

Original Paper

# AI-Enhanced Automatic Life Story Structuring for Reminiscence Therapy in Older Adults: Technical Feasibility Study

Fang Gui<sup>1\*</sup>, PhD; Mengchen Yang<sup>1\*</sup>, BE; Liuqi Jin<sup>2\*</sup>, PhD; Jing Qian<sup>1\*</sup>, PhD

<sup>1</sup>School of Artificial Intelligence, Chuzhou University, Chuzhou, Anhui Province, China

<sup>2</sup>School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China

\*all authors contributed equally

## Corresponding Author:

Jing Qian, PhD  
School of Artificial Intelligence  
Chuzhou University  
No. 1 Huifeng West Road  
Chuzhou, Anhui Province  
China  
Phone: 86 13692166474  
Email: [1254809540@qq.com](mailto:1254809540@qq.com)

## Abstract

**Background:** Storytelling interventions have demonstrated substantial potential in improving emotional well-being, cognitive function, and quality of life for older adults. However, its effectiveness is often limited by the challenges of processing disorganized and redundant life stories, which impose substantial cognitive demands on caregivers. Although storytelling interventions are a well-established therapeutic approach, current practices depend heavily on manual narrative organization, restricting both the scalability and consistency of treatment delivery. Prior research has primarily focused on validating the clinical outcomes of storytelling interventions, with insufficient attention given to technological solutions that could enhance narrative processing while preserving therapeutic integrity. Digital approaches to life story structuring remain underexplored, despite their potential to amplify storytelling benefits by reducing cognitive load and improving recall accuracy.

**Objective:** This study aims to design an event timeline generation algorithm to optimize the prior work of the Story Mosaic system. The optimized system enables (1) the automatic extraction of event elements from life narratives, (2) the automatic organization of fragmented life stories into structured timelines, and (3) the preservation of clinically relevant contextual details during compression. The goal is to reduce manual intervention costs while increasing treatment efficacy through artificial intelligence–driven narrative structuring.

**Methods:** We have designed a novel method, CARE event timeline (CARE-ET), which combines a temporal attention mechanism with graph-based event relationship modeling. Furthermore, we used the CARE-ET algorithm to optimize existing story collage systems. The system uses multifeature extraction technology to capture event clues from oral histories, prioritizes the 6 elements of events through a hierarchical attention mechanism, and uses adaptive compression algorithms to reduce redundancy while maintaining narrative continuity. To verify the effectiveness of the CARE-ET method, this paper adopts a multidimensional evaluation framework, which encompasses event summary assessment, timeline quality evaluation, and usability testing of the optimized system.

**Results:** The proposed CARE-ET algorithm outperforms the baseline in both narrative flow and temporal accuracy. The Story Mosaic system, optimized by the CARE-ET algorithm, underwent usability evaluation by 10 caregivers recruited for this study. Based on standardized assessment metrics, the system received an A rating for usability. The comprehensive experimental results demonstrate that the CARE-ET method can effectively structure fragmented narratives from older adults, enhancing the usability of the Story Mosaic system.

**Conclusions:** The proposed method enables the structured extraction of representative event summaries, transforming disorganized life stories into an event timeline for caregiver-supported older adult well-being interventions. Future research should investigate longitudinal effects on cognitive preservation and explore integration with existing dementia care protocols. This work establishes a critical foundation for intelligent assistive technologies in geriatric mental health interventions.

**Keywords:** storytelling; life story organization; caregiver support; Story Mosaic system; older adult well-being interventions

## Introduction

Cognitive impairment is a common condition affecting the health of older adults, and it is difficult to treat through medication [1]. People with mild cognitive impairment may encounter the following issues: (1) impaired short-term and long-term memory with reduced recall efficiency [2]; (2) deteriorated social functioning, exhibiting social withdrawal and diminished emotional interaction, which can lead to social isolation [3, 4]; and (3) declined executive function in daily activities [5,6].

Through the analysis of over a hundred studies, Zhang et al [7] demonstrated that digital health technologies can effectively address the limitations of traditional medical models by delivering personalized and highly accessible intervention strategies, thereby facilitating timely intervention during the early stages of cognitive decline. Storytelling is an important application of digital health technology, which plays a significant role in the field of reminiscence [8]. Harnessing the power of storytelling, people with mild cognitive impairment can be encouraged to participate more proactively in their health care [9]. Additionally, life stories can serve as a database and decision support tool for understanding older adults and providing personalized services [10-13]. In practice, Zhu et al [14] have demonstrated that storytelling requires older adults to recall, organize, and narrate personal experiences—a process that naturally exercises cognitive functions such as memory, attention, and language abilities.

Interviews and daily interactions with older adults are a common way to access their life stories [15]. However, these life stories often have problems such as redundant content and disorganized structure, which increase the cognitive load on caregivers when learning them [16]. Moreover, as the number of exchanges grows and the volume of life stories continues to expand, the scattered information becomes increasingly difficult for caregivers to process [17,18]. In our prior work, our team developed the Story Mosaic system, designed to assist caregivers in integrating and organizing life stories for older adults. The findings were published in the international journal *JMIR Aging*. However, this version of the system has significant limitations in functional implementation: the extraction of story elements and the arrangement of narrative logic remain highly dependent on manual operations. This not only substantially increases the workload for caregivers but also necessitates the assistance of volunteers in practical applications to complete the recording and organization of older adults' life stories [17].

An effective solution is to extract the core events from older adults' life stories, rearrange the events in the order in which they occurred, and generate an event timeline structure of older adults' life stories. Then, based on the algorithm, develop a visualization system to present the timeline of older adults. An event timeline is a series of events, activities, or

important moments recorded and displayed in chronological order, with a simple and clear presentation that allows users to quickly understand the temporal relationship and development of events. The structuring of life stories lies at the critical intersection point between the basic data infrastructure and artificial intelligence applications. It addresses the current problem of “fragmented data and lack of depth” in personalized intervention research, laying the foundation for the development of an intelligent intervention system that truly “understands you” [19].

Existing studies have made some progress in personal event timeline generation, but there are still limitations that do not meet the requirements of older adults' life stories. The Twitter (subsequently rebranded X) event extraction method proposed by Li et al [20] treats frequent mentions of events as general events, which contradicts the characteristics of older adults' narratives. Althoff et al's [21] timeline construction method relies on precise temporal information, which is difficult to apply to the temporally ambiguous life stories. Sun et al's [22] stage-by-stage story line generation method ensures plot integrity but lacks a generalized description of key events. The pretraining with extracted gap-sentences for abstractive summarization model learns to generate summaries by pretraining with masked key sentences, but its reliance on document structural integrity and sentence coherence makes it difficult to handle issues such as temporal confusion, information redundancy, or fragmentation in the life stories of older adults [23]. The long text-to-text transfer transformer model uses a local-global attention mechanism to efficiently process long texts and avoid excessive information repetition, but its limited ability to capture fine-grained temporal relationships constrains its performance in strictly reorganizing fragmented life events into chronological timelines [24]. The pyramid-based masked sentence pretraining for multidocument summarization model enhances multidocument information integration through pyramid-based masked pretraining; however, as the life stories of older adults are often fragmented and narrated individually, the lack of cross-document structural associations limits the effectiveness of its hierarchical modeling approach [25]. The life stories of older adults exhibit intrinsic particularities, such as ambiguous temporal information, conflicting event elements, and difficulty in determining event significance. These characteristics directly contribute to the limitations of existing methods in processing such narratives.

To address limitations in existing approaches to organizing life stories for older adults, this paper proposes a CARE event timeline (CARE-ET), an innovative event timeline generation method that systematically structures narratives through optimized event relevance analysis and implements the practical Story Mosaic system for caregiver application. Our research makes the following three key contributions. First, we developed a multimodal event fusion method using a 2-layer encoder-decoder architecture that integrates event

triggers and timestamps. This structured representation model significantly improves accuracy in extracting key events from older adult narratives. Second, the theoretical novelty lies in the synergistic integration of graph attention networks (GATs) and deep belief networks (DBNs). In contrast to existing hybrids where DBNs serve merely as static feature extractors, the CARE-ET model uses a joint self-supervised module. Here, the latent clustering probability distribution generated by the DBN acts as a global before regularizing the GAT attention coefficients. This ensures that the learned values are optimized not only for local semantic alignment but also for global thematic coherence across the life story timeline. Finally, building upon the CARE-ET method, we have optimized and enhanced the previously developed Story Mosaic system. The improved system is designed to construct more logically coherent life story timelines for older adults, thereby providing reliable data support for personalized care.

## Methods

This section first elaborates on the overall algorithmic framework of the CARE-ET method and concludes with an introduction to the optimized and enhanced functionalities of the Story Mosaic system.

### *Ethical Considerations*

This study aims to propose a novel algorithm to optimize the Story Mosaic system, thereby better facilitating the automatic organization of fragmented life stories of older adults. This study has been approved by the research ethics committees of Hefei University of Technology (HFUT20220921001) and Chuzhou University (CZSC2025-021), respectively. All

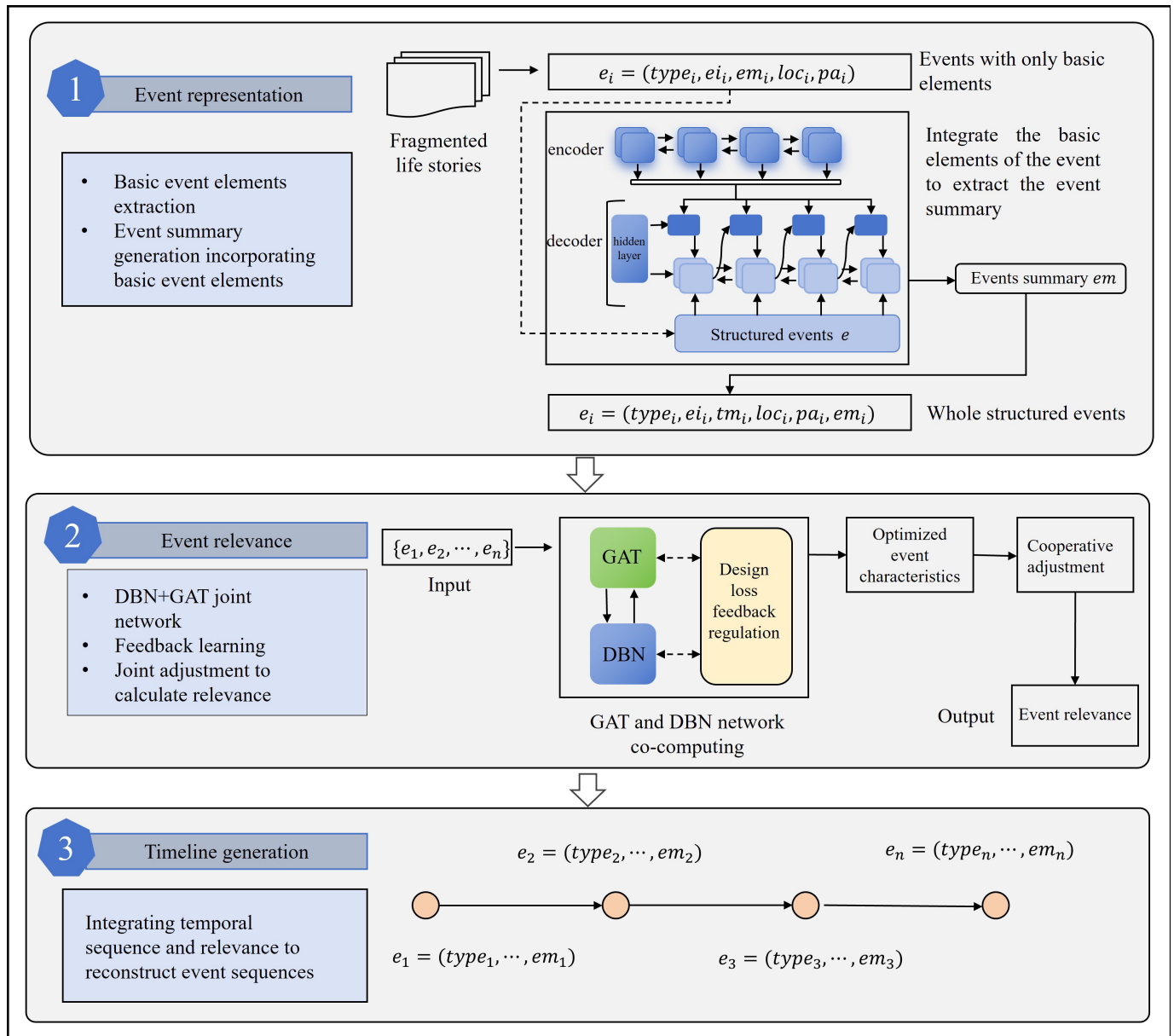
participants, including older adults and caregivers, signed a written informed consent form before the initiation of the study. The informed consent form explicitly included authorization for audio recording of the interviews, verbatim transcription of the recorded content, and the use of such materials for academic research purposes. Participation is entirely voluntary. Participants have the right to withdraw unconditionally at any stage of the study without assuming any liability or providing a reason. Given the nature and objectives of this study, there are no circumstances that necessitate the provision of financial or any other form of compensation to participants. All data collected during the research process (including interview recordings and transcribed texts) have undergone rigorous anonymization. All personally identifiable information has been removed or replaced, and comprehensive privacy protection measures have been implemented. The relevant information and images related to the older adults presented in this paper do not involve any personal privacy.

## **CARE-ET Method**

### **Overview**

The CARE-ET method automatically generates life story timelines through a three-stage computational process, as shown in [Figure 1](#): (1) bidirectional long short-term memory (BiLSTM)-based event extraction and representation, which synthesizes triggers, time stamps, and semantic context into structured events; (2) hybrid DBN-GAT modeling for self-supervised event correlation computation using optimized loss functions and distance-density joint adjustment; and (3) integrating temporal features with semantic correlations to produce coherent event timelines.

**Figure 1.** The CARE event timeline (CARE-ET) algorithmic framework for automated life story structuring. The framework comprises three stages: (1) event extraction via bidirectional long short-term memory (BiLSTM), (2) relevance modeling using the joint network of graph attention network (GAT) and deep belief network (DBN), and (3) timeline generation. Rectangles represent computational modules, while arrows indicate the flow of feature vectors and processed narrative data.



### Event Extraction and Presentation

Event extraction and representation consist of 2 main processes: the extraction of basic event elements and the generation of event summaries incorporating the basic event elements. Event information is extracted from older adults’ life stories. The structured event representations include event type, event trigger word, time, location, participant, and event summary, and the formal expression is:  $e_i = (type_i, ei_i, tm_i, loc_i, pa_i, em_i)$  (1)

### Structured Event Extraction

The module achieves coreference resolution by using ALBERT (A Lite Bidirectional Encoder Representations from Transformers for Self-supervised Learning of Language Representations) embeddings with a 512-token sliding window for context management. Similarity between

candidate expressions is measured via cosine distance, with a threshold of 0.8 established to determine coreferentiality [26,27]. We adopt a semisupervised open domain event extraction approach to expand event templates through manual annotation of triggers and types, enabling systematic extraction of basic event elements (type, trigger, time, location, and participants) for structured representation of older adults’ life stories [28].

### Context-Aware Summary Generation

This module builds a dual-level encoder-decoder architecture. It uses ALBERT-BiLSTM for encoding and combines BiLSTM decoding with dynamic attention mechanisms [16]. The model prioritizes event-centric content by jointly analyzing word or sentence-level features alongside spatiotemporal and participant context. For event-level

summarization, the system selects the top three most salient sentences using maximal marginal relevance based on ALBERT's cross-attention scores. This ensures that the selected content maximizes coverage of the event trigger while minimizing information redundancy between sentences.

## Event Relevance Calculation

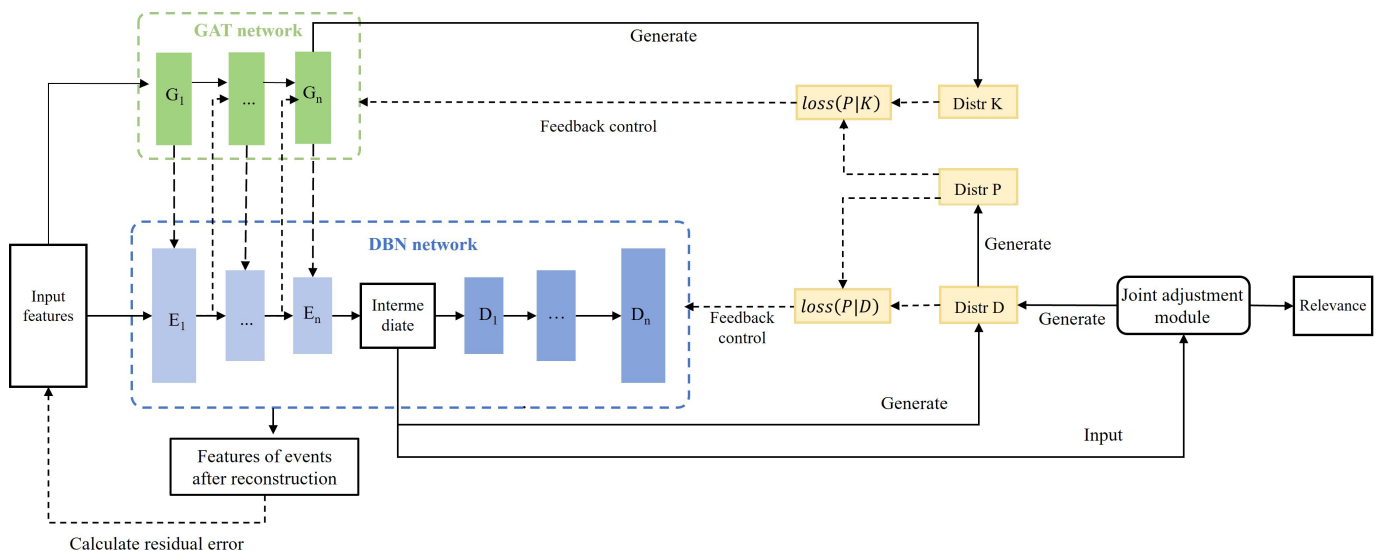
### Overview

After the event extraction module extracts the structured events for each life story, we need to calculate event correlations to arrange the event nodes on the timeline.

The CARE-ET method uses a novel GAT-DBN collaborative model to compute event relationships for

timeline construction [29]. The architecture combines: (1) an enhanced GAT network that captures multihop event dependencies through dynamic attention weights, extending influence beyond immediate neighbors; (2) a DBN component that learns compressed event representations while preserving relational semantics; and (3) a hybrid k-means or density-based spatial clustering of applications with noise (DBSCAN) self-supervision module that optimizes feature clustering through distance-density joint optimization [30], as shown in Figure 2. This integrated approach automatically learns hierarchical event relationships while maintaining computational efficiency through backpropagation-based training.

**Figure 2.** Detailed architecture of the event relevance computing model. This diagram illustrates the collaborative workflow between the graph attention network (GAT; which captures global dependencies) and the deep belief network (DBN; which performs feature compression). The joint adjustment module integrates the results to output refined correlation coefficients used for timeline sequencing.

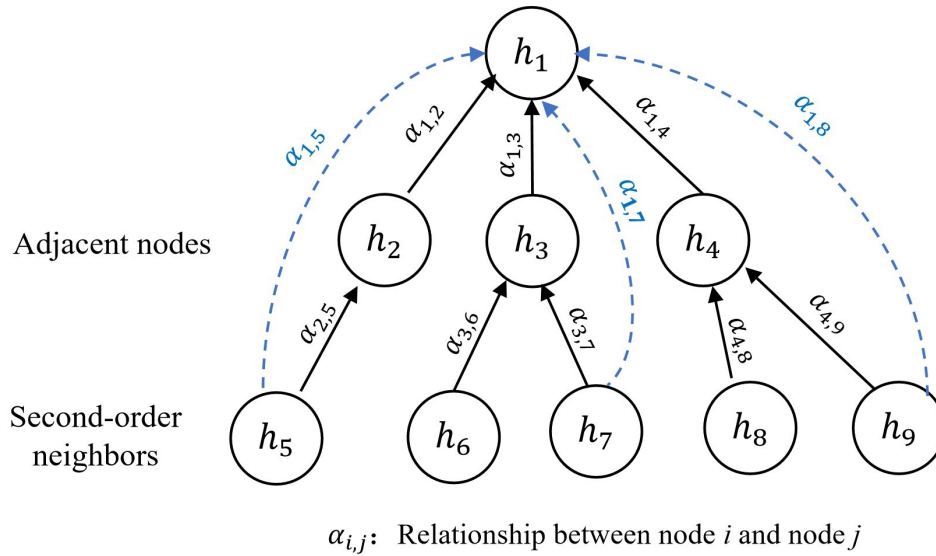


### The Improved GAT Network

The proposed GAT extension structure advances traditional attention mechanisms by incorporating multihop node relationships during state updates. While standard GAT computes attention weights based solely on direct neighbors [31], our method dynamically integrates deeper network contexts through an adaptive depth strategy, enabling more comprehensive feature representation while maintaining computational efficiency. This enhancement allows each node to capture richer structural patterns by considering both immediate connections and semantically relevant indirect relationships within life story narratives.

Figure 3 illustrates the GAT network's enhanced state update mechanism, where each node's output (eg,  $h_1$ ) incorporates weighted influences from both direct neighbors ( $h_2, h_3, h_4$ ) and secondary neighbors ( $h_5, h_7, h_9$ ). The model automatically filters insignificant connections (eg,  $h_6$  and  $h_8$ ) through threshold-based attention weights, enabling efficient multihop relationship modeling while maintaining computational efficiency. This dynamic depth perception improves event relationship characterization without compromising network performance.

**Figure 3.** The mechanism of graph attention network node updates based on multihop relationships. Circles represent events  $h_i$ . Solid arrows indicate direct influence from immediate neighbors, while dashed arrows denote filtered indirect connections identified by the adaptive depth strategy. Gray nodes (eg.  $h_6$  and  $h_8$ ) represent connections that have been discarded due to the similarity threshold.



Assuming that a total of  $n$  events exist,  $e$  denotes the event, and each event is characterized as an  $m$ -dimensional vector. All event features can be represented as the set  $\Omega = \{e_1, e_2, \dots, e_n\}$ . Based on this, the features of all events can be integrated into a sample feature matrix  $\Phi$  of  $n \times m$ :

$$\Phi = \begin{bmatrix} e_{11} & e_{21} & \dots & e_{n1} \\ e_{12} & e_{22} & \dots & e_{n2} \\ \vdots & \vdots & \dots & \vdots \\ e_{1m} & e_{2m} & \dots & e_{nm} \end{bmatrix} \quad (2)$$

The GAT network can be described as :

$$O = G(\Phi) \quad (3)$$

The improved GAT network facilitates node feature evolution through linear transformations and attention-weighted aggregation. The feature update for each node in a specific layer is defined as:

$$h_i^l = \sigma \left( \sum_{j \in (N_i \cup I_i)} a_{ij}^l W^l \vec{h}_j^l \right) \quad (4)$$

where  $\Phi$  is the network input,  $G$  denotes the network function, and  $O$  denotes the network output. Thus, the GAT network can be described as :

$$\mathfrak{D} = \begin{bmatrix} o_{11} & o_{21} & \dots & o_{n1} \\ o_{12} & o_{22} & \dots & o_{n2} \\ \vdots & \vdots & \dots & \vdots \\ o_{1m'} & o_{2m'} & \dots & o_{nm'} \end{bmatrix} \quad (5)$$

where  $m'$  is the new dimension of each event vector element, the actual output is a matrix consisting of new vectors describing each event.

The core of the GAT model lies in the graph attention layer, which maps the upper-layer graph node states to the lower-layer graph node states. The input to the graph attention layer,  $\mathcal{H}^{l-1}$ , is the set of graph node feature vectors output from the upper layer.

$$\mathcal{H}^{l-1} = \{e_1^{\mathcal{H}^{l-1}}, e_2^{\mathcal{H}^{l-1}}, \dots, e_n^{\mathcal{H}^{l-1}}\} \quad (6)$$

$n$  is the number of nodes in the network, and  $f^{l-1}$  is the dimension of the output features of the previous layer. Thus, the size of the matrix  $\mathcal{H}^{l-1}$  is  $n \times f^{l-1}$ :

$$H^{l-1} = \begin{bmatrix} h_{11}^{H^{l-1}} & h_{21}^{H^{l-1}} & \dots & h_{n1}^{H^{l-1}} \\ h_{12}^{H^{l-1}} & h_{22}^{H^{l-1}} & \dots & h_{n2}^{H^{l-1}} \\ \vdots & \vdots & \dots & \vdots \\ h_{1f^{l-1}}^{H^{l-1}} & h_{2f^{l-1}}^{H^{l-1}} & \dots & h_{nf^{l-1}}^{H^{l-1}} \end{bmatrix} \quad (7)$$

The output of this layer is a new set of node feature vectors  $\mathfrak{M}'$ , denoted as :

$$\mathfrak{M}' = \{e_1^{\mathfrak{M}'}, e_2^{\mathfrak{M}'}, \dots, e_n^{\mathfrak{M}'}\} \quad (8)$$

The attention coefficient  $\theta_{ij}$  is computed using a single-layer feed-forward neural network, parameterized by a weight vector  $a$ , which operates on the concatenated features of the node pair:

$$\theta_{ij} = \vec{a} \left( W \vec{h}_i, W \vec{h}_j \right) \quad (9)$$

$$\theta_{ij} = \text{LeakyReLU} \left( \vec{a}^T \left[ W \vec{h}_i \parallel W \vec{h}_j \right] \right) \quad (10)$$

The proposed model extends traditional GAT attention computation through a dynamic coefficient function  $\alpha$  that

incorporates both direct neighbors  $N_i$  and validated indirect connections  $I_i$ , as shown in Figure 3. By enhancing the masked attention mechanism with SoftMax regularization, the system captures multi-hop dependencies while maintaining stable gradient propagation, significantly improving relationship modeling for life event networks compared to standard single-hop approaches.

This formula represents the attention coefficient of node  $i$  to node  $j$ , where  $\alpha$  is not a constant but a node-dependent function, and  $W$  denotes the trainable weight parameters. The key innovation of our GAT enhancement lies in its refined attention allocation mechanism.

Unlike traditional GAT models, which compute attention solely over the direct neighbors of node  $i$ , our improved masked attention mechanism extends the scope of attention coefficient computation, enabling the exploration of indirect node relationships. Additionally, we introduce a SoftMax function to normalize attention scores across both direct neighbors and selected indirect nodes meeting specific criteria (equation 9). This modification enhances information propagation and dependency modeling between nodes:

$$\alpha_{ij} = \text{softmax}_j(\theta_{ij}) = \frac{\exp(\theta_{ij})}{\sum_{(k \in N_i \cup I_i)} \exp(\theta_{ik})} \quad (11)$$

where  $N_i$  denotes the set of adjacent nodes for node  $i$ , while  $I_i$  represents the set of qualified indirect nodes for node  $i$ . According to the definition of  $N_i$  and  $I_i$ , only events with a similarity exceeding a predefined threshold are established as related events.

Our GAT extension introduces an adaptive second-hop set  $I_i$  based on semantic relevance thresholds. While heterogeneous graph attention typically focuses on predefined meta-paths [32], our approach dynamically identifies high-influence indirect neighbors in the narrative graph. By redefining the training of attention coefficients  $\alpha$  over  $I_i$ , the model captures long-range dependencies in disorganized oral histories while mitigating the oversmoothing issues common in deep GNNs.

The network function  $\mathcal{G}$  determines the node influence weights based on the correlation between the 2 events. The cosine similarity function is used to calculate the event similarity  $\varphi$  between the 2 events  $e_i$  and  $e_j$ :

$$\varphi(e_i, e_j) = \frac{e_i \cdot e_j}{|e_i| |e_j|} \quad (12)$$

To model the structural relevance of life events, we define the neighbor set  $N_i$  for a given event node  $e_i$  as:

$$N_i = \{e_j \mid \cos\_sim(h_i, h_j) \geq \tau_1\} \quad (13)$$

where  $\tau_1 = 0.6$ .

For global consistency adjustment, the second-order inference set  $I_i$  is constructed as:

$$I_i = \{e_k \mid \exists e_j \in N_i, \cos\_sim(h_j, h_k) \geq \tau_2\} \quad (14)$$

where  $\tau_2 = 0.8$ .

Next, the LeakyReLU function is adopted as the single-layer feed-forward function of the attention mechanism, and the improved attention mechanism is shown below :

$$\alpha_{ij} = \text{softmax}_j(\sigma_{ij}) = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [W \vec{h}_i \parallel W \vec{h}_j]))}{\sum_{(k \in N_i \cup I_i)} \exp(\text{LeakyReLU}(\vec{a}^T [W \vec{h}_i \parallel W \vec{h}_k]))} \quad (15)$$

The attention function can calculate the attention coefficients between different nodes and predict the output characteristics of each node. The output characteristics are :

$$\vec{h}'_i = \sigma \left( \sum_{j \in (N_i \cup I_i)} \alpha_{ij} W \vec{h}_j \right) \quad (16)$$

Meanwhile, this method adopts the multihead attention mechanism, which can make the attention training more stable. The core idea is to execute  $m$  attention functions independently and use the average distribution of  $m$  attention outputs as the output of the overall attention mechanism :

$$\vec{h}'_i = \sigma \left( \frac{1}{m} \sum_{l=1}^m \sum_{j \in (N_i \cup I_i)} \alpha_{ij}^m W^m \vec{h}_j \right) \quad (17)$$

### GAT Synergy With DBN Networks

The model integrates GAT and DBN through a bidirectional encoding-decoding framework. GAT captures interevent correlations via attention-weighted node interactions, while DBN performs hierarchical feature compression through multilayer transformations. Their synergistic operation—alternating GAT's encoded outputs with DBN's hidden states during each encoding round—preserves individual event characteristics while enhancing semantic relationships. The complementary decoding process ensures reconstructed features maintain fidelity to original inputs while optimizing them for correlation computation, achieving both dimensionality reduction and relationship-aware representation learning.

In this synergistic interaction model, the inputs and outputs of each layer of the DBN coding part are shown in equation 16 :

$$h_i^+ = \text{sigmoid} \left( \sum_j |\mathcal{G}_j| w_j v_j + b_i \right) \quad (18)$$

Also, the inputs and outputs of each GAT network can be adapted in the following form:

$$\mathcal{D}_i = G(\mathcal{D}_{i-1}, h_{i-1}^+) \quad (19)$$

### Self-Monitoring Module

To address the coordination challenge between GAT and DBN networks, we introduce a novel self-supervised joint

adjustment module that serves as an integrative optimizer. This module dynamically aligns the feature representations from both networks during prediction by: (1) generating target distributions through hybrid k-means or DBSCAN clustering, (2) constructing a unified loss function for backpropagation, and (3) enabling simultaneous parameter optimization across both architectures. The resulting feature representations are specifically tuned for accurate event correlation computation while maintaining the distinct advantages of each network's processing paradigm.

If the 2 events are highly correlated, their features should also be highly similar. Multiple events describing similar contexts will form clusters in the feature space. In this paper, we use the output features of the DBN as inputs to the joint adjustment module to compute the interevent correlation and cluster structure. This is done by calculating the similarity between each event feature within a cluster and the center event feature of that cluster, which in turn generates a distribution  $D$  that contains the correlations of all the events, where  $h_i$  denotes the DBN feature of the  $i^{\text{th}}$  event,  $c_j$  is the center event of the  $j^{\text{th}}$  cluster, and  $h_{c_j}$  is the center event feature and  $v_c$  is a fixed parameter. The construction of the correlation distribution is accomplished by traversing all clusters and their center events. The correlation distribution of the  $i^{\text{th}}$  event belonging to the  $j^{\text{th}}$  class is shown in equation 18:

$$d_{ij} = \frac{\left(1 + \frac{|h_i - h_{c_j}|^2}{v_c}\right)^{-(v_c + 1)}}{\sum_n \left(1 + \frac{|h_i - h_{cn}|^2}{v_c}\right)^{-(v_c + 1)}} \quad (20)$$

The self-supervised module enhances event feature representations by optimizing correlations between individual events and their cluster centroids, thereby improving semantic similarity grouping among related events. The target distribution of event relevance is first computed  $P$ :

$$p_{ij} = \frac{d_{ij}^2 / f_i}{\sum_{j'} d_{ij'}^2 / f_i} \quad (21)$$

where  $f_j = \sum_i d_{ij}$  is the sum of the frequencies of a category of event correlation, and  $\sum_{j'} d_{ij'}^2 / f_i$  is the sum of the frequencies of all categories of event correlation, where  $j'$  denotes the individual categories in the computation. The following loss function is constructed from the distributions  $P$  and  $D$  for inverse training the DBN. The self-supervision module uses the Kullback-Leibler divergence to quantify the discrepancy

between the predictive and target distributions. The joint loss function is formally defined as:

$$L = KL(P|D) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{d_{ij}} \quad (22)$$

GAT also requires backpropagation training. Since both distributions  $P$  and  $D$  depend on the features generated by the DBN, more noise may be introduced if the loss( $P|D$ ) is used directly to compute the loss of GAT. The GAT network itself can generate a distribution  $K$  (which is also the output of the last GAT network  $\mathcal{O}$ ), using  $K$  instead of the DBN-generated distribution  $D$ , using  $K$  and  $P$  to compute the loss. In this way, the GAT-generated distribution  $K$  is supervised by the distribution  $P$ , and the loss is designed to be used for the backpropagation of the GAT, where  $k_{ij}$  is  $o_{ij}$  in the last network output matrix  $\mathcal{O}$ .

$$\text{loss}(P|K) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{k_{ij}} \quad (23)$$

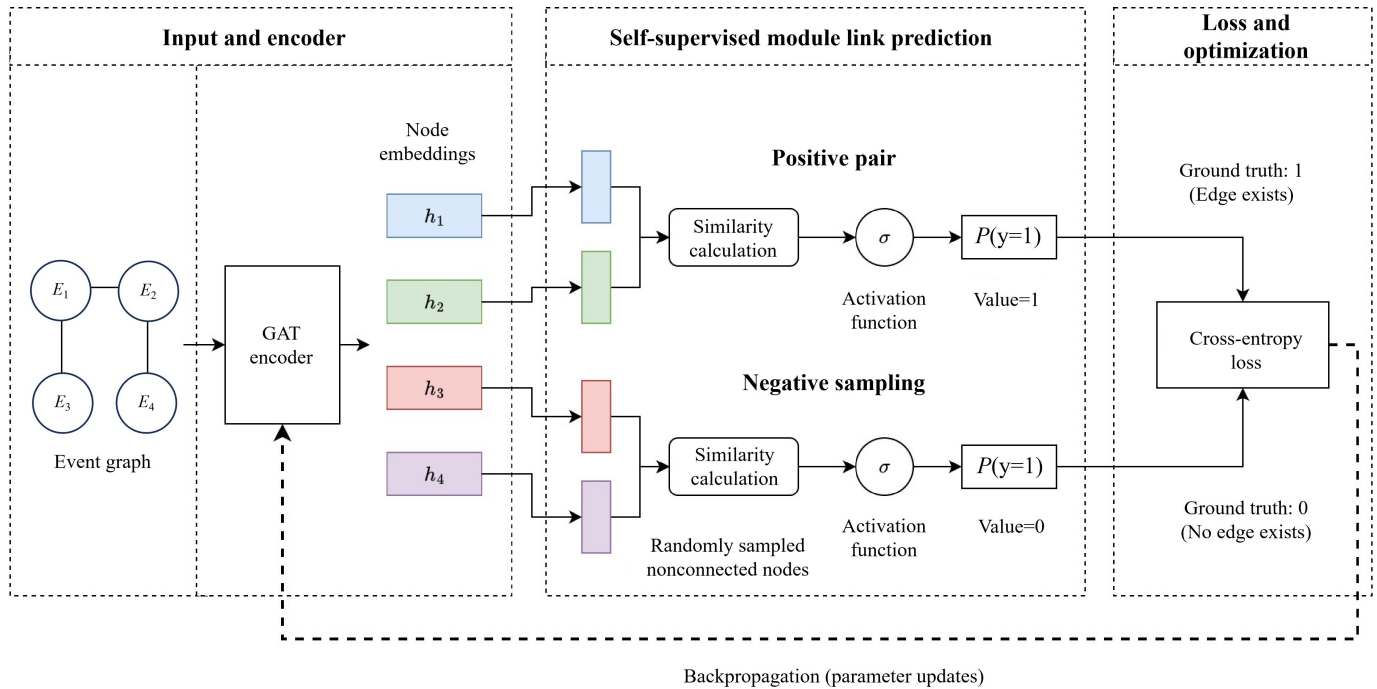
Intuitively, the self-monitoring module acts as an "internal auditor." While GAT makes initial predictions about event relevance based on local connections, the DBN provides a global thematic template. The module calculates the "surprise" or discrepancy between these two (via Kullback-Leibler divergence) and forces the model to self-correct, ensuring that local narrative links do not deviate from the overall life story logic.

The optimal number of clusters  $K$  in the joint adjustment module is determined using the silhouette coefficient and the elbow method applied to a validation subset of narratives. To prevent overfitting, we use L2 regularization and dropout (rate=0.5) within the GAT layers. Furthermore, the self-supervised loss function itself acts as a structural regularizer, penalizing the model when it focuses excessively on transient noise in individual narratives at the expense of global consistency.

The joint loss function is configured as  $L_{total} = \lambda_1 L_{GAT} + \lambda_2 L_{joint}$ , where  $\lambda_1 = 0.7$  and  $\lambda_2 = 0.3$  were empirically chosen to prioritize structural accuracy while maintaining global coherence. Optimization was performed using the AdamW routine with a weight decay of 0.01 and a linear learning rate scheduler that warmed up for the first 10% of total training steps.

Figure 4 illustrates the operational mechanism of the self-supervised module.

**Figure 4.** Self-supervised link prediction. A graph attention network (GAT) encoder generates node embeddings  $h_i$  from the input graph. Positive and negative pairs are evaluated via dot products and sigmoid activation. The model minimizes cross-entropy loss through backpropagation to update encoder parameters and optimize latent representations.

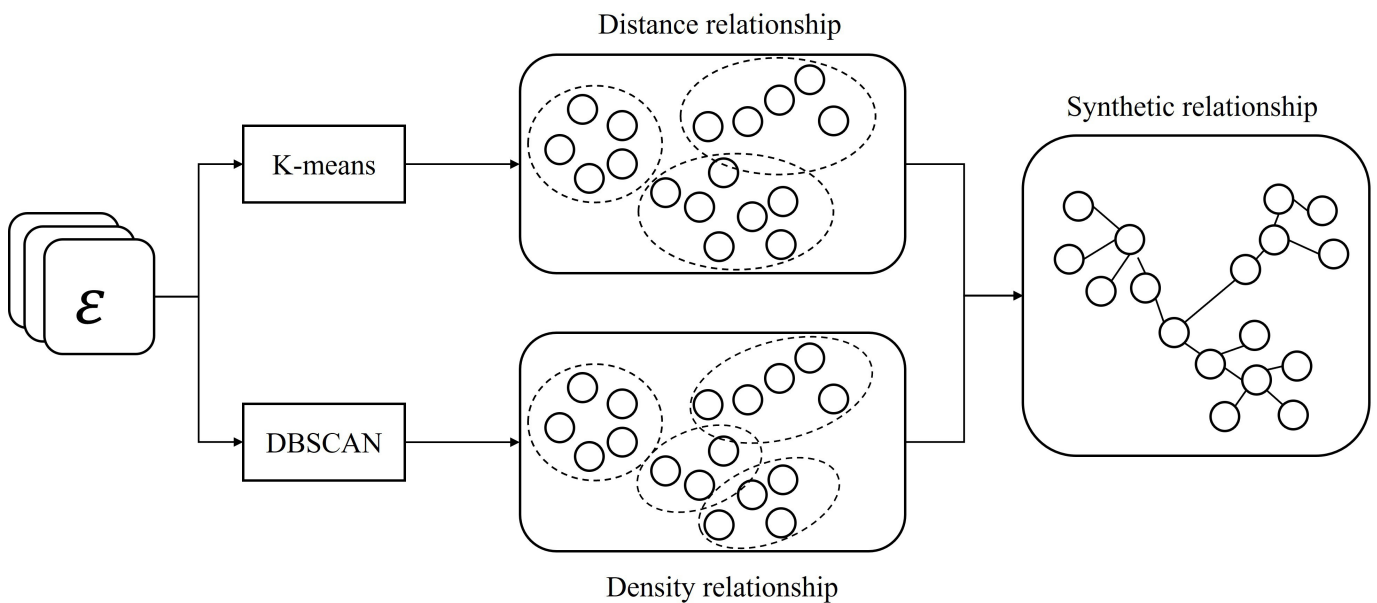


**Joint Relevance Adjustment**

The proposed method integrates k-means and DBSCAN clustering to compute event correlations through a dual-metric approach, as shown in Figure 5. K-means

evaluates co-clustered events using relative centroid distances, while DBSCAN assesses connectivity through minimum path lengths between nodes. Both algorithms assign zero correlation to nonclustered event pairs.

**Figure 5.** Framework of the joint adjustment module combining geometric and density metrics. The process integrates k-means (based on Euclidean distance to centroids) and density-based spatial clustering of applications with noise (DBSCAN; based on local density connectivity). The final correlation score is a weighted fusion of these 2 metrics to ensure structural robustness in narrative clustering.



The final correlation coefficient synergistically combines both clustering results, leveraging k-means’ geometric precision and DBSCAN’s density awareness. This hybrid approach captures complementary aspects of event relationships while maintaining computational efficiency through conditional zero assignments for unrelated events.

**Timeline Generation**

To address temporal ambiguity in life stories, we first standardize diverse time expressions (eg, relative dates like “three years ago” or historical references like “Liberation War”). For events lacking absolute dates, we calculate cosine

similarity between their text features and a curated historical event database, assigning standardized timestamps when similarity exceeds 0.85. This process ensures consistent temporal representation across both precise and ambiguous event references.

The generation algorithm first arranges temporally annotated events chronologically as the timeline backbone. For undated events, it calculates semantic similarity with existing timeline nodes to determine optimal insertion points between adjacent events. To integrate undated events  $E_u$  into the anchor timeline  $T_{anchor} = \{e_1, e_2, \dots, e_n\}$ , we propose a relevance-based greedy insertion algorithm. For each  $e_u \in E_u$ , the optimal position  $i$  is determined by:

$$i^* = \arg \max_{0 \leq i \leq n} (\alpha \cdot Rel(e_i, e_u) + (1 - \alpha) \cdot Rel(e_u, e_{i+1})) \quad (21)$$

where  $e_0$  and  $e_{n+1}$  are virtual start and end nodes. If multiple positions yield the same maximum score, the tie is resolved by selecting the position that maintains the original narrative proximity in the transcript. This dual consideration of temporal sequence and semantic coherence produces timelines that maintain both chronological accuracy and narrative flow, essential for effective reminiscence therapy applications.

The cosine similarity threshold of 0.85 was empirically determined through a sensitivity analysis. In our pilot tests, thresholds above 0.95 resulted in a 36% loss of valid temporal anchors, while thresholds below 0.75 introduced significant “semantic drift,” where personal anecdotes were incorrectly mapped to unrelated historical events. To mitigate the impact of potential misalignment on timeline validity, we introduced a “confidence flag” mechanism: events with similarity scores between 0.70 and 0.85 are flagged for human-in-the-loop verification by caregivers. Error analysis

indicates that inaccuracies primarily occur in narratives with vague temporal adverbs (eg, “back then”), which the system addresses by placing them in a relative sequence based on narrative flow rather than fixed historical coordinates.

The Historical Event Database used for mapping undated narratives contains approximately 5000 major historical events spanning from 1900 to 2000, curated from open-source repositories including Wikipedia and Baidu Baike. It focuses on events highly relevant to the Chinese older adult cohort (eg, social movements, economic reforms, and significant natural disasters). The database is maintained in Chinese and is licensed under CC-BY-SA 4.0 for research purposes.

## Story Mosaic System

The core architecture of the Story Mosaic system is built around three principal functional modules: (1) a story input module supporting multiple formats, including text, images, and videos, with temporal annotations available for manual or automatic addition; (2) a theme-based content organization module that allows for customizable categorization (eg, “Career” and “Family Life”); and (3) a chronological timeline generation module for key life events [17].

In the preliminary work of the Story Mosaic system, the generation of life event timelines, while algorithmically automated, exhibited limited accuracy, with event elements largely reliant on manual annotation during story input. In this study, the timeline generation module has been specifically optimized based on the CARE-ET method. This enhanced module leverages the structured framework of the CARE-ET approach, integrating three core functions: event extraction, temporal sequencing, and narrative coherence, as shown in Figure 6. It enables the automated generation of accurate event summaries and timelines, thereby alleviating the data entry burden on caregivers.

**Figure 6.** Schematic diagram of the automatic event feature generation module, supporting 1-click generation of event features.

The screenshot displays the 'Story Mosaic' application interface. At the top, a navigation bar contains 'Story Mosaic', 'Story Visualization', 'Story Management', and 'Settings'. The main content area is titled 'Life Story > Mr. Zhang'. A prominent 'One-click generation' button is located at the top right of the form. The form includes several input fields: 'Topic' (with a placeholder 'Please set a topic'), 'Life Story' (a text area containing a paragraph about Grandpa Zhang), 'Type' (set to 'Study'), 'Time' (set to 'Autumn'), 'Location' (set to 'Calligraphy class'), 'Participants' (set to 'Grandpa Zhang, Teacher Li, Grandma Wang'), and 'Summary' (set to 'After retiring, Grandpa Zhang joins a calligraphy class.'). At the bottom, there are sections for 'Picture' and 'Video' with upload instructions and buttons for 'OK' and 'Cancel'.

## Results

We conducted a comprehensive evaluation of the proposed system, organized into two key aspects: (1) quantitative assessment of event representation and timeline quality, and (2) systematic usability testing.

### Datasets

We selected 2 datasets for the experiments to verify the effectiveness of the event timeline generation method. The first is the self-constructed OALS (Older Adults' Life Stories) dataset, which contains the event data TimID (timeline identity) that has been annotated with timelines. The second is the Chinese news summary dataset LCSTS (Large Scale Chinese Short Text Summarization), organized by the Harbin Institute of Technology, which is constructed based on Sina microblog text. For detailed information about the datasets, please refer to [Multimedia Appendix 1](#).

### Experimental Setup

Our model was implemented using PyTorch 1.12 (PyTorch Foundation) and trained on an NVIDIA Tesla T4 GPU (16GB VRAM). The ALBERT-base-v2 model was fine-tuned for 10 epochs with a learning rate of  $2 \times 10^{-5}$ , a batch size of 16, and a weight decay of 0.01 using the AdamW optimizer. Early stopping was implemented with a patience of 3 epochs based on validation loss. Random seeds were fixed at 42 for reproducibility. The more detailed parameters are listed in [Table 1](#). For the CARE-ET framework, the GAT neighbor selection threshold was set to 0.6, and the joint adjustment threshold to 0.8. An 80/10/10 split was used for training, validation, and testing. To facilitate reproducibility, the code is provided as supplementary material for review and will be made public upon acceptance. The more detailed parameters are listed in [Table 2](#).

**Table 1.** Training hyperparameters.

Parameter	Value
Framework	PyTorch 1.12 (PyTorch Foundation)
Hardware	NVIDIA Tesla T4 GPU (16GB VRAM)
Base model	ALBERT <sup>a</sup> -base-v2
Fine-tuning epochs	10
Batch size	16
Learning rate	$2.00 \times 10^{-5}$
Weight decay	0.01 (AdamW)
Optimizer	AdamW
Dropout rate	0.5
Early stopping	Patience=3 epochs
Random seed	42
Dataset split	80% / 10% / 10% (Train/Val/Test)

<sup>a</sup>ALBERT: A Lite Bidirectional Encoder Representations from Transformers for Self-supervised Learning of Language Representations.

**Table 2.** CARE-ET (CARE event timeline) and timeline generation parameters.

Parameter	Symbol and description	Value
GAT <sup>a</sup> neighbor selection threshold	$N_i$ (selection)	0.6
Joint adjustment threshold	$I_i$ (second-order neighbor)	0.8
Structural loss weight	$\lambda_1$ (structural accuracy)	0.7
Global coherence weight	$\lambda_2$ (consistency)	0.3
Temporal anchor similarity	Cosine threshold for history	0.85
Confidence flag range	For human-in-the-loop verification	[0.70, 0.85]
Coreference resolution window size	Sliding window for ALBERT <sup>b</sup>	512
Coreference resolution threshold	Cosine distance threshold	0.8

<sup>a</sup>GAT: graphic attention layer.

<sup>b</sup>ALBERT: A Lite Bidirectional Encoder Representations from Transformers for Self-supervised Learning of Language Representations.

The CARE-ET model was trained on a dataset of 3142 annotated narrative segments derived from 118 sessions. The annotation workflow involved 2 gerontology specialists who independently labeled event triggers and temporal relations; interannotator agreement reached a Cohen  $\kappa$  of 0.77. For model evaluation, we used a 5-fold cross-validation

strategy to ensure the robustness of the results across different storytelling styles.

To ensure the statistical rigor of our results, we conducted bootstrap significance testing ( $n=1000$ ) and used the paired

2-tailed  $t$  test to compare CARE-ET against the strongest baseline. Statistical significance is  $P < .05$ .

## Event Summary Assessment

Life story event summarization enables caregivers to efficiently retrieve relevant narratives while providing the structural foundation for organizing fragmented life stories, with evaluation conducted using Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metrics [33]. ROUGE is a commonly used evaluation method for automated text summarization, which measures the quality of summaries by calculating the overlap between automated summaries and human-referenced summaries for unigram words (ROUGE-1), bigram words (ROUGE-2), and the longest common subsequence (ROUGE-L). The method is simple, intuitive, highly consistent with manual evaluation results, and is one of the most effective evaluation metrics for current text summarization tasks.

For automated summarization evaluation, we report the  $F_1$ -scores of ROUGE-1, ROUGE-2, and ROUGE-L. All metrics were computed using the Python *rouge-score* package with Porter stemming enabled and standard English stop-words removed. To ensure the reliability of performance gains, we conducted bootstrap significance testing ( $n=1000$ ) to report 95% CIs and  $P$  values against the baselines.

Table 3 and Table 4 present the experimental evaluations on both the OALS and LCSTS datasets [34], demonstrating CARE-ET's superior performance in event summarization tasks compared to several baseline methods (evolutionary timeline summarization, Timeline-sumy, temporally sensitive submodularity framework for timeline summarization, crisis local timeline line-up summarization, pretraining with extracted gap-sentences for abstractive summarization, LongT5, and pyramid-based masked sentence pretraining for multidocument summarization) [23-25,35-38]. Using median and mean text lengths as input parameters (58/210 bytes for OALS; 234/248 bytes for LCSTS), CARE-ET consistently achieved higher accuracy across all test conditions, with particularly notable advantages in full-length input scenarios, showing 0.6% (SD 1.20%) and 1.6% (SD 0.69%) average improvements over the strongest baseline in OALS and LCSTS, respectively. These results validate the model's effectiveness in extracting structured event summaries from both domain-specific life story data (OALS) and general narrative texts (LCSTS), confirming its robust generalization capabilities while maintaining processing accuracy across variable input lengths.

**Table 3.** Performance on the OALS (Older Adults' Life Stories) dataset using different lengths.

Method	Length limited within 60 bytes (%), mean (SD)			Length limited within 220 bytes (%), mean (SD)			Full length (%), mean (SD)		
	ROUGE-1 <sup>a</sup>	ROUGE-2 <sup>b</sup>	ROUGE-L <sup>c</sup>	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
ETS <sup>d</sup>	26.4 (1.15)	13.9 (0.72)	19.5 (0.98)	31.1 (1.28)	14.3 (0.65)	26.3 (1.05)	34.9 (1.42)	15.1 (0.74)	30.1 (1.25)
Timeline-Sumy	26.4 (1.08)	13.9 (0.68)	19.5 (0.91)	31.1 (1.21)	14.3 (0.61)	26.3 (0.98)	34.9 (1.35)	15.1 (0.71)	30.1 (1.18)
TSSF-TLS <sup>e</sup>	26.9 (1.02)	15.5 (0.58)	19.9 (0.85)	31.3 (1.15)	15.9 (0.55)	26.9 (0.92)	35.8 (1.28)	15.9 (0.65)	30.8 (1.12)
CrisisLTLSum <sup>f</sup>	27.5 (0.95)	16.9 (0.62)	20.9 (0.78)	32.1 (1.08)	17.7 (0.52)	27.5 (0.88)	37.2 (1.15)	16.9 (0.68)	31.6 (1.05)
PEGASUS <sup>g</sup>	27.3 (0.91)	17.2 (0.55)	21.5 (0.82)	32.5 (1.02)	18.1 (0.48)	27.9 (0.85)	37.4 (1.12)	17.2 (0.58)	31.9 (0.98)
LongT5 <sup>h</sup>	27.8 (0.88)	17.1 (0.52)	21.3 (0.75)	32.4 (0.98)	17.9 (0.45)	28.1 (0.82)	37.7 (1.08)	17.1 (0.55)	32.1 (0.95)
PRIMERA <sup>i</sup>	28.1 (0.82)	17.1 (0.48)	21.7 (0.68)	32.6 (0.92)	17.9 (0.42)	28.2 (0.78)	37.6 (0.98)	17.9 (0.52)	32.6 (0.88)
CARE-ET <sup>j</sup> (Word2Vec)	27.9 (0.85)	17.6 (0.51)	21.9 (0.72)	32.7 (0.95)	18.2 (0.45)	28.4 (0.81)	37.9 (1.02)	16.9 (0.55)	32.4 (0.91)
CARE-ET (LSTM <sup>k</sup> )	28.5 (0.78)	17.7 (0.45)	21.7 (0.65)	33.1 (0.88)	18.6 (0.38)	28.7 (0.75)	38.1 (0.95)	17.4 (0.48)	32.8 (0.85)
CARE-ET	28.3 (0.72) <sup>l</sup>	17.5 (0.42) <sup>l</sup>	22.1 (0.61) <sup>l</sup>	33.2 (0.85) <sup>l</sup>	18.8 (0.35) <sup>l</sup>	28.9 (0.72) <sup>l</sup>	38.5 (0.92) <sup>l</sup>	17.7 (0.45) <sup>l</sup>	33.2 (0.82) <sup>l</sup>

<sup>a</sup>ROUGE-1: Recall-Oriented Understudy for Gisting Evaluation, variant 1.

<sup>b</sup>ROUGE-2: Recall-Oriented Understudy for Gisting Evaluation, variant 2.

<sup>c</sup>ROUGE-L: Recall-Oriented Understudy for Gisting Evaluation–Longest Common Subsequence.

<sup>d</sup>ETS: evolutionary timeline summarization.

<sup>e</sup>TSSF-TLS: temporally sensitive submodularity framework for timelines summarization.

<sup>f</sup>CrisisLTLSum: crisis local timeline line-up summarization.

<sup>g</sup>PEGASUS: pretraining with extracted gap-sentences for abstractive summarization.

<sup>h</sup>LongT5: long text-to-text transfer transformer.

<sup>i</sup>PRIMERA: pyramid-based masked sentence pretraining for multidocument summarization.

<sup>j</sup>CARE-ET: CARE event timeline.

<sup>k</sup>LSTM: long short-term memory.

<sup>l</sup>Statistically significant,  $P < .05$ .

**Table 4.** Performance on the LCSTS (Large Scale Chinese Short Text Summarization) dataset using different lengths.

Method	Length limited within 234 bytes (%), mean (SD)			Length limited within 248 bytes (%), mean (SD)			Full length (%), mean (SD)		
	ROUGE-1 <sup>a</sup>	ROUGE-2 <sup>b</sup>	ROUGE-L <sup>c</sup>	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
ETS <sup>d</sup>	25.4 (0.72)	12.8 (0.35)	22.1 (0.52)	27.2 (0.78)	14.1 (0.38)	24.5 (0.61)	29.1 (0.82)	16.3 (0.45)	26.2 (0.68)
Timeline-Sumy	25.5 (0.68)	12.9 (0.32)	22.2 (0.48)	27.3 (0.74)	14.2 (0.35)	24.6 (0.55)	29.2 (0.78)	16.4 (0.42)	26.3 (0.65)
TSSF-TLS <sup>e</sup>	26.4 (0.65)	14.1 (0.28)	23.3 (0.45)	28.5 (0.71)	15.6 (0.32)	25.7 (0.52)	30.2 (0.75)	17.5 (0.41)	27.4 (0.62)
CrisisLTLSum <sup>f</sup>	29.1 (0.58)	16.5 (0.25)	26.2 (0.38)	31.4 (0.62)	18.2 (0.30)	28.3 (0.48)	33.1 (0.68)	20.2 (0.38)	30.1 (0.58)
PEGASUS <sup>g</sup>	34.2 (0.55)	21.3 (0.32)	31.1 (0.42)	36.5 (0.65)	23.1 (0.38)	33.2 (0.51)	38.1 (0.71)	24.9 (0.42)	34.8 (0.61)
LongT5 <sup>h</sup>	34.6 (0.51)	21.6 (0.28)	31.5 (0.38)	36.8 (0.58)	23.4 (0.32)	33.7 (0.48)	38.5 (0.65)	25.2 (0.38)	35.3 (0.55)
PRIMERA <sup>i</sup>	34.5 (0.48)	21.8 (0.25)	31.2 (0.35)	36.7 (0.55)	23.6 (0.28)	33.5 (0.45)	38.4 (0.62)	25.4 (0.35)	35.1 (0.52)
CARE-ET <sup>j</sup> (Word2Vec)	34.9 (0.45)	22.1 (0.22)	31.8 (0.32)	37.1 (0.52)	23.9 (0.25)	33.9 (0.42)	38.9 (0.58)	25.8 (0.32)	35.8 (0.48)
CARE-ET (LSTM <sup>k</sup> )	35.3 (0.42)	22.4 (0.21)	32.2 (0.28)	37.4 (0.48)	24.2 (0.22)	34.2 (0.38)	39.2 (0.55)	26.1 (0.31)	36.2 (0.45)
CARE-ET	35.7 (0.38) <sup>l</sup>	22.7 (0.18) <sup>l</sup>	32.8 (0.25) <sup>l</sup>	37.8 (0.42) <sup>l</sup>	24.5 (0.21) <sup>l</sup>	34.8 (0.35) <sup>l</sup>	39.5 (0.52) <sup>l</sup>	26.4 (0.28) <sup>l</sup>	36.9 (0.42) <sup>l</sup>

<sup>a</sup>ROUGE-1: Recall-Oriented Understudy for Gisting Evaluation, variant 1.

<sup>b</sup>ROUGE-2: Recall-Oriented Understudy for Gisting Evaluation, variant 2.

<sup>c</sup>ROUGE-L: Recall-Oriented Understudy for Gisting Evaluation–Longest Common Subsequence (ROUGE-L).

<sup>d</sup>ETS: evolutionary timeline summarization.

<sup>e</sup>TSSF-TLS: temporally sensitive submodularity framework for timeline summarization.

<sup>f</sup>CrisisLTLSum: crisis local timeline line-up summarization.

<sup>g</sup>PEGASUS: pretraining with extracted gap-sentences for abstractive summarization.

<sup>h</sup>LongT5: long text-to-text transfer transformer.

<sup>i</sup>PRIMERA: pyramid-based masked sentence pretraining for multidocument summarization.

<sup>j</sup>CARE-ET: CARE event timeline.

<sup>k</sup>LSTM: long short-term memory.

<sup>l</sup>Statistically significant,  $P < .05$ .

To evaluate the specific contributions of each component, we conducted a rigorous ablation study. Table 5 indicates that replacing ALBERT with Word2Vec leads to a 1% drop in  $F_1$ -score, highlighting the necessity of contextual embeddings. Furthermore, the GAT-DBN synergy outperforms GAT-only and DBN-only configurations by 2.8% and

3.4%, respectively. The self-supervised module provides an additional 0.9% gain in timeline consistency. Our model contains 64M parameters, and the average training time per fold is approximately 4.5 (SD 0.34) hours on a Tesla T4 GPU, representing a balanced tradeoff between complexity and performance.

**Table 5.** Ablation experiment results.

Configuration	ROUGE-1 <sup>a</sup>	ROUGE-2 <sup>b</sup>	ROUGE-L <sup>c</sup>	SLEU <sup>d</sup>
CARE-ET <sup>e</sup> (Full)	38.5 <sup>f</sup>	17.7	33.2	0.72
Without ALBERT <sup>g</sup> (Word2Vec)	37.4	16.8	32.2	0.69
Without BiLSTM <sup>h</sup> (LSTM <sup>i</sup> )	37.3	16.5	32.0	0.68
Without DBN <sup>j</sup> (GAT <sup>k</sup> only)	35.7	15.2	30.4	0.61
Without GAT (DBN only)	35.1	14.8	29.8	0.58
Without self-supervised module	37.6	17.2	32.3	0.65
Without all (baseline Seq2Seq)	34.2	13.5	28.5	0.52

<sup>a</sup>ROUGE-1: Recall-Oriented Understudy for Gisting Evaluation, variant 1.

<sup>b</sup>ROUGE-2: Recall-Oriented Understudy for Gisting Evaluation, variant 2.

<sup>c</sup>ROUGE-L: Recall-Oriented Understudy for Gisting Evaluation–Longest Common Subsequence.

<sup>d</sup>SLEU: sequence link evaluation.

<sup>e</sup>CARE-ET: CARE event timeline.

<sup>f</sup>Italicized numbers indicate the best result among all experimental results.

<sup>g</sup>ALBERT: A Lite Bidirectional Encoder Representations from Transformers for Self-supervised Learning of Language Representations.

<sup>h</sup>BiLSTM: bidirectional long short-term memory.

<sup>i</sup>LSTM: long short-term memory.

<sup>j</sup>DBN: deep belief network.

<sup>k</sup>GAT: graphic attention layer.

## Qualitative Analysis

To further understand the performance of CARE-ET, we present 2 anonymized case studies comparing the generated timelines against the therapeutic ground truth.

### Case 1 (Success in Long-Range Dependencies)

During an interview, the participant mentioned family-related milestones (eg, marriage and grandchildren) and their prime years in a later segment. CARE-ET effectively captured the cross-segment dependencies, chronologically positioning these events after related life stages while maintaining the semantic links between the memories.

### Case 2 (Failure in Near-Duplicate Events)

During a session discussing a wedding, the participant mentioned the date and event details in different segments. The model failed to merge these, creating two independent nodes for the same event (a redundancy error). This failure suggests that current GAT-based relevance thresholding remains insufficient for fine-grained deduplication in repetitive, fragmented oral testimonies.

## Quality Assessment of Timelines

In this paper, sequence link evaluation (SLEU) metrics and TimID metrics are used to assess the quality of event timelines.

### Sequence Link Evaluation

In this paper, the evaluation metrics of SLEU are used to verify the quality of the timeline [22]. The sentences in the system-generated storyline are  $s^1, s^2, \dots, s^n$ , where  $s^1$  is the beginning of the story and  $s^n$  is the end of the story. The sentences of the reference storyline are  $s'_1, s'_2, \dots, s'_n$ .  $n$  is the number of stages, and  $h$ -gram and  $h$ -gram' denote a sequence of  $h$  sentences in a storyline, not necessarily consecutive, generated by the reference storyline and the generated storyline. The collocation of information between sentences in a storyline is utilized to compute the probability of the occurrence of the storyline and thus determine whether the storyline is complete or coherent.

The cosine similarity between  $s^i (i \in [1, n])$  and  $s^j (j \in [1, n])$  is first calculated. Then the probability of  $h$ -gram matching between the 2 storylines  $P_h$  is calculated separately:

$$P_h = \frac{\min(\text{Count}_{\text{match}}(h\text{-gram}), \text{Count}_{\text{match}}(h\text{-gram}'))}{\text{Count}(h\text{-gram})} \quad (22)$$

The possible values of the variable  $h$  depend on the specific case and are represented by the set  $H$ . The set  $H$  is a set of 1-grams. Storylines using 1-grams tend to satisfy the synthesis requirement more easily. Longer  $h$ -gram matches indicate better coherence. The count ( $h$ -gram) denotes the number of comparable  $h$ -grams (eg, if  $h=4$ , count(2-gram)=6). The count( $h$ gram') equals count( $h$ -gram). The count<sub>match</sub>( $h$ -gram) denotes the number of  $h$ -grams that can

be matched (ie,  $h$ -grams against  $h$ -grams). The count<sub>match</sub>( $h$ -gram) denotes the number of  $h$ -grams that can be matched (ie, the number of  $h$ -gram pairs for which the similarity of the corresponding sentences exceeds a specific threshold  $\lambda$ ). In the experiments, we set  $\lambda=0.2$ . Then  $\text{SLEU} = \exp(\sum_{h \in H} w_h \log p_h)$  is computed. The SLEU metric value ranges from 0 to 1 and is used to quantitatively assess the quality of the storyline.

We standardized the timeline consistency metric as SLEU. It evaluates the structural integrity of the generated timeline by calculating the weighted sum of correct temporal transitions:

$$\text{SLEU} = \sum_{h=1}^H w_h \cdot \exp(-\lambda \cdot |pos_{\text{pred}} - pos_{\text{ref}}|) \quad (23)$$

where  $H = 3$  denotes the maximum historical hop distance considered,  $w_h$  represents the weight assigned to the  $h$ -th hop (set as  $1/h$ ), and  $\lambda = 0.5$  is the decay constant for distance penalties.

### TimID

The study introduces TimID, a temporal annotation marker embedded within a novel life story dataset for older adults, enabling precise evaluation of event chronology through 2 complementary dimensions. Absolute position accuracy quantifies deviations between generated event nodes and their corresponding TimID reference points, where smaller positional differences indicate higher timeline fidelity. Simultaneously, relative order preservation assesses the correctness of pairwise event sequences by verifying their alignment with ground-truth temporal relationships. This dual-metric approach ensures comprehensive validation of both event placement precision and chronological logic in life story reconstruction.

TimID is defined to measure the precision of event placement. The "absolute position" of an event is defined as its normalized rank  $r_i = i/N$ , where  $N$  is the total number of events. The relative order preservation is computed using the Kendall  $\tau$  coefficient, which quantifies the ordinal correlation between the predicted sequence and the gold-standard therapeutic timeline.

## Experimental Results Based on SLEU and TimID

The study uses 2 complementary assessment frameworks: SLEU evaluates storyline coherence through sentence-level probability analysis, while TimID assesses timeline quality via absolute position accuracy and relative order consistency. As shown in Table 6, CARE-ET outperforms baseline methods across both metrics, achieving superior SLEU scores (0.021-0.061 higher) and TimID performance (0.08%-0.24% % better relative order accuracy, with 0.8-1.7 lower positional deviations). These results demonstrate significant improvements in both narrative flow and temporal precision.

**Table 6.** Performance on the dataset OALS (Older Adults' Life Stories) using sequence link evaluation (SLEU) and timeline identity (TimID).

Method	SLEU	TimID-ap <sup>a</sup>	TimID-rp <sup>b</sup>
ETS <sup>c</sup>	0.47 (0.02)	4.51 (0.38)	0.43 (0.03)
Timeline-Sumy	0.48 (0.02)	4.12 (0.25)	0.51 (0.03)
TSSF-TLS <sup>d</sup>	0.47 (0.02)	4.21 (0.31)	0.48 (0.03)
CrisisLTLSum <sup>e</sup>	0.51 (0.02)	3.64 (0.18)	0.59 (0.02)
CARE-ET <sup>f</sup>	0.53 (0.01) <sup>g</sup>	2.79 (0.15) <sup>g</sup>	0.67 (0.02) <sup>g</sup>

<sup>a</sup>TimID-ap: timeline identity–absolute position.

<sup>b</sup>TimID-rp: timeline identity–relatable position.

<sup>c</sup>ETS: evolutionary timeline summarization.

<sup>d</sup>TSSF-TLS: temporally sensitive submodularity framework for timeline summarization.

<sup>e</sup>CrisisLTLSum: crisis local timeline line-up summarization.

<sup>f</sup>CARE-ET: CARE event timeline.

<sup>g</sup>Statistically significant,  $P < .05$ .

The metrics validate CARE-ET's dual capacity to generate clinically meaningful timelines: SLEU confirms enhanced narrative coherence, which is crucial for therapeutic applications, while TimID verifies improved chronological accuracy that aligns with human cognition. This combination of strong textual fluency and precise event sequencing enhances both the interpretability of life stories for caregivers and their therapeutic use in reminiscence interventions, advancing digital health solutions for older adult care.

### System Usability Assessment

This study has established cooperative relationships with the Hefei Tianyu Elderly Care Institution and the Chuzhou Lekang Elderly Care Institution. The system has been deployed within the daily care management platforms of both institutions, serving a total of 36 older adult users (21 at the Tianyu Institution and 15 at the Lekang Institution). For the optimized Story Mosaic system, we invited 4 caregivers and 1 manager from both the Tianyu and Lekang senior

care facilities to complete the System Usability Scale (SUS) [39]. The statistical information for the 10 participants is shown in Table 7. Four participants (A1, C2, C3, and C4) had previously participated in the feasibility study of our published Story Mosaic system, while the other 6 were newly recruited for this study. The participants included 2 core roles—administrators and caregivers—with balanced gender representation. Aged 35 to 62 years, they represented young, middle-aged, and senior practitioner cohorts, while their elder care tenure (4–23 y) reflected experiential disparities among novice, proficient, and senior workers. Their educational qualifications showed a gradient distribution, ranging from no formal education to master's degrees. Participant selection ensured diversity across roles, sex, age, tenure, education, and affiliated institutions. This sampling strategy represented the target user groups and provided empirical support for the scientific rigor and objectivity of the system usability evaluation.

**Table 7.** Demographic characteristics of participants.

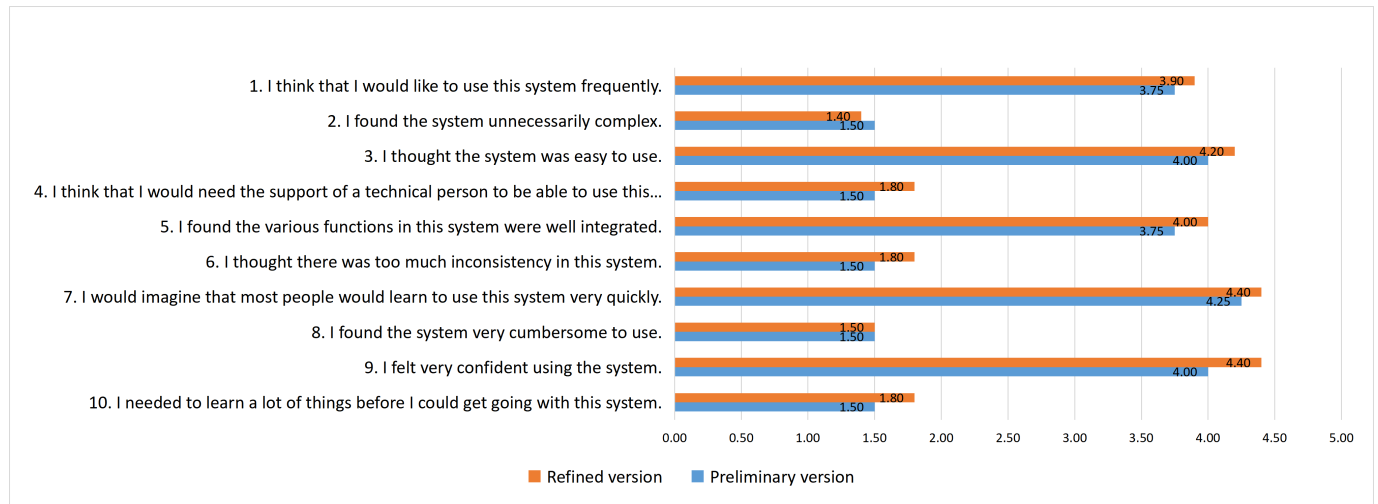
ID	Identity	Sex	Age (y)	Tenure	Education	Affiliated institution
A1	Administrator	Male	37	13	Master's degree	Tianyu
A2	Administrator	Female	52	23	Junior high school	Lekang
C1	Caregiver	Female	56	10	Primary school	Tianyu
C2	Caregiver	Female	48	6	Junior high school	Tianyu
C3	Caregiver	Male	36	5	Junior high school	Tianyu
C4	Caregiver	Female	57	8	Primary school	Tianyu
C5	Caregiver	Male	35	4	Bachelor's degree	Lekang
C6	Caregiver	Female	45	9	Junior high school	Lekang
C7	Caregiver	Female	62	14	Null <sup>a</sup>	Lekang
C8	Caregiver	Female	58	8	Primary school	Lekang

<sup>a</sup>Null: no education has been received.

In the preliminary work of the Story Mosaic system, the average SUS score was 82.63, achieving an A– grade. After enhancing the system based on the CARE-ET algorithm,

a usability test was conducted again, and the evaluation results are shown in Figure 7 with SUS scores from both the preliminary and refined versions.

**Figure 7.** System Usability Scale scores from the preliminary and refined versions.



A comparison of the refined and preliminary versions of the SUS scale shows the following: for positively worded items (1, 3, 5, 7, 9), the refined version scored higher, reflecting more positive evaluations of the system’s usability. For negatively worded items (2, 4, 6, 8, 10), the refined version still had relatively high scores, indicating that users perceived some remaining operational complexity. This is mainly because the refined version’s evaluators included caregivers with primary education or no formal education. While these users gave high scores on item 7 (“Most people can learn to use this system quickly”), their actual onboarding speed was slower. Moving forward, it will be necessary to simplify processes and add basic guidance for this group to improve their efficiency.

In addition to the above questions, we also asked an open-ended question: “What other ways do you think the timelines can help you care for older adults”? One participant noted:

*Timelines are critical in the care of older adults with cognitive impairment. These older adults' behavioral and psychological symptoms are closely related to their*

*past experiences. Timelines help us understand their experiences and act appropriately to the situation.*

### Objective Comparisons

To objectively evaluate the impact of the Story Mosaic system on cognitive load, we conducted a comparative study between the artificial intelligence–assisted workflow and a traditional manual workflow. We used the NASA Task Load Index, which assesses cognitive demand across 6 dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration [40]. The comparative evaluation results are shown in Table 8. Focusing on the 3 core dimensions of mental demand, temporal demand, and physical effort prioritized by participants, the Story Mosaic system achieved load reductions of 37.5%, 36.4%, and 42.9%, respectively, in these metrics. It precisely addressed users’ key needs by effectively alleviating cognitive pressure, time constraints, and energy input. Meanwhile, the system also optimized loads across other dimensions, cutting the overall workload by 37.4% and demonstrating excellent human-machine compatibility.

**Table 8.** NASA Task Load Index (NASA-TLX) evaluation of the manual workflow and the Story Mosaic system.

Metric	Manual workflow, mean (SD)	Story Mosaic system, mean (SD)	Relative reduction (%)
Mental demand	76.0 (13.1)	47.5 (23.6)	37.5
Physical demand	27.0 (7.50)	25.0 (8.5)	7.4
Temporal demand	77.0 (10.1)	49.0 (19.3)	36.4
Performance	43.0 (16.7)	31.0 (16.3)	27.9
Effort	73.5 (13.8)	42.0 (15.1)	42.9
Frustration	62.5 (14.4)	34.0 (14.3)	45.6
Overall workload	65.7 (13.6)	41.2 (16.9)	37.4

## Discussion

### Principal Results

The GAT framework incorporates dynamic node filtering using multilevel semantic similarity thresholds, automatically

preserving clinically meaningful life event relationships. This adaptive graph optimization enhances event relationship modeling precision while maintaining computational efficiency for real-world care applications.

CARE-ET demonstrates robust performance across both specialized life story datasets and general narratives through its novel integration of event triggers, spatiotemporal markers, and contextual encoding. The system outperforms conventional methods in extracting key life events from older adult narratives, providing a solid foundation for timeline construction.

Rigorous evaluation confirms that the generated timelines meet clinical standards for both narrative coherence and temporal accuracy. Caregiver assessments highlight superior readability and content generalization, validating the system's ability to transform fragmented life stories into structured therapeutic materials for cognitive reminiscence therapy.

## Limitations

While CARE-ET demonstrates promising results, several limitations warrant consideration. First, the method's performance relies heavily on the quality of temporal annotations (TimID) in the training data, which may limit generalizability to life stories with sparse or ambiguous time references. Second, the current implementation focuses primarily on textual narratives, potentially underutilizing multimodal cues (eg, emotional valence in vocal tones or visual context) that could enhance event interpretation. Third, the evaluation involved relatively small-scale caregiver assessments (n=10), necessitating broader clinical validation to confirm real-world utility across diverse older adult populations and care settings. Furthermore, the collection and processing of older adult individuals' personal life stories involve highly sensitive data with significant privacy implications. Although this study followed ethical protocols including approval, informed consent, and data anonymization, several challenges persist. These include

ensuring long-term data security, preserving participants' ongoing control over their narratives, and addressing potential emotional impacts from recalling personal memories—each representing key ethical considerations in this research domain.

Addressing these current limitations, future work will focus on advancing research in the following areas: First, the development of weakly supervised temporal annotation enhancement algorithms will leverage cross-modal temporal reasoning and the mining of implicit temporal cues to improve the model's robustness in parsing life stories with sparse temporal information. Second, the construction of a multimodal fusion framework that integrates speech emotion analysis, visual scene understanding, and textual semantic representation to establish a cross-modally aligned paradigm for life event comprehension. Third, collaboration with more elder care institutions to evaluate the long-term use of the system in real-world care scenarios.

## Conclusions

This study proposes the CARE-ET method, which enables the structured extraction of core events from life stories through multimodal event element fusion-based summary generation techniques and improves the accuracy of event relevance calculation via a collaborative computation model of GAT and DBN. Experiments on the OALS and LCSTS datasets show that CARE-ET significantly reduces redundancy and disorder in narrative texts, generating event timelines that outperform baseline methods in readability and structural rationality. This study provides an effective technical solution for the digital presentation of older adult life stories, with important application value for enhancing the personalization and precision of older adult care services.

## Funding

This work is supported by the Scientific Research Initiation Project (2024qd12 and 2025qd10) and the Anhui Provincial Department of Education Project (2025AHGXZK40376 and 2025AHGXZK40409).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Ethical approval of the dataset, data sources, construction basis, construction process, sample size, recruitment criteria, annotation specifications and processes for event elements and timelines, dataset public statement, etc.

[[DOCX File \(Microsoft Word File\), 57 KB-Multimedia Appendix 1](#)]

## References

1. Morley JE. An overview of cognitive impairment. *Clin Geriatr Med*. Nov 2018;34(4):505-513. [doi: [10.1016/j.cger.2018.06.003](https://doi.org/10.1016/j.cger.2018.06.003)] [Medline: [30336985](https://pubmed.ncbi.nlm.nih.gov/30336985/)]
2. Xu L, Fields NL, Daniel KM, Cipher DJ, Troutman BA. Reminiscence and digital storytelling to improve the social and emotional well-being of older adults with Alzheimer's disease and related dementias: protocol for a mixed methods study design and a randomized controlled trial. *JMIR Res Protoc*. Sep 7, 2023;12:e49752. [doi: [10.2196/49752](https://doi.org/10.2196/49752)] [Medline: [37676706](https://pubmed.ncbi.nlm.nih.gov/37676706/)]
3. Piolatto M, Bianchi F, Rota M, Marengoni A, Akbaritabar A, Squazzoni F. The effect of social relationships on cognitive decline in older adults: an updated systematic review and meta-analysis of longitudinal cohort studies. *BMC Public Health*. Feb 11, 2022;22(1):278. [doi: [10.1186/s12889-022-12567-5](https://doi.org/10.1186/s12889-022-12567-5)] [Medline: [35148704](https://pubmed.ncbi.nlm.nih.gov/35148704/)]
4. Barsuglia JP, Nedjat-Haiem FR, Shapira JS, et al. Observational themes of social behavioral disturbances in frontotemporal dementia. *Int Psychogeriatr*. Sep 2014;26(9):1475-1481. [doi: [10.1017/S104161021400091X](https://doi.org/10.1017/S104161021400091X)] [Medline: [24846824](https://pubmed.ncbi.nlm.nih.gov/24846824/)]

5. Chehrehnegar N, Shati M, Esmaeili M, Foroughan M. Executive function deficits in mild cognitive impairment: evidence from saccade tasks. *Aging Ment Health*. May 2022;26(5):1001-1009. [doi: [10.1080/13607863.2021.1913471](https://doi.org/10.1080/13607863.2021.1913471)] [Medline: [33928806](https://pubmed.ncbi.nlm.nih.gov/33928806/)]
6. Che J, Cheng N, Jiang B, et al. Executive function measures of participants with mild cognitive impairment: systematic review and meta-analysis of event-related potential studies. *Int J Psychophysiol*. Mar 2024;197:112295. [doi: [10.1016/j.ijpsycho.2023.112295](https://doi.org/10.1016/j.ijpsycho.2023.112295)] [Medline: [38266685](https://pubmed.ncbi.nlm.nih.gov/38266685/)]
7. Zhang Y, Zhu X, Yang S, Wong AKC, Chen X. Digital health technologies applied in patients with early cognitive change: scoping review. *J Med Internet Res*. Dec 29, 2025;27(1):e82881. [doi: [10.2196/82881](https://doi.org/10.2196/82881)] [Medline: [41461066](https://pubmed.ncbi.nlm.nih.gov/41461066/)]
8. Sommerlad A, Kivimäki M, Larson EB, et al. Social participation and risk of developing dementia. *Nat Aging*. May 2023;3(5):532-545. [doi: [10.1038/s43587-023-00387-0](https://doi.org/10.1038/s43587-023-00387-0)] [Medline: [37202513](https://pubmed.ncbi.nlm.nih.gov/37202513/)]
9. Gucciardi E, Jean-Pierre N, Karam G, Sidani S. Designing and delivering facilitated storytelling interventions for chronic disease self-management: a scoping review. *BMC Health Serv Res*. Jul 11, 2016;16(1):249. [doi: [10.1186/s12913-016-1474-7](https://doi.org/10.1186/s12913-016-1474-7)] [Medline: [27401836](https://pubmed.ncbi.nlm.nih.gov/27401836/)]
10. Sljivic H, Sutherland I, Stannard C, Ioppolo C, Morrisby C. Changing attitudes towards older adults: eliciting empathy through digital storytelling. *Gerontol Geriatr Educ*. 2022;43(3):360-373. [doi: [10.1080/02701960.2021.1900838](https://doi.org/10.1080/02701960.2021.1900838)] [Medline: [35837695](https://pubmed.ncbi.nlm.nih.gov/35837695/)]
11. Stargatt J, Bhar S, Bhowmik J, Al Mahmud A. Digital storytelling for health-related outcomes in older adults: systematic review. *J Med Internet Res*. Jan 12, 2022;24(1):e28113. [doi: [10.2196/28113](https://doi.org/10.2196/28113)] [Medline: [35019845](https://pubmed.ncbi.nlm.nih.gov/35019845/)]
12. Rios Rincon AM, Miguel Cruz A, Daum C, Neubauer N, Comeau A, Liu L. Digital storytelling in older adults with typical aging, and with mild cognitive impairment or dementia: a systematic literature review. *J Appl Gerontol*. Mar 2022;41(3):867-880. [doi: [10.1177/07334648211015456](https://doi.org/10.1177/07334648211015456)] [Medline: [34009053](https://pubmed.ncbi.nlm.nih.gov/34009053/)]
13. Zhu D, Al Mahmud A, Liu W. Digital storytelling intervention for enhancing the social participation of people with mild cognitive impairment: co-design and usability study. *JMIR Aging*. Jan 17, 2024;7(1):e54138. [doi: [10.2196/54138](https://doi.org/10.2196/54138)] [Medline: [38231541](https://pubmed.ncbi.nlm.nih.gov/38231541/)]
14. Zhu D, Al Mahmud A, Liu W. Effects of a mobile storytelling app (Huiyou) on social participation among people with mild cognitive impairment: pilot randomized controlled trial. *JMIR Hum Factors*. Jun 18, 2025;12(1):e70177. [doi: [10.2196/70177](https://doi.org/10.2196/70177)] [Medline: [40532206](https://pubmed.ncbi.nlm.nih.gov/40532206/)]
15. Ferstad K, Rykkje L. Understanding the significance of listening to older people's life stories in whole person care-an interview study of nurses in gerontology. *SAGE Open Nurs*. 2023;9:23779608231164077. [doi: [10.1177/23779608231164077](https://doi.org/10.1177/23779608231164077)] [Medline: [36949851](https://pubmed.ncbi.nlm.nih.gov/36949851/)]
16. An N, Gui F, Jin L, Ming H, Yang J. Toward better understanding older adults: a biography brief timeline extraction approach. *Int J Hum Comput Interact*. Mar 16, 2023;39(5):1084-1095. [doi: [10.1080/10447318.2022.2077278](https://doi.org/10.1080/10447318.2022.2077278)]
17. Gui F, Yang J, Wu Q, Liu Y, Zhou J, An N. Enhancing caregiver empowerment through the Story Mosaic system: human-centered design approach for visualizing older adult life stories. *JMIR Aging*. Nov 8, 2023;6:e50037. [doi: [10.2196/50037](https://doi.org/10.2196/50037)] [Medline: [37962517](https://pubmed.ncbi.nlm.nih.gov/37962517/)]
18. Valtolina S, Pugliese A. AI-assisted cognitive support for caregivers: a RAG and EUD framework for geriatric care. In: Santoro C, Schmidt A, Matera M, Bellucci A, editors. *End-User Development: 10th International Symposium, IS-EUD 2025, Munich, Germany, June 16–18, 2025, Proceedings*. 2025:205-220. [doi: [10.1007/978-3-031-95452-8\\_13](https://doi.org/10.1007/978-3-031-95452-8_13)]
19. Almeida PV, Rocha RP. AI-powered social robot for cognitive stimulation in the elderly through personalized storytelling. Presented at: 2025 11th International Conference on Automation, Robotics, and Applications (ICARA); Feb 12-14, 2025:122-126; Zagreb, Croatia. 2025.[doi: [10.1109/ICARA64554.2025.10977621](https://doi.org/10.1109/ICARA64554.2025.10977621)]
20. Li J, Cardie C. Timeline generation: tracking individuals on twitter. Presented at: WWW '14: 23rd International World Wide Web Conference; Apr 7-11, 2014:643-652; Seoul, Korea. [doi: [10.1145/2566486.2567969](https://doi.org/10.1145/2566486.2567969)]
21. Althoff T, Dong XL, Murphy K, Alai S, Dang V, Zhang W. TimeMachine: timeline generation for knowledge-base entities. Presented at: KDD '15: The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 10-13, 2015:19-28; Sydney, Australia. [doi: [10.1145/2783258.2783325](https://doi.org/10.1145/2783258.2783325)]
22. Sun W, Wang Y, Gao Y, Li Z, Sang J, Yu J. Comprehensive event storyline generation from microblogs. Presented at: Proceedings of the 1st ACM International Conference on Multimedia in Asia; Dec 16-18, 2019; Beijing China. Dec 15, 2019.[doi: [10.1145/3338533.3366601](https://doi.org/10.1145/3338533.3366601)]
23. Zhang J, Zhao Y, Saleh M, Liu P. PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. Presented at: Proceedings of the 37th International Conference on Machine Learning PMLR; Jul 13-18, 2020:11328-11339; 2020.URL: <https://proceedings.mlr.press/v119/zhang20ae/zhang20ae.pdf> [Accessed 2026-03-10]
24. Guo M, Ainslie J, Uthus D, et al. LongT5: efficient text-to-text transformer for long sequences. Presented at: Findings of the Association for Computational Linguistics; Jul 10-15, 2022:724-736; Seattle, United States. 2022.[doi: [10.18653/v1/2022.findings-naacl.55](https://doi.org/10.18653/v1/2022.findings-naacl.55)]

25. Xiao W, Beltagy I, Carenini G, Cohan A. PRIMERA: pyramid-based masked sentence pre-training for multi-document summarization. Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); May 22-27, 2022:5245-5263; Dublin, Ireland. 2022.[doi: [10.18653/v1/2022.acl-long.360](https://doi.org/10.18653/v1/2022.acl-long.360)]
26. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. ALBERT: a lite BERT for self-supervised learning of language representations. Presented at: 8th International Conference on Learning Representations; Apr 26-30, 2020. URL: <https://openreview.net/pdf?id=H1eA7AEtvS> [Accessed 2026-03-10]
27. Che W, Feng Y, Qin L, Liu T. N-LTP: an open-source neural language technology platform for Chinese. Presented at: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing; Nov 7-11, 2021:42-49; Online and Punta Cana, Dominican Republic. 2021.[doi: [10.18653/v1/2021.emnlp-demo.6](https://doi.org/10.18653/v1/2021.emnlp-demo.6)]
28. Liu X, Huang H, Zhang Y. Open domain event extraction using neural latent variable models. Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Jul 28 to Aug 2, 2019:2860-2871; Florence, Italy. 2019.[doi: [10.18653/v1/P19-1276](https://doi.org/10.18653/v1/P19-1276)]
29. Suo W, Muhammad BA, Zhang Z, et al. GAT-LS: a graph attention network for learning style detection in online learning environment. *J Netw Netw Appl.* 2024;4(2):60-72. [doi: [10.33969/J-NaNA.2024.040202](https://doi.org/10.33969/J-NaNA.2024.040202)]
30. Singh NA, Kishore K, Deo RN, Lu Y, Kodikara J. Automated segmentation framework for asphalt layer thickness from GPR data using a cascaded k-means–DBSCAN algorithm. *J Environ Eng Geophys.* Dec 1, 2022;27(4):179-189. [doi: [10.32389/JEEG22-019](https://doi.org/10.32389/JEEG22-019)]
31. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. Presented at: International Conference on Learning Representations (ICLR 2018); Apr 30 to May 3, 2017; Vancouver, Canada. 2017.URL: <https://openreview.net/pdf?id=rJXMpikCZ> [Accessed 2026-03-10]
32. Li J, Sun Q, Zhang F, Yang B. Meta-structure-based graph attention networks. *Neural Netw.* Mar 2024;171:362-373. [doi: [10.1016/j.neunet.2023.12.025](https://doi.org/10.1016/j.neunet.2023.12.025)] [Medline: [38134599](https://pubmed.ncbi.nlm.nih.gov/38134599/)]
33. Lin CY. Rouge: a package for automatic evaluation of summaries text summarization branches out. Presented at: Text Summarization Branches Out: Proceedings of the ACL-04 Workshop; Jul 25-27, 2004:74-81; Barcelona, Spain. 2004.URL: <https://aclanthology.org/W04-1013.pdf> [Accessed 2026-03-10]
34. Hu B, Chen Q, Zhu F. LCSTS: a large scale chinese short text summarization dataset. Presented at: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015); Sep 17-21, 2015:1967-1972; Lisbon, Portugal. [doi: [10.18653/v1/D15-1229](https://doi.org/10.18653/v1/D15-1229)]
35. Yan R, Wan X, Otterbacher J, Kong L, Li X, Zhang Y. Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. Presented at: SIGIR '11: The 34th International ACM SIGIR Conference on Research and Development in Information Retrieval; Jul 24-28, 2011:745-754; Beijing, China. [doi: [10.1145/2009916.2010016](https://doi.org/10.1145/2009916.2010016)]
36. Chang Y, Tang J, Yin D, Yamada M, Liu Y. Timeline summarization from social media with life cycle models. Presented at: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16); Jul 9-15, 2016:3698-3704; New York, NY. URL: <https://www.ijcai.org/Proceedings/16/Papers/520.pdf> [Accessed 2026-03-30]
37. Martschat S, Markert K. A temporally sensitive submodularity framework for timeline summarization. Presented at: Proceedings of the 22nd Conference on Computational Natural Language Learning; Nov 1-2, 2018; Brussels, Belgium. 2018.[doi: [10.18653/v1/K18-1023](https://doi.org/10.18653/v1/K18-1023)]
38. Rajaby Faghihi H, Alhafni B, Zhang K, Ran S, Tetreault J, Jaimes A. CrisisLTLsum: a benchmark for local crisis event timeline extraction and summarization. Presented at: Findings of the Association for Computational Linguistics: EMNLP 2022; Dec 7-11, 2022:5455-5477; Abu Dhabi, United Arab Emirates. URL: <https://aclanthology.org/2022.findings-emnlp.400.pdf> [Accessed 2026-03-10] [doi: [10.18653/v1/2022.findings-emnlp.400](https://doi.org/10.18653/v1/2022.findings-emnlp.400)]
39. Vlachogianni P, Tselios N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): a systematic review. *J Res Technol Educ.* May 27, 2022;54(3):392-409. [doi: [10.1080/15391523.2020.1867938](https://doi.org/10.1080/15391523.2020.1867938)]
40. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv Psychol.* 1988;52:139-183. [doi: [10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)]

## Abbreviations

- ALBERT:** A Lite Bidirectional Encoder Representations from Transformers for Self-supervised Learning of Language Representations
- BiLSTM:** bidirectional long short-term memory
- CARE-ET:** CARE event timeline
- DBN:** deep belief network
- DBSCAN:** density-based spatial clustering of applications with noise
- GAT:** graph attention network

**LCSTS:** Large Scale Chinese Short Text Summarization

**OALS:** Older Adults' Life Stories

**ROUGE:** Recall-Oriented Understudy for Gisting Evaluation

**SLEU:** sequence link evaluation

**SUS:** System Usability Scale

---

---

*Edited by Mamoun Mardini; peer-reviewed by Anne-Sophie Rigaud, Jiaoyun Yang, Lin Lyu, Lingling Yan, Yutao Yang, Zhenyu Liu; submitted 28.Aug.2025; final revised version received 15.Jan.2026; accepted 10.Feb.2026; published 06.Apr.2026*

*Please cite as:*

*Gui F, Yang M, Jin L, Qian J*

*AI-Enhanced Automatic Life Story Structuring for Reminiscence Therapy in Older Adults: Technical Feasibility Study*

*JMIR Aging 2026;9:e83122*

*URL: <https://aging.jmir.org/2026/1/e83122>*

*doi: [10.2196/83122](https://doi.org/10.2196/83122)*

© Fang Gui, Mengchen Yang, Liuqi Jin, Jing Qian. Originally published in JMIR Aging (<https://aging.jmir.org>), 06.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Aging, is properly cited. The complete bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.