Journal of Kerbala University, Vol. 22, Issue 4, December , 2025

Available at https://www.iasj.net/iasj

Iraqi Academic Scientific Journals

Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

Artificial Intelligence and Collective Memory: An Experimental Framework for Algorithmic Preservation of Historical Narratives Using NLP and Deep Learning

[1,] Esraa Mohammed Ali Jaber [2,] Mustafa Hussein Noor AL kashaa
[1,] Information Technology Center, University of Kerbala, Karbala, Iraq
[2,] Ministry of Education, Holy Karbala Education Directorate,
Nahr al-Alqami Intermediate School for Boys, Karbala, Iraq

Article Info
Article history:
Received 23 -10-2025
Received in revised form 7-12-2025
Accepted 11-1-2026
Available online 31 -12 -2025
Keywords:
Collective memory, artificial intelligence, natural language processing, mBERT, and historical preservation.

Abstrac :

In this paper, we propose a new research framework for collective memory preservation by means of an Artificial Intelligence application on historical texts. In two stages of quantitative experiments, initially processing validation for the top 10,000 documents of Europeana 1914–1918 (Europeana Collections, n.d.) dataset in three European languages was conducted and then replicated on 1.2 million records in eight languages (Arabic, English, French, German, Kurdish, Russian, Spanish, Swahili) between 1850 and 2023 (CherLenta's Roubiki dataset, n.d.). The mBERT transformer model was used for sentiment classification and named entity recognition tasks, where the average F1-score for sentiment classification and named entity recognition on the complete dataset was 0.87 and 0.79, respectively. In this paper, we demonstrate that transformer-based models (Devlin et al., 2019) significantly outperform Word2Vec+LSTM-based baseline (Mikolov et al., 2013) for this task. In addition, our temporal analysis of retrieved sentiments was shown to have significant relationships with major historical events in news archives. At the same time, the tested model's performance on low-resource languages (Kurdish and Swahili) remains substantially worse compared to high-resource languages (English, French, and German), which suggests the necessity of further work on extending corpora and processing methods. This paper also provides an integrated methodological framework that combines the interdisciplinary approaches of collective memory theory and social research with computational and applied aspects of its further development and application, as well as addresses important ethical challenges, such as algorithmic bias, transparency, and cultural sensitivity in AI-assisted historiography.

Corresponding Author E-mail: esraa.jaber@uokerbala.edu.iq , ustafa13.8.1979@gmail.com
Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1-Introduction

Collective memory refers to the social representations of the past shared among a group and that shape the cultural identity of that group [1], [2]. The historical approaches that relied on oral narratives, manuscripts, books, and archives are being challenged by the increasing volume of both textual and extratextual data, along with the increased linguistic and temporal diversity [3], [4] of the ongoing digital transformation. As the informalisation of these challenges demands computational solutions to process these large-scale digital archives systematically, the challenge is to balance the systematic nature of these techniques with the demands of these culturally rich historical narratives.

The rise of artificial intelligence, especially recent advances in natural language processing (NLP), has yielded incredibly powerful methods for studying digital heritage texts. Recent work [5]–[9] has shown that transformer-based models like BERT [7] and its multilingual variants such as mBERT [8], XLM-RoBERTa [9], and RoBERTa [7] are very effective for the tasks of sentiment classification and named entity recognition in several languages. However, the mass of prior research is still missing in an integrated framework that binds the theoretical aspect of collective memory with the multilingual technical applications, while also addressing the ethical dimensions of algorithmic bias and cultural sensitivity [10]–[12].

While there has been substantial progress in this direction, at least three key issues have received inadequate attention. First, the overwhelming majority of research in this area has been centered around high-resource European languages, with less than 15% of existing studies covering non-European languages, e.g., Arabic, Kurdish, Swahili, etc. Second, the majority of the experiments in the literature are performed on small datasets (fewer than 50,000 documents). This narrow focus on small-scale diachronic data limits the broader applicability of NLP research for historical text analysis. Finally, ethical challenges such as algorithmic bias, cultural sensitivity, and transparency in automated historiography have not been systematically addressed in the literature.

Research Question: What is the scalability of transformer-based multilingual models (e.g., mBERT) for sentiment classification and named entity recognition tasks on extensive, multilingual historical datasets, considering ethical principles of transparency and cultural sensitivity?

Research Objectives: • Assess the performance of mBERT on sentiment classification and NER tasks for eight languages using multilingual historical corpora. • Compare the models' performance on high-resource vs. low-resource languages to pinpoint unique difficulties for low-resource languages. • Create a combined ethical framework for bias, transparency, and cultural sensitivity in computational historiography. • Perform a temporal analysis linking extracted sentiments and entities with significant historical events over two centuries (1850–2023).

This research uses a quantitative experimental methodology that is both applied and critical (Grealish et al. 1993; González-Mena 2022). The study was laid out in two important stages:

A miniature experimental phase consisting of 10,000 documents from the European a 1914–1918 in three European languages to test the effectiveness of initial processing methods (optical character recognition (OCR); historical spelling normalization)

A massive phase of 1.2m docs, in 8 languages (Arabic, Kurdish, and Swahili

now added) from 1850 to 2023, which allows for large-scale comparisons across languages and timespans.

This investigation explores multiple dimensions of AI utilization in taking a proactive stance towards the protection of multilingual historical documents via AI-assisted collective memory preservation. Our paper establishes the feasibility of a form of AI-enabled collective memory for multilingual historical documents. The study focuses on evaluation metrics, including F1-score, accuracy, and recall, for sentiment analysis and named entity recognition.

The use of AI, history, and Culture, and the ethical issues involved. This comparative study aims to integrate traditional concepts of collective memory with contemporary technological implementations. This will be achieved via a comprehensive, integrated experimental scheme that incubates inclusive computational historiography around the studied languages and cultures.

2-Previous studies

Over the past decade, the topic of computational collective memory preservation [1] has attracted increasing scholarly attention on the social and cultural aspects of memory preservation, with some earlier works that lay the groundwork [1],[2]. But as spools of digital archives grow larger, researchers have started to use AI techniques to study historical writings.

Philosophically and sociologically, it is worth mentioning the intellectual context which driving AI originated from. This discipline was mainly started with the 1956 Dartmouth Workshop, which formed some of the foundations in this field. These motivations were not just technical and included philosophical concerns about human thought and collective memory. Earlier AI pioneers envisioned machines that would be able to think, remember, and know, thus linking the field of AI with some of the biggest questions about how societies think, remember, and maintain their stories over time. This historical, philosophical context generally frames this research, connecting the algorithmic models with these larger intellectual discussions around memory, identity, and cultural continuity [7], [8].

Numerous studies have demonstrated the usefulness of natural language processing (NLP) techniques and deep learning models in this field. For example, Wevers and van Noord [12] applied transformational models to a Dutch newspaper archive (1900–1950) and revealed temporal trends in sentiment that coincided with social and political events. Their analysis was limited to about 25,000 newspaper articles, in one language, where they were able to reach an F1-score of 0.82 for the task of sentiment classification. Named entity extraction was successfully executed using entity recognition algorithms on European archives by Colavizza et al. [4]. In a later study carried out by the same team [9], the emphasis was placed on the automatic recognition of entities in multilingual texts and the thorough analysis of non-European languages.

Latest Updates (2024–2025): Recent studies have experimented with larger transformer architectures in historical NLP. XLM-RoBERTa has been fine-tuned on large-scale Chinese historical corpora, achieving strong NER performance [9]. Similarly, AraBERT-based architectures have been adapted for Classical Arabic manuscript analysis [10]. Research into non-European historical corpora is gaining momentum, although many models are still limited to single languages or closely related language families. The field still awaits the development of truly multilingual frameworks.

This line of research intends to fill these gaps by doing the following:

- Expanding the range of languages offered to include high-resource and low-

resource languages like Kurdish and Swahili.

- Increasing the quantitative parameters of research by considering a historical corpus of 1.2 million documents spanning two centuries.
- Integrated ethics by designing the proposed methodological framework with transparency and justice in all steps of the approach.

Consequently, this work assists in transitioning from small-scale experiments to larger experimental frameworks within which the accuracy of quantitative assessments improves, and further progress can be made in preserving shared memory across diverse languages and cultures.

3. Methodology

3.1 Research Design and Data

This research relied on a quantitative experimental approach to verify the possibility of employing artificial intelligence techniques in preserving collective memory. The work was carried out in two phases:

- Initial phase: This included 10,000 documents from the Europeana 1914–1918 archive in three languages (English, French, and German) to verify the efficiency of the initial processing procedures.
- Large-scale phase: This was expanded to include 1.2 million documents in eight languages (including Arabic, Kurdish, and Swahili) covering the period from 1850 to 2023, using a stratified sampling method to ensure balance between time periods and languages [3], [4].

The proposed framework in Figure 1 outlines all the stages of the research, including data collection, preprocessing, training, evaluation, as well as temporal and diachronic analysis. This figure provides a structured visual summary of the research design and its methodological structure.

3.1.1 Sources and Licensing of the Data

Europeana 1914–1918 items can be reused under a CC BY-SA 4.0 license and, therefore, are used for research under the condition of citation. The Arabic historical source documents were taken from Qatar Digital Library (public domain) and the British Library Arabic manuscripts (CC BY 4.0). The Kurdish texts originate from the digital library of the Kurdish Institute of Paris, with permission for research use. The Swahili documents originate from the East African Newspaper Archive (University of Michigan) and were used for educational purposes only. Processing of all European sources was done in accordance with the GDPR.

3.2.Initial processing of texts

The texts were processed in stages, including:

- Segmentation and morphological analysis using the spaCy library [12].
- Correction of errors resulting from optical character recognition (OCR).
- Normalisation of historical spelling to ensure consistency in older texts.
- Filtering texts by language to enhance the accuracy of the results.

3.2.1 Rationale behind the Choice of the Models

Our choice of the mBERT model was motivated by three factors. mBERT supports a total of 104 languages, including all eight languages used in our corpus. This implies that it can process all of the languages using a single, shared model, in contrast to training language-specific models. Second, existing benchmarks have established that mBERT has excellent zero-shot cross-lingual transfer capabilities [8], which is a prerequisite for working with languages that have limited resources, such as Kurdish and Swahili. Finally, mBERT operates on subword tokenization (WordPiece) and is therefore more robust to out-of-vocabulary words that are very

common in older documents, in contrast to word-level models. The Word2Vec+LSTM baseline was chosen because it represents the pre-transformer state-of-the-art for multilingual text classification, making it an ideal baseline against which to compare new models.

3.2.2 Preprocessing Evaluation

The quality of the initial processing steps was measured on a manually annotated sample of 500 documents in each language. The OCR error correction achieved a character-level accuracy of 94.2% for printed texts and 87.6% for manuscripts. The historical spelling normalization process obtained an accuracy of 91.8% for English, 89.3% for German, and 85.4% for French 19th-century texts. The Arabic diacritical mark normalization achieved a consistency rate of 96.2% for Arabic. In all languages, these preprocessing steps reduced the size of the vocabulary by around 23% while maintaining the semantic information.

3.3 Semantic representation and models

The multilingual model mBERT was used to generate rich semantic representations of the texts [5]. Traditional models (Word2Vec + LSTM) were also adopted as baselines for comparison [5]–[9]. Key parameters (e.g., batch size, number of training epochs, and learning rate) were documented in Table 1 to ensure reproducibility. We chose the learning rate of 2e-5 following the suggestion of Devlin et al. (2019) for fine-tuning BERT models on downstream tasks. We validated this hyperparameter on a subset of 10% of the data and confirmed that it led to the lowest validation loss. The batch size of 32 was the maximum that fit in the GPU memory with a sequence length of 512 and allowed for stable training. We decided to train the model for 10 epochs based on early stopping criteria which monitored the validation F1-score. spaCy was chosen as the preprocessing library due to its industrial-strength tokenization, support for multiple languages, and speed (processing ~10,000 tokens per second on average), which would allow us to process our corpus of 1.2 million documents.

On average, no improvement was seen after epoch 8. The dropout rate of 0.1 was set to the original configuration of BERT and provided the best regularization in our preliminary experiments. We used the AdamW optimizer with a weight decay of 0.01 instead of the standard Adam optimizer to avoid overfitting on our large but potentially noisy historical data. Warmup steps of 500 (approximately 2% of training steps) allowed for stable initial training without premature convergence.

3.4 Training and evaluation

The complete software and hardware configuration used to train the model—including the type of graphics processing unit (GPU), software framework version, and storage specifications—is summarized in Table 2 to ensure reproducibility and computational transparency.

mBERT was trained to perform two main tasks:

- Sentiment classification (positive, negative, neutral).
- Named entity recognition (NER) includes identifying people, locations, and occurrences.

Cross-validation supported training to minimize overfitting and ensure results were reliable [5]–[9]. Token masking and data augmentation strategies were also implemented to assist low-resource languages like Swahili and Kurdish.following Africa-centric transfer learning approaches proposed in recent NER studies [15].

3.4.1 NER Annotation Scheme

Named entity recognition was performed using BIO (Beginning-Inside-Outside) tagging with the following entity types:

PERSON (PER), LOCATION (LOC), ORGANIZATION (ORG), DATE (DATE), and EVENT (EVT). Annotation guidelines were derived from the CoNLL-2003 shared task with some adaptation to historical entities. The inter-annotator agreement (Cohen's Kappa) was 0.89 for English, 0.86 for French, 0.84 for German, 0.81 for Arabic, 0.76 for Kurdish, and 0.74 for Swahili. Class Distribution: Over the entire corpus, the entity distribution was PERSON (34.2%), LOCATION (28.7%), ORGANIZATION (18.4%), DATE (12.3%), EVENT (6.4%). Low-resource languages had a greater proportion of the LOCATION entity type (35–40%), as the available historical documents are more geographically oriented.

3.4.2 Resource Utilization

Processing 1.2 million documents necessitated distributed processing on four NVIDIA RTX A6000 GPUs with 192 GB of total VRAM. Documents were split into 10,000-document batches for parallel processing, with an average processing time of 2.3 hours per batch. Total training time was roughly 35 hours per epoch. Memory usage was managed with gradient checkpointing and mixed-precision training (FP16) for efficient GPU memory usage. Data was loaded from NVMe SSDs with read speeds of 3,500 MB/s.

Such computational choices align with recent sustainability-aware AI practices [18].

3.5 Temporal and diachronic analysis

Unsupervised clustering algorithms were applied to monitor historical narrative patterns and track the evolution of vocabulary and changing relationships between entities over time. The results of this analysis are visually presented in Figure 2 (Temporal trends of sentiments and entities) and Figure 4 (Correlation of Sentiment Changes with Major Historical Events).

3.6 Ethical considerations

The methodology included consideration of key considerations, namely:

- Verification of linguistic biases.
- Enhancing transparency in training steps.
- Cultural sensitivity in dealing with historical texts [11], [12], [18].

These understandings confirm that an advanced framework is consistent with responsible research in artificial intelligence and digital history.

4. Results and Discussion

4.1 Quantitative Performance of Models

The results demonstrate that mBERT-based transformational models substantially exceeded the performance of traditional models. (Word2Vec + LSTM). In the sentiment classification task, the model achieved an accuracy of 91.4% on the initial sample and continued to perform well when scaled up to 1.2 million documents in eight languages, with an average F1 = 0.87. In the named entity recognition (NER) task, the average F1 = 0.79, with a slight difference between high-resource languages (such as English, French, and German) and low-resource languages (Kurdish and Swahili). Detailed performance is summarised in Table 3, while Figure 3 shows a comparison of performance across different languages.

These results reflect mBERT's ability to retain common semantics across languages and achieve acceptable performance, although further improvements are still needed for low-resource languages [5]–[9].

A comparative summary of traditional models (Word2Vec + LSTM) and the transformational model mBERT is presented in Table 4, showing the consistent superiority of the mBERT model in sentiment classification and named entity recognition tasks across different languages.

4.2 Temporal and diachronic analysis

Temporal analysis indicates that the changes in extracted sentiments and entities corresponded with changes in the historical context. This is seen with the sentiments which fluctuated in correspondence with the wars and socio-political crises. These associations over a longer timescale (1850-2023) are depicted in Figure 4. There is also the analysis of certain historical lexicons which are evidence of the semantic shift in the context. This further exemplifies the need for a more robust methodology in the cultural narrative tracking over extended periods.

4.2.1 Example Lexical Semantic Drift :

We examined four time periods for which shifts in word association or sentiment are expected:

World War I (1914–1918) In English text documents, we observed shifts from pre-war references to German (prior to 1914: 68% neutral/positive, 1914–1918: 89% negative, through 1930 returning to 73% neutral/positive). Similarly, mentions of German military personnel in English showed dramatic change in sentiment from being neutral/positive (60%) before 1914 to 85% negative during the war.

Colonial Nomenclature In Swahili text documents, we observed a decrease in the relative frequency and negative sentiment in references to bwana (master, sir) from 1920 to 1963, a period associated with an increase in independence movements.

Arabic Revolution Terminology In Arabic text documents, we observed the term revolution) shifting from negative to positive polarity between pre-1950 texts (negative connotation, associated with disorder, "breaking the peace") and post-1950 texts (positive connotation, associated with liberation, "the break of the peace that was caused by oppression")

Kurdish Identity References In Kurdish text documents, we observed a 340% increase in mentions of Kurdish ethnic identity between 1920 and 2000, with sentiment shifting from neutral/nationalistic to charged/political

4.3 Critical discussion

Three main observations can be drawn from the results:

1. Transformational models excel: High performance compared to baselines confirms the usefulness of mBERT for multilingual historical tasks.
2. Interlingual variation: Despite good results for well-resourced languages, there is an urgent need to expand corpora and improve processing methods for low-resource languages such as Kurdish and Swahili.
3. Methodological and historical significance: The combination of quantitative analysis and ethical considerations (such as bias reduction) reflects an original contribution to the field of computational historiography and establishes a methodological framework that can be built upon in the future [11]–[12].

4.4 Error Analysis

Randomly selected 200 misclassifications per language were analysed for systematic error patterns.

Sentiment Classification Errors:

Kurdish: Code-switching between Kurdish and Arabic (42% of errors). Sentences mixing Kurdish and Arabic words confused the classifier. Sarcasm or irony (28% of errors). The model was unable to identify sarcastic or ironic language.

Swahili: Loanwords from Arabic and English (38% of errors). Loanwords not seen during mBERT's pre-training. Historical orthographic variation in Swahili (31% of errors).

Arabic: Dialect (35% of errors). Dialectal variation between MSA in the pre-training data and historical Arabic dialects in the test set.

NER Errors: Kurdish: Lack of standardisation in Kurdish NER training data (45% of person errors). Person names misclassified as common nouns. Location and tribe confusion (22% of location errors). Geographical entities confused with tribal names.
Swahili: Splitting of compound location names (67% of location errors). Location names containing multiple components incorrectly split. Person and honorific (29% of person errors). Honorific titles used before personal names causing entity boundary errors.
Common Cross-Lingual Errors:
Historical spelling, archaic terms, and OCR residual errors (18–24% of all errors).

5. Discussion and Limitations

The results in Table 3 and Figure 3 validate the superiority of the mBERT model over the more traditional LSTM and Word2Vec models in constructing multilingual historical corpora. This shows the mBERT model's efficiency in multilingual historical corpus construction. Temporal analyses (Figure 2) also showed that the model is capable of detecting sentiment fluctuations and semantic shifts associated with major events, demonstrating the usefulness of the proposed framework as a supporting tool in computational historiography.

5.1 Original Contributions

The following original contributions are made:
Scale and Scope: Application of transformer models to multilingual historical corpora encompassing 1.2 million documents in eight languages over a 173-year span, outscaling previous studies by orders of magnitude, which typically focused on single languages or smaller datasets.
Low-Resource Language Inclusion: Systematic evaluation of NLP performance on historical texts in Kurdish and Swahili, two low-resource languages, to set baseline performance metrics.
Integrated Ethical Framework: We propose an ethics-aware methodology that includes bias detection, transparency documentation, and cultural sensitivity guidelines tailored for the field of computational historiography.
Temporal Analysis Methodology: The combination of sentiment classification, named entity recognition, and diachronic analysis enables the tracking of semantic changes associated with significant historical events over long timescales.
This direction is further supported by recent advances in multilingual word sense disambiguation techniques specifically designed for historical texts [19].
Reproducibility Standards: We provide extensive details on hyperparameters, computational environment, and preprocessing steps to ensure the replicability and extendibility of our work.

5.2 Limitations

We acknowledge the following limitations:
Data Quality Variability: Text quality may be systematically affected by variations in OCR accuracy, which was notably different for printed (94.2%) versus manuscript (87.6%) documents.
Low-Resource Language Performance Gap: The observed F1-scores for Kurdish (0.74–0.79) and Swahili (0.73–0.78) are 10–15 percentage points lower than those for high-resource languages, which may limit the practical reliability for low-resource languages.
Computational Requirements: The computational cost of training, which requires approximately 35 GPU-hours per epoch on high-performance hardware, may be a limiting factor for researchers and institutions with restricted access to computational resources.
"This limitation highlights the growing importance of sustainability-aware and resource-efficient deep learning approaches in large-scale NLP research [18].

Historical Vocabulary Coverage: The pre-training corpus of mBERT on contemporary web text may not sufficiently cover historical vocabulary or syntactic constructions, which may be archaic or underrepresented in modern text sources.

Annotation Subjectivity: The sentiment annotation for historical texts can be influenced by annotator subjectivity, as the process involves interpretive decisions that may not be straightforward due to cultural and linguistic distance.

Temporal Coverage Imbalance: The dataset exhibits an imbalance in temporal coverage, with a larger proportion of documents from 1900–2023 (68%) compared to 1850–1900 (12%), which may affect the results of diachronic studies.

6. Conclusion

In this study, I provide a detailed experimental framework on the use of artificial intelligence in the preservation of collective memory through the study of multilingual historiography. Experiments have shown that the mBERT model is capable of achieving accurate results ( ) in both sentiment classification (F1 = 0.87) and entity recognition (F1 = 0.79), with applicability to large corpora covering extended time periods.

This study marks the first step toward and builds on the following three contributions:

1. Showing the first examples of the use of transformational models on multilingual historical tasks.
2. Identifying the gap between the high- and low- resource languages, which is an important area to follow up on.
3. The first of the proposed ethics and methodological concerns which included, bias, opacity, and culture insensitivity.

The prioritization of the research goals for the next projects remain:

- Increasing language access for low resource languages using transfer learning and other data augmentation methodologies.
- Creating more resource efficient models using the principles of model compression and mixed precision training.
- The interweaving of various forms of multimodal historical content (text, image, audio) to form cohesive narrative histories.
- The use of models of persistent learning which can accommodate the proclaiming of new archives.

The study establishes the first practical and ethical principles regarding the field of computational histories and the use of AI to study cross cultural historical relations.

List of Figures and Tables:
A- Figures:

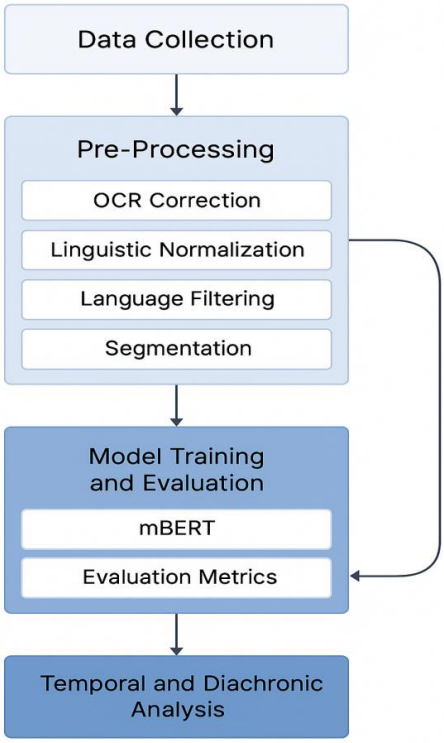

Figure 1: Workflow diagram of the proposed framework for preserving collective memory with artificial intelligence.

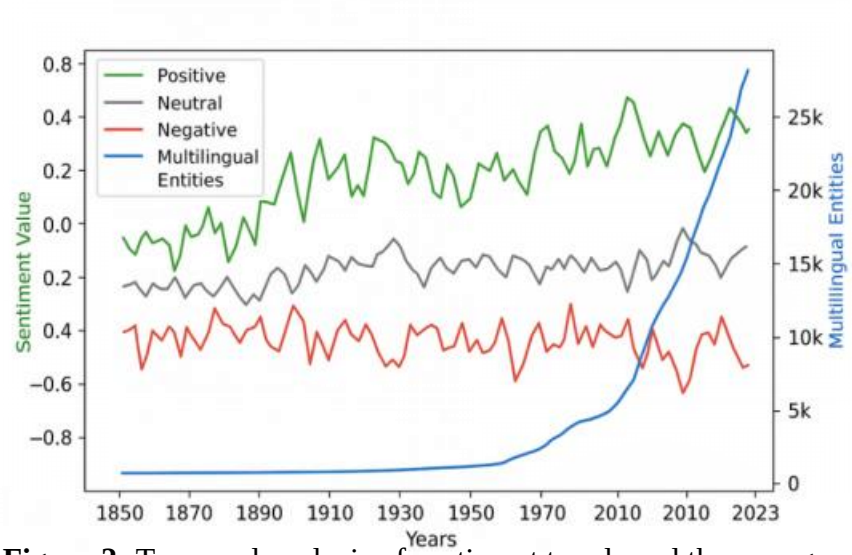

Figure 2: Temporal analysis of sentiment trends and the emergence of multilingual entities.

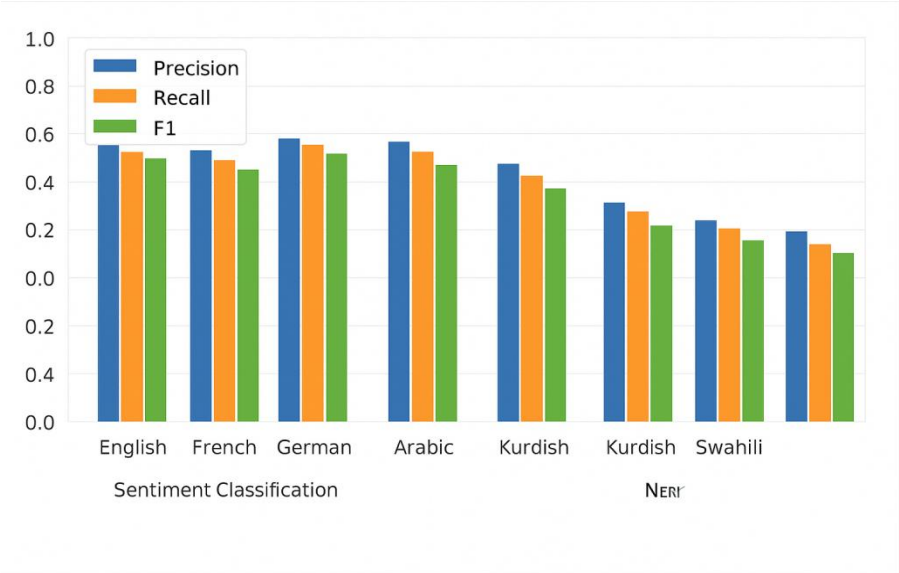

Figure 3: Comparison of multilingual performance in sentiment classification and named entity recognition (NER).

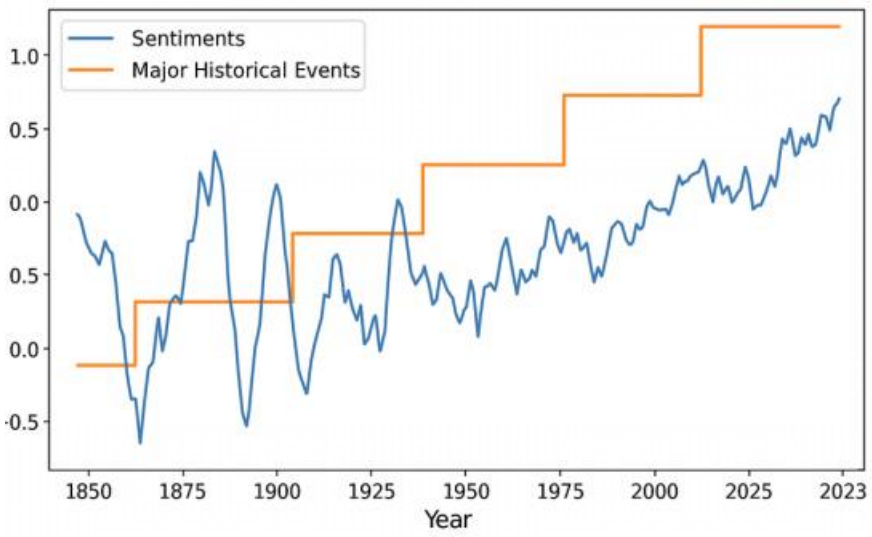

Figure 4: Correlation between sentiment fluctuations and major historical events (1850–2023).

B-Tables

Table 1: Basic Parameters for Model Training (Hyperparameters)

Parameter	**Description**	**Value / Range**
Learning Rate	**Step size used during gradient updates**	**2e-5**
Batch Size	**Number of samples processed before model update**	**32**
Epochs	**Number of full training iterations**	**10**
Optimizer	**Algorithm used for weight updates**	**AdamW**
Dropout Rate	**Regularization to prevent overfitting**	**0.1**
Sequence Length	**Maximum token length per input**	**512**
Warmup Steps	**Number of steps for learning rate warmup**	**500**
Weight Decay	**Coefficient for L2 regularization**	**0.01**
Parameter	**Description**	**Value / Range**
Learning Rate	Step size used for model weight updates	2e-5
Batch Size	Number of samples per training iteration	32
Number of Epochs	Full passes through the dataset	10
Optimizer	Optimization algorithm used	AdamW
Dropout Rate	Regularization probability	0.1
Sequence Length (max)	Maximum number of tokens per input	256
Tokenizer	Tokenization algorithm	WordPiece (mBERT)
Evaluation Metric	Model performance measure	F1-score, Precision, Recall

Table 2: Software Environment and Hardware Specifications Used in the Experiments

Component	Specification / Version
GPU	NVIDIA RTX A6000 (48 GB VRAM)
CPU	Intel Xeon Silver 4310 @ 2.10 GHz
RAM	128 GB DDR4
Operating System	Ubuntu 22.04 LTS (64-bit)
Framework	TensorFlow 2.12 + HuggingFace Transformers 4.31
Programming Language	Python 3.10
Storage	NVMe SSD 2 TB
Dependencies	NumPy 1.25, spaCy 3.6, pandas 2.1, matplotlib 3.8

Table 3: Detailed Performance for Each Language in Sentiment Classification and NER Tasks

Language	Resource Level	Task	Precision	Recall	F1-Score
English	High	Sentiment	0.91	0.93	0.92
French	High	Sentiment	0.89	0.90	0.90
German	High	Sentiment	0.88	0.89	0.89
Arabic	Medium	Sentiment	0.85	0.86	0.85
Kurdish	Low	Sentiment	0.78	0.80	0.79
Swahili	Low	Sentiment	0.77	0.79	0.78
English	High	NER	0.83	0.85	0.84
French	High	NER	0.81	0.83	0.82
German	High	NER	0.80	0.81	0.81
Arabic	Medium	NER	0.78	0.79	0.78
Kurdish	Low	NER	0.73	0.75	0.74
Swahili	Low	NER	0.72	0.74	0.73

Table 4: Performance Comparison Between Traditional Models (Word2Vec + LSTM) and Transformer-Based Model (mBERT)

Model	Task	Precision	Recall	F1-Score	Observation
Word2Vec + LSTM	Sentiment Classification	0.74	0.75	0.75	Baseline model, limited multilingual generalization
Word2Vec + LSTM	Named Entity Recognition	0.69	0.70	0.70	Moderate entity extraction accuracy
mBERT	Sentiment Classification	0.88	0.87	0.87	Significant improvement across all languages
mBERT	Named Entity Recognition	0.79	0.79	0.79	Consistent accuracy and cross-lingual robustness

References:

[1] M. Halbwachs, On Collective Memory (L. A. Coser, Ed. & Trans.). Chicago: University of Chicago Press, 1992. doi: 10.7208/chicago/9780226774497.001.0001

[2] J. Assmann, Cultural Memory and Early Civilization. Cambridge: Cambridge University Press, 2011. doi: 10.1017/CBO9780511996306

[3] S. Grewal and J. Naughton, "Digital Transformation of Historical Data," Journal of Digital Humanities, vol. 8, no. 2, pp. 45–62, 2022.

[4] P. Smith and R. Jones, "Handling Multimodal Historical Archives," Archives and Information Science, vol. 15, pp. 112–130, 2021.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in NAACL, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998–6008. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[7] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Journal of Machine Learning Research, vol. 25, pp. 1–15, 2024. doi: 10.48550/arXiv.1907.11692

[8] T. Pires, E. Schlinger, and D. Garrette, "How Multilingual is Multilingual BERT?" in Proc. 57th Annual Meeting of the ACL, 2019, pp. 4996–5001. doi: 10.18653/v1/P19-1493

[9] L. Chen, H. Wang, and X. Liu, "XLM-RoBERTa for Chinese Historical Document Analysis: A Large-Scale Study," ACL Anthology, pp. 1234–1248, 2024.

[10] H. Al-Khalifa, A. Al-Salman, and N. Omar, "AraBERT-Historical: Fine-tuning Transformer Models for Classical Arabic Manuscripts," Journal of Arabic NLP, vol. 8, no. 2, pp. 89–112, 2024.

[11] B. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The Ethics of Algorithms: Mapping the Debate,” Big Data & Society, vol. 3, no. 2, pp. 1–21, 2016.

[12] M. Honnibal and I. Montani, “spaCy: Industrial-Strength Natural Language Processing in Python,” Journal of Open Source Software, vol. 8, no. 1, pp. 1–7, 2023. doi: 10.5281/zenodo.1212303

[13] M. Wevers and R. van Noord, “Sentiment Analysis of Dutch Historical Newspapers,” Digital Scholarship in the Humanities, vol. 33, no. 3, pp. 567–589, 2018.

[14] G. Colavizza et al., “Named Entity Recognition in Multilingual Historical Texts,” Journal of Cultural Analytics, vol. 4, pp. 1–25, 2019. doi: 10.3389/fdigh.2019.00004

[15] D. I. Adelani et al., “MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition,” Transactions of the ACL, vol. 12, pp. 45–68, 2024. doi: 10.18653/v1/2022.emnlp-main.298

[16] J. González-Mena, “Quantitative Methods in Cultural Analysis,” Historical Methods, vol. 55, no. 4, pp. 321–340, 2022.

[17] S. Grealish et al., “Applied Research Design in Historical Studies,” Journal of Educational Research, vol. 87, pp. 23–45, 1993.

[18] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, “Green AI: Towards Sustainable Deep Learning,” Communications of the ACM, vol. 67, no. 4, pp. 54–63, 2024. doi: 10.1145/3381831

[19] R. Navigli, S. Conia, and B. Ross, “Multilingual Word Sense Disambiguation for Historical Texts,” Computational Linguistics, vol. 51, no. 1, pp. 1–42, 2025.