

Final Presentation

Jingheng Pan

jingheng.pan@studium.uni-hamburg.de

Adrian Lindloff

adrian.lindloff@studium.uni-hamburg.de

Lijunnan Bai

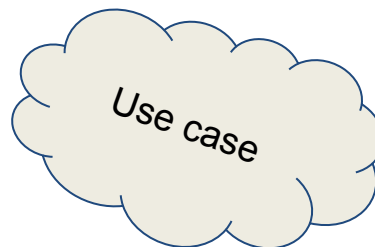
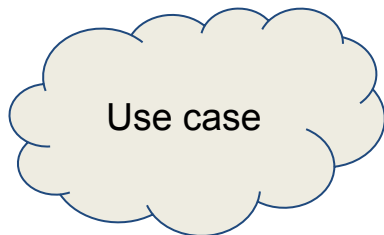
lijunnan.bai@studium.uni-hamburg.de

Structure

1. What is the Adaptive Storyfinder?
2. How did the project go?
3. Core functionality
4. Demo
5. Outlook

Use cases:

- People want to see stuff about things that they like
→ discover interesting new information
- A person would like to read an article and focus on desirable sections
→ made aware of interesting parts



=> Goal of the Adaptive Storyfinder:

- Suggest reading material which matches the preferences
→ User preference-oriented ranking of search results
- Find interesting parts of a website according to user preferences

“Adaptive Storyfinder”

What is the Adaptive Storyfinder?

- A System which leverages user preferences to fit the users needs
- “Adaptive” Storyfinder assists in finding interesting content

Why “adaptive”?

- Different users get different answers tailored to their interests

How did the project go?

- A lot of struggles on finding approaches on what exactly to do
- What do we want to implement? Where do we want to take the idea?
 - ~~Behavioral data~~
 - ~~Eyetracking~~

=> User-related content to determine user preferences

Backend:

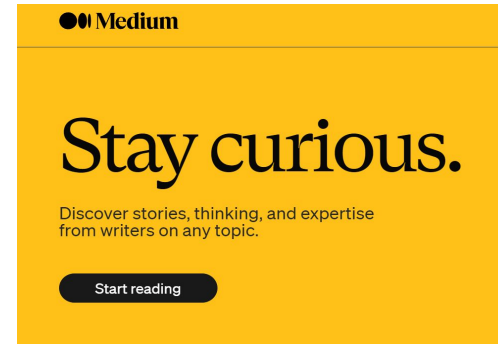
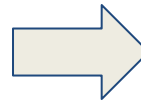
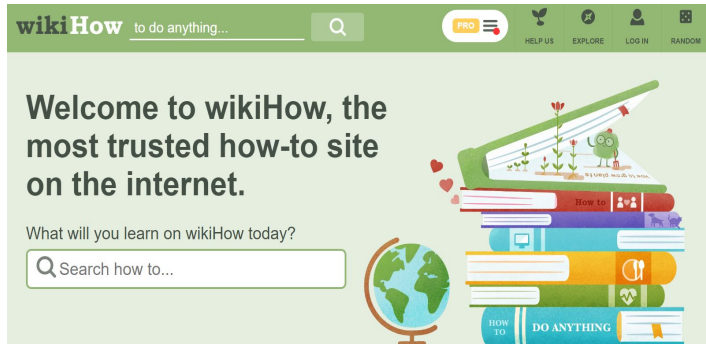
- FastAPI
- Docker
- PyTorch for ML + HuggingFace

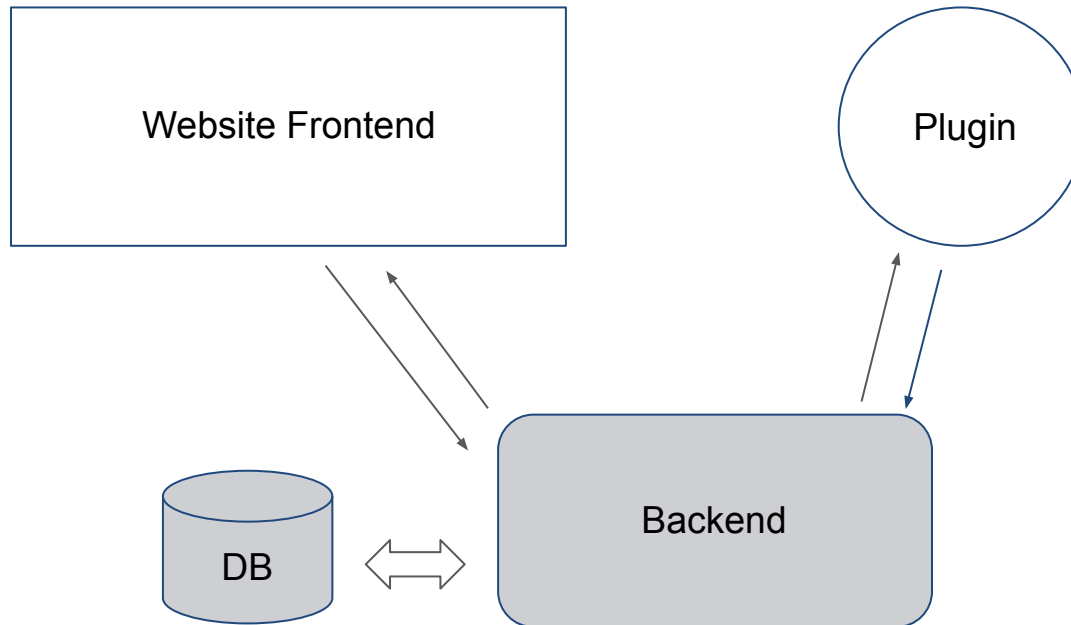
Frontend:

- Vue.js with mainly Vuetify & Bootstrap

Article database:

- From WikiHow API & dataset to a medium dataset





Backend deals with:

- User Management
- User Histories

For the Website Frontend:

- Re-ranking
- Topic recommendation


For the plugin:

- Highlighting of paragraphs

Core functionality

- Search feature
- History feature
- Topic feature
- Plugin

Search feature

 **Mediur**


Food Beverly Hills Recipe

Going Down the Restaurant
Memory Lane of My Childhood

text1

Danna Reich Colman Thursday June 30, 2016

SUMMARIZE

 **Mediur**


Business Loyalty Program Restaurant Business Rewards Prog

I ordered chole bhature and
received customer experience in
return

text1

Danna Reich Colman Thursday June 30, 2016

SUMMARIZE

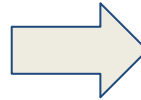
 **Mediur**

Food Beverly Hills Recipe

Is Your Loyalty Program
Rewarding the Right
Customers?

text1

SUMMARIZE



Shiv Pratap Singh Rajawat

Wednesday December 9, 2020



Electric Car Platform: Bi-directional Charging

Hyundai on Tuesday revealed more about the platform, called E-GMP, that will underpin next-generation electric vehicles sold across Hyundai Motor Group brands — including Hyundai, Kia, and Genesis. A high-performance model based on E-GMP will be able to dash from 0–62 mph in less than 3.5 seconds, and...

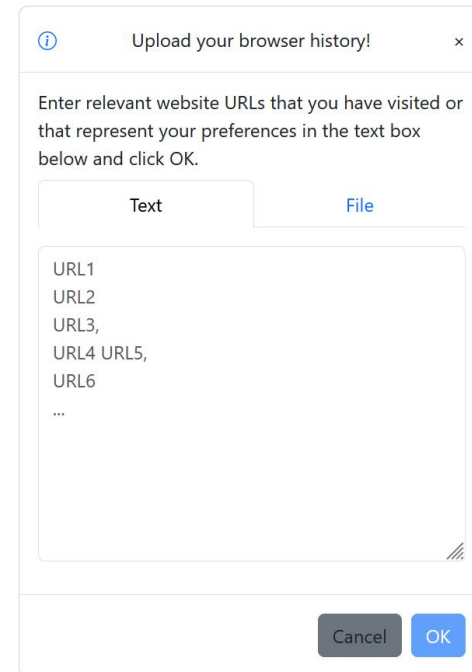
BUSINESS

ENVIRONMENT

TECH

History feature

- A browser history partly contains a users interests
→ can be leveraged to enhance the pursuit of interesting new articles
- Users can specify their favorite websites
- Users can upload their favorite websites
- Users can automatically send their chrome history



The screenshot shows a web interface for uploading browser history. It has a title bar with an information icon, the text 'Upload your browser history!', and a close button. Below the title bar is a text area with the instruction: 'Enter relevant website URLs that you have visited or that represent your preferences in the text box below and click OK.' There are two tabs: 'Text' (selected) and 'File'. The 'Text' tab contains a text input field with the following text: 'URL1', 'URL2', 'URL3,', 'URL4 URL5,', 'URL6', and '...'. At the bottom right of the form are two buttons: 'Cancel' and 'OK'.

History feature

- Management of user histories

Search

Date		Upload Number	Delete
▼	14.09.2023	1	
▲	18.09.2023	2	

Number	Title	URL	Delete
0	Medizin-Nobelpreis für Wegbereiter der Covid-19-Impfung	https://www.welt.de/wissenschaft/article247774502/Medizin-Nobelpreis-fuer-Grundlagenforschung-zu-Covid-19-Impfung.html	
1	Taktische Züge 2	https://www.spiegel.de/politik/deutschland/news-des-tages-eu-aussenminister-in-kiew-aerztestreik-nobelpreis-fuer-medizin-a-1494d6e8-689c-48df-9d26-be635b06061a	

Rows per page: 10 ▼ 1-2 of 2 < >

Inspiration: Predefined tags in the original dataset do not scale efficiently

Topic model:

- Using Bert-based pre-trained model
 - News-category-classification (42 categories)
- Applied to our dataset → Categorize websites consistently
- Reliable way to capture interest groups
 - Used to recommend new articles & to aid in the re-ranking process



WELLNESS

PARENTING

HEALTHY LIVING

WOMEN

Topic feature

TECH



In the beginning, there was void
*

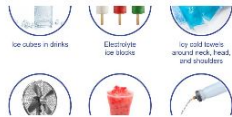


It's notification time... Slack
Notifications, that is!



How to install multiple version
of Ruby and PostgreSQL on
Mac OS Catalina 🐉

WELLNESS



Nutrition 101 for Junior
Cyclists



From "MD" to "MKT": digital
marketing as a kick-start for
career change

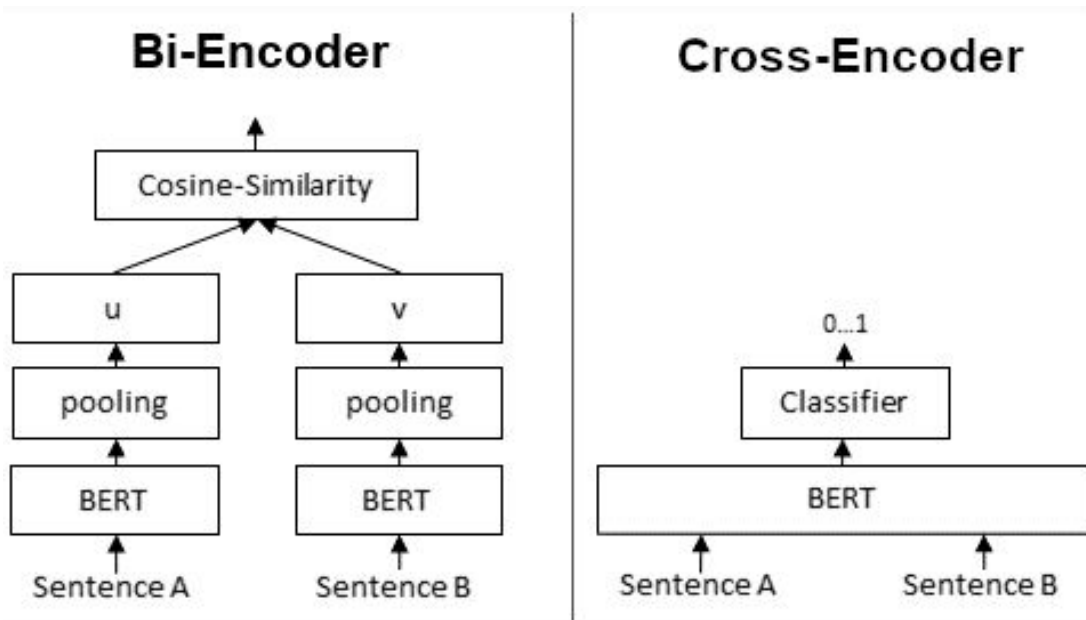


This glorious 65-inch 4K Roku
TV is a jaw-dropping \$400
today

Top topics

WELLNESS BUSINESS SPORTS WORLD NEWS
ENVIRONMENT TECH POLITICS HOME & LIVING

Sentence-transformers



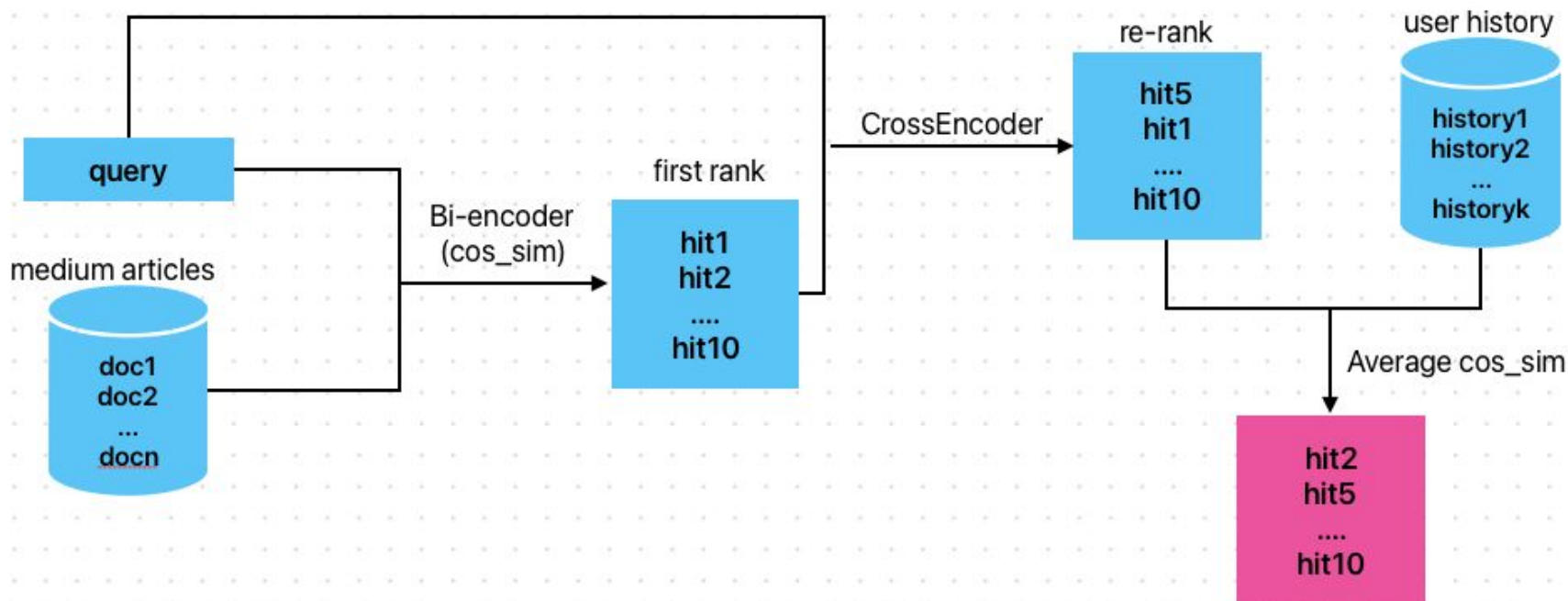
Bi-encoder:

- Pass sentence independently
- Output: embeddings

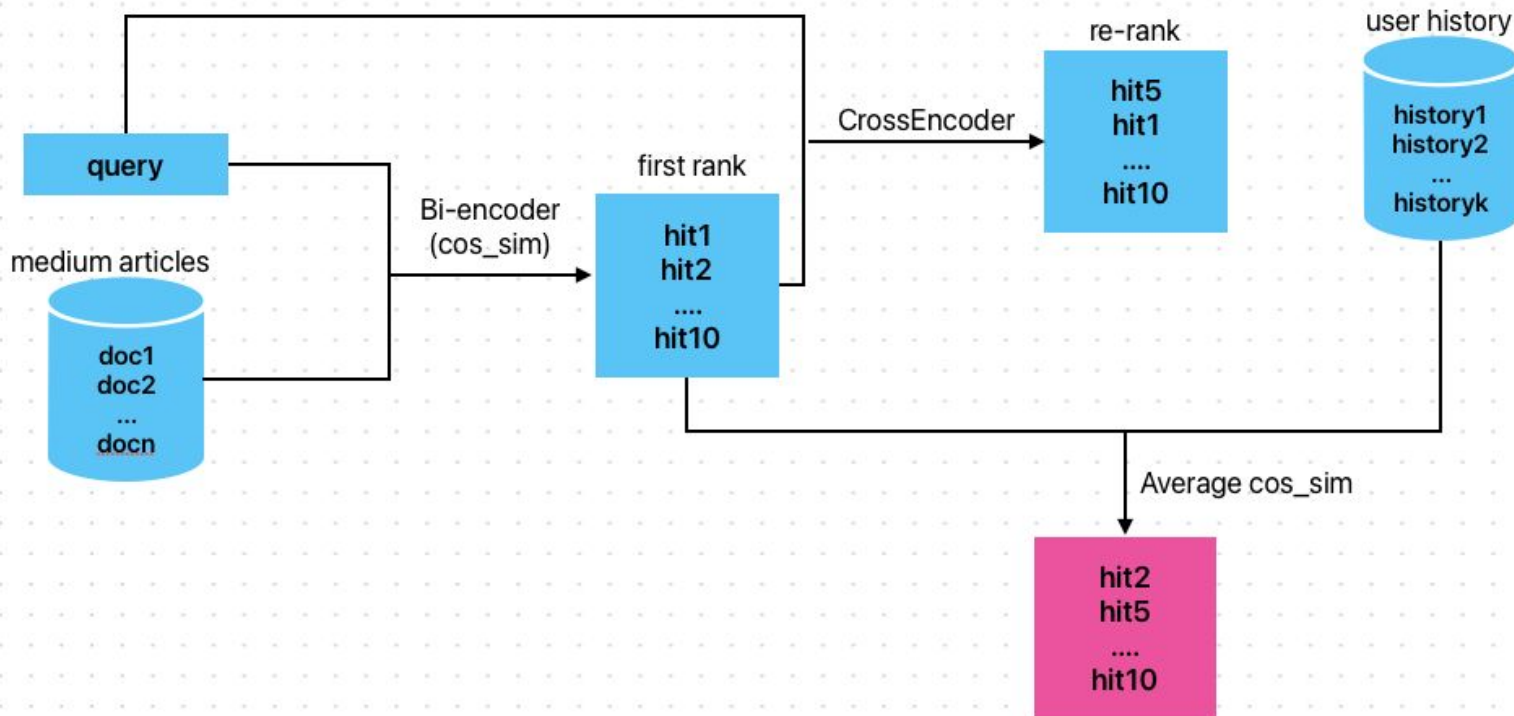
Cross-encoder:

- Pass sentence simultaneously
- Output: possibility of the similarity of input sentences

Ideally previous schema



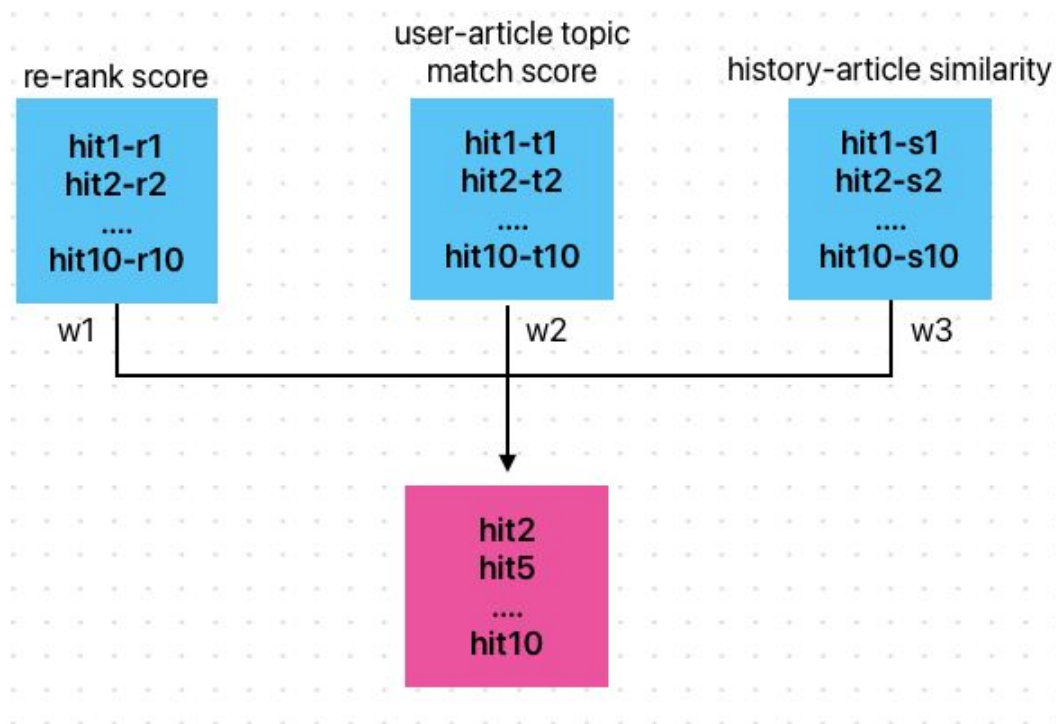
Real previous schema



Old pipeline:

- Query search: Re-ranker for query retrieval
- Preference search: Cosine-similarity between history and candidates
- Preference search has nothing to do with the results from query search

Optimized schema



User-article topic match score

user topic occurrence ratio

$[r_1, r_2, \dots, r_{42}]$



article topic occurrence

article 1: $[o_1, o_2, \dots, o_{42}]$
article 2: $[o_1, o_2, \dots, o_{42}]$
....
article 10: $[o_1, o_2, \dots, o_{42}]$

=

user-article topic
match score

hit1-t1
hit2-t2
....
hit10-t10

history-article similarity

user history text

history 1
history 2
...
history 30



article candidates

article 1
articles 2
...
articles 10

=

article history similarity

article 1: [h1,h2,...h30]
article 2: [h1,h2,...h30]
...
article 10: [h1,h2,...h30]

avg

history-article similarity

hit1-s1
hit2-s2
...
hit10-s10

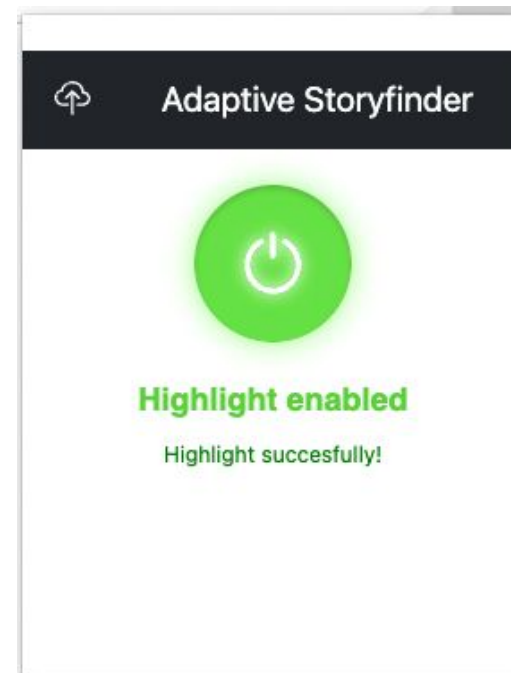
Old pipeline:

- Query search: Re-ranker for query retrieval
- Preference search: Cosine-similarity between history and candidates
- Preference search has nothing to do with the results from query search

New pipeline:

- Query search: Re-ranker for query retrieval
- Preference search: consider **re-rank score, history similarity and topic similarity score**

- Highlight the paragraphs
 - Paragraphs that might be interesting to user
 - According to history-paragraphs cos-similarity
- Automatically upload browser history
- Shares User Authentication with the Website



Demo

- Pipeline
 - Scale of the dataset
 - Re-Rank weights trade-offs
 - User-article topic match score is less convincing due to topic sparsity of articles
- Topic Model
 - Not very accurate to assign a topic
- Plugin
 - Paragraphs-article similarity is insufficient to suggest interesting paragraphs to user

- Include other sources of information beyond the medium dataset
→ expand dataset with more sources
- Generate AI Images for articles
- Implement user/ data privacy or at least inform the users about the scope of data usage

Any Questions?