



NLP: AUTOMATING THE CREATION OF NEWS DIGESTS

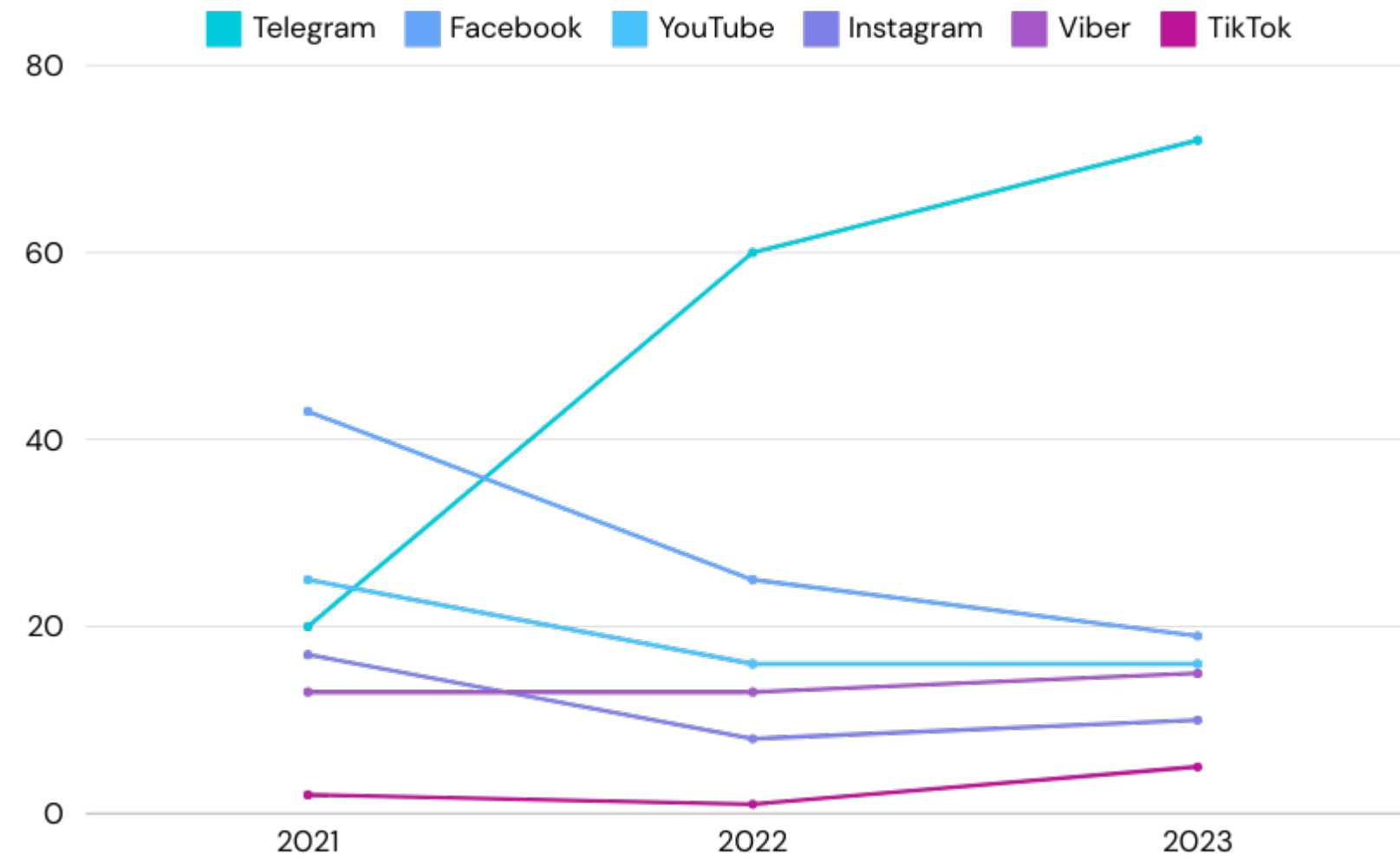
Stanislav Orlov,
SE-3 student

Scientific Advisor:
Oleh Smysh, Senior Lecturer



2024

REASONING



Social media the Ukrainians use as a primary news source
USAID- Internews surveys on media consumption

OBJECT

processing of Ukrainian- language news text data

SUBJECT

text analysis using various natural language processing instruments and methods and their usage to create a chatbot for generating news digests

GOAL

Filling the gap in the currently available resources designed to automate the creation of digests based on the content of Ukrainian-language Telegram news channels

TASKS

01

researched available tools and methods of Ukrainian NLP

02

analyzed available models

03

investigated the creation of digests

04



developed a chatbot for automatic regular summarization



STRUCTURE OF DIGEST

Digests with the same regularity	
Hyperlink to the full article on the main word	
Template headline	
Template ending	
Template bullet points	
Template news divider	

-  Template headline
-  Template bullet points
-  Template ending
-  Main words as hyperlinks

-  Channels with declared characteristics
-  Channels without declared characteristics

Ukraine NOW

Головні новини за 9 травня:

- Верховна Рада **посилила** штрафи за порушення правил мобілізації
- Дві українські ГЕС **вивели** з експлуатації після російського обстрілу

...

- Як тільки обіцяні США об'єми допомоги надійдуть - ми **зупинимо** ініціативу військ РФ на Сході, - Зеленський
- Німеччина **придбас** у США три установки HIMARS і передасть їх Україні

✓ Підписуйся на Ukraine NOW

📅 9.05.2024, 23:40 👁️ 99.6K

СУСПІЛЬНЕ КУЛЬТУРА

Оксана Забужко на «Берлінале», Ріпін українець та «Євробачення» — **головні новини тижня від Суспільне Культура:**

- ◆ Оксана Забужко **увійшла** до складу журі основного конкурсу «Берлінале-2024».
- ◆ Переможницями Нацвідбору на «Євробачення-2024» **стали** alyona alyona & Jerry Heil з піснею Maria & Teresa.

...

- ◆ У Великій Британії **опублікують** книгу про українське мистецтво від початку XX століття до сьогодні.
- ◆ Universal Music Group не **продовжив** контракт з TikTok, і тому із застосунку видалили пісні Тейлор Свіфт, Біллі Айліш та Гаррі Стайлза. Видалення **торкнулося** і музикантів лейблу ENKO.

Щоб не пропустити головні новини культури, підпишіться на Telegram-канал Суспільне Культура.

📅 4.02.2024, 21:00 👁️ 146.8K

REFERENCE-BASED EVALUATION

Агент Києва

! У Києві терміново евакуюють дві лікарні, одна з яких призначена для дітей, розташованих на вулиці Богатирській.

Цей захід є реакцією на поширення відео в інтернет-медіа, де натякають на можливий удар ворога по цим медичним установам та неправдиві твердження про присутність військових у лікарнях.

Щоб забезпечити безпеку хворих дітей, їхніх батьків і медичного персоналу, влада Києва здійснює заходи з перевезення пацієнтів та медиків до інших медичних установ міста.

[Агент Києва | Підписатись](#)

📅 26.04.2024, 16:32

👁️ 10.5K

Агент Києва

Головне за день:

• 80% міських водопровідних мереж потребують негайного відновлення через аварійний стан, попереджає депутат Київради Андрій Вітренко.

• У Києві терміново евакуюють дві лікарні, одна з яких призначена для дітей, розташованих на вулиці Богатирській.

• Шмигаль оголосив про затвердження постанови, що продовжує пільгову ціну на газ для постачальників тепла та гарячої води до кінця літа.

• США оголосили про запровадження найбільшого пакету допомоги на \$6 мільярдів на засіданні "Рамштайну".

[Агент Києва | Підписатись](#)

📅 26.04.2024, 22:31

👁️ 16.7K

Reference: Соловейко в темнім гаї сонце зустрічає.

Prediction: Соловейко зустрічав світанок у темному гаї.

ROUGE

ROUGE-N

Counts overlap of n-grams

ROUGE-1

ROUGE-2

ROUGE-L

Counts overlap of the LCS



Overlap: 5
 Recall: 5/6
 Precision: 5/6
 F1: 0.83



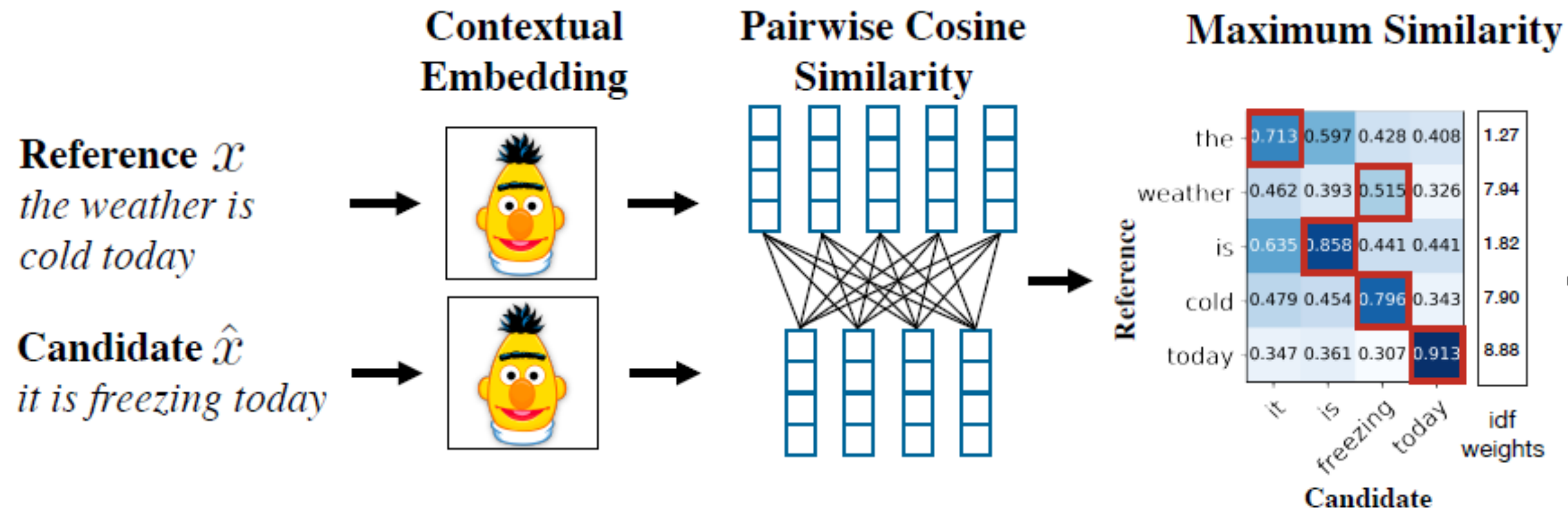
Overlap: 2
 Recall: 2/6
 Precision: 2/6
 F1: 0.33



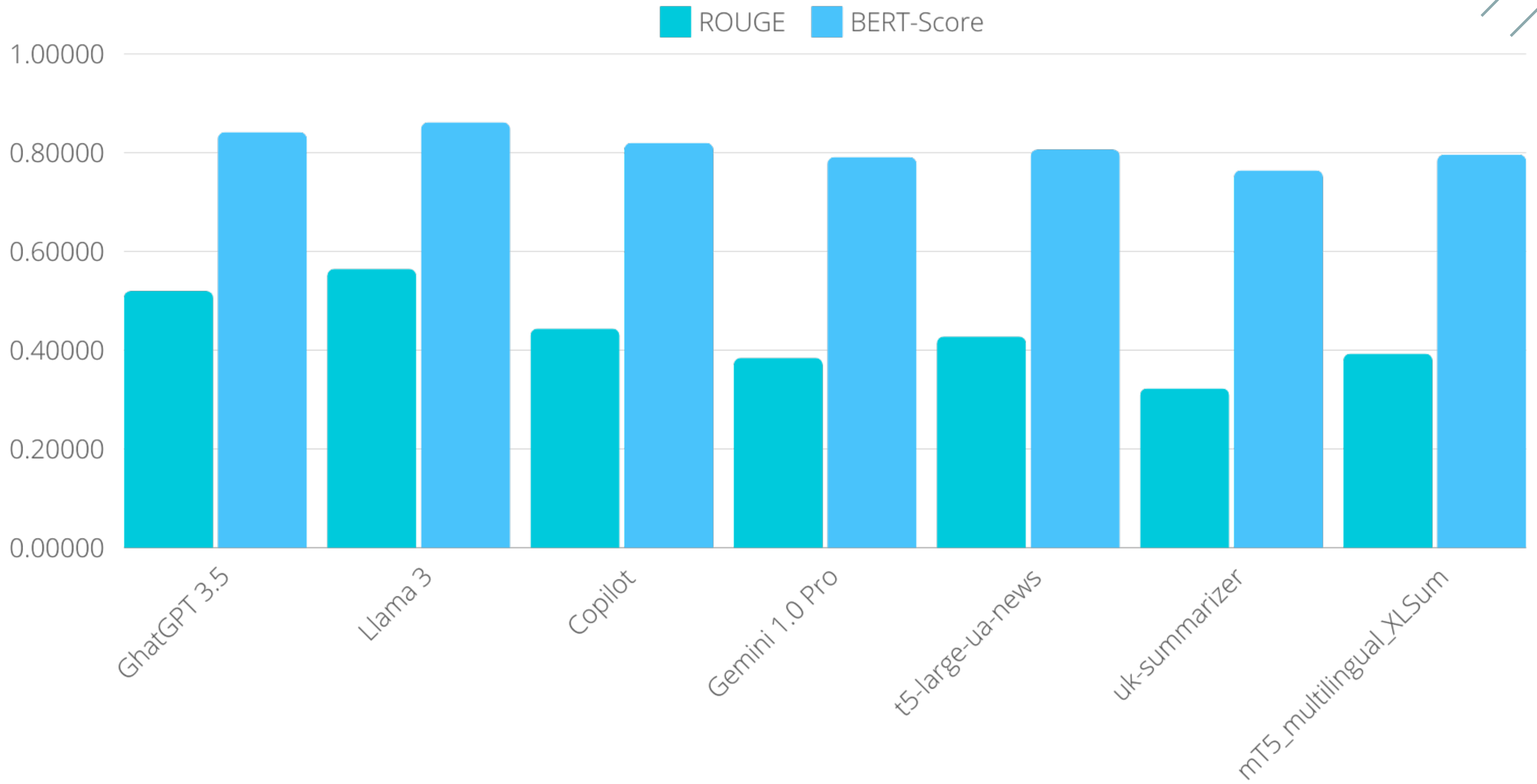
Overlap: 4
 Recall: 4/6
 Precision: 4/6
 F1: 0.67

BERT-SCORE

Calculates pairwise cosine similarity between each word



RESULT OF EVALUATION



BOT'S FUNCTIONALITY

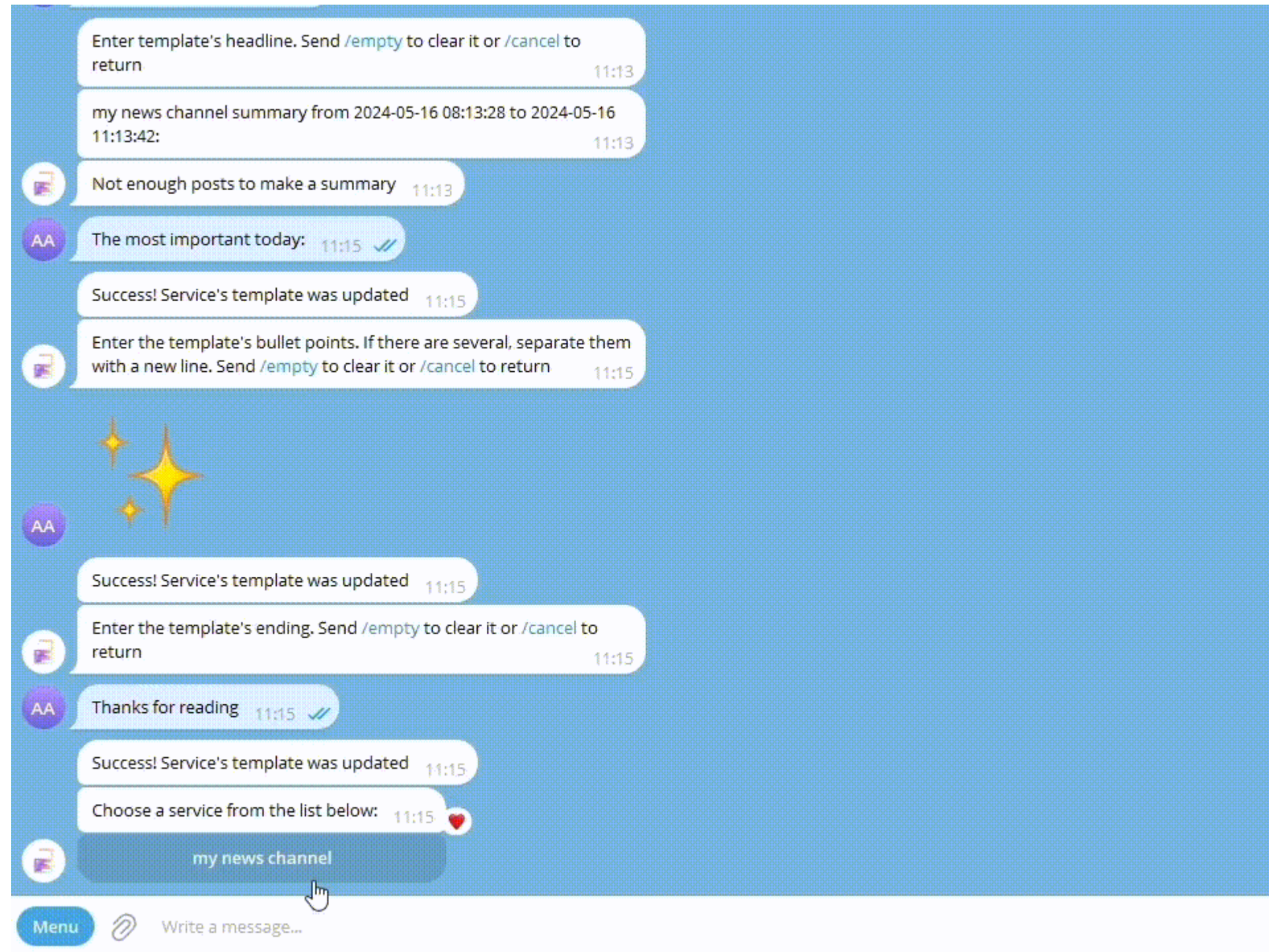
MANAGING CHANNELS

- Add a channel to the bot
- Remove the channel from the bot
- Set a limit on the amount of news
- Set digests' frequency
- Set the minimal positivity of news
- Set the headline for the channel's digests
- Set the bullet points for the channel's digests
- Set the ending for the channel's digests
- Generate a digest

CREATING DIGESTS

- Calculate the message's positivity
- Define sentences' main words to attach links
- Shorten news to their headlines
- Sort the news in order of their importance
- Set a timer for the next digest

DEMONSTRATION



APPROBATION

The abstract of the coursework was included in the conference paper "Scientific Horizons of the 21st Century: Interdisciplinary Studies"

CONCLUSIONS

- The goal was achieved, as the deficiency in the available resources was addressed.
- The tasks were accomplished due to the following:
 - Ukrainian NLP tools were examined.
 - dataset comprising news headlines and full texts was gathered;
 - digests were analyzed to pinpoint the common structural elements;
 - accuracy of summarization was evaluated based on the collected dataset;
 - Telegram chatbot for generating news digests was developed with NLP techniques.
- The developed instrument optimizes the process of creating news digests, while the accompanying research lays the groundwork for future advancements and applications in this area.

THANK YOU

