From Fragments to Facts: A Curriculum-Driven DPO Approach for Generating Hindi News Veracity Explanations

Anonymous ACL submission

Abstract

In an era of rampant misinformation, generating reliable news explanations is vital, especially for underrepresented languages like Hindi. Lacking robust automated tools, Hindi faces challenges in scaling misinformation detection. To bridge this gap, we propose ***DeFactoX***, a novel framework integrating Direct Preference Optimization (DPO) with Curriculum learning to align machine-generated explanations with human reasoning. Fact-checked explanations from credible sources serve as preferred responses, while LLM outputs highlight system limitations and serve as non-preferred responses. At the core of this framework lies ***Hin-DPO***, an enhanced variant of DPO that enriches the loss function with two novel parameters, *Actuality* and *Finesse*, enhancing explanation quality and consistency. Experiments with LLMs (Mistral, Llama, Gemma) and PLMs (mBART, mT5) confirm the framework's effectiveness in generating coherent, contextually relevant explanations. Moreover, owing to the data augmentation capability of our framework, it can be effectively adapted to other low-resource languages, thereby broadening the scope of automated explanation generation for combating misinformation.

1 Introduction

The rise of fake news, fuelled by its low production cost and widespread digital dissemination, poses a significant threat to the integrity of journalism. As Toomas Hendrik Ilves aptly states, "*Fake news is cheap to produce. Genuine journalism is expensive.*" This disparity is evident in the societal disruptions caused by misinformation, particularly during crises like the COVID-19 pandemic, where unverified claims about treatments and preventive measures fuelled panic and confusion (Barua et al., 2020; Roozenbeek et al., 2020). Misinformation also exacerbates political polarization, creating deep societal divides (Cantarella et al., 2023; Bovet and Makse, 2019). Fact-checking platforms face substantial challenges in scaling their efforts, especially in languages like Hindi. Despite over 600 million Hindi speakers[1], automated tools for generating credible, human-like explanations in Hindi remain underdeveloped. Addressing this gap is crucial to supporting fact-checking initiatives and reducing the impact of fake news on society.

Overview: This paper introduces ***DeFactoX***, a framework for Hindi news veracity prediction and explanation generation. The framework first constructs a synthetic preference dataset in which human-written, fact-checked explanations serve as preferred responses and LLM-generated explanations serve as rejected responses, thereby grounding alignment in reliable human reasoning while exposing the model to common erroneous patterns.

Building on this dataset, we employ Curriculum learning (Pattnaik et al., 2024) to progressively train the model to distinguish between explanations of increasing difficulty. To further strengthen alignment, we extend Direct Preference Optimization (DPO) (Rafailov et al., 2024) with a novel loss function, termed ***Hin-DPO***, which incorporates two additional parameters: ***Actuality***, quantifying the factual correctness of responses, and ***Finesse***, measuring hallucination through output instability across generations. Together, these innovations enable *DeFactoX* to generate explanations that are factually accurate, consistent, and trustworthy, while providing a scalable approach to misinformation detection in low-resource languages.

Research Gap: While NLP has made significant strides, most automated explanation generation systems focus on high-resource languages like English and Chinese (Wang et al., 2020; Zhang et al., 2020; Xu et al., 2024; Hsu et al., 2023) , leaving Hindi largely underserved.

[1] https://www.britannica.com/topic/languages-by-total-number-of-speakers-2228881

Pre-trained LLMs, trained on generalized datasets, struggle to assess the veracity of Hindi news and generate contextually relevant, factually grounded explanations for the veracity predicted. Moreover, fact-checking in Hindi remains predominantly manual, lacking scalable automated solutions. Given Hindi's vast speaker base and the increasing spread of misinformation, it is crucial to develop robust, scalable methods for *automated veracity prediction and explanation generation.*

Research Questions: This research aims to address the following questions:

- **RQ-1:** How can automated systems reliably assess the veracity of Hindi news and generate human-like explanations that are coherent and contextually relevant?
- **RQ-2:** How effective is incorporating parameters such as *Actuality* (factual accuracy) and *Finesse* (hallucination sensitivity) in aligning model outputs with human preferences?
- **RQ-3:** How can Curriculum learning be integrated with DPO to refine veracity prediction in under-represented languages?

Research Motivation: The rapid spread of misinformation in underrepresented languages like Hindi highlights the need for scalable systems that assess veracity and generate reliable, human-like explanations. Unlike high-resource languages, Hindi lacks robust fact-checking tools, and existing LLMs often struggle with coherence, factual accuracy, and human alignment, thereby motivating the need for novel frameworks that ensure trustworthy explanations in low-resource settings.

This work addresses these gaps by refining veracity explanation generation through Direct Preference Optimization (DPO) (Rafailov et al., 2024), Curriculum learning (Pattnaik et al., 2024), and our enhanced loss function ***Hin-DPO***, which integrates the ***Actuality*** score (inspired by FactScore (Min et al., 2023)) and the ***Finesse*** score (a variance-based measure of hallucination), ensuring both accuracy and scalability.

Contributions:

- We create a *synthetic, ranking-based Hindi preference dataset*, where human fact-checked explanations serve as top-ranked responses, and LLM outputs are ranked using an *automated scoring mechanism* that closely aligns with human preferences.
- We propose ***DeFactoX***, a two-stage framework that combines curriculum learning with an enhanced preference optimization objective, ***Hin-DPO***, which integrates two novel parameters: ***Actuality*** (to quantify factual correctness) and ***Finesse*** (a variance-based measure of hallucination).
- To the best of our knowledge, *DeFactoX* is the first framework for automated veracity-driven explanation generation in Hindi, and it can be extended to other low-resource languages, providing a scalable approach to combating misinformation globally.

2 Related Works

Automated Misinformation Detection and Explanation Generation: Recent studies have advanced misinformation detection and explanation generation. Joshi et al. (2023) integrated Domain Adversarial Neural Networks (DANN) with LIME to enhance COVID-19 misinformation detection. Chi and Liao (2022) proposed QA-AXDS, a scalable, interpretable fake news detection system using dialogue trees. Yao et al. (2023) introduced MOCHEG, a multimodal fact-checking benchmark incorporating textual and visual evidence. Zhou et al. (2023) examined AI-generated misinformation, highlighting linguistic nuances and proposing updated detection guidelines. Gong et al. (2024) emphasized socio-contextual cues in "social explanation" to combat misinformation. Russo et al. (2023) showed that extractive steps improve abstractive summarization for claim verification. Bilal et al. (2024) developed a GNN-based rumour verification model leveraging opinion-guided summaries. Yue et al. (2024) introduced RARG, combining evidence retrieval with RLHF-tuned LLMs, excelling in COVID-19 misinformation detection. The study by Yang et al. (2022) employed a QA-based framework with attention-driven comparisons for interpretable fact-checking. For a comprehensive review, readers are pointed to the work by Kotonya and Toni (2020).

Our Novelty: *While previous works primarily target high-resource languages like English and Chinese, our focus is on Hindi, an under-represented language.*

Applications and Advancements in Preference Optimization and Curriculum Learning: Recent advancements in preference optimization and Curriculum learning have enhanced model performance across domains. Pattnaik et al. (2024) introduced Curry-DPO, a Curriculum learning-based enhancement of DPO, achieving up to 7.5% im-

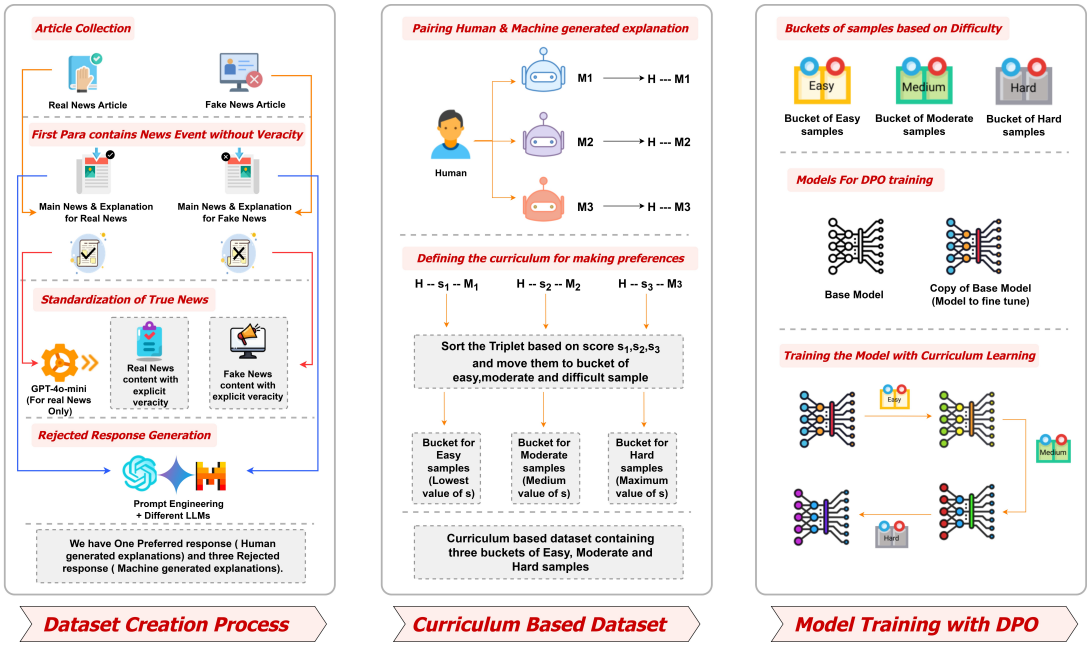

Figure 1: Overview of ***DeFactoX*** framework. **(Left)** Dataset creation with human-written explanations as preferred responses and LLM-generated explanations as rejected responses. **(Center)** Curriculum-based dataset construction, where samples are ranked and bucketed into easy, moderate, and hard levels. **(Right)** Model training with **Hin-DPO** under curriculum learning, fine-tuning the model to generate human aligned explanations.

provement across datasets. Chen et al. (2024a) proposed a multi-stage Curriculum framework optimizing humour and structure preferences in LLMs. Yin et al. Yin et al. (2024) developed Self-Augmented Preference Optimization (SAPO), surpassing DPO and SPIN across multiple benchmarks. Morimura et al. (2024) introduced filtered DPO (fDPO), refining datasets for better training efficiency. Wang et al. (2024) proposed Balanced Preference Optimization (BPO), enhancing knowledge depth while maintaining efficiency. Zeng et al. (2024) presented Token-DPO, improving alignment and diversity in LLMs through token-level fine-tuning. Wallace et al. (2024) introduced Diffusion-DPO, aligning diffusion models with human preferences for superior visual appeal. Chen et al. (2024b) proposed Softmax-DPO to enhance recommender systems via user preference modelling. Lai et al. (2024) introduced Step-DPO, improving mathematical reasoning in LLMs with minimal data. For a comprehensive review, readers are referred to the study by Xiao et al. (2024).

Our Novelty: *While DPO has been applied across various domains, our approach is specifically tailored to generating veracity claims and explanations for Hindi news. The research gap in this area highlights challenges faced by under-represented languages like Hindi, including data scarcity and the limitations of LLMs trained on multilingual datasets. Our study offers a scalable and effective solution to address these challenges.*

3 Preference Dataset Creation

To construct our synthetic preference dataset, as shown in Figure 1, we followed a systematic multi-step approach to ensure data quality, uniformity, and relevance to the task. Below, we outline the process in a structured manner.

3.1 Dataset Selection and Sampling

Multiple veracity claim misinformation detection datasets Bhardwaj et al. (2020); Kumar and Singh (2022); Sharma and Arya (2024); Bansal et al. (2024); Badam et al. (2022); Sharma and Arya (2023); Kumar et al. (2025) provide Hindi news articles sourced from fact-checking websites, ensuring authentic veracity labels. We selected data instances from Sharma and Arya (2024), a comprehensive dataset featuring over 15,000 articles in the fake news category and 13,000+ in the real news category. This dataset was chosen for its extensive coverage of news, spanning from older to recent events, and its sourcing from fact-checking websites, which provide verified veracity labels. To maintain a balanced and manageable dataset, we extracted the **most recent** 5,000 articles from

each class (fake and real). The selection ensures a healthy mix of data samples without being overbearing in size. Further stats are presented in the appendix (Section A.9).

3.2 Characteristics of the Selected Data

(1) Each article is sourced from fact-checking websites that not only classify news as fake or real but also provide comprehensive, well-reasoned explanations justifying these classifications. These **explanations serve as ground truth references** for evaluating model-generated outputs.

(2) The first paragraph of every article **strictly contains only the core news content**, intentionally excluding any veracity or reasoning. This neutral presentation ensures that readers, and more importantly, models cannot determine whether the news is fake or real based solely on this segment.

This design choice is crucial for constructing the preference dataset, as these initial news passages act as inputs for models, which must then generate both veracity predictions (fake or real) and coherent supporting explanations.

3.3 Observations on Explanations

A key observation in the dataset is the **distinct difference in the writing styles of explanations for true and fake news**.

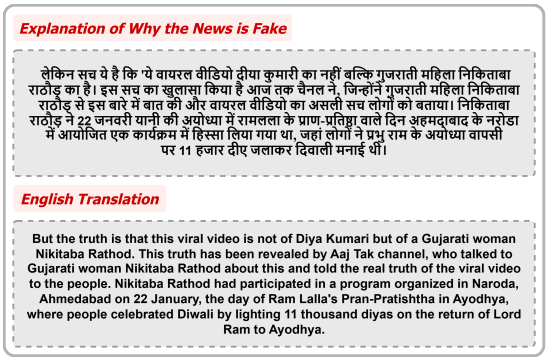

Figure 2: Snippet of fake news explanation with explicit reasoning for its veracity.

(1) True news explanations: During our manual verification, we observed that many human-written explanations were primarily informational in nature. They tended to summarize the facts of the news story but did not include explicit reasoning or assertive statements confirming its authenticity. Such explanations typically present the news content in a descriptive manner, without additional justification or emphasis on truthfulness. For an example, see appendix (Section A.2).

(2) Fake news explanations: In contrast, explanations for fake news are far more detailed and explicit. Fact-checkers provide strong declarative statements, rejecting falsehoods and supplementing them with clear justifications, such as evidence-based counterarguments, source verification, and logical reasoning. This explicit mention of veracity is illustrated in Figure 2. For a deeper understanding of these explanations, readers may consult the original fact-checking sources: OneIndia[2], Vishvas News[3], and Aaj Tak[4], which serve as the primary references for this dataset and are certified by the International Fact-Checking Network (IFCN).

3.4 Standardizing True News Explanations

While explanations for fake news naturally include explicit reasoning and veracity statements, true news explanations often lack such clarity. Fact-checking sources assume that factual news is self-evident, leading to minimal justifications. This inconsistency makes it difficult for models to learn a uniform veracity-based explanation structure. To address this, we standardized true news explanations by ensuring they explicitly affirm their veracity while preserving factual integrity. This step aligns true news explanations with the structured reasoning seen in fake news explanations, thereby creating a more balanced training dataset. For implementation details and examples, including the standardization prompt, please refer to the appendix (Section A.3 & A.5).

3.5 Generating Rejected Responses

To construct the rejected responses, we used the **first paragraph of each article, presenting only the core news content,** as input for three state-of-the-art LLMs: `gpt-4o-mini` (Achiam et al., 2023), `Mistral-7B-v0.1` (Jiang et al., 2023), and `gemini-1.5-flash` (Team et al., 2024). These models were selected due to their strong reasoning capabilities and proven performance in NLP tasks (Liu et al., 2024; Mathur et al., 2024; Siino, 2024a,b; Trott and Rivière, 2024; Sato et al., 2024).

Unlike existing approaches that aim to minimize hallucinations in synthetic preference datasets, our goal was fundamentally different: we intentionally generated **weaker, non-preferred explanations** using the prompt in appendix (Section A.4). We use a simple prompt without additional fine-tuning

[2] https://hindi.oneindia.com/fact-check/
[3] https://www.vishvasnews.com/
[4] https://www.aajtak.in/fact-check

or safeguards, letting the generated responses naturally reflect the inherent limitations of LLMs.

Observations: Through manual verification of 1,500 sampled rejected responses, we found recurring issues that distinguished them from human-written explanations. Specifically, LLM-generated explanations often:

- Over-emphasized superficial linguistic elements or unusual words in the news text.
- Displayed bias towards stylistic or surface-level framing rather than factual grounding.
- Focused on a single aspect of news instead of covering multiple factual dimensions.
- Lacked the depth, balance, and context-awareness consistently present in human-authored explanations.

These weaknesses confirmed their suitability as the negative class for preference optimization.

3.6 Final Preference Dataset Composition

The final synthetic preference dataset comprises:

- **Preferred outputs:** Explanations for fake news, sourced from fact-checking websites, and true news explanations, standardized using the prompt described in the appendix (Section A.3).
- **Rejected outputs:** Machine-generated explanations, based on the first paragraph of the news articles, produced by three state-of-the-art LLMs: `gpt-4o-mini`, `Mistral-7B-v0.1`, and `gemini-1.5-flash`.

Each data sample contains one positive (preferred) explanation and three negative (rejected) explanations, ensuring a balanced dataset by maintaining a consistent 1:3 ratio. This structure provides equal exposure to both high-quality and suboptimal explanations, helping to improve distinction and generalization. Furthermore, examples of input news and their corresponding non-preferred outputs, are provided in the appendix (Section A.10).

4 Methodology

We propose ***DeFactoX***, a unified framework for news veracity prediction and explanation generation that emphasizes both factual reliability and explanatory robustness. It builds on two complementary ideas: a curriculum learning strategy that ranks explanations by alignment with ground truth and trains progressively from easier to harder cases, and a domain-aware extension of the standard DPO objective, ***Hin-DPO***, which incorporates signals for factual consistency and stability. Together, these components help *DeFactoX* better align with human preferences (see Appendix A.11).

4.1 Explanation Ranking for Curriculum Learning

To integrate curriculum learning into our framework, we require a mechanism to distinguish explanations by their quality and progressively guide the model from easier to more challenging training cases. We achieve this by scoring non-preferred explanations, ranking them according to their alignment with ground-truth rationales, and then organizing the training sequence based on these ranks.

Scoring Function for Explanation Ranking: The scoring function ***fs*** combines BERTScore, ROUGE-L, and METEOR, as shown below:

$$\boldsymbol{fs} = \frac{\text{BERTScore} + 3 \times (\text{ROUGE-L} + \text{METEOR})}{4}$$

Justification for the Scoring Function (*fs*): Our scoring function was designed to rank explanations in a way that closely reflects human preferences. Details of the empirical study validating the weighting strategy are provided in Appendix A.6.

Curriculum Learning Strategy: Explanations were ranked using scoring function ***fs*** into three levels: **rank-0** (least aligned with ground truth, having minimum fs), **rank-1** (moderately aligned), and **rank-2** (most aligned, maximum fs). This curriculum strategy guided the model from easier to harder distinctions: it was first trained on rank-0 explanations, which are clearly different from the preferred responses and thus easier to identify, followed by rank-1 explanations with moderate similarity, and finally on rank-2 explanations. The rank-2 cases are the most challenging, as they differ only subtly from the preferred responses, requiring the model to capture fine-grained distinctions. This structured progression, akin to human learning, enhances robustness and generalizability.

4.2 Our *Hin-DPO* Loss function

We extend the standard DPO objective by proposing ***Hin-DPO***, a modified loss function tailored for news veracity prediction and explanation generation. Hin-DPO incorporates two additional parameters: ***Actuality***, which captures the factual correctness of explanations, and ***Finesse***, a variance-based measure that quantifies the degree of hallucination.

By integrating these factors, *Hin-DPO* refines preference optimization to emphasize both factual accuracy and consistency in explanation generation.

4.3 Explaining Actuality

Rationale: The reliability of AI-generated explanations for news tasks relies heavily on factual accuracy. Given the rise of misinformation and its societal consequences, explanations must align with verifiable facts. To encourage this, we introduce the *Actuality* score as a reward signal, prioritizing factually accurate explanations over those containing incorrect or unverifiable content. This score is directly integrated into our loss function to enforce factual grounding.

Calculation: *Actuality* score is designed to ensure factual accuracy in news explanations. It leverages GPT-4o-mini's internal knowledge, with access to a web search tool, to assess factual correctness. The score is computed by extracting key factual statements from a news article, evaluating their correctness, and averaging the results into a single numerical value between 0 and 1.

Prompt for *Actuality* Score

Task: *You will be given a news article. Follow these steps:*

1. *Extract up to 15 of the most important and factually relevant sentences from the article.*
2. *For each extracted sentence, assess its factual correctness:*
   - *Label each sentence as **1** if it is factually accurate.*
   - *Label it as **0** if it contains factual errors.*
3. *Compute the **average** of all the labels (1s and 0s).*

Output: *Return only the factual consistency score as a single numerical value (e.g., 0.75). Do not include any additional explanations, calculations, or extracted sentences.* ***Here is the news article:*** *{article}* ***Answer:***

Justification: Highly fluent but misleading explanations can appear convincing, leading to misinformation. The *Actuality* score penalizes inconsistent explanations and rewards factual ones. Using an LLM for factuality detection enables reliable identification of subtle inaccuracies. Incorporating this score into training ensures fact-based alignment. Experimental validation of the *Actuality* score are provided in appendix (Section A.7).

4.4 Explaining *Finesse*

Rationale: In veracity prediction and explanation tasks, models often hallucinate when uncertain. This hallucination appears as divergent explanations across repeated generations of the same input, especially in ambiguous cases. To capture this phenomenon, we introduce the *Finesse* score, which quantifies the variance of token-level probability distributions across multiple responses. A higher score reflects greater instability in model outputs, i.e., more hallucination, while a lower score reflects stability and reliability.

Calculation: To compute *Finesse*, we regenerate each explanation five times for the same input using a high temperature ($T = 0.9$). This introduces controlled diversity while preserving coherence. For each generation, we extract the token-level probability distribution over the vocabulary. We then compute the variance of these probabilities across the five runs, and finally average across the vocabulary to obtain a single scalar score:

$$\text{Finesse} = \frac{1}{|V|} \sum_{j=1}^{|V|} \text{Var}\big(P(y_j \mid x)\big)$$

where V denotes the model's vocabulary and $P(y_j \mid x)$ is the probability assigned to token j across generations. Higher *Finesse* shows that the model produced more diverse and unstable explanations for the same input, signaling hallucination.

Integration: The *Finesse* score serves as a measure of explanation uncertainty.In Hin-DPO loss, it rescales log probabilities so that stable explanations (low variance) are prioritized, while unstable or hallucinatory ones (high variance) are down-weighted. This improves factual reliability by penalizing outputs prone to hallucination.

Justification: The choice of a high-temperature value each of the five times strikes a balance between introducing diversity and maintaining response quality. Additionally, the computational cost of generating five responses was manageable for our dataset size.

4.5 Modified DPO Loss Function

This section presents the modifications made to the DPO loss function by incorporating domain-specific parameters such as *Actuality* and *Finesse* scores. These additions aim to balance factual accuracy and uncertainty, ensuring that the model generates reliable and contextually relevant outputs aligned with human preferences.

$$\text{Let } r_w = \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)}, \quad r_l = \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}$$

$$S(x, y_w, y_l) = \frac{1}{v+\epsilon}\Big[(1+s_w)\log r_w - \max(0.01, s_l)\log r_l\Big]$$

Then the *Hin-DPO* loss function is defined as:

$$L_{\text{Hin-DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim D}\Big[\log\sigma\big(\beta \cdot S\big)\Big]$$

Here $\pi_\theta(y \mid x)$ and $\pi_{\text{ref}}(y \mid x)$ denote the probabilities assigned by the learned policy and the reference policy, respectively, for a response y given input x. The dataset sample $(x, y_w, y_l) \sim D$ consists of an input x, a preferred response y_w, and a rejected response y_l. The *Actuality* scores s_w and s_l quantify the factual accuracy of the preferred and rejected responses, respectively. The hyperparameter β controls the preference weighting, while v represents the *Finesse* score. ϵ is a learnable parameter. The gradient analysis of *Hin-DPO* is represented in appendix (Section A.8). It should be noted that despite the inclusion of these additional parameters, the core preference-based alignment in DPO remains intact.

Utilization of *Actuality*: We leverage *Actuality* scores to modulate the weighting of preferred and rejected responses. Specifically, $(1 + s_w)$ amplifies the log probability of the preferred response, ensuring that factually accurate explanations are prioritized. Conversely, $\max(0.01, s_l)$ is applied to the rejected response, penalizing factually weak explanations without over-penalization.

Impact of *Finesse*: The *Finesse* score, computed via variance in predictions for an input, regulates the scaling factor in DPO loss. Low variance (indicating stable predictions) amplifies the reward, while high variance (≈ 1) retains standard DPO.

5 Experimental Setup

We fine-tuned five models, comprising three Large Language Models (LLMs): Gemma-2-9B-It (Team, 2024), Llama-3.1-8B-Instruct (Dubey et al., 2024), and Mistral-7B-Instruct-v0.3 (Jiang et al., 2023) and two Pre-trained Language Models (PLMs): mBART-large-50 (Tang et al., 2020) and mT5-large (Xue et al., 2021). The quality of the generated explanations was assessed using three key metrics: BERTSCORE (Zhang et al., 2019), ROUGE-1,2, L score (Lin, 2004) and METEOR score (Banerjee and Lavie, 2005). Given the involvement of Hindi, we utilized the Polyglot tokenizer (Al-Rfou' et al., 2013) to evaluate ROUGE-1, 2, L, and METEOR scores. Hyperparameters are presented in the appendix (Section A.1).

6 Results and Analysis

In this section, we present our findings and provide a comprehensive analysis of how our proposed framework, ***DeFactoX***, compares against a range

Model →	**mBART**					**mT5**					**Gemma2-9B**		
Config↓	**R-1**	**R-2**	**R-L**	**MT**	**BS**	**R-1**	**R-2**	**R-L**	**MT**	**BS**	**R-1**	**R-2**	**R-L**
Base	13.39	6.33	9.89	18.49	70.14	15.41	7.11	10.21	19.29	70.01	29.51	18.03	22.17
Base+SFT	14.68	8.21	10.99	20.52	71.37	16.83	8.09	11.15	20.15	72.78	30.12	18.74	23.11
DPO	16.21	9.61	12.18	20.60	73.17	17.93	9.14	13.59	22.87	73.61	30.92	19.86	25.19
DPO+Act	17.66	10.50	11.73	22.12	75.32	19.23	9.51	14.22	**24.67**	76.35	32.68	20.55	27.13
DPO+Fin	17.96	10.95	**13.73**	22.63	76.09	19.81	**9.53**	15.97	24.11	75.33	33.41	21.26	26.59
Hin-DPO	**18.98**	**11.85**	13.19	**23.50**	**77.19**	**20.29**	9.45	**16.89**	24.39	**77.32**	**33.55**	**21.64**	**27.91**
Δ (vs DPO)	+2.77	+2.24	+1.01	+2.90	+4.02	+2.36	+0.31	+3.30	+1.52	+3.71	+2.63	+1.78	+2.72
Model →	**Mistral-7B**					**Llama3.1-8B**					**Gemma2-9B**		
Config↓	**R-1**	**R-2**	**R-L**	**MT**	**BS**	**R-1**	**R-2**	**R-L**	**MT**	**BS**	**MT**	**BS**	
Base	26.21	16.69	22.72	26.11	76.34	32.12	19.91	25.02	30.61	77.52	29.12	76.88	
Base+SFT	26.89	18.02	24.36	27.15	78.27	33.46	21.03	27.45	32.97	79.82	30.62	78.45	
DPO	28.04	19.77	25.76	28.92	79.34	34.56	22.00	28.73	34.53	80.98	31.21	79.59	
DPO+Act	29.65	20.76	27.13	**30.18**	81.94	35.71	23.12	29.23	36.13	82.42	33.10	82.23	
DPO+Fin	29.79	21.12	27.34	29.51	81.55	36.51	22.33	30.19	36.23	83.79	33.17	82.52	
Hin-DPO	**30.82**	**22.07**	**28.13**	29.87	**82.95**	**37.13**	**24.07**	**31.22**	**37.25**	**84.73**	**33.84**	**83.67**	
Δ (vs DPO)	+2.78	+2.30	+2.37	+0.95	+3.61	+2.57	+2.07	+2.49	+2.72	+3.75	+2.63	+4.08	

Table 1: Performance comparison across models. **Abbreviations:** R-1: ROUGE-1, R-2: ROUGE-2, R-L: ROUGE-L, MT: METEOR, BS: BERTScore, Act: Actuality, Fin: Finesse. Bold values denote best performance. The blue-shaded row corresponds to ***Hin-DPO***. The Δ rows indicate relative improvement of *Hin-DPO* over DPO.

Method	mT5					LLaMA3.1-8B				
	R-1	R-2	R-L	MT	BS	R-1	R-2	R-L	MT	BS
Hin-DPO (w/o CL)	18.78	9.26	15.22	23.50	76.19	35.14	22.17	29.74	35.22	82.01
Hin-DPO (with CL)	20.29	9.45	16.89	24.39	77.32	37.13	24.07	31.22	37.25	84.73

Table 2: Ablation study on mT5 and LLaMA3.1-8B models with and without Curriculum Learning. Curriculum Learning consistently improves performance. **Abbreviations:** R-1: ROUGE-1, R-2: ROUGE-2, R-L: ROUGE-L, MT: METEOR, BS: BERTScore. The blue-shaded row corresponds to our proposed ***Hin-DPO*** method.

of baseline approaches. Table 1 illustrates the effectiveness of *Hin-DPO* across multiple evaluation metrics, showing its ability to generate contextually relevant, and factually accurate explanations.

A key observation is that incorporating DPO-based techniques (*DPO*, *DPO+Actuality*, and *DPO+Finesse*) consistently enhances model performance. This directly addresses **RQ1** by showcasing improvements in coherence, contextual relevance, and factual accuracy, all of which are critical for reliable news summarization. Among these, *Hin-DPO* performs best, highlighting the benefits of language-specific fine-tuning and the need for tailored approaches for resource-constrained languages like Hindi (**RQ2**).

Additionally, the consistent improvements across multiple evaluation metrics, including ROUGE, METEOR, and BERTSCORE, further emphasize that fine-tuning with DPO techniques enhances semantic alignment, fluency, and contextual accuracy. The particularly high BERTSCORE results achieved by *Hin-DPO* provide strong evidence of its ability to generate human-like explanations that are both contextually grounded and comprehensible, effectively addressing **RQ3**. This demonstrates that leveraging Curriculum learning alongside DPO-based training can systematically refine a model's factual correctness.

Ablation Study: Quantifying the Benefits of Curriculum Learning To assess the impact of Curriculum Learning, we compared models trained with and without it, mT5 and LLaMA3.1-8B. As shown in Table 2 Curriculum Learning provides consistent improvements across ROUGE, METEOR, and BERTScore for mT5. For LLaMA3.1-8B, the gains are even more pronounced, indicating that Curriculum Learning effectively enhances both semantic alignment and factual correctness.

It is worth noting that the impact of the *Actuality* and *Finesse* parameters is already reported in the main results table (Table 1).

Human Evaluation: We manually evaluated 800 explanations (400 real, 400 fake) using three student evaluators, who rated each explanation on a 0–5 scale for factual accuracy, relevance, clarity, justification, and conciseness. Scores were averaged across evaluators. Focusing on **Llama3.1-8B** and **Gemma2-9B**, the results in Table 3 show that both models consistently produce higher-quality explanations, supporting the automatic metrics and confirming improvements in factual correctness and expressiveness.

Method	Gemma2-9B	Llama3.1-8B
Base+SFT	3.29	3.07
DPO	3.92	3.87
Hin-DPO	**4.12**	**4.23**

Table 3: Human Evaluation Scores (0–5) for Gemma2-9B and Llama3.1-8B. Bold indicates best performance. The blue-shaded row corresponds to our method.

7 Conclusion

In this work, we presented ***DeFactoX***, a framework for misinformation detection and veracity-focused explanation generation in Hindi. At its core, *Hin-DPO* extends preference optimization by incorporating *Actuality* to ensure factual correctness and *Finesse* to reduce hallucination, while Curriculum Learning progressively aligns model outputs with human reasoning. *DeFactoX* offers practical utility: media houses and fact-checkers can use it to detect misleading narratives, while end-users benefit from contextually accurate explanations that strengthen their ability to distinguish truth from misinformation. Despite these advances, challenges remain, including limited availability of high-quality fact-checked data and difficulty handling highly complex or domain-specific claims. Future directions include extending *DeFactoX* to other low-resource languages through multilingual transfer and incorporating human-in-the-loop feedback to further enhance explanation quality.

Limitations

DeFactoX faces several limitations. High-quality, fact-checked Hindi datasets are limited, restricting the diversity and domain coverage of training data. The framework may struggle with highly specialized or technical news where background knowledge is crucial. While the approach is designed for Hindi, its adaptation to other low-resource languages requires further validation. Additionally, the Actuality score relies on external LLMs, which may themselves contain biases or inaccuracies, and Finesse computation requires multiple output generations, increasing computational costs.

Ethics Statement

DeFactoX is designed to support responsible misinformation mitigation by generating reliable, human-aligned explanations for underrepresented languages like Hindi. Human evaluators were involved to ensure alignment with human judgment and reduce automated bias. Users are advised to exercise caution and verify high-stakes claims, particularly in politically or medically sensitive contexts. While Actuality and Finesse reduce hallucination and factual errors, source data biases may persist. All data used is publicly available fact-checked news, and no private or sensitive user information was included.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Rami Al-Rfou', Bryan Perozzi, and Steven Skiena. 2013. Polyglot: Distributed word representations for multilingual NLP. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 183–192, Sofia, Bulgaria. Association for Computational Linguistics.

Jathin Badam, Akash Bonagiri, Kvln Raju, and Dipanjan Chakraborty. 2022. Aletheia: A fake news detection system for hindi. In *Proceedings of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)*, CODS-COMAD '22, page 255–259, New York, NY, USA. Association for Computing Machinery.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Shubhi Bansal, Nishit Sushil Singh, Shahid Shafi Dar, and Nagendra Kumar. 2024. Mmcfnd: Multimodal multilingual caption-aware fake news detection for low-resource indic languages. *arXiv preprint arXiv:2410.10407*.

Zapan Barua, Sajib Barua, Salma Aktar, Najma Kabir, and Mingze Li. 2020. Effects of misinformation on covid-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Progress in Disaster Science*, 8:100119.

Mohit Bhardwaj, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2020. Hostility detection dataset in hindi. *arXiv preprint arXiv:2011.03588*.

Iman Munire Bilal, Preslav Nakov, Rob Procter, and Maria Liakata. 2024. Generating unsupervised abstractive explanations for rumour verification. *arXiv preprint arXiv:2401.12713*.

Alexandre Bovet and Hernán A Makse. 2019. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1):7.

Michele Cantarella, Nicolò Fraccaroli, and Roberto Volpe. 2023. Does fake news affect voting behaviour? *Research Policy*, 52(1):104628.

Yang Chen, Chong Yang, Tu Hu, Xinhao Chen, Man Lan, Li Cai, Xinlin Zhuang, Xuan Lin, Xin Lu, and Aimin Zhou. 2024a. Are U a joke master? pun generation via multi-stage curriculum learning towards a humor LLM. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 878–890, Bangkok, Thailand. Association for Computational Linguistics.

Yuxin Chen, Junfei Tan, An Zhang, Zhengyi Yang, Leheng Sheng, Enzhi Zhang, Xiang Wang, and Tat-Seng Chua. 2024b. On softmax direct preference optimization for recommendation. *arXiv preprint arXiv:2406.09215*.

Haixiao Chi and Beishui Liao. 2022. A quantitative argumentation-based automated explainable decision system for fake news detection on social media. *Knowledge-Based Systems*, 242:108378.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Yeaeun Gong, Lanyu Shang, and Dong Wang. 2024. Integrating social explanations into explainable artificial intelligence (xai) for combating misinformation: Vision and challenges. *IEEE Transactions on Computational Social Systems*.

Chiaming Hsu, Changtong Zan, Liang Ding, Longyue Wang, Xiaoting Wang, Weifeng Liu, Fu Lin, and Wenbin Hu. 2023. Prompt-learning for cross-lingual relation extraction. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, and 1 others. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Gargi Joshi, Ananya Srivastava, Bhargav Yagnik, Mohammed Hasan, Zainuddin Saiyed, Lubna A Gabralla, Ajith Abraham, Rahee Walambe, and Ketan Kotecha. 2023. Explainable misinformation detection across multiple social media platforms. *IEEE Access*, 11:23634–23646.

Neema Kotonya and Francesca Toni. 2020. Explainable automated fact-checking: A survey. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5430–5443, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Raghvendra Kumar, Pulkit Bansal, Raunak Kumar Singh, and Sriparna Saha. 2025. Sifting truth from spectacle! a multimodal hindi dataset for misinformation detection with emotional cues and sentiments. *Authorea Preprints*.

Sudhanshu Kumar and Thoudam Doren Singh. 2022. Fake news detection on hindi news dataset. *Global Transitions Proceedings*, 3(1):289–297.

Xin Lai, Zhuotao Tian, Yukang Chen, Senqiao Yang, Xiangru Peng, and Jiaya Jia. 2024. Step-dpo: Step-wise preference optimization for long-chain reasoning of llms. *arXiv preprint arXiv:2406.18629*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Xiang Liu, Peijie Dong, Xuming Hu, and Xiaowen Chu. 2024. LongGenBench: Long-context generation benchmark. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 865–883, Miami, Florida, USA. Association for Computational Linguistics.

Suyash Vardhan Mathur, Jainit Sushil Bafna, Kunal Kartik, Harshita Khandelwal, Manish Shrivastava, Vivek Gupta, Mohit Bansal, and Dan Roth. 2024. Knowledge-aware reasoning over multimodal semi-structured tables. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14054–14073, Miami, Florida, USA. Association for Computational Linguistics.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2023. FActScore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12076–12100, Singapore. Association for Computational Linguistics.

Tetsuro Morimura, Mitsuki Sakamoto, Yuu Jinnai, Kenshi Abe, and Kaito Ariu. 2024. Filtered direct preference optimization. *arXiv preprint arXiv:2404.13846*.

Pulkit Pattnaik, Rishabh Maheshwary, Kelechi Ogueji, Vikas Yadav, and Sathwik Tejaswi Madhusudhan. 2024. Enhancing alignment using curriculum learning & ranked preferences. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 12891–12907.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. 2020. Susceptibility to misinformation about covid-19 around the world. *Royal Society open science*, 7(10):201199.

Daniel Russo, Serra Sinem Tekiroğlu, and Marco Guerini. 2023. Benchmarking the generation of fact checking explanations. *Transactions of the Association for Computational Linguistics*, 11:1250–1264.

Ayako Sato, Kyotaro Nakajima, Hwichan Kim, Zhousi Chen, and Mamoru Komachi. 2024. TMU-HIT's submission for the WMT24 quality estimation shared task: Is GPT-4 a good evaluator for machine translation? In *Proceedings of the Ninth Conference on Machine Translation*, pages 529–534, Miami, Florida, USA. Association for Computational Linguistics.

Richa Sharma and Arti Arya. 2023. Lfwe: Linguistic feature based word embedding for hindi fake news detection. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(6).

Richa Sharma and Arti Arya. 2024. Mmhfnd: Fusing modalities for multimodal multiclass hindi fake news detection via contrastive learning. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Marco Siino. 2024a. Mcrock at semeval-2024 task 4: Mistral 7b for multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 53–59.

Marco Siino. 2024b. Mistral at semeval-2024 task 5: Mistral 7b for argument reasoning in civil procedure. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 155–162.

C Spearman. 1904. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2020. Multilingual translation with extensible multilingual pretraining and finetuning.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, and 1 others. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

Gemma Team. 2024. Gemma.

Sean Trott and Pamela Rivière. 2024. Measuring and modifying the readability of English texts with GPT-4. In *Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024)*, pages 126–134, Miami, Florida, USA. Association for Computational Linguistics.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. 2024. Diffusion model alignment using direct preference optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8228–8238.

Dongyang Wang, Junli Su, and Hongbin Yu. 2020. Feature extraction and analysis of natural language processing for deep learning english language. *IEEE Access*, 8:46335–46345.

Sizhe Wang, Yongqi Tong, Hengyuan Zhang, Dawei Li, Xin Zhang, and Tianlong Chen. 2024. Bpo: Towards balanced preference optimization between knowledge breadth and depth in alignment. *arXiv preprint arXiv:2411.10914*.

Wenyi Xiao, Zechuan Wang, Leilei Gan, Shuai Zhao, Wanggui He, Luu Anh Tuan, Long Chen, Hao Jiang, Zhou Zhao, and Fei Wu. 2024. A comprehensive survey of datasets, theories, variants, and applications in direct preference optimization. *arXiv preprint arXiv:2410.15595*.

Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, Yang Wang, and Enhong Chen. 2024. Large language models for generative information extraction: A survey. *Frontiers of Computer Science*, 18(6):186357.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Jing Yang, Didier Vega-Oliveros, Taís Seibt, and Anderson Rocha. 2022. Explainable fact-checking through question answering. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8952–8956.

Barry Menglong Yao, Aditya Shah, Lichao Sun, Jin-Hee Cho, and Lifu Huang. 2023. End-to-end multimodal fact-checking and explanation generation: A challenging dataset and models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2733–2743.

Yueqin Yin, Zhendong Wang, Yujia Xie, Weizhu Chen, and Mingyuan Zhou. 2024. Self-augmented preference optimization: Off-policy paradigms for language model alignment. *arXiv preprint arXiv:2405.20830*.

Zhenrui Yue, Huimin Zeng, Yimeng Lu, Lanyu Shang, Yang Zhang, and Dong Wang. 2024. Evidence-driven retrieval augmented response generation for online misinformation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5628–5643, Mexico City, Mexico. Association for Computational Linguistics.

Yongcheng Zeng, Guoqing Liu, Weiyu Ma, Ning Yang, Haifeng Zhang, and Jun Wang. 2024. Token-level direct preference optimization. *arXiv preprint arXiv:2404.11999*.

Peng Zhang, Yunlu Xu, Zhanzhan Cheng, Shiliang Pu, Jing Lu, Liang Qiao, Yi Niu, and Fei Wu. 2020. Trie: end-to-end text reading and information extraction for document understanding. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1413–1422.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20.

A Appendix

A.1 Major Hyperparameters 12
A.2 Example of a human-written true news explanation 12
A.3 Prompt for Standardizing True News Explanations 12
A.4 Prompt for Generating Non-Preferred Explanations 13
A.5 Example of Standardizing True News Explanations 13
A.6 Validation of the Scoring Function (*fs*) 13
A.7 Experimental Validation of *Actuality* Score: 13
A.8 Gradient Analysis of Hin-DPO . . 14
A.9 Dataset Statistics 14
A.10 Example of Non-Preferred Outputs 15
A.11 Algorithms for Dataset Creation and Hin-DPO Training 17

A.1 Major Hyperparameters

This section details the resources and configurations used in our experiments. Dataset generation with `Mistral-7B-v0.1` was performed on an NVIDIA RTX 3090 GPU (24 GB), whereas `gpt-4o-mini` and `gemini-1.5-flash` relied on smaller GPUs, as their API-based tasks required minimal computational resources. Fine-tuning of the LLMs was carried out on an L40S GPU with 45 GB of memory to ensure efficient processing. The dataset was split into 75% training, 5% validation, and 20% testing subsets. For data generation, `gpt-4o-mini` and `Mistral-7B-v0.1` employed a temperature of 0.7, top_k of 50, and top_p of 0.95, while `gemini-1.5-flash` used default parameters. Direct Preference Optimization (DPO) alignment was performed over 10 epochs with a learning rate of 1×10^{-4}, a batch size of 2, and a beta value of 0.6, taking approximately 48 hours, with fine-tuning conducted in parallel. All experiments were reproducible, and hyperparameters were carefully tuned to balance performance and computational efficiency.

A.2 Example of a human-written true news explanation

Figure 3 illustrates a typical human-written explanation for true news that we got from websites, summarizing factual content without explicit reasoning or assertive claims.

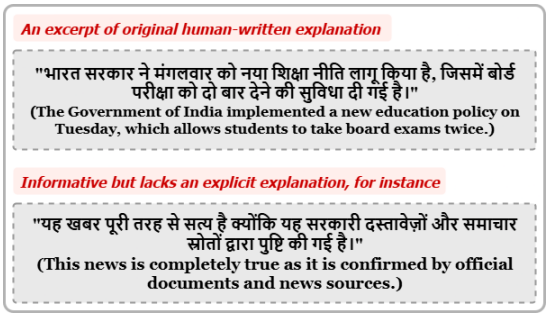

Figure 3: Example of a human-written true news explanation that is primarily informational, summarizing facts without explicitly confirming authenticity.

A.3 Prompt for Standardizing True News Explanations

To ensure uniformity in true news explanations, we employed prompt engineering with the **GPT-4o-mini model** (Achiam et al., 2023). The goal was to make true news explanations explicitly state their veracity, aligning them with the structured reasoning found in fake news explanations.

Prompt for Explicit True News Reasoning

I will provide you with an article containing a verified news story along with related contextual information. Your task is to rewrite the article as an explanation in Hindi, explicitly emphasizing that the news is true. Follow these rules strictly:

1. Use only the information provided in the article. Do not add, remove, or fabricate any content.
2. Insert at least two explicit affirmations stating that the news is true, and repeat this emphasis naturally within the explanation.
3. Preserve the original paragraphing and logical sequence of the article, without introducing unnecessary structural changes.
4. Ensure the tone is factual, clear, and objective, avoiding exaggeration or speculation.
5. The final output must be entirely in Hindi.

This prompt ensures:

1. **Consistency:** True news explanations explicitly affirm their veracity, aligning them with fake news explanations.
2. **Factual Integrity:** The process avoids intro-

ducing biases while preserving the original content.

3. **Linguistic Uniformity:** All explanations are generated in Hindi for consistency.

A.4 Prompt for Generating Non-Preferred Explanations

The following prompt was used to generate weaker, non-preferred explanations for comparison with preferred model outputs. It was intentionally kept simple, without additional fine-tuning or safeguards, to capture the natural shortcomings of LLMs.

Prompt for News Explanation

Task: I will provide you with a news article. Your task is to do the following:
1. Predict whether the news is fake or real.
2. Provide a detailed explanation for your prediction.
Ensure your response is written as a flowing paragraph, avoiding bullet points, numbering, or any other structured format. The explanation should naturally justify your prediction, without adding any extraneous information.
Here is the news article: `article`
Answer in the form of a detailed paragraph.

A.5 Example of Standardizing True News Explanations

To illustrate our standardized prompting approach, Figure 4 presents an example with the input (a verified Hindi news article on PM Narendra Modi's 2024 election speech) and the corresponding standardized explanation generated using GPT-4o-mini.

Input (Original Article): The article reports PM Modi's rally in Andhra Pradesh with Chandrababu Naidu and Pawan Kalyan, highlighting NDA's unity, the "400+ seats" slogan, and themes of progress and development. While factually accurate, the input provides no explicit reasoning about its veracity.

Output (Standardized Explanation): The standardized explanation explicitly affirms the truthfulness of the news (e.g., "This news is completely true"), while retaining all factual details, restructuring content for clarity, and maintaining strict Hindi language consistency. Strategic repetition of verification statements reinforces authenticity without altering the original content.

A.6 Validation of the Scoring Function (*fs*)

To validate the choice of our scoring function, we conducted an empirical study over 300 randomly sampled explanations (150 True News and 150 Fake News). Each sample was shown to human annotators, who were asked to rank the three rejected responses in order of quality. We then compared three weighting strategies for aligning automatic scores with these human rankings: (i) equal weighting of all metrics (1:1:1), (ii) a moderate weighting giving twice the importance to (ROUGE-L + METEOR) relative to BERTScore (1:2), and (iii) a stronger weighting giving three times the importance to (ROUGE-L + METEOR) relative to BERTScore (1:3).

The degree of alignment between automatic scores and human-provided rankings was quantified using the Spearman rank correlation (Spearman, 1904), with results summarized in Table 4.

Scoring Strategy	Spearman ρ
1:1 weighted average	0.63
1:2 weighted average	0.74
1:3 weighted average	**0.81**

Table 4: Alignment of scoring strategies with human-annotated rankings. Here, the ratios indicate the relative weight assigned to BERTScore versus the combined contribution of ROUGE-L and METEOR. For example, 1:3 means BERTScore is given weight 1, while the sum of ROUGE-L and METEOR is given weight 3.

The 1:3 weighting achieved the strongest alignment ($\rho = 0.81$) and was therefore adopted in our framework. *Importantly, this scoring mechanism is not proposed as a novel contribution, but rather as a supporting utility to align curriculum-based training with human judgments.* Other weighting schemes or evaluation metrics may be equally suitable in different datasets or applications.

A.7 Experimental Validation of *Actuality* Score:

Since the *Actuality* score relies directly on the factual predictions of GPT model, its reliability is strongly tied to the efficacy of this model. To evaluate this, we conducted a human study comparing GPT's predictions with independent human judgments, as shown in Table 5 & 6. A total of 200 explanations were randomly sampled, with 100

from true news and 100 from fake news. From each explanation, two sentences were extracted, yielding 400 factual claims. Each claim was fact-checked by three independent student evaluators, who had access to Google Search and ChatGPT with browsing tools, and assigned binary labels as factual (1) or non-factual (0). These labels were then compared against GPT-4o-mini's binary predictions used in the computation of the *Actuality* score.

Label Type	**(1)**	**(0)**	**Total**
Human Labels	230	170	400
GPT Predictions	260	140	400

Table 5: Distribution of human labels and GPT-4o-mini predictions across 400 factual claims. '1' represents Correct and '0' represents Incorrect.

Comparison	**Count**
True Positives (TP)	205
False Positives (FP)	55
True Negatives (TN)	115
False Negatives (FN)	25

Table 6: Confusion matrix comparison between GPT-4o-mini predictions and human judgments.

Overall, GPT-4o-mini achieved an accuracy of 80.0%, precision of 78.8%, recall of 89.1%, and F1-score of 83.7%. These results demonstrate strong alignment with human fact-checking and validate the use of GPT-4o-mini as the underlying model for computing the *Actuality* score. Nevertheless in principle, any capable language model could be employed to generate the factuality assessments on which the score depends.

A.8 Gradient Analysis of Hin-DPO

A.8.1 Objective Function

The Hin-DPO objective is given by:

$$L_{\text{Hin-DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\big[\log \sigma(u)\big], \tag{1}$$

where

$$u = \frac{\beta}{v+\epsilon}\Big((1+s_w)r_w - \max(0.01, s_l)\, r_l\Big). \tag{2}$$

where we define the log-ratio terms as

$$r_w = \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)}, \qquad r_l = \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}. \tag{3}$$

A.8.2 Loss Function Formulation

$$r(x,y) = \beta \log \frac{\pi_\theta(y \mid x)}{\pi_{\text{ref}}(y \mid x)} + \beta \log Z(x). \tag{4}$$

The reward models for our context, associated with the preferred and rejected responses, respectively, are expressed as:

$$r_w(x,y) = (1+s_w)\, r(x,y), \tag{5}$$

$$r_l(x,y) = \max(0.01, s_l)\, r(x,y). \tag{6}$$

The preference probability is

$$P(y_w \succ y_l \mid x) = \frac{\exp(r_w(x,y_w))}{\exp(r_w(x,y_w)) + \exp(r_l(x,y_l))} \tag{7}$$

$$= \frac{1}{1+\exp\big(r_l(x,y_l) - r_w(x,y_w)\big)} \tag{8}$$

$$= \sigma\Big(\beta\big[(1+s_w)r_w - \max(0.01, s_l)r_l\big]\Big). \tag{9}$$

A.8.3 Gradient Derivation

Let the sigmoid argument be

$$u = \frac{\beta\big((1+s_w)r_w - \max(0.01, s_l)\, r_l\big)}{v+\epsilon}. \tag{10}$$

The gradient of the objective is

$$\nabla_\theta L_{\text{Hin-DPO}} = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\big[(1-\sigma(u))\, \nabla_\theta u\big]. \tag{11}$$

We compute

$$\nabla_\theta u = \frac{\beta}{v+\epsilon}\Big[(1+s_w)\nabla_\theta r_w - \max(0.01, s_l)\nabla_\theta r_l\Big]. \tag{12}$$

Substituting back gives

$$\nabla_\theta L_{\text{Hin-DPO}} = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\frac{\beta^2\sigma(u)}{v+\epsilon}\, C\right], \tag{13}$$

where

$$C = (1+s_w)\nabla_\theta r_w - \max(0.01, s_l)\, \nabla_\theta r_l. \tag{14}$$

A.9 Dataset Statistics

In this section, we provide a more comprehensive overview of the dataset, covering label distribution and model-specific explanation statistics (Tables 7 & 8). This structured overview captures the source and average length of explanations, offering transparency into how explanations were curated and generated. It also helps readers understand the composition of training samples used in preference optimization.

Explanation Type	Source	Count	Avg. Length (tokens)
Preferred Explanations	Human-curated	10000	124.7
Non-Preferred (GPT-4o-mini)	LLM Output	10000	110.4
Non-Preferred (Gemini-1.5-Flash)	LLM Output	10000	112.0
Non-Preferred (Mistral-7B)	LLM Output	10000	98.1

Table 7: Breakdown of explanation sources with counts and average token length.

Label Type	Count
Fake News Samples	5000
Real News Samples	5000
Total	**10000**

Table 8: Distribution of Fake and Real news samples in the dataset.

A.10 Example of Non-Preferred Outputs

To illustrate the composition of the synthetic preference dataset, Figure 5 presents an input article about Rahul Gandhi's viral speech video, which critics claimed distorted Mahatma Gandhi's philosophy. The controversy stemmed from a selectively cropped clip that misrepresented his statement, drawing sharp political reactions.

Non-Preferred Outputs (Machine-Generated Explanations) The bottom section of Figure 5 shows explanations generated by GPT-4o-mini, Mistral-7B-v0.1, and Gemini-1.5-Flash. Their key shortcomings include:

Model 1: GPT-4o-mini
The explanation is considered non-preferred due to its lack of precision and depth. While it correctly asserts that the news is authentic, it fails to address the core misinformation issue: the selective editing of Rahul Gandhi's speech. A preferred explanation would note that the viral video was intentionally cropped to exclude his immediate correction, which distorted the intended meaning. Moreover, the response does not cite fact-checking sources or verifiable evidence, instead relying on vague statements such as "typical for political figures." It also misses a critical analysis of how the manipulated video spread through political reactions. Although it mentions reactions from leaders, it does not contextualize their role in amplifying the misinformation. A stronger explanation would explicitly identify the manipulation, reference fact-checking timelines, and clarify Gandhi's corrected statement, ensuring a clear, fact-based debunking of the viral claim.

Model 2: Mistral-7B-v0.1
This explanation lacks relevance and critical engagement with the article's content. It dismisses the piece as "fake" based purely on writing style, arguing that its "simple and clear" or "formal and professional" tone indicates spuriousness. These are not valid indicators of news credibility, as professional reporting often uses such tones. The explanation further fails to address the actual issue: the misleading nature of the viral video and the misrepresentation of Gandhi's words. A more appropriate response would highlight the intentional cropping of his speech and its distortion. Additionally, the model does not engage with fact-checking resources or logical evidence, instead relying on unsupported stylistic judgments. This lack of reasoning makes the explanation inadequate for identifying and debunking misinformation.

Model 3: Gemini-1.5-Flash
This model produces a more balanced explanation, but it still suffers from key reasoning gaps, making it non-preferred. While it acknowledges the political event and Gandhi's misstatement, it overlooks the critical detail that the viral video was edited to distort his words. The response treats Gandhi's correction as proof of authenticity, but misses the fact that the correction was intentionally omitted in the circulated clip. The explanation also briefly mentions "verifying through reliable sources," which is positive, but it does not specify how verification should be done or reference fact-checking outlets. A stronger explanation would emphasize the role of editing, highlight the political motivations behind the distortion, and recommend consulting fact-checking platforms. Such an approach would provide a more precise, context-aware, and fact-based debunking.

Overall, these explanations lack precision, factual grounding, and critical reasoning, underscoring why they are deemed non-preferred.

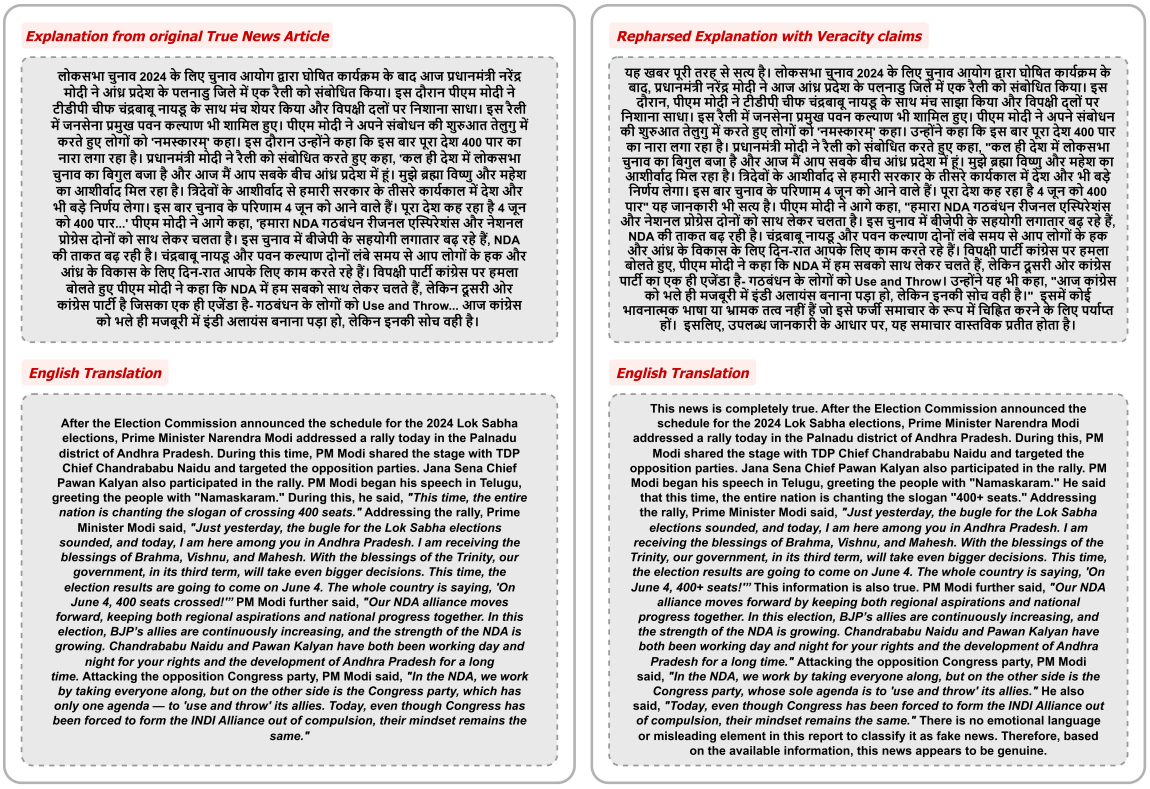

Figure 4: Snippet of True news transformation.

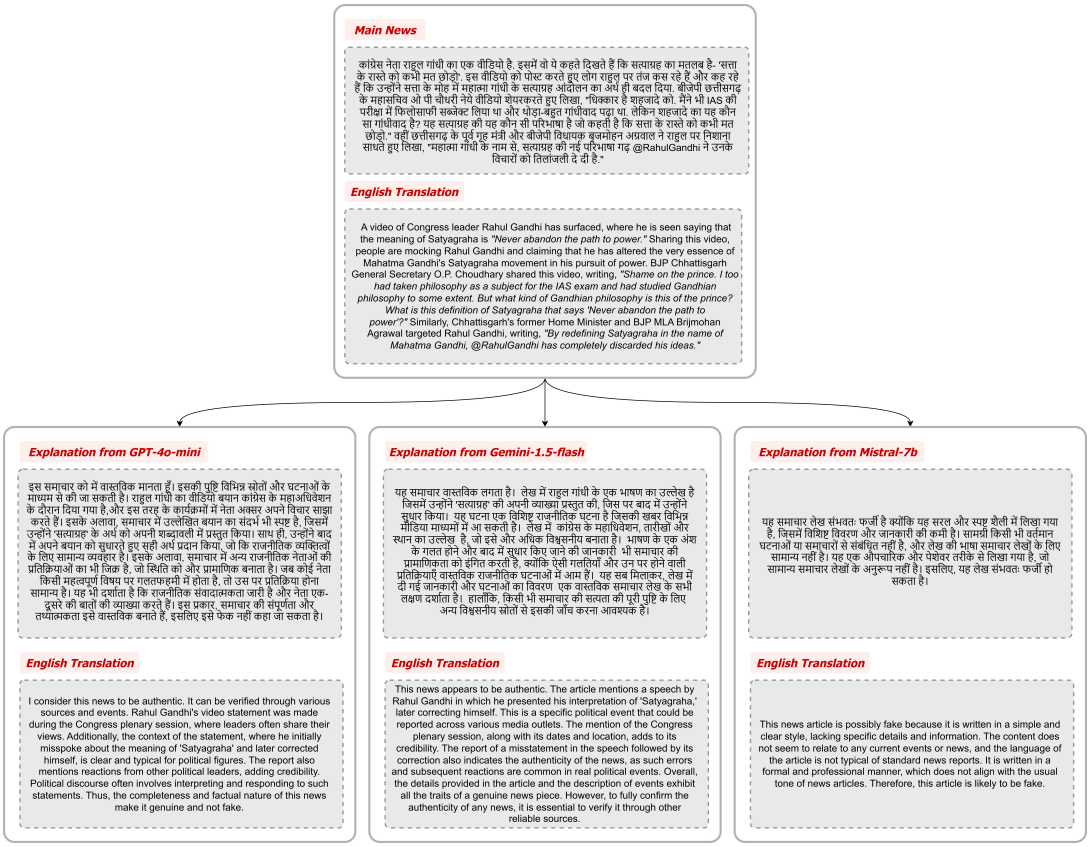

Figure 5: Snippet of Non-Preferred Response Generation.

A.11 Algorithms for Dataset Creation and Hin-DPO Training

Algorithm 1 Dataset Creation from Fact-Checked News Articles

Require: Scraped news article A containing main news N and explanation E, Large Language Models $\{LLM_1, LLM_2, LLM_3\}$, Scoring function S

Ensure: Dataset D with explanations categorized by quality

Step 1: Segregate Explanation and Main News

Extract main news N and ground-truth explanation E_{GT} from article A

Step 2: Generate LLM Explanations

for each model LLM_i in $\{LLM_1, LLM_2, LLM_3\}$ **do**

  Provide N as input to LLM_i and obtain predicted explanation E_i

end for

Step 3: Compute Scores for Explanations

for each predicted explanation E_i **do**

  Compute score $S(E_i)$ using scoring function S

end for

Step 4: Rank Explanations

Sort explanations $\{E_1, E_2, E_3\}$ in ascending order based on $S(E_i)$

Step 5: Bucketize Explanations

Define three score-based categories:

- **Low-quality bucket** $B_L \leftarrow$ Explanations with lowest scores
- **Medium-quality bucket** $B_M \leftarrow$ Explanations with mid-range scores
- **High-quality bucket** $B_H \leftarrow$ Explanations with highest scores

Step 6: Construct Final Dataset

Form dataset D by concatenating explanations in the order:

$$D = B_L \cup B_M \cup B_H$$

return D

Algorithm 2 Our Hin-DPO Training Algorithm

Require: Training dataset ad Dataloader D with win and lose samples (paired or unpaired), initial model parameters θ_0, reference model π_{ref}, number of iterations T, scaling factor β, temperature parameter τ

Initialize model π_θ with parameters θ_0
Set π_θ to training mode and π_{ref} to evaluation mode
for iteration = 1 to T **do**
  for each batch in D **do**
    Initialize running mean $\mu \leftarrow 0$ and running variance $\sigma^2 \leftarrow 0$ {Running statistics for probability distribution}
    Set num_iter $= 5$ {Number of iterations for variance computation}
    for iter = 1 to num_iter **do**
      Compute logits for the preferred response:
      logits $\leftarrow \pi_\theta(\text{pref_ids}, \text{pref_mask}).\text{logits}$
      Compute probabilities:
      probs $\leftarrow \exp(\text{log_probs(logits, pref_ids)})$
      Update Mean and Variance:
      $\mu \leftarrow \mu + \frac{\text{probs}-\mu}{\text{iter}}$
      $\sigma^2 \leftarrow \sigma^2 + (\text{probs} - \mu) \times (\text{probs} - \mu)$
    end for
    Compute final variance:
    $\sigma^2 \leftarrow \frac{\sigma^2}{\text{num_iter}-1}$
    Get log probabilities for preferred and dispreferred responses using π_θ:
    model_pref_log $\leftarrow \text{log_prob}(\pi_\theta(\text{pref_ids}, \text{pref_mask}), \text{pref_ids})$
    model_dispref_log $\leftarrow \text{log_prob}(\pi_\theta(\text{dispref_ids}, \text{dispref_mask}), \text{dispref_ids})$
    Get log probabilities for preferred and dispreferred responses using reference model π_{ref}:
    ref_pref_log $\leftarrow \text{log_prob}(\pi_{\text{ref}}(\text{pref_ids}, \text{pref_mask}), \text{pref_ids})$
    ref_dispref_log $\leftarrow \text{log_prob}(\pi_{\text{ref}}(\text{dispref_ids}, \text{dispref_mask}), \text{dispref_ids})$
    Compute Hin-DPO loss:
    loss $\leftarrow$ Hin-DPO_loss(model_pref_log, model_dispref_log, ref_pref_log, ref_dispref_log, σ^2, β)

    Backpropagate loss:
    loss.backward()
    Update model parameters:
    $\theta \leftarrow$ optimizer.step()
  end for
end for