Article

Not peer-reviewed version

Simulated User Behavior for Recommender Systems Applied to the MIND Dataset

Anthony S. , Ahmad Farooq * , Charnelle B. , Surbhi Rathore , Amanda P.

Posted Date: 2 April 2025

doi: 10.20944/preprints202504.0155.v1

Keywords: simulated user behavior; recommender systems; reinforcement learning (RL); large language models (LLMs); mind dataset

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

Article

Simulated User Behavior for Recommender Systems Applied to the MIND Dataset

Anthony S. [1], Ahmad Farooq [2,*], Charnelle B. [1], Surbhi Rathore [3] and Amanda P. [1]

1 U.S. Department of Defense
2 University of Arkansas at Little Rock
3 University of Rhode Island
* Correspondence: afarooq@ualr.edu

Abstract: The SUBER (Simulated User Behavior for Recommender Systems) framework, a novel approach in the field, has been extended and applied to the MIND (Microsoft News Dataset) to improve the performance of news recommender systems. This study addresses the need for more personalized news recommendations by leveraging SUBER to simulate realistic user interactions, especially in areas with sparse data. Utilizing the Stable Baselines3 library, extensive trials with various models were conducted to identify the most effective configurations. Our approach involved generating synthetic interaction data through simulations, which were then used to train and evaluate different recommendation models. Although the primary goal was to get SUBER to work effectively with the MIND dataset, the extension successfully integrated and operated within this new context. This study underscores the potential of advanced simulation techniques, like those provided by SUBER, to enhance the capabilities of news recommendation systems. Key achievements include successfully applying SUBER to the MIND dataset and demonstrating its robustness and adaptability in a new domain. This work sets a new benchmark for future research in news recommendation systems. It contributes to the broader field of recommender systems by showcasing the utility of user behavior simulation in optimizing recommendation algorithms, highlighting the importance of using simulated data to complement fundamental user interactions for improved user experiences and satisfaction with recommended content.

Keywords: simulated user behavior; recommender systems; reinforcement learning (RL); large language models (LLMs); MIND dataset

---

1. Introduction

Recommender systems are crucial components of contemporary digital platforms, offering personalized content suggestions to users. News platforms face distinct challenges, such as the ever-changing content and diverse user preferences, which make it difficult to provide accurate and timely recommendations. The effectiveness of news recommender systems is pivotal for enhancing user engagement and satisfaction, making them a unique and significant area of study. Using simulated user behavior has emerged as a promising technique to improve recommender systems' performance. The SUBER framework [1], applied initially to movie and book recommendations, incorporates large language models (LLMs) to generate synthetic user interactions to train recommender systems.

While systems such as news sites and book platforms often have access to user information and interactions for training, many do not release this data for research and experimentation purposes [2]. This limitation hinders the ability to conduct extensive research across multiple datasets. The benefit of simulated user data lies in its ability to generate user interactions for datasets where such data is unavailable, thus enabling more comprehensive experimentation and development of recommender systems. By simulating user behavior, SUBER provides a versatile tool for researchers,

 doi:10.20944/preprints202504.0155.v1

allowing them to work with various datasets without needing proprietary user data, thereby advancing research in the field.

This study extends the SUBER framework to the Microsoft News Dataset (MIND) [3] to enhance news recommender systems. Although the MIND dataset already includes user interaction histories, we utilize it to demonstrate the capability of SUBER to simulate user behavior effectively. This showcases SUBER's potential in enhancing recommender systems for datasets lacking user interaction data and its adaptability to different domains. Utilizing the Stable-Baselines3 library, we conducted extensive trials with various models to determine the most effective configurations. Our research demonstrates the versatility and robustness of SUBER, showcasing its applicability to different datasets and domains.

2. Materials and Methods

This study aims to extend the SUBER (Simulated User Behavior for Recommender Systems) framework to the Microsoft News Dataset (MIND) to enhance the performance of news recommender systems. We utilized frameworks and methodologies from the original work, including Stable-Baselines3 (SB3) [4], General Purpose Tensor Quantization (GPTQ) [5], Advantage Actor-Critic (A2C) [6], and large language Models (LLMs) [7]. These components were integral to achieving our objectives.

2.1. Materials

2.1.1. SB3

Stable-Baselines3 is a set of reliable implementations of RL algorithms in Python, built on top of PyTorch [8]. It provides a standardized interface for developing and testing RL algorithms. SB3 is considered the standard in RL research due to its flexibility, ease of use, and extensive support for a wide range of RL algorithms.

2.1.2. GPTQ

GPTQ is a technique used to compress large language models by quantizing the weights and reducing model size while maintaining performance. This method was essential for optimizing the large language models integrated into the SUBER framework, ensuring efficient processing and interaction simulation within the news recommendation context. GPTQ was a vital aspect of the original SUBER framework to balance model size, accuracy, and resource demands.

The maximum GPU size available for this research project was 24GB. This is comparable to the GPU used in the original work. The type of EC2 instance used was G6.4xlarge with a 24GB L4 GPU, 16 cores and 64GB of memory.

2.1.3. A2C

A2C is a synchronous, deterministic variant of the Asynchronous Advantage Actor-Critic (A3C) RL algorithm. It combines the benefits of policy gradient and value-based methods, making it suitable for training models that simulate user interactions within the news recommender system. A2C was chosen for its stability and efficiency, as it was a fundamental component in the original SUBER framework.

While other RL algorithms, such as Deep Q-Networks (DQN) [9], Proximal Policy Optimization (PPO) [10], and Trust Region Policy Optimization (TRPO) [11] were evaluated in the original work, we focused on A2C only. Recent advancements in reinforcement learning for text-based tasks, particularly those involving summarization and user feed- back modeling, have demonstrated the effectiveness of both policy gradient methods and hierarchical architectures [12,13]. These insights further validate the relevance of using A2C within the SUBER framework for simulating user interactions and learning effective recommendation strategies.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

2.1.4. LLMs

Large language models, such as GPT-3 and its variants, simulate realistic user interactions by generating synthetic interaction data. Incorporating LLMs in RL represents a novel approach, leveraging their capability to understand and generate human-like text to simulate complex user behaviors. These models were employed in two key ways:

Generating User Descriptions

LLMs created detailed descriptions of users and their news interests. This process is similar to the approach used for movies in the original SUBER work. Creating realistic and diverse user profiles helped the models understand user preferences and behaviors, which is essential for personalizing the recommendation system.

Shaping Rewards During Training

LLMs were vital in shaping rewards during the training of our recommendation models. They provided feedback by generating ratings that the described users would have given if they were real. This feedback was combined with the scores provided by the recommender system, ensuring a more accurate and realistic learning process. This methodology follows the original SUBER framework, leveraging LLMs' capabilities to generate human-like text and simulate complex user behaviors.

2.2. Methods

2.2.1. Initial Evaluation of SUBER

We began by evaluating SUBER against a movie dataset and reviewing the provided source code [12], cross-referencing with the paper. This initial step was crucial for gaining insight into the framework's structure and organization.

2.2.2. Dataset Preparation

The MIND dataset provided a comprehensive set of user interactions and news articles. Both the small and large MIND datasets were used. The small dataset is representative of 50,000 users, 7.5 million impressions, and 161000 news articles and is approximately 4GB in size. The large dataset is representative of 1 million users, 24 million impressions, 161,000 news articles, and 48GB in size.

The data comes in two files: news and behaviors. These files were transformed and combined into a single data file for use with the SUBER framework. Key metrics were created and provided to each news article. Analyzing the user behavior for each news article gained a deeper insight into user interactions with news articles. These metrics include click-through rate (CTR), vote average, vote count, and read Frequency:

- **Click-Through Rate (CTR):** This metric represents the ratio of clicks to impressions for each news article. It serves as a crucial indicator of user engagement, showing how often users click on an article when it is displayed. A higher CTR signifies greater interest and relevance of the article to the users, making it a valuable metric for assessing article performance in the recommender system.
- **Vote Average and Vote Count:** These metrics were maintained from the original work. The Vote Average provides a normalized measure of article engagement, reflecting users' average ratings based on clicks. The Vote Count indicates the number of times an article was displayed to users. This metric shows the article's exposure and reach within the dataset. Together, these metrics help in understanding the articles' popularity and visibility. The vote average and vote count were necessary for working in SUBER, as they were used in the original work, providing critical inputs for the system's reinforcement learning algorithms.
- **Read Frequency:** This metric shows how often an article appears in users' read history, providing insight into its recurrence and sustained interest among users. It is determined by

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

counting the occurrences of each article in the user history. High read frequency indicates that an article consistently attracts user attention, making it a critical factor for personalized recommendations.

By integrating these metrics into the preprocessed dataset, the new dataset offers a comprehensive view of user interactions with news articles. These enhanced insights are essential for developing and testing more accurate and personalized news recommender systems.

2.2.3. User Description Generation

The user generation technique originally used [1] was adapted to use GPT4o, asking for a user description consisting of name, age, gender, primary news interest, secondary news interest, and a description. An example prompt and result are provided in Appendix A.

2.2.4. Model Implementation and Usage

Using Stable-Baselines3, a custom A2C network for processing users and their interactions with news was constructed. It comprised a policy network for the actor and a value network for the critic. Additional layers were added to the models, and the input layers were adjusted to account for changes in the data format. The LLMs used include:

- TheBloke/Llama-2-7b-Chat-GPTQ
- TheBloke/Llama-2-13B-chat-GPTQ
- TheBloke/vicuna-7B-v1.3-GPTQ
- TheBloke/vicuna-13b-v1.3.0-GPTQ
- TheBloke/vicuna-7B-v1.5-GPTQ
- TheBloke/vicuna-13B-v1.5-GPTQ
- gpt-4o
- TheBloke/Mistral-7B-Instruct-v0.2-GPTQ

A provided LAS API gave access to the GPT4o model and seamlessly integrated into our experiments. It also decreased requirements for GPU resources. However, making the API calls took longer than accessing a model loaded in the local GPU, limiting experimentation.

The other models were available on Hugging Face and were implemented using the GPTQ (General Purpose Tensor Quantization) technique. This approach allowed the usage of more complex models that would have required more resources than were available. By leveraging both API-based and Hugging Face models, we were able to incorporate a diverse set of LLMs into our study.

2.2.5. Prompt Engineering

The movie prompts from the original work [1] were modified to align with news article recommendations. The system prompt establishes the role of the LLM. Here is an example:

> *You are a highly sophisticated news rating assistant equipped with an advanced understanding of human behavior. Your mission is to deliver personalized news recommendations by carefully considering the unique characteristics, tastes, and previously seen news of each individual. When presented with information about a specific news article, you will diligently analyze its primary categories, persons, places, and average rating. Using this comprehensive understanding, your role is to provide thoughtful and accurate ratings for news on a scale of 0 to 9, ensuring they resonate with the person's preferences. Remain impartial and refrain from introducing any biases in your predictions. You are an impartial and reliable source of news rating predictions for the given individual.*

The intention behind providing the system prompt to the LLM is to clearly define its role, capabilities, and objectives in the context of delivering personalized news recommendations. The user description is appended to the system prompt, along with the news articles the user has seen and rated. An example of a complete prompt is provided in Appendix B.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

2.2.6. Training and Evaluation

Extensive trials were conducted with different system configurations to identify the most effective setups. To account for the vast number of hyper-parameters, Optuna [13], a hyper-parameter optimization framework, was utilized. The metric used for selecting optimal configurations was the mean reward. Detailing the materials and methods used in this study provides a comprehensive guide for replicating and building upon our results. This approach ensures transparency and facilitates future research in news recommender systems.

3. Results

In this study, we aimed to achieve three primary objectives: firstly, to extend the SUBER framework to be compatible with the MIND dataset; secondly, to integrate and evaluate LLMs into a reinforcement learning environment; and third, to enhance the effectiveness of news recommendations.

The approach to enhancing recommender systems using LLMs is new and innovative, making it crucial to understand the underlying code thoroughly. We successfully applied the SUBER technique to the MIND dataset, demonstrating its potential. Although the SUBER framework mentions extensibility, its implementation can be improved. We found it necessary to duplicate Python modules that could have been more abstract and reusable. We believe the framework could be rapidly improved with community interest and support to become more user-friendly and extensible.

LLMs were used in two primary ways in the original work and in our study: first, to generate user descriptions and their news interests, as previously discussed, and second, as the source of truth in rewards-shaping. The recommender model provides a rating that is compared with the recommendation from the integrated LLM. This comparison is used to shape the rewards. The weights and biases of the actor's policy and critic's value networks in the A2C network are adjusted through many iterations.

In our experiments, we consistently received positive rewards, with the training process stabilizing over time. In other words, the networks were trending upward. However, we could not conclusively determine if the models improved significantly because we could not compare the model to any baseline or existing models due to time constraints. Therefore, while the initial results are promising, further work is needed to compare the approach's effectiveness comprehensively.

4. Discussion

Our study extended the SUBER framework to work with the MIND dataset, evaluated a variety of LLMs, and aimed to enhance news recommendations. The results demonstrated that SUBER can be adapted to the MIND dataset, and LLMs can effectively simulate user behavior and shape rewards.

The successful adaptation of SUBER to the MIND dataset indicates its versatility. Our findings align with the work in [1] which highlighted the framework's utility in movie and book recommendations. Extending this to the news domain showed that simulated user behavior can be incorporated into recommender systems.

Given detailed user descriptions, LLMs in RL can simulate user behavior effectively. LLMs also showed that consistent positive rewards and stable training were achieved, suggesting that the reinforcement learning models were improving.

Our study had limitations, primarily the inability to compare the model with baselines due to time constraints, which prevented conclusive determinations of improvement. Additionally, while useful, synthetic data may not capture the complexity of actual user behavior fully.

5. Future Work

Future work should compare the SUBER framework with existing baseline models, such as those produced at SCADS, to see its effectiveness. This includes benchmarking against traditional

 doi:10.20944/preprints202504.0155.v1

recommender systems and the latest approaches in different areas. These comparisons will help us understand how much SUBER improves performance and where it can improve. Metrics like precision, recall, F1-score, and user engagement can be used for evaluation.

While LLMs have shown promise, experimenting with sophisticated neural network architectures like transformers or graph neural networks could capture user behavior patterns more effectively. Techniques like meta-learning or transfer learning also improve adaptability.

Reinforcement learning has been promoted to balance exploration and exploitation of user interests, potentially combating feedback loops. Investigating whether the RL model can effectively serve this purpose will be an important area of future research.

Testing the SUBER framework with actual user data in real-world settings is necessary to uncover any issues that synthetic data might not reveal. Collaborating with industry partners or using publicly available datasets with real user interactions can enable this testing.

6. Conclusions

Our study successfully extended the SUBER framework to the MIND dataset, integrated LLMs into reinforcement learning, and demonstrated the potential of simulated user behavior in enhancing news recommendations. While limited by time constraints, the initial results are promising and suggest valuable directions for future research. We can further validate and improve the SUBER framework by addressing the outlined research directions and establishing a way forward for more effective and personalized recommender systems across.

Acknowledgments: The author would like to thank the Laboratory for Analytic Sciences (LAS) at North Carolina State University for providing the resources and support necessary to conduct this research.

Conflicts of Interest: The author declares no conflict of interest.

Disclaimer: This material is based upon work done, in whole or in part, in coordination with the Department of Defense (DoD). Any opinions, findings, conclusions, or recommen- dations expressed in this material are those of the author(s) and do not necessarily reflect the views of the DoD and/or any agency or entity of the United States Government.

Appendix A. User Generation Prompt

The following illustrates the prompt constructed and provided to the GPT4o model, resulting in the generation of news enthusiast and their descriptions.

Prompt provided to GPT4o through the API

Generate a persona interested in news.

Name: Generate a unique first and last name for a persona interested in news. Age: 46

Gender: Generate a gender (M/F) for a persona interested in news. Primary News Interest: Sports

Secondary News Interest: Food and Cooking Job: Web Developer

Description: Provide a detailed description of their news consumption habits, secondary interests, and reasons for following news.

Example Result returned by GPT4o

name: David Harrison age: 46

gender: Male primary_news_interest: Sports

secondary_news_interest: Food and Cooking job: Web Developer

description: David Harrison is a web developer by profession and a sports enthusiast by heart. He is the kind of person who wakes up to the smell of fresh coffee and the rustling sound of morning newspapers. David spends about an hour each morning reading through the sports section, catching up on the latest news about football, basketball, and tennis, his favorite sports. He loves to analyze the strategies used by different teams and often engages in online debates and discussions about the same.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

Aside from that, David has a secondary interest in food and cooking. He enjoys reading articles and blogs about new recipes, cooking techniques, and emerging food trends. He usually dedicates his weekends to experimenting with these new recipes in his kitchen. David follows news primarily to stay updated with the latest happenings around the world and to fuel his passion for sports

and cooking. He believes that being informed helps him make better decisions, both personally and professionally.

Appendix B. A Complete Prompt

This is an example of a complete prompt provided to an LLM.

You are a highly sophisticated news rating assistant, equipped with an advanced understanding of human behavior. Your mission is to deliver personalized news recommendations by carefully considering the unique characteristics, tastes, and previously seen news of each individual. When presented with information about a specific news article, you will diligently analyze its primary categories, persons, places, and average rating. Using this comprehensive understanding, your role is to provide thoughtful and accurate ratings for news on a scale of 0 to 9, ensuring they resonate with the person's preferences. Remain impartial and refrain from introducing any biases in your predictions. You are an impartial and reliable source of news rating predictions for the given individual.

Alex is a 25-year-old man, he is Alex is a voracious consumer of news, particularly in the realm of law and justice. As a journalist, he feels a responsibility to stay current on the latest developments in legal matters, court decisions, and societal implications of justice-related issues. He subscribes to various law journals and frequently visits websites like The New York Times, BBC News, and The Guardian for their coverage of legal affairs.

In addition to his primary interest, Alex follows football news closely. Being a lifelong fan of the sport, he enjoys keeping up with his favorite teams and players, as well as the general state of the sport. He checks ESPN and other sports news websites daily and often watches live matches during his free time.

Alex follows the news not only for his job but also due to a deep-rooted interest in current affairs. He believes that staying informed about the world helps him to be a responsible citizen and to understand the context of his own life better. He often engages in discussions about news stories on social media, participates in local town hall meetings, and occasionally writes opinion pieces for his local newspaper.

Alex has previously read the following news articles (in parentheses are the ratings he gave on a scale of 0 to 9):

- "50 Worst Habits For Belly Fat" (1)
- "The Cost of Trump's Aid Freeze in the Trenches of Ukraine's War" (6)
- "I Was An NBA Wife. Here's How It Affected My Mental Health." (3)

Consider the news article entitled "The Brands Queen Elizabeth, Prince Charles, and Prince Philip Swear By". It is described as follows: shop the notebooks, jackets, and more that the royals can't live without.

The news article is categorized as "lifestyle" with a subcategory of "lifestyleroyals". On average, the new article has a click-through rate of 0 and the news article has been read 4 times.

Alex has read the news article, "The Brands Queen Elizabeth, Prince Charles, and Prince Philip Swear By", for the first time.

What can you conclude about Alex's rating for the news article "The Brands Queen Elizabeth, Prince Charles, and Prince Philip Swear By" on a scale of 0 to 9, where 0 represents a low rating and 9 represents a high rating, based on available information and logical reasoning?

Response

Based on Alex's preferences and tastes, I conclude that he will assign a rating of Given Alex's reading habits and previous ratings, it is likely that he would rate the article 'The Brands Queen

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

Elizabeth, Prince Charles, and Prince Philip Swear By' a 3 on a scale of 1 to 10. Despite his primary interest in legal and justice-related news, Alex has demonstrated an appreciation for current affairs and lifestyle topics, albeit to a lesser extent. The article on the Royal

Family's preferred brands offers an exciting glimpse into their personal preferences, which might capture Alex's curiosity to some degree. However, his focus on more impactful news, such as legal matters and football, means that this lifestyle article would only moderately engage him, leading him to rate it a 3.

Appendix C. Optuna Results for the Model, TheBloke/Llama-2-13B-Chat-GPTQ

Trial #	Mean Reward	Embedding Dim	Gamma	Items Retrieval	Learning Rate	LLM Rater	Perturbator	Reward Shaping
0	415.1	256	0.95	most_similar_3_title	0.00001	2Shot_system_our_1_10	greedy	identity
1	431	512	0.8	none	0.0001	2Shot_system_our_1_10	greedy	random_watch
2	392.8	256	0.8	none	0.01	0Shot_system_our	greedy	exp_decay_time
3	409.3	256	0.8	last_3	0.01	0Shot_system_our_1_10	greedy	random_watch
4	347	300	0.9	most_similar_3_title	0.001	2Shot_system_our	none	identity
5	387.4	100	0.99	none	0.001	2Shot_system_our	greedy	random_watch
6	340.3	100	0.99	most_similar_3_title	0.001	2Shot_system_our	greedy	random_watch
7	392.1	256	0.9	most_similar_3_abstract	0.0001	1Shot_invert_system_our	gaussian	random_watch
8	415.6	128	0.95	most_similar_3_title	0.00001	2Shot_system_our_1_10	greedy	exp_decay_time
9	357.1	50	0.99	last_3	0.01	1Shot_system_our	none	exp_decay_time
10	360.4	512	0.8	simple_3	0.0001	2Shot_invert_system_our	gaussian	random_watch
11	427.8	128	0.95	none	0.00001	2Shot_system_our_1_10	greedy	exp_decay_time
12	431.8	128	0.95	none	0.0001	2Shot_system_our_1_10	greedy	exp_decay_time
13	411.8	512	0.95	none	0.0001	1Shot_system_our_1_10	greedy	exp_decay_time
14	431.4	512	0.8	none	0.0001	2Shot_system_our_1_10	none	identity
15	412.1	128	0.8	most_similar_3_abstract	0.0001	2Shot_system_our_1_10	none	identity
16	407.5	50	0.95	simple_3	0.0001	0Shot_system_our_1_10	none	identity
17	385.4	300	0.9	none	0.0001	0Shot_system_our	none	identity
18	368.3	512	0.95	none	0.0001	2Shot_invert_system_our	gaussian	identity
19	410.6	128	0.8	none	0.0001	1Shot_system_our_1_10	none	exp_decay_time
20	377.1	128	0.95	most_similar_3_abstract	0.0001	1Shot_invert_system_our	none	identity
21	436.1	512	0.8	none	0.0001	2Shot_system_our_1_10	greedy	random_watch
22	431.4	512	0.8	none	0.0001	2Shot_system_our_1_10	greedy	random_watch
23	430.7	512	0.8	none	0.0001	2Shot_system_our_1_10	gaussian	exp_decay_time
24	394	512	0.8	none	0.0001	1Shot_system_our	greedy	exp_decay_time
25	407.8	512	0.8	last_3	0.0001	2Shot_system_our_1_10	none	random_watch
26	409.3	50	0.9	simple_3	0.00001	2Shot_system_our_1_10	greedy	identity
27	434.1	300	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
28	428.9	300	0.99	none	0.001	2Shot_system_our_1_10	greedy	random_watch
29	429.2	300	0.99	none	0.001	2Shot_system_our_1_10	greedy	exp_decay_time
30	380	300	0.99	most_similar_3_abstract	0.001	1Shot_system_our	greedy	identity
31	431.5	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
32	426.9	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
33	430.5	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2025 doi:10.20944/preprints202504.0155.v1

34	385.7	100	0.99	none	0.01	0Shot_system_our	greedy	identity
35	405.1	256	0.99	last_3	0.001	0Shot_system_our_1_10	greedy	random_watch
36	430.7	300	0.95	none	0.001	2Shot_system_our_1_10	greedy	identity
37	394.8	100	0.99	most_similar_3_title	0.01	1Shot_invert_system_our	greedy	random_watch
38	389.4	128	0.99	none	0.001	2Shot_system_our	greedy	exp_decay_time
39	354.6	100	0.9	simple_3	0.00001	2Shot_invert_system_our	greedy	identity
40	411.4	256	0.95	most_similar_3_title	0.001	1Shot_system_our_1_10	gaussian	random_watch
41	429.2	512	0.8	none	0.0001	2Shot_system_our_1_10	none	identity
42	432.4	512	0.8	none	0.0001	2Shot_system_our_1_10	greedy	identity
43	432.6	300	0.8	none	0.01	2Shot_system_our_1_10	greedy	identity
44	431.5	300	0.8	none	0.0001	2Shot_system_our_1_10	greedy	identity
45	414	300	0.8	last_3	0.01	2Shot_system_our_1_10	greedy	exp_decay_time
46	390.1	300	0.8	none	0.01	0Shot_system_our	greedy	identity
47	387.1	128	0.8	none	0.01	2Shot_system_our	greedy	random_watch
48	420.8	512	0.8	most_similar_3_title	0.01	0Shot_system_our_1_10	greedy	identity
49	428.1	50	0.95	none	0.00001	2Shot_system_our_1_10	greedy	exp_decay_time
50	409.7	300	0.8	most_similar_3_abstract	0.0001	2Shot_system_our_1_10	gaussian	identity
51	429.8	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
52	430.8	128	0.9	none	0.01	2Shot_system_our_1_10	greedy	identity
53	431.8	512	0.8	none	0.0001	2Shot_system_our_1_10	greedy	identity
54	410.7	512	0.8	none	0.0001	1Shot_invert_system_our	greedy	identity
55	372.1	512	0.8	none	0.0001	2Shot_invert_system_our	greedy	identity
56	411.9	512	0.8	simple_3	0.0001	2Shot_system_our_1_10	greedy	random_watch
57	413.8	512	0.8	none	0.0001	1Shot_system_our_1_10	greedy	exp_decay_time
58	375.9	256	0.95	none	0.0001	1Shot_system_our	gaussian	identity
59	413.2	512	0.8	last_3	0.0001	2Shot_system_our_1_10	greedy	random_watch
60	426.3	128	0.8	none	0.0001	2Shot_system_our_1_10	greedy	exp_decay_time
61	425.8	300	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
62	430.5	512	0.95	none	0.00001	2Shot_system_our_1_10	greedy	identity
63	430.4	50	0.8	none	0.0001	2Shot_system_our_1_10	greedy	identity
64	431.7	300	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
65	412.3	300	0.9	most_similar_3_abstract	0.0001	2Shot_system_our_1_10	greedy	identity
66	392.2	300	0.8	none	0.01	2Shot_system_our	greedy	identity
67	393.4	300	0.95	none	0.001	0Shot_system_our	greedy	identity
68	431.1	300	0.99	none	0.0001	2Shot_system_our_1_10	greedy	exp_decay_time
69	409.9	512	0.8	simple_3	0.0001	0Shot_system_our_1_10	gaussian	identity
70	429.7	128	0.99	none	0.001	2Shot_system_our_1_10	greedy	random_watch
71	432.9	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
72	430.9	300	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
73	431.4	512	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
74	429.2	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
75	385	256	0.99	most_similar_3_title	0.001	1Shot_invert_system_our	none	identity
76	430.3	300	0.95	none	0.0001	2Shot_system_our_1_10	greedy	identity
77	385.7	512	0.8	none	0.001	1Shot_system_our	greedy	identity
78	345.2	128	0.99	last_3	0.01	2Shot_invert_system_our	greedy	random_watch

79	408.1	50	0.9	none	0.00001	1Shot_system_our_1_10	greedy	exp_decay_time
80	413.8	100	0.8	most_similar_3_abstract	0.0001	2Shot_system_our_1_10	greedy	identity
81	429.9	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
82	431.9	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
83	430.7	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
84	429.6	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
85	428.5	300	0.99	none	0.001	2Shot_system_our_1_10	none	identity
86	429.9	100	0.99	none	0.001	2Shot_system_our_1_10	greedy	identity
87	432.6	512	0.8	none	0.01	2Shot_system_our_1_10	greedy	random_watch
88	396.2	512	0.8	simple_3	0.01	0Shot_system_our	greedy	random_watch
89	351.8	512	0.8	most_similar_3_title	0.01	2Shot_system_our	gaussian	random_watch
90	398.6	512	0.8	none	0.01	0Shot_system_our_1_10	greedy	random_watch
91	426.6	512	0.8	none	0.01	2Shot_system_our_1_10	greedy	random_watch
92	429.2	512	0.8	none	0.01	2Shot_system_our_1_10	greedy	random_watch
93	429.8	128	0.95	none	0.0001	2Shot_system_our_1_10	greedy	exp_decay_time
94	432	300	0.8	none	0.001	2Shot_system_our_1_10	greedy	identity
95	432.2	512	0.8	none	0.0001	2Shot_system_our_1_10	greedy	identity
96	432.5	100	0.8	none	0.01	2Shot_system_our_1_10	greedy	random_watch
97	433.1	100	0.8	none	0.01	2Shot_system_our_1_10	greedy	random_watch
98	430.6	100	0.8	none	0.01	2Shot_system_our_1_10	greedy	random_watch
99	379.8	100	0.8	last_3	0.01	1Shot_invert_system_our	greedy	random_watch

References

1. Corecco, N.; Piatti, G.; Lanzendörfer, L.A.; Fan, F.X.; Wattenhofer, R. An LLM-based Recommender System Environment. *arXiv* **2024**, arXiv:2406.01631.
2. Digital News Fact Sheet—Pew Research. Available online: https://www.pewresearch.org/journalism/ (accessed on 19 September 2024).
3. Wu, F.; Qiao, Y.; Chen, J.H.; Wu, C.; Qi, T.; Lian, J.; Liu, D.; Xie, X.; Gao, J.; Wu, W.; et al. MIND: A large-scale dataset for news recommendation. In *Proceedings of the 58th annual meeting of the association for computational linguistics*; 2020, pp.3597–3606.
4. Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; Dormann, N. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *J. Mach. Learn. Res.* **2021**, 22, 1–8.
5. Frantar, E.; Ashkboos, S.; Hoefler, T.; Alistarh, D. GPTQ: Accurate post-training quantization for generative pre-trained transformers. *arXiv* **2022**, arXiv:2210.17323.
6. Mnih, V. Asynchronous Methods for Deep Reinforcement Learning. *arXiv* **2016**, arXiv:1602.01783.
7. Brown, T.B. Language Models Are Few-Shot Learners. *arXiv* **2020**, arXiv:2005.14165.
8. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, 32.
9. Mnih, V. Playing Atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602.
10. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
11. Schulman, J. Trust Region Policy Optimization. *arXiv* **2015**, arXiv:1502.05477.
12. Farooq, A. Hierarchical Reinforcement Learning for Adaptive Text Summarization. *Preprints* **2025**, https://www.preprints.org/manuscript/202503.2300/v1.
13. Shakil, H.; Farooq, A.; Kalita, J. Abstractive Text Summarization: State of the Art, Challenges, and Improvements. *Neurocomputing* **2024**, *603*, 128255. https://doi.org/10.1016/j.neucom.2024.128255.

doi:10.20944/preprints202504.0155.v1

14. GitHub—SUBER-Team/SUBER: This repository accompanies our research paper titled "An LLM-based Recommender System Environment". Available online: https://github.com/SUBER-Team/SUBER, accessed on 19 September 2024.
15. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp.2623–2631.