Reducing Hallucinations in Abstractive Summarization via Verifier‑Reranking

Proposal

Abstract summarizers often produce fluent but unsupported statements that limit practical use. This project targets factuality for news summarization with a simple, reproducible approach. I will fine‑tune BART-base on the public CNN/DailyMail dataset. At inference, I will generate K candidate summaries per article and rerank them using an automatic factuality score. Using **FactCC** as the primary verifier and reporting **QAGS** as a secondary check. If reranking harms ROUGE, I will test constrained decoding that favors copying from the source as a fallback. Implementations rely on Hugging Face Transformers, Datasets, and Evaluate, and will be cited.

Dataset

CNN/DailyMail (Hugging Face ID: **ccdv/cnn_dailymail, config 3.0.0**) with standard train/validation/test splits.

Metrics and Success

Quality: ROUGE‑1/2/L and BERTScore on validation and test.

Factuality: FactCC (primary) and QAGS on validation.

Human check: 50 examples with an error taxonomy.

Success criterion: ≥ +2.0 FactCC points over the BART baseline on validation with ≤ 1.0 ROUGE‑L drop.

Risks / Challenges

Verifier may mis‑score paraphrases; mitigate with human spot checks. Compute limits handled by base‑size models and modest K. If reranking underperforms, tune decoding and report ablations.

Why this matters

Abstractive models often sound good but insert mistakes. Picking the most factual draft reduces those hallucinations without heavy engineering.

This focused design addresses a known failure mode with minimal engineering and clear metrics. It fits the class scope, uses a single public dataset with stable splits, and supports deep analysis. I will release code, configs, and small data samples for reproducibility.

References

- **BART (Lewis et al., 2020, ACL)**
  ACL Anthology page: https://aclanthology.org/2020.acl-main.703/
  PDF: https://aclanthology.org/2020.acl-main.703.pdf

- **PEGASUS (Zhang et al., ICML 2020)**
  ICML / MLR page: https://proceedings.mlr.press/v119/zhang20ae.html
  PDF: https://proceedings.mlr.press/v119/zhang20ae/zhang20ae.pdf
  Project page: https://jingqingz.github.io/publication/2019-PEGASUS

- **FactCC (Kryściński et al., EMNLP 2020)**
  Anthology (paper “Evaluating the Factual Consistency of Abstractive Text Summarization”): https://aclanthology.org/2020.emnlp-main.750/
  PDF: https://aclanthology.org/2020.emnlp-main.750.pdf

- **QAGS (Wang et al., ACL 2020)**
  ACL Anthology page: https://aclanthology.org/2020.acl-main.450/
  PDF: https://aclanthology.org/2020.acl-main.450.pdf

- **BERT (Devlin et al., NAACL-HLT 2019)**
  ArXiv: https://arxiv.org/abs/1810.04805
  Conference reference: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL-HLT 2019.