

Detecting Inconsistencies in Narrative Elements of Cross Lingual Nakba Texts

Nada Hamarsheh^{1,a} Zahia Elabour^{1,b} Aya Murra^{1,c} Adnan Yahya^{1,d}

¹ Electrical and Computer Engineering Department, Birzeit University, Birzeit, Palestine

a: hnada.edu@gmail.com
b: zahia-ali@hotmail.com
c: ayaibrahim065@gmail.com
d: yahya@birzeit.edu

Abstract

This paper proposes a methodology for contradiction detection in cross lingual texts about the Nakba. We outline a pipeline that includes text translation using Google’s Gemini for context-aware translations, followed by a fact extraction task using either Gemini or the TextRank algorithm. We then apply Natural Language Inference (NLI) by using models trained for this task, such as XLM-RoBERTa and BART to detect contradictions from different texts about the Nakba. We also describe how the performance of such NLI models is affected by the complexity of some sentences as well as the unique syntactic and semantic characteristics of the Arabic language. Additionally, we suggest a method using cosine similarity of vector embeddings of facts for identifying missing or underrepresented topics in historical narrative texts. This work is a proof-of-concept, and the results are preliminary. However, they offer initial insights into biases, contradictions, and gaps in narratives surrounding the Nakba, providing a foundation for future research into contradictions in historical perspectives.

1 Introduction

Nakba (Arabic for catastrophe) refers to the displacement of hundreds of thousands of Palestinians from Palestine during the 1948 war between Arabs and Israel (United Nations, 2024). As a result of this catastrophe, approximately 750,000 Palestinians were forcefully displaced, 15,000 killed, and more than 531 towns and villages destroyed (Palestinian Central Bureau of Statistics (PCBS), 2008). With this catastrophe happening 76 years ago, the narratives that we have now surrounding it vary, with belief, denial, or skepticism.

Conflicting narratives regarding the history of the Nakba preceded it, and eventually led to it. A clear example that displays how a conflict in narratives contributed to the events leading up to Nakba

is the famous phrase “A land without a people for a people without a land”. This phrase, believed by many historical references to be a “Zionist slogan” (Muir, 2008), was in stark contrast to the 690,000 people who lived in Palestine in 1914 (Palestinian Central Bureau of Statistics (PCBS), 2021) before the Balfour Declaration and the waves of immigration to Palestine that followed.

By considering such examples of how a narrative can influence and be influenced by remarkable events of history, we establish the need for a systematic approach to review such narratives, and point to pieces of historical evidence that have been tampered with, dropped or manipulated.

In this paper, we propose a method that incorporates the use of Natural Language Processing (NLP) to review pieces of written text, compare texts from contrasting backgrounds, and lastly presents sentences where this contrast in facts happen, which in return indicates contradiction of information between the subject sources. The system also presents what statements have been left out in one source but mentioned in another. This highlighting of the contradictions found in the subject text, as well as some text failing to mention certain facts, can lead to better detect biases. By flagging such contradictions, historians and experts on the Nakba can make informed remedial decisions based on the approach outlined in this paper.

The system we propose begins by taking two input texts, the focus languages here are English and Arabic, and the texts are expected to reflect differing historical narratives about the Nakba. Since Arabic-speaking historians are the target, texts in English are first translated to Arabic. In a previous work of ours (Murra et al., 2024), we reached the conclusion that using Gemini for translation performed better than using machine translation models such as MarianMT. Therefore, we propose using Gemini to perform this step.

Then we extract facts that represent the ideas of

the texts. For this we suggest two methods of providing summaries, once by prompting Gemini, and the other by using the TextRank algorithm. The facts extracted by this step are assumed to reflect the contradictions and overlooked facts between the texts.

Lastly, we perform contradiction detection by utilizing the labels provided by NLI models such as XLM-RoBERTa (xlm-roberta-large-xnli) and BART (bart-large-mnli). This gives a score for the relationship between a pair of sentences, to indicate either an entailment, neutral, or contradiction relationship.

For finding gaps in fact representation over the texts, we propose using embedding techniques on the facts extracted from the texts, then use similarity metrics, such as cosine similarity, to find unique sentences that are missing in the corresponding texts.

2 Background

In this paper, we suggest the use of automatic summarization of the subject texts. Automatic text summarization extracts the main ideas of a text document (Mallick et al., 2019). Summarization can happen in two types, extractive and abstractive summaries. An extractive summary produces a summary based on the sentences used in the text itself rather than producing a unique summary. It looks for identical information in documents and assigns a score to each statement based on how well it explains the other facts of the document. An abstractive summary, in contrast, generates a unique summary by rephrasing and restructuring the most important information from the original text (N. et al., 2022). However, summarization is a complex NLP problem influenced by text type, length, vocabulary level, and named entities. However, due to its abstract nature, it remains a long-standing issue (Bongale et al., 2022).

Google’s Gemini is designed to improve automated text summarization by mimicking human-like sentence-level styles. It combines extractive and abstractive techniques, allowing for fine control over summary style and quality, which reduces risks like factual inaccuracies in summaries. Gemini introduces a “Fusion Index” to analyze and adjust sentence styles within new datasets, enhancing flexibility for different applications (Bao et al., 2023). While effective, this approach still lacks a clear way to measure its ability to generate fully

abstracted summaries, marking an area for further exploration (Bao et al., 2023).

TextRank is an unsupervised extractive text summarization algorithm that ranks sentences by extracting the main ideas from a document depending on their importance. It is related to Google’s PageRank, which ranks web pages for online search results (Bongale et al., 2022).

With that said, it should be noted that research points to limitations in Arabic summarization techniques. This is linked to many reasons including the scarcity of annotated datasets in Arabic, the complexity of Arabic linguistics, morphology, and syntax. All these reasons lead to difficulties in obtaining a coherent summary that does not change the intended tone and meaning of the original text (Souri et al., 2023)

A notable technique in relationship inference from texts is the Natural Language Inference (NLI) task. NLI classifies the relationship between a pair of premise and hypothesis as either entailment, contradiction, or neutral. It forms the basis for higher-level NLP tasks like question-answering and summarization (Nie et al., 2019). In Arabic however, NLI is a challenging task because of the lexical ambiguity of the language, lack of large entailment datasets, etc. (Jallad and Ghneim, 2022)

3 System Implementation

3.1 Data Collection

To ensure the capability of achieving the aim of this paper, the text data collected employs samples of different and oftentimes contradicting perspectives on topics related to the Nakba, sourcing Arabic and/or English texts on the issue.

The texts used originate from different sources, such as articles from Wikipedia, academic journals by Arabs, Israeli, or others writing on the Nakba and the events of 1948, official reports of historians, governmental figures, or organization like the United Nations (UN), and lastly some articles from websites discussing the topic.

3.2 Translation

As our focus is mainly texts in Arabic or English, a translation task is crucial. We are particularly interested in presenting findings to Arabic readers, including historians and field experts. Therefore, English texts are translated to Arabic to ensure the results are relevant to the specific audience.

The process begins by detecting the language of

the text, ensuring that only content in English is processed for translation. Once identified, the English text is split into smaller chunks for translation using the Gemini-pro model.

The model is prompted to provide a translation in Arabic that preserves the tone and meaning of the original article. The model then translates each chunk from English to Arabic, removing unnecessary elements. This translation will be the source of the information and factual statements that will be processed for any contradictions.

To illustrate the process further, we referenced two articles to explain the proof-of-concept: "The Nakba: Something That Did Not Occur (Although It Had to Occur)" (Bronstein, 2009), and "The Nakba: More than just a historical event" (Original: النكبة أكبر من مجرد حدث تاريخي (Naim, 2023)). The first article is originally in English, and as mentioned all English texts must be translated to Arabic. Since the other article is in Arabic, no work is needed to be done there at this point. The following example is a text excerpt from (Bronstein, 2009), followed by the translation Gemini provided.

"From early on, Zionism ignored the existence of the Arab inhabitants of Palestine. It is, therefore, not possible that some 800,000 persons were ethnically cleansed from the country and that more than 500 Palestinian villages were destroyed".

" تجاهلت الصهيونية منذ وقت مبكر وجود السكان العرب في فلسطين. ومن ثم، فإنه ليس من الممكن أن يتم تطهير ما يقرب من 800000 شخص عرقياً من البلاد، وأن يتم تدمير أكثر من 500 قرية فلسطينية".

3.3 Summarization

The aim of using Summarization is to reduce data complexity and improve the efficiency of similarity and contradiction analysis. Summarization is used to extract the main points from the text and focus on relevant information. This process separates content into actionable, high-quality content, avoiding irrelevant details. We experimented with different summarization approaches on the same text about Nakba, and below is a discussion of the results of this experimentation.

Using TextRank:

As mentioned earlier, the TextRank algorithm is used to retrieve sentences from the original text, after it ranks the sentences with importance scores, providing an extractive summary of the text at hand. The example below shows a group of facts

(sentences) that were extracted and retrieved by the TextRank algorithm when applied to (Naim, 2023).

- "أن كل الأهوال التي مر بها شعبنا الفلسطيني في تلك الفترة كانت جزءاً من مخطط متعمد من القيادة العليا للحركة الصهيونية، - تم تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدينة،"

- "وتهجير أكثر من 750 ألف فلسطيني (نحو 50% من الفلسطينيين الذين كانوا يقيمون في فلسطين التاريخية)،" - "فقد سيطر اليهود بالقوة عام 1948 على نحو 78% من

أراضي فلسطين التاريخية، وأنشؤوا دولتهم على هذه المساحة."

Using Gemini:

The second approach we used is Google's Gemini. Gemini, unlike traditional summarization techniques, can be prompted to summarize text in a list of the main points or ideas from the text, providing an abstractive summary that uses paraphrasing of the original text. For illustration purposes, Gemini was prompted to summarize (Naim, 2023), and the example below shows a sample of the extracted facts.

- "النكبة هي كلمة عربية تعني الكارثة، وهي حدث أليم استمر نحو سنتين من منتصف عام 1947 إلى منتصف عام 1949."

- "تضمنت النكبة تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدينة وتهجير أكثر من 750 ألف فلسطيني."

- "رُوجّ الصهاينة أن فلسطين كانت صحراء خالية من السكان،

لكن هذه الرواية لا تصمد أمام حقائق التاريخ."

- "منذ النكبة، وضعت القيادة الصهيونية خطة للسيطرة على الشعب الفلسطيني وإحباط قدرته على الثورة والرفض."

Using both of these models comes with advantages and disadvantages. TextRank for example is unsupervised and can extract the exact "problem" sentences which can hold some contradiction to others. However, if the original text holds some ambiguity in its wording, the sentence will be passed as is. In contrast, Gemini reduces the complexity of some sentences by paraphrasing, which makes comparing and finding contradictions easier. This, however, can be problematic if paraphrasing changes the intended meaning.

3.4 Similarity and Contradictions

The core of our work is to determine if two texts of different sources have contradicting statements. Two sentences are said to be contradicting each other if they are generally about the same idea, with conflicting information found in both. An example on contradicting sentences can be seen in the following example:

Sentence 1 (Originally in English): "From early on, Zionism ignored the existence of the Arab inhabitants of Palestine. It is, therefore, not possible that some 800,000 persons were ethnically cleansed from the country and that more than 500 Palestinian villages were destroyed." (Bronstein, 2009)

(Translation:

تجاهلت الصهيونية منذ وقت مبكر وجود السكان العرب في فلسطين. ومن ثم، فإنه ليس من الممكن أن يتم تطهير ما يقرب من 800000 شخص عرقياً من البلاد، وأن يتم تدمير أكثر من 500 قرية فلسطينية.

Sentence 2 (Originally in Arabic):

في تلك الفترة، تم تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدينة، وتهجير أكثر من 750 ألف فلسطيني (نحو 50% من الفلسطينيين الذين كانوا يقيمون في فلسطين التاريخية). (Naim, 2023)

(Translation: "During that time, more than 530 Palestinian villages were destroyed, over 700 villages and towns were seized, and more than 750,000 Palestinians (approximately 50% of those living in historic Palestine) were displaced.")

Both sentences are on the same topic, but present contradictory statements.

To detect such contradictions, the facts extracted from both subject texts are compared using two Natural Language Inference (NLI) models, XLM-RoBERTa (xlm-roberta-large-xnli) and BART (bart-large-mnli). This is done by pairing one fact (sentence) from each list of the extracted facts at a time and identifying their relationships as Contradiction, Neutral, or Entailment. Each pair created is assigned different probabilities representing each of these labels, with the highest probability determining the final label given to the pair. If the highest probability is given to the contradiction label, then the sentences are considered to be contradicting.

For comparison purposes, this process is performed twice: once for the facts extracted from both subject texts using Gemini, and again for the facts extracted using the Textrank algorithm. As there are differences in the type of sentences retrieved by each approach, not all facts found in one approach are necessarily found in the other. Tables 1 and 2 shows a sample of pairs with the labels assigned by each NLI model and their corresponding probabilities. In both tables, the premise comes from the facts extracted from (Naim, 2023), while the hypothesis comes from the facts extracted from

(Bronstein, 2009).

From table 1, the reader can infer that the premise and the hypothesis are contradictory based on the overall meanings of the sentences of examples 1 and 2 (i.e. The Nakba led to contrasting outcomes for both Palestinians and Zionists). Both models correctly labeled the pair as a contradiction.

In the third example, the reader can also infer a contradiction (i.e. Describing the Nakba as a catastrophe in the premise, and denying the Nakba in the hypothesis), but because the wording can be complex as seen in the hypothesis, the models did not assign the same label to the pair.

Finally, the last example also shows a case of disagreement between the models. Even though the sentences are contradictory (i.e. the premise discusses the expulsion of 750,000 Palestinians, while the hypothesis doubts both the expulsion of said refugees and even their existence), XLM-RoBERTa (xlm-roberta-large-xnli) was not able to detect the contradiction, and BART-NLI (bart-large-mnli) assigned a lower contradiction score. This indicates that the models were not able to identify this contradiction correctly.

From table 2, the reader can clearly see that the premise and the hypothesis of example 1 are contradictory (i.e. Premise detailing the expulsion of 750,000 refugees, and the hypothesis denying the possibility of that happening). Both models resulted in labeling the pair as a contradiction.

Example 2 is more complicated in the sense that the contradiction is inferred from the meaning of the pair (i.e premise references torture and detainment of more than 1,000,000 Palestinians, while the hypothesis claims Zionists ignored Arabs in Palestine). This contradiction is more subtle than the first, therefore the models did not assign the same label, and the contradiction score given was low. In contrast, the last example shows an inferred contradiction (i.e premise states that the tragedies happening at the time (Nakba) was a deliberate plan by Zionist leaders, and the hypothesis clearly implies the Nakba did not happen in the past but continues to happen today (because of the discourse around the topic). The models successfully detected a contradiction between the pair.

The models however failed to detect that the pair in example 3 were actually not contradictory (i.e both the premise and the hypothesis state that refugees had been evicted, in the premise by force, and in the hypothesis as a means for the Zionist

Table 1: Samples of Comparison Results for facts Extracted Using Gemini

#	Premise	Hypothesis	XLM-RoBERTa Label	XLM-RoBERTa Probability	BART Label	BART Label Probability
1	تضمنت النكبة تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدنية وتهجير أكثر من 750 ألف فلسطيني.	النكبة هي حدث ضروري، لأنها حققت ذاتية صهيونية نقية مغلقة ومستقلة من الناحية العرقية.	Contradiction	0.999	Contradiction	0.902
2	أدت النكبة إلى تقسيم الشعب الفلسطيني إلى معازل مفصولة وتعرض لأكثر من مليون فلسطيني للاعتقال والتعذيب والحرمان منذ عام 1967.	النكبة هي حدث ضروري، لأنها حققت ذاتية صهيونية نقية مغلقة ومستقلة من الناحية العرقية.	Contradiction	0.999	Contradiction	0.906
3	النكبة هي كلمة عربية تعني الكارثة، وهي حدث أليم استمر نحو سنتين من منتصف عام 1947 إلى منتصف عام 1949.	تستمر النكبة تحدث لم يحدث في الماضي في الحدوث أيضاً اليوم.	Neutral	0.931	Contradiction	0.885
4	تضمنت النكبة تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدنية وتهجير أكثر من 750 ألف فلسطيني.	إذا لم يكن الفلسطينيون متواجدين 'حقاً'، فلا يمكن أن يحدث طردهم أيضاً.	Neutral	0.996	Contradiction	0.546

Table 2: Samples of Comparison Results for facts Extracted Using TextRank

#	Premise	Hypothesis	XLM-RoBERTa Label	XLM-RoBERTa Probability	BART Label	BART Label Probability
1	وتهجير أكثر من 750 ألف فلسطيني،	فإنه ليس من الممكن أن يتم تطهير ما يقرب من 800000 شخص عرقياً من البلاد،	Contradiction	0.982	Contradiction	0.946
2	أكثر من مليون فلسطيني تعرض للاعتقال والتعذيب والحرمان منذ احتلال عام 1967،	تجاهلت الصهيونية منذ وقت مبكر وجود السكان العرب في فلسطين.	Neutral	0.779	Contradiction	0.506
3	فقد سيطر اليهود بالقوة عام 1948 على نحو 78% من أراضي فلسطين التاريخية، وأنشؤوا دولتهم على هذه المساحة .	كان على المشروع الصهيوني إخلاء سكان البلاد من أجل تحقيق ذاته.	Neutral	0.998	Contradiction	0.810
4	أن كل الأهوال التي مر بها شعبنا الفلسطيني في تلك الفترة كانت جزءاً من مخطط متعمد من القيادة العليا للحركة الصهيونية،	تستمر النكبة تحدث لم يحدث في الماضي في الحدوث أيضاً اليوم.	Contradiction	0.967	Contradiction	0.855

project to “realize itself”). This pair was mislabeled by both models.

Since comparing the different summarization approaches (using Gemini or TextRank) was not feasible due to the limitations of the NLI task, it is more effective to select the appropriate approach based on the specific use case. We recommend that if a researcher wants to quote the translation as it is and point to contradictions without external paraphrasing, then using the TextRank algorithm is preferred. However, if the text requires paraphrasing or further explanation, then benefiting from the abilities of Google’s Gemini might be a better choice. In other words, if the researcher prefers an extractive summary, TextRank should be used, but if an abstractive summary is needed, then Gemini is the appropriate choice.

3.5 Finding Missing Facts

This section addresses the other important aspect of the paper: determining whether a specific fact can be found in one of the subject texts but not in the other. This approach allows the reader to detect biases in narratives, and find patterns of what facts tend to be omitted frequently.

Each sentence in the extracted facts lists is embedded using a sentence transformer model. This generates a vector embedding for each sentence, and therefore comparing a pair of sentences can be done using cosine similarity between their respective vectors. Every pair combination from the two lists are compared. If a specific sentence scores a low similarity when paired with every sentence of the other list, this sentence is considered to be a unique facts, meaning that the other text does not contain a similar sentence, indicating that such a fact is dropped or left out from the other narrative, or simply that the topic of the fact is not found in the other text.

Some facts mentioned in (Naim, 2023) but not in (Bronstein, 2009) discuss some crucial implications of the Nakba. This includes the statement “7,000,000 Palestinian (Refugees) in the diaspora suffer from deprivation of the most basic rights and are subjected to persecution and harassment” (original: “أن 7 ملايين فلسطيني في الشتات يعانون الحرمان من أبسط الحقوق الأساسية”). Such an important fact was not found in (Bronstein, 2009), which might be an indication of bias. Other facts from (Naim, 2023) that are not mentioned in (Bronstein, 2009) can be found in tables 4 and 6 of the appendix.

On the other hand, (Bronstein, 2009) speaks more on the views of the Zionists leaders regarding the Nakba, a stance which is not found in (Naim, 2023). An example is “Attitudes of the leaders and architects of Zionism towards the indigenous inhabitants of ‘Zion’ were situated between their perception as (temporary) guardians or holders of the land on one end”. Other facts from (Bronstein, 2009) that are not mentioned in (Naim, 2023) can be found in tables 3 and 5 in the appendix.

4 Limitations and Future Work

4.1 Limitations

One of the main limitations of the summarization approach using Gemini is that it may not capture all the key points accurately, potentially omitting or misrepresenting some of them. This can lead to contradictions when comparing the summarized information with other texts. The accuracy of the model remains a concern, as it may struggle to extract or interpret some of the essential details correctly, affecting the reliability of the summaries.

A significant limitation to address is the translation model’s inability to distinguish effectively between past and present contexts. For example, the sentence “it is, therefore, not possible that some 800,000 persons were ethnically cleansed from the country” was translated as “أن الممكن من ليس فإنه البلاد, من عرقياً شخص 800000 من يقرب ما تطهير يتم” which does not properly capture the past context of the event, and instead reflects a present action. Such subtle contradictions may be harder to detect.

Some sentences in Arabic have different syntax structures not found in English. For example, a sentence like “الشعب الفلسطيني تحت الاحتلال” (The Palestinian people (is) under occupation) lacks a verb but conveys a clear fact. Many tools rely on complete sentence structures, making it hard to handle nominal or verbless sentences while preserving their meaning.

4.2 Future Work

A key direction for future work includes a quantitative evaluation of the proposed approach to measure its performance and reliability in contradiction detection.

Future work could focus on improving the translated model to handle text translation from any language to Arabic. This enhancement would help ensure that the model can process a broader range of content. Additionally, by refining the model, it

may be possible to detect contradictions more effectively across various texts. For instance, when discussing historical events like the Nakba, the model can be trained to prioritize Arabic sources, as they are more likely to contain the correct points about the event compared to other languages.

Furthermore, the approach we presented in this paper can be the foundation for other work focusing on other pieces of Nakba history, such as oral history related to the Nakba and the 1948 war, providing a more comprehensive understanding of this historical event.

5 Conclusion

In this paper we proposed a method of contradiction detection in historical texts about the Nakba, and the sensitivity of dealing with the different narratives surrounding the issue. We proposed the use of Google's Gemini to provide context-aware translations of texts in English, as the audience this work is directed towards is Arabic-speaking historians and experts. In addition, we also compared prompting Gemini to provide facts summarized from textual content, in addition to using the TextRank algorithm for the same purpose. The core of the paper then employs NLI models such as XLM-RoBERTa (xlm-roberta-large-xnli) and BART (bart-large-mnli) to detect contradictions in pairs of statements taken from different texts about the Nakba. The findings suggest that the performance of these models on this specific task is influenced by the complexity of the sentences and the Arabic linguistic features in general. Another important part of this paper is a suggested method of finding sentences or topics that are not mentioned in a specific text with a specific narrative. Our aim was that the methodologies suggested in this paper enable an expert in history to gain deeper analysis of biases, contradictions, and gaps in historical narratives from both sides of history.

References

- Guangsheng Bao, Zebin Ou, and Yue Zhang. 2023. Gemini: controlling the sentence-level summary style in abstractive text summarization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 831–842.
- Rugved Bongale, Talha Chafekar, Mayank Chowdhary, and Tushar Bapecha. 2022. Au-

tomatic news summarizer using textrank. In *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, pages 1–5. IEEE.

Eitan Bronstein. 2009. The Nakba: Something that did not occur (although it had to occur). https://www.zochrot.org/publication_articles/view/50644/en?The_Nakba_Something_That_Did_Not_Occur__Although_It_Had_to_Occur__.

Khlood Al Jallad and Nada Ghneim. 2022. Arnli: Arabic natural language inference for entailment and contradiction detection. *arXiv preprint arXiv:2209.13953*.

Chirantana Mallick, Ajit Kumar Das, Madhurima Dutta, Asit Kumar Das, and Apurba Sarkar. 2019. Graph-based text summarization using modified textrank. In *Soft Computing in Data Analytics: Proceedings of International Conference on SCDA 2018*, pages 137–146. Springer.

Diana Muir. 2008. A land without a people for a people without a land. *Middle East Quarterly*, 15(2).

Aya Murra, Nada Hamarsheh, and Zahia Elabour. 2024. *Automated Generation and Improvement of Arabic Wikipedia Web Pages*. Graduation project report, Electrical and Computer Engineering Department, Birzeit University.

Balaji N., Deepa Kumari, Bhavatarini N., Megha N., Sunil Kumar, et al. 2022. Text summarization using nlp technique. In *2022 International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, pages 30–35. IEEE.

Basem Naim. 2023. النكبة... أكبر من مجرد حدث تاريخي [the Nakba ..more than just a historical event]. <https://www.aljazeera.net/opinions/2023/6/4/%D8%A7%D9%84%D9%86%D9%83%D8%A8%D8%A9-%D8%A3%D9%83%D8%A8%D8%B1-%D9%85%D9%86-%D9%85%D8%AC%D8%B1%D8%AF-%D8%AD%D8%AF%D8%AB-%D8%AA%D8%A7%D8%B1%D9%8A%D8%AE%D9%8A>.

Yixin Nie, Yicheng Wang, and Mohit Bansal. 2019. Analyzing compositionality-sensitivity of nli models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6867–6874.

Palestinian Central Bureau of Statistics (PCBS). 2008. Special Report on the 60th Anniversary of the Nakba. https://www.pcbs.gov.ps/Portals/_pcbs/PressRelease/nakba%2060.pdf.

Palestinian Central Bureau of Statistics (PCBS). 2021. Statistical Review on the 73rd Annual Commemoration of the Palestinian Nakba. https://www.pcbs.gov.ps/portals/_pcbs/PressRelease/Press_En_10-5-2021-nakba-en.pdf.

Adnan Souri, Mohammed Al Achhab, Badr Ed-dine El Mohajir, Mohamed Naoum, Outman El Hichami, and Abdelali Zbakh. 2023. Arabic text summarization challenges using deep learning techniques: A review. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(11s):134–142.

United Nations. 2024. About the Nakba - Question of Palestine. <https://www.un.org/unispal/about-the-nakba/>. Retrieved November 10, 2024.

A Appendix

Unique TextRank with original excerpts from (Bronstein, 2009)		
نُظمت مراسم لإحياء الذكرى قرب 'سينما سيتي' (هيرتسليا) عن قرية إجليل الفلسطينية التي كانت قائمة في ذلك الموقع حتى عام 1948.	"In March 2004 a commemoration was held near the 'Cinema City' (Herzliya) for the Palestinian village of Ijlil which existed at the site until 1948"	(Bronstein, 2009)
كانت مواقف قادة ومعماري الصهيونية تجاه السكان الأصليين لـ 'صهيون' في مكان ما بين اعتبارهم أوصياء (مؤقتين) أو حائزين للأرض في أحد طريقي نقيض.	"Attitudes of the leaders and architects of Zionism towards the indigenous inhabitants of 'Zion' were situated between their perception as (temporary) guardians or holders of the land on one end"	(Bronstein, 2009)
"تم الإشارة إلى ماضي عام 1948 فقط بما يتماشى مع السرد الصهيوني الذي يرى أنه 'كما لم يقبلونا هنا في الماضي (على سبيل المثال وفقاً لخطة تقسيم الأمم المتحدة)'"	"Reference to the past of 1948 is made only in line with the Zionist narrative which holds that, 'just like they did not accept us here in the past (e.g. according to the UN Partition Plan)'"	(Bronstein, 2009)

Table 3: Unique TextRank facts from (Bronstein, 2009)

Unique TextRank facts with translations from (Naim, 2023)		
"تم تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدينة"	"More than 530 Palestinian villages were destroyed, and control was established over more than 700 villages and towns."	(Naim, 2023)
"أسست أول جمعية نسائية في فلسطين عام 1903،"	"The first women's association in Palestine was established in 1903."	(Naim, 2023)
"فقد سيطر اليهود بالقوة عام 1948 على نحو 78% من أراضي فلسطين التاريخية، وأنشؤوا دولتهم على هذه المساحة."	"In 1948, Jews seized nearly 78% of historical Palestine by force and established their state on this land."	(Naim, 2023)
"أكثر من 76% من مساحة الضفة الغربية تحت السيطرة الإسرائيلية الكاملة،"	"More than 76% of the West Bank is under full Israeli control."	(Naim, 2023)
"أن 7 ملايين فلسطيني في الشتات يعانون الحرمان من أبسط الحقوق الأساسية"	"Seven million Palestinians in the diaspora suffer from deprivation of basic human rights."	(Naim, 2023)
"ويتعرضون بشكل متكرر لموجات من الاضطهاد والملاحقة في الدول المختلفة،"	"They are repeatedly subjected to waves of persecution and harassment in various countries."	(Naim, 2023)

Table 4: Unique TextRank facts from (Naim, 2023)

Unique Gemini facts with translations from (Bronstein, 2009)		
<p>بُنيت الهوية الصهيونية منذ البداية على نفي مزدوج: نفي زمن ومكان اليهود خارج صهيون، ونفي زمن ومكان السكان الأصليين لإقليم صهيون.</p>	<p>”Zionist identity was built from the beginning on a two-fold negation: it negates time and space of the Jews outside Zion, a ‘negation of exile’ which extends beyond the realm of religion, and it negates time and space of those indigenous to the territory of Zion”</p>	(Bronstein, 2009)
<p>وفقاً للصهيونية، فقد وقعت الأحداث العنيفة التي حدثت في حوالي عام 1948 بالفعل، ولكن فقط في شكل رد غير قابل للتجنب على الاضطرابات التي تسبب فيها السكان المحليون الذين لم يقبلوا بإنشاء الكيان الجديد، الدولة اليهودية.</p>	<p>”According to Zionism, the violent events around 1948 did in fact occur, but only in form of an unavoidable response to the disturbance caused by the ‘locals,’ who did not accept the establishment of the new entity, the Jewish State.”</p>	(Bronstein, 2009)
<p>النكبة هي حدث ضروري، لأنها حققت ذاتية صهيونية نقية مغلقة ومستقلة من الناحية العرقية.</p>	<p>”On the other hand, and paradoxically, the Nakba – the violent expulsion of the inhabitants of the country and the transformation of those remaining into refugees in their homeland, or into second-class citizens – is a necessary event, because it brought about the realization of the ethnically pure, closed and autonomous Zionist subject which builds itself in the framework of a state aimed exclusively for him/her.”</p>	(Bronstein, 2009)
<p>يصف يوسف فايتس، أحد رؤساء الصندوق القومي اليهودي في ذلك الوقت، تدمير قرية زرنوقة بعد طرد سكانها، على الرغم من دعوات عديدة من اليهود بالامتناع عن طردهم.</p>	<p>”Yosef Weitz, one of the heads of the Jewish National Fund at the time, provides evidence which is surprising in its honesty. He tells of the destruction of the village of Zarnuqa after its inhabitants had been expelled, despite of numerous calls by Jews to abstain from their expulsion.”</p>	(Bronstein, 2009)

Table 5: Unique Gemini facts with original excerpts from (Bronstein, 2009)

Unique Gemini facts with translation from (Naim, 2023)		
تضمنت النكبة تدمير أكثر من 530 قرية فلسطينية والسيطرة على أكثر من 700 قرية ومدينة وتهجير أكثر من 750 ألف فلسطيني.	"The Nakba included the destruction of more than 530 Palestinian villages, the control of over 700 villages and towns, and the displacement of more than 750,000 Palestinians."	(Naim, 2023)
أكثر من 76% من مساحة الضفة الغربية تحت السيطرة الإسرائيلية الكاملة، وقطاع غزة تحت حصار إسرائيلي خانق منذ أكثر من 17 عاماً.	"More than 76% of the West Bank is under full Israeli control, while the Gaza Strip has been under a suffocating Israeli blockade for more than 17 years."	(Naim, 2023)
يعاني 7 ملايين فلسطيني في الشتات من الحرمان من أبسط الحقوق الأساسية ويتعرضون للاضطهاد والملاحقة.	"Seven million Palestinians in the diaspora suffer from deprivation of basic rights and are subjected to persecution and harassment."	(Naim, 2023)

Table 6: Unique Gemini facts with translation from (Naim, 2023)