this, we first learn the reward model for polarization, denoted as $r_{\phi,P}$. This process begins with a single-objective DPO procedure, as outlined in Equation (8), to derive the LM policy $\pi_{\phi,P}$ based on the comparison data C_P. We then parametrize the reward model as $r_{\phi,P} = \beta log \frac{\pi_{\phi,P}(y|x)}{\pi_{ref}(y|x)}$, following the approximation in Equation (14).

6.3.4 Phase 3: Language Modeling

Comparison Data for Engagement Objective, C_E: For language modeling in Phase 3, we leverage comparison data to capture the agent's preferences for engagement (C_E). Specifically, we denote this dataset as $C_E = \left\{(x, y_w, y_l)^{(i)}\right\}_{i=1}^{N}$, where each observation consists of an original content piece (x) and two modified versions (y_w and y_l) evaluated by the agent based upon engagement. Since our objective is to enhance engagement, y_w represents the more engaging version, while y_l corresponds to the less engaging one; as before both versions are obtained from the two versions of the articles generated in §4.

Language Modeling (LM): Here, we incorporate the reward model obtained from Phase 2 as an adjustment term into the objective function in Equation (15) to directly optimize the language model $\pi_{\theta_\mathbf{w}}$. Essentially, what happens under the hood during language modeling is that we use an intermediate, less polarizing language model $\pi_{\phi,P}$ to safeguard the training of the multi-objective language model's $\pi_{\theta_\mathbf{w}}$ on the engaging preference dataset C_E. Similar to Phase 1, the language modeling process uses the LoRA procedure for the actual optimization.

In summary, we obtain a Pareto-optimal language model $\pi_{\theta_\mathbf{w}}$ by leveraging the capability from pre-trained models in Phase 0, supervised fine-tuning in Phase 1, marginal reward modeling in Phase 2, and language modeling in Phase 3. This approach balances the dual objectives of enhancing engagement and controlling polarization, which allows firms to tailor content that aligns with diverse business goals.

6.3.5 Other Implementation Details

Lastly, we discuss a few remaining points related to our implementation.

Article Generation: We note that all modified versions of articles in the demonstration datasets used in Phase 1, as well as the comparison datasets used in Phase 2 and Phase 3, are generated with the help of ChatGPT, as described in §4. This approach aligns with the growing trend of using synthetic data generated by LLMs for model training, given their remarkable zero-shot learning proficiency and low cost in data generation (Liu et al. 2024a; Ubani et al. 2023). Notably, the GPT-3 technical report (Brown et al. 2020) demonstrates that GPT-3 can generate news articles that human evaluators often find indistinguishable from those written

explore cases where the media outlet adjusts an article's polarization/slant to a specific preferred level.

by journalists. With recent advancements in LLMs, these capabilities have only improved, making synthetic data an increasingly ubiquitous and, in many cases, preferred alternative to purely human-generated data. For example, the training process of the groundbreaking "DeepSeek-R1" model used a substantial amount of synthetic data, including 600,000 reasoning examples and 200,000 non-reasoning instances, generated using internal models and the "DeepSeek-V3" model (DeepSeekAI 2025).

However, our approach is quite general and does not inherently require AI-generated content. It can be easily adapted to settings where all the demonstration and comparison data consist of human-written articles. In §8, we present an application of the MODPO approach to news articles published by various media outlets, which serve as our demonstration and comparison datasets. These articles are sourced from *AllSides* (Allsides 2012a), a news aggregator that presents multiple perspectives by curating news coverage from outlets across the political spectrum (left, center, and right) alongside its own factual summaries of the same news events.

Preference Generation: In our empirical context, the preference ratings for which articles are more engaging and which ones are more polarizing are established using ChatGPT, as described in §4 and Web Appendix §B. There has been ongoing discussion about the pros and cons of using humans vs. AI agents/-models to generate feedback used to align language models. While the early stream of work primarily relied on human feedback, collecting human feedback is often time-consuming and costly, making it hard to scale. It is also subject to inconsistencies due to individual differences and can inadvertently introduce human biases into the training process. Therefore, recent studies (Lee et al. 2024; Bai et al. 2022b) have increasingly relied on AI-generated feedback as a viable alternative to human input. Notably, AI-generated feedback has demonstrated performance comparable to human input in tasks like summarization (Lee et al. 2024), offering advantages such as rapid scalability, greater consistency, and facilitation of large-scale training. Indeed, even the original DPO paper (Rafailov et al. 2024) employs AI-generated preference data for tasks such as controlled sentiment generation and single-turn dialogue. Naturally, if AI preferences are biased/misleading, then the LM obtained from our alignment procedure will perform poorly in the eventual content generation. As such, it is good practice to ensure that the content produced after LLM alignment using MODPO (or any other alignment procedure) is consistent with human evaluations on the task of interest. To that end, we validate our final results using human ratings (in addition to evaluation results from an independent LLM); see Web Appendix §E.3 for details.

Training, Validation, and Test Data: We split our demonstration and comparison data into training (80%), validation (10%), and test sets (10%). The test set serves a corpus to conduct evaluations using independent LLMs (Gemini and Claude3) and human raters.[19]

Weight Selection: For our main results in §7.1, we set equal weights (i.e., $w = [0.5, 0.5]$) for engagement and polarization. We examine the impact of modifying the weight vectors on the language model policy and content generated in §7.2.1.

[19]In our implementation, we exclude articles with identical polarization ratings, resulting in a comparison dataset of 11,814 articles. To determine the number of training epochs for all three phases, we use 10% of the dataset for validation. Within the test set, which initially consists of 1,181 articles (10% of the total), a small fraction fails to generate due to the SFT step and certain idiosyncrasies of the Llama model. As a result, the final test set contains 983 news articles. When using an alternative base LLM—Alpaca, as discussed in §7.4.1—the final test set includes 1,028 news articles.

Table 5: Gemini Evaluation Results (with *engaging SFT* as the reference policy)

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$
Mean	2.13^{***}	2.57^{***}	-0.44^{***}	0.80^{***}	1.75^{***}	-0.95^{***}
Std	1.15	1.58	1.28	1.66	2.08	1.37
Median	2.00	3.00	-1.00	1.00	2.00	-1.00
N	983.00	983.00	983.00	983.00	983.00	983.00

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

7 Results

We present the main results from our analysis in §7.1. In §7.2, we document the model's sensitivity to the weight vector and data used for supervised finetuning. In §7.3, we provide some insights into how the *multi-objective DPO model* achieves multiple objectives by examining how it leverages different types content strategies. Finally, in §7.4, we present a series of robustness checks.

7.1 Evaluation Results for MODPO

We now examine how the MODPO approach performs on our test data on the two objectives of interest – engagement and polarization. To that end, for each article in the test data, we generate a new version of the article based on the MODPO approach outlined above. To aid comparison, we also consider two benchmark versions of each article:

- *Original version* – This is the original version of the article, as published originally.
- *Engaging DPO version* – This is the version of the article generated using a single-objective DPO approach that only optimizes engagement without controlling polarization.[20]

We use "Gemini-1.5-Flash" to evaluate the engagement and polarization of all three versions of each article in the test data. Further, to ensure consistent evaluations, we provide all three versions of an article within the same prompt and ask Gemini to rate each one in direct comparison to one another. This ensures better consistency of evaluations across the three versions, capturing the relative differences more effectively.[21] Please refer to Web Appendix §E.1 for details on the exact prompts used.

Table 5 shows how the engagement and polarization scores of the MODPO and engaging DPO versions change compared to the original version, as well as to each other. We find that the articles recreated using both *engaging DPO* and *multi-objective DPO* are significantly more engaging than the original articles. ($\Delta_{\text{EngDPO_ORIG}} = 2.57, p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 2.13, p < 0.001$). Interestingly, articles generated by the *multi-objective DPO model* exhibit only slightly lower engagement than those generated by the engaging DPO, which suggests that moving from a single-objective setting (of solely maximizing engagement) to a multi-objective setting does not severely harm performance on engagement, which tends to be the primary metric of interest for firms. Further, the *multi-objective DPO model* is more effective in limiting excessive polarization

[20] In terms of implementation, Figure 2 simplifies to a single objective ($K = 1$) without requiring the reward modeling phase. Additionally, the weight parameter is set to $w = 1$, and the loss function for language modeling simplifies to Equation (8).

[21] We report the results in terms of relative differences in engagement and polarization scores, as our primary interest lies in understanding how the MODPO and engaging DPO versions compare to the original version and each other. The raw scores assigned by "Gemini-1.5-Flash" to the original versions are available in Web Appendix §F.

(unlike the *engaging DPO model*, which exhibits the highest level of polarization) and adheres more closely to the editorial stance in the original version ($\Delta_{\text{MODPO_EngDPO}} = -0.95$, $p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 0.80$, $p < 0.001$). Nevertheless, it is not able to fully avoid any increase in polarization, which is understandable given the desire to boost engagement. In sum, our results demonstrate that the *multi-objective DPO model* effectively balances engagement while not deviating too far from the preferred editorial stance of the firm. As such, it allows firms and content creators to mindfully leverage open-source LLMs in combination with multi-objective alignment to successfully generate content that is engaging for users while minimizing unintended negative consequences such as excessive polarization.

Finally, we make note of a few points regarding the evaluation process. As noted earlier, we use the "Gemini-1.5-Flash" model to score all articles on engagement and polarization for the results shown in the main text. The choice is guided by two reasons. First, "Gemini-1.5-Flash" has demonstrated superior performance in various benchmarks such as enhanced Massive Multitask Language Understanding (MMLU-Pro), challenging math problems (MATH), language translation (WMT23), and multimodal college-level understanding and reasoning tasks (MMMU) (Google 2024). Second, we intentionally avoid using GPT for evaluation because both the demonstration data and comparison data were generated by GPT. Prior research suggests that LLMs can exhibit self-preference, meaning they may favor their own output over those of other models or humans, even when human evaluators rate these as having equally high quality (Panickssery et al. 2024; Xu et al. 2024). If GPT exhibits any inherent tendencies that lead it to generate or rate certain types of content as more engaging or polarizing, using a different LLM for evaluation ensures that our results remain independent of the data generation process for model training.

To further validate our results and ensure that they are not specific to Gemini, we also consider three alternative evaluation approaches to score engagement and polarization for the test data. First, we use an alternative LLM, "claude-3-5-sonnet-20240620," a top-ranked and widely used model known for its strong performance on major NLP benchmarks. The results from this evaluation are reported in Web Appendix §E.2. Second, we also conduct a small-scale Qualtrics experiment to gather human evaluations on a subset of articles; see Web Appendix §E.4. Finally, we use a human rater to score all the articles in the full test dataset; details are provided in Web Appendix §E.3. Our key findings remain consistent across all the evaluation methods, reinforcing the robustness of our results.

7.2 Impact of Firm's Preferences for Maximizing Engagement vs. Controlling Editorial Stance

We now examine how different baseline preferences for engagement vs. polarization affect the language model's content creation strategy. In the results presented so far, the weight vector for the relative importance of engagement and polarization was set to $\mathbf{w} = \{0.5, 0.5\}$, and the baseline reference policy π_{ref} was set to maximize engagement. In §7.2.1, we examine how changing the relative weights on engagement vs. polarization affects the behavior of the language model. Next, in §7.2.2, we examine how changing the reference policy used in the first phase of MODPO affects the language model.

7.2.1 Effect of Modifying the Weight Vector in *Multi-Objective DPO*

Media firms can have differing priorities when balancing the objective of increasing engagement with the need to adhere to a preferred editorial stance or limit polarization. We now examine how adjusting the weight

vector influences the language model's output. This exercise is valuable for two reasons. First, it allows us to assess the sensitivity of the MODPO model to different input weights, helping us calibrate the extent to which the language model's policy shifts in response to these weight changes. Second, from a managerial perspective, it presents a natural way for firms to align their editorial strategy with the language model employed. For example, a firm that cares significantly about engagement and only marginally about its editorial stance may choose a lower weight on polarization and vice-versa. By iterating over different weight vectors, we can approximate the Pareto front of the optimal language models (LM), denoted as $\{\pi_{\mathbf{w}^\intercal \mathbf{r}^*} | \mathbf{w} \in \Omega\}$. This provides firms with a flexible framework to select solutions that best align with their strategic priorities, enabling a more tailored and adaptive content creation strategy.

We vary the polarization weight over a range of values ($w_p \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$), adjusting the engagement weight accordingly ($1 - w_p \in [0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1]$), and fine-tune nine versions of the MODPO model. Next, for each original article in the test set, we generate nine versions using these different models. We then evaluate each recreated version on both engagement and polarization using "Gemini-1.5-Flash" model. Finally, we regress the polarization and engagement scores of the recreated articles on the corresponding weight w_p used to fine-tune the language model. As shown in column (1) and column (3) of Table 6, we find that the coefficient for the polarization weight (w_p) is negative and significant for both the polarization ($b = -0.754$, $p < 0.001$) and engagement ($b = -1.428$, $p < 0.001$) scores. This indicates that increasing the polarization weight in the MODPO model reduces the polarization score, but also reduces the engagement score of the generated articles (consistent with our earlier findings reported in §4). To ensure that our findings are not driven by baseline differences in the original articles, we control for the polarization and engagement scores of the original versions. The coefficient for the original article's polarization score is positive and significant ($b_{\text{Polarization}_{\text{ORIG}}} = 0.260$, $p < 0.001$), suggesting that articles with higher initial polarization levels tend to become more polarizing, even after regenerating the article using the MODPO process to balance engagement and polarization.

A potential concern is that polarization may be easier to control for some articles than others; for example, it might be easier to avoid polarization in articles on architecture than in those on politics. To examine whether these results hold across different types of content, we control for article fixed-effects in column (2) and column (4) of Table 6. The results remain largely unchanged. In other words, we find that our approach effectively controls polarization as a function of the weight vector across all types of articles.

7.2.2 Changing the Reference Policy to Focus on Limiting Polarization

Recall that the first step of MODPO consists of supervised fine-tuning to generate a reference policy that focuses on a single objective, denoted as π_{ref}. In our main analysis, we use engagement as the objective on which this reference policy is fine-tuned. Intuitively, the reference policy π_{ref} reflects the firm's primary objective, while additional objectives are incorporated in later phases (Phase 2 and Phase 3). Furthermore, due to the KL divergence term in Equation (15), the final policy generated by MODPO cannot move too far from the reference policy. As such, the choice of the reference policy plays a crucial role in the final policy.

We now examine how changing the reference policy from one that enhances engagement to one that minimizes polarization affects the language model. To implement this, we change the demonstration data from D_E to D_P, training a *less polarizing SFT model* as the new reference policy π_{ref}. In the demonstration data

Table 6: Polarization and Engagement Level of MODPO Versions

	(1) $\text{Polarization}_{\text{MODPO}}$	(2) $\text{Polarization}_{\text{MODPO}}$	(3) $\text{Engagement}_{\text{MODPO}}$	(4) $\text{Engagement}_{\text{MODPO}}$
Weight (w_p)	-0.754***	-0.755***	-1.428***	-1.430***
	(0.0659)	(0.0597)	(0.0717)	(0.0727)
$\text{Polarization}_{\text{ORIG}}$	0.260***			
	(0.00806)			
$\text{Engagement}_{\text{ORIG}}$			0.00750	
			(0.0230)	
Constant	5.166***	6.754***	7.945***	8.003***
	(0.0616)	(0.0336)	(0.182)	(0.0409)
Article FE	No	Yes	No	Yes
N	8623	8623	8623	8623
R^2	0.1156	0.1795	0.0425	0.0559

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(D_P), the original content serves as the input (x), while a less polarized version of the content is used as the ideal output (y). If the firm's sole objective is to reduce polarization, we can further refine this model by using comparison data that capture agent preference for lower polarization levels. Employing DPO on this data yields a *less polarizing DPO model*, which recreates articles optimized specifically to reduce polarization.[22]

If the firm also values improving engagement as a secondary objective (in addition to controlling for polarization), we can use *multi-objective DPO* with the *less polarizing SFT model* as the reference policy. In this case, we first use the comparison data C_E, where the agent ranks different content versions based upon engagement, as described in §6.3.3, to train a reward model for the second objective—enhancing engagement in Phase 2. This reward model is then incorporated as an adjustment term in the objective function of the *multi-objective DPO model* in Phase 3, which optimize Equation (15) using C_P. Finally, to evaluate the outcomes, we score the engagement and polarization levels of all three versions of each article from the test data using Gemini as before; see Web Appendix §G for details.

Table 7: Evaluation Results using *less polarized SFT* as the Reference Policy

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{LPolDPO_ORIG}}$	$\Delta_{\text{MODPO_LPolDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{LPolDPO_ORIG}}$	$\Delta_{\text{MODPO_LPolDPO}}$
Mean	-0.05	-3.24***	3.19***	-0.78***	-3.29***	2.51***
Std	2.73	2.15	2.12	2.70	2.65	2.39
Median	1.00	-3.00	3.00	-2.00	-4.00	3.00
N	1154.00	1154.00	1154.00	1154.00	1154.00	1154.00

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The evaluation results are shown in Table 7. Compared to the original articles, articles generated using

[22] In this case, since there is only a single objective, Figure 2 simplifies to represent a single objective k without the need for reward modeling steps for multiple objectives.

the *less polarizing DPO* and *multi-objective DPO* are significantly less polarizing ($\Delta_{\text{MODPO_ORIG}} = -0.78$, $p < 0.001$; $\Delta_{\text{LPolDPO_ORIG}} = -3.29$, $p < 0.001$). That is, both these models make the original content even more neutral than the original version and reduce the left-bias significantly. However, MODPO is able to avoid a significant drop in engagement ($\Delta_{\text{LPolDPO_ORIG}} = -0.05$, $p > 0.1$) while the less polarizing DPO loses significant ground on the engagement objective ($\Delta_{\text{LPolDPO_ORIG}} = -3.24$, $p < 0.001$). This suggests that media firms can successfully prioritize their primary goal to fine-tune the *SFT model*, and the secondary goals can be incorporated as marginal rewards during the fine-tuning process of the *multi-objective DPO*. Further, depending on the exact editorial stance that the firm wants to adopt, it can decide which goal to fine-tune on.

7.3 Uncovering Content Strategies for Enhancing Engagement and Controlling for Polarization

We now provide some insights into how the *multi-objective DPO model* enhances engagement while simultaneously avoiding excessive left-polarization.

7.3.1 Theory-Driven Content Strategies

First, we identify three sets of theory-driven features that can affect polarization and engagement, based on prior literature – (1) content structure and flow, (2) emotional and persuasive content elements, and (3) informational and factual content elements.

Content Structure and Flow. Prior research has shown that three key aspects of content structure can affect engagement and polarization – *strong opening*, *narrative structure*, and *framing issues as questions*. A *strong opening*, also referred to as a hook, is designed to capture the audience's attention and set the tone for the content that follows (Ting and Chin-Chin Siaw 2020; TikTok 2024). While widely recognized for their effectiveness in driving engagement, strong openings may also have implications for polarization. For example, hooks tailored to resonate with specific audiences might inadvertently reinforce pre-existing beliefs, potentially heightening polarization (Smith et al. 2024). Beyond a strong opening, *narrative structure* can also play a critical role in driving engagement. This refers to how the story is organized and presented, including its flow, themes, and techniques used to attract readers. Past research highlights storytelling elements such as the speed of plot progression, the build-up of cognitive tension, and the circuitous nature of a story as effective tools for keeping audiences engaged (Toubia et al. 2021; Laurino Dos Santos and Berger 2022). Another important content feature, particularly in news articles, is *framing issues as questions*, where topics or issues are presented through thought-provoking or guiding questions. However, the impact of this characteristic on engagement remains inconclusive. For example, Scacco and Muddiman (2020) find that news articles with question headlines tend to generate less intention to engage, while Lai and Farbrot (2014) observe that such headlines often spark greater readership interest. Moreover, the way questions are framed is critical. When questions emphasize division or stereotypes, they may reinforce existing biases and deepen polarization, whereas those framed inclusively can encourage diverse perspectives, foster empathy and reduce polarization (Sinnott-Armstrong and Skorburg 2019).

Emotional and Persuasive Techniques. We explore six content features along this dimension – *emotion level*, *dramatic metaphors*, *provocative language*, *definitive language*, *call to action*, and *urgency level* – and their influence on engagement and polarization. *Emotion level* refers to the degree to which the content

(A) Original Version
Engagement Score: 6.0 Polarization Score: 6.0

Capitol Lacks Tests for Returning Senators While White House Tests Many in Trump' s Circle

Dr. Brian P. Monahan, the tight-lipped doctor who attends to Congress, sent up on Thursday what some have construed as a warning: His office, he told senior Republican officials on a private conference call, can not screen all 100 senators for the coronavirus when they return to work on Monday. Two miles down Pennsylvania Avenue at the White House, the story is very different. President Trump and Vice President Mike Pence are tested frequently, aides who come into close contact with them are tested weekly and the list of people who need to be tested daily keeps expanding, according to officials familiar with the process. The stark contrast between the testing haves at the White House and the have-nots on Capitol Hill, first reported in Politico, makes clear that Mr. Trump' s pronouncement that "anybody that wants a test can get a test," as he said on March 6 at the Centers for Disease Control and Prevention in Atlanta, is far from true. Although the rich and powerful are clearly favored, not even all the powerful have equal access. And beyond whether people can be tested, there are questions about the tests available. At the White House, the medical unit is using a rapid-testing kit developed by Abbott, which yields results in about five minutes. But Dr. Monahan told the Republican aides on Thursday that he lacked such equipment, and that it would take at least two days to get test results. The disparity highlights the fundamental tension between the natural impulse to protect a president' s health and the desire by most politicians to project that they are not receiving any special treatment, Robert Dallek, the presidential historian, said in an interview Friday. By flaunting his own access to tests and making false claims about the availability of testing, Mr. Trump, he said, was only hurting his own credibility with voters. "When you add it to the fact that people on Capitol Hill, who after all form an essential part of the government as well, can not get testing as readily, it just underscores the feeling that this man is principally self-serving," Mr. Dallek said.

(B) MODPO Version
Engagement Score: 8.0 Polarization Score: 6.0

Capitol's Testing Dilemma: Senators Left in the Dark While the White House Shines

As the nation grapples with the coronavirus pandemic, a stark divide has emerged in the heart of American politics. Dr. Brian P. Monahan, the chief medical officer for Congress, delivered a sobering message on Thursday: his office can not screen all 100 senators for COVID-19 as they prepare to return to work on Monday. Meanwhile, just two miles away at the White House, President Trump and Vice President Mike Pence are receiving frequent tests, while aides in close contact with them are tested weekly. The contrast is glaring. Trump' s assertion that "anybody that wants a test can get a test" rings hollow in light of this reality. The powerful are not all treated equally, and the lack of access to testing raises critical questions about fairness and safety in our government. The disparity is not just about numbers; it highlights a deeper issue. Dr. Monahan revealed that his office lacks rapid-testing kits, with results taking at least two days. This revelation underscores a fundamental tension: the desire to protect the president' s health versus the need for transparency and accountability among elected officials. "When you add it to the fact that people on Capitol Hill, who after all form an essential part of the government, can not get testing as readily, it just underscores the feeling that this man is principally self-serving," said presidential historian Robert Dallek. The stakes are high. As Congress prepares to resume business, the health and safety of lawmakers are paramount. With many already infected, the urgency for testing is clear. The House of Representatives has secured a rapid-testing kit from Abbott, but the Senate is left scrambling, with Dr. Monahan' s office unable to provide the same level of access. The situation is further complicated by a lack of communication. Dr. Monahan' s office has not been in touch with Senate officials, and Senate Majority Leader Mitch McConnell has not commented on the situation. This silence raises eyebrows, especially given the critical need for clarity and cooperation during a public health crisis.

(C) Engaging DPO Version
Engagement Score: 8.0 Polarization Score: 7.0

Capitol Lacks Tests for Returning Senators While White House Tests Many in Trump' s Circle

In a startling revelation, Dr. Brian P. Monahan, the physician who attends to Congress, sent a chilling message during a private conference call with senior Republican officials: his office can not screen all 100 senators for the coronavirus as they prepare to return to work on Monday. Meanwhile, just two miles down Pennsylvania Avenue, the White House tells a different story. President Trump and Vice President Mike Pence are tested frequently, while aides who come into close contact with them undergo weekly screenings. The list of those needing daily tests is expanding, according to insiders. This stark contrast raises serious questions about the integrity of Trump' s claim that "anybody that wants a test can get a test." The reality is far from that. While the rich and powerful are clearly favored, even among them, not everyone has equal access. The disparity in testing capabilities is alarming. At the White House, a rapid-testing kit developed by Abbott yields results in just five minutes. Yet, Dr. Monahan revealed that he lacks such equipment, with test results taking at least two days to arrive. This glaring gap highlights a fundamental tension: the desire to protect the president' s health versus the need for politicians to project that they are not receiving special treatment. Robert Dallek, a presidential historian, weighed in, stating, "By flaunting his own access to tests and making false claims about their availability, Trump is undermining his credibility with voters." The fact that lawmakers, who form an essential part of our government, struggle to access testing underscores a troubling perception: that Trump is self-serving. As the nation grapples with the pandemic, the White House is in the spotlight, while the Capitol remains in the shadows. The Senate is set to reconvene on Monday, with senators expected to return to their offices. Yet, the reality is that many will not be tested, and those who are will face a lengthy wait for results. In a statement, Dr. Monahan defended his office' s testing protocols, claiming they are "robust and thorough."

Figure 3: Three versions of an example article.

elicits emotional responses, such as joy, anger, fear, or hope. Highly emotional content often sparks increased sharing and commenting, reflecting a higher level of engagement (Berger and Milkman 2012). However, it may also reinforce echo chambers by amplifying preexisting biases and encouraging users to interact primarily with like-minded individuals (Garrett 2009). *Dramatic metaphors* use vivid and figurative language to frame events or issues, making complex situations more relatable (Lakoff and Johnson 2008). Although this can improve engagement, it can also deepen polarization by aligning individuals with groups that share similar identities (Thibodeau et al. 2019). Similarly, both *provocative language* and *definitive language* may influence how engaging the content is (Banerjee and Urminsky 2024). *Provocative language* evokes strong emotional reactions and provokes thoughts, while *definitive language* projects authority and certainty. However, they may also polarize audiences by leaving little room for alternative interpretations, thus discouraging opposing views. *Call to action* – explicit or implicit instructions or suggestions, urging readers to act, can significantly influence engagement by creating urgency or aligning with audience motivations. Yet, calls to action, particularly those encouraging sharing, may amplify the dissemination of content aligned with users' preexisting beliefs, thereby fostering echo chambers and reinforcing polarized opinions (Bakshy et al. 2015). *Urgency level*, which emphasizes immediacy and the need to act quickly, has been found to drive higher engagement (Perez Vega 2016), though its impact on polarization has not been documented before.

Informational and Factual Elements. We consider three content features along these dimensions – *fact level*, *relevant statistics*, *balanced perspective*. *Fact level* refers to specific and verifiable information included in an article, that serve as the foundational building blocks of its narrative. This feature helps distinguish factual claims from opinions, interpretations, or commentary, providing clarity about what is being presented as a fact, and potentially reduce polarization. However, prior research has shown that it is possible to cherry-pick which factual sources are cited to present a slanted view of events (Groseclose and Milyo 2005). Complementing this, *relevant statistics* provide numerical data to support the factuality of the article, enabling readers to assess the magnitude or significance of the presented information. Together, these two features can increase content credibility and trustworthiness, often driving higher engagement (Van Krieken 2020). *Balanced perspective*, is characterized by fair, objective, and impartial presentation of diverse information. By appealing to readers who value objectivity, this element can help mitigate polarization (Gentzkow and Shapiro 2006; Allen et al. 2020). A balanced perspective is particularly effective in reducing confirmation bias that often fuels divisive discourse (Stroud 2011).

7.3.2 Illustrative Example of Content Strategies

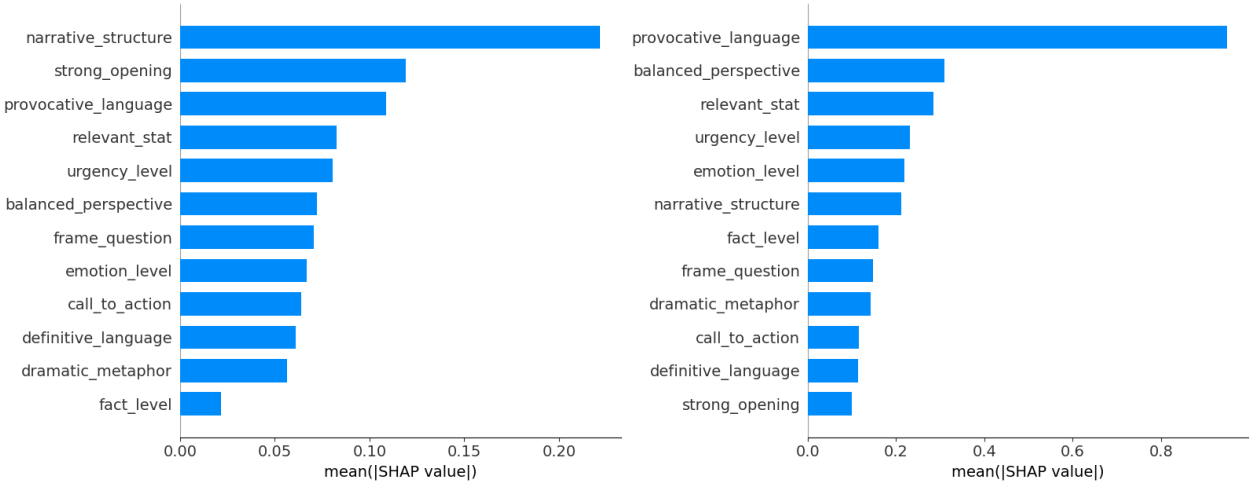
We now highlight how different LMs (i.e., *the multi-objective DPO model* vs. the engaging DPO model) leverage the discussed content strategies using an illustrative example from the test data in Figure 3. Due to space constraints, we only provide one example here and refer readers to Web Appendix §H.3 for additional examples.

In general, all three versions highlight the disparity in access to COVID-19 testing between the White House, where rapid and frequent testing was prioritized, and Capitol Hill, where senators faced limited resources. Both MODPO and engaging DPO versions enhance engagement by having a strong opening, although they employ distinct approaches. The MODPO version begins with a striking contrast in its title, "Senators Left in the Dark While the White House Shines," which immediately grabs attention through vivid

imagery. Meanwhile, the engaging DPO version opens with dramatic phrases like "startling revelation" and "chilling message" to evoke an emotional response. By contrast, the original article starts more descriptively, which, while factual, lacks the same immediate appeal. The narrative structure also varies significantly across the versions. The original article is linear, fact-driven, and dense, offering comprehensive detail but lacking the same immediate appeal. In comparison, the MODPO version is more logically organized, guiding readers through the content with greater clarity and flow. The engaging DPO, on the other hand, prioritizes emotional impact and critiques.

Polarization management differs across the versions as well. The MODPO version employs moderately provocative language, such as "stark divide" and "glaring contrast," which effectively highlight inequities without excessive dramatization. By comparison, the engaging DPO version uses more emotive terms, such as "startling," "chilling," and "self-serving," alongside dramatic metaphors like "The Capitol remains in the shadows." While this approach evokes a strong sense of injustice, it also intensifies polarization. For its part, the original article remains relatively neutral, using mild language to point out disparities, such as, "Although the rich and powerful are clearly favored, not even all the powerful have equal access." Finally, the tone and perspective also set the versions apart. The original article maintains a relatively factual tone, though its implicit critique of inequality and the Trump administration shows partisanship and leans left. The MODPO version also achieves a somewhat balanced perspective by focusing on systemic issues, as reflected in statements like, "The powerful are not all treated equally, and the lack of access to testing raises critical questions about fairness and safety in our government." Conversely, the engaging DPO version is the most polarizing, with direct criticism of Trump. For example, it asserts, "This stark contrast raises serious questions about the integrity of Trump's claim that 'anybody that wants a test can get a test."

7.3.3 Relationship between Content Strategies and Engagement/Polarization in the Original Version



Engagement (Original Version) Polarization (Original Version)

Figure 4: The Importance of Content Strategies on Engagement and Polarization (SHAP value)

While the example discussed gives us some insights into the content strategies of different LMs, we now perform a more systematic evaluation. Specifically, we use Gemini to assess the use of the 12 theory-driven

strategies for each news article in the test set. As before, we provide all three versions of an article within the same prompt, and use "Gemini-1.5-Flash" to compare and rate the content strategies for each version. For details on the exact prompts used, please refer to Web Appendix §H.1. We now examine how much of the variance in the two measures of interest – engagement and polarization – can be explained by the content strategies discussed above. To that end, we specify the following regression:

$$Y_i = \alpha + \sum_{j=1}^{12} \beta_i X_i + \varepsilon_i, \tag{16}$$

where Y_i denotes the polarization score or engagement score of the original article i in the test data and X_i represents the 12 content features summarized in §7.3.1. The regression results are shown in Table A7 in Web Appendix §H.2. We see that the content strategies from §7.3.1 explain a larger proportion of the variance in polarization scores than in engagement scores (61.44% vs. 15.54%, respectively). Further, we find that *strong openings* and *narrative structure* emerge as important drivers of engagement, exhibiting a significant positive relationship with engagement scores. In contrast, *provocative language* and *urgency level* are positively associated with polarization while a *balanced perspective* is negatively associated with polarization.

To gain further insights, we apply SHAP (Lundberg and Lee 2017), which quantifies the importance of each factor in explaining engagement and polarization. As illustrated in Figure 4, *narrative structure* and *strong opening* are the two most influential content features driving engagement, while *provocative language* and *balanced perspective* are the top two features that explain content polarization. These findings align closely with the regression results.

7.3.4 Content Strategies in the *Multi-Objective DPO Model*

Intuitively, for the *multi-objective DPO model* to recreate content that is more engaging without exacerbating polarization, it should leverage content strategies that significantly impact polarization but have minimal impact on engagement. Therefore, we plot the Pearson correlation coefficients between content strategies with each other as well as with polarization and engagement scores. As shown in Figure 5, and consistent with the SHAP analysis results in Figure 4, two content strategies stand out: *provocative language* and *balanced perspective*. These strategies are strongly associated with polarization but have a relatively smaller impact on engagement (provocative language: p_{eng} = 0.24, p_{pol} = 0.68; balanced perspective: p_{eng} = -0.11, p_{pol} = -0.58). Interestingly, Figure 6, which highlights the percentage change in each content strategy generated by the *multi-objective DPO model* relative to the single-objective *engaging DPO model*, reveals significant differences in these two strategies. This observation underscores their critical role in limiting polarized language in the *multi-objective DPO model*. Similarly, we observe substantial changes in *definitive language*, *urgency level*, and *relevant statistics*, as these features also have a greater influence on polarization than on engagement.

Beyond identifying "what needs to change", it is equally important to consider "what can easily be changed." For instance, *dramatic metaphor*, which has a very low baseline in the original content, sees notable adjustments in the *multi-objective DPO model*, likely due to the ease of adjusting this aspect. At the same time, to maintain the content's engagement level, the model should preserve features that primarily drive

engagement – such as *narrative structure* and *strong opening* – at levels similar to those in the *engaging DPO model*. However, some trade-offs arise. Content strategies that influence polarization, such as *provocative language* and *dramatic metaphor*, often positively correlate with those affecting engagement, like *narrative structure* and *strong opening*. As a result, efforts to control polarization by adjusting these features can also negatively affect engagement. This interdependence likely explains the observed changes in *narrative structure* and *strong opening* in Figure 6.

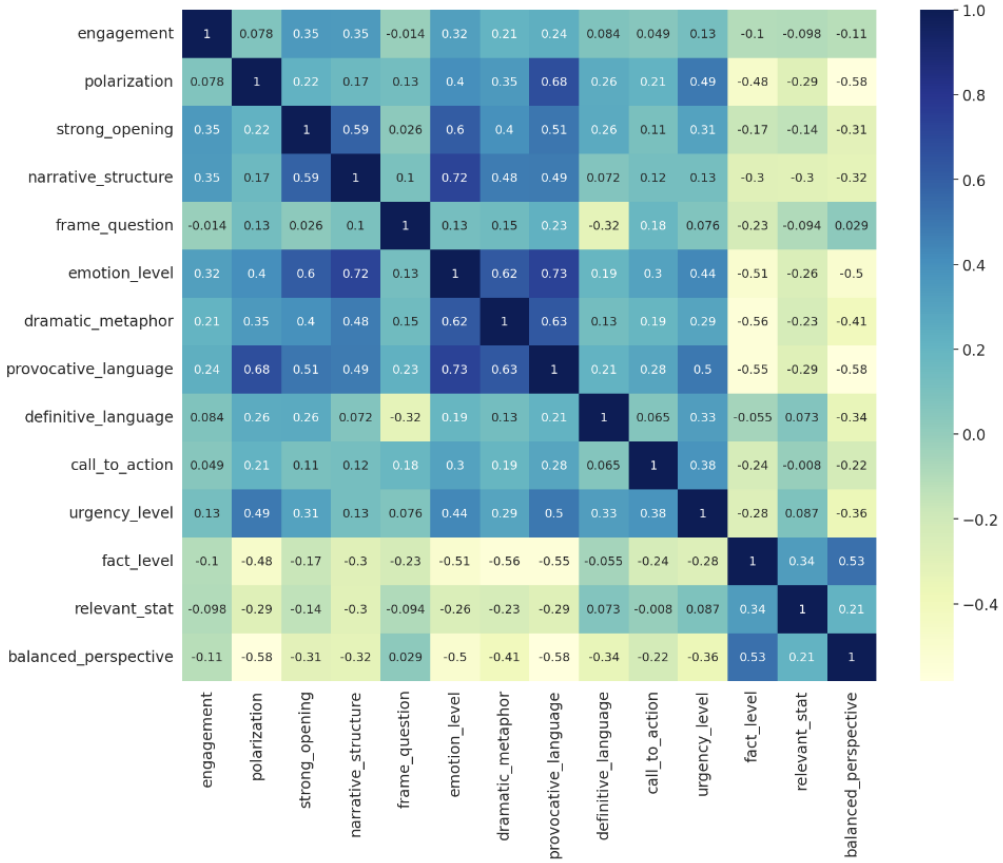


Figure 5: Correlation Coefficient Plot Between Content Strategies

7.4 Robustness Checks

We now present a series of robustness checks and extensions to provide additional support for our results.

7.4.1 Alternative base LLM – Alpaca

In the main analysis, we use Llama 3-8b as the base model on which we perform the multi-objective alignment. However, we can use any other open-source LLM in Phase 0 in Figure 2. To test whether our results are robust to the use of alternative LLMs, we implement the MODPO approach (with the same data) using Alpaca in the first phase illustrated in Figure 2. Alpaca is a fine-tuned model based upon the Llama 2-7B model, trained on 52,000 instruction-following demonstrations. Despite its smaller size, it outperforms the base Llama model in various instruction-following tasks, including summarization, text generation, and question answering (Taori et al. 2023). Additionally, Alpaca is often used as a foundation model for LLM

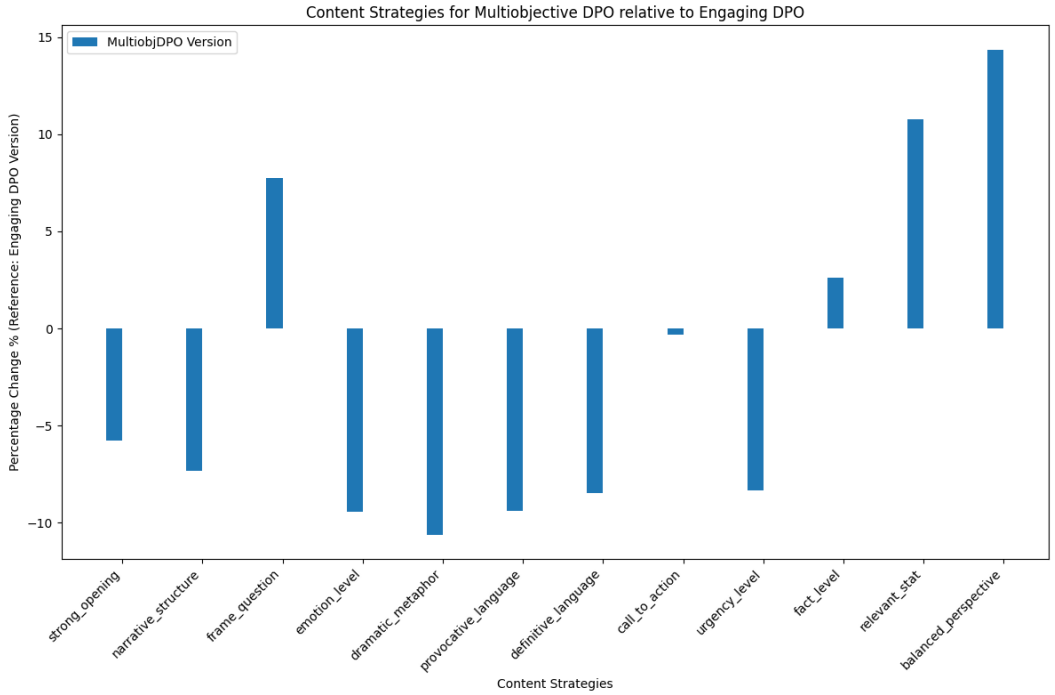


Figure 6: Relative Differences in Content Strategies Between the Multi-Objective and Engaging DPOs

safety alignment, aiming to optimize both helpfulness and harmlessness (Dai et al. 2023).

Building upon Alpaca as our base model, we generate articles using the following three models: the *engaging SFT model*, the *engaging DPO model*, and the *multi-objective DPO model*, as illustrated in §7.1. We then evaluate these articles using another LLM – Gemini – to measure the engagement and polarization levels following the same approach described earlier. We find that the results with Alpaca as the base model are similar to those with Llama 3-8B as the base model. For detailed results, please refer to Web Appendix I.1.

7.4.2 Hallucination

One key concern when using LLMs to rewrite news content is hallucination, which refers to cases where the generated content is nonsensical and/or unfaithful to the source material (Filippova 2020). To see if this is an issue in our context, we examine the extent to which our exercise suffers from factual hallucinations, which refers to the situation where the generated content conflicts with verifiable real-world facts, often manifesting as factual inconsistencies (Ji et al. 2023b).[23] To assess the prevalence of factual hallucinations, we conduct a manual review of all the 983 generated news articles in our evaluation set. We find that approximately 3% of these articles exhibit signs of factual hallucination. Among these, 0.75% contained verifiable errors that directly contradicted reality, while 2.25% contain potentially fabricated information that could not be corroborated through publicly available sources. These fabrications included a range of elements, including fictitious characters, fabricated scientific studies, invented details, and unverifiable data. For detailed examples

[23]Factual hallucinations differ from faithfulness hallucinations, which refer to instances where the generated content diverges from user input or contradicts itself within the generated response. For example, if a user prompts the model to translate the English question "What is the capital of France?" into Spanish, and the model instead responds with "The capital of France is Paris," it has failed to follow the task correctly, providing an answer to the question rather than a translation.

illustrating the breakdown of hallucination types, please refer to Web Appendix §I.2. In sum, while we see evidence of hallucination in a few rare instances, it does not seem to be a prevalent problem in our setting.

7.4.3 Validation of Results Using a Real-World Dataset

In addition to generating synthetic output from the GPT model to build training data, we collect real-world news articles from AllSides (Allsides 2012a), a news aggregator that presents multiple perspectives by featuring coverage from outlets with varying political stances (i.e., left, center, and right) alongside its own factual summaries of the same news events. More details about AllSides data collection and an example news event from its website are provided in §8.1 and Web Appendix §I.3.

Using this real-world dataset from AllSides, we generated articles with two language models: the *engaging DPO model*, and the *multi-objective DPO model*, as described in §7.1. We then evaluate these articles using "Gemini-1.5-Flash", to measure the engagement and polarization levels, following the same approach described in Web Appendix §E.1. We find that the results using the real-world Allsides data yield similar results to those trained on synthetic data. For detailed evaluation results, please refer to Web Appendix §I.4. Furthermore, when we apply the model trained on AllSides data to generate articles for *The New York Times*, we find that it can also produce more engaging and less polarizing articles, highlighting the model's generalizability across different datasets. Please note that the generated articles tended to be shorter, likely due to the relatively brief nature of AllSides' leading paragraphs compared to the more extensive coverage from *The New York Times*.

8 Extensions

So far, we have considered a setting where a firm seeks to craft content that is engaging without deviating too far from its current level of polarization. However, our framework is general and can be adapted to other objectives and settings. In §8.1, we present an extension where the media firm seeks to generate content to achieve a specific level of left- or right-leaning slant. Next, in §8.2, we outline general algorithms for two additional natural extensions: one where the firm seeks to personalize news content to different user segments and another that aligns a language model for multiple modalities such as images, in addition to text.

8.1 A Specific Level of Slant: the Case of Targeting Left-Leaning Audiences

Different media organizations often have distinct editorial goals regarding the slant of their news content. For instance, outlets like *CBS News* and *CNN* typically lean left, whereas *Fox News* and *Newsmax* tend to lean right (Flaxman et al. 2016). In our main analysis, we assume that the original news article already reflects the preferred editorial stance of the news outlet, however, our framework can easily extend to settings where the firm seeks to recreate content to achieve a desired level of left- or right-leaning polarization to better reflect its editorial stance. To show how this can be achieved, we still use the real-world news articles from the news aggregator AllSides illustrated in Web Appendix §I.3. Specifically, we collect data on 9,982 news events from *AllSides* headline roundups (Allsides 2012b), covering the period from June 1, 2012, to October 13, 2024. For each news event, we obtain the leading paragraph coverage from left-, center-, and right-leaning media outlets.[24] To ensure completeness of the final sample, we retain 5,656 events where coverage from all three perspectives is available.

[24] As outlined in Web Appendix §I.3, AllSides (Allsides 2012a) provides media bias ratings for news outlets, which we validate using polarization scores generated by ChatGPT.

Consider a firm that has some neutral content that it seeks to align with its moderate left-leaning ideological stance, i.e., make the content left-leaning, but incorporate some right-leaning language.[25] The news articles from *AllSides* serve as a real-world benchmark for constructing demonstration and comparison datasets across different modeling phases. To train a model that reflects this firm's editorial preferences, we first construct a demonstration dataset, denoted as $D_L = \left\{(x, y)^{(i)}\right\}_{i=1}^{N}$. Here, the content from the center outlet serves as the input (x), and the corresponding coverage from a left-leaning outlet is used as the ideal output (y). If the firm seeks to reinforce its left-leaning stance, we can construct a comparison dataset that captures its preference for left-leaning content ($C_L = \left\{(x, y_w, y_l)^{(i)}\right\}_{i=1}^{N}$), where each observation consists of an article from the center outlet (x), a corresponding left-leaning piece (y_w), and a corresponding right-leaning content piece (y_l). Alternatively, we also consider another comparison dataset to capture firms' preference for right-leaning content ($C_R = \left\{(x, y_w, y_l)^{(i)}\right\}_{i=1}^{N}$), where each observation consists of an article from the center outlet (x), a corresponding right-leaning piece (y_w), and a corresponding left-leaning piece (y_l).

Next, using these different datasets, we construct two language models and generate three versions of each article in the test data:

- *Center version:* The original version of the article from the center outlet.
- *Left-leaning DPO version:* This version is generated using a single-objective DPO approach that only optimizes for left-leaning perspectives without incorporating right-leaning viewpoints. Specifically, we first fine-tune a SFT model using D_L, and then train a left-leaning DPO model using C_L.
- *MODPO version:* This version is generated using the MODPO approach, which optimizes left-leaning perspectives while also accounting for right-leaning perspectives. Specifically, we first fine-tune a SFT model using D_L, then use C_R for the marginal reward modeling step (as illustrated in Figure 2), and finally use C_L for the language modeling step. We set the weight vector for the relative importance of left- and right-leaning preferences to $\mathbf{w} = \{0.5, 0.5\}$.

Similar to our main analysis, we use "Gemini-1.5-Flash" to evaluate the left-leaning polarization level and engagement of all three versions of each article in the test data.[26]

Table 8: Gemini Evaluation Results – Difference (with *left-leaning SFT* as the reference policy)

	Difference in Left-leaning Polarization Scores			Difference in Engagement Scores		
	$\Delta_{\text{MODPO_CENTER}}$	$\Delta_{\text{LeftDPO_CENTER}}$	$\Delta_{\text{MODPO_LeftDPO}}$	$\Delta_{\text{MODPO_CENTER}}$	$\Delta_{\text{LeftDPO_CENTER}}$	$\Delta_{\text{MODPO_LeftDPO}}$
Mean	0.16*	1.29***	-1.12***	0.62***	1.50***	-0.88***
Std	1.28	1.86	2.00	1.60	2.64	2.73
Median	0.00	1.00	-1.00	1.00	2.00	-2.00
N	534.00	534.00	534.00	534.00	534.00	534.00

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The left panel of Table 8 shows how the polarization of different models compare to each other. Note that the left-leaning polarization score is a cumulative score that captures the overall slant of the article. Our findings suggest that both the *left-leaning DPO* and *multi-objective DPO* models generate articles

[25] In practice, media firms may be uncertain about the polarization/slant level of their original content. They can first assess its ideological slant and then adjust it to align with their preferred editorial stance.

[26] As before, to ensure consistency in evaluations, we provide all three versions of an article within the same prompt and ask Gemini to rate each one in direct comparison to the others.

that are significantly more left-leaning than the original articles ($\Delta_{\text{LeftDPO_CENTER}} = 1.29$, $p < 0.001$; $\Delta_{\text{MODPO_CENTER}} = 0.16$, $p < 0.05$). However, the *multi-objective DPO model*, which incorporates right-leaning perspectives, exhibits a lower left-leaning polarization level compared to the *left-leaning DPO model* ($\Delta_{\text{MODPO_LeftDPO}} = -1.12$, $p < 0.001$). Notably, by adjusting the relative weights on left- and right-leaning preferences, we can achieve a specific level of left-leaning polarization.

Interestingly, while our fine-tuning process only targets different types of polarization (left and right), we also observe shifts in engagement scores (see right panel of Table 8). The left-leaning DPO version, which exhibits the highest level of left-leaning polarization, also has the highest engagement rating compared to the original content from the center outlet. In contrast, the MODPO version, which is moderately left-leaning, shows lower engagement compared to the center version. This pattern re-affirms our earlier finding (from §4) that engagement and polarization tend to move together, and modifying one of these without directly controlling for the other can inadvertently shift the other metric as well.

8.2 Other Extensions

News Personalization: So far, we have not considered user-level characteristics or how to personalize content based on users' behavioral features (primarily because we do not have user-specific information in either the *The New York Times* or the *AllSides* datasets). However, if firms have access to such data, personalization can be achieved.[27] Indeed, early experiments with continuous personalization using in-context learning in ChatGPT suggest that LLMs can be used effectively for personalization (Matz et al. 2024).

Our approach can easily be adapted to continuous multi-objective content personalization tailored to segments of consumers or individual user preferences. To implement a multi-objective content personalization algorithm, we would combine a learnable user model with a base LM to define new learning objectives aimed at jointly minimizing both user-agnostic and user-specific preference losses. A natural way to do this is to use different weight vectors for different segments of consumers in the MODPO optimization and fine-tune different language models for each segment.

Multi-modalities – Images: Our current analysis focuses only on text. A natural extension would be to incorporate images, which, like text, can serve multiple objectives and have broader applications in ad copy and social media posts. Notably, visual content is not neutral. As Mosaffa et al. (2025) demonstrate, images accompanying news articles often exhibit significant polarization, with notable variations across media outlets and political figures. This suggests that media firms may want to align not just text, but also visual content to address multiple objectives. If the firm has access to visual data that captures both editorial stance and user preferences, it is feasible to perform multimodal MODPO alignment using LLMs such as Llama 3.2 (Meta 2024) that support both text and vision.

9 Conclusion

In conclusion, this paper studies the relationship between content engagement and polarization and explores how media firms can leverage and adapt LLMs to create content that balances engagement and desired editorial slant. Using a large dataset from NYTimes, we first show that more engaging articles also tend

[27] Prior research has shown that behavioral targeting can enhance the match between users and the content they engage with across a variety of settings; see Rafieian and Yoganarasimhan (2023).

to be more biased/polarizing. Further, we show that using naive prompt-based or single objective alignment approaches (e.g., DPO) to improve engagement can inadvertently heighten the level of bias/polarization.

We present a constructive solution to this problem. We adopt the Multi-Objective Direct Preference Optimization (MODPO) algorithm, a newly proposed language alignment approach in the LLM safety literature, which combines Direct Preference Optimization (DPO) with multi-objective optimization techniques for aligning LLM objectives. We adopt this approach to our setting and data and show that the content created by the newly aligned language model effectively balances the trade-off between enhancing engagement and adhering to the preferred editorial policy of the firm. We further show that the approach is flexible, and the language model policy is responsive to the weight vector specified by the firm and the demonstration data used for supervised fine-tuning. We also provide some insights into how the *multi-objective DPO model* leverages different content strategies to balance between the two objectives. Lastly, using an alternative data source from *Allsides*, we show that the model can be easily applied to other conflicting objectives as well as other data sources.

More broadly, our general framework can be adapted to a variety of other business applications where LLM-based approaches are increasingly used for content creation; e.g., advertising agencies now use LLMs to craft ad copies (Forbes 2023), e-commerce platforms use them to automate product descriptions (Marvik 2023), and content-creators in social media utilize LLMs to generate posts (Lyu et al. 2024). In all these settings, content creators and firms often seek to balance multiple objectives simultaneously. For example, Jasper, a leading artificial intelligence (AI) marketing content platform, strives to ensure its generated content remains true to a brand's voice while also effectively engaging targeted audiences (Jasper 2024). Social media creators seek to produce posts that promote brands they are collaborating with while maintaining authenticity with followers (Forbes 2023). E-commerce companies seek to craft product descriptions that not only align with brand values but also convey accurate product information. We expect our approach to be applicable and effective across all these marketing applications.

Funding and Competing Interests Declaration

Author(s) have no competing interests to declare.

References

J. Allen, B. Howland, M. Mobius, D. Rothschild, and D. J. Watts. Evaluating the fake news problem at the scale of the information ecosystem. *Science advances*, 6(14):eaay3539, 2020.

Allsides. Allsides website, 2012a. Link. Accessed on November, 2024.

Allsides. Allsides headline roundup website, 2012b. Link. Accessed on November, 2024.

P. A. Andersen and T. R. Blackburn. An experimental study of language intensity and response rate in e mail surveys. *Communication Reports*, 2004.

P. Angelopoulos, K. Lee, and S. Misra. Causal alignment: Augmenting language models with a/b tests. *Available at SSRN 4781850*, April 2024.

Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022a.

Y. Bai, S. Kadavath, S. Kundu, A. Askell, J. Kernion, A. Jones, A. Chen, A. Goldie, A. Mirhoseini, C. McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022b.

E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

A. Banerjee and O. Urminsky. The language that drives engagement: A systematic large-scale analysis of headline experiments. *Marketing Science*, 2024.

D. P. Baron. Persistent media bias. *Journal of Public Economics*, 90(1-2):1–36, 2006.

R. M. Batista and J. Ross. Words that work: Using language to generate hypotheses. *Available at SSRN 4926398*, 2024.

M. Ben-Akiva, D. McFadden, M. Abe, U. Böckenholt, D. Bolduc, D. Gopinath, T. Morikawa, V. Ramaswamy, V. Rao, D. Revelt, et al. Modeling methods for discrete choice analysis. *Marketing Letters*, 8:273–286, 1997.

J. Berger and K. L. Milkman. What makes online content viral? *Journal of marketing research*, 49(2):192–205, 2012.

J. Berger and G. Packard. Are atypical things more popular? *Psychological science*, 29(7):1178–1184, 2018.

J. Berger, Y. D. Kim, and R. Meyer. What makes content engaging? how emotional dynamics shape success. *Journal of Consumer Research*, 48(2):235–250, 2021.

J. Berger, W. W. Moe, and D. A. Schweidel. What holds attention? linguistic drivers of engagement. *Journal of Marketing*, 87(5):793–809, 2023.

J. Bezos. Jeff bezos' note to the washington post team, Feburary 2025. Link. Accessed on March, 2025.

R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

J. Brand, A. Israeli, and D. Ngwe. Using gpt for market research. *Harvard Business School Working Paper*, 23(062), 2023.

T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901, 2020.

T. Chu, Y. Zhai, J. Yang, S. Tong, S. Xie, D. Schuurmans, Q. V. Le, S. Levine, and Y. Ma. Sft memorizes, rl generalizes: A comparative study of foundation model post-training. *arXiv preprint arXiv:2501.17161*, 2025.

J. Dai, X. Pan, R. Sun, J. Ji, X. Xu, M. Liu, Y. Wang, and Y. Yang. Safe rlhf: Safe reinforcement learning from human feedback. *arXiv preprint arXiv:2310.12773*, 2023.

J. Dai, X. Pan, R. Sun, J. Ji, X. Xu, M. Liu, Y. Wang, and Y. Yang. Safe rlhf: Safe reinforcement learning from human feedback. In *The Twelfth International Conference on Learning Representations*, 2024. Link. Accessed on November, 2024.

DeepSeekAI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36, 2024.

K. N. Dörr. Mapping the field of algorithmic journalism. *Digital journalism*, 4(6):700–722, 2016.

P. B. Ellickson, W. Kar, J. C. Reeder III, and G. Zeng. Using contextual embeddings to predict the effectiveness of novel heterogeneous treatments. *Available at SSRN 4845956*, 2024.

K. Filippova. Controlled hallucinations: Learning to generate faithfully from noisy data. *arXiv preprint arXiv:2010.05873*, 2020.

S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320, 2016.

Forbes. How will chatgpt affect your job if you work in advertising and marketing?, January 2023. Link. Accessed on December, 2024.

Futurism. Buzzfeed is quietly publishing whole ai-generated articles, not just quizzes, March 2023a. Link. Accessed on November, 2024.

Futurism. Cnet is quietly publishing entire articles generated by ai, January 2023b. Link. Accessed on November, 2024.

L. Gao, J. Schulman, and J. Hilton. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, pages 10835–10866. PMLR, 2023.

R. K. Garrett. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of computer-mediated communication*, 14(2):265–285, 2009.

M. Gentzkow and J. M. Shapiro. Media bias and reputation. *Journal of Political Economy*, 114(2):280–316, 2006.

M. Gentzkow and J. M. Shapiro. Competition and truth in the market for news. *Journal of Economic perspectives*, 22 (2):133–154, 2008.

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

M. Gentzkow, J. M. Shapiro, and M. Taddy. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340, 2019.

A. S. Gerber, D. Karlan, and D. Bergan. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52, 2009.

Google. Updated production-ready gemini models, reduced 1.5 pro pricing, increased rate limits, and more, September 2024. Link. Accessed on December, 2024.

T. Groseclose and J. Milyo. A measure of media bias. *The quarterly journal of economics*, 120(4):1191–1237, 2005.

G. Gui and O. Toubia. The challenge of using llms to simulate human behavior: A causal inference perspective. *arXiv preprint arXiv:2312.15524*, 2023.

Harvard Law Review. Nyt v. openai: The times's about-face, April 2024. Link, Accessed on November, 2024.

E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

K. Iizuka, Y. Seki, and M. P. Kato. The effect of news article quality on ad consumption. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3107–3111, 2021.

Jasper. The content marketer's guide to jasper, May 2024. Link. Accessed on December, 2024.

J. Ji, M. Liu, J. Dai, X. Pan, C. Zhang, C. Bian, B. Chen, R. Sun, Y. Wang, and Y. Yang. Beavertails: Towards improved safety alignment of LLM via a human-preference dataset. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023a. Link. Accessed on November, 2024.

Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023b.

L. Lai and A. Farbrot. What makes you click? the effect of question headlines on readership in computer-mediated communication. *Social Influence*, 9(4):289–299, 2014.

G. Lakoff and M. Johnson. *Metaphors we live by*. University of Chicago press, 2008.

H. Laurino Dos Santos and J. Berger. The speed of stories: Semantic progression and narrative success. *Journal of Experimental Psychology: General*, 151(8):1833, 2022.

H. Lee, S. Phatale, H. Mansoor, T. Mesnard, J. Ferret, K. R. Lu, C. Bishop, E. Hall, V. Carbune, A. Rastogi, et al. Rlaif vs. rlhf: Scaling reinforcement learning from human feedback with ai feedback. In *Forty-first International Conference on Machine Learning*, May 2024.

K. Li, T. Zhang, and R. Wang. Deep reinforcement learning for multiobjective optimization. *IEEE transactions on cybernetics*, 51(6):3103–3114, 2020.

P. Li, N. Castelo, Z. Katona, and M. Sarvary. Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Science*, 43(2):254–266, 2024.

F. Liu et al. Learning to summarize from human feedback. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.

H. Liu, C. Li, Q. Wu, and Y. J. Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024a.

Z. Liu, X. Sun, and Z. Zheng. Enhancing llm safety via constrained direct preference optimization. *arXiv preprint arXiv:2403.02475*, 2024b.

S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. Link. Accessed on November, 2024.

Y. Lyu, H. Zhang, S. Niu, and J. Cai. A preliminary exploration of youtubers' use of generative-ai in content creation. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2024.

F. Marconi. *Newsmakers: Artificial intelligence and the future of journalism*. Columbia University Press, 2020.

G. J. Martin and A. Yurukoglu. Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9): 2565–2599, 2017.

Marvik. Finetuning LLMs: Enhancing product descriptions efficiently, December 2023. Link. Accessed on December, 2024.

S. Matz, J. Teeny, S. S. Vaid, H. Peters, G. Harari, and M. Cerf. The potential of generative ai for personalized persuasion at scale. *Scientific Reports*, 14(1):4692, 2024.

Meta. Introducing meta llama 3: The most capable openly available llm to date, April 2024. Link. Accessed on November, 2024.

Meta. Llama 3.2: Revolutionizing edge ai and vision with open, customizable models, September 2024. Link. Accessed on November, 2024.

M. Mosaffa, O. Rafieian, and H. Yoganarasimhan. Visual polarization measurement using counterfactual image generation. *working paper*, 2025.

S. Mullainathan and A. Shleifer. The market for news. *American economic review*, 95(4):1031–1053, 2005.

O. Netzer, A. Lemaire, and M. Herzenstein. When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, 56(6):960–980, 2019.

A. News. Rupert murdoch's attempt to change his family's trust over fox news media empire control rejected, December 2024. Link. Accessed on March, 2025.

S. Nishal and N. Diakopoulos. Envisioning the applications and implications of generative ai for news media. *arXiv preprint arXiv:2402.18835*, 2024.

OpenAI. GPT-4o mini: advancing cost-efficient intelligence, July 2024. Link. Accessed on December, 2024.

L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

A. Panickssery, S. R. Bowman, and S. Feng. Llm evaluators recognize and favor their own generations. *arXiv preprint arXiv:2404.13076*, 2024.

R. Perez Vega. *Measuring the effect of immediacy on consumer engagement behaviours in social media settings*. PhD thesis, Heriot-Watt University, 2016.

R. Rafailov, A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

O. Rafieian and H. Yoganarasimhan. Ai and personalization. *Artificial Intelligence in Marketing*, pages 77–102, 2023.

O. Rafieian, A. Kapoor, and A. Sharma. Multiobjective personalization of marketing interventions. *Marketing Science*, 2024.

J. Rahme and R. P. Adams. A theoretical connection between statistical physics and reinforcement learning. *arXiv preprint arXiv:1906.10228*, 2019.

A. Rame, G. Couairon, C. Dancette, J.-B. Gaya, M. Shukor, L. Soulier, and M. Cord. Rewarded soups: towards pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards. *Advances in Neural Information Processing Systems*, 36, 2024.

J. M. Scacco and A. Muddiman. The curiosity effect: Information seeking in the contemporary news environment. *New Media & Society*, 22(3):429–448, 2020.
J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
W. Sinnott-Armstrong and J. Skorburg. How to cure political polarization by asking questions, April 2019. Link. Accessed on November, 2024.
G. Skelley and H. Fuong. 3 in 10 americans named political polarization as a top issue facing the country, 2022. Link. Accessed on November, 2024.
SmartNews. Why news consumers are a goldmine for advertisers, December 2024. Link. Accessed on March, 2025.
L. G. Smith, E. F. Thomas, A.-M. Bliuc, and C. McGarty. Polarization is the psychological foundation of collective engagement. *Communications Psychology*, 2(1):41, 2024.
N. J. Stroud. Niche news: The politics of news choice, 2011.
R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T. B. Hashimoto. Alpaca: a strong, replicable instruction-following model; 2023. 2023. Link. Accessed on November, 2024.
The Guardian. This article is more than 4 years old. a robot wrote this entire article. are you scared yet, human?, September 2020. Link. Accessed on November, 2024.
The Wall Street Journal. Sports illustrated publisher taps ai to generate articles, story ideas, February 2023. Link. Accessed on November, 2024.
P. H. Thibodeau, T. Matlock, and S. J. Flusberg. The role of metaphor in communication and thought. *Language and Linguistics Compass*, 13(5):e12327, 2019.
TikTok, April 2024. Link. Accessed on November, 2024.
S.-H. Ting and J. Chin-Chin Siaw. Hooks and endings in magazine articles to engage readers. *Human Behavior, Development & Society*, 21(4), 2020.
O. Toubia and O. Netzer. Idea generation, creativity, and prototypicality. *Marketing science*, 36(1):1–20, 2017.
O. Toubia, J. Berger, and J. Eliashberg. How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences*, 118(26):e2011695118, 2021.
S. Ubani, S. O. Polat, and R. Nielsen. Zeroshotdataaug: Generating and augmenting training data with chatgpt. *arXiv preprint arXiv:2304.14334*, 2023.
K. Van Krieken. Do reconstructive and attributive quotes in news narratives influence engagement, credibility and realism? *Journalism Studies*, 21(2):145–161, 2020.
K. Van Moffaert and A. Nowé. Multi-objective reinforcement learning using sets of pareto dominating policies. *The Journal of Machine Learning Research*, 15(1):3483–3512, 2014.
K. Van Moffaert, M. M. Drugan, and A. Nowé. Scalarized multi-objective reinforcement learning: Novel design techniques. In *2013 IEEE symposium on adaptive dynamic programming and reinforcement learning (ADPRL)*, pages 191–199. IEEE, 2013.
Y. Wang, L. Tao, and X. X. Zhang. Recommending for a multi-sided marketplace: A multi-objective hierarchical approach. *Marketing Science*, 2024.
J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
M. Williams. Multi-objective reinforcement learning from ai feedback. *arXiv preprint arXiv:2406.07295*, 2024.
S. M. Xie, A. Raghunathan, P. Liang, and T. Ma. An explanation of in-context learning as implicit bayesian inference. *arXiv preprint arXiv:2111.02080*, 2021.
W. Xu, G. Zhu, X. Zhao, L. Pan, L. Li, and W. Wang. Pride and prejudice: Llm amplifies self-bias in self-refinement. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15474–15492, 2024.

Z. Ye, H. Yoganarasimhan, and Y. Zheng. Lola: Llm-assisted online learning algorithm for content experiments. *arXiv preprint arXiv:2406.02611*, 2024.

H. Yoganarasimhan and I. Iakovetskaia. From feeds to inboxes: A comparative study of polarization in facebook and email news sharing. *Management Science*, 70(9):6461–6472, 2024.

Z. Zhao, E. Wallace, S. Feng, D. Klein, and S. Singh. Calibrate before use: Improving few-shot performance of language models. In *International conference on machine learning*, pages 12697–12706. PMLR, 2021.

Z. Zhou, J. Liu, J. Shao, X. Yue, C. Yang, W. Ouyang, and Y. Qiao. Beyond one-preference-fits-all alignment: Multi-objective direct preference optimization. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 10586–10613, 2024.

E. Zvornicanin. What makes large language models expensive?, June 2024. Link. Accessed on December, 2024.

Appendices

A Prevalence of Topics in New York Times Articles

Figure A1 shows the prevalence of the 40 most popular topics across all articles in the corpus from the LDA analysis, in decreasing order of popularity. We see that the most prevalent topics in our corpus are *Family*, *Politics*, and *Emotions and Feelings*, while the least prevalent include *Judaism*, *Pets and Animals*, and *Horse Racing and Farms*. For further details of the LDA analysis and the keywords recovered for each topic, please see Yoganarasimhan and Iakovetskaia (2024).

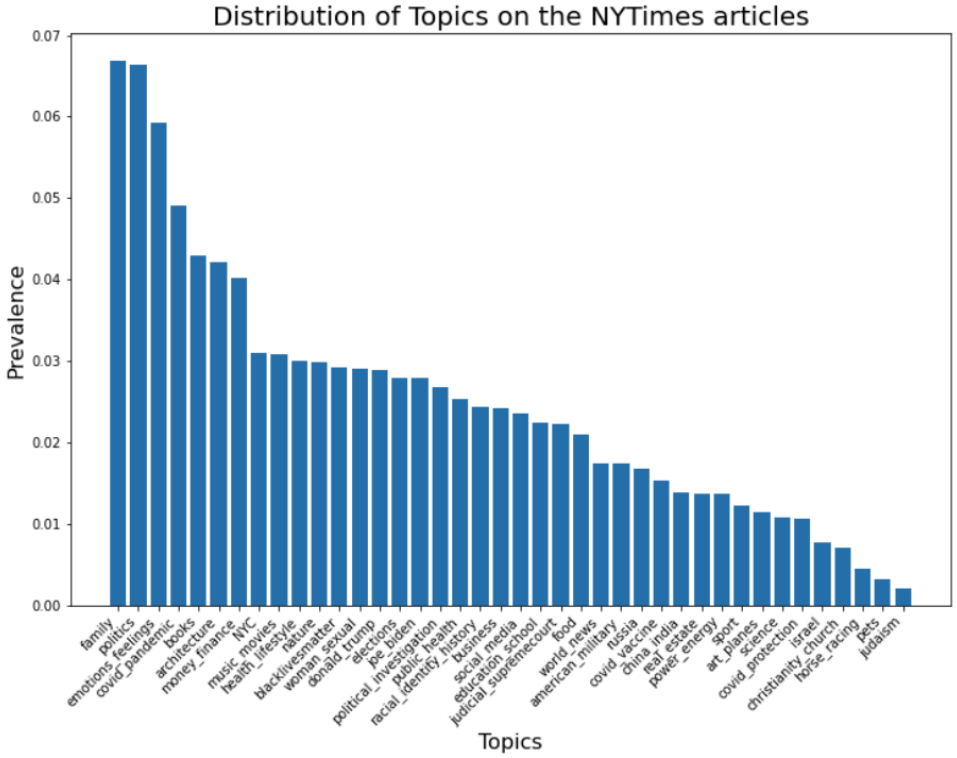


Figure A1: Topic Prevalence on the New York Times Corpus

B Prompting Procedure to Generate Synthetic Data

In our study, we use the "GPT-4o mini" model to generate responses in a zero-shot manner, enabling us to 1) evaluate the polarization and engagement levels of the original *New York Times* articles, 2) create both a more engaging and a less engaging version of each article, and 3) assess the engagement and polarization levels of these generated versions. These tasks would be largely infeasible using traditional survey methods. However, GPT-4o mini model can easily analyze the large corpus of 13508 articles, extract polarization and engagement scores, and generate articles with varying engagement levels (Gilardi et al. 2023). Additionally, the GPT-4o mini model has been optimized for chat using the Chat Completion API and can produce responses based on the provided chat history. Therefore, for consistency, we use the same prompt below as the input whenever we

run the model on a new article.[28] Building on recent works (Zhang et al. 2021; He et al. 2023) that leverage LLM-generated synthetic data for model training, our prompting procedure to produce synthetic data is as follows:

- Step 1: Evaluating the polarization and engagement levels of the original article.

  *User: I will provide you with some text from a news article. Please rate from 1 to 10 how polarizing this article is. Let's call this article as "original version":*article text**

  *User: I will provide you with some text from a news article. Please rate from 1 to 10 how engaging this article is. Let's call this article as "original version":*article text**

- Step 2: Generating two new versions of the original article by making it more engaging or less engaging.

  User: Could you help make the text of this article more engaging and attract more attention? Let's call this new version "version A".

  User: Could you help make the text of this article less engaging and attract less attention? Let's call this new version "version B".

- Step 3: Evaluating the engagement levels of the two generated versions.

  User: Please rate from 1 to 10 how engaging "version A" is and "version B" is, and explain which version of the article is more engaging and likely to attract more attention:
  *"version A": *more engaging version generated from Step 2**
  *"version B": *less engaging version generated from Step 2*.*

- Step 4: Evaluating the polarization levels of the two generated versions.

  User: Please rate from 1 to 10 how polarizing "version A" is and "version B" is, and explain which version of the article is less polarizing:
  *"version A": *more engaging version generated from Step 2**
  *"version B": *less engaging version generated from Step 2**

C Example of Article modified by GPT

We provide an example from *The New York Times*, published in May 2020, which reports on so-called "coronavirus parties" as a potential means of gaining immunity. The original article presents the issue with a cautious tone:

[28] We do not explicitly define engagement and polarization in the LLM prompt used for our main analysis. However, we perform a robustness check by including the same definitions provided to human raters in Web Appendix §E.3, where "engaging" is defined as "a news article is considered engaging if it attracts your attention and keeps you interested from start to finish," and "polarizing" is defined as "a news article is considered politically polarizing if it contains extreme views from either Democrats or Republicans." We find that the ratings generated by the GPT-4o mini model remain consistent regardless of whether these definitions are included.

"Covid-19 Parties" Probably Didn't Involve Intentional Spread

Amid growing impatience over stay-at-home orders and rising unemployment, public health experts have worried that some people may try to expose themselves to the coronavirus in a risky bid to gain immunity. One fear is the prospect of "coronavirus parties," much like the chickenpox parties of the past that preceded the development of a chickenpox vaccine, designed to deliberately spread infection. County health officials in southeastern Washington State reported this week that they had evidence that one or more such gatherings had been linked to at least two new coronavirus cases. But on Wednesday night, the officials retracted those comments and said the so-called Covid-19 parties may have been more innocent gatherings. Meghan DeBolt, the director of community health for Walla Walla County, said county officials were learning more about the cases that have emerged from the recent social gatherings. She said they were still hearing reports of parties where infected people were present but do not have evidence that the people who became ill after the gatherings had attended out of a desire to be exposed. The county had said in a news release on Monday that the authorities were "receiving reports of Covid-19 parties occurring in our community, where non-infected people mingle with an infected person in an effort to catch the virus." Officials later elaborated on those reports in interviews, saying the parties were discovered after tracing the paths of people who were found to be infected after the gatherings. Officials in Walla Walla have been working in recent weeks to contain a large outbreak at a meat processing facility in the area. Ms. DeBolt said the county was not close to halting the rise in infections, and health officials are concerned that more people are engaging in ill-advised social interactions despite stay-at-home requirements. "We know that people are exhausted from isolation and quarantine," Ms. DeBolt said. "We want to be able to reopen, too. We want to be able to go to restaurants and socialize with friends and family members. We need our community's help to be diligent for a little bit longer so that we can get ahead of this." The prospect of infection parties for people who wish to quickly contract the disease in the hope of gaining immunity has been a fear among some health experts because the country has a long history of people choosing purposeful infection. In the case of chickenpox parties, the Centers for Disease Control and Prevention "strongly recommends against hosting or participating in these events.

By contrast, the LLM-generated "more engaging" version adopts more emotionally charged language and creates a heightened sense of urgency. It frames the "Covid-19 parties" as a major public threat, which aligns with Democrats' political narratives advocating for strong government intervention to protect public health based on expert recommendations:

The Shocking Rise of "Covid-19 Parties": A Dangerous Gamble for Immunity?
As frustration mounts over relentless stay-at-home orders and skyrocketing unemployment, a troubling trend is emerging: the rise of "Covid-19 parties." In a desperate bid for immunity, some individuals are reportedly risking it all to expose themselves to the coronavirus, reminiscent of the infamous chickenpox parties from years past. Imagine a gathering where the goal is not to celebrate, but to contract a virus. Public health experts are sounding the alarm as reports surface of these reckless events, where non-infected individuals mingle with those who have tested positive for COVID-19, hoping to catch the virus and gain immunity. Recently, health officials in southeastern Washington State revealed they had evidence linking such gatherings to at least two new cases of the virus. However, just as quickly, they walked back those statements, suggesting that these so-called "Covid-19 parties" might not be as sinister as initially thought. Meghan DeBolt, the director of community health for Walla Walla County, explained, "We're still piecing together what's happening. While we've heard reports of gatherings with infected individuals, we don't have clear evidence that attendees are intentionally trying to get sick." The situation is further complicated by a significant outbreak at a local meat processing facility, prompting officials to urge the community to remain vigilant. "We know that people are exhausted from isolation and quarantine," DeBolt continued. "We want to reopen, too. But we need everyone's help to stay diligent just a little longer. The stakes are too high." The concept of infection parties raises serious concerns among health experts, who recall a troubling history of purposeful infections. The Centers for Disease Control and Prevention has strongly advised against such gatherings, warning of the potential consequences that could arise from this dangerous trend. As the pandemic drags on, the question looms: Are we witnessing a desperate attempt to reclaim normalcy, or are these gatherings a reckless gamble with lives at stake? Only time will tell, but one thing is clear: the fight against COVID-19 is far from over, and the choices we make today could shape the future of our communities.

In contrast, the LLM-generated "less engaging" version is more neutral and factual:

Reports of Gatherings During COVID-19

Some individuals have been reported to be attending gatherings during the COVID-19 pandemic, which are sometimes referred to as "Covid-19 parties." These events are thought to involve people trying to expose themselves to the coronavirus in order to gain immunity, similar to past chickenpox parties. Recently, health officials in southeastern Washington State mentioned that there may be a connection between these gatherings and a few new coronavirus cases. However, they later clarified that the gatherings might not have been as intentional as previously suggested. Meghan DeBolt, the director of community health for Walla Walla County, stated that officials are still investigating the situation. She noted that while there have been reports of parties with infected individuals, there is no definitive evidence that attendees are deliberately trying to contract the virus. The county is currently dealing with an outbreak at a local meat processing facility, and officials are concerned about the potential for more infections due to social interactions that go against stay-at-home orders. DeBolt mentioned that the community needs to continue following health guidelines to help control the spread of the virus. Health experts have expressed concerns about the idea of infection parties, given the history of similar practices. The Centers for Disease Control and Prevention advises against hosting or participating in such events. In summary, while there are reports of gatherings during the pandemic, the motivations behind them are still unclear, and health officials continue to monitor the situation.

D Details of the LoRA Fine Tuning Technique

We illustrate LoRA (Hu et al. 2021), a Parameter-Efficient Fine-Tuning (PEFT) method that enables fine-tuning on hardware with limited GPU memory while maintaining competitive model performance, in Figure A2. Let W represent a pre-trained weight matrix, such as W_Q, W_K, or W_V, which represent the query, key, and value vectors in the architecture of the Transformer's attention mechanism. These matrices store most of the information and knowledge learned by a LLM, and play a crucial role in calculating the attention scores and subsequently determining the relevance of different tokens in the input sequence. We use $d \times d_W$ to represent the dimension of the weight matrix and r to represent the rank of the low-rank matrices. LoRA modifies the weight matrix W during fine-tuning as follows:

$$W' = W + \Delta W, \text{where} \quad \Delta W = AB, \quad \text{with} \quad A \in R^{d \times r,} \quad B \in R^{r \times d_w} \tag{A1}$$

Essentially, LoRA introduces low-rank matrices A and B to model the weight matrix change ΔW during fine-tuning, while keeping the original weight matrix W frozen. This strategy significantly reduces the number of trainable parameters from $d \times d_W$ to $r \times (d + d_W)$, as the rank r is typically set to a small value, such as 4 or 8, to balance the trade-off between model capacity and computational efficiency.

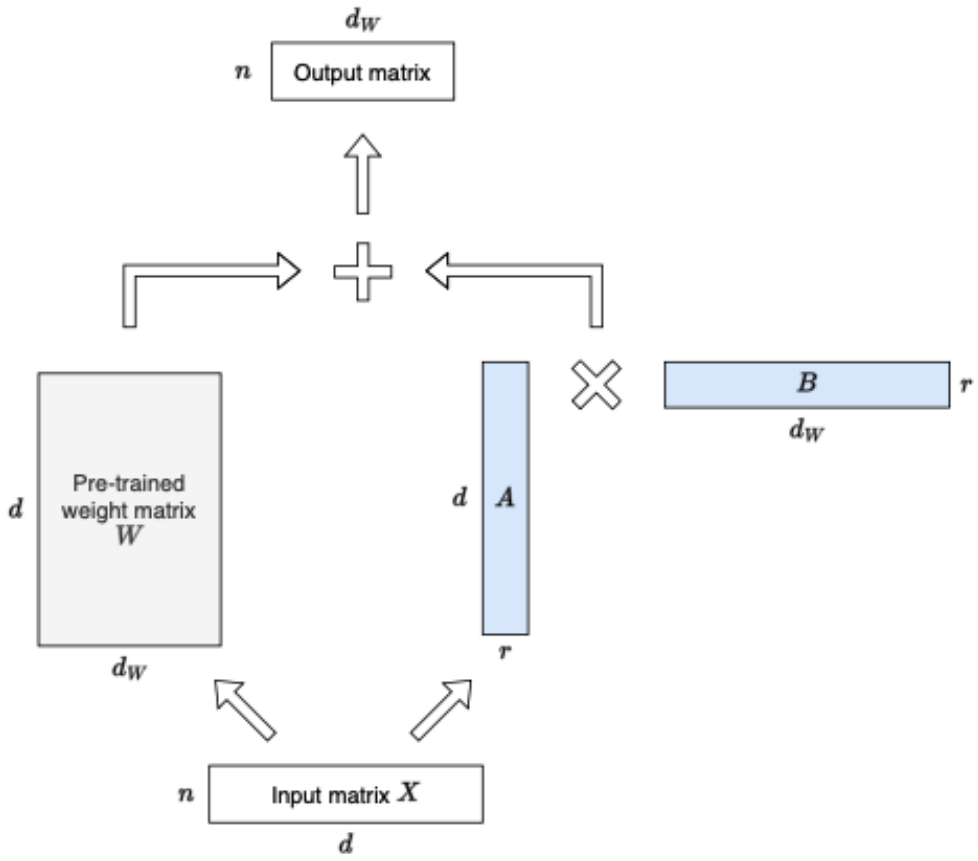


Figure A2: Illustration of LoRA Fine-Tuning

E Evaluation Ratings for Polarization and Engagement for the Main Model

In this section, we provide details of four approaches for obtaining article evaluation ratings on engagement and polarization levels: (a) Gemini, (b) Claude, (c) human coders, and (d) lab experiments on Qualtrics.

E.1 Prompts for Gemini Evaluation

We use Gemini to assess the polarization and engagement levels of the articles in §7.1. To evaluate polarization, we use the following prompt:

"I will provide you with three versions of lead paragraphs from a news article. Please rate each one ("Version A", "Version B", and "Version C") on a scale of 1 to 10 for how polarizing it is. Afterward, explain which version you find the most polarizing, which version you find the least polarizing, or if you feel all versions are equally polarizing:

"Version A": the original article

"Version B": the article generated by the multi-objective DPO model

"Version C": the article generated by the engaging DPO model

Similarly, to evaluate engagement, we used the following prompt:

"I will provide you with three versions of lead paragraphs from a news article. Please rate each one ("Version

A", "Version B", and "Version C") on a scale of 1 to 10 for how engaging it is. Afterward, explain which version you find the most engaging, which version you find the least engaging, or if you feel all versions are equally engaging:
"Version A": the original article
"Version B": the article generated by the multi-objective DPO model
"Version C": the article generated by the engaging DPO model

E.2 Ratings from Claude Evaluation

We also use an alternative LLM – "claude-3-5-sonnet-20240620," a top-ranked and widely used model known for its strong performance on major NLP benchmarks. To evaluate polarization, we use the following prompt:

"I will provide you with three versions of lead paragraphs from a news article. Please rate each one ("Version A", "Version B", and "Version C") on a scale of 1 to 10 for how polarizing it is. Afterward, explain which version you find the most polarizing, which version you find the least polarizing, or if you feel all versions are equally polarizing:
"Version A": the original article
"Version B": the article generated by the multi-objective DPO model
"Version C": the article generated by the engaging DPO model

Similarly, to evaluate engagement, we used the following prompt:
"I will provide you with three versions of lead paragraphs from a news article. Please rate each one ("Version A", "Version B", and "Version C") on a scale of 1 to 10 for how engaging it is. Afterward, explain which version you find the most engaging, which version you find the least engaging, or if you feel all versions are equally engaging:
"Version A": the original article
"Version B": the article generated by the multi-objective DPO model
"Version C": the article generated by the engaging DPO model

Table A1 shows how the engagement and polarization scores of the MODPO and engaging DPO versions change compared to the original version, as well as to each other. We find that the articles modified using both *the engaging DPO* and *multi-objective DPO* are significantly more engaging than the original articles ($\Delta_{\text{EngDPO_ORIG}} = 1.49$, $p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 1.82$, $p < 0.001$). Interestingly, articles generated by *the multi-objective DPO model* exhibit even higher engagement than those generated by *the engaging DPO*, which suggests that moving from a single-objective setting (focused solely on maximizing engagement) to a multi-objective setting does not necessarily compromise engagement, which tends to be the primary metric of interest for firms. Further, *the multi-objective DPO model* is effective in controlling polarization compared to *the engaging DPO model* ($\Delta_{\text{MODPO_EngDPO}} = -0.27$, $p < 0.001$). However, it still leads to some increase in polarization compared to the original version ($\Delta_{\text{MODPO_ORIG}} = 0.88$, $p < 0.001$), indicating that it is not able to fully avoid any increase in polarization, which is understandable given the desire to simultaneously

boost engagement. In sum, our results demonstrate that *the multi-objective DPO model* effectively balances engagement and polarization, allowing firms and content creators to mindfully leverage open-source LLMs in combination with multi-objective alignment to successfully generate content that is engaging for users while managing unintended negative consequences such as polarization.

Table A1: Claude Evaluation Results (with *engaging SFT* as the reference policy)

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$
Mean	1.82***	1.49***	0.32**	0.88***	1.15***	-0.27***
Std	1.34	1.27	1.11	2.10	1.96	1.24
Min	-4.00	-4.00	-3.00	-6.00	-6.00	-5.00
25%	2.00	1.00	-1.00	1.00	1.00	-1.00
50%	2.00	1.00	1.00	1.00	2.00	-1.00
75%	3.00	2.00	1.00	2.00	2.00	1.00
Max	5.00	5.00	4.00	5.00	5.00	2.00
N	983.00	983.00	983.00	983.00	983.00	983.00

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

E.3 Ratings from Human Raters

We hired a human rater to evaluate news articles on polarization and engagement. To ensure consistency, transparency, and reliability in this annotation process, we developed a detailed annotation rubric. In the rubric, political polarization is defined as the degree to which a news article contains extreme views from either Democrats or Republicans. Additionally, it also includes some specific cues to guide raters in evaluating polarization:

1. Framing of Issues
    - 1-3: The article lacks a clear framing of polarization issues or pays minimal attention to them.
    - 4-6: Issues are framed in a basic manner, with mild divisive undertones.
    - 7-8: The framing acknowledges multiple perspectives but creates divisions and makes assertive claims.
    - 9-10: The framing strongly highlights clear divisions or creates larger divisions.
2. Amount of Perspectives Included
    - 1-3: The article actively presents different perspectives equally, maintaining an informational tone without promoting a specific agenda.
    - 4-6: Diverse perspectives are integrated, demonstrating an effort toward inclusion.
    - 7-8: Some perspectives are included, but the article subtly reinforces a particular agenda or framing.
    - 9-10: Some key voices are marginalized, or significant perspectives are omitted entirely.
3. Language and Tone

- 1-3: The language is neutral, factual, and unbiased.
- 4-6: The tone is slightly subjective, reflecting underlying opinions without overt bias; speculative language may be present.
- 7-8: Clear biases emerge, either positively or negatively, with a more definitive tone.
- 9-10: The language is heavily charged and intentionally inflammatory or manipulative.

4. Informativeness

    - 1-3: The article is factually rich, comprehensive, and balanced.
    - 4-6: It is well-researched and substantial but may lack key details or alternative viewpoints.
    - 7-8: While including relevant facts, the content lacks depth or clarity.
    - 9-10: Factual information is very limited.

In the rubric, engagement is defined as the extent to which an article attracts reader attention and keeps the reader interested from beginning to end. The rubric also provides some specific cues to guide raters in assessing engagement:

1. Emotional Resonance

    - 1-3: Lacks emotional appeal or connection.
    - 4-6: Exhibits basic emotional appeal with limited resonance.
    - 7-8: Demonstrates moderate emotional engagement, resonating with some readers.
    - 9-10: Creates a strong emotional connection, deeply resonating with a broad audience.

2. Call to Action or Reflection

    - 1-3: Lacks a clear takeaway or purpose.
    - 4-6: Provides minimal call to action or a vague reflection.
    - 7-8: Provides clear call to action or a modest reflection.
    - 9-10: Offers a strong, compelling call to action or encourages deep reflection, effectively motivating the audience.

3. Opening

    - 1-3: Starts weakly, offering little context or interest.
    - 4-6: Moderately engaging, though it lacks a strong hook.
    - 7-8: Opens with a compelling question, statement, or scenario that grabs attention.
    - 9-10: Delivers an exceptionally compelling opening, immediately drawing the readers in.

4. Irrelevance

- 1-3: Overly lengthy with significant irrelevant content.
- 4-6: Contains sections that feel too long or irrelevant but offer some value.
- 7-8: Includes minor irrelevant parts that detract from the main message.
- 9-10: Concise and focused, staying relevant and to the point throughout.

5. Storytelling

   - 1-3: Lacks coherence and a clear flow.
   - 4-6: Features a basic narrative structure but may lack fluidity or strong transitions.
   - 7-8: Provides a well-balanced narrative with good integration of storytelling and information.
   - 9-10: Delivers a compelling narrative with a strong structure, smoothly guiding the reader through the key points.

The polarization and engagement ratings from human rating are summarized in Table A3. Consistent with our observations discussed in §3, we also find a significant positive relationship between polarization and engagement, regardless of whether we control for the topic distribution of each article (column (1): β = 0.728, $p < 0.001$; column (2): β = 0.724, $p < 0.001$). Furthermore, the results show that articles generated using *the engaging DPO model* or *the multi-objective DPO model* are significantly more engaging compared to the original articles ($\Delta_{\text{EngDPO_ORIG}} = 1.39$, $p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 1.31$, $p < 0.001$), and there is no significant difference in engagement between the two methods ($\Delta_{\text{MODPO_EngDPO}} = -0.07$, $p > 0.1$). Notably, while both methods enhance engagement, articles generated using *the multi-objective DPO* exhibit a substantial reduction in polarization ($\Delta_{\text{MODPO_EngDPO}} = -0.46$, $p < 0.001$) compared to *the engaging DPO*.

Table A2: Relationship between Polarization and Engagement in Human Annotation Data

	(1) Polarization		(2) Polarization	
Engagement	0.728***	(0.0250)	0.724***	(0.0258)
Constant	1.217***	(0.0795)	1.835**	(0.704)
Topics	No		Yes	
N	983		983	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E.4 Ratings from Qualtrics Experiments

Our experiment was aimed at understanding how people rate news articles in terms of polarization and engagement. To achieve this, we conducted a lab experiment with 230 undergraduate students from a major public U.S. university.[29] Each participant was randomly assigned to review three different articles presented in a randomized order: one original article, one MODPO article, and one Engaging DPO article. Participants

[29] Of these, 180 provided valid responses; the remaining students indicated that they did not consider their responses to be of good quality and chose not to include them in the study.

Table A3: Human Annotation Evaluation Results (with *engaging SFT* as the reference policy)

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$
Mean	1.31***	1.39***	-0.07	1.42***	1.89***	-0.46***
Std	1.13	1.24	1.24	1.34	1.61	1.57
Min	-2.00	-3.00	-3.00	-4.00	-5.00	-4.00
25%	1.00	1.00	-1.00	1.00	1.00	-2.00
50%	1.00	1.00	0.00	1.00	2.00	-1.00
75%	2.00	2.00	1.00	2.00	3.00	1.00
Max	5.00	6.00	4.00	6.00	6.00	5.00
N	983.00	983.00	983.00	983.00	983.00	983.00

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

were asked to review snippets of these articles and evaluate them on two dimensions: (1) how engaging or interesting they found the article, and (2) how politically polarizing it appeared. To ensure thoughtful evaluations, we first provided participants with clear definitions of both political polarization and engagement, along with illustrative examples that demonstrated high and low levels of each dimension. Later, they were required to spend a minimum of 30 seconds reviewing and rating each article. Figures A3 and A4 provide the definitions and illustrative examples shown to participants for polarization and engagement, respectively. The user interface used for article rating is presented in Figure A5.

The polarization and engagement ratings from the Qualtrics experiment are summarized in Table A4. The results show that articles generated using *the engaging DPO* or *the multi-objective DPO* were rated as significantly more engaging compared to the original articles ($\Delta_{\text{EngDPO_ORIG}} = 0.62$, $p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 0.62$, $p < 0.001$). Notably, while both methods enhance engagement, articles generated using *the multi-objective DPO model* exhibit a substantial reduction in polarization compared to *the engaging DPO* ($\Delta_{\text{MODPO_EngDPO}} = -0.43$, $p < 0.001$).

Table A4: Ratings from Qualtrics Experiment (with *engaging SFT* as the reference policy)

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$
Mean	0.62***	0.62***	-0.01	-0.20***	0.23***	-0.43***
Std	2.72	2.55	2.49	2.73	2.93	2.62
Min	-7.00	-6.00	-7.00	-9.00	-9.00	-6.00
25%	-1.00	-1.00	-1.25	-2.00	-1.00	-2.00
50%	1.00	1.00	0.00	0.00	0.00	-0.50
75%	2.00	2.00	1.25	1.00	2.00	1.00
Max	8.00	6.00	8.00	9.00	8.00	9.00
N	180.00	180.00	180.00	180.00	180.00	180.00

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Please read the following definition of **Political Polarization**: A news article is considered politically polarized if it contains extreme views from either Democrats or Republicans.

To help you calibrate and rate the extent of polarization in a given article, we provide you with two versions of the same article that use different phrasing and then show you some typical ratings.

Version A: Trumpism Is a Con Game, and Steve Bannon Was Its Master Manipulator

In the MAGA cult, you're either a predator or a patsy. The latest Senate Intelligence report on Russian campaign interference exposes the rot at the core of Trumpism. A footnote quotes Steve Bannon, the devious architect of Donald Trump's 2016 campaign, mocking Trump's oldest son. Bannon, who reportedly thought "very highly" of Donald Trump Jr., privately sneered that Trump Jr. was "a guy who believes everything on Breitbart is true."

Bannon, the puppeteer behind Breitbart, the infamous propaganda machine, knew better. He orchestrated a toxic narrative that fueled Trump's rise, all while cynically despising the very people he manipulated. He knew that Trump's base — who he derogatorily called "hobbits" — were just pawns in his game of deceit.

Shameless and cunning, Bannon cloaked himself in the image of a populist hero, even as he looked down on the working-class Americans he pretended to champion. In "The Brink," a 2019 documentary about Bannon, he's shown pandering to supporters in a cluttered living room adorned with crosses. As he fawns over the room's "working-class" vibe, he's already thinking of his next move: stepping into his chauffeured car and jetting off to a private airport. "You couldn't pay me a million dollars a year to live in that house," snickered one of his cronies.

Version B: Examining the Complexities of Trumpism and Steve Bannon's Role

Navigating the Dynamics of the MAGA Movement The recent Senate Intelligence report on Russian campaign interference sheds light on various aspects of the 2016 presidential campaign. Within the report, a footnote mentions Steve Bannon, the former chief executive of Donald Trump's campaign, expressing his thoughts on Donald Trump Jr. Bannon noted that he thought "very highly" of Trump Jr., while also commenting on his belief in the news presented by Breitbart.

Bannon, who led Breitbart, a media outlet known for its strong right-wing stance, played a significant role in shaping the narrative that resonated with Trump's base. His ability to connect with this audience was evident, though some of his private comments suggest a more nuanced view of the movement he helped drive.

In "The Brink," a 2019 documentary about Bannon, he is portrayed interacting with supporters in a modest living room. He relates the space to his grandmother's house and acknowledges the role of working-class and middle-class individuals in nationalist movements. However, the documentary also highlights moments where Bannon and his associate exhibit behavior that contrasts with the image they project publicly.

○ I have read both articles and am ready to proceed

Figure A3: Definition and Illustrative Example of Polarization Rating Task

Please read the following definition of **Engaging**: A news article is considered engaging if it attracts your attention and keeps you interested from start to finish.

To help you calibrate and rate the extent of how engaging a given article is, we provide you with two versions of the same article that use different phrasing and then show you some typical ratings.

Version A: Showdown on Fifth Avenue: 'Black Lives Matter' to be Painted Outside Trump Tower

In a bold move that's sure to spark national conversation, New York City Mayor Bill de Blasio has ordered the words "Black Lives Matter" to be emblazoned in massive yellow letters right outside Trump Tower. The mural, which will stretch across Fifth Avenue between 56th and 57th Streets, is set to be completed in the coming week, making a powerful statement in one of the city's most iconic locations.

"The president is a disgrace to the values we cherish in New York City," declared Julia Arredondo, a spokeswoman for Mayor de Blasio, in a fiery statement released Thursday. "He can't run or deny the reality we are facing, and any time he wants to set foot in the place he claims is his hometown, he should be reminded: Black Lives Matter."

Predictably, President Trump didn't take the news lightly. He took to Twitter, lambasting the plan to paint what he described as "the fabled & beautiful Fifth Avenue, right in front of Trump Tower/Tiffany," and in a provocative twist, attempted to stoke tensions between the Black Lives Matter movement and the police. Mr. Trump falsely claimed that BLM protesters in New York had chanted for the killing of police officers, referencing a 2015 protest in Minnesota with the chant, "Pigs in a Blanket, Fry 'Em Like Bacon," which was recently aired on Tucker Carlson's show on Fox News. "NYC Police are furious!" Trump declared.

Version B: 'Black Lives Matter' Will Be Painted on Street Outside Trump Tower

Mayor Bill de Blasio has decided that the words "Black Lives Matter" will be painted in large yellow letters on the street outside Trump Tower. The painting is expected to be done in the coming week on Fifth Avenue, between 56th and 57th Streets, according to the city.

Julia Arredondo, a spokeswoman for Mayor de Blasio, stated on Thursday that "The president is a disgrace to the values we cherish in New York City. He can't run or deny the reality we are facing, and any time he wants to set foot in the place he claims is his hometown, he should be reminded Black Lives Matter."

In response, President Trump tweeted about Mr. de Blasio's plan to paint "the fabled & beautiful Fifth Avenue, right in front of Trump Tower/Tiffany." He also mentioned animosity between the Black Lives Matter movement and the police in New York City. Mr. Trump claimed in the tweet that Black Lives Matter protesters in the city had shouted chants encouraging the killing of police officers. He referred to a chant from a 2015 protest in Minnesota, saying "'Pigs in a Blanket, Fry 'Em Like Bacon', referring to killing Police, is their chant." A scene from that protest had recently been shown in a segment on Tucker Carlson's show on Fox News. Mr. Trump added in his tweet that "NYC Police are furious!"

○ I have read both articles and am ready to proceed

Figure A4: Definition and Illustrative Example of Engagement Rating Task

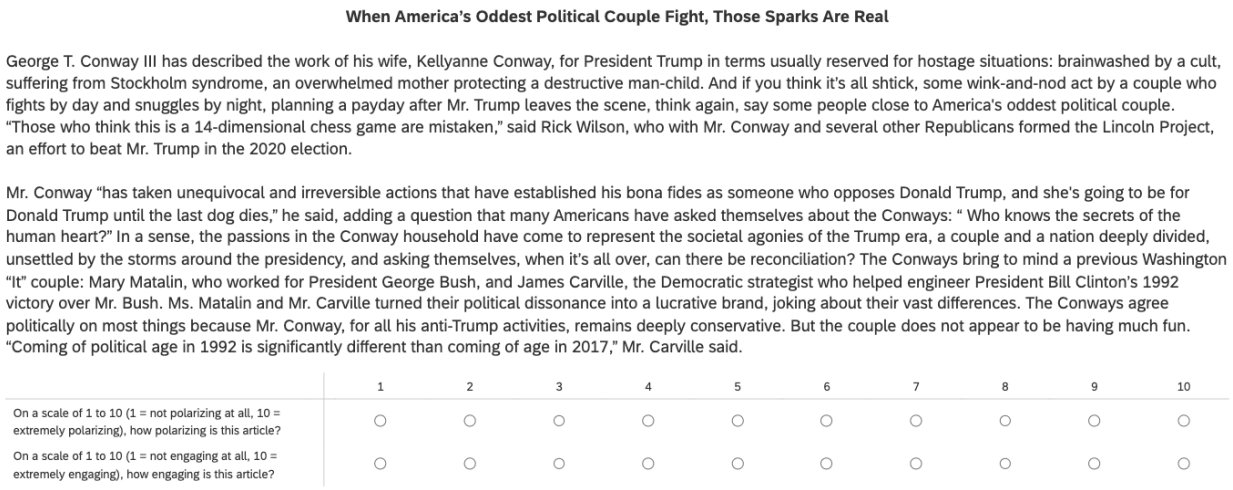


Figure A5: User Interface for Article Rating

F Raw Values for the Evaluation Results

Table A5 presents the raw polarization levels and engagement scores obtained from "Gemini-1.5-Flash" for the original article, the MODPO version, and the less polarizing DPO version we illustrate in §7.2.2.

Table A5: Evaluation Results Using the *less-polarized SFT* as the Reference Policy

	Engagement Score			Polarization Score		
	Eng_{ORIG}	$\text{Eng}_{\text{MODPO}}$	$\text{Eng}_{\text{LPolDPO}}$	Pol_{ORIG}	$\text{Pol}_{\text{MODPO}}$	$\text{Pol}_{\text{LPolDPO}}$
Mean	6.81	6.77	3.58	6.40	5.62	3.11
Std	1.45	1.78	1.35	2.09	1.46	1.83
Min	1.00	1.00	1.00	1.00	1.00	1.00
25%	6.00	5.00	3.00	5.00	5.00	2.00
50%	7.00	7.00	3.00	7.00	6.00	3.00
75%	8.00	8.00	4.00	8.00	7.00	4.00
Max	10.00	9.00	9.00	10.00	9.00	10.00
N	998.00	998.00	998.00	998.00	998.00	998.00

Table A6 presents the raw left-leaning polarization levels and engagement scores obtained from "Gemini-1.5-Flash" for the original article, the MODPO version, and the left-leaning DPO version we illustrate in §8.1.

Table A6: Evaluation Results Using the *left-leaning SFT* as the Reference Policy

	Left-leaning Polarization Score			Engagement Score		
	Pol_{ORIG}	$\text{Pol}_{\text{MODPO}}$	$\text{Pol}_{\text{LeftDPO}}$	Eng_{ORIG}	$\text{Eng}_{\text{MODPO}}$	$\text{Eng}_{\text{LeftDPO}}$
Mean	4.30	4.46	5.59	6.29	6.91	7.79
Std	1.02	1.29	1.69	1.02	1.05	2.33
Min	1.00	1.00	1.00	2.00	3.00	1.00
25%	4.00	4.00	5.00	6.00	6.00	8.00
50%	4.00	5.00	6.00	6.00	7.00	9.00
75%	5.00	5.00	7.00	7.00	8.00	9.00
Max	8.00	8.00	9.00	8.00	9.00	10.00
N	534.00	534.00	534.00	534.00	534.00	534.00

G Prompts for Polarization and Engagement Evaluation when Changing the Reference Policy

We use Gemini to assess the polarization and engagement levels of the articles in §7.2.2. To evaluate polarization, we use the following prompt:

"I will provide you with three versions of lead paragraphs from a news article. Please rate each one ("Version A", "Version B", and "Version C") on a scale of 1 to 10 for how polarizing it is. Afterward, explain which version you find the most polarizing, which version you find the least polarizing, or if you feel all versions are equally polarizing:

*"Version A": *the original article**

*"Version B": *the article generated by the multi-objective DPO model**

*"Version C": *the article generated by the less polarizing DPO model**

Similarly, to evaluate engagement, we used the following prompt:

"I will provide you with three versions of lead paragraphs from a news article. Please rate each one ("Version A", "Version B", and "Version C") on a scale of 1 to 10 for how engaging it is. Afterward, explain which version you find the most engaging, which version you find the least engaging, or if you feel all versions are equally engaging:

*"Version A": *the original article**

*"Version B": *the article generated by the multi-objective DPO model**

*"Version C": *the article generated by the less polarizing DPO model**

H Appendix for Content Strategies

H.1 Prompting Procedure to Generate Ratings for Content Strategies

Here, we present the exact prompts used to generate ratings for the 12 theory-driven content strategies. These ratings are based on three key sets of characteristics outlined in §7.3.1: (1) content structure and flow, (2) emotional and persuasive elements, and (3) informational and factual content elements. To ensure clarity, coherence, and high-quality responses, we structured the prompts to focus on a set of content features within each query rather than incorporating all 12 content features at once. This approach helps maintain a

manageable input context length while optimizing the effectiveness of each prompt. Below, we provide the exact prompts used:

- Content Structure and Flow:
  User: I'll provide you with three versions of some text from a news article, labeled "Version A," "Version B," and "Version C." Please evaluate each version on the following aspects, using a scale of 1 to 10:
  Strength of opening – How strong is the opening? Rate this under the variable strong_opening.
  Frame the issues as questions – To what extent are issues presented as questions? Rate this under frame_question.
  Narrative structure – How much storytelling does the content incorporate? Rate this under narrative_structure.
  Below are the three versions of the article for evaluation:
  *"Version A": *the original article**
  *"Version B": *the version generated by the multi-objective DPO model**
  *"Version C": *the version generated by the engaging DPO model**

- Emotional and Persuasive Techniques:
  I'll provide you with three versions of some text from a news article, labeled "Version A," "Version B," and "Version C." Please evaluate each version on the following aspects, using a scale of 1 to 10:
  Emotion level – How emotional is the content? Rate this under emotion_level.
  Dramatic metaphor – To what extent does the article use dramatic metaphors? Rate this under dramatic_metaphor.
  Provocative language – How provocative is the language? Rate this under provocative_language.
  Definitive language – How definitive and assertive is the language used? Rate this under definitive_language.
  Call to action – Does the content prompt the readers to take specific action? Rate this under call_to_action.
  Urgency – How urgently does the content present the issues? Rate this under urgency_level.
  Below are the three versions of the article for evaluation:
  *"Version A": *the original article**
  *"Version B": *the version generated by the multi-objective DPO model**
  *"Version C": *the version generated by the engaging DPO model**

- Informational and Factual Elements:
  I'll provide you with three versions of some text from a news article, labeled "Version A," "Version B," and "Version C." Please evaluate each version on the following aspects, using a scale of 1 to 10:
  Fact Level – How factual is the content? Rate this under fact_level.
  Relevant statistics – How much relevant statistics is included in the content? Rate this under relevant_stat.
  Balanced perspective – How balanced are the perspectives toward the issues discussed? Rate this under balanced_perspective.

Below are the three versions of the article for evaluation:
*"Version A": *the original article**
*"Version B": *the version generated by the multi-objective DPO model**
*"Version C": *the version generated by the engaging DPO model**

H.2 Relationship between Content Strategies and Engagement/Polarization

Table A7 presents the regression results on the relationship between content strategies and engagement/polarization. The findings indicate that the content features discussed in §7.3.1 account for 61.44% of the variance in an article's polarization score, but only 15.54% of the variance in its engagement score. Beyond overall variance explained, we also examine how specific content features influence engagement and polarization. *Strong openings* and *narrative structure* emerge as important drivers of engagement, exhibiting a significant positive relationship with engagement scores ($b_{\text{strong_opening}} = 0.267$, $p < 0.001$; $b_{\text{narrative_structure}} = 0.109$, $p < 0.001$). In contrast, *provocative language* and *urgency level* are positively associated with polarization ($b_{\text{provocative_language}} = 0.718$, $p < 0.001$; $b_{\text{urgency_level}} = 0.287$, $p < 0.001$), while a *balanced perspective* appears to reduce polarization, as indicated by its negative coefficient ($b_{\text{balanced_perspective}} = -0.360$, $p < 0.001$).

H.3 More Examples of Content Strategies

We provide two additional examples to illustrate the content strategies across the three article versions. Figure A6 presents the first example of the three versions of the created articles, where version A is the

(A) Original Version	(B) Multi-Objective DPO Version	(C) Engaging DPO Version
How Voters Turned Virginia From Deep Red to Solid Blue	How Voters Turned Virginia From Deep Red to Solid Blue	From Red to Blue: The Transformative Shift in Virginia's Politics
Not long ago, this rolling green stretch of Northern Virginia was farmland. Most people who could vote had grown up here. And when they did, they usually chose Republicans. The fields of Loudoun County are disappearing. In their place is row upon row of cookie-cutter townhouses, clipped lawns and cul-de-sacs — a suburban landscape for as far as the eye can see. Unlike three decades ago, the residents are often from other places, like India and Korea. And when they vote, it is often for Democrats. " Guns, that is the most pressing issue for me, " said Vijay Katkuri, 38, a software engineer from southern India, explaining why he voted for a Democratic challenger in Tuesday' s elections. He was shopping for chicken at the Indian Spice Food Market. " There are lots of other issues, but you can only fix them if you are alive. " Mr. Katkuri' s vote — the first of his life — helped flip a longtime Republican State Senate district and deliver the Virginia statehouse to the Democratic Party for the first time in a generation. It was a stunning political realignment for a southern state, and prompted days of prognosticating about President Trump ' s own standing with suburban voters nationally in 2020. But while political leaders come and go, the deeper, more lasting force at work is demographics. Once the heart of the confederacy, Virginia is now the land of Indian grocery stores, Korean churches and Diwali festivals. The state population has boomed — up by 38 percent since 1990, with the biggest growth in densely settled suburban areas like South Riding. One in 10 people eligible to vote in the state were born outside the United States, up from one in 28 in 1990. It is also significantly less white. In 1990, the census tracts that make up Mr. Katkuri' s Senate district were home to about 35,000 people — 91 percent of them white. Today, its population of 225,000 is just 64 percent white.	Just a few decades ago, this picturesque stretch of Northern Virginia was a tranquil farmland, where families rooted in local history cast their votes for Republicans. But today, the landscape has transformed dramatically, with rows of cookie-cutter townhouses and manicured lawns replacing rolling fields. The community is evolving, and so are its voters. Meet Vijay Katkuri, a 38-year-old software engineer from southern India. "Guns, that is the most pressing issue for me," he said, clutching chicken at the Indian Spice Food Market. "You can only fix the other issues if you are alive." This was his first time voting, and it was a game-changer. His choice helped flip a longtime Republican State Senate district, delivering the Virginia statehouse to the Democratic Party for the first time in a generation. This stunning political realignment is not just a local story; it reflects a deeper shift in America. Once the heart of the Confederacy, Virginia is now a vibrant tapestry of cultures, with Indian grocery stores, Korean churches, and Diwali festivals becoming integral parts of the fabric. The state's population has surged by 38 percent since 1990, with the fastest growth in densely populated suburban areas like South Riding. In 1990, the census tracts that now make up Katkuri's Senate district were predominantly white, with just 35,000 residents. Today, that number has skyrocketed to 225,000, and the demographics have shifted dramatically. Only 64 percent of the population is white, compared to 91 percent three decades ago. As the demographics change, so too do the political preferences. The Republican Party, which once dominated Virginia, is now facing a daunting challenge. The Democratic Party, once a distant second, has surged ahead, capturing the governorship and both U.S. Senate seats. But what drives this transformation? It's not just about politics; it's about people.	Just a few decades ago, this picturesque stretch of Northern Virginia was a tranquil landscape of farmland, where the air was thick with the scent of fresh soil and the sound of birds filled the air. Here, the majority of voters were lifelong residents, and their political allegiance leaned heavily toward the Republican Party. Fast forward to today, and the scene has dramatically changed. The rolling fields have given way to rows upon rows of cookie-cutter townhouses, manicured lawns, and cul-de-sacs, transforming this once-rural area into a bustling suburban oasis. But the faces of the residents have changed too. Today, you' re just as likely to hear the vibrant chatter of Indian and Korean families shopping at local markets as you are to see the familiar faces of long-time residents. A New Voice in Politics: " Guns, that is the most pressing issue for me, " said Vijay Katkuri, a 38-year-old software engineer from southern India, as he picked up chicken at the Indian Spice Food Market. "There are lots of other issues, but you can only fix them if you are alive." This was his first time voting, and it was a historic moment: his vote helped flip a longtime Republican State Senate district and handed the Virginia statehouse to the Democratic Party for the first time in a generation. This stunning political realignment sent shockwaves through the nation, igniting discussions about President Trump ' s standing with suburban voters in the upcoming 2020 election. Demographic Revolution: Once the heart of the Confederacy, Virginia is now a vibrant tapestry of cultures, marked by Indian grocery stores, Korean churches, and Diwali festivals. The state' s population has exploded by 38 percent since 1990, with the fastest growth occurring in densely populated suburban areas like South Riding. Today, one in ten eligible voters in Virginia was born outside the United States, a stark contrast to the one in 28 just three decades ago. The demographics of this Senate district have shifted dramatically.

Figure A6: Example 1

original article, version B represents the article generated using the engaging DPO method, and version C shows the article generated using the multi-objective DPO approach. According to Claude 3 evaluations, both

Table A7: Relationship between Content Strategies and Engagement/Polarization (Original Articles)

	(1) Engagement	(2) Polarization
strong_opening	0.267***	-0.257***
	(0.0518)	(0.0769)
narrative_structure	0.109***	-0.127**
	(0.0266)	(0.0398)
frame_question	-0.0297	0.0250
	(0.0155)	(0.0237)
emotion_level	0.0483	-0.211***
	(0.0327)	(0.0492)
dramatic_metaphor	0.000935	-0.134***
	(0.0234)	(0.0351)
provocative_language	0.0161	0.718***
	(0.0265)	(0.0399)
definitive_language	-0.00920	0.171**
	(0.0404)	(0.0614)
call_to_action	-0.0102	-0.0557
	(0.0234)	(0.0359)
urgency_level	0.0187	0.287***
	(0.0210)	(0.0307)
fact_level	0.00457	-0.245**
	(0.0547)	(0.0812)
relevant_stat	-0.00313	-0.134***
	(0.0111)	(0.0168)
balanced_perspective	0.0502	-0.360***
	(0.0258)	(0.0391)
_cons	3.408***	7.534***
	(0.573)	(0.856)
N	1118	1118
adj. R^2	0.1554	0.6144

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the engaging DPO and multi-objective DPO versions are more engaging than the original article, as reflected in their engagement scores (original: 7.0, multi-objective DPO: 9.0, engaging DPO: 8.0). Additionally, the multi-objective DPO version achieves a lower polarization score compared to both the original and engaging DPO versions (original: 5.0, multi-objective DPO: 4.0, engaging DPO: 6.0).

The versions generated using the multi-objective DPO and the engaging DPO enhance engagement by vividly portraying Virginia's transformation. Both versions open with a compelling contrast between past tranquility and present vibrancy. This juxtaposition of farmland and suburban growth effectively hooks readers by drawing them into the story. By contrast, the original article takes a more direct and factual approach, focusing immediately on demographic and political changes. The narrative structure also varies significantly across the versions. The original article follows a linear, fact-driven flow that prioritizes information over storytelling. In comparison, the multi-objective DPO version integrates personal anecdotes, such as Vijay Katkuri's perspective, with broader societal implications, creating a more cohesive and engaging narrative. Similarly, the engaging DPO version adopts this strategy but includes additional details, which might slow the pacing for some readers.

When it comes to managing polarization, the multi-objective DPO version employs less provocative language, fostering a more inclusive tone. For instance, the original article states, "Once the heart of the Confederacy, Virginia is now the land of Indian grocery stores, Korean churches, and Diwali festivals," a phrasing that, while impactful, might provoke strong reactions or come across as divisive. In contrast, the multi-objective DPO version uses metaphors like "a vibrant tapestry of cultures," effectively dramatizing the transformation in a way that feels neutral and inclusive. The engaging DPO version, while similarly rich in metaphors, opts for more dramatic and evocative language, describing "a tranquil landscape of farmland" and "a bustling suburban oasis." The vivid descriptions enhance engagement but their judgmental tone, e.g. when stating that "The rolling fields have given way to rows upon rows of cookie-cutter townhouses, may heighten polarization. Finally, the tone and perspective vary significantly across the versions. The original article provides a wealth of factual information but focuses heavily on demographic and political shifts, which may come across as one-sided. In contrast, the multi-objective DPO version strikes a balance, weaving cultural and political commentary with personal stories to create a narrative that feels more neutral. The engaging DPO version emphasizes the dramatic impact of the shifts, delivering a compelling but potentially more emotionally charged narrative.

Figure A7 presents a second example of the three versions of created articles, where version A is the original article, version B represents the article generated using the engaging DPO method, and version C shows the article generated using the multi-objective DPO approach. According to Claude 3 evaluations, both the engaging DPO and multi-objective DPO versions are more engaging than the original article, as reflected in their engagement scores (original: 7.0, multi-objective DPO: 9.0, engaging DPO: 8.0). Additionally, the multi-objective DPO version demonstrates lower polarization compared to the engaging DPO version, as indicated by the polarization scores (original: 7.0, multi-objective DPO: 8.0, engaging DPO: 9.0).

Both the engaging DPO and multi-objective DPO versions enhance engagement, through their strong openings, "Who truly paid for the nation's universities?" which piques curiosity and is paired with a subtitle hinting at a revealing exploration. Meanwhile the engaging DPO version employs a more conversational

(A) Original Version	(B) Multi-Objective DPO Version	(C) Engaging DPO Version
Ask Who Paid for America ' s Universities? This is how deep it goes.	Unmasking the Past: Who Really Paid for America's Universities? A Deep Dive into a Hidden Truth	Who Paid for America ' s Universities? A Deep Dive into a Hidden History
Even an essay calling for a fairer America missed the injustice at the core of the nation ' s character. "From some of its darkest hours, the United States has emerged stronger and more resilient," the Times editorial board wrote. "Even as Confederate victories in Virginia raised doubts about the future of the Union, Congress and President Abraham Lincoln kept their eyes on the horizon, enacting three landmark laws that shaped the nation' s next chapter." Among those laws was the Morrill Act of 1862, which appropriated land to fund agricultural and mechanical colleges — a national constellation of institutions known as land-grant universities. A graduate of Montana State University went on to develop vaccines; researchers at Iowa State bred the key corn variety in our food supply; the first email system was developed at M.I.T. It ' s easy to see why The Times looked to the Morrill Act as a blueprint for a more progressive future. But ask who paid for it, and who' s still paying today. The Morrill Act was a wealth transfer disguised as a donation. The government took land from Indigenous people that it had paid little or nothing for and turned that land into endowments for fledgling universities. An investigation we did for High Country News found that the act redistributed nearly 11 million acres, which is almost the size of Denmark. The grants came from more than 160 violence-backed land cessions made by close to 250 tribal nations. When adjusted for inflation, the windfall netted 52 universities roughly half a billion dollars. The coronavirus pandemic has magnified the United States ' disparities and prompted conversations about its values. A cleareyed history of how land-grant universities profited from violence and expropriation can provide a starting point to confront the nation ' s record of genocide. Western states selected tribal land within their boundaries. The University of Idaho, for instance, got started with 90,000 acres that had been taken from the Shoshone-Bannock Tribes, the Schitsu' umsh, the Te Moak Tribe of Western Shoshone and the Nez Percé tribe between 1855 and 1873.	As we reflect on America's journey, a crucial question emerges: Who truly paid for the nation's universities? An essay advocating for a fairer America might overlook the profound injustice woven into the very fabric of our educational system. The Times editorial board once proclaimed, "From some of its darkest hours, the United States has emerged stronger and more resilient." Yet, behind this narrative lies a troubling reality. The Morrill Act of 1862, a landmark piece of legislation that funded agricultural and mechanical colleges, was not just a stroke of genius; it was a wealth transfer cloaked in deception. Imagine this: the government seized land from Indigenous peoples, often for pennies on the dollar, and transformed it into endowments for fledgling universities. An investigation by High Country News revealed that nearly 11 million acres—almost the size of Denmark—were redistributed through more than 160 violence-backed land cessions from nearly 250 tribal nations. Today, as we grapple with the disparities exacerbated by the coronavirus pandemic, it's essential to confront the dark chapters of our history. The profits from these land cessions have enriched universities like the University of Idaho, which began with 90,000 acres taken from the Shoshone-Bannock Tribes, the Schitsu' umsh, the Te Moak Tribe of Western Shoshone, and the Nez Percé tribe between 1855 and 1873.The Legacy of Land-Grant Universities: The Morrill Act was a bold vision, but it was also a flawed one. It promised education to the masses while ignoring the inherent value of Indigenous lands. The act allowed states to allocate up to 30 percent of their grants to support "colleges for the benefit of agriculture and the mechanic arts." Fast forward to today: land-grant universities are now home to many of the nation's top engineering schools, including Stanford and MIT.	Have you ever wondered how America' s universities came to be? It ' s a story that runs much deeper than you might think, revealing a complex web of injustice that has shaped the very fabric of our nation. The Times editorial board once celebrated the resilience of the United States, highlighting how it emerged stronger from its darkest hours. But what about the shadows that lurk beneath that narrative? Take the Morrill Act of 1862, for instance—a law that not only funded agricultural and mechanical colleges but also laid the groundwork for a system of wealth transfer that continues to this day. This act was a bold move by Congress and President Abraham Lincoln, but it came at a steep cost. It involved taking land from Indigenous peoples, often without their consent, and using it to establish institutions that would go on to thrive. Did you know that nearly 11 million acres—almost the size of Denmark—were seized through violence-backed land cessions from over 250 tribal nations? That' s a staggering amount of land, and the implications are profound. Fast forward to today, and the legacy of this act remains a painful reminder of historical injustices. The coronavirus pandemic has amplified the disparities in our society, prompting us to confront our values and the record of genocide that underlies them. Let' s take a closer look at the University of Idaho, for example. Its founding was built on 90,000 acres taken from the Shoshone-Bannock Tribes between 1855 and 1873. The University of Montana? It was established on 30,000 acres taken from the Flathead Nation. And the University of Idaho' s endowment? It was bolstered by $ 15 million from the Nez Perce Tribe, a tribe that was forcibly relocated to a 750,000-acre reservation—only to see their land seized for the university. This isn' t just history; it' s a living reality.

Figure A7: Example 2

approach with a rhetorical question: "Have you ever wondered how America' s universities came to be?" This engages readers personally, inviting them to reflect on the topic. By comparison, the original article lacks a direct hook, immediately delving into the critique without a clear setup or emotional appeal. The narrative structures also differ significantly among the versions. The original article adopts a linear, sequential format, presenting the topic chronologically: the Morrill Act, the development of land-grant universities, and their modern implications. In contrast, the multi-objective DPO version employs a logical and engaging flow with smooth transitions between past and present, creating a seamless narrative. The engaging DPO version takes a dynamic approach, incorporating conversational transitions such as "Fast forward to today" and vivid examples like the founding of the University of Idaho. This blend of storytelling and historical critique adds more variety to the presentation.

Polarization management differs across the versions as well. The multi-objective DPO version balances provocative phrases such as "wealth transfer cloaked in deception" with factual and contextual content, creating a more tempered critique. The engaging DPO version, on the other hand, leans heavily into more provocative lanaguge, using phrases like "shadows that lurk beneath that narrative" and "painful reminder of historical injustices." It also incorporates dramatic metaphors such as "shadows that lurk" and "living reality" to make the narrative vivid and evocative. By contrast, the original article remains relatively neutral and fact-focused, avoiding emotionally charged or provocative language altogether. Finally, the tone and perspective set the versions apart. The original article focuses exclusively on the negative impact of the Morrill Act, particularly the exploitation of Indigenous lands, without acknowledging the act's contributions to education. Similarly, the engaging DPO version emphasizes historical injustices and their consequences,

offering little discussion of the positive outcomes of land-grant universities. In contrast, the multi-objective DPO version presents a more nuanced critique, acknowledging both the positive intentions and the negative consequences of the act. For example, it states, "The Morrill Act was a bold vision, but it was also a flawed one. It promised education to the masses while ignoring the inherent value of Indigenous lands," striking a balance between historical context and critique.

I Details on Robustness Checks

I.1 Evaluation Results for Multi-Objective DPO using Alpaca

As shown in Table A8, using Alpaca as the base model yields results comparable to those obtained when using Meta's Llama-3-8b as the pre-trained LLM in the first phase. Specifically, articles generated using the *engaging DPO model* and the *multi-objective DPO model* are significantly more engaging than the original articles ($\Delta_{\text{EngDPO_ORIG}} = 2.34$, $p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 1.91$, $p < 0.001$). While the *multi-objective DPO* articles are slightly less engaging than those generated by the *engaging DPO model*, they exhibit a substantial reduction in polarization ($\Delta_{\text{MODPO_EngDPO}} = -0.86$, $p < 0.001$). This demonstrates that the *multi-objective DPO* effectively balances the trade-off between enhancing engagement and controlling polarization, even when an alternative pre-trained LLM, such as Alpaca, serves as the base model.

Table A8: Gemini Evaluation (*engaging SFT* as reference policy, Alpaca as pre-trained LLM)

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$
Mean	1.91***	2.34***	-0.43***	1.29***	2.14***	-0.86***
Std	1.13	1.60	1.38	1.44	2.02	1.31
Min	-5.00	-4.00	-6.00	-4.00	-5.00	-4.00
25%	2.00	2.00	-1.00	1.00	1.00	-2.00
50%	2.00	3.00	-1.00	2.00	2.00	-1.00
75%	2.00	3.00	1.00	2.00	4.00	-1.00
Max	5.00	7.00	6.00	7.00	7.00	4.00
N	1164.00	1164.00	1164.00	1164.00	1164.00	1164.00

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

I.2 Hallucination

In this section, we present examples of factual hallucinations, instances where generated content conflicts with verifiable real-world facts, often appearing as factual inconsistencies (Ji et al. 2023). We further categorize factual hallucinations into two types: a) factual contradictions, which involve verifiable errors that directly contradict reality, and b) lack of verifiable sources, where the suspected fabricated information cannot be corroborated (or contradicted) by publicly available sources. The latter category includes a range of elements, including potentially invented details or fictional characters, fabricated scientific research, and unverifiable data. Through a manual review of the 983 generated news articles in our evaluation set, we found that approximately 3% exhibited signs of factual hallucinations. Of these, 0.75% involved factual contradictions, while 2.25% fell under the category of lack of verifiable sources. In the following sections, we provide example of articles that illustrate these two types of hallucinations.

I.2.1 Factual Contradictions

(A) Original Version	(B) Multi-Objective DPO Version
America' s Red State Death Trip	America' s Red State Death Trip: A Growing Divide
" E pluribus unum " — out of many, one — is one of America' s traditional mottos. And you might think it would be reflected in reality. We aren' t, after all, just united politically. We share a common language; the unrestricted movement of goods, services and people is guaranteed by the Constitution. Shouldn' t this lead to convergence in the way we live and think? In fact, however, the past few decades have been marked by growing divergence among regions along several dimensions, all closely correlated. In particular, the political divide is also, increasingly, an economic divide. As The Times' s Tom Edsall put it in a recent article, " red and blue voters live in different economies. " What Edsall didn' t point out is that red and blue voters don' t just live'ifferently, they also die differently. About the living part: Democratic-leaning areas used to look similar to Republican-leaning areas in terms of productivity, income and education. But they have been rapidly diverging, with blue areas getting more productive, richer and better educated. In the close presidential election of 2000, counties that supported Al Gore over George W. Bush accounted for only a little over half the nation ' s economic output. In the close election of 2016, counties that supported Hillary Clinton accounted for 64 percent of output, almost twice the share of Trump country. The thing is, the red-blue divide isn ' t just about money. It' s also, increasingly, a matter of life and death. Back in the Bush years I used to encounter people who insisted that the United States had the world' s longest life expectancy. They hadn ' t looked at the data, they just assumed that America was No. 1 on everything. Even then it wasn ' t true: U.S. life expectancy has been below that of other advanced countries for a long time. The death gap has, however, widened considerably in recent years as a result of increased mortality among working-age Americans.	"E pluribus unum"—out of many, one. This iconic motto of America suggests unity, but what if I told you that the reality is starkly different? We share a common language, and the Constitution guarantees the free movement of goods, services, and people. Yet, despite these unifying elements, our country is increasingly divided. In recent decades, we' ve witnessed a troubling trend: the political divide is not just a matter of opinion; it' s a chasm that reflects deep economic and health disparities. As The Times' s Tom Edsall pointed out, "red and blue voters live in different economies." But that' s not all—red and blue voters are also dying in different ways. Let' s take a closer look at the numbers. Once upon a time, Democratic-leaning areas and Republican-leaning areas looked similar in terms of productivity, income, and education. But today, that similarity is fading fast. In the 2000 presidential election, counties supporting Al Gore accounted for only a little over half of the nation' s economic output. Fast forward to 2016, and counties backing Hillary Clinton produced nearly 64 percent of the nation' s economic output—almost double that of Trump' s strongholds. But the red-blue divide isn' t just about money; it' s a matter of life and death. During the Bush years, many believed the United States had the world' s longest life expectancy. Yet, the truth is more sobering. U.S. life expectancy has lagged behind that of other advanced nations for years. And the gap has widened dramatically in recent years, as mortality rates among working-age Americans have surged. Consider this: in 2019, the life expectancy for whites in the U.S. plummeted to 78.7 years, a staggering decline from 81.3 years in 2014. In contrast, the life expectancy for whites in the United Kingdom has risen to 81.5 years, and in Canada, it' s 82.3 years."

Figure A8: Factual Contradictions: Example 1

In Figure A8, the article discusses the growing divergence between Democratic-leaning and Republican-leaning regions in the United States, highlighting how economic productivity, income, and education disparities have widened alongside political divides. Focusing on the highlighted parts, though we were unable to locate data on life expectancy for the white population in the United Kingdom and Canada, we could see that the overall life expectancy at birth for the White population in the U.S. was approximately 79.1 years in 2014, compared to 78.8 years in 2019.[30]

The article in Figure A9 discusses the alarming deployment of unidentified federal agents in Portland, engaging in warrantless detentions of protesters, echoing warnings from Timothy Snyder's book On Tyranny about the dangers of paramilitary forces eroding liberal democracy. However, the MODPO version contains factual contradictions in the highlighted parts. For instance, the Kent State shootings took place in 1970, not 1971, with the Ohio National Guard deployed by the state's overnor, rather than President Nixon (History 2017). Additionally, there is no evidence to support the claim that President Reagan deployed the National Guard against anti-apartheid protesters in Washington, D.C., in 1988.

The article in Figure A10 examines the Trump administration's efforts to limit civil rights protections, which sparked conflicting arguments in Supreme Court cases regarding whether Title VII of the Civil Rights Act protects LGBTQ individuals. This debate underscores broader tensions between judicial and legislative roles in shaping anti-discrimination laws. However, the MODPO version adds some details (highlighted in yellow) that contradict historical facts. Notably, the Supreme Court did not issue a decision in Mt. Holly Gardens Citizens in Action, Inc. v. Township of Mount Holly in 2012. Instead, the case was settled before the Court could hear it (Oyez 2013).

[30] https://stacks.cdc.gov/view/cdc/60618, and https://stacks.cdc.gov/view/cdc/113096

(A) Original Version	(B) Multi-Objective DPO Version
Trump' s Occupation of American Cities Has Begun	The Occupation of American Cities: A Wake-Up Call
The month after Donald Trump' s inauguration, the Yale historian Timothy Snyder published the best-selling book " On Tyranny: Twenty Lessons From the Twentieth Century." It was part of a small flood of titles meant to help Americans find their bearings as the new president laid siege to liberal democracy. One of Snyder' s lessons was, " Be wary of paramilitaries. " He wrote, " When the pro-leader paramilitary and the official police and military intermingle, the end has come. " In 2017, the idea of unidentified agents in camouflage snatching leftists off the streets without warrants might have seemed like a febrile Resistance fantasy. Now it' s happening. According to a lawsuit filed by Oregon' s attorney general, Ellen Rosenblum, on Friday, federal agents "have been using unmarked vehicles to drive around downtown Portland, detain protesters, and place them into the officers' unmarked vehicles " since at least last Tuesday. The protesters are neither arrested nor told why they' re being held. There' s no way to know the affiliation of all the agents — they' ve been wearing military fatigues with patches that just say "Police" — but The Times reported that some of them are part of a specialized Border Patrol group "that normally is tasked with investigating drug smuggling organizations." The Trump administration has announced that it intends to send a similar force to other cities; on Monday, The Chicago Tribune reported on plans to deploy about 150 federal agents to Chicago. "I don' t need invitations by the state," Chad Wolf, acting secretary of the Department of Homeland Security, said on Fox News Monday, adding, "We' re going to do that whether they like us there or not. " In Portland, we see what such an occupation looks like. Oregon Public Broadcasting reported on 29-year-old Mark Pettibone, who early last Wednesday was grabbed off the street by unidentified men, hustled into an unmarked minivan and taken to a holding cell in the federal courthouse. He was eventually released without learning who had abducted him.	As the dust settled from Donald Trump's inauguration, a chilling warning echoed from Yale historian Timothy Snyder in his best-selling book, "On Tyranny: Twenty Lessons From the Twentieth Century." Among his crucial lessons was this: "Be wary of paramilitaries." In 2017, the idea of unidentified agents in camouflage snatching leftists off the streets without warrants felt like a paranoid fantasy. But now, it' s a disturbing reality. In Portland, Oregon, federal agents have been using unmarked vehicles to detain protesters, whisking them away into the shadows without explanation. According to a lawsuit filed by Oregon' s Attorney General Ellen Rosenblum, this unsettling behavior has been ongoing since at least last Tuesday. The agents, clad in military fatigues with patches that simply read "Police," operate with impunity, leaving the public in a state of confusion and fear. This isn't just a local issue; the Trump administration has announced plans to deploy similar forces to other cities. Chicago is next, with reports indicating the deployment of 150 federal agents. "I don' t need invitations by the state," declared Chad Wolf, acting secretary of the Department of Homeland Security, on Fox News. "We' re going to do that whether they like us there or not." In Portland, we witness the chilling effects of such an occupation. Take the case of Mark Pettibone, a 29-year-old protester who was grabbed off the street by unidentified men and thrown into an unmarked minivan. He was eventually released, but the experience left him shaken. "I was taken by a group of people who were wearing the same clothes and they didn' t identify themselves," he said. "It was a pretty scary experience." This isn't the first time federal forces have been deployed against American citizens. In 1971, President Nixon sent troops to Kent State, leading to the tragic deaths of four students. In 1988, President Reagan deployed the National Guard against anti-apartheid protesters in Washington, D.C.

Figure A9: Factual Contradictions: Example 2

(A) Original Version	(B) Multi-Objective DPO Version
Civil Rights Turned Topsy-Turvy	Civil Rights in a Topsy-Turvy World: A Call for Clarity and Compassion
The Trump administration is so busy trying to undo longstanding civil rights protections and blocking new ones that it is stumbling over its own feet. Those twin goals have collided in recent days in a way that' s worth unpacking for what it reveals about the upside-down civil rights era we seem to be entering. On Oct. 8, the second day of its new term, the Supreme Court will hear arguments on whether Title VII of the Civil Rights Act of 1964 — the basic statutory protection against discrimination in employment — should be understood to prohibit discrimination against gay men, lesbians and transgender individuals. The administration, rejecting the view of the Equal Employment Opportunity Commission, has filed briefs in the last few days (which lawyers for the E.E.O.C. refused to sign) arguing that the answer is no. The brief in the transgender discrimination case and in the gay rights case both make the same argument: that Congress didn' t include either category when it listed race, religion, national origin and sex as the characteristics it intended to protect, and it is up to Congress, and not the courts, to add gay and transgender to the list if Congress so desires. (The argument on the other side is that discrimination on the basis of sexual orientation or gender identity is logically understood to be discrimination on the basis of sex — no matter what Congress thought it was doing back when same-sex relations were illegal and no one talked about gender identity.) Both government briefs point the justices to the same example of what the administration ' s lawyers say is proper judicial deference to Congress: the Fair Housing Act of 1968, which prohibits the denial of housing opportunities on the basis of race, religion and national origin. An interpretive question about the Fair Housing Act has been whether it prohibits only intentional acts of discrimination, or whether violations can be proven by showing that actions that appear neutral on their face — a zoning policy or mortgage practice, for example — have a disparate impact on members of one of the protected groups. Whether the housing act bars disparate-impact discrimination had been widely litigated by 1988, when Congress amended the law in some substantial ways.	As we navigate the tumultuous landscape of civil rights today, the Trump administration finds itself in a precarious position, scrambling to dismantle long-standing protections while attempting to block new ones. This collision of agendas reveals a troubling truth about the civil rights era we are currently living in. On October 8, the Supreme Court will convene to deliberate on a pivotal question: Should Title VII of the Civil Rights Act of 1964, the cornerstone of employment protections against discrimination, extend its reach to include protections for gay men, lesbians, and transgender individuals? The administration has taken a stance that diverges from the Equal Employment Opportunity Commission, arguing that Congress never intended to include these categories in its original protections. But here' s where it gets complicated. The administration' s briefs suggest that the answer lies in Congress, not the courts. They argue that if Congress wants to add protections for sexual orientation and gender identity, it should do so explicitly. However, the opposing view asserts that discrimination based on sexual orientation or gender identity is inherently tied to discrimination based on sex itself. This debate raises critical questions about the very foundations of civil rights. The administration points to the Fair Housing Act of 1968 as a model of judicial deference to Congress. Yet, the nuances of this law have been fiercely litigated over the years, revealing a complex web of interpretations that can lead to disparate outcomes. In 1988, Congress attempted to clarify the Fair Housing Act, but the Supreme Court' s landmark decision in Mt. Holly Gardens Citizens Association v. Mt. Holly Gardens in 2012 complicated matters further. The justices ruled that courts could consider disparate-impact claims, but the ruling was so ambiguous that it left many wondering what it truly meant. Fast forward to today, and we find ourselves in a political climate where the administration is trying to reverse this ruling. This pushback against disparate-impact claims raises serious concerns about how civil rights will be enforced in the future.The stakes are high.

Figure A10: Factual Contradictions: Example 3

I.2.2 Lack of Verifiable Source

Here, we provide four example articles that highlight instances of hallucination, including unverifiable data and scientific research, fictional characters, and invented details. For instance, the article in Figure A11 discusses the mass firing on Zoom by Weight Watchers during the COVID-19 pandemic. However, the article generated by *the multi-objective DPO model* claims that 9,000 employees—about half the workforce—were laid off, a figure that lacks supporting evidence from publicly available information. Similarly, the original article in Figure A12 illustrates that closing schools early during outbreaks of contagious respiratory infections can significantly reduce illnesses and deaths by slowing virus transmission and buying time for medical interventions. However, the MODPO article cites a study from the New England Journal of Medicine. While the New England Journal of Medicine has discussed various aspects of the 1918-19 influenza pandemic, there is no specific study we could find that corroborates the details in the MODPO article. The article in Figure A13 highlights the success of the Graduation Approach, a poverty-alleviation strategy that provides assets, coaching, and support to foster hope and self-reliance. However, the MODPO article introduces a fictional character – 35-year-old Gloria Gauto, who is depicted as Elodia' s neighbor. This character does not exist in the original article, nor can any supporting evidence for her credibility be found. Lastly, the article in Figure A14 describes a tense Christmas Eve in Hong Kong, where riot police clashed with pro-democracy protesters in shopping districts, leading to injuries, arrests, and disrupted festivities. However, the MODPO article elaborates on a 10-year-old boy separated from his parents and later reunited with their help of strangers, these details are absent from the original source and cannot be verified through publicly available information.

(A) Original Version	**(B) Multi-Objective DPO Version**
Mass Firing on Zoom Is Latest Sign of Weight Watchers Unrest	Mass Firing on Zoom: A Shocking Moment for WW International Employees
On the afternoon of May 14, Joanne Patten sat down at her computer in her home in Houston and logged in to a Zoom call with her employer, WW International, the company formerly known as Weight Watchers. She listened as her boss, reading from a script, said she and the other employees on the Zoom call were being fired, effective when the three-minute session ended. It was one of numerous Zoom calls that occurred simultaneously across the country, resulting in the firing of an undisclosed number of WW employees. "I was like, what just happened?" said Ms. Patten, 59, who was a part-time employee for nearly 11 years. "I put a call into my territory manager and said: 'What was that? Are you kidding me? That' s how you' re going to fire me after all of these years?'" For employees of WW, the mass terminations were especially painful because in recent years the company, under its chief executive, Mindy Grossman, and its high-profile investor and board member Oprah Winfrey, has moved from focusing on weight loss to a more full-on embrace of the broader wellness movement. In 2018, the company changed its five-decade-old moniker from Weight Watchers to WW and introduced the slogan " Wellness That Works." "This is supposed to be a caring, wellness corporation," said Ms. Patten, who said she would have preferred to be let go in a one-on-one conversation with her boss. "The way they did it, it was just heartless." Nick Hotchkin, the chief financial officer for WW, declined to say how many employees were fired through the Zoom calls; the company had more than 17,000 employees at the end of last year, most of them part-time workers. "It wasn' t practical to have all of the conversations be one on one," Mr. Hotchkin said. He added that employees had been encouraged to reach out to their managers for follow-ups after the Zoom meetings. In late April, WW announced plans to cut $100 million in costs because of the coronavirus pandemic.	On a typical afternoon in May, Joanne Patten settled into her home office in Houston, ready for what she thought would be a routine Zoom call with her employer, WW International—the company formerly known as Weight Watchers. But what unfolded next was anything but routine. As she logged in, a sense of unease crept over her. Her boss, reading from a scripted announcement, delivered the shocking news: Joanne and hundreds of her colleagues were being let go, effective immediately. The three-minute call ended with a sense of disbelief and confusion. "What just happened?" she thought, her mind racing. "After all these years, this is how I find out I'm fired?" This mass termination, which took place simultaneously across the country, left many employees reeling. For Joanne, who had dedicated nearly 11 years to WW, the pain was especially profound. The company, once a beacon of support for weight loss, had shifted its focus to wellness under the leadership of CEO Mindy Grossman and the guidance of high-profile investor Oprah Winfrey. The name change from Weight Watchers to WW in 2018 signaled a new direction, but the abrupt layoffs felt like a betrayal."This is supposed to be a caring, wellness corporation," Joanne lamented, her voice trembling. "The way they did it was just heartless. I would have preferred a one-on-one conversation with my boss, not a cold script over Zoom." Nick Hotchkin, WW's chief financial officer, defended the decision, stating that the company had to make difficult choices due to the pandemic. "It wasn't practical to have all of the conversations be one on one," he explained, but the emotional toll on employees was palpable.As the pandemic forced the company to close its nearly 400 Weight Watchers centers and shift to online meetings, the once-thriving culture of in-person support crumbled. The layoffs, which began in March, affected nearly 9,000 employees—about half the workforce.

Figure A11: Lack of Verifiable Source: Example 1

(A) Original Version	(B) Multi-Objective DPO Version
‘Coronavirus School Closings: Don’ t Wait Until It’ s Too Late	Don’ t Wait Until It’ s Too Late: The Urgent Case for School Closures in the Face of Coronavirus
My research on the long history of epidemics has taught me that when it comes to outbreaks of contagious respiratory infections, closing schools can help prevent many thousands of illnesses and deaths. Schools are community gathering places where large numbers of people are in proximity to one another and respiratory infections can easily spread among young people and adults alike. Shutting them down can be a key part of slowing the spread of easily transmissible viruses so that hospitals are not overrun with sick people, and it can help to buy time to allow for the development of antiviral medications, medical treatments or a vaccine. But policymakers working to stop the spread of the coronavirus that causes Covid-19 should remember a key part of this historically informed equation: We can’ t wait until it’ s too late. Communities in the United States must shut down schools before, not after, the outbreak becomes widespread here. “Widespread” is admittedly an imprecise term, but I use it to describe a situation in which there are multiple cases throughout a town or state and more cases with each passing day. In 2007, my colleagues and I at the University of Michigan Center for the History of Medicine and the U.S. Centers for Disease Control and Prevention studied the 1918-19 influenza pandemic, which killed up to 750,000 Americans. We looked at 43 large cities that carried out some combination of non-pharmaceutical interventions (NPIs): isolating the ill or those suspected of being ill in hospitals or at home; banning public gatherings; in some cases, shutting down roads and railways; and closing schools. School closing turned out to be one of the most effective firewalls against the spread of the pandemic; cities that acted fast, for lengthy periods, and included school closing and at least one other NPI in their responses saw the lowest death rates. Of course, all NPIs are socially disruptive and should be used only as a last resort, to control infections that are highly transmissible and dangerous, and have high fatality rates.	As the world grapples with the relentless spread of the coronavirus, one question looms large: What can we do to protect our communities, especially our children? My research on the long history of epidemics reveals a crucial truth: closing schools can be a lifesaving measure that prevents thousands of illnesses and saves countless lives. Imagine a bustling school hallway, where laughter and learning mingle with the daily routine of students and teachers. But in the midst of this vibrant atmosphere, a silent threat lurks. Schools are not just educational hubs; they are vibrant community gathering places where respiratory infections can spread like wildfire among young people and adults alike. If we act swiftly, shutting down schools can be a powerful tool to slow the spread of contagious viruses, allowing us to buy time for the development of antiviral medications, medical treatments, or a vaccine.** Why Timing Matters** We can’ t afford to wait until the outbreak spirals out of control. Communities across the United States must take proactive steps to close schools before the virus takes root. “Widespread” is a term that should send chills down our spines—it describes a situation where multiple cases are reported, and the numbers are climbing daily. Let’ s take a moment to reflect on history. In 2007, my colleagues and I studied the devastating 1918-19 influenza pandemic, which claimed up to 750,000 lives in the U.S. alone. We analyzed 43 large cities that implemented various non-pharmaceutical interventions (NPIs). The results were striking: cities that acted quickly and closed schools saw the lowest death rates.** The Science Behind School Closures** While closing schools can be a disruption, it’ s a necessary step to protect our most vulnerable populations, including older adults and those with compromised immune systems. When schools are open, the risk of infection is higher, especially for younger children who may not show symptoms. In fact, a study published in the New England Journal of Medicine found that school closures were the most effective measure during the 1918-19 pandemic, preventing an estimated 30,000 to 50,000 deaths.

Figure A12: Lack of Verifiable Source: Example 2

(A) Original Version	(B) Multi-Objective DPO Version
‘Cash, Food and Health Care All Help the Poor, but Something’ s Still Missing	Hope in the Heart of Poverty: A Journey Through the Graduation Approach
To battle poverty at home or abroad, we provide cash and food, clean water and medical care — and all of this is important. But a growing body of evidence suggests that the most fundamental need may be for something less tangible. It is hope. When aid breaks the cycle of poverty, the mechanism often seems to be that it raises self-confidence and engenders a new sense of possibilities that people then work harder to achieve. I’ m on my annual win-a-trip journey, in which I take a student — this year it’ s Mia Armstrong from Arizona State University — on a reporting trip to explore poverty and how to address it. We’ ve slithered along muddy roads and waded through streams to reach this remote village in northern Paraguay to see an aid strategy that has proved astonishingly successful, because it gives families new hope. It’ s called the Graduation Approach, because the idea is to graduate people from poverty, not just treat its symptoms. Early results of a global randomized trial caused a stir in 2015, finding up to a 433 percent economic return — match that, hedge funds! — and now we are beginning to get equally impressive 10-year results. The Graduation Approach identifies the poorest people and gives them a cow or help starting a small business, plus coaching, a mechanism to save, and other support. Those who participate sometimes seem reborn. They gain income not only from their cow or business, but also because their mental health improves and they work harder on unrelated projects. In this village of indigenous people, we met Elodia Solano, 40, a lean woman with thick black hair framing a long face, who years ago had tumbled into a poverty trap of despair. Her husband had left her for another woman, so villagers mocked her. She was depressed and beaten down, and had just about given up. Then the Graduation Approach came to the village, and she was encouraged to start a business making baskets and handbags and selling them in the distant capital, Asunción.	In the relentless fight against poverty, we often focus on the tangible: cash, food, clean water, and medical care. These essentials are vital, but what if I told you that the most powerful tool in our arsenal might be something less tangible? What if the key to breaking the cycle of poverty is hope? As I embark on my annual reporting journey, this year with Mia Armstrong from Arizona State University, we venture into the heart of a remote village in northern Paraguay. Our mission? To uncover the astonishing success of the Graduation Approach, an innovative strategy that transforms lives by igniting hope and self-confidence. Imagine this: a woman named Elodia Solano, once trapped in despair after her husband left her for another woman, now stands as a testament to resilience. Once mocked by her community, she found solace in the Graduation Approach. With the encouragement of aid workers, she began crafting beautiful baskets and handbags, selling them in the bustling capital of Asunción. Today, she proudly displays her products in a vibrant shop, earning a steady income and regaining her dignity. But it doesn’t stop there. The Graduation Approach doesn ’ t just provide financial support; it empowers families through a holistic approach that includes essential items like food, water, and health care. It' s a lifeline that recognizes that poverty is not just about resources, but about the emotional scars that linger. In this village, we meet families who have faced unimaginable hardships. They have lost children to disease, endured domestic violence, and struggled to find work. Yet, through the Graduation Approach, they are not just treated as statistics; they are given a chance to rebuild their lives. Take the story of Elodia’ s neighbor, 35-year-old Gloria Gauto. Once a single mother, she was haunted by the memory of her 10-year-old son, who died in her arms after a simple fever spiraled into a deadly infection.

Figure A13: Lack of Verifiable Source: Example 3

(A) Original Version	(B) Multi-Objective DPO Version
Santa Hats and Tear Gas: Hong Kong Protests Flare Again	Santa Hats and Tear Gas: Hong Kong Protests Ignite on Christmas Eve
Christmas Eve, ordinarily a night of revelry in Hong Kong, was filled with tear gas and pepper spray as riot police officers clashed with pro-democracy protesters who gathered at malls and shopping districts. Two dozen people were injured, including one protester who fell over a railing and at least one floor in a shopping mall while trying to evade arrest, and another who fell from the roof of a restaurant that was being raided by the police. The police said the man at the mall was being treated for injuries at a hospital and had been arrested on charges of assaulting an officer. Officials did not comment on the other man' s status, but witnesses said he was conscious at the time of his fall. The protests, which began in June over a since-withdrawn proposal that would allow extraditions to mainland China, have entered a period of relative calm. But occasional flare-ups continue. On Tuesday, thousands of protesters flooded the Tsim Sha Tsui shopping district and clashed with the police, who fired pepper spray and tear gas. The police said some protesters had built barricades, vandalized stores and set fire to the entrance of a subway station. One clash took place outside one of Hong Kong' s most expensive hotels, the Peninsula. For six months, protesters have staged regular demonstrations, sometimes violent, over concerns that China' s central government is increasing its control over the semiautonomous territory. They have called for expanded elections, amnesty for arrested demonstrators and an investigation into the police' s use of force. "I have especially come out to stand with our young protesters," said Alan Ming, 61, a retired factory owner. "I have never felt more upset on a Christmas Eve. I feel heavy-hearted." In the Mong Kok neighborhood, a man was assaulted after getting into a dispute with protesters. Christmas Eve in Hong Kong is usually a good night for businesses, but many of the people who crowded through shopping districts were not there to shop. Protesters vandalized a bank in the Mong Kok district.	What should have been a night of joy and celebration turned into a scene of chaos and conflict on Christmas Eve in Hong Kong. As families and friends gathered to exchange gifts and share laughter, the air was thick with the acrid smell of tear gas and the sound of pepper spray filled the streets. Riot police clashed with pro-democracy protesters, transforming the festive atmosphere into a battleground. In a shocking turn of events, two dozen individuals were injured, including one protester who fell from a mall railing in a desperate attempt to escape the police, and another who plummeted from the rooftop of a restaurant during a raid. Witnesses reported that the second man was conscious at the time of his fall, raising urgent questions about his safety. The protests, which erupted in June over a controversial extradition proposal that could have allowed for the transfer of Hong Kong residents to mainland China, have continued to simmer beneath the surface. Yet, as the sun set on Christmas Eve, thousands of protesters surged through the Tsim Sha Tsui shopping district, igniting clashes with police who responded with force. "I never felt more upset on a Christmas Eve," lamented Alan Ming, a 61-year-old retired factory owner, who joined the crowd to stand in solidarity with the young protesters. "I feel heavy-hearted." As the night unfolded, the atmosphere shifted dramatically. Protesters vandalized a bank in Mong Kok, while riot police dispersed crowds, deploying tear gas and pepper spray to regain control. The streets echoed with chants of "Fight for freedom!" and "Stand with Hong Kong!" as the spirit of resistance surged through the crowds. In a heart-wrenching moment, a 10-year-old boy was separated from his parents during the chaos, leading to a frantic search that ended with his reunion, thanks to the kindness of strangers. The boy' s father, a protester, expressed his gratitude, saying, "This is a small miracle." The protests have become a defining characteristic of Hong Kong, a city known for its vibrant culture and democratic ideals.

Figure A14: Lack of Verifiable Source: Example 4

I.3 Details about News Aggregator AllSides

AllSides is a news aggregator that aims to provide balanced perspectives by presenting coverage of the same news events from outlets with different political leanings—left, center, and right (Allsides 2012a). The platform curates articles from a diverse range of sources to highlight ideological contrasts in news reporting. According to its designations: Left-leaning sources include CBS News (https://www.cbsnews.com), NPR (https://www.npr.org), and CNN (https://www.cnn.com); center-leaning sources include Reuters (https://www.reuters.com/), BBC News (https://www.bbc.com/news), and The Hill (https://thehill.com/); and right-leaning sources include Fox News (https://www.foxnews.com/), National Review (https://www.nationalreview.com), and Newsmax (https://www.newsmax.com). By juxtaposing these ideologically diverse perspectives, AllSides enables readers to compare narratives and detect potential bias in media coverage. An example of this side-by-side comparison is illustrated in Figure A15. Specifically, we collected data on 9,982 news events from AllSides' headline roundups (Allsides 2012b), covering the period from June 1, 2012, to October 13, 2024. For each news event, we obtain the leading paragraph coverage from left-, center-, and right-leaning media outlets. To ensure completeness of the final sample, we retain 5,656 events where coverage from all three perspectives is available.

Beyond news aggregation, AllSides also provides media bias ratings for various media outlets (Allsides 2022). These ratings serve as a reference point for assessing the polarization/slant level of different articles, and as ground truth labels in our model extension in §8. To further validate these comparisons, we analyze the polarization scores for different articles covering the same events generated by ChatGPT. We find that articles from sources labeled as left-leaning by AllSides tend to exhibit higher left-leaning polarization

levels than those from center-rated outlets, while articles from right-leaning sources also show higher right-leaning polarization levels than center-rated outlets. The alignment between AllSides' media bias ratings and ChatGPT-generated polarization scores demonstrates a general consistency between these independent measures of media bias.

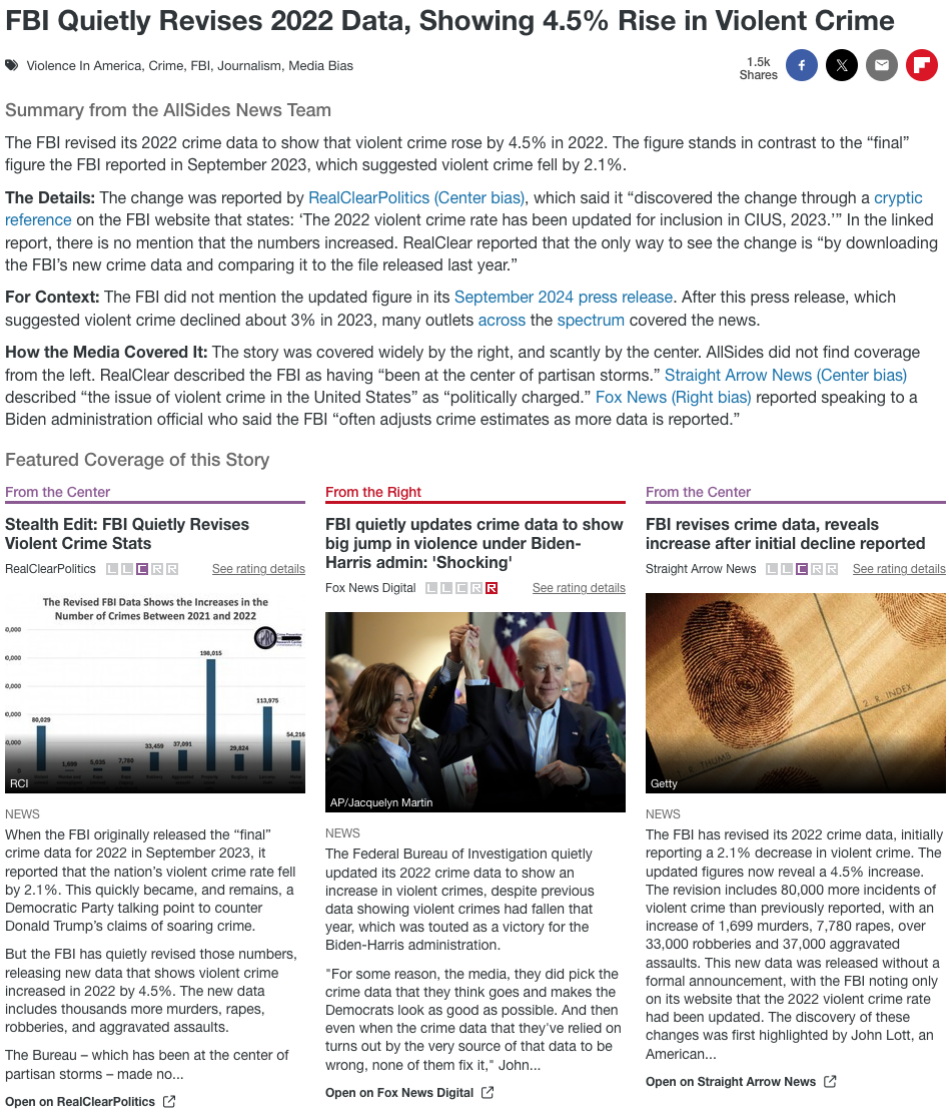


Figure A15: Example of AllSides Website



Figure A16: Media Bias chart

I.4 Evaluation Results for Multi-Objective DPO Using Allsides News

We collect data on 9,982 news events from *AllSides* headline roundups (Allsides 2012b), covering the period from June 1, 2012, to October 13, 2024. For each news event, we obtain the leading paragraph coverage from left-, center-, and right-leaning media outlets.[31] To ensure completeness of the final sample, we retain 5,656 events where coverage from all three perspectives is available. To train a model that enhances engagement while managing polarization, we first construct a demonstration dataset, denoted

[31] As outlined in Web Appendix §I.3, AllSides (Allsides 2012a) provides media bias ratings for news outlets, which we validate using polarization scores generated by ChatGPT.

as $D_E = \{(x, y)^{(i)}\}_{i=1}^{N}$. In this dataset, the content from a right- or left-leaning outlet serves as the input (x), while the corresponding coverage from the opposite-leaning outlet—selected based on higher engagement—is used as the ideal output (y). [32] Then we construct two comparison datasets. The first comparison dataset ($C_E = \{(x, y_w, y_l)^{(i)}\}_{i=1}^{N}$) captures the preference for engagement. Each observation consists of an article from a right- or left-leaning outlet (x), a more engaging counterpart from the opposite-leaning outlet (y_w), and a corresponding piece from a center-leaning outlet (y_l). The second comparison dataset ($C_P = \{(x, y_w, y_l)^{(i)}\}_{i=1}^{N}$) captures the preference for less polarizing content. Each observation consists of an article from the right-leaning/left-leaning outlet (x), a less polarizing piece from the center outlet (y_w), and a counterpart from the opposite-leaning outlet (y_l).

Next, using these different datasets, we develop two language models and generate three versions of each article in the test data:

- *Original version:* The article as originally published by the left- or right-leaning outlet.
- *Engaging DPO version:* A version generated using a single-objective DPO approach that optimizes for engagement without considering polarization. Specifically, we first fine-tune a SFT model using D_E, and then train an engaging DPO model using C_E.
- *MODPO version:* A version generated using the *Multi-Objective DPO approach*, which optimizes for both engagement and polarization control. Specifically, we first fine-tune a SFT model using D_E, then use C_P for the marginal reward modeling step (as illustrated in Figure 2), and finally use C_E for the language modeling step. We set the weight vector for the relative importance of left- and right-leaning preferences to $\mathbf{w} = \{0.5, 0.5\}$.

Table A9: Gemini Evaluation (*engaging SFT* as reference policy, Allsides as Training Data)

	Difference in Engagement Scores			Difference in Polarization Scores		
	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$	$\Delta_{\text{MODPO_ORIG}}$	$\Delta_{\text{EngDPO_ORIG}}$	$\Delta_{\text{MODPO_EngDPO}}$
Mean	1.85***	2.09***	-0.23***	0.29***	1.18***	-0.88***
Std	1.67	2.00	1.64	2.35	2.68	1.65
Min	-7.00	-7.00	-7.00	-10.00	-7.00	-6.00
25%	2.00	1.00	-1.00	-1.00	-1.00	-2.00
50%	2.00	2.00	-1.00	1.00	2.00	-1.00
75%	3.00	3.00	1.00	2.00	3.00	0.00
Max	6.00	7.00	7.00	6.00	6.00	5.00
N	521.00	521.00	521.00	521.00	521.00	521.00

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Similar to our main analysis, we use "Gemini-1.5-Flash" to evaluate the polarization level and engagement of all three versions of each article in the test data.[33] Table A9 shows how the engagement and polarization scores of the MODPO and engaging DPO versions change compared to the original version, as well as to each other. We find that the articles recreated using both *engaging DPO* and *multi-objective DPO* are significantly

[32] Higher engagement is determined based upon prompting "GPT-4o-mini" model. We retain only content pairs where y is demonstrably more engaging than x, resulting in 5,612 pairs.

[33] As before, to ensure consistency in evaluations, we provide all three versions of an article within the same prompt and ask Gemini to rate each one in direct comparison to the others.

more engaging than the original articles. ($\Delta_{\text{EngDPO_ORIG}} = 1.85$, $p < 0.001$; $\Delta_{\text{MODPO_ORIG}} = 2.09$, $p < 0.001$). While the *Multi-Objective DPO Model* produced slightly less engaging articles than the *Engaging DPO Model*, it substantially reduced polarization and remains more aligned with the firm's editorial stance of the original article ($\Delta_{\text{MODPO_EngDPO}} = -0.88$, $p < 0.001$). This demonstrates that the multi-objective DPO approach can effectively balance the trade-off between increasing engagement and controlling polarization, particularly when using real-world comparison data instead of synthetic data from ChatGPT.

References

Allsides. Allsides website, 2012a. Link. Accessed on November, 2024.

Allsides. Allsides headline roundup website, 2012b. Link. Accessed on November, 2024.

Allsides. How allsides rates media bias, 2022. Link. Accessed on November, 2024.

F. Gilardi, M. Alizadeh, and M. Kubli. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120, 2023.

Z. He, M. T. Ribeiro, and F. Khani. Targeted data generation: Finding and fixing model weaknesses. *arXiv preprint arXiv:2305.17804*, 2023.

History. Kent state shooting, 2017. Link. Accessed on December, 2024.

E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023.

Oyez. Mount holly v. mt. holly gardens citizens in action, inc., 2013. Link. Accessed on December, 2024.

H. Yoganarasimhan and I. Iakovetskaia. From feeds to inboxes: A comparative study of polarization in facebook and email news sharing. *Management Science*, 70(9):6461–6472, 2024.

Y. Zhang, H. Ling, J. Gao, K. Yin, J.-F. Lafleche, A. Barriuso, A. Torralba, and S. Fidler. Datasetgan: Efficient labeled data factory with minimal human effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10145–10155, 2021.