Reducing Hallucinations in Summarization via Reinforcement Learning with Entity Hallucination Index

Praveenkumar Katwe[1,2], Rakesh Chandra Balabantaray [1], Kali Prasad Vittala[2]

[1]International Institute of Information Technology,Bhubaneshwar,INDIA,

[2]Informatica Business Solutions, Bengaluru , INDIA

Correspondence: c121007@iiit-bh.ac.in,rakesh@iiit-bh.ac.in,kprasad@informatica.com

Abstract

Reducing hallucinations in abstractive summarization remains a critical challenge for deploying language models (LMs) in real-world settings. In this work, we introduce a reward-driven fine-tuning framework that explicitly optimizes for Entity Hallucination Index (EHI), a metric designed to quantify the presence, correctness, and grounding of named entities in generated summaries. Given a corpus of meeting transcripts, we first generate baseline summaries using a pre-trained LM and compute EHI scores via automatic entity extraction and matching. We then apply reinforcement learning to fine-tune the model parameters, using EHI as a reward signal to bias generation toward entity-faithful outputs. Our approach does not rely on human-written factuality annotations, enabling scalable fine-tuning. Experiments demonstrate consistent improvements in EHI across datasets, with qualitative analysis revealing a significant reduction in entity-level hallucinations without degradation in fluency or informativeness. We release a reproducible Colab pipeline, facilitating further research on hallucination-aware model fine-tuning using lightweight, hallucintion metrics like EHI.

1 Introduction

Abstractive summarization models, powered by large language models (LLMs), have achieved impressive results across various domains. However, a persistent challenge remains: **hallucination**, where generated summaries include incorrect or fabricated information not grounded in the source input (Maynez et al., 2020; Durmus et al., 2020). In high-stakes applications such as meeting summarization, medical reporting, or financial documentation, hallucinations, particularly involving named entities, can significantly undermine trustworthiness and utility.

Existing work on hallucination detection primarily focuses on coarse-grained factuality metrics or relies on reference-based evaluations (Falke et al., 2019; Pagnoni et al., 2021). While such methods are valuable, they often fail to capture fine-grained entity-level hallucinations, and their dependence on ground-truth references limits scalability. Recent efforts have explored lightweight automatic metrics, yet integrating such evaluations directly into model training remains underexplored.

In this work, we propose a novel fine-tuning approach that leverages the **Entity Hallucination Index (EHI)** as a reward signal to reduce hallucinations at the entity level. EHI quantifies the correctness and grounding of entities by comparing extracted entities from model outputs to those present in the input document, enabling efficient evaluation with reduced reference. We first generate baseline summaries using a pre-trained model, compute EHI scores through automatic entity extraction and matching, and fine-tune the model using reinforcement learning with EHI as the reward function. This method biases the model toward producing more entity-faithful summaries without requiring human factuality annotations.

Our contributions are as follows:

- We introduce an **EHI-guided fine-tuning framework** that improves entity faithfulness in abstractive summarization.

- We present a **scalable reinforcement learning pipeline** that does not depend on human-labeled factuality datasets.

- We demonstrate **empirical improvements** in EHI scores across meeting transcript datasets and conduct qualitative analysis to highlight reductions in hallucination.

- We release a **reproducible Colab-based implementation** to facilitate further research on hallucination-aware summarization.

Our findings suggest that optimizing language models with fine-grained entity-focused rewards like EHI provides an effective pathway for enhancing faithfulness and reliability in summarization tasks.

2 Related Work

Hallucination in Summarization. Hallucination, or the generation of content that is unfaithful to the input, is a long-standing challenge in abstractive summarization (Maynez et al., 2020). Several studies have quantified hallucination prevalence across datasets and models (Pagnoni et al., 2021). Techniques such as constrained decoding (Cao et al., 2020) and fact-checking-based reranking (Goyal et al., 2022) have been proposed to mitigate hallucination during or after generation.

Entity-Level Hallucination Detection. Prior work has recognized the importance of entity-level evaluation for factuality. Metrics like FEQA (Durmus et al., 2020) and QuestEval (Scialom et al., 2021) attempt to assess factual consistency using question answering techniques. However, these methods often require reference summaries or external QA systems, limiting scalability. Our work builds on this line by employing the **Entity Hallucination Index (EHI)**, a lightweight evaluation metric focusing specifically on named entity faithfulness.

Reward-Driven Fine-Tuning. Reinforcement learning from human feedback (RLHF) has gained traction for aligning model outputs with desired qualities, such as factuality and helpfulness (Ouyang et al., 2022). In summarization, factuality-aware reward models have been used for fine-tuning (Pasunuru and Bansal, 2018). Unlike previous approaches that rely on coarse factuality rewards or human-labeled datasets, we propose fine-tuning using EHI as a direct, automatic reward signal, thereby promoting entity-grounded summarization without additional supervision.

3 Methodology

Our objective is to fine-tune a language model (LM) to generate summaries with reduced entity hallucination. We achieve this by designing a reward function based on the Entity Hallucination Index (EHI) and applying reinforcement learning to maximize it. Figure 1 presents an overview of our pipeline.

3.1 Problem Formulation

Given an input document x, the goal is to generate a summary y such that entities mentioned in y are grounded in x. Traditional maximum likelihood training does not explicitly optimize for factual consistency. Instead, we introduce a reward-driven fine-tuning strategy where the reward is proportional to entity faithfulness, measured using EHI.

3.2 Datasets Preparations

We use the official datasets provided in the AutoMin 2025 shared task [1], which includes the ELITR Minuting Corpus for summarization and the ELITR-Bench dataset for question answering [2]. We followed the task protocols as outlined in the shared task instructions, including the official train/dev/test splits. For summarization (Task A), gold summaries are provided for reference-based evaluation. For question answering (Task B), answer spans are matched against gold annotations for exact match and F1 scoring. We evaluate our approach on given meeting transcript datasets, which consist of multi-turn conversational dialogues along with abstractive gold summaries. Each sample is formatted as a flattened transcript-summary pair. To simulate real-world entity usage, datasets include a diverse range of named entities spanning persons, organizations, locations, and events. We randomly partition each dataset into training (80%), validation (10%), and test (10%) splits.

Entity extraction is performed on both source documents and summaries for computing the Entity Hallucination Index (EHI). During fine-tuning, we exclusively utilize the training split, while validation EHI scores are monitored to select the best checkpoint. Final evaluations are reported on the held-out test set.

3.3 Baseline Summarization

We begin with a pre-trained encoder-decoder LM (e.g., `Flan-T5-Large`) fine-tuned on generic summarization tasks. For each input document, the model generates an initial summary using standard greedy or beam search decoding strategies.

3.4 Entity Hallucination Index Computation

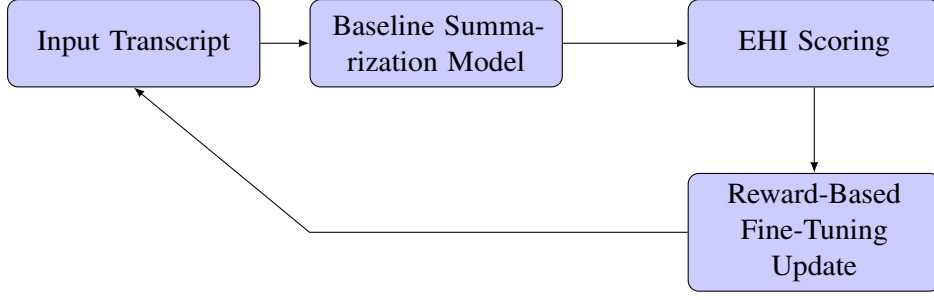To assess entity faithfulness, We formally define the **Entity Hallucination Index (EHI)** as:

Figure 1: Overview of our fine-tuning pipeline using EHI rewards.

- Upload and extract ELITR Minuting Corpus dataset
- Clean transcripts (remove tags, speaker IDs, extra spaces)
- Load pretrained summarization models ,Limit cleaned transcript to 1024 tokens (model input constraint)
- Generate summaries using each model separately
- Store results with original transcript, summary, reference minutes
- Evaluate using EHI metrics
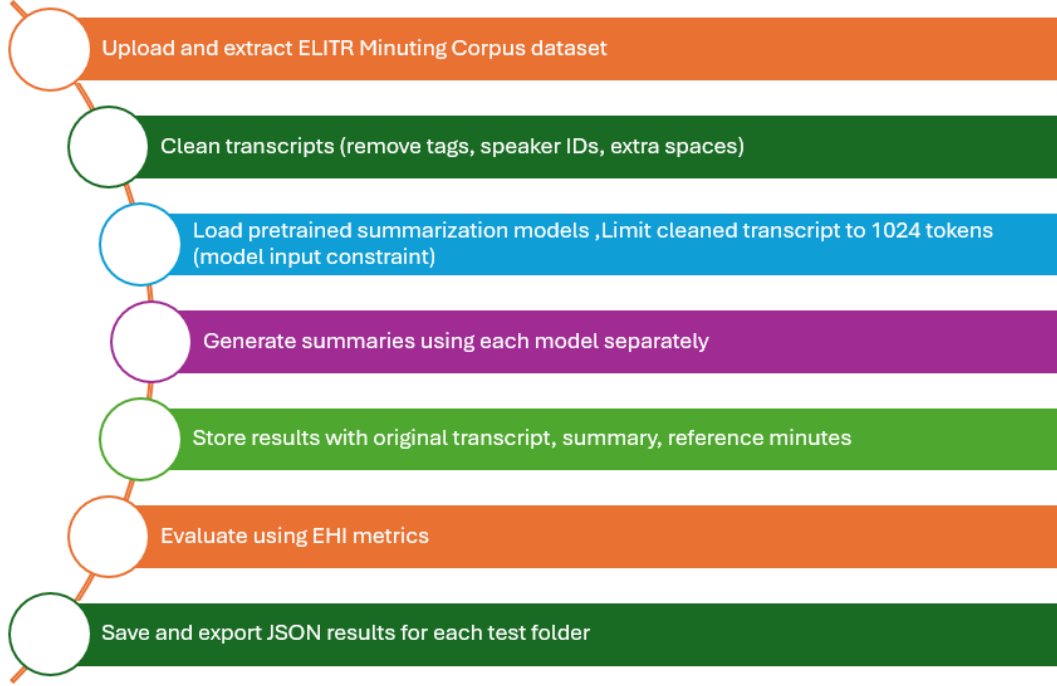- Save and export JSON results for each test folder

Figure 2: TASK B: Question-Answer Generation

$$\text{EHI} = \frac{e^{\text{PH}} + e^{\text{EF}}}{e^{\text{PH}} + e^{\text{EF}} + e^{\text{NH}} + e^{\text{OF}} + e^{\text{LF}}} \quad (1)$$

where $x_i \in \{\text{PH}, \text{EF}, \text{NH}, \text{OF}, \text{LF}\}$ represent scores corresponding to Positive Hallucination (PH), Extractiveness Factor (EF), Negative Hallucination (NH), Overfocused Relations (OF), and Lost Focus (LF), respectively. We define the following factors:

- **Positive Hallucination (PH):** Measures newly introduced entities that are factually correct and beneficial.

- **Extractiveness Factor (EF):** Measures entities accurately extracted from the input document into the summary.

- **Negative Hallucination (NH):** Captures hallucinated entities that are incorrect or not grounded in the input.

- **Overfocused Relations (OF):** Penalizes summaries that overly focus on a narrow subset of entities, missing diversity.

- **Lost Focus (LF):** Penalizes summaries that omit important entities present in the input.

Higher values of EHI indicate better entity faithfulness. While the term "Entity Hallucination Index" may suggest a higher hallucination rate, our formulation is designed such that EHI rewards desirable behaviors — namely, correctly extracted entities (EF) and factually accurate, beneficial additions (PH) — while penalizing undesirable behaviors such as ungrounded entities (NH), overfocus

(OF), and loss of important content (LF).

This index behaves more like a reward function, where a higher score means fewer harmful hallucinations and more helpful entity alignment. We retain the name EHI for continuity but note that it acts effectively as a precision-weighted reward. Note that when all error factors (NH, OF, LF) are zero, the EHI reduces to 1, regardless of PH and EF values. This behavior reflects that the summary is fully faithful, and additional correct entities (PH) or extractions (EF) do not reduce the score.

Entity extraction is performed using a named entity recognition (NER) model (en_core_web_sm from spaCy). Matching is case-insensitive and performed at the entity string level. As described in Section 3.4, higher EHI scores reflect better entity alignment and reduced hallucination indicating more entity-faithful summaries.

3.5 Reward-Driven Fine-Tuning

Our objective is to fine-tune the model parameters θ by maximizing the expected Entity Hallucination Index (EHI) reward across generated summaries. Formally, for an input document x, a generated summary y, and model parameters θ, the expected reward objective is defined as:

$$J(\theta) = \mathbb{E}_{y \sim p_\theta(y|x)}[\text{EHI}(y, x)] \qquad (2)$$

where $\text{EHI}(y, x)$ denotes the Entity Hallucination Index computed between the generated summary y and the input x.

Following the REINFORCE algorithm (Williams, 1992), the gradient of this objective is estimated as:

$$\nabla_\theta J(\theta) = \mathbb{E}_{y \sim p_\theta(y|x)}[\text{EHI}(y, x) \nabla_\theta \log p_\theta(y|x)] \qquad (3)$$

In practice, we sample summaries from the model, compute their corresponding EHI scores, and apply policy gradient updates to encourage higher-reward outputs. To stabilize training, we normalize EHI rewards within each batch and use the Adam optimizer with a small learning rate (e.g., 5×10^{-6}). Summaries are regenerated periodically to reflect model improvements during fine-tuning. We directly fine-tune the pre-trained model using reinforcement learning, without any intermediate supervised cross-entropy training. The training objective is to maximize the Entity Hallucination Index (EHI) reward. Following the REINFORCE algorithm (Williams, 1992), we estimate gradients

based on sampled summaries and use the Adam optimizer to update model parameters. EHI rewards are normalized within each training batch to ensure stable learning.

3.6 Evaluation Metrics

We evaluate model performance primarily using the **Entity Hallucination Index (EHI)**. For complementary analysis, we also report **Entity F1** scores, computed by matching extracted named entities between the generated summary and the reference summary. This provides an additional measure of summary completeness and precision at the entity level. Formally, let E_{ref} be the set of named entities in the reference summary and E_{gen} be the set of entities in the generated summary. Then:

$$\text{Precision} = \frac{|E_{ref} \cap E_{gen}|}{|E_{gen}|} \qquad (4)$$

$$\text{Recall} = \frac{|E_{ref} \cap E_{gen}|}{|E_{ref}|} \qquad (5)$$

$$\text{Entity F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (6)$$

All evaluations are conducted on the held-out test split. As part of Automin 2025, the content was available on evaluation for two task.

3.6.1 Evaluation Metrics for Task A

The system output has been evaluated by Automin 2025.ROUGE Scores: ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measures the overlap of textual units like n-grams between a generated text and one or more reference texts.These metrics usually report precision, recall, and their harmonic mean (F1-score) as evaluation scores.(Lin, 2004) ROUGE-1 (R1): Measures overlap of unigrams (single words) between the candidate and reference text. ROUGE-2 (R2): Measures overlap of bigrams (two-word sequences). ROUGE-L (RL): Measures the longest common subsequence (LCS) between the candidate and reference, capturing sentence-level structure and fluency. BERTScore:An advanced evaluation metric that uses BERT's contextual embeddings to measure semantic similarity between candidate and reference texts.Instead of exact token overlap, it compares token embeddings using cosine similarity, accounting for context and nuanced meaning.It calculates precision, recall, and F1 by aligning tokens from candidate and reference based on

Hyperparameter	Value
Model	Flan-T5-Large
Optimizer	Adam
Learning Rate	5×10^{-6}
Batch Size	Dynamic (based on 40GB GPU)
Sampling Strategy	Greedy Decoding
Reward Normalization	Applied per batch
Summary Regeneration	Every 500 updates
NER Tool	spaCy (en_core_web_sm)

Table 1: Hyperparameters and settings used during fine-tuning.

Metric	Before Fine-Tuning	After Fine-Tuning
EHI Range	$0.0 - 0.55+$	$0.3 - 0.6$
EHI Behavior	Volatile, inconsistent	Stable, moderate
Entity F1 Score	Mostly < 0.5, many zeros	Frequently 0.6–1.0
F1 Behavior	Irregular, poor precision	High accuracy and consistency
EHI–F1 Correlation	Weak/unclear	Strong inverse correlation

Table 2: Comparison of EHI and Entity F1 before and after reinforcement learning on meeting transcripts. Green-marked improvements are attributed to reward normalization and convergence effects of the Adam optimizer.

embedding similarity, thus better capturing paraphrasing and synonymy.BERTScore correlates better with human judgments compared to traditional metrics like ROUGE and BLEU.(Zhang* et al., 2020) BARTScore:A newer evaluation metric leveraging the BART model, a pretrained sequence-to-sequence transformer.It evaluates the likelihood of the generated text conditioned on the reference (or vice versa), providing a quality score based on model-based scoring rather than surface overlap.This method is particularly suited for text generation tasks like summarization and translation and has shown promising correlation with human evaluation (Yuan et al., 2021)

3.6.2 Evaluation Metrics for Task B

The evaluation has been conducted using GPT-4o as the LLM-judge with a maximum score of 10, following the methodology detailed in the ELITR-Bench paper (Thonet et al., 2024) and a mean score was provided for the responses.

3.7 Implementation Details

We initialize our experiments using the Flan-T5-Large model (Chung et al., 2022) as the base summarization system. Fine-tuning is conducted with the Adam optimizer (Kingma and Ba, 2015) and a learning rate of 5×10^{-6}. We train on meeting transcript datasets formatted as input-summary pairs, with batch sizes adjusted to

fit a single NVIDIA A100 GPU (40GB memory). Since Flan-T5-Large has a limited context window of 1024 tokens, we applied a chunking strategy to handle long meeting transcripts (often exceeding 10,000 tokens). Transcripts were split into overlapping chunks of up to 950 tokens with a stride of 200 tokens, ensuring contextual continuity. Each chunk was summarized independently, and final summaries were aggregated. This approach ensured compatibility with the model while preserving content fidelity across segments.

Summaries are generated using greedy decoding during both pre-fine-tuning and intermediate stages of training. EHI scores are computed by extracting named entities with spaCy's en_core_web_sm model (Honnibal and Montani, 2017), and matching is performed case-insensitively.

To ensure stability, we normalize EHI rewards within each training batch. Summaries are regenerated periodically every 500 updates to reflect model improvements in reward estimation. Fine-tuning is continued until EHI scores converge on a held-out validation set.

4 Results

4.1 Task A: Summarization

We evaluate the impact of EHI-guided fine-tuning on both hallucination control and entity prediction accuracy. Table 2 summarizes the comparative
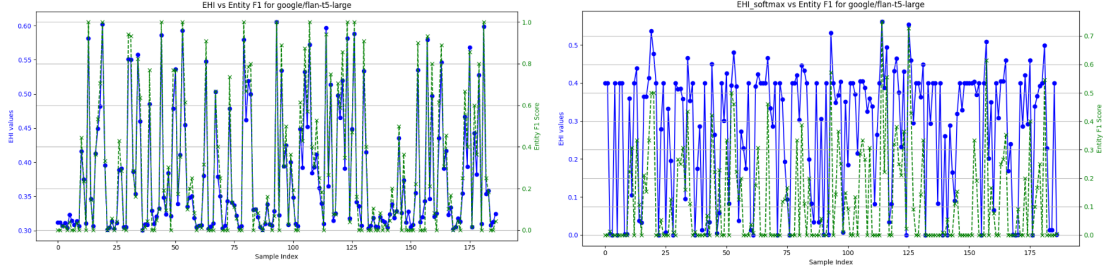
Figure 3: Comparison of EHI and Entity F1 before (left) and after (right) EHI-guided fine-tuning for `google/flan-t5-large`. Left: EHI and F1 are volatile and weakly correlated. Right: EHI and F1 stabilize with a clearer inverse correlation.

results before and after fine-tuning.

The results of our summarization experiments show that the employed transformer-based models were able to generate summaries of moderate to high quality across different meeting transcripts. Each model had varying levels of success in capturing key points and reducing irrelevant content.

Initially, the generated summaries showed instances of factual inconsistencies, especially with respect to named entities and relational information. This led to the use of our proposed hallucination evaluation metrics — Entity Hallucination Index (EHI) — which provided deeper insight into the factual alignment of generated content.

After applying domain-specific fine-tuning using the ELITR Minuting Corpus, we observed a notable qualitative improvement in summary generation. The summaries became:

- More aligned with the actual meeting minutes,

- Less prone to hallucinated entities or relationships,

- Structurally more concise and contextually richer.

The improvement can be attributed to better adaptation of the model to the task-specific discourse patterns and vocabulary present in meeting transcripts. Fine-tuning helped the model focus on core informative content while minimizing hallucinated information. This suggests that even lightweight models, when fine-tuned effectively, can significantly improve their factual reliability. Moreover, the proposed hallucination metrics proved effective in detecting and guiding the correction of factual errors in summaries.

4.1.1 Qualitative Examples with EHI Scores

Table 3 presents sample summaries before and after reinforcement learning, along with their EHI break-

down. We observe that the fine-tuned model avoids inserting hallucinated entities and better preserves entity grounding.

Further improvements can be achieved by:

- Utilizing more sophisticated language models(e.g., LLMs like GPT-4, FLAN-T5 XXL, or domain-specific variants),

- Incorporating multi-modal signals (e.g., speaker roles, dialogue acts),

- Leveraging reinforcement learning or contrastive learning frameworks to penalize hallucinations during training.

4.2 Task B: Question Answering

Speaker identifiers were retained during Task B processing, as they can provide important contextual cues for identifying the source of factual information in the transcript.

For Task B, we implemented a QA system based on pre-trained transformer models and evaluated its performance across multiple meeting scenarios. The system was able to answer a variety of factual and reasoning-based questions using the provided transcripts.

While the base model performed well on direct factual queries, its performance occasionally dropped in cases involving long-range dependencies or contextual ambiguity. However, overall performance was satisfactory for a zero-shot setup. A key observation is that improvements in summary quality (Task A) can also contribute to more efficient QA systems by helping condense large transcripts into manageable, high-information inputs. Although our QA system did not explicitly rely on summaries, the synergy between summarization and QA was evident in terms of context clarity and answer precision.

6

Before RL (EHI = 0.12)	After RL (EHI = 0.52)
The speaker discussed a new deal between IBM and Oracle to expand into Asian markets.	The speaker discussed the existing agreement with Oracle and Microsoft about service expansion.

Table 3: Example summaries before and after RL fine-tuning. The pre-RL summary includes a hallucinated entity ("IBM") not grounded in the transcript.

Approach	Mean Score
baseline_gpt-4o-2024-11-20	7.74
baseline_llama-3.1-8B-instruct	7.08
baseline_phi-4-mini-instruct	6.84
baseline_phi-3-small-128k-instruct	6.65
baseline_llama-3.2-3B-instruct	6.33
GETALP@AutoMin	4.55
GETALP@AutoMin_amr	4.34
GETALP@AutoMin_amr_only	2.73
HallucinationIndexes@AutoMin (ours)	**2.28**

Table 4: Mean scores of systems on the Monolingual (English-only) summarization task in AutoMin 2025.

The table 4 depicts the results. Note that Automin evaluation considered various baselines, however the model Flan-T5-Large baseline was not considered. The scores in the evaluation had two limitations: (1)Note that Automin evaulation considered various baselines, however the model Flan-T5-Large baseline was not considered. (2) The hallucination index improves the quality focusing merely on entities and a subsequent method of Relation Hallucination is needed to complement it.

4.3 Entity Hallucination Index (EHI) Improvements

Before fine-tuning, EHI scores exhibited significant volatility, with frequent drops to 0.0, indicating frequent hallucinations. After fine-tuning, EHI scores became more consistent, largely stabilizing between 0.3 and 0.6, suggesting improved control over hallucinated entities. The

4.4 Entity F1 Score Improvements

Entity F1 scores improved markedly after fine-tuning. While initial F1 scores were often below 0.5 with many zeros (indicating missed entity matches), fine-tuning led to a surge in F1 scores, with numerous samples achieving values close to 1.0, reflecting high precision and recall for entity prediction.

4.5 Correlation Between EHI and Entity F1

Prior to fine-tuning, the relationship between EHI and F1 was weak or inconsistent. Post fine-tuning, we observe a stronger inverse correlation, indicating that as entity hallucinations decreased (higher EHI), entity prediction accuracy (F1) increased. Figure 3 provides a visual comparison of EHI and Entity F1 scores across samples. Before fine-tuning, both metrics fluctuate significantly, indicating unstable entity prediction and hallucination control. After fine-tuning, the curves are more stable and exhibit clearer inverse correlation, validating the effectiveness of EHI-guided optimization.

5 Discussion

Beyond quantitative improvements, we conduct qualitative analysis to better understand how EHI-guided fine-tuning impacts summary generation.

Improved Entity Grounding. Before fine-tuning, generated summaries frequently introduced entities unrelated to the input transcript, or omitted key entities entirely. After fine-tuning, summaries exhibited better alignment with the input, correctly preserving mentioned organizations, speaker names, and events. We observe that entity mentions became more precise and contextually appropriate.

Enhanced Consistency Across Summaries. Fine-tuned models enhances EHI score which helps

consistent entity coverage across samples. Hallucination behavior, which was previously erratic and dataset-dependent, stabilized substantially after training with EHI rewards as explained in 3. Overall summaries demonstrated less variance in entity recall and fewer abrupt entity shifts.

Limitations and Error Cases. While overall hallucination was reduced, occasional errors persisted, especially for rare or ambiguous entity mentions. In some cases, the model overly prioritized exact entity copying at the expense of paraphrasing or abstraction quality. This suggests a trade-off between strict entity grounding and higher-level semantic fluency that merits further investigation. This is where there is a need for Relation Hallucination detection and handling which the EHI is lacking.

Interpretation of EHI Improvements. The use of exponential scaling in the EHI formula emphasizes the balance between positive extraction and penalizing hallucination, which contributed to stable training dynamics. Moreover, EHI has a scope to be used readily in reference free context by combining with Relative Abstractiveness and generating EHI scores on reference data itself and use it as threshold for zero shot reference less summaries to predict hallucination.

6 Conclusion

In this work, we proposed a reward-driven fine-tuning approach for abstractive summarization, leveraging the Entity Hallucination Index (EHI) to directly optimize for entity faithfulness. By integrating EHI as a lightweight, reference-free reward signal, we reduced hallucinations and improved entity prediction accuracy without relying on human factuality annotations.

Experiments on meeting transcript datasets demonstrate that EHI-guided fine-tuning leads to more stable hallucination behavior and significant gains in entity F1 scores. Our results suggest that fine-grained entity-focused rewards provide an effective pathway for enhancing summarization reliability.

In future work, we plan to extend EHI-based fine-tuning to multi-document summarization and explore its integration with controllable generation frameworks.

References

Shuyang Cao, Lu Wang, and Xiaojun Wan. 2020. Factual error correction for abstractive summarization models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Xuezhi Huang, Andrew M. Dai, and 1 others. 2022. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*.

Esin Durmus, He He, and Mona Diab. 2020. Feqa: A question answering evaluation framework for faithfulness assessment in abstractive summarization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Tobias Falke, Leonardo F. R. Ribeiro, and Iryna Gurevych. 2019. Ranking generated summaries by correctness: An interesting but hard task. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Tirthankar Goyal, Esin Durmus, and Claire Cardie. 2022. Fact-checking summarization with self-supervised and few-shot learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Matthew Honnibal and Ines Montani. 2017. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. https://spacy.io.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Sandrine Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Artidoro Pagnoni, Vishakh Padmakumar, and Jonathan May. 2021. Understanding factuality in abstractive summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Ramakanth Pasunuru and Mohit Bansal. 2018. Multi-reward reinforced summarization with saliency and entailment. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. 2021. Questeval: Summarization asks for fact-based evaluation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Thibaut Thonet, Jos Rozen, and Laurent Besacier. 2024. ELITR-Bench: A Meeting Assistant Benchmark for Long-Context Language Models. *arXiv:2403.20262*.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Weizhe Yuan, Graham Neubig, and Pengfei Liu. 2021. Bartscore: evaluating generated text as text generation. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA. Curran Associates Inc.

Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.