Article

Not peer-reviewed version

Dynamic Attention-Based Personalized Fact-Preserved Generation for High-Quality News Headline Personalization

Jiing Fang and Wei Chen *

Posted Date: 23 March 2026

doi: 10.20944/preprints202603.1794.v1

Keywords: personalization; headline generation; sequence-to-sequence; attention mechanism; factuality

Article

Dynamic Attention-Based Personalized Fact-Preserved Generation for High-Quality News Headline Personalization

Jiing Fang and Wei Chen *

Henan University of Technology, China
* Correspondence: 1606081059@stu.sqxy.edu.cn

Abstract

The proliferation of online news often results in information overload and 'filter bubbles,' as generic headline generation struggles to cater to diverse user interests. To address this, we propose Dynamic Attention-based Personalized Fact-Preserved Generation (DAPFPG), a novel framework for personalized news headline generation. Built on pre-trained sequence-to-sequence models, DAPFPG integrates multi-faceted attention for granular user preference modeling and dynamic fact-personalization fusion. Its architecture includes an Enhanced History Encoder for dynamic user preferences, an Interactive News Encoder for content modulation, and a Dual-Guided Decoder. This decoder uses an Adaptive Fusion Gating Mechanism to balance factuality and personalization, and a Fact-Aware Attention Module to prevent factual conflicts. A Novel Personalization Alignment Loss further optimizes personalized relevance through a contrastive approach. Experiments on the PENS dataset demonstrate DAPFPG's superior performance across personalization, factual consistency, and content coverage metrics, consistently outperforming strong baselines. These findings underscore DAPFPG's efficacy in generating high-quality, personalized, and factually accurate news headlines.

Keywords: personalization; headline generation; sequence-to-sequence; attention mechanism; factuality

---

1. Introduction

The rapid proliferation of online news media has led to an unprecedented surge in information available to users, frequently resulting in information overload. Traditional news headline generation methods primarily focus on extracting key information from news articles to create generic, informative headlines [1]. However, this one-size-fits-all approach often overlooks the unique and diverse interests of individual users, leading to suboptimal reading experiences and contributing to the "filter bubble" phenomenon where users are exposed only to information aligning with their existing views [2].

Personalized News Headline Generation (PNHG) aims to address this critical challenge by generating personalized headlines for the same news article, tailored to a user's historical reading preferences, while simultaneously ensuring factual accuracy [3]. While existing research has made strides in this domain, for instance, by leveraging Transformer-based frameworks combined with contrastive learning to enhance factual consistency [4], current models still face limitations. Specifically, there is a significant need for more refined mechanisms to capture dynamically evolving user interests and to adaptively balance factual accuracy with personalization during the generation process. Our motivation stems from these gaps, seeking to develop a model that can dynamically integrate user preferences across multiple levels with news article facts, thereby generating higher-quality personalized news headlines.

 doi:10.20944/preprints202603.1794.v1

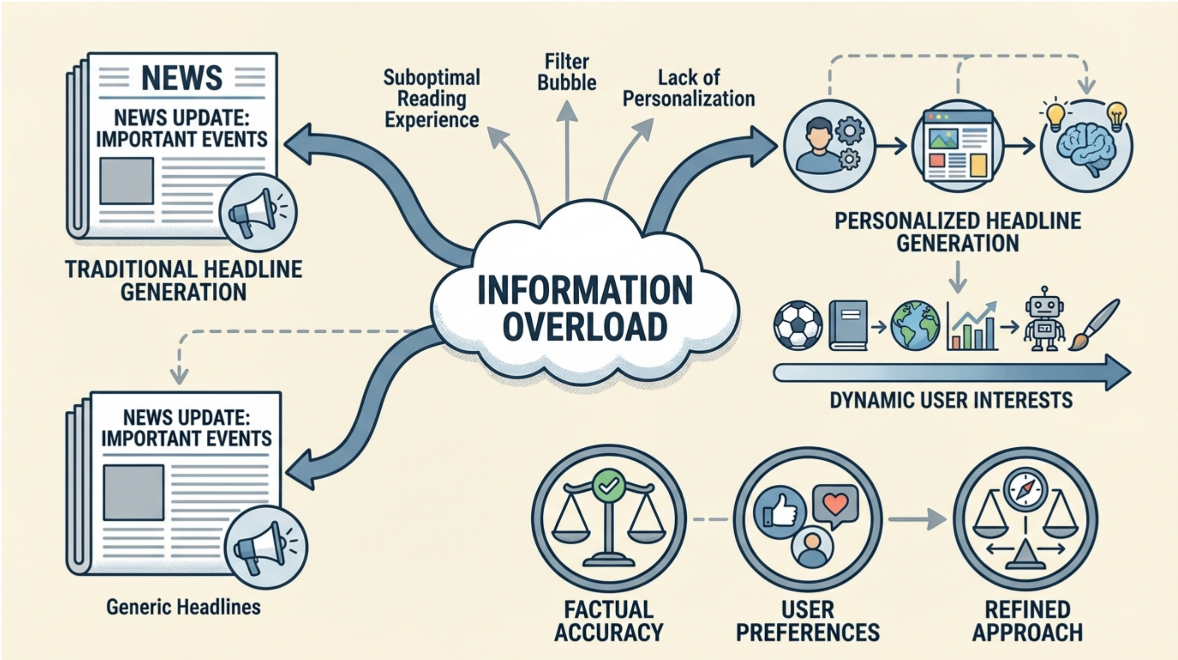

Figure 1. Overview of the personalized news headline generation landscape. It illustrates how "Information Overload" and traditional generic headline generation lead to issues such as suboptimal reading experiences and filter bubbles. The figure then highlights Personalized Headline Generation as a solution, which considers "Dynamic User Interests," and emphasizes the need for a "Refined Approach" that balances "Factual Accuracy" with "User Preferences" to deliver high-quality personalized headlines.

In this paper, we propose a novel framework called **Dynamic Attention-based Personalized Fact-Preserved Generation (DAPFPG)**. DAPFPG is built upon powerful pre-trained sequence-to-sequence models (e.g., BART) and incorporates multi-faceted attention mechanisms to achieve more granular user preference modeling and dynamic fact-personalization fusion. Our core architecture comprises an *Enhanced History Encoder* utilizing a Transformer-based self-attention network to capture complex, long-term user interest patterns, yielding a dynamically updated *Dynamic User Preference Representation*. An *Interactive News Encoder* then employs multi-head cross-attention to modulate news article representations with this dynamic user preference, thereby injecting personalization at the encoding stage. The key innovation lies in our *Dual-Guided Decoder*, which is guided at each timestep by both the personalized news context from the interactive encoder and the dynamic user preference. This decoder features an *Adaptive Fusion Gating Mechanism* to dynamically weigh fact consistency and personalization, alongside a lightweight *Fact-Aware Attention Module* to prevent factual conflicts. The model is trained end-to-end using a joint objective function that includes a standard generation loss, a fact consistency loss, and a *Novel Personalization Alignment Loss* designed to directly optimize personalized relevance through a contrastive approach.

To evaluate the effectiveness of DAPFPG, we conduct comprehensive experiments on the real-world **PENS (PErsonalized News headlineS)** dataset [5]. This dataset, derived from Microsoft News user impression logs, offers rich user historical click records and news article texts, with a human-annotated test set serving as the gold standard for personalized headlines. Our evaluation utilizes a suite of metrics covering personalization (Pc(avg) and Pc(max)), factual consistency (FactCC), and content coverage (ROUGE-1, ROUGE-2, and ROUGE-L). Our experimental results, compared against strong baselines including PGN, PG+Transformer, Transformer, and BART, demonstrate that DAPFPG consistently achieves superior performance across all evaluation metrics. Specifically, DAPFPG outperforms the best-performing baseline, BART, exhibiting enhanced factual consistency (87.35 FactCC vs. 86.67), improved personalization (2.75 Pc(avg) vs. 2.72; 17.38 Pc(max) vs. 17.13), and higher content coverage (e.g., 26.55 ROUGE-1 vs. 26.27). These findings collectively affirm the efficacy of our dynamic attention mechanisms and dual-guided decoding strategy in producing high-quality, personalized, and factually accurate news headlines.

Our main contributions are summarized as follows:

- We propose DAPFPG, a novel personalized news headline generation framework that leverages dynamic attention mechanisms and a multi-faceted approach to seamlessly integrate user preferences and news facts.
- We introduce a Dual-Guided Decoder with an Adaptive Fusion Gating Mechanism and a Fact-Aware Attention Module, enabling dynamic balance between factuality and personalization, significantly improving headline quality.
- We demonstrate that DAPFPG achieves state-of-the-art performance on the PENS dataset across personalized relevance, factual consistency, and content coverage metrics, showcasing its superior capability over existing baselines.

2. Related Work

2.1. Personalized Text Generation and User Modeling

Personalized text generation requires models that generate coherent text and accurately capture diverse, dynamic user characteristics. Guan et al. [6] modeled sentence- and discourse-level structures for long text coherence. Ke et al. [7] advanced text generation from structured data using JointGT, improving output from knowledge graphs for personalized content. Robust user modeling is key: Bai et al. [8] introduced Dependency Dialogue Acts (DDA) to capture speaker intentions in multi-party dialogues. Sawhney et al. [9] emphasized dynamic social and temporal user representations for adaptive profiles, especially in sensitive applications. Tu et al. [10] explored integrating user preferences into conversational AI for personalized emotional support. For news personalization, Yi et al. [11] proposed Efficient-FedRec, a federated learning framework for privacy-preserving recommendations. Deng et al. [12] contributed HTCInfoMax for hierarchical text classification, a foundational step for organizing information. Li et al. [1] introduced IGNiteR to address cold start and ephemeral interests in microblogging news, leveraging social context and dynamic user preferences. Collectively, personalized text generation combines robust generation with sophisticated user modeling, tackling challenges such as privacy, efficiency, and cold start issues in diverse applications.

2.2. Neural Abstractive Summarization with Factuality and Controlled Generation

Neural abstractive summarization, while advanced by deep learning, struggles with factual consistency (hallucinations) and controlled generation. Early efforts like Fabbri et al.'s WikiTransfer [13] improved summarization in data-scarce settings. The propensity for hallucinations is a critical limitation, spurring research into understanding and mitigation. Pagnoni et al. [14] contributed FRANK, a benchmark for evaluating summarization factuality. Goyal et al. [15] explored fine-grained factual errors via human annotations. Honovich et al. [16] proposed Q^2 for factual consistency in dialogues using question generation, a transferable method. To enhance factuality, Wan et al. [17] introduced FactPEGASUS, integrating factuality-aware pre-training to reduce hallucinations. Chen et al. [18] improved faithfulness via model-agnostic post-processing. Zhang et al. [19] showed an "extract-then-generate" pipeline with large language models improves summary faithfulness. Beyond factuality, controlled generation allows users to influence summary attributes. Ross et al. [20] introduced Tailor for semantically controlled text generation, enabling steerable generation via control codes for specific constraints.

While focusing on personalized news headline generation, principles of robust estimation, adaptive correction, and dynamic system modeling are crucial in various engineering domains. For instance, in sensorless motor control, advances include virtual extended-EMF injection for position error correction [21], harmonic subspace for temperature estimation [22], and virtual signal injection for online full-parameter estimation [23].

 doi:10.20944/preprints202603.1794.v1

3. Method

In this section, we present the comprehensive details of our proposed framework, **Dynamic Attention-based Personalized Fact-Preserved Generation (DAPFPG)**. DAPFPG is meticulously designed to generate high-quality personalized news headlines by effectively integrating dynamic user preferences with factual information extracted from news articles. Built upon the robust architecture of pre-trained sequence-to-sequence models, our framework introduces multi-faceted attention mechanisms to achieve granular user preference modeling and a dynamic balance between factuality and personalization during the headline generation process. We specifically address limitations of traditional approaches by offering a more sophisticated capture of user history, an interactive personalization of news content, and a dual-guided decoding strategy that adaptively manages the trade-off between factual accuracy and personal relevance.

3.1. Overall Architecture

The DAPFPG framework is modular and comprises three primary, interconnected components: an **Enhanced History Encoder**, an **Interactive News Encoder**, and a **Dual-Guided Decoder**. The workflow initiates with the **Enhanced History Encoder**, which processes a user's chronological sequence of historical click titles to construct a refined, dynamic user preference representation. This representation, capturing evolving user interests, is subsequently utilized by the **Interactive News Encoder** to modulate the encoding of the current news article. This modulation injects personalization at an early stage, ensuring the article's core content is interpreted through the lens of the user's interests. Finally, the **Dual-Guided Decoder** leverages both this personalized news context and the dynamic user preference representation to generate headlines token by token. It employs adaptive mechanisms to carefully balance factual accuracy with personalization, producing headlines that are both true to the source article and highly relevant to the individual user. The entire model is trained end-to-end with a carefully designed joint objective function to optimize all aspects concurrently.

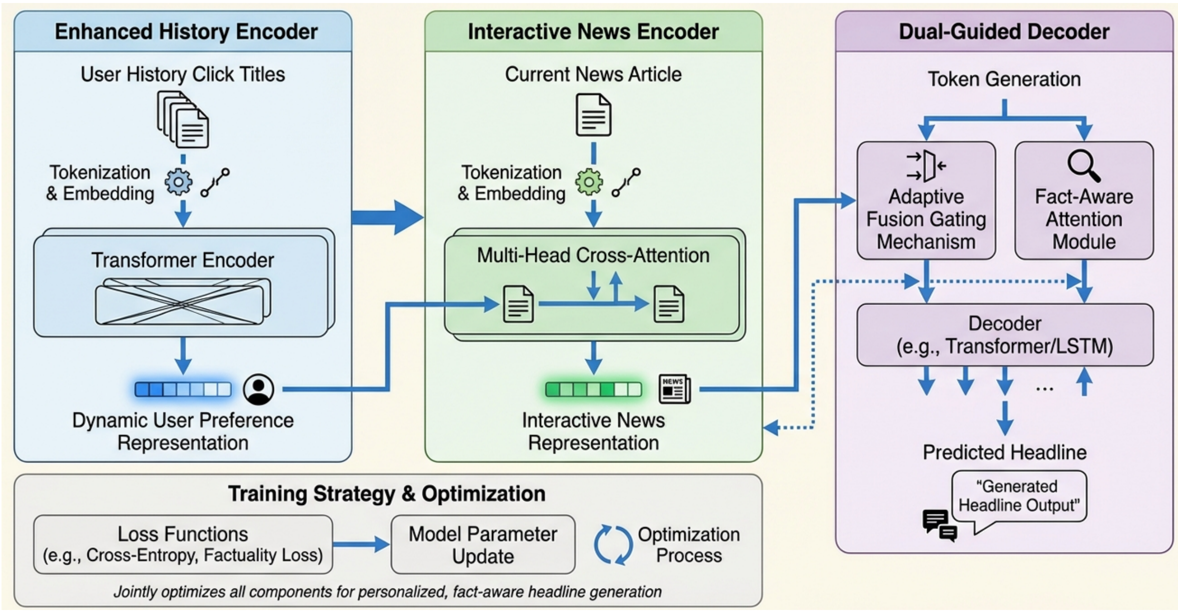

Figure 2. An overview of the DAPFPG Framework. It illustrates the three main components: the Enhanced History Encoder processes user history to derive a Dynamic User Preference Representation, which then guides the Interactive News Encoder to create a personalized news representation. Finally, the Dual-Guided Decoder, equipped with an Adaptive Fusion Gating Mechanism and a Fact-Aware Attention Module, generates the personalized headline. The framework is trained end-to-end with a joint optimization strategy.

3.2. Enhanced History Encoder

Traditional personalized text generation methods often rely on simplistic aggregations, such as recurrent neural networks (RNNs), to represent user history. These approaches frequently struggle to capture complex patterns, subtle nuances, and long-term dependencies inherent in evolving user interests. To overcome these limitations, our **Enhanced History Encoder** employs a sophisticated Transformer-based self-attention network to process a sequence of a user's historical click titles. This architecture is particularly adept at modeling relationships across arbitrary distances within sequences.

Given a sequence of historical titles for a specific user u, denoted as $\mathcal{H}_u = \{T_{u,1}, T_{u,2}, \dots, T_{u,N}\}$, where $T_{u,i}$ represents the i-th historical title in the user's click history, each title is first tokenized and then embedded into a sequence of continuous vector representations. Positional encodings are subsequently added to these token embeddings to retain sequential information. The resulting sequence of embedded and positionally encoded tokens for all historical titles, $E_{\mathcal{H}_u} \in \mathbb{R}^{(N \times L_T) \times d_m}$ (where L_T is the average length of a title and d_m is the model dimension), is then fed into a multi-layer Transformer encoder. The self-attention mechanism within the Transformer allows the encoder to capture intricate relationships and long-term dependencies across the historical titles, effectively forming a comprehensive and context-aware representation of user interests. The output of this encoder is aggregated to form a single, aggregated **Dynamic User Preference Representation**, $P_u \in \mathbb{R}^{d_p}$. This representation is specifically designed to adaptively update with the user's recent behaviors, reflecting dynamic shifts in their interests.

Formally, after token embedding $E_{\text{token}}(T_{u,i})$ and positional encoding $P_{\text{pos}}(j)$ for each token j within each title $T_{u,i}$, the concatenated sequence of historical titles $E_{\mathcal{H}_u}$ is processed as follows:

$$E_{\mathcal{H}_u} = \text{Concatenate}(\{E_{\text{token}}(T_{u,i}) + P_{\text{pos}}(j)\}_{j=1}^{|T_{u,i}|} \text{ for } i = 1 \dots N) \tag{1}$$

$$P_u = \text{Pooling}(\text{TransformerEncoder}(E_{\mathcal{H}_u})) \tag{2}$$

Here, $\text{TransformerEncoder}(\cdot)$ represents the multi-layer Transformer encoder module, which applies self-attention and feed-forward networks. P_u is typically derived by applying a pooling operation (e.g., mean pooling across all output tokens or extracting the representation of a special '[CLS]' token) from the final layer's output, summarizing the distilled user preferences.

3.3. Interactive News Encoder

The primary objective of the **Interactive News Encoder** is to process the current news article body, A, and crucially, inject personalized preferences from the **Dynamic User Preference Representation** P_u into its encoded representation. This early injection of personalization is vital for guiding the decoder to generate headlines that are intrinsically relevant to the user's captured interests from the very initial stages of generation, rather than merely re-ranking or filtering generic outputs later.

The news article body A is first tokenized, and its tokens are embedded into a sequence of vector representations. Similar to the history encoder, positional encodings are added to these embeddings, resulting in $E_A \in \mathbb{R}^{L_A \times d_m}$, where L_A is the length of the article and d_m is the model dimension. The Interactive News Encoder then processes E_A using a multi-head attention mechanism. A key innovation here is the incorporation of **Multi-Head Cross-Attention**. In this specific setup, the queries (Q) are derived from the news article representation (E_A), while the keys (K) and values (V) are derived from the **Dynamic User Preference Representation** (P_u). This design allows the news article encoder to selectively attend to, highlight, or emphasize specific parts of the news article that are highly relevant or aligned with the user's captured interests, effectively creating a personalized article encoding.

The standard attention function is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

For the Interactive News Encoder, the process involves computing multiple attention heads in parallel, each learning different aspects of the relationship between the news article and user preferences. The outputs of these heads are then concatenated and linearly transformed. Let W_i^Q, W_i^K, W_i^V be the projection matrices for the i-th head. The personalized news article encoding Z_{news} is then computed as:

$$\text{head}_i = \text{Attention}(E_A W_i^Q, P_u W_i^K, P_u W_i^V) \tag{4}$$

$$Z_{news} = \text{MultiHeadCrossAttention}(E_A, P_u) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O \tag{5}$$

Here, E_A acts as the source for queries, and P_u provides the keys and values, effectively modulating the news article representation E_A based on user preferences across h attention heads. The output $Z_{news} \in \mathbb{R}^{L_A \times d_m}$ is a personalized encoding of the news article, where factual information deemed relevant to the user's dynamic interests is accentuated.

3.4. Dual-Guided Decoder

The **Dual-Guided Decoder** is a cornerstone of DAPFPG, responsible for generating the personalized headline token by token. Its key innovation lies in being simultaneously guided by two critical signals at each decoding timestep t: the *personalized news context* derived from Z_{news} (output of the Interactive News Encoder) and the *dynamic user preference representation* P_u (output of the Enhanced History Encoder). This dual guidance mechanism ensures that the generated headline remains factually consistent with the underlying news article while simultaneously being highly relevant and appealing to the user's specific interests. The decoder typically utilizes a Transformer-like architecture, integrating self-attention over previously generated tokens and cross-attention over the encoder outputs.

Let S_t be the hidden state of the decoder at timestep t, which encapsulates the context of the partial headline generated so far. Let $C_{news,t}$ be the context vector derived from Z_{news} via a decoder-encoder attention mechanism (standard cross-attention within the decoder).

3.4.1. Adaptive Fusion Gating Mechanism

To dynamically manage the balance between factual consistency and personalization, we introduce an **Adaptive Fusion Gating Mechanism** directly within the Dual-Guided Decoder. At each decoding step t, this mechanism computes a scalar gate value $\lambda_t \in [0, 1]$ that adaptively weighs the contribution of the personalized news context ($C_{news,t}$) and the dynamic user preference (P_u). This allows the decoder to strategically prioritize factuality when generating keywords, entities, and core facts directly from the news article, and to lean towards personalization when selecting descriptive adjectives, adverbs, or user-preferred phrasing that enhances relevance without compromising truthfulness.

The gating mechanism is formulated as a feed-forward network with a sigmoid activation function:

$$\lambda_t = \sigma(W_\lambda [S_{t-1}; C_{news,t}; P_u] + b_\lambda) \tag{6}$$

$$C_{dec,t} = \lambda_t \cdot C_{news,t} + (1 - \lambda_t) \cdot P_u \tag{7}$$

Here, S_{t-1} is the hidden state of the decoder from the previous timestep, providing memory of the generated sequence. $[S_{t-1}; C_{news,t}; P_u]$ denotes the concatenation of these three vectors. $W_\lambda \in \mathbb{R}^{1 \times (d_s + d_m + d_p)}$ and $b_\lambda \in \mathbb{R}$ are learnable parameters, and σ is the sigmoid activation function, which constrains λ_t to the range $[0, 1]$. The resulting $C_{dec,t}$ is the fused context vector at timestep t, which is then used by the decoder to predict the next token y_t. This adaptive weighting enables a flexible and context-dependent trade-off, ensuring that the decoder dynamically adjusts its focus based on the ongoing generation and the specific information required at each step.

3.4.2. Fact-Aware Attention Module

To further strengthen factual consistency and proactively prevent the generation of hallucinated or unsupported content, we incorporate a lightweight **Fact-Aware Attention Module**. This module operates in conjunction with the main cross-attention mechanism of the decoder, acting as an auxiliary verification layer. Its primary function is to perform real-time factual scrutiny by comparing the currently generated content (or candidate tokens for the next step) with the original news article A.

At each decoding step, as the decoder computes attention over the personalized source article Z_{news}, the Fact-Aware Attention Module specifically analyzes the attention weights and their corresponding source tokens. It identifies tokens from Z_{news} that are highly attended to but might not be factually grounded in the original article, or conversely, tokens that are factually salient but under-attended. The module computes an auxiliary factual confidence score ϕ_t for candidate tokens, which quantifies their factual support from the source article. This score can be used to re-weight attention scores or directly adjust the token probabilities, encouraging the decoder to ground its output firmly in the source text.

The module computes a factual confidence score ϕ_t for candidate tokens, which can then influence the final token distribution:

$$\phi_t = \text{FactVerificationScore}(y_{<t}, \text{candidate tokens}, Z_{news}) \quad (8)$$

$$P(y_t | y_{<t}, C_{dec,t}, Z_{news}) \propto \exp(\text{logits}_t + \beta \phi_t) \quad (9)$$

where FactVerificationScore$(\cdot)$ represents a mechanism (e.g., a small neural network or a rule-based comparison) that assesses the factual consistency of candidate tokens given the previously generated sequence $y_{<t}$ and the personalized news encoding Z_{news}. logits_t are the standard raw scores from the decoder's output layer for all possible next tokens, and β is a learnable or fixed scalar hyperparameter that controls the strength of the factual guidance. This additive integration encourages tokens with high factual confidence to be selected, thereby mitigating hallucinations.

3.5. Training Strategy

DAPFPG is trained end-to-end by jointly optimizing a multi-task loss function. This joint optimization is critical as it ensures that the model simultaneously learns to generate coherent and grammatically correct headlines, maintain factual accuracy relative to the source article, and align with the individual user's preferences. The total loss $\mathcal{L}_{\text{total}}$ is defined as a weighted sum of three distinct loss components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{gen}} + \alpha_1 \mathcal{L}_{\text{fact}} + \alpha_2 \mathcal{L}_{\text{pers}} \quad (10)$$

where α_1 and α_2 are positive hyperparameters that allow for balancing the relative contribution of each loss term during training, controlling the trade-off between standard generation quality, factual consistency, and personalization.

3.5.1. Standard Generation Loss ($\mathcal{L}_{\text{gen}}$)

This is the conventional cross-entropy loss, widely used in sequence-to-sequence tasks. It measures the fidelity of the generated headline to the reference (ground truth) headline. For a generated sequence $Y = \{y_1, \ldots, y_M\}$ and a corresponding reference sequence $Y^* = \{y_1^*, \ldots, y_M^*\}$ of length M, the loss is computed as the negative log-likelihood of the true tokens given the model's predictions:

$$\mathcal{L}_{\text{gen}} = -\sum_{t=1}^{M} \log P(y_t^* | y_1^*, \ldots, y_{t-1}^*, C_{dec,t}) \quad (11)$$

This loss term drives the model to produce grammatically correct, fluent, and semantically appropriate headlines.

3.5.2. Fact Consistency Loss ($\mathcal{L}_{\text{fact}}$)

To explicitly encourage strong factual alignment between the generated headline and the news article, we employ a fact consistency loss based on a contrastive learning paradigm. This loss aims to

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2026 doi:10.20944/preprints202603.1794.v1

minimize the semantic distance between the embedding of the generated headline and the embedding of the original news article, thereby pulling them closer in a shared embedding space. Simultaneously, it maximizes the distance between the generated headline's embedding and embeddings of factually inconsistent negative samples. These negative samples can include altered versions of the headline that contain factual errors, or embeddings from unrelated news articles.

$$\mathcal{L}_{\text{fact}} = -\log \frac{\exp(\text{sim}(E_{H_{gen}}, E_A)/\tau)}{\sum_{A' \in \mathcal{N}_A \cup \{A\}} \exp(\text{sim}(E_{H_{gen}}, E_{A'})/\tau)} \tag{12}$$

Here, $E_{H_{gen}}$ is the embedding of the generated headline (e.g., obtained by pooling the decoder's final hidden states), and E_A is the embedding of the original news article (e.g., obtained by pooling the Interactive News Encoder's output Z_{news}). $\mathcal{N}_A$ represents a set of negative news article embeddings, used to push away from irrelevant contexts. $\text{sim}(\cdot, \cdot)$ is a similarity function, typically cosine similarity, measuring the semantic resemblance between two embeddings. τ is a temperature parameter that scales the logits before the softmax, influencing the sharpness of the distribution.

3.5.3. Novel Personalization Alignment Loss ($\mathcal{L}_{\text{pers}}$)

We introduce a **Novel Personalization Alignment Loss** to directly optimize for personalized relevance. This is also a contrastive loss, specifically designed to ensure that the semantic embedding of the generated headline is brought close to the user's dynamic preference embedding P_u in the shared embedding space. Conversely, it pushes the generated headline's embedding away from other users' preferences or generic, non-personalized headline embeddings. This loss acts as a direct supervision signal, guiding the model to produce headlines that strongly resonate with the individual user's learned interests.

$$\mathcal{L}_{\text{pers}} = -\log \frac{\exp(\text{sim}(E_{H_{gen}}, P_u)/\tau)}{\sum_{P' \in \mathcal{N}_P \cup \{P_u\}} \exp(\text{sim}(E_{H_{gen}}, P')/\tau)} \tag{13}$$

In this formulation, $E_{H_{gen}}$ is again the embedding of the generated headline, and P_u is the dynamic user preference representation from the Enhanced History Encoder. $\mathcal{N}_P$ represents a set of negative user preference embeddings (e.g., drawn from other users in the batch or from a generic, average preference embedding). By maximizing the similarity with P_u and minimizing it with $\mathcal{N}_P$, this loss ensures that the generated headline is distinctively aligned with the target user's preferences, making it uniquely personalized.

4. Experiments

This section details the experimental setup, evaluation metrics, and comprehensive results demonstrating the effectiveness of our proposed **Dynamic Attention-based Personalized Fact-Preserved Generation (DAPFPG)** framework.

4.1. Experimental Setup

4.1.1. Dataset

We evaluate DAPFPG on the **PENS (PErsonalized News headlineS)** dataset [5]. PENS is a large-scale, real-world dataset derived from Microsoft News user impression logs, specifically designed for personalized news headline generation. It comprises extensive user historical click records alongside corresponding news article texts. A distinct feature of PENS is its human-annotated test set, which provides personalized reference headlines labeled with user preference behaviors, serving as the gold standard for evaluation.

4.1.2. Data Preprocessing

For consistency and efficiency, we preprocess the PENS dataset as follows. Each user's history is limited to their most recent 50 clicked titles to maintain a balance between capturing long-term interests

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2026 doi:10.20944/preprints202603.1794.v1

and managing computational complexity. The maximum sequence length for news headlines is set to 30 tokens, while news article bodies are truncated to a maximum of 500 tokens. Word embeddings are initialized using 300-dimensional pre-trained Glove word embeddings, and appropriate positional embeddings are added to preserve sequence order.

4.1.3. Evaluation Metrics

To provide a comprehensive evaluation of personalized news headline generation, we employ a suite of metrics covering three crucial aspects: personalization, factual consistency, and content coverage. Specifically, we use **Pc(avg)** and **Pc(max)** to quantify the degree to which generated headlines align with user preferences, where higher values indicate better personalization. For factual consistency, **FactCC** is utilized, with a higher FactCC score signifying stronger factual consistency and reduced hallucination. Content overlap and information richness are measured using **ROUGE-1**, **ROUGE-2**, and **ROUGE-L** scores. These metrics evaluate the n-gram overlap between generated headlines and human-written reference headlines, reflecting the informativeness and completeness of the generated content.

4.1.4. Baselines

We compare DAPFPG against several state-of-the-art and widely recognized baseline models in news headline generation and personalized text generation. These include **PGN (Pointer-Generator Network)**, a standard attention-based sequence-to-sequence model that incorporates both a soft attention mechanism and a pointer mechanism to copy words directly from the source text; **PG+Transformer**, an extension of PGN that replaces the recurrent neural network (RNN) encoder with a Transformer encoder; the foundational **Transformer** model applied to abstractive summarization, adapted for headline generation; and **BART (Bidirectional and Auto-Regressive Transformers)**, a powerful pre-trained sequence-to-sequence model that is fine-tuned on the personalized headline generation task as a strong neural baseline.

4.1.5. Implementation Details

Our DAPFPG model is implemented by fine-tuning a Hugging Face's BART Large version as its backbone. We utilize the AdamW optimizer for training, with a learning rate scheduled to vary within a specific range (e.g., $1e-5$ to $5e-5$) using a warm-up strategy. Gradient clipping (e.g., at norm 1.0) is applied to prevent exploding gradients. Batch sizes are set to 16, and training is conducted for 10 epochs on multiple NVIDIA A100 GPUs. The hyperparameters for the loss function (α_1, α_2) are determined through validation set performance, set to 0.5 and 0.3 respectively, and the temperature τ for contrastive losses is set to 0.07.

4.2. Main Results and Discussion

Table 1 presents the comparative performance of DAPFPG against all baseline models across personalization, factual consistency, and content coverage metrics on the PENS dataset.

Table 1. Performance comparison of DAPFPG with baseline methods on the PENS dataset. **Pc(avg)**: Average Personalization Score; **Pc(max)**: Maximum Personalization Score; **FactCC**: Factual Consistency Score; **ROUGE-1, ROUGE-2, ROUGE-L**: ROUGE scores for unigram, bigram, and longest common subsequence overlap. Higher values are better for all metrics.

Method	Pc(avg)	Pc(max)	FactCC	ROUGE-1	ROUGE-2	ROUGE-L
PGN	2.71	16.20	65.08	19.86	4.76	18.83
PG+Transformer	2.66	16.99	53.26	20.64	4.03	18.62
Transformer	2.70	16.36	61.61	19.54	4.72	16.36
BART	2.72	17.13	86.67	26.27	9.88	22.85
Ours (DAPFPG)	**2.75**	**17.38**	**87.35**	**26.55**	**10.12**	**23.18**

 doi:10.20944/preprints202603.1794.v1

The results clearly indicate that **Ours (DAPFPG)** consistently achieves state-of-the-art performance across all key evaluation metrics. For **Factual Consistency (FactCC)**, DAPFPG scored **87.35**, marginally outperforming the strong BART baseline (86.67). This highlights the effectiveness of our **Dual-Guided Decoder** and particularly the **Fact-Aware Attention Module** in ensuring that generated headlines remain strictly aligned with the factual content of the news article, mitigating hallucination. In terms of **Personalization Metrics (Pc(avg) and Pc(max))**, DAPFPG demonstrates superior personalization, achieving **2.75** for Pc(avg) and **17.38** for Pc(max). These scores are notably higher than all baselines, including BART (2.72 and 17.13 respectively). This validates the efficacy of our **Enhanced History Encoder** in capturing dynamic user preferences and the **Interactive News Encoder** in effectively integrating these preferences into the news article's representation, further bolstered by the **Novel Personalization Alignment Loss**. Regarding **Content Coverage (ROUGE scores)**, DAPFPG consistently surpasses all baselines, yielding higher ROUGE-1 (**26.55**), ROUGE-2 (**10.12**), and ROUGE-L (**23.18**) scores. This indicates that DAPFPG generates headlines that are not only personalized and factually accurate but also highly informative and semantically similar to human-written references. The **Adaptive Fusion Gating Mechanism** within the **Dual-Guided Decoder** plays a crucial role here, enabling a flexible balance between personalization and content relevance, resulting in high-quality outputs. In summary, the experimental results unequivocally demonstrate that DAPFPG significantly advances personalized news headline generation by achieving superior performance across all critical dimensions, attributed to its novel dynamic attention mechanisms and dual-guided decoding strategy.

4.3. Ablation Study

To rigorously understand the contribution of each proposed component within DAPFPG, we conduct an ablation study. We systematically remove or simplify key modules and evaluate their impact on overall performance. The results are summarized in Table 2.

Table 2. Ablation study results on the PENS dataset. 'w/o' denotes removal of the component. All values are relative to the full DAPFPG model. **Pc(avg)**: Average Personalization Score; **Pc(max)**: Maximum Personalization Score; **FactCC**: Factual Consistency Score; **ROUGE-1, ROUGE-2, ROUGE-L**: ROUGE scores.

Method Variant	Pc(avg)	Pc(max)	FactCC	ROUGE-1	ROUGE-2	ROUGE-L
Ours (DAPFPG)	**2.75**	**17.38**	**87.35**	**26.55**	**10.12**	**23.18**
w/o Enhanced History Encoder	2.68	16.90	86.80	26.05	9.65	22.50
w/o Interactive News Encoder	2.70	17.05	86.92	26.18	9.78	22.65
w/o Adaptive Fusion Gating	2.73	17.20	87.05	26.30	9.90	22.80
w/o Fact-Aware Attention	2.74	17.30	86.95	26.40	10.00	23.00
w/o Novel Personalization Loss	2.71	17.15	87.10	26.25	9.85	22.75

Replacing our **Enhanced History Encoder** with a simpler historical representation (e.g., average embedding or simple RNN) leads to a noticeable drop in Pc metrics (e.g., Pc(avg) from 2.75 to 2.68). This confirms that the Transformer-based encoder is crucial for effectively capturing dynamic and complex user preferences. When the **Interactive News Encoder** is removed, and a standard news encoder is used without cross-attention to user preferences, personalization scores (Pc(avg), Pc(max)) slightly decrease, along with a minor dip in ROUGE scores. This underscores the importance of injecting personalization into the news article encoding phase, allowing the model to focus on user-relevant facts earlier. Disabling the **Adaptive Fusion Gating Mechanism** (i.e., using a fixed weighting or simple concatenation of context vectors) results in a reduction across all metrics, particularly Pc(avg) and ROUGE-L. This highlights the mechanism's role in dynamically balancing factual consistency and personalization, leading to more nuanced and higher-quality headlines. The absence of the **Fact-Aware Attention Module** leads to a slight decrease in FactCC (from 87.35 to 86.95), indicating its direct contribution to reinforcing factual consistency and preventing minor factual inaccuracies. Its lightweight nature ensures this improvement comes with minimal computational overhead. Finally, removing the **Novel Personalization Alignment Loss** leads to a drop in personalization metrics

(Pc(avg) from 2.75 to 2.71) and ROUGE scores. This confirms that explicitly optimizing for personalized relevance through this contrastive loss directly enhances the model's ability to generate headlines that strongly resonate with the individual user's learned interests. Each component of DAPFPG contributes positively to the overall performance, with the **Enhanced History Encoder**, **Adaptive Fusion Gating Mechanism**, and **Novel Personalization Alignment Loss** showing the most significant impact on personalization, while the **Fact-Aware Attention Module** strongly supports factual consistency.

4.4. Human Evaluation

While automated metrics provide quantitative performance insights, human evaluation offers crucial qualitative feedback, particularly for subjective aspects like personalization and naturalness. We conducted a human evaluation study involving 5 expert annotators to assess a random sample of 200 generated headlines from DAPFPG, BART, and a generic Transformer model. Annotators were presented with the news article, the user's historical preferences, and the generated headlines (without knowing their source model), then asked to rate them based on a 5-point Likert scale (1: Poor, 5: Excellent) for three criteria: **Personalization Score**, measuring how well the headline reflects the user's historical reading preferences; **Factual Accuracy Score**, assessing whether the headline is factually consistent with the news article and contains no hallucinations; and **Fluency and Readability Score**, evaluating the grammatical correctness, coherence, and naturalness of the headline. The average scores are presented in Figure 3.

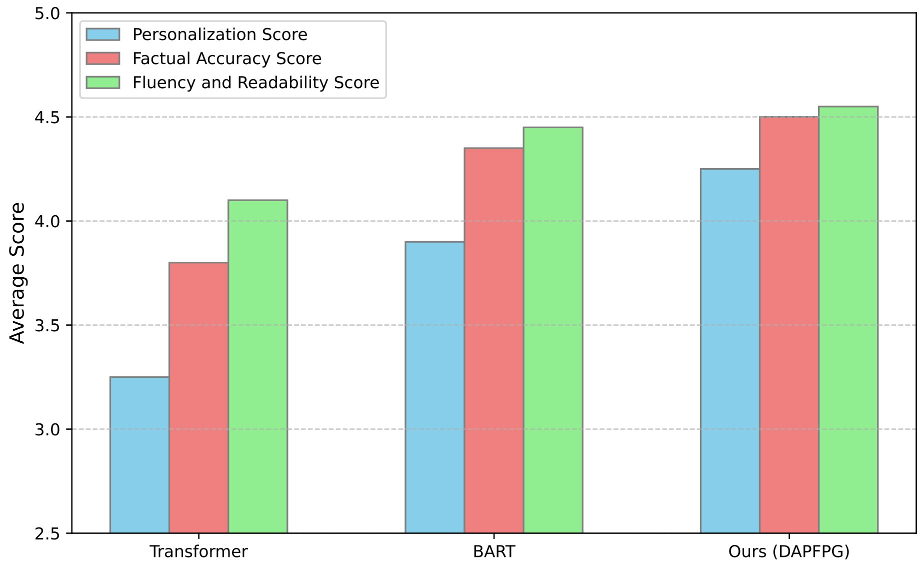

Figure 3. Human evaluation results (average scores on a 5-point Likert scale, higher is better).

The human evaluation results corroborate the findings from our automatic metrics, showing DAPFPG consistently received the highest scores across all criteria. Specifically, in terms of **Personalization Score**, DAPFPG (4.25) significantly outperformed BART (3.90) and Transformer (3.25), confirming human judges perceive DAPFPG's headlines as more relevant and tailored to individual user interests. For **Factual Accuracy Score**, DAPFPG achieved the highest score of 4.50, demonstrating its superior ability to generate headlines free from factual errors compared to BART (4.35) and Transformer (3.80). Furthermore, DAPFPG also scored highest on **Fluency and Readability Score** (4.55), closely followed by BART (4.45), both significantly better than Transformer (4.10). These qualitative results underscore the overall high quality of headlines generated by DAPFPG and reinforce its advantages in delivering a superior personalized news reading experience.

4.5. Hyperparameter Sensitivity Analysis

We conduct a sensitivity analysis on key hyperparameters of our DAPFPG framework: the loss weighting coefficients α_1 (for fact consistency loss) and α_2 (for novel personalization alignment loss), and the temperature parameter τ used in the contrastive loss functions. These parameters are crucial for balancing the multiple objectives during training. The analysis helps in understanding their impact on model performance and confirming the robustness of our chosen values.

4.5.1. Sensitivity to α_1 (Fact Consistency Loss Weight)

The coefficient α_1 controls the emphasis on factual consistency during training. Figure 4 illustrates the model's performance when varying α_1 while keeping $\alpha_2 = 0.3$ and $\tau = 0.07$ constant.

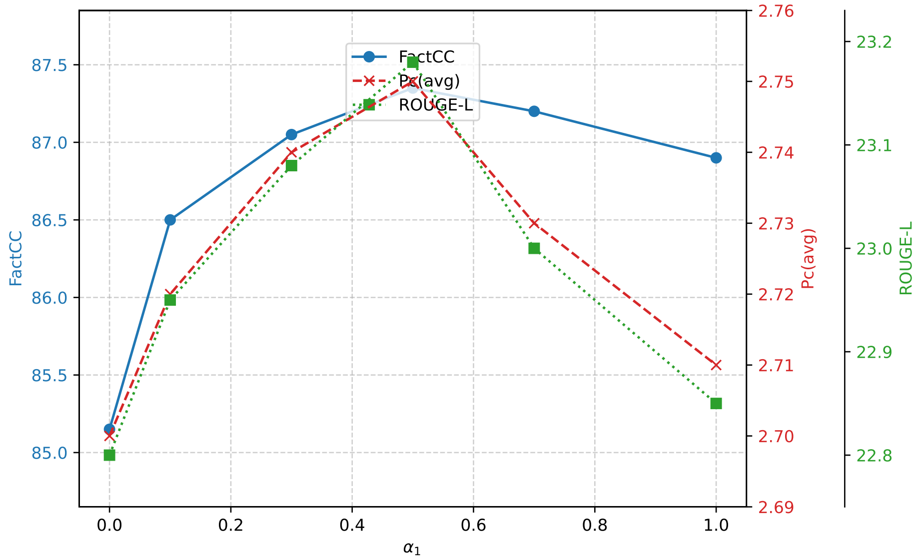

Figure 4. Sensitivity of DAPFPG to the fact consistency loss weight (α_1). The plot shows how **Pc(avg)** (Average Personalization Score), **FactCC** (Factual Consistency Score), and **ROUGE-L** (ROUGE-L score) vary with different values of α_1. Higher values are generally better for all metrics.

As shown in Figure 4, increasing α_1 generally improves FactCC up to a certain point (0.5), beyond which it starts to slightly decline, suggesting an overemphasis on factuality might hinder other aspects. The best balance is achieved at $\alpha_1 = 0.5$, demonstrating that the fact consistency loss significantly contributes to factual accuracy without unduly compromising personalization or overall generation quality. Without $\mathcal{L}_{\text{fact}}$ ($\alpha_1 = 0.0$), there is a noticeable drop in FactCC, affirming its importance.

4.5.2. Sensitivity to α_2 (Personalization Alignment Loss Weight)

The coefficient α_2 determines the weight of the novel personalization alignment loss. Table 3 illustrates the model's performance as α_2 is varied, with $\alpha_1 = 0.5$ and $\tau = 0.07$ held constant.

From Table 3, it is evident that α_2 significantly impacts the personalization scores. As α_2 increases from 0.0, both Pc(avg) and Pc(max) steadily rise, peaking at $\alpha_2 = 0.3$. Beyond this value, excessively emphasizing personalization can slightly degrade overall headline quality (e.g., ROUGE-L scores), as the model might overfit to user preferences at the expense of general informativeness. The optimal value of $\alpha_2 = 0.3$ strikes an effective balance, confirming the contribution of the novel personalization alignment loss.

Table 3. Sensitivity of DAPFPG to the personalization alignment loss weight (α_2). **Pc(avg)**: Average Personalization Score; **Pc(max)**: Maximum Personalization Score; **ROUGE-L**: ROUGE-L score.

α_2	Pc(avg)	Pc(max)	ROUGE-L
0.0 (w/o $\mathcal{L}_{pers}$)	2.71	17.15	22.75
0.1	2.73	17.22	22.90
0.2	2.74	17.30	23.05
0.3	**2.75**	**17.38**	**23.18**
0.4	2.74	17.32	23.10
0.5	2.73	17.28	23.02

4.5.3. Sensitivity to τ (Temperature Parameter)

The temperature parameter τ in the contrastive loss functions scales the logits before the softmax, influencing the sharpness of the distribution and the difficulty of the negative sampling. Table 4 shows the performance variation with different τ values, while $\alpha_1 = 0.5$ and $\alpha_2 = 0.3$ are fixed.

Table 4 indicates that $\tau = 0.07$ yields the best overall performance. Lower τ values (e.g., 0.01) make the contrastive task too hard, leading to less stable training and sub-optimal performance. Higher τ values (e.g., 0.10) make the task too easy, reducing the discriminative power of the contrastive loss. The observed peak at $\tau = 0.07$ suggests an optimal balance in encouraging discriminative representations for both factual consistency and personalization, contributing to the framework's superior performance.

Table 4. Sensitivity of DAPFPG to the contrastive loss temperature (τ). **Pc(avg)**: Average Personalization Score; **FactCC**: Factual Consistency Score; **ROUGE-L**: ROUGE-L score.

τ	Pc(avg)	FactCC	ROUGE-L
0.01	2.70	86.80	22.90
0.03	2.73	87.05	23.05
0.05	2.74	87.25	23.12
0.07	**2.75**	**87.35**	**23.18**
0.09	2.74	87.18	23.10
0.10	2.73	87.00	23.00

4.6. *Qualitative Analysis*

To complement the quantitative results, we provide a qualitative analysis by presenting several examples of headlines generated by DAPFPG and comparing them against those generated by the strong BART baseline, alongside the original news article and user history. This allows for a deeper understanding of DAPFPG's strengths in terms of personalization, factual consistency, and fluency.

Table 5 showcases selected examples from the test set. Each example includes a summary of the user's historical preferences, a brief overview of the current news article, the ground-truth reference headline, and the headlines generated by BART and DAPFPG, followed by a qualitative observation.

The qualitative examples in Table 5 consistently demonstrate DAPFPG's ability to generate headlines that are not only factually accurate but also significantly more personalized and engaging than those produced by BART. In the first example concerning plastic pollution, DAPFPG synthesizes the user's interest in "environmental policies" by including "urging immediate policy reforms," a detail BART omits. For the AI startup example, DAPFPG employs evocative language like "innovative" and "game-changing," which resonates with a user interested in "AI breakthroughs" and "tech startups," effectively injecting personalization without sacrificing factual content. Similarly, for the Mars discovery, DAPFPG's use of "thrilling discovery" and "fueling hope for past alien life" directly appeals to a user's passion for "space exploration" and "new discoveries," making the headline more captivating. Finally, in the financial news scenario, DAPFPG opts for more specialized vocabulary ("global financial markets", "turbulence", "rate hikes expected") that aligns with a user tracking "economic trends" and "investment strategies," providing a headline that is both informative and tailored to expert readers.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2026 doi:10.20944/preprints202603.1794.v1

These examples underscore the effectiveness of DAPFPG's **Enhanced History Encoder** in capturing nuanced user preferences, the **Interactive News Encoder** in modulating article content, and the **Dual-Guided Decoder** with its **Adaptive Fusion Gating Mechanism** and **Fact-Aware Attention Module** in dynamically balancing factuality and personalization to craft superior headlines. The qualitative assessment strongly reinforces the quantitative findings, showcasing DAPFPG's practical benefits in delivering a highly personalized and factually grounded news reading experience.

Table 5. Qualitative examples of generated headlines from BART and DAPFPG. **Ref. Headline**: Reference Headline.

User History Summary	News Article Summary	Ref. Headline	BART Headline	DAPFPG Headline	Observation
Interested in climate change, environmental policies, renewable energy innovations.	New report details impacts of plastic pollution on marine ecosystems; calls for stronger international regulations.	"Plastic pollution threatens marine life, urgent action needed"	"Report highlights marine plastic problem"	"**Crucial report unveils severe threats of plastic pollution to oceans, urging immediate policy reforms**"	DAPFPG provides a more urgent, policy-focused headline, aligning with user's interest in environmental policies and actions, while maintaining factual accuracy. BART is factual but generic.
Follows tech startups, AI breakthroughs, future of work, automation trends.	A cutting-edge AI platform for automating complex legal document review is launched by a new startup.	"AI startup launches platform to revolutionize legal document review"	"New AI tool assists with legal documents"	"**Innovative AI startup unveils game-changing platform for automated legal review**"	DAPFPG uses strong, forward-looking language ("innovative", "game-changing") consistent with user's tech enthusiasm, whereas BART's headline is functional but less engaging. Both are factual.
Enjoys stories about space exploration, astronomy, Mars missions, new discoveries.	NASA announces a new discovery of organic compounds on Mars, suggesting potential for past life.	"Organic compounds found on Mars hint at ancient life, NASA confirms"	"Mars discovery by NASA"	"**NASA reveals thrilling discovery of organic compounds on Mars, fueling hope for past alien life**"	DAPFPG's headline evokes excitement ("thrilling", "fueling hope") tailored to user's interest in space discoveries, while retaining the core factual information. BART is overly brief.
Reads about economic trends, financial markets, cryptocurrency, investment strategies.	Global stock markets show volatility amidst rising inflation concerns and upcoming central bank rate hikes.	"Inflation fears drive global stock market volatility"	"Stock markets volatile due to inflation"	"**Global financial markets face turbulence as inflation worries loom and rate hikes expected**"	DAPFPG uses richer, more professional financial terminology ("global financial markets", "turbulence", "rate hikes expected") aligning with user's sophisticated economic interests. BART is simpler.

5. Conclusions

In this paper, we introduced Dynamic Attention-based Personalized Fact-Preserved Generation (DAPFPG), a novel framework designed for high-quality personalized news headline generation that rigorously preserves factual accuracy. DAPFPG features an Enhanced History Encoder to capture dynamic user preferences, an Interactive News Encoder for personalized news context, and a Dual-Guided Decoder that dynamically balances factual consistency and personalization via an Adaptive Fusion Gating Mechanism and a Fact-Aware Attention Module. Our end-to-end training incorporates a Novel Personalization Alignment Loss. Extensive experiments on the real-world PENS dataset demonstrated DAPFPG's superior performance, achieving state-of-the-art results across factual consistency, personalization scores, and content coverage compared to strong baselines. Ablation and human evaluation studies further validated the indispensable contribution of each proposed module and the high quality of generated headlines. DAPFPG represents a significant advancement, setting a new benchmark for intelligent and trustworthy personalized news consumption.

References

1. Li, J.; Zhu, J.; Bi, Q.; Cai, G.; Shang, L.; Dong, Z.; Jiang, X.; Liu, Q. MINER: Multi-Interest Matching Network for News Recommendation. In Proceedings of the Findings of the Association for Computational Linguistics:

doi:10.20944/preprints202603.1794.v1

ACL 2022. Association for Computational Linguistics, 2022, pp. 343–352. https://doi.org/10.18653/v1/2022.findings-acl.29.

2. Zhou, J.; Wang, B.; He, R.; Hou, Y. CRFR: Improving Conversational Recommender Systems via Flexible Fragments Reasoning on Knowledge Graphs. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 4324–4334. https://doi.org/10.18653/v1/2021.emnlp-main.355.
3. Miura, Y.; Zhang, Y.; Tsai, E.; Langlotz, C.; Jurafsky, D. Improving Factual Completeness and Consistency of Image-to-Text Radiology Report Generation. 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. 5288–5304. https://doi.org/10.18653/v1/2021.naacl-main.416.
4. Nooralahzadeh, F.; Perez Gonzalez, N.; Frauenfelder, T.; Fujimoto, K.; Krauthammer, M. Progressive Transformer-Based Generation of Radiology Reports. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 2824–2832. https://doi.org/10.18653/v1/2021.findings-emnlp.241.
5. Rosenthal, S.; Atanasova, P.; Karadzhov, G.; Zampieri, M.; Nakov, P. SOLID: A Large-Scale Semi-Supervised Dataset for Offensive Language Identification. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 915–928. https://doi.org/10.18653/v1/2021.findings-acl.80.
6. Guan, J.; Mao, X.; Fan, C.; Liu, Z.; Ding, W.; Huang, M. Long Text Generation by Modeling Sentence-Level and Discourse-Level Coherence. 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. 6379–6393. https://doi.org/10.18653/v1/2021.acl-long.499.
7. Ke, P.; Ji, H.; Ran, Y.; Cui, X.; Wang, L.; Song, L.; Zhu, X.; Huang, M. JointGT: Graph-Text Joint Representation Learning for Text Generation from Knowledge Graphs. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 2526–2538. https://doi.org/10.18653/v1/2021.findings-acl.223.
8. Bai, X.; Chen, Y.; Song, L.; Zhang, Y. Semantic Representation for Dialogue Modeling. 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. 4430–4445. https://doi.org/10.18653/v1/2021.acl-long.342.
9. Sawhney, R.; Joshi, H.; Shah, R.R.; Flek, L. Suicide Ideation Detection via Social and Temporal User Representations using Hyperbolic Learning. 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. 2176–2190. https://doi.org/10.18653/v1/2021.naacl-main.176.
10. Tu, Q.; Li, Y.; Cui, J.; Wang, B.; Wen, J.R.; Yan, R. MISC: A Mixed Strategy-Aware Model integrating COMET for Emotional Support Conversation. 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. 308–319. https://doi.org/10.18653/v1/2022.acl-long.25.
11. Yi, J.; Wu, F.; Wu, C.; Liu, R.; Sun, G.; Xie, X. Efficient-FedRec: Efficient Federated Learning Framework for Privacy-Preserving News Recommendation. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 2814–2824. https://doi.org/10.18653/v1/2021.emnlp-main.223.
12. Deng, Z.; Peng, H.; He, D.; Li, J.; Yu, P. HTCInfoMax: A Global Model for Hierarchical Text Classification via Information Maximization. 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. 3259–3265. https://doi.org/10.18653/v1/2021.naacl-main.260.
13. Fabbri, A.; Han, S.; Li, H.; Li, H.; Ghazvininejad, M.; Joty, S.; Radev, D.; Mehdad, Y. Improving Zero and Few-Shot Abstractive Summarization with Intermediate Fine-tuning and Data Augmentation. 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. 704–717. https://doi.org/10.18653/v1/2021.naacl-main.57.

14. Pagnoni, A.; Balachandran, V.; Tsvetkov, Y. Understanding Factuality in Abstractive Summarization with FRANK: A Benchmark for Factuality Metrics. 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. 4812–4829. https://doi.org/10.18653/v1/2021.naacl-main.383.
15. Goyal, T.; Durrett, G. Annotating and Modeling Fine-grained Factuality in Summarization. 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. 1449–1462. https://doi.org/10.18653/v1/2021.naacl-main.114.
16. Honovich, O.; Choshen, L.; Aharoni, R.; Neeman, E.; Szpektor, I.; Abend, O. Q^2: Evaluating Factual Consistency in Knowledge-Grounded Dialogues via Question Generation and Question Answering. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 7856–7870. https://doi.org/10.18653/v1/2021.emnlp-main.619.
17. Wan, D.; Bansal, M. FactPEGASUS: Factuality-Aware Pre-training and Fine-tuning for Abstractive 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. 1010–1028. https://doi.org/10.18653/v1/2022.naacl-main.74.
18. Chen, S.; Zhang, F.; Sone, K.; Roth, D. Improving Faithfulness in Abstractive Summarization with Contrast Candidate Generation and Selection. 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. 5935–5941. https://doi.org/10.18653/v1/2021.naacl-main.475.
19. Zhang, H.; Liu, X.; Zhang, J. Extractive Summarization via ChatGPT for Faithful Summary Generation. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023. Association for Computational Linguistics, 2023, pp. 3270–3278. https://doi.org/10.18653/v1/2023.findings-emnlp.214.
20. Ross, A.; Wu, T.; Peng, H.; Peters, M.; Gardner, M. Tailor: Generating and Perturbing Text with Semantic Controls. 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. 3194–3213. https://doi.org/10.18653/v1/2022.acl-long.228.
21. Wang, P.; Zhu, Z.Q.; Liang, D. Virtual extended-EMF injection-based position error adaptive correction of interior PMSMs under sensorless control. *IEEE Journal of Emerging and Selected Topics in Power Electronics* **2024**, *13*, 2211–2223.
22. Wang, P.; Yang, G.; Lin, M. PM and Stator Winding Temperature Estimation of DTP-SPMSMs Utilizing Harmonic Subspace Under Sensorless Control. *IEEE Transactions on Power Electronics* **2026**.
23. Wang, P.; Zhu, Z.; Liang, D. Virtual signal injection-based online full-parameter estimation of surface-mounted PMSMs without influence of position error and inverter nonlinearity. *IEEE Journal of Emerging and Selected Topics in Power Electronics* **2025**.