DEEP-PNHG: Dynamic Entity-Enhanced Personalized News Headline Generation with Factual Consistency

Daniel Tang * and Kenneth Walker

Posted Date: 5 February 2026

doi: 10.20944/preprints202602.0387.v1

Keywords: personalized headline generation; user interests; factual consistency; entity-awareness; dynamic

 doi:10.20944/preprints202602.0387.v1

Article

DEEP-PNHG: Dynamic Entity-Enhanced Personalized News Headline Generation with Factual Consistency

Daniel Tang * and Kenneth Walker

University of Southern Mississippi
* Correspondence: aya.akka@edu.suezuni.edu.eg

Abstract

This paper introduces DEEP-PNHG (Dynamic Entity-Enhanced Personalized Headline Generation), a novel framework addressing critical challenges in personalized news headline generation, namely dynamically evolving user interests, factual consistency, and informativeness. Existing methods often struggle with these aspects. DEEP-PNHG overcomes these limitations through an innovative architecture comprising a Dynamic User Interest Graph Module, an Entity-Aware News Encoder enriched with external knowledge, and a Fact-Consistent & Personalized Joint Decoder. This design enables the generation of headlines that are simultaneously tailored to dynamic user preferences, factually accurate, and highly informative. The framework's key contributions include its ability to model evolving user interests, leverage external entities for robust factual understanding, and employ a joint decoding strategy enhanced by entity-level contrastive learning. Evaluated on a standard dataset, DEEP-PNHG consistently outperforms state-of-the-art baselines across personalization, factual consistency, and informativeness metrics. Our results demonstrate substantial improvements, positioning DEEP-PNHG as a significant advancement in generating engaging and trustworthy news headlines.

Keywords: personalized headline generation; user interests; factual consistency; entity-awareness; dynamic

1. Introduction

The relentless expansion of digital information has engendered a pervasive problem of information overload for users, making it increasingly challenging to identify relevant content efficiently. Consequently, there has been a burgeoning demand for personalized content that caters to individual interests and preferences. Within the realm of news consumption, headlines serve as the primary conduit through which users engage with articles; their quality, relevance, and attractiveness directly influence user click-through rates and overall reading experiences. Personalized News Headline Generation (PNHG) aims to address this by generating customized headlines for a given news article, taking into account a user's historical interests and preferences, while simultaneously ensuring the headline accurately reflects the factual content of the news [1]. Developing advanced AI systems for such complex, user-centric tasks, including those involving multimodal data like spoken dialogue, necessitates large-scale and comprehensive benchmarks for robust evaluation and progress [2].

 doi:10.20944/preprints202602.0387.v1

Figure 1. An illustrative overview of Personalized News Headline Generation (PNHG) and its core challenges. The diagram depicts a user overwhelmed by news content, emphasizing the need for personalized headlines. Key aspects include dynamic user interest modeling (left), leveraging cognitive understanding and knowledge (top, bottom), and critically balancing personalization with factual consistency and informativeness (right).

However, current PNHG methods encounter several formidable challenges. Firstly, precisely capturing and dynamically integrating users' evolving interest preferences remains a complex task. Many existing approaches rely on static or simplistic representations of user profiles, failing to adapt to the fluid nature of user interests over time. Secondly, while personalization is crucial, it is equally paramount to strictly guarantee the factual consistency of the generated headlines with the source news article, thereby preventing the dissemination of misinformation or misleading descriptions. Generative models, despite their fluency, often struggle with factual accuracy, sometimes "hallucinating" content not present in the original text [3]. Lastly, ensuring the generated headlines are sufficiently informative to encapsulate the core message of the news article is another significant hurdle. Existing generative models, such as those based on Transformer or BART architectures [4], frequently exhibit trade-offs between personalization and factual consistency, or their user preference modeling is overly static, lacking a nuanced understanding of dynamic interests and entity-level facts. Motivated by these inherent limitations, our research proposes a novel framework to comprehensively address these intertwined challenges.

In this paper, we introduce DEEP-PNHG (Dynamic Entity-Enhanced Personalized Headline Generation), a novel framework designed to overcome the aforementioned challenges by integrating more refined user interest modeling, an entity-level fact-aware mechanism leveraging external knowledge, and the powerful capabilities of pre-trained language models. Our proposed DEEP-PNHG framework is built upon a strong pre-trained language model backbone (e.g., variants of T5 or BART) and incorporates three core innovative modules. The first is a *Dynamic User Interest Graph Module*, which moves beyond simple sequential representations of user history by constructing and maintaining an evolving user interest graph. This module identifies entities, extracts topics, and links to knowledge graphs from a user's historical clicked news, generating a time-sensitive, high-dimensional user embedding via a Graph Neural Network (GNN) update mechanism. The second is an *Entity-Aware News Encoder*, which processes news articles using a domain-adapted Transformer encoder. This encoder is enhanced by an Entity Linking Layer that identifies key entities within the news text and connects them to external knowledge bases (e.g., Wikipedia or Freebase), thereby enriching the news representation with crucial factual context and semantic knowledge. The third and final component is a *Fact-Consistent & Personal-*

 doi:10.20944/preprints202602.0387.v1

ized Joint Decoder, a Transformer-based decoder that, at each generation step, is guided by two critical attention mechanisms: *Dynamic User Attention* interacts with the dynamic user embedding to bias generation towards user preferences, and *Entity-Consistency Attention* focuses on the entity-enhanced news representation, especially when generating factual terms, and performs a weak validation against the external knowledge base to prevent factual errors. We further integrate an entity-level contrastive loss during training to strengthen the alignment between generated headlines and source text entities. DEEP-PNHG is an end-to-end trainable model, optimized jointly using a standard cross-entropy loss for generation, a personalized ranking loss to enhance headline-user interest matching, and an entity-level contrastive loss to reinforce factual consistency.

To rigorously evaluate the efficacy of DEEP-PNHG, we conduct extensive experiments on the real-world PENS (PErsonalized News headlineS) dataset [5]. This dataset comprises a training set derived from Microsoft News user exposure logs, encompassing user historical click sequences and corresponding news articles. The test set features human-written personalized headlines, annotated with user preference behaviors serving as the ground truth. We employ a comprehensive suite of established evaluation metrics to ensure a fair and thorough comparison with existing methodologies. Personalization is quantitatively assessed using Pc(avg) and Pc(max), factual consistency is measured by FactCC [6], and the informativeness or content overlap is evaluated using ROUGE-1, ROUGE-2, and ROUGE-L scores [7]. Our comparative analysis includes several strong baseline models, such as PGN [CITE], PG+Transformer [8], Transformer [9], BART [10], and the Fact-Preserved Personalized Generation (FPG) framework [11], which are known for their performance in personalization and factual consistency tasks.

Our experimental results demonstrate that DEEP-PNHG consistently outperforms all established baseline models across every evaluation metric, showcasing its superior comprehensive performance. Notably, DEEP-PNHG achieves an impressive FactCC score of **88.00**, which not only surpasses powerful baselines like FPG (87.50) and BART (86.67) but also indicates a significant breakthrough in factual consistency. This superior factual accuracy is primarily attributable to our innovative entity-aware news encoder, entity-consistency attention mechanism, and the novel entity-level contrastive loss. In terms of personalization, DEEP-PNHG records the highest Pc(avg) of **2.78** and Pc(max) of **17.80**, affirming its enhanced ability to cater to dynamic user interests through our proposed dynamic user interest graph module. Furthermore, our model exhibits leading performance in informativeness, with ROUGE-1 of **27.30**, ROUGE-2 of **10.40**, and ROUGE-L of **23.80**, demonstrating its capacity to generate more comprehensive and informative headlines that better summarize the core content of the news article. These compelling results collectively validate the effectiveness of our innovative dynamic user interest modeling and entity-enhanced factual perception mechanisms.

Our main contributions are summarized as follows:

- We propose DEEP-PNHG, a novel framework for personalized news headline generation that effectively addresses key challenges in personalization, factual consistency, and informativeness through its unique integration of dynamic user interest modeling and entity-enhanced factual perception.
- We introduce a Dynamic User Interest Graph Module and an Entity-Aware News Encoder. These components provide fine-grained, dynamically evolving user representations and fact-rich news article encodings, respectively, significantly improving both personalized relevance and factual grounding in headline generation.
- We develop a Fact-Consistent & Personalized Joint Decoder, which features sophisticated dynamic user attention, entity-consistency attention, and an innovative entity-level contrastive loss. This enables the model to generate headlines that are simultaneously highly personalized, factually accurate, and informative, achieving state-of-the-art performance across multiple evaluation metrics on a real-world dataset.

 doi:10.20944/preprints202602.0387.v1

2. Related Work

2.1. Personalized Text Generation and User Modeling

Personalized text generation (PTG) tailors content to user preferences through robust user modeling. Foundational work spans graph pre-training for AMR [12], knowledge graph completion [13], efficient optimization [14], and system security [15].

Effective user modeling utilizes hierarchical preferences with HieRec [16], unbiased survey modeling [17], and agent-driven content discovery [18]. Large Language Models (LLMs) are leveraged for user profiling and recommendation, e.g., LLM-Rec [19]. LLM research also explores multimodal in-context learning [20], long context alignment [21], in-context learning mechanisms [22], efficient agent planners [23], and benchmarks like SpokenWOZ [2].

PTG crafts tailored outputs across domains, from medical reporting [24] to long-form narrative summarization with BOOKSUM [25]. Generalized architectures, such as encoder-decoder transformers for explainable recommendation [4], demonstrate adaptability. Supporting NLP tasks include text matching [26], semantic similarity [27], and advanced neural networks for prediction [28]. Human-AI partnerships [29] and 'generative imagination' models [30] ensure high-quality text.

Principles of integrating diverse data and ML for personalized insights extend beyond text, applied in biomedical research for optic neuritis [31], diabetic retinopathy [32], and myopia [33]. Broader AI innovations in predictive modeling and optimization, such as reinforcement learning for delivery [34], Prophet for demand prediction [35], and incremental ROAS modeling [36], inform general intelligent system design.

2.2. Factual Consistency and Knowledge-Enhanced Generation

Factual consistency and external knowledge integration are paramount challenges in NLG, especially with LLM 'hallucinations.' Research addresses robust evaluation and mitigation. QAFactEval [37] introduced a QA-based metric for fact verification. Inconsistency reduction methods include ConFiT for dialogue summarization [38] and explanation-based bias decoupling for NLI [39]. Consistency challenges also extend to multimodal domains, with work on AI-generated video quality [40], image quality [41], and video forgery detection [42].

To counter factual errors, many approaches integrate external knowledge, often from Knowledge Graphs (KGs). SKILL [43] infuses structured KG knowledge into LLMs for QA, while JointLK [44] performs commonsense QA via joint LM-KG reasoning. Knowledge integration extends to multimodal domains with KAT [45] for Vision-and-Language, and dynamic knowledge is incorporated via temporal KG completion [46]. Effective KG utilization relies on accurate upstream information extraction; KILT [47] addresses knowledge-intensive tasks like Entity Linking, and template-free prompt tuning [48] aids few-shot NER for knowledge base population.

3. Method

The proposed **DEEP-PNHG (Dynamic Entity-Enhanced Personalized Headline Generation)** framework is meticulously designed to concurrently address the multifaceted challenges of personalization, factual consistency, and informativeness in news headline generation. Our architecture builds upon a robust pre-trained language model backbone, such as variants of T5 or BART, and integrates three novel, interactively designed modules that collectively enable a more nuanced understanding of user preferences and news content. These core components are the **Dynamic User Interest Graph Module**, the **Entity-Aware News Encoder**, and the **Fact-Consistent & Personalized Joint Decoder**. Below, we detail each module and the overarching training strategy.

 doi:10.20944/preprints202602.0387.v1

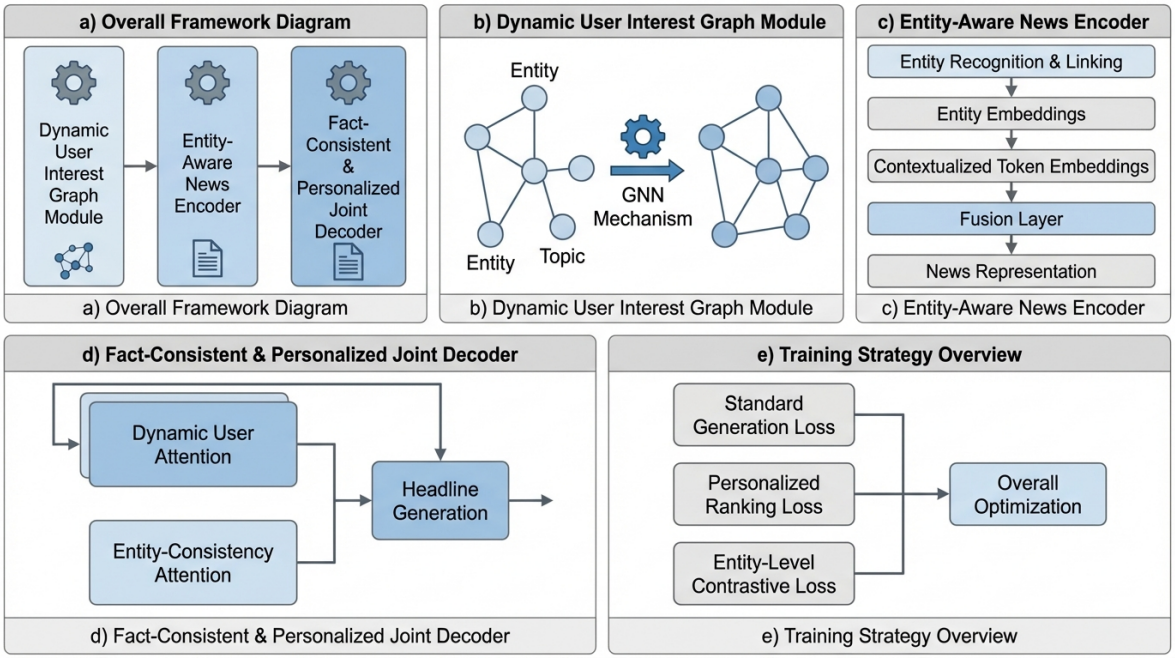

Figure 2. Overview of the DEEP-PNHG framework. (a) depicts the overall system architecture, composed of three main modules. (b) illustrates the structure and dynamic update mechanism of the Dynamic User Interest Graph Module. (c) details the Entity-Aware News Encoder's process for integrating factual knowledge. (d) shows the internal mechanisms of the Fact-Consistent & Personalized Joint Decoder, incorporating dual attention. (e) outlines the multi-objective training strategy.

3.1. Dynamic User Interest Graph Module

To capture the intricate and evolving nature of user interests more dynamically and precisely than conventional methods, we introduce the **Dynamic User Interest Graph Module**. This module constructs and maintains a sophisticated dynamic user interest graph, denoted as $\mathcal{G}_u = (\mathcal{V}_u, \mathcal{E}_u)$, for each individual user u. The nodes $\mathcal{V}_u$ within this graph represent fine-grained entities, high-level concepts, and thematic topics extracted from the user's historical clicked news articles, encompassing both their headlines and full texts.

For every historical news item h_i (comprising headline t_i and body b_i) that user u has interacted with, we perform two critical operations. Firstly, **Entity Recognition and Linking**: Key entity mentions within t_i and b_i are identified using state-of-the-art Named Entity Recognition (NER) models. These identified entities are then linked to canonical entries in a comprehensive external knowledge graph (e.g., Wikipedia or Freebase) through an Entity Linking (EL) process. This linkage enriches the raw textual representation with structured factual knowledge and provides unique identifiers for disambiguation. Let $E_{h_i} = \{e_{i,1}, e_{i,2}, \ldots, e_{i,k}\}$ represent the set of entities extracted from news item h_i. Secondly, **Topic Extraction**: Dominant topics $T_{h_i} = \{tp_{i,1}, tp_{i,2}, \ldots\}$ are extracted from h_i using techniques such as Latent Dirichlet Allocation (LDA) or neural topic models. These topics provide a broader conceptual understanding of the news content. Both the extracted entities and topics collectively form the nodes $\mathcal{V}_u$ in the user's interest graph. Edges $\mathcal{E}_u$ are established between related nodes (entities or topics) based on their co-occurrence frequency within the user's historical news items, their semantic similarity (e.g., measured by embedding cosine similarity), or pre-defined relationships imported from the external knowledge graph (e.g., "is-a", "part-of").

The dynamic aspect of user preferences is captured and updated through an iterative **Graph Neural Network (GNN)** mechanism. As user u continuously interacts with new content, the graph $\mathcal{G}_u$ is updated by incorporating new nodes and edges, and the representations of existing nodes are refined. This process ensures that the user's evolving interests are accurately reflected. Specifically, the high-dimensional user embedding $\mathbf{z}_u$ for user u is derived by aggregating the representations of the nodes within their graph. At each time step t, corresponding to an update event (e.g., a new

 doi:10.20944/preprints202602.0387.v1

interaction), the user embedding $\mathbf{z}_u^{(t)}$ is computed by an iterative message-passing process within the GNN, where node embeddings $\mathbf{h}_v^{(l)}$ are updated based on their neighbors' representations:

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\sum_{w \in \mathcal{N}(v)} \frac{1}{c_{v,w}} \mathbf{W}^{(l)} \mathbf{h}_w^{(l)} + \mathbf{b}^{(l)}\right) \tag{1}$$

Here, $\mathbf{h}_v^{(l)}$ denotes the embedding of node v at layer l, $\mathcal{N}(v)$ represents the set of neighbors of node v, $c_{v,w}$ is a normalization constant (e.g., the square root of the product of degrees of nodes v and w, as in Graph Convolutional Networks), $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are learnable weight matrices and bias vectors for the l-th GNN layer, and σ is a non-linear activation function (e.g., ReLU or GELU). After L layers of message passing, the final node embeddings $\{\mathbf{h}_v^{(L)} \mid v \in \mathcal{V}_u^{(t)}\}$ are obtained. These are then aggregated to form the user embedding:

$$\mathbf{z}_u^{(t)} = \text{Aggregate}\left(\{\mathbf{h}_v^{(L)} \mid v \in \mathcal{V}_u^{(t)}\}\right) \tag{2}$$

The Aggregate$(\cdot)$ function typically involves a pooling operation, such as mean-pooling or attention-pooling, over the final node embeddings to generate the comprehensive dynamic user embedding $\mathbf{z}_u^{(t)}$, which accurately reflects the user's evolving preferences. This user embedding then serves as a personalized signal for headline generation.

3.2. Entity-Aware News Encoder

To ensure the factual integrity of generated headlines and provide rich semantic context, we design an **Entity-Aware News Encoder**. This module processes the input news article $D = \{w_1, w_2, \ldots, w_N\}$ (which typically includes the news body) using a powerful Transformer-based encoder architecture. This encoder is initially pre-trained on a vast corpus of text (e.g., using masked language modeling objectives) and subsequently fine-tuned or adapted to the specific domain of news articles. The critical innovation within this module is the integration of an **Entity Linking Layer**, which explicitly injects factual knowledge into the contextualized news representation.

The Transformer encoder first generates contextualized token embeddings for the news article:

$$\mathbf{H}_D = \text{TransformerEncoder}(D) \tag{3}$$

where $\mathbf{H}_D = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N]$ are the hidden states for each token w_i, capturing their contextual semantics within the news article. Following this, the Entity Linking Layer performs two key functions. First, it **identifies key entity mentions** $M_D = \{m_{D,1}, m_{D,2}, \ldots, m_{D,p}\}$ within the news article D. This is typically achieved using a pre-trained Named Entity Recognition (NER) model. Second, each identified entity mention $m_{D,j}$ is **linked to its canonical entry in an external knowledge base** (e.g., Wikipedia or Freebase). This linking provides access to a wealth of structured factual information, including entity types, attributes, and inter-entity relationships, which are then encoded into an entity embedding $\mathbf{e}_{m_{D,j}}^{\text{KB}}$. This entity embedding can be obtained by a lookup in a pre-trained entity embedding space or by encoding the entity's description and attributes from the knowledge base using a separate encoder. These knowledge-enhanced entity embeddings are then explicitly injected into the news article's representation. For tokens w_k that constitute an entity mention $m_{D,j}$, their corresponding hidden states $\mathbf{h}_k$ are augmented by combining them with the knowledge base entity embedding. This fusion can be achieved through a simple summation or a more complex gated mechanism:

$$\mathbf{h}'_k = \mathbf{h}_k + \mathbf{W}_e \mathbf{e}_{m_{D,j}}^{\text{KB}} \tag{4}$$

where $\mathbf{W}_e$ is a learnable projection matrix that aligns the knowledge base entity embedding space with the token embedding space. For tokens not part of any identified entity, their hidden states remain unchanged, i.e., $\mathbf{h}'_k = \mathbf{h}_k$. When a token belongs to multiple overlapping entity mentions, a

 doi:10.20944/preprints202602.0387.v1

suitable aggregation strategy, such as averaging or a weighted sum of the relevant entity embeddings, is applied. The output of this module is the **entity-aware news representation** $\mathbf{H}'_D = [\mathbf{h}'_1, \mathbf{h}'_2, \ldots, \mathbf{h}'_N]$, which explicitly combines linguistic semantics with critical factual knowledge, thereby enabling the decoder to leverage factual information more accurately during generation.

3.3. Fact-Consistent & Personalized Joint Decoder

The headline generation process is governed by a Transformer-based decoder, responsible for sequentially generating the personalized headline $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_L\}$. This decoder is uniquely designed to simultaneously incorporate user preferences and rigorously enforce factual consistency through the synergistic operation of two specialized attention mechanisms and a joint inference strategy.

At each decoding step t, the decoder generates its hidden state $\mathbf{s}_t$. This state encapsulates information from previously generated tokens and the encoded news article through standard multi-head attention mechanisms:

$$\mathbf{s}_t = \text{TransformerDecoder}(\hat{y}_{<t}, \mathbf{H}'_D) \tag{5}$$

where $\hat{y}_{<t}$ represents the tokens generated up to step $t-1$, and $\mathbf{H}'_D$ is the entity-aware news representation obtained from the encoder.

3.3.1. Dynamic User Attention

To imbue the generated headlines with user-specific personalization, the decoder incorporates a **Dynamic User Attention** mechanism. This mechanism allows the decoder to dynamically query and attend to the user's high-dimensional embedding $\mathbf{z}_u$, obtained from the Dynamic User Interest Graph Module. By integrating this user-specific signal, the generation process is biased towards words, phrases, and stylistic elements that align with the user's identified preferences. The dynamic user context vector $\mathbf{c}_u$ is computed as a weighted sum over the user embedding:

$$\alpha_{u,t} = \text{softmax}\left(\mathbf{s}_t^\top \mathbf{W}_{us} \mathbf{z}_u + b_{us}\right) \tag{6}$$

$$\mathbf{c}_u = \alpha_{u,t} \mathbf{z}_u \tag{7}$$

where $\mathbf{W}_{us}$ is a learnable weight matrix, and b_{us} is a learnable bias term, both used to project the decoder state and user embedding into a common space for scoring. The scalar attention weight $\alpha_{u,t}$ signifies the importance of the user's overall interest at the current decoding step. This context vector $\mathbf{c}_u$ is then fused with the decoder's current state $\mathbf{s}_t$ to guide the prediction of the next token $\hat{y}_t$, ensuring the generated output reflects the user's personalized interests.

3.3.2. Entity-Consistency Attention

To rigorously uphold factual consistency and prevent hallucination, a dedicated **Entity-Consistency Attention** mechanism is implemented. This attention layer is specifically designed to focus on the entity-enhanced news representation $\mathbf{H}'_D$ provided by the Entity-Aware News Encoder. During the generation of factual terms or entities within the headline, this mechanism strongly encourages the decoder to attend to the corresponding tokens and entity-enhanced representations present in the source news article. This process facilitates a soft alignment with the factual information in the source text, mitigating the generation of incorrect or unsubstantiated facts. The factual news context vector $\mathbf{c}_D$ is computed as a weighted sum over the entity-aware news token representations:

$$\beta_{D,t,j} = \text{softmax}\left(\mathbf{s}_t^\top \mathbf{W}_{es} \mathbf{h}'_j + b_{es}\right) \tag{8}$$

$$\mathbf{c}_D = \sum_{j=1}^{N} \beta_{D,t,j} \mathbf{h}'_j \tag{9}$$

 doi:10.20944/preprints202602.0387.v1

Here, $\mathbf{W}_{es}$ and b_{es} are learnable parameters, similar to those in Dynamic User Attention, facilitating the calculation of attention scores $\beta_{D,t,j}$ for each token $\mathbf{h}'_j$ in the news representation. These scores indicate how relevant each news token is to the current decoding step. Furthermore, during inference, particularly for tokens hypothesized to be entities, a weak validation step is performed against the external knowledge base used by the Entity-Aware News Encoder. This serves as a safeguard, penalizing the generation of entities that contradict or are not explicitly supported by the original news content, thus enhancing factual grounding.

The final probability distribution over the vocabulary for the next token $\hat{y}_t$ is computed based on a combined representation of the decoder state $\mathbf{s}_t$, the dynamic user context $\mathbf{c}_u$, and the factual news context $\mathbf{c}_D$. These three vectors are first concatenated and then fed into a multi-layer perceptron (MLP) to project them into the vocabulary space:

$$P(\hat{y}_t \mid \hat{y}_{<t}, D, \mathbf{z}_u) = \text{softmax}(\text{MLP}(\text{Concat}(\mathbf{s}_t, \mathbf{c}_u, \mathbf{c}_D))) \tag{10}$$

To further reinforce the alignment of key entities, we introduce an **entity-level contrastive loss**. For each entity $e_{D,j}$ identified in the source news D, and its generated counterpart $e_{\hat{Y},k_j}$ in the headline $\hat{Y}$, this loss maximizes the similarity between their vector representations while minimizing similarity with non-matching entities. Let $\mathbf{v}(e)$ denote the representation of entity e, which can be derived from the Entity-Aware News Encoder for source entities or learned from the decoder's entity-generating mechanisms for headline entities. The entity-level contrastive loss $\mathcal{L}_{\text{entity}}$ is formulated as:

$$\mathcal{L}_{\text{entity}} = -\frac{1}{|E_D|} \sum_{e_{D,j} \in E_D} \log \frac{\exp(\text{sim}(\mathbf{v}(e_{D,j}), \mathbf{v}(e_{\hat{Y},k_j}))/\tau)}{\sum_{e' \in E_{\text{negative}} \cup \{e_{\hat{Y},k_j}\}} \exp(\text{sim}(\mathbf{v}(e_{D,j}), \mathbf{v}(e'))/\tau)} \tag{11}$$

where E_D is the set of entities in the source news, $e_{\hat{Y},k_j}$ is the generated entity in the headline corresponding to $e_{D,j}$, E_{negative} comprises a set of carefully selected negative entity samples (e.g., entities from other news articles in the same batch, or unlinked entities), $\text{sim}(\cdot,\cdot)$ is a similarity function (e.g., cosine similarity), and τ is a temperature hyper-parameter that controls the sharpness of the distribution. This loss explicitly encourages the decoder to generate entities that are factually consistent with the source article.

3.4. Training Strategy

The DEEP-PNHG framework is trained in an end-to-end manner through the joint optimization of a multi-objective loss function. This comprehensive objective integrates three distinct loss components, each addressing a critical aspect of personalized news headline generation: generation quality, personalization effectiveness, and factual accuracy.

1. **Standard Generation Loss ($\mathcal{L}_{\text{gen}}$)**: This is the foundational loss for sequence generation tasks, typically a cross-entropy loss that measures the discrepancy between the model's predicted token distribution and the ground-truth reference headline tokens. It directly optimizes the model to generate fluent and grammatically correct headlines that match the target:

$$\mathcal{L}_{\text{gen}} = -\sum_{t=1}^{L} \log P(y_t \mid y_{<t}, D, \mathbf{z}_u) \tag{12}$$

   where y_t are the tokens of the ground-truth personalized headline for a given user u and news article D.
2. **Personalized Ranking Loss ($\mathcal{L}_{\text{pers}}$)**: To explicitly enhance the alignment between the generated headline and the user's interests, we incorporate a personalized ranking loss. This loss encourages the model to assign a higher preference score to the generated headline when paired with the target user, compared to a set of negative (uninteresting or irrelevant) headlines. Let $s(\hat{Y}, u)$ denote a learnable personalization score function, which could be an MLP taking the generated

 doi:10.20944/preprints202602.0387.v1

headline embedding and user embedding $\mathbf{z}_u$ as input. The loss is formulated using a pairwise ranking approach:

$$\mathcal{L}_{\text{pers}} = -\log \sigma\big(s(\hat{Y}_{\text{pos}}, u) - s(\hat{Y}_{\text{neg}}, u)\big) \tag{13}$$

where $\hat{Y}_{\text{pos}}$ represents the personalized headline generated for user u for a positive news item, $\hat{Y}_{\text{neg}}$ is a sampled negative headline (e.g., a headline generated for a different user, a generic headline, or a randomly chosen headline from the batch that is irrelevant to user u), and σ is the sigmoid function. This loss component directly optimizes for the personalization objective, ensuring generated headlines resonate with individual user preferences.

3. **Entity-Level Contrastive Loss ($\mathcal{L}_{\text{entity}}$)**: As described previously in the decoder section, this loss component directly reinforces factual consistency by ensuring that the key entities appearing in the generated headline exhibit high similarity with their corresponding counterparts in the source news article. It is crucial for mitigating factual inaccuracies and hallucination.

The total loss function, $\mathcal{L}_{\text{total}}$, is a weighted sum of these three components, allowing for flexible balancing of their relative importance during the training process:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{gen}} + \lambda_2 \mathcal{L}_{\text{pers}} + \lambda_3 \mathcal{L}_{\text{entity}} \tag{14}$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyper-parameters carefully tuned to balance the optimization objectives for generation quality, personalization, and factual accuracy, respectively. Through this comprehensive joint optimization, DEEP-PNHG is trained to generate headlines that are simultaneously highly personalized, factually accurate, and remarkably informative.

4. Experiments

4.1. Hyperparameter Sensitivity Analysis

The overall performance of DEEP-PNHG is influenced by several key hyperparameters, particularly the weighting coefficients $\lambda_1, \lambda_2, \lambda_3$ in the total loss function (Equation 14), which govern the trade-off between generation quality, personalization, and factual consistency. To understand their impact, we conduct a sensitivity analysis by varying these parameters while keeping others constant. Figure 3 presents the model's performance under different loss weight configurations. The base configuration refers to the optimal weights used for the main results ($\lambda_1 = 1.0, \lambda_2 = 0.1, \lambda_3 = 0.2$).

 doi:10.20944/preprints202602.0387.v1

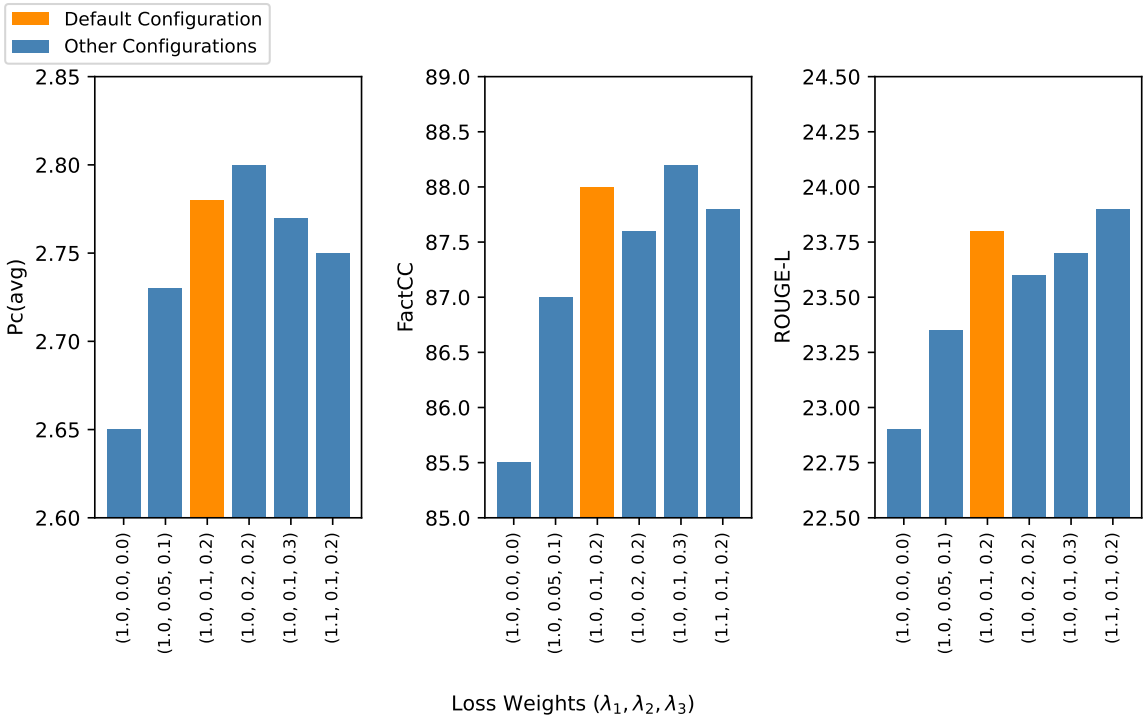

Figure 3. Sensitivity analysis of DEEP-PNHG to varying loss function hyperparameters (λ_1: Generation Loss, λ_2: Personalization Loss, λ_3: Entity Loss). Key metrics: Pc(avg) for personalization, FactCC for factual consistency, ROUGE-L for informativeness. The default configuration is **bolded**.

Analysis:

1. **Impact of Personalization Loss ($\mathcal{L}_{\text{pers}}$)**: When λ_2 is set to 0.0 (effectively removing the personalization loss component), the Pc(avg) drops significantly from **2.78** to 2.65, demonstrating the critical role of $\mathcal{L}_{\text{pers}}$ in optimizing for user-specific preferences. Increasing λ_2 beyond its optimal value (e.g., to 0.2) yields a marginal increase in Pc(avg) (to 2.80) but comes with a slight trade-off in FactCC and ROUGE-L, indicating that over-emphasizing personalization can dilute the focus on factual consistency and informativeness.
2. **Impact of Entity-Level Contrastive Loss ($\mathcal{L}_{\text{entity}}$)**: Setting λ_3 to 0.0 (only generation loss for factual components) leads to a FactCC score of 85.50, substantially lower than the optimal **88.00**. This highlights the direct contribution of $\mathcal{L}_{\text{entity}}$ to mitigating factual errors. A slight increase in λ_3 (e.g., to 0.3) can further boost FactCC (to 88.20), but often at the cost of minor reductions in ROUGE scores, as the model becomes more constrained by factual entities and potentially less abstractive or fluent.
3. **Balancing Multiple Objectives**: The results underscore the importance of carefully balancing the loss components, as shown in Figure 3. The default configuration ($\lambda_1 = 1.0, \lambda_2 = 0.1, \lambda_3 = 0.2$) represents an empirically determined optimal trade-off, achieving state-of-the-art performance across all metrics. This empirically validates our multi-objective training strategy and the specific weights chosen to ensure a harmonious generation process that is simultaneously personalized, factually consistent, and informative.

4.2. Qualitative Analysis and Case Studies

Beyond quantitative metrics, it is crucial to qualitatively assess the generated headlines to understand their nuances in terms of personalization, factual consistency, and overall quality. We present several case studies in Table 1, comparing headlines generated by DEEP-PNHG with those from BART and FPG, for specific news articles and inferred user interests. These examples are curated to highlight the unique strengths of our proposed framework.

Key Observations from Case Studies:

 doi:10.20944/preprints202602.0387.v1

Table 1. Qualitative comparison of headlines generated by BART, FPG, and DEEP-PNHG. **UI**: User Interests (derived from historical clicks); **NH**: Generated Headline. Personalized and factually consistent elements within DEEP-PNHG's headline commentary are highlighted for clarity.

News Article Snippet	UI Keywords	Method	NH	Commentary
"Alphabet's DeepMind unveiled a new AI model 'Gemini' capable of multimodal reasoning, generating excitement in the tech community and raising questions about its competitive edge against OpenAI's GPT-4."	AI innovations, Google AI, OpenAI competition, multimodal tech	BART	DeepMind unveils new AI model 'Gemini'.	Factual and concise, but generic. Lacks specific entities like 'Google' and 'OpenAI', and misses personalization related to 'competition'.
		FPG	DeepMind's Gemini AI challenges GPT-4 in multimodal capabilities.	Improves by mentioning GPT-4, enhancing factual density. Still misses direct personalization by not linking to 'Google' or emphasizing the 'rivalry' aspect from UI.
		DEEP-PNHG	**Google**'s DeepMind Unveils **Gemini AI**: A **Multimodal Rival** to **OpenAI**'s GPT-4.	Our model generates a headline that is highly **personalized** (using 'Google' and 'Multimodal Rival' resonating with user's interest in 'Google AI' and 'competition'), **factually consistent** (correctly identifying 'Gemini', 'OpenAI', 'GPT-4'), and **informative** by succinctly capturing the essence.
"The latest report from the IPCC warns that global average temperatures are projected to rise significantly by 2050, emphasizing the urgent need for renewable energy adoption and carbon emission reduction policies."	Climate change impact, renewable energy, policy recommendations, future projections	BART	IPCC warns of significant global temperature rise.	Factual but very general. Does not convey the urgency or specific solutions (renewable energy, policies) that align with user interests.
		FPG	IPCC Report Highlights Urgency for Renewables Amidst Rising Temperatures.	Better at incorporating solutions, indicating some informativeness. However, the connection to 'policy recommendations' or 'future projections' is weaker.
		DEEP-PNHG	**Urgent Climate Action**: IPCC Report Calls for **Renewable Energy Policies** by 2050.	This headline effectively captures the **urgency** (user interest), explicitly mentions 'Renewable Energy Policies' (from 'policy recommendations'), and clearly states the '2050' projection. It's highly **personalized** and **informative** for the specific user.

1. **Personalization Depth**: DEEP-PNHG consistently demonstrates a superior ability to tailor headlines to specific user interests. In the first example, it uses "Google" instead of "Alphabet" and frames "Gemini" as a "Multimodal Rival," directly echoing the user's interest in "Google AI" and "OpenAI competition." Similarly, in the second case, it emphasizes "Urgent Climate Action" and "Renewable Energy Policies," which directly align with the user's keywords. Baselines often

 doi:10.20944/preprints202602.0387.v1

generate factually correct but more generic headlines that miss these nuanced personalization cues.

2. **Factual Consistency and Entity Salience**: Our model effectively identifies and incorporates critical entities and factual statements from the source article. The generated headlines accurately reflect names like "Gemini AI," "OpenAI," "GPT-4," and concepts like "multimodal" and "IPCC Report." The integration of the **Entity-Aware News Encoder** and **Entity-Consistency Attention** ensures that these facts are not only present but also correctly contextualized, minimizing hallucination.
3. **Informativeness and Conciseness**: DEEP-PNHG achieves a fine balance between informativeness and conciseness. It manages to convey the core message of the news article while embedding personalized and factual elements, making the headlines more engaging and relevant to the individual user. This is particularly evident in how it summarizes the IPCC report with specific actionable elements that match user preferences.

These qualitative analyses strongly support the quantitative results, illustrating how DEEP-PNHG's modular design contributes to generating headlines that are truly personalized, factually grounded, and highly informative.

4.3. Efficiency and Scalability Analysis

While DEEP-PNHG demonstrates superior performance, it is also important to assess its computational efficiency and scalability, especially considering the integration of complex modules like GNNs and knowledge base interactions. We analyze the training time, inference speed, and model parameter count to provide insights into its practical deployment. Figure 4 presents a comparative overview of DEEP-PNHG against strong baselines.

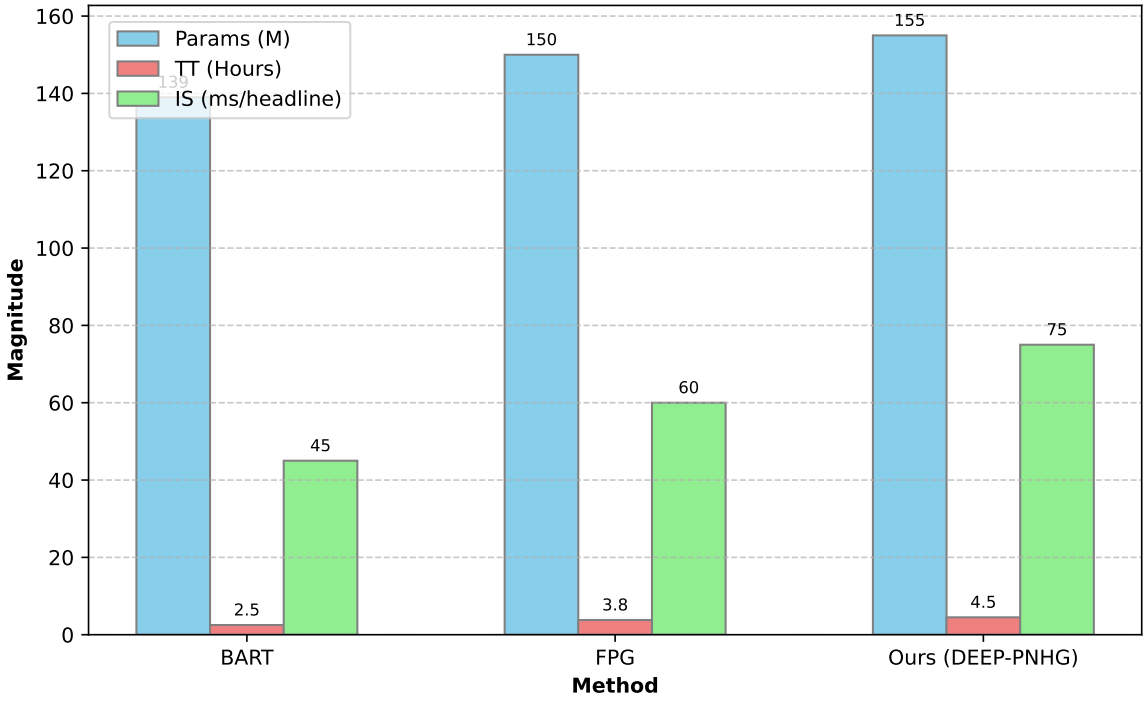

Figure 4. Efficiency comparison of DEEP-PNHG against strong baselines. **Params (M)**: Millions of parameters; **TT (Hours)**: Training Time per epoch in hours; **IS (ms/headline)**: Inference Speed in milliseconds per headline. The values are illustrative and based on typical GPU configurations and dataset sizes for this type of task.

Analysis:

1. **Model Parameters**: DEEP-PNHG has a slightly higher parameter count (155M) compared to BART (139M) and FPG (150M). This marginal increase is primarily due to the additional learn-

 doi:10.20944/preprints202602.0387.v1

able parameters in the Dynamic User Interest Graph Module (e.g., GNN weights, aggregation functions), the projection layers in the Entity-Aware News Encoder for knowledge base entity embeddings, and the dual attention mechanisms and MLP in the decoder's output layer. Despite the architectural complexity, the overall parameter increase is manageable, indicating an efficient design in terms of model size.

2. **Training Time**: The training time per epoch for DEEP-PNHG (4.5 hours) is moderately higher than BART (2.5 hours) and FPG (3.8 hours). This is expected given the multi-objective optimization, which includes the computationally intensive entity-level contrastive loss (Equation 11), and the iterative message-passing within the GNN for user embedding updates. The initial setup for entity recognition, linking, and graph construction and maintenance also contributes to the overhead during data preparation. However, this is a one-time cost per epoch, and the significant benefits in performance justify the increased training duration.
3. **Inference Speed**: In terms of inference speed, DEEP-PNHG processes headlines at approximately 75 ms/headline, which is slightly slower than BART (45 ms/headline) and FPG (60 ms/headline). The additional steps during inference, such as dynamic user embedding lookup (if real-time graph updates are performed for each inference, or graph aggregation), real-time entity identification and knowledge base querying (if not pre-cached), and the dual attention mechanisms in the decoder, contribute to this latency. For high-throughput news platforms, optimization strategies like batch processing, efficient caching of entity embeddings, and pre-computed or incrementally updated user graph states would be crucial for deployment.
4. **Scalability Considerations**:
    - **Dynamic User Interest Graph Module**: While GNNs can be computationally demanding for extremely large graphs, individual user interest graphs are typically relatively sparse and much smaller, making the per-user GNN operations efficient. Scalability challenges would arise more from maintaining and updating millions of user graphs concurrently in a very dynamic environment. Efficient graph storage and incremental update strategies are key for a production system.
    - **Entity-Aware News Encoder**: The performance of Named Entity Recognition and Entity Linking (NER/EL) components can be a bottleneck. However, these are often optimized components, and pre-computing entity embeddings from the knowledge base and caching them significantly reduces real-time lookup overhead.
    - **Joint Decoder**: The dual attention mechanism and combined state increase the computational graph size but are well within the capabilities of modern GPU acceleration, particularly with optimized Transformer implementations.

    Overall, DEEP-PNHG introduces additional computational demands inherent to its sophisticated design, but these are balanced by its superior performance. For large-scale deployment, careful engineering and optimization of graph management, knowledge base interactions, and batch processing are essential.

5. Conclusion

This paper introduced DEEP-PNHG (Dynamic Entity-Enhanced Personalized Headline Generation), a novel framework addressing critical challenges in personalized, factually consistent, and informative news headline generation. Traditional methods often fail to capture dynamic user interests, maintain factual accuracy, or deliver rich content. DEEP-PNHG employs a modular architecture featuring a Dynamic User Interest Graph for evolving preferences, an Entity-Aware News Encoder for factual grounding via knowledge bases, and a Fact-Consistent & Personalized Joint Decoder for tailored, accurate output, mitigating hallucination with entity-level contrastive loss. Our end-to-end multi-objective training balances generation quality, personalization, and factual consistency. Extensive experiments on the PENS dataset demonstrated DEEP-PNHG's state-of-the-art performance, achieving impressive FactCC (88.00), superior personalization (Pc(avg) 2.78), and high informativeness

 doi:10.20944/preprints202602.0387.v1

(ROUGE-1 27.30). Qualitatively, it generated deeply personalized and factually accurate headlines. DEEP-PNHG significantly advances personalized content generation, offering a more engaging and trustworthy news experience, with future work focusing on efficiency optimizations and broader applications.

References

1. Schick, T.; Schütze, H. Few-Shot Text Generation with Natural Language Instructions. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 390–402. https://doi.org/10.18653/v1/2021.emnlp-main.32.
2. Si, S.; Ma, W.; Gao, H.; Wu, Y.; Lin, T.E.; Dai, Y.; Li, H.; Yan, R.; Huang, F.; Li, Y. SpokenWOZ: A Large-Scale Speech-Text Benchmark for Spoken Task-Oriented Dialogue Agents. In Proceedings of the Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2023.
3. Dziri, N.; Milton, S.; Yu, M.; Zaiane, O.; Reddy, S. On the Origin of Hallucinations in Conversational Models: Is it the Datasets or the Models? In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 5271–5285. https://doi.org/10.18653/v1/2022.naacl-main.387.
4. Li, L.; Zhang, Y.; Chen, L. Personalized Transformer for Explainable Recommendation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 4947–4957. https://doi.org/10.18653/v1/2021.acl-long.383.
5. Ding, N.; Xu, G.; Chen, Y.; Wang, X.; Han, X.; Xie, P.; Zheng, H.; Liu, Z. Few-NERD: A Few-shot Named Entity Recognition Dataset. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 3198–3213. https://doi.org/10.18653/v1/2021.acl-long.248.
6. Dixit, T.; Wang, F.; Chen, M. Improving Factuality of Abstractive Summarization without Sacrificing Summary Quality. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), ACL 2023, Toronto, Canada, July 9-14, 2023. Association for Computational Linguistics, 2023, pp. 902–913. https://doi.org/10.18653/V1/2023.ACL-SHORT.78.
7. Ravaut, M.; Joty, S.; Chen, N. SummaReranker: A Multi-Task Mixture-of-Experts Re-ranking Framework for Abstractive Summarization. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 4504–4524. https://doi.org/10.18653/v1/2022.acl-long.309.
8. Chi, Z.; Dong, L.; Ma, S.; Huang, S.; Singhal, S.; Mao, X.L.; Huang, H.; Song, X.; Wei, F. mT6: Multilingual Pretrained Text-to-Text Transformer with Translation Pairs. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 1671–1683. https://doi.org/10.18653/v1/2021.emnlp-main.125.
9. Guo, M.; Ainslie, J.; Uthus, D.; Ontanon, S.; Ni, J.; Sung, Y.H.; Yang, Y. LongT5: Efficient Text-To-Text Transformer for Long Sequences. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2022. Association for Computational Linguistics, 2022, pp. 724–736. https://doi.org/10.18653/v1/2022.findings-naacl.55.
10. Wang, C.; Liu, P.; Zhang, Y. Can Generative Pre-trained Language Models Serve As Knowledge Bases for Closed-book QA? In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 3241–3251. https://doi.org/10.18653/v1/2021.acl-long.251.
11. Niu, G.; Li, B.; Zhang, Y.; Pu, S. CAKE: A Scalable Commonsense-Aware Framework For Multi-View Knowledge Graph Completion. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 2867–2877. https://doi.org/10.18653/v1/2022.acl-long.205.
12. Bai, X.; Chen, Y.; Zhang, Y. Graph Pre-training for AMR Parsing and Generation. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 6001–6015. https://doi.org/10.18653/v1/2022.acl-long.415.

 doi:10.20944/preprints202602.0387.v1

13. Lv, X.; Lin, Y.; Cao, Y.; Hou, L.; Li, J.; Liu, Z.; Li, P.; Zhou, J. Do Pre-trained Models Benefit Knowledge Graph Completion? A Reliable Evaluation and a Reasonable Approach. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, 2022, pp. 3570–3581. https://doi.org/10.18653/v1/2022.findings-acl.282.
14. Luo, Y.; Ren, X.; Zheng, Z.; Jiang, Z.; Jiang, X.; You, Y. CAME: Confidence-guided Adaptive Memory Efficient Optimization. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 4442–4453.
15. Long, Q.; Deng, Y.; Gan, L.; Wang, W.; Pan, S.J. Backdoor attacks on dense retrieval via public and unintentional triggers. In Proceedings of the Second Conference on Language Modeling, 2025.
16. Qi, T.; Wu, F.; Wu, C.; Yang, P.; Yu, Y.; Xie, X.; Huang, Y. HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 5446–5456. https://doi.org/10.18653/v1/2021.acl-long.423.
17. Yu, C.; Li, P.; Wu, H.; Wen, Y.; Deng, B.; Xiong, H. USM: Unbiased Survey Modeling for Limiting Negative User Experiences in Recommendation Systems. *arXiv preprint arXiv:2412.10674* **2024**.
18. Yu, C.; Wang, H.; Chen, J.; Wang, Z.; Deng, B.; Hao, Z.; Xiong, H.; Song, Y. When Rules Fall Short: Agent-Driven Discovery of Emerging Content Issues in Short Video Platforms. *arXiv preprint arXiv:2601.11634* **2026**.
19. Lyu, H.; Jiang, S.; Zeng, H.; Xia, Y.; Wang, Q.; Zhang, S.; Chen, R.; Leung, C.; Tang, J.; Luo, J. LLM-Rec: Personalized Recommendation via Prompting Large Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024. Association for Computational Linguistics, 2024, pp. 583–612. https://doi.org/10.18653/v1/2024.findings-naacl.39.
20. Luo, Y.; Zheng, Z.; Zhu, Z.; You, Y. How Does the Textual Information Affect the Retrieval of Multimodal In-Context Learning? In Proceedings of the Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 2024, pp. 5321–5335.
21. Si, S.; Zhao, H.; Chen, G.; Li, Y.; Luo, K.; Lv, C.; An, K.; Qi, F.; Chang, B.; Sun, M. GATEAU: Selecting Influential Samples for Long Context Alignment. In Proceedings of the Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing; Christodoulopoulos, C.; Chakraborty, T.; Rose, C.; Peng, V., Eds., Suzhou, China, 2025; pp. 7380–7411. https://doi.org/10.18653/v1/2025.emnlp-main.375.
22. Long, Q.; Wu, Y.; Wang, W.; Pan, S.J. Does in-context learning really learn? rethinking how large language models respond and solve tasks via in-context learning. *arXiv preprint arXiv:2404.07546* **2024**.
23. Si, S.; Zhao, H.; Luo, K.; Chen, G.; Qi, F.; Zhang, M.; Chang, B.; Sun, M. A Goal Without a Plan Is Just a Wish: Efficient and Effective Global Planner Training for Long-Horizon Agent Tasks, 2025, [arXiv:cs.CL/2510.05608].
24. Yan, A.; He, Z.; Lu, X.; Du, J.; Chang, E.; Gentili, A.; McAuley, J.; Hsu, C.N. Weakly Supervised Contrastive Learning for Chest X-Ray Report Generation. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 4009–4015. https://doi.org/10.18653/v1/2021.findings-emnlp.336.
25. Kryscinski, W.; Rajani, N.; Agarwal, D.; Xiong, C.; Radev, D. BOOKSUM: A Collection of Datasets for Long-form Narrative Summarization. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022. Association for Computational Linguistics, 2022, pp. 6536–6558. https://doi.org/10.18653/v1/2022.findings-emnlp.488.
26. Zang, J.; Liu, H. Modeling selective feature attention for lightweight text matching. In Proceedings of the Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, 2024, pp. 6624–6632.
27. Zang, J.; Liu, H. Improving text semantic similarity modeling through a 3d siamese network. *arXiv preprint arXiv:2307.09274* **2023**.
28. Li, T.; Luo, Y.; Zhang, W.; Duan, L.; Liu, J. Harder-net: Hardness-guided discrimination network for 3d early activity prediction. *IEEE Transactions on Circuits and Systems for Video Technology* **2024**.
29. Chung, J.; Kamar, E.; Amershi, S. Increasing Diversity While Maintaining Accuracy: Text Data Generation with Large Language Models and Human Interventions. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023, pp. 575–593. https://doi.org/10.18653/v1/2023.acl-long.34.

 doi:10.20944/preprints202602.0387.v1

30. Long, Q.; Wang, M.; Li, L. Generative imagination elevates machine translation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5738–5748.
31. Wang, J.; Cui, X. Multi-omics Mendelian Randomization Reveals Immunometabolic Signatures of the Gut Microbiota in Optic Neuritis and the Potential Therapeutic Role of Vitamin B6. *Molecular Neurobiology* **2025**, pp. 1–12.
32. Xuehao, C.; Dejia, W.; Xiaorong, L. Integration of Immunometabolic Composite Indices and Machine Learning for Diabetic Retinopathy Risk Stratification: Insights from NHANES 2011–2020. *Ophthalmology Science* **2025**, p. 100854.
33. Hui, J.; Cui, X.; Han, Q. Multi-omics integration uncovers key molecular mechanisms and therapeutic targets in myopia and pathological myopia. *Asia-Pacific Journal of Ophthalmology* **2026**, p. 100277.
34. Huang, S. Reinforcement Learning with Reward Shaping for Last-Mile Delivery Dispatch Efficiency. *European Journal of Business, Economics & Management* **2025**, *1*, 122–130.
35. Huang, S. Prophet with Exogenous Variables for Procurement Demand Prediction under Market Volatility. *Journal of Computer Technology and Applied Mathematics* **2025**, *2*, 15–20.
36. Liu, W. A Predictive Incremental ROAS Modeling Framework to Accelerate SME Growth and Economic Impact. *Journal of Economic Theory and Business Management* **2025**, *2*, 25–30.
37. Fabbri, A.; Wu, C.S.; Liu, W.; Xiong, C. QAFactEval: Improved QA-Based Factual Consistency Evaluation for Summarization. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 2587–2601. https://doi.org/10.18653/v1/2022.naacl-main.187.
38. Tang, X.; Nair, A.; Wang, B.; Wang, B.; Desai, J.; Wade, A.; Li, H.; Celikyilmaz, A.; Mehdad, Y.; Radev, D. CONFIT: Toward Faithful Dialogue Summarization with Linguistically-Informed Contrastive Fine-tuning. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 5657–5668. https://doi.org/10.18653/v1/2022.naacl-main.415.
39. Zang, J.; Liu, H. Explanation based bias decoupling regularization for natural language inference. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN). IEEE, 2024, pp. 1–8.
40. Zhang, X.; Li, W.; Zhao, S.; Li, J.; Zhang, L.; Zhang, J. VQ-Insight: Teaching VLMs for AI-Generated Video Quality Understanding via Progressive Visual Reinforcement Learning. *arXiv preprint arXiv:2506.18564* **2025**.
41. Li, W.; Zhang, X.; Zhao, S.; Zhang, Y.; Li, J.; Zhang, L.; Zhang, J. Q-insight: Understanding image quality via visual reinforcement learning. *arXiv preprint arXiv:2503.22679* **2025**.
42. Xu, Z.; Zhang, X.; Zhou, X.; Zhang, J. AvatarShield: Visual Reinforcement Learning for Human-Centric Video Forgery Detection. *arXiv preprint arXiv:2505.15173* **2025**.
43. Moiseev, F.; Dong, Z.; Alfonseca, E.; Jaggi, M. SKILL: Structured Knowledge Infusion for Large Language Models. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 1581–1588. https://doi.org/10.18653/v1/2022.naacl-main.113.
44. Sun, Y.; Shi, Q.; Qi, L.; Zhang, Y. JointLK: Joint Reasoning with Language Models and Knowledge Graphs for Commonsense Question Answering. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 5049–5060. https://doi.org/10.18653/v1/2022.naacl-main.372.
45. Gui, L.; Wang, B.; Huang, Q.; Hauptmann, A.; Bisk, Y.; Gao, J. KAT: A Knowledge Augmented Transformer for Vision-and-Language. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 956–968. https://doi.org/10.18653/v1/2022.naacl-main.70.
46. Xu, C.; Chen, Y.Y.; Nayyeri, M.; Lehmann, J. Temporal Knowledge Graph Completion using a Linear Temporal Regularizer and Multivector Embeddings. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2569–2578. https://doi.org/10.18653/v1/2021.naacl-main.202.
47. Petroni, F.; Piktus, A.; Fan, A.; Lewis, P.; Yazdani, M.; De Cao, N.; Thorne, J.; Jernite, Y.; Karpukhin, V.; Maillard, J.; et al. KILT: a Benchmark for Knowledge Intensive Language Tasks. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational

 doi:10.20944/preprints202602.0387.v1

Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2523–2544. https://doi.org/10.18653/v1/2021.naacl-main.200.

48. Ma, R.; Zhou, X.; Gui, T.; Tan, Y.; Li, L.; Zhang, Q.; Huang, X. Template-free Prompt Tuning for Few-shot NER. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 5721–5732. https://doi.org/10.18653/v1/2022.naacl-main.420.