

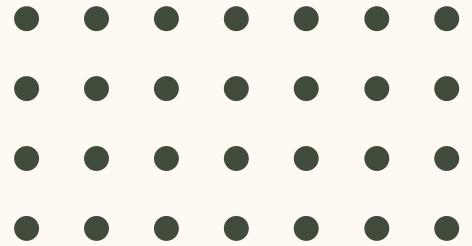


Mahidol University
Faculty of Information
and Communication Technology



大阪大学
THE UNIVERSITY OF OSAKA

Institute of Science and Technology
NAIST



Social Media Reactions to Open Source Promotions: AI-Powered GitHub Projects on Hacker News



P. Meakpaiboonwattana, W. Tarntong, T. Mekratanavorakul,
C. Ragkhitwetsagul, P. Sangaroonsilp, M. Choetkertikul,
T. Sunetnanta

R.G. Kula

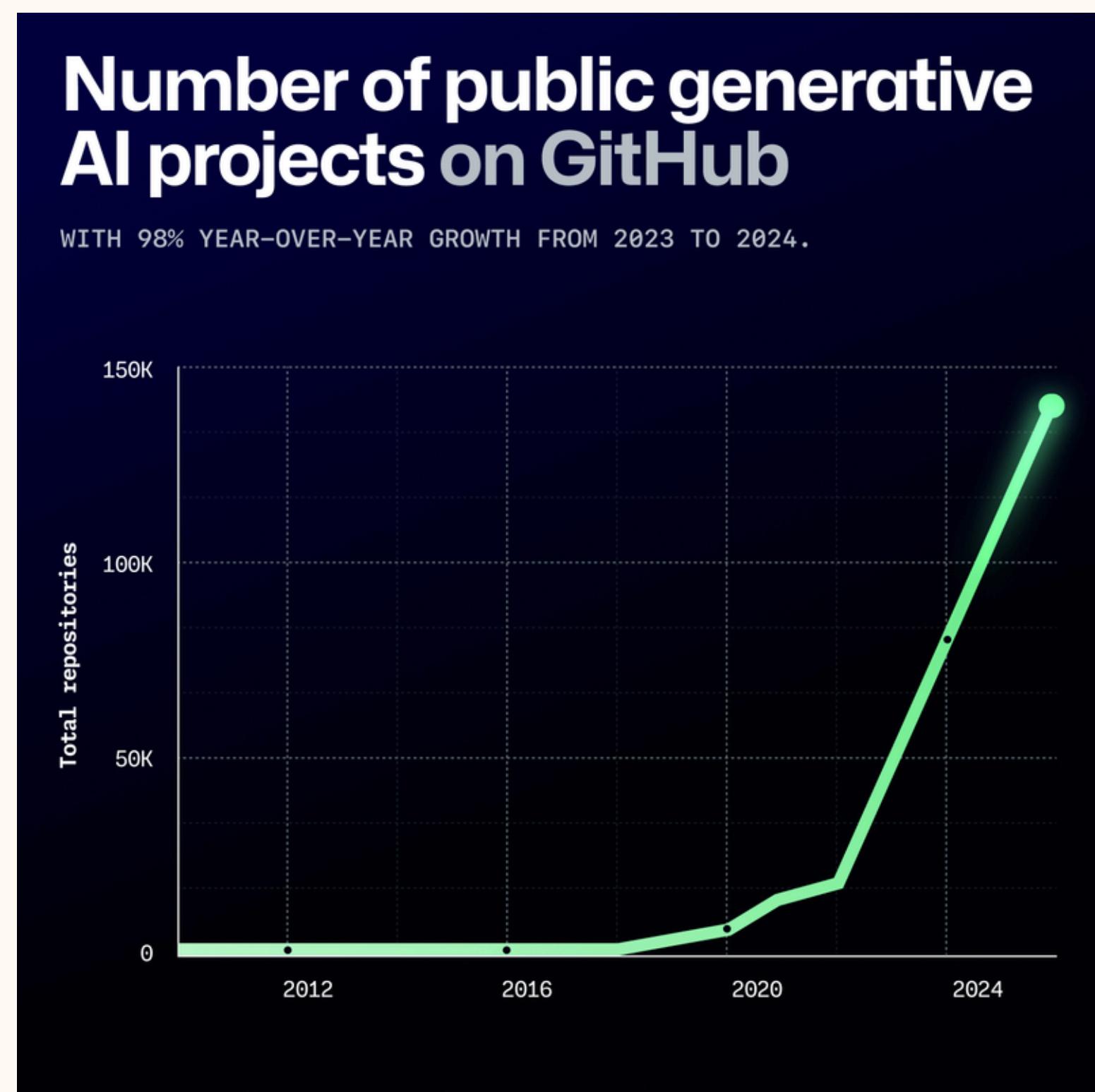
K. Matsumoto

Recent trends in AI/LLM

After the introduction of ChatGPT in November 2022, the number of **AI projects has significantly increased** over the years [1].

Contributions to generative AI projects **surged by 59%** in 2023 [2].

More than **double** the number of open-source **generative AI projects** on GitHub from 2023 to 2024 [3].



[1] B. Marr, "A Short History Of ChatGPT: How We Got To Where We Are Today," Forbes. Accessed: Apr. 28, 2025. [Online]

[2] K. D. Staff GitHub, "Octoverse: The state of open source and rise of AI in 2023," The GitHub Blog. Accessed: Nov. 21, 2024

[3] Staff GitHub "Octoverse: AI leads Python to top language as the number of global developers surges" , The GitHub Blog. Accessed: Apr. 28,2025

Problems

Despite these increasing trends in AI/LLM

- Limited understanding of AI projects promotions on social media platforms
- Reception of AI technologies remains ambiguous among IT practitioners
- Lack comprehensive data on how social media promotions could influence AI/LLM projects on its attraction and development

HackerNews Anatomy

The screenshot shows a Hacker News story page for a project called "Yami". The top navigation bar includes links for new, past, comments, ask, show, jobs, and submit. The main title of the story is "Show HN: Yami - An Open Source Music Player with Spotal Integration" with a GitHub URL: github.com/dever-m. A red box highlights the "Story title". A blue box highlights the "Story URL (GitHub repository URL in this case)". Below the title is a large empty image placeholder. A red button labeled "add comment" is visible. The story has 35 points and was posted 1 hour ago.

Story title

Story URL
(GitHub repository URL in this case)

Comment (depth 1)

This is my high school project btw, i would like some feedback as well as some feature requests it is also available on pypi <https://pypi.org/project/yami-music-player/>

Comment (depth 2)

Having some issues with it, but might be my error ;) The pip-installed (in a venv) version will complain about missing data/theme.json, and then crash. So then I did the git-clone, pip-install variant, that starts. Interface is very minimal. After pointing it to my sshfs mount (~mnt/Audio), it will list directories, but won't find a single file, strange.

Copy the an directory of audiofiles to /tmp, browse there, it works. Very strange.

Comment (depth 3)

I-M 2 hours ago | root | parent | next [-]
if possible can you create an issue about this in github with more details(screen shots)

 **Hacker News** new | past | comments | ask | show | jobs | submit

login

▲ Show HN: OpenLIT – Open-Source LLM Observability with OpenTelemetry (github.com/openlit)

62 points by aman_041 on April 28, 2024 | hide | past | favorite | 22 comments

Hey HN, we're super excited to share something we've been working on: OpenLIT. After an engaging preview that some of you might recall, we are now proudly announcing our first stable release!

What's OpenLIT? Simply put, OpenLIT is an open-source tool designed to make monitoring your Large Language Model (LLM) applications straightforward. It's built on OpenTelemetry, aiming to reduce the complexities that come with observing the behavior and usage of your LLM stack.

Beyond Basic Text Generation: OpenLIT isn't restricted to just text and chatbot outputs. It now includes automatic monitoring capabilities for GPT-4 Vision, DALL·E, and OpenAI Audio. Essentially, we're prepared to assist you with your multi-modal LLM projects all through a single platform and we're not stopping here; more updates and model support are on their way!

Key Features:

- *Instant Alerts:* Offers immediate insights on cost & token usage, in-depth usage analysis, and latency metrics.
- *Comprehensive Coverage:* Supports a range of LLM Providers, Vector DBs, and Frameworks - everything from OpenAI and AnthropicAI to ChromaDB, Pinecone, and LangChain.

Aligned with Standards: OpenLIT follows the OpenTelemetry Semantic Conventions for GenAI, ensuring your monitoring efforts meet the community's best practices.

Wide Integration Compatibility: For those already utilizing observability tools, OpenLIT integrates with various telemetry destinations, including OpenTelemetry Collector, Jaeger, Grafana Cloud, and more, expanding your data's reach and utility.

Getting Started: Check our quickstart guide and explore how OpenLIT can enhance your LLM project monitoring:
<https://docs.openlit.io/latest/quickstart>

We genuinely believe OpenLIT can change the game in how LLM projects are monitored and managed. Feedback from this community could be invaluable as we continue to improve and expand. So, if you have thoughts, suggestions, or questions, we're all ears.

Let's push the boundaries of LLM observability together.

Check out OpenLIT here: <https://github.com/openlit/openlit>

Thanks for checking it out!

H Hacker News new | past | comments | ask | show | jobs | submit

login

▲ Show HN: OpenLIT – Open-Source LLM Observability with OpenTelemetry (github.com/openlit)

62 points by aman_041 on April 28, 2024 | hide | past | favorite | 22 comments

Hey HN, we're super excited to share something we've been working on: OpenLIT. After an engaging preview that some of you might recall, we are now proudly announcing our first stable release!

What's OpenLIT? Simply put, OpenLIT is an open-source tool designed to make monitoring your Large Language Model (LLM) applications straightforward. It's built on OpenTelemetry, aiming to reduce the complexities that come with observing the behavior and usage of your LLM stack.

Beyond Basic Text Generation. OpenLIT isn't restricted to just text and chatbot outputs. It now includes automatic monitoring capabil

Vision
not s

Key

- *In:
Supp
*Alig
comm

*Wid
Open

*Gett
https

We g
invalid

Let's
Check

▲ LLMtools on April 29, 2024 | prev | next [-]

Very cool. QQ: How does this differ from Langtrace (<https://www.langtrace.ai>) / (<https://github.com/Scale3-Labs/langtrace>)?

▲ Areibman on April 29, 2024 | parent | next [-]

Some others have mentioned before, but there are actually several projects working on exactly this:

<https://github.com/Arize-ai/openinference>

<https://github.com/traceloop/opentelemetry>

<https://github.com/Scale3-Labs/langtrace>

▲ patcher99 on April 29, 2024 | root | parent | next [-]

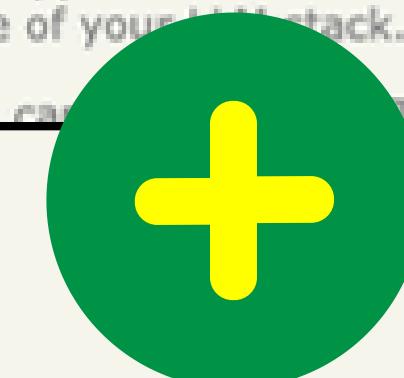
Yup that's correct, Love as more tools come up! We though tried to follow the actual OpenTelemetry's Semantic Convention for GenAI, the others dont.

Additionally none of these libraries "then"(Can check again) seemed to send the cost attribute which was what most of our user's mentioned, So we added that.

▲ kakaly0403 on April 29, 2024 | root | parent | next [-]

Langtrace core maintainer here. We capture input/output/total tokens for both streaming and non streaming across various LLM providers. For streaming we calculate using tiktoken library as OpenAI does not show that with their outputs. Cost is calculated on the client side as it may keep changing. But all you need is the cost table for the llm vendor and you can get the cost for token usage

Thanks for checking it out!



Study of AI-Powered GitHub Projects on Hacker News



**Identify the trend
of each different
AI/LLM discussion
and development
pattern**



**Understand the
developers
community's
receptions and
responses on AI
projects**



**Understand the
impact and
effectiveness of
promoting AI/LLM
project on Hacker
News**

Research Questions



What is the **spread** of AI/LLM projects on HackerNews?

RQ1



What are the **social reactions** to HN GH-AI stories?

RQ2



What are the **changes of activities** in GitHub AI projects after being mentioned in HN?

RQ3

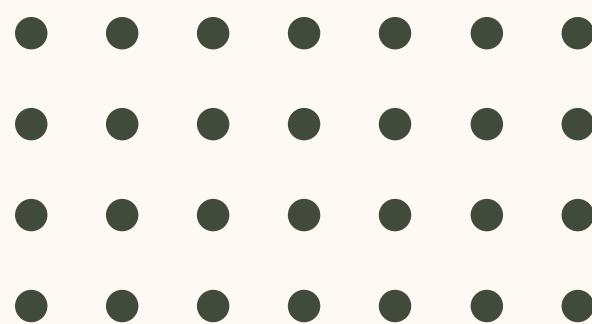
• INTRODUCTION

• BACKGROUND & RELATED WORKS

• RQ1 - SPREAD

• RQ2 - REACTION

• RQ3 - ACTIVITIES

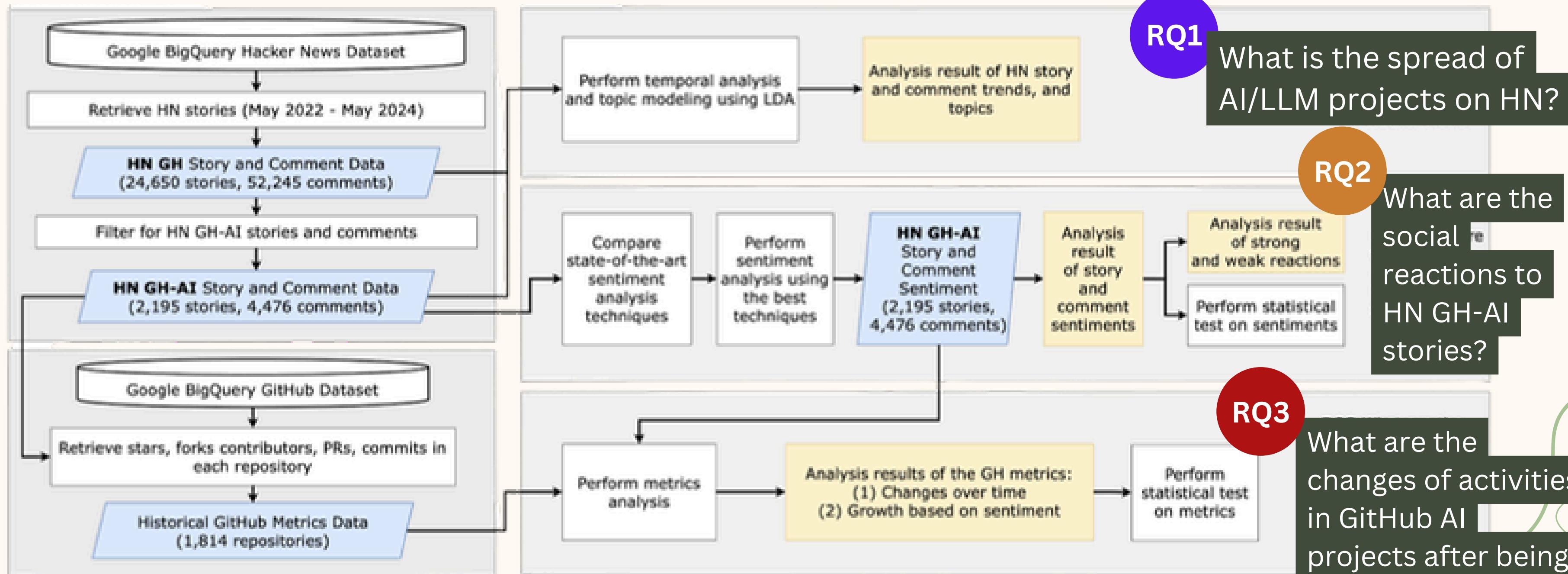


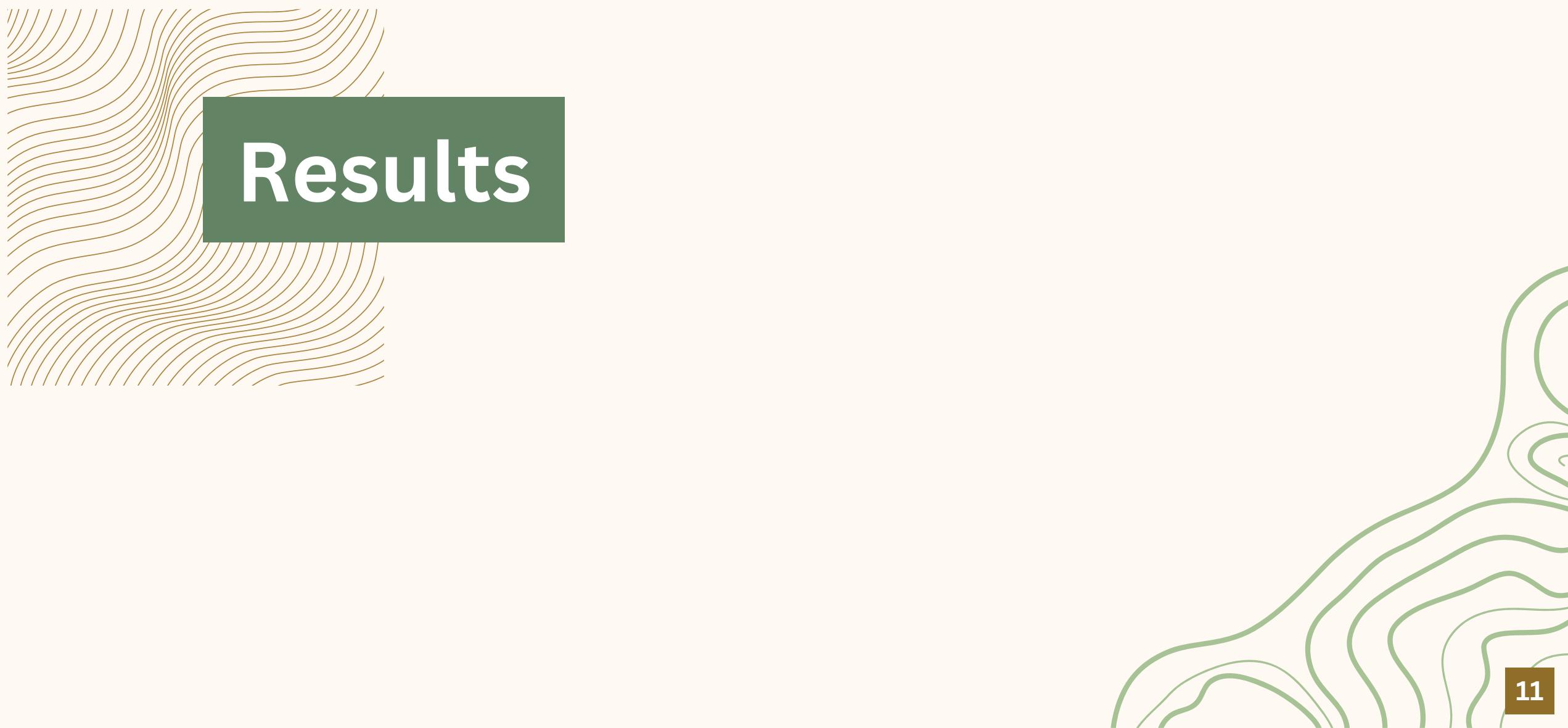
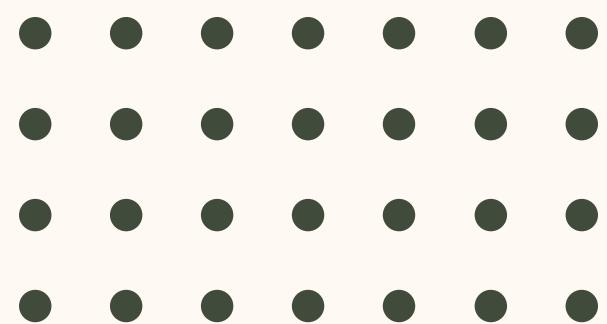
Methodology

The slide features abstract background graphics on the right side, consisting of several sets of thin, wavy lines in gold and green colors, creating a sense of depth and motion.

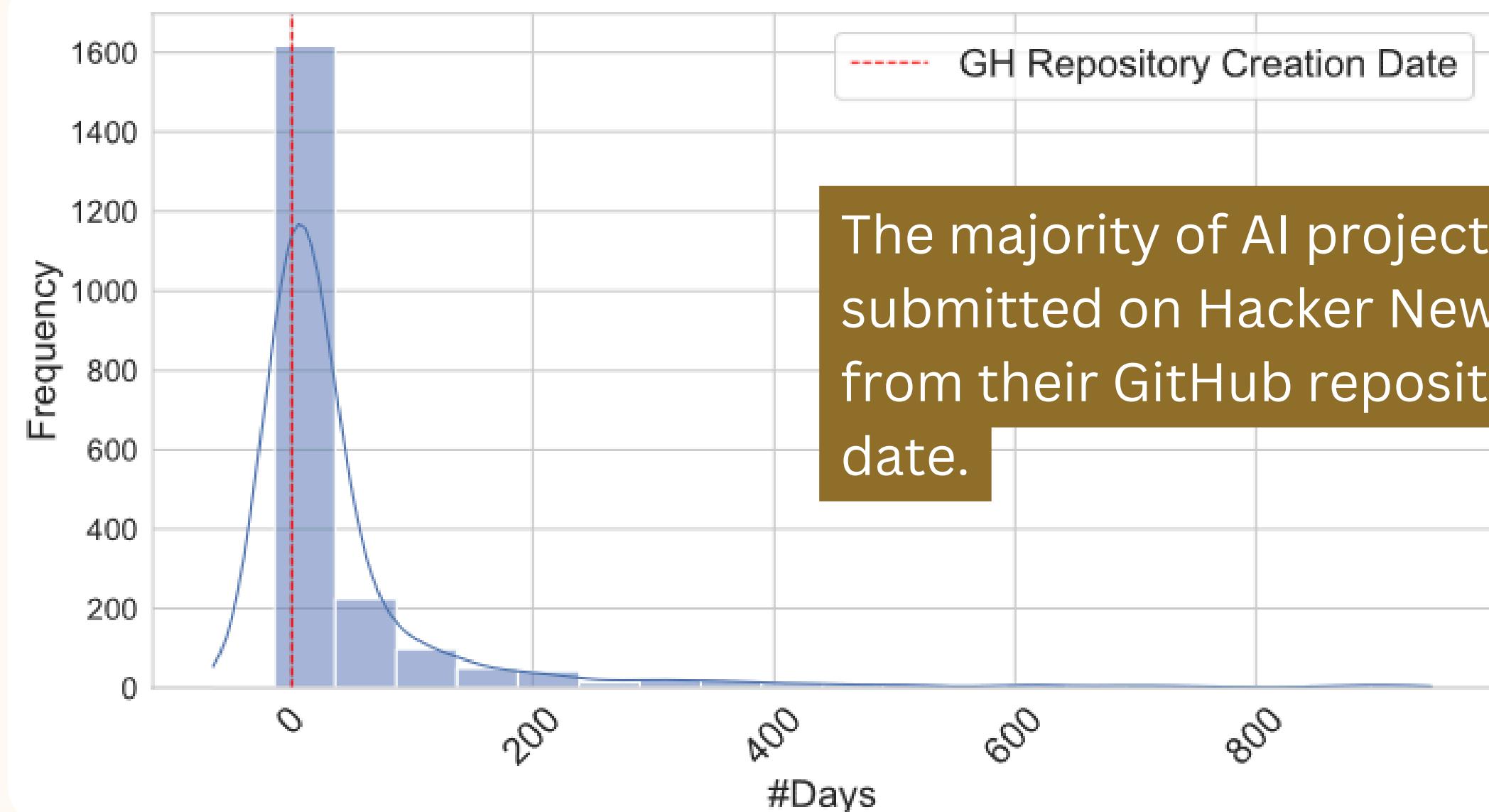
Methodology Overview

Data Collection





Results

RQ1**What is the spread of AI/LLM projects on HN?**

The majority of AI projects were submitted on Hacker News within 1 week from their GitHub repository creation date.

Analysis of the Hacker News Posters and Posting Time

Observation 1: Majority of GitHub AI projects submitted on Hacker News close to the time of their creation on GitHub

RQ1

What is the spread of AI/LLM projects on HN?



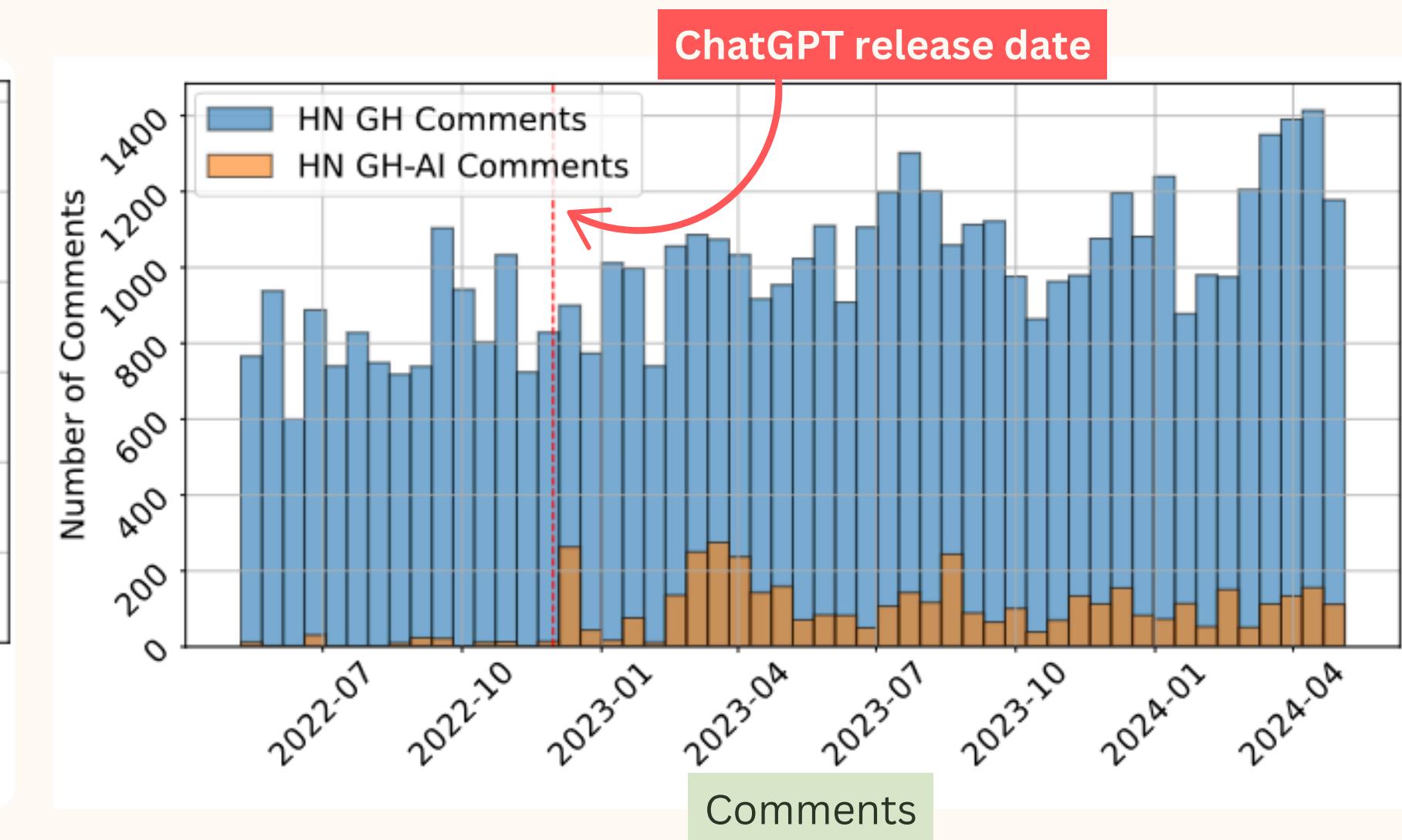
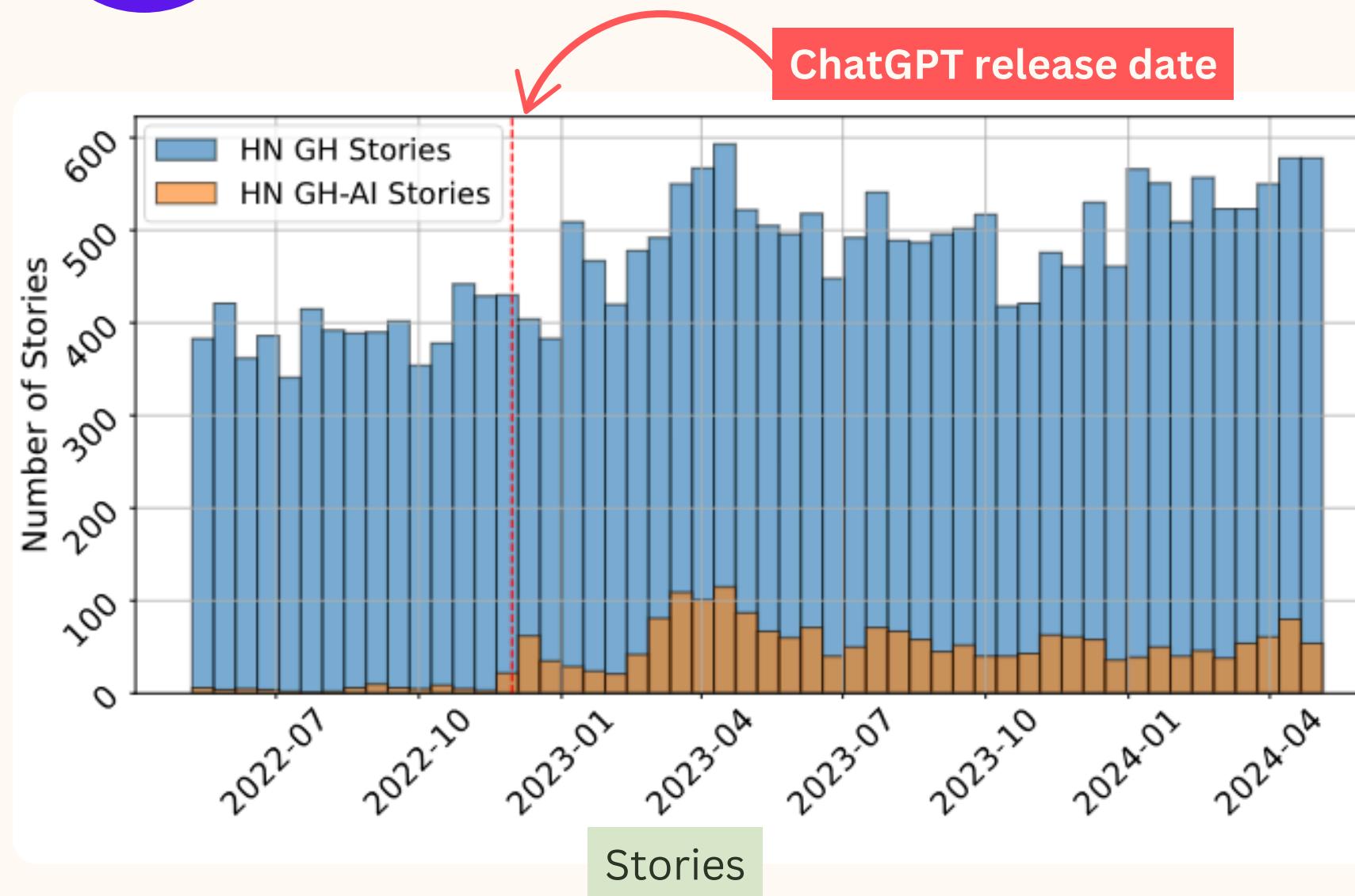
Time to HN submission vs. GH repo creation date

- Recent projects (2023 and 2024)
 - **Submitted on fewer days** after their creation
- Projects created before 2023
 - **Broader range of days** before submission date

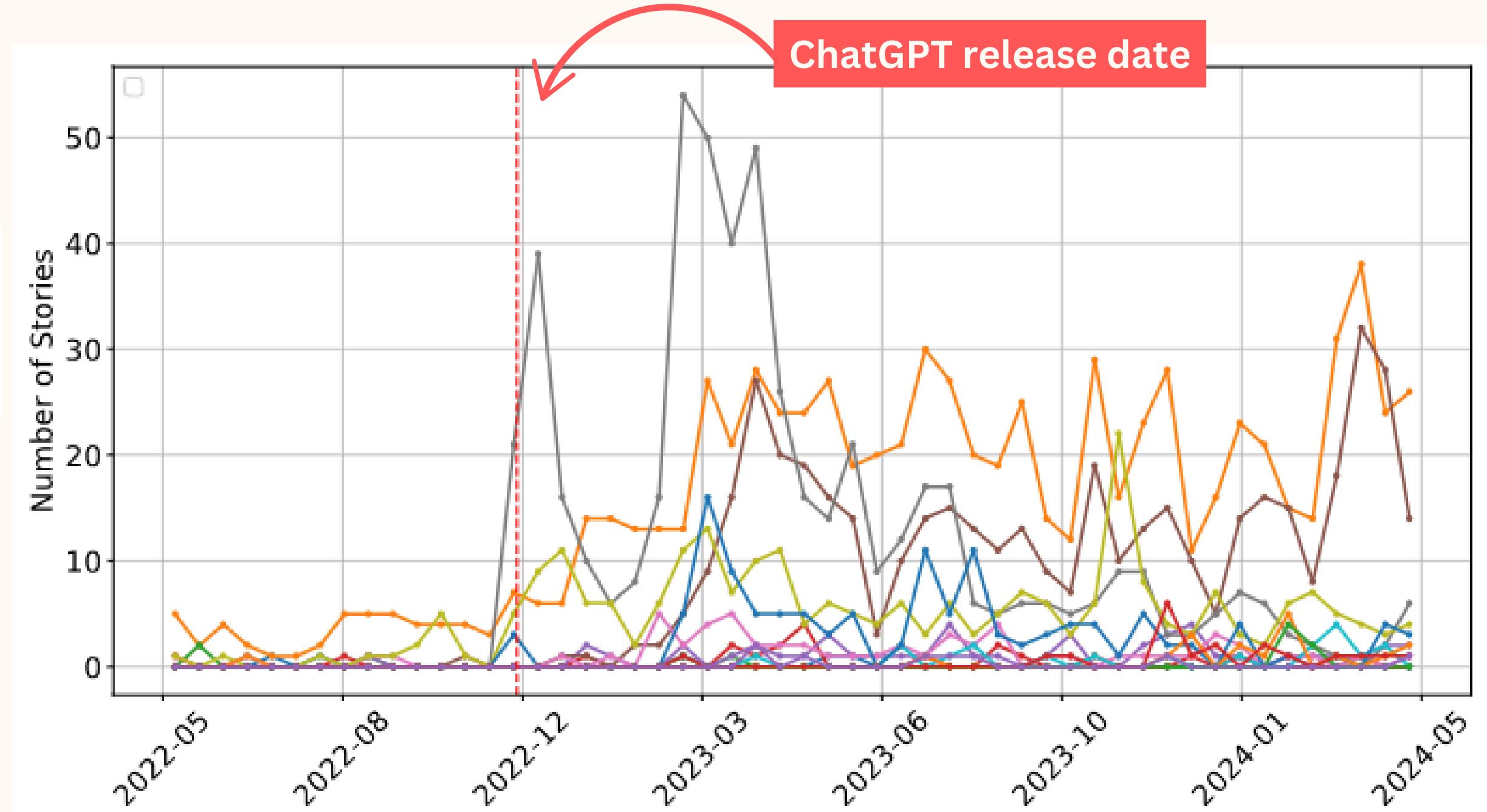
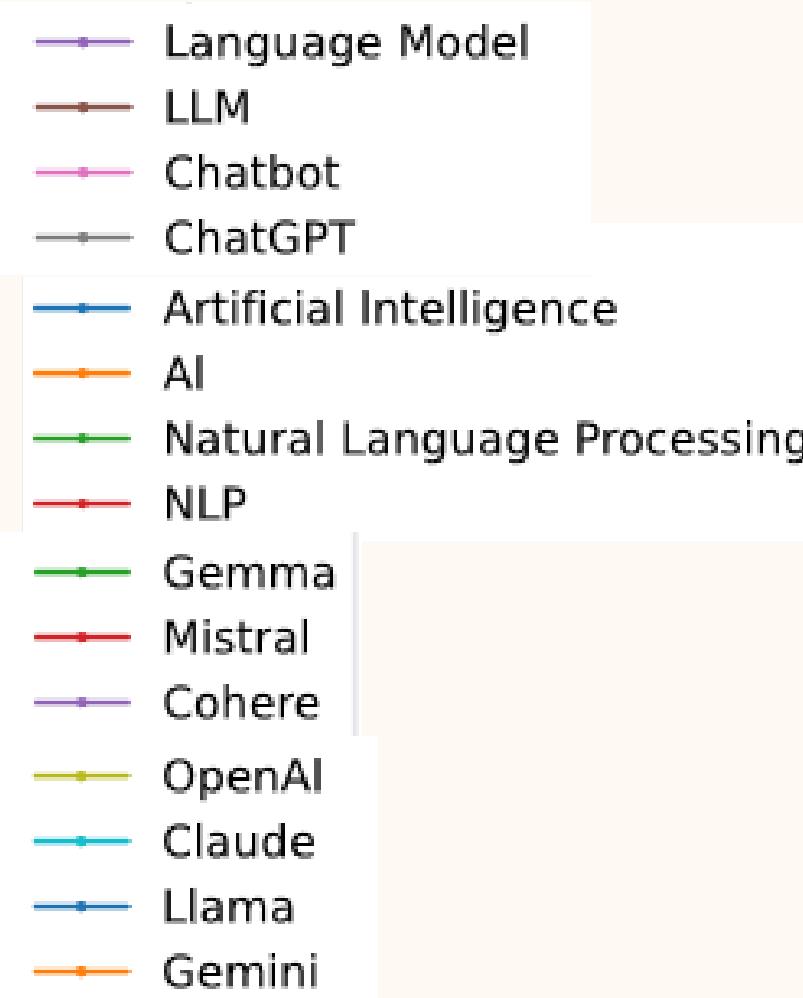
RQ1**What is the spread of AI/LLM projects on HN?****“About” section on Hacker News**

user:	muehuhu
created:	9 months ago
karma:	1
about:	my profile might contain github username

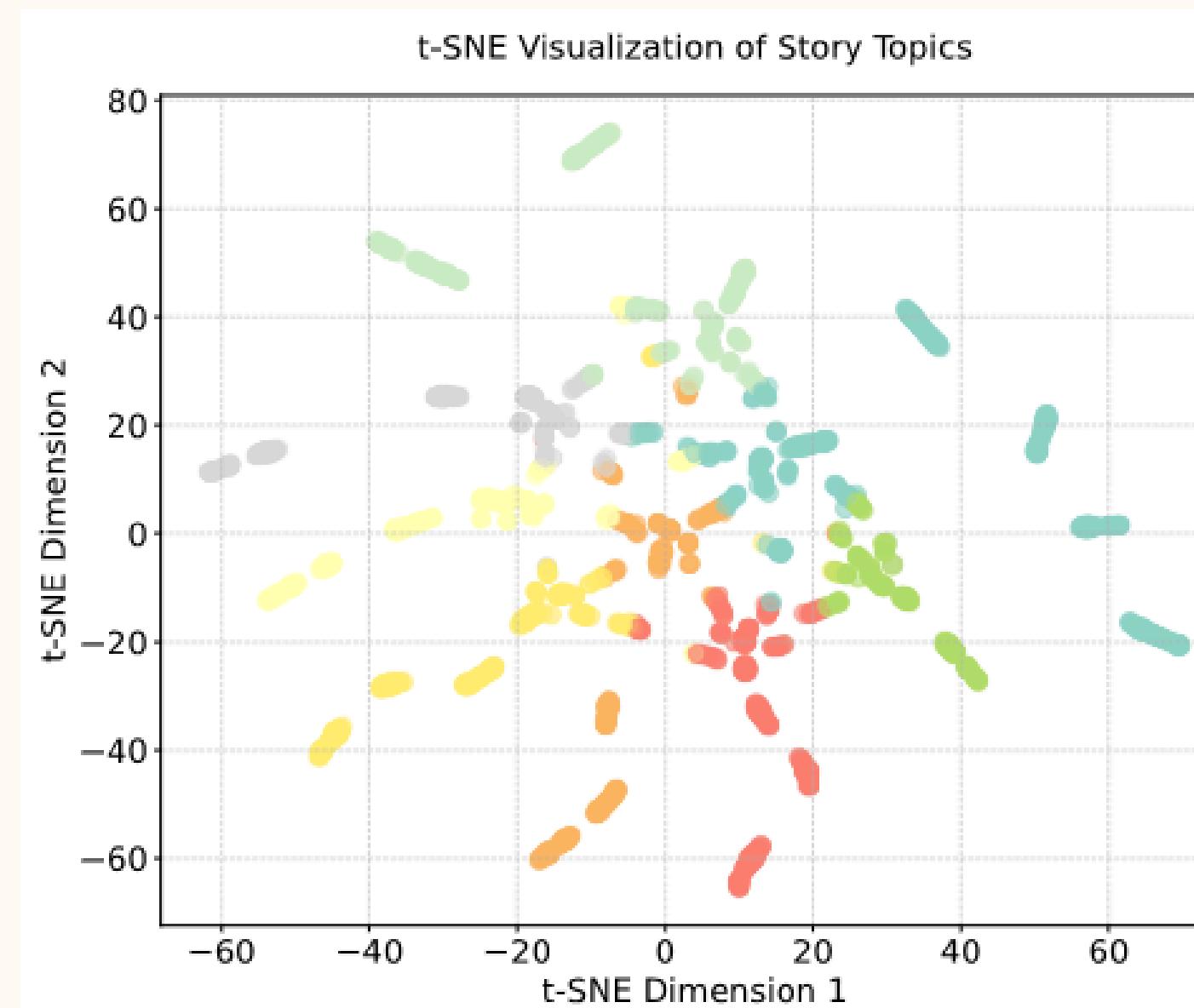
Observation 2: At least 19% of the posters of HN GH-AI stories were contributors to the GitHub project themselves, implying self-promotion

RQ1**What is the spread of AI/LLM projects on HN?**

Observation 3: HN GH-AI stories increased sharply after the ChatGPT release, confirming the impact of the LLMs on AI projects

RQ1**What is the spread of AI/LLM projects on HN?**

Observation 3: HN GH-AI stories increased sharply after the ChatGPT release, confirming the impact of the LLMs on AI projects

RQ1**What is the spread of AI/LLM projects on HN?**

Topic	Top Terms	Interpretation
T1	openai, chatgpt, code, llm, function	Projects using OpenAI's models, e.g., ChatGPT
T2	open, sourc, chatbot, convers, llm	Open-source chatbot projects and conversational models
T3	llm, chatgpt, openai, python, whisper	LLM implementations, including OpenAI tools like Whisper and ChatGPT
T4	chatgpt, plugin, command, use, line	ChatGPT plugins, command-line usage, and integrations
T5	list, ui, project, llm, chatgpt	UI-driven projects and curated lists related to LLMs and ChatGPT
T6	chatgpt, game, commit, automat, llama	AI-powered automation using ChatGPT including games and commit reviews
T7	languag, model, opensourc, llm, framework	Open-source language models, frameworks, and LLM development
T8	chatgpt, termin, llm, gpt, altern	GPT alternatives, terminology, and model evaluations

Observation 4: HN GH-AI stories cover 8 types of AI/LLM applications that range from LLM usage to model evaluations

RQ2 What are the social reactions to HN GH-AI stories?

1.1 Compared SOTA Sentiment Analysis Techniques

3 SOTA pretrained transformer models



Twitter-roBERTa
from CardiffNLP



RoBERTa
from Meta



BERT
from Google

1 LLM



GPT-4o mini
from OpenAI

Fine-tuned using 5 fold cross validation and
Optuna to optimize the hyperparameters

Used **few shot
prompting
techniques**

RQ2 What are the social reactions to HN GH-AI stories?

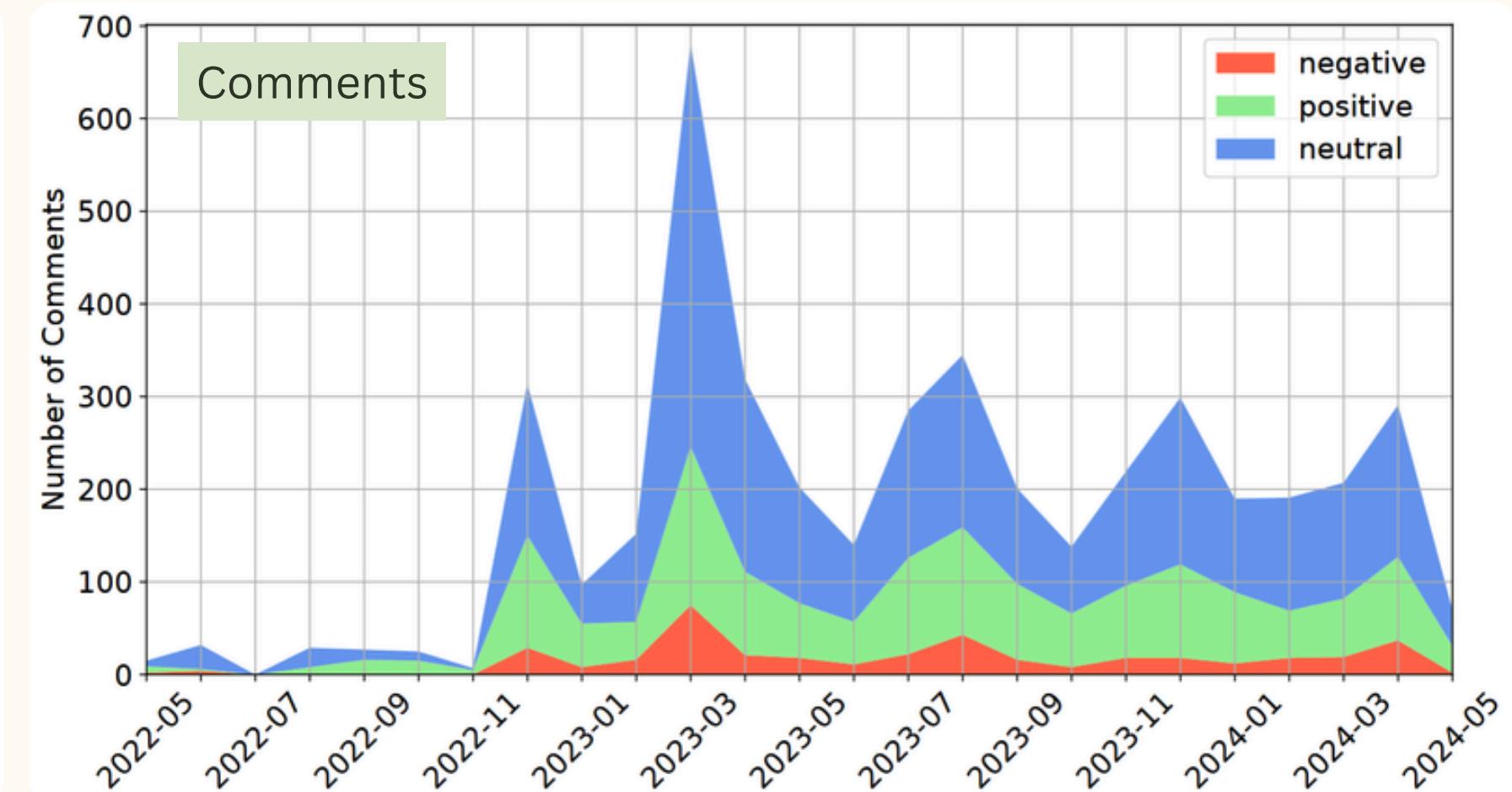
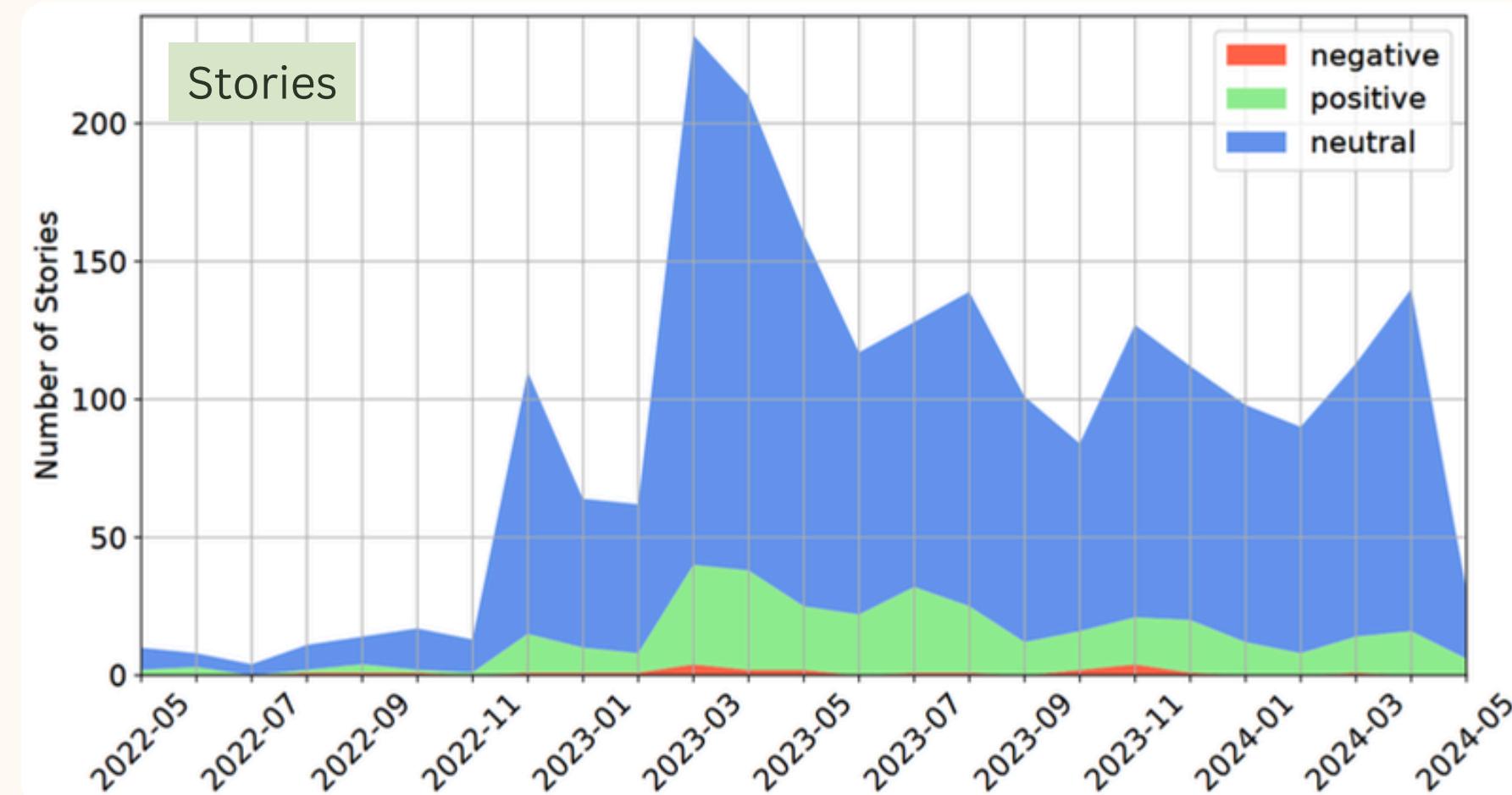
1.1 Compared SOTA Sentiment Analysis Techniques

Model	Stories	Comments
BERT	0.683	0.599
RoBERTa	0.724	0.681
Twitter-RoBERTa	0.733	0.709
GPT-4o mini	0.762	0.763

Select to perform
sentiment
analysis

RQ2 What are the social reactions to HN GH-AI stories?

2.1 Analysis of HN GH-AI Story and Comment Sentiment



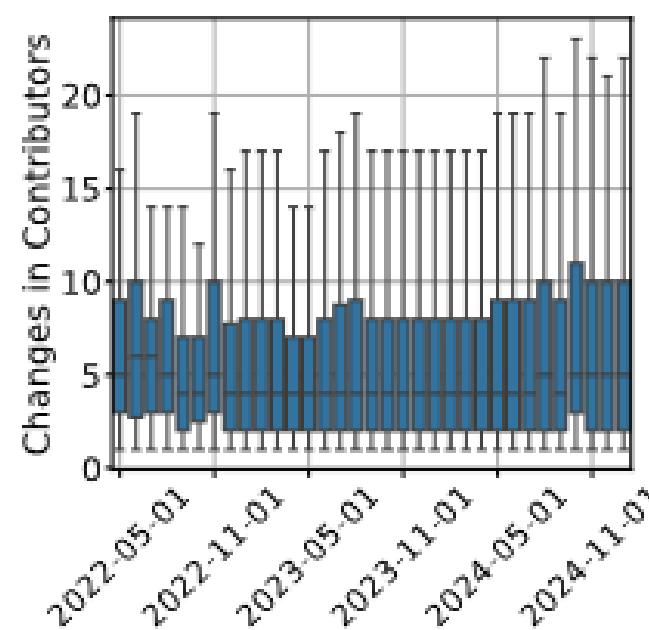
Observation 5: Hacker News community tend to be more positive than negative towards stories related to AI project on GitHub.

RQ2 What are the social reactions to HN GH-AI stories?

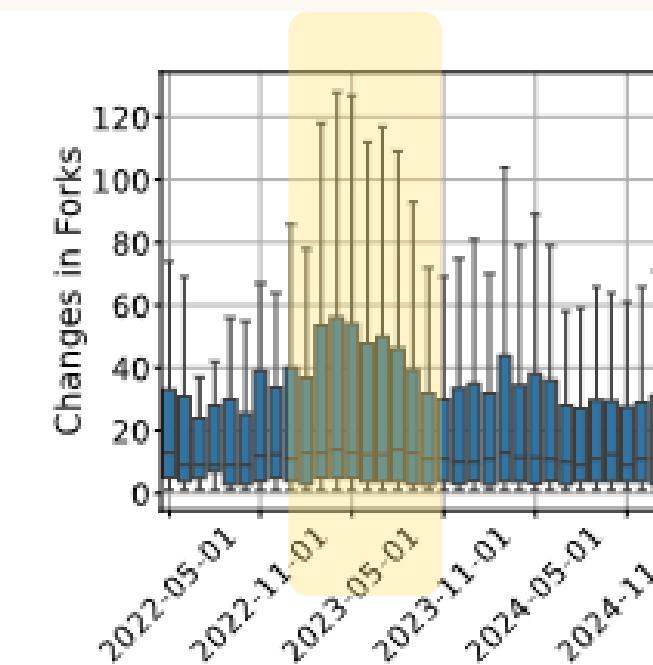
Story	Comment
<p>Show HN: Developer Friendly Natural Language Processing</p>	<p>Huh, interesting. It gets POS tagging correct in a few edge cases I've used to test other frameworks. Bookmarked!</p>
<p>Show HN: Chrome extension to summarize blogs and articles using ChatGPT</p>	<p>Great job, I was doing it manually before! It would be nice if it would be possible to continue chatting with chatGPT after the summary. I always find it interesting to probe chatGPT about the article after summarizing it</p>

RQ3

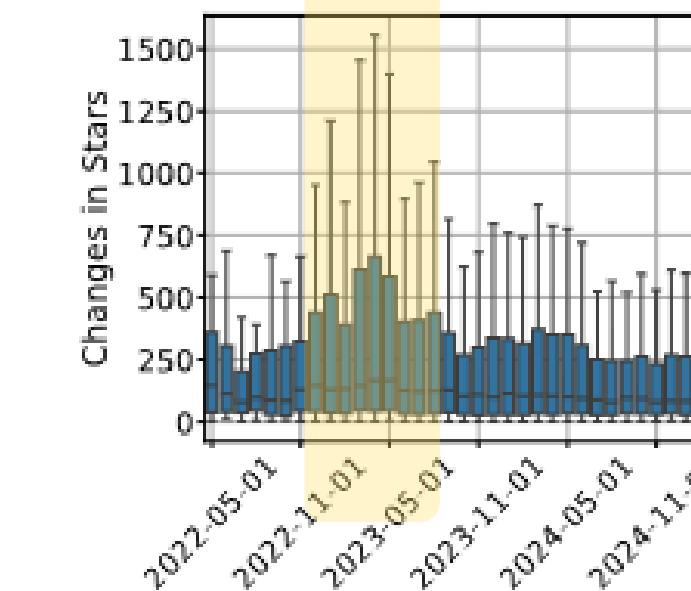
What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?



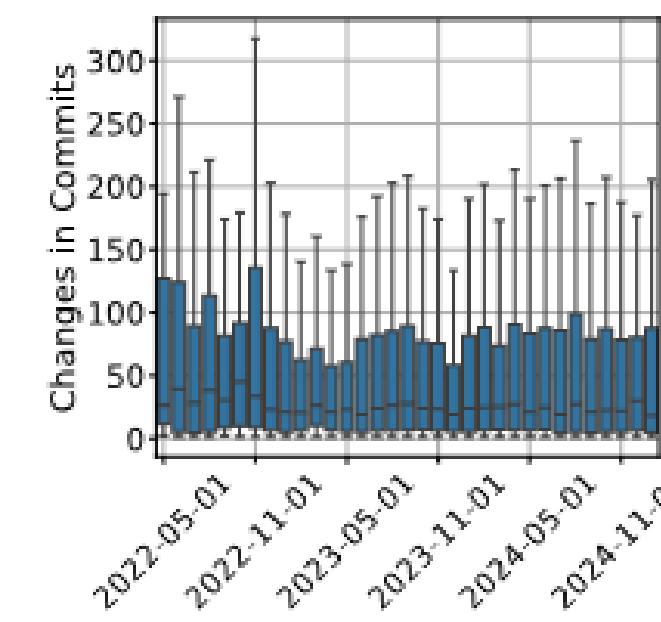
(a) Contributors



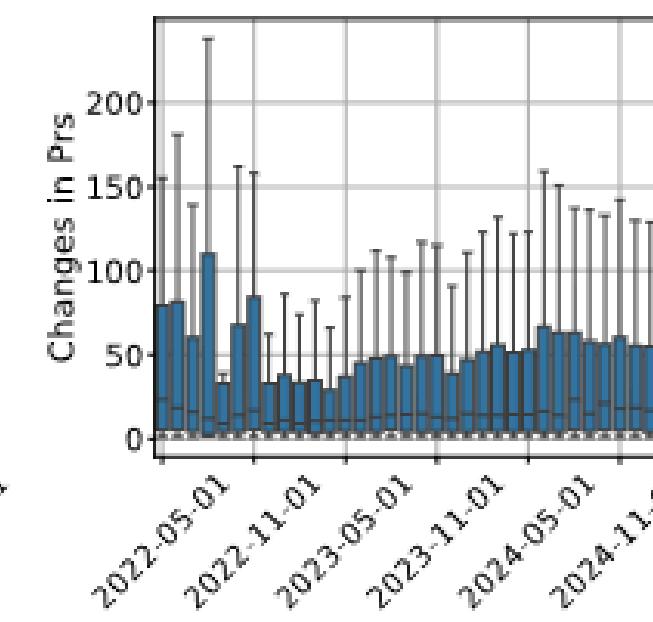
(b) Forks



(c) Stars



(d) Commits

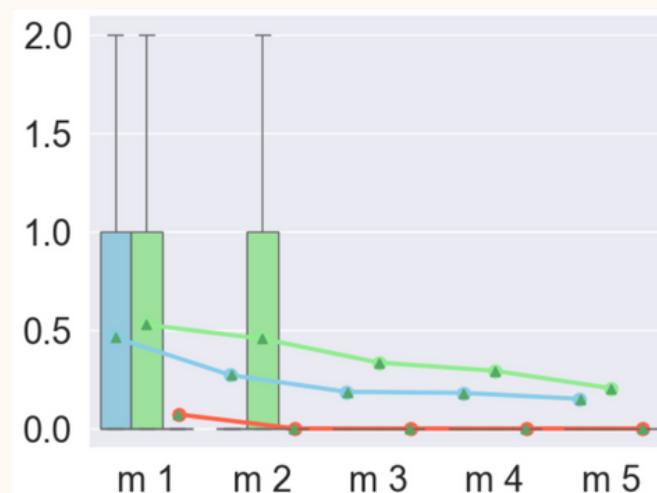


(e) Pull Requests

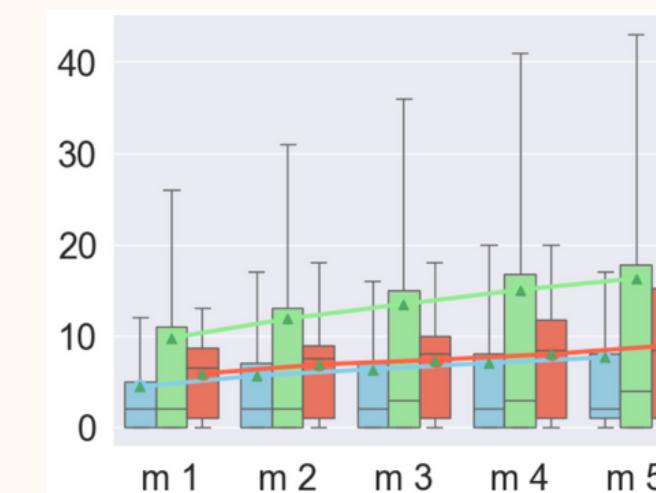
Observation 7: GitHub AI projects on Hacker News received an **increase in activity on stars and forks shortly after the release of ChatGPT.**

RQ3

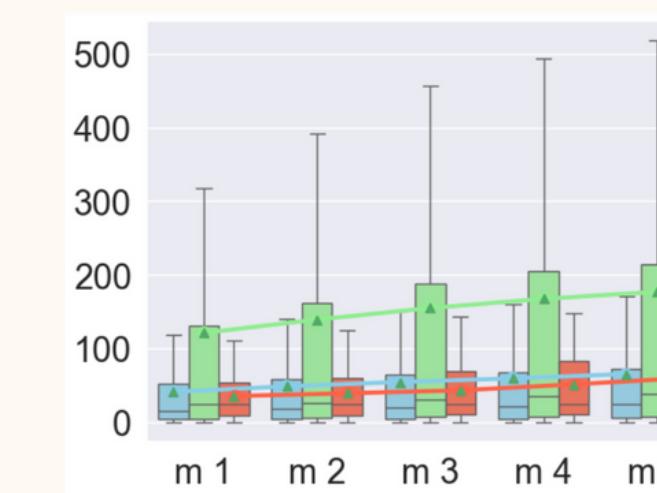
What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?



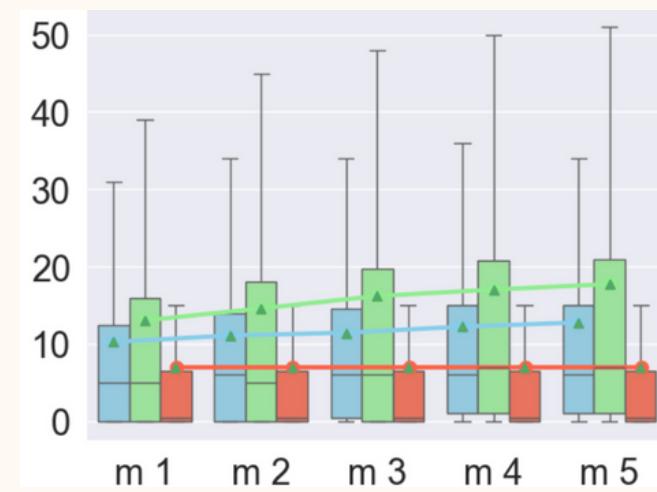
Contributors changes



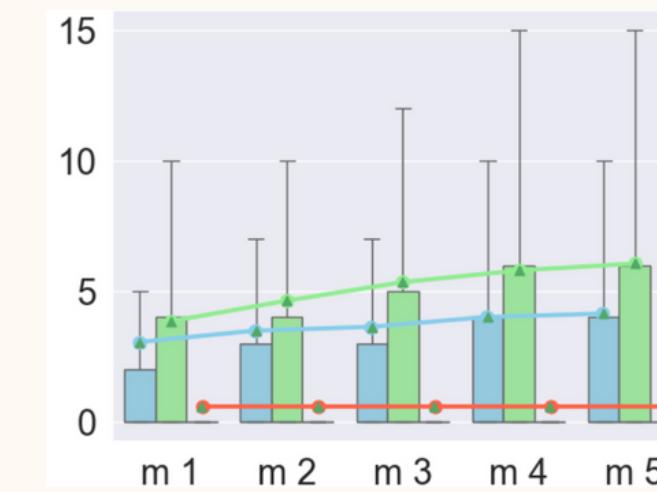
Forks Growth



Stars Growth



Commits Growth



PRs Growth

- : AI projects with **Neutral** sentiment
- : AI projects with **Positive** sentiment
- : AI projects with **Negative** sentiment

Observation 8: the **positive** group generally have more metrics growth or changes than the **negative** group and the **neutral** group after their projects have been submitted on Hacker News

Summary

RQ1

What is the **spread** of AI/LLM projects on HN?

AI developers actively **promoted** their AI GitHub projects on Hacker News

RQ2

What are the social **reactions** to HN GH-AI stories?

These AI GitHub projects usually receive **positive** responses.

RQ3

What are the changes of **activities** in GitHub AI projects after being mentioned in HN?

These projects **gain more popularity** with an increasing number of GitHub *stars* and *forks*, **after** Hacker News submission.

Research Questions



What is the spread of AI/LLM projects on HackerNews?

RQ1



What are the social reactions to HN GH-AI stories?

RQ2

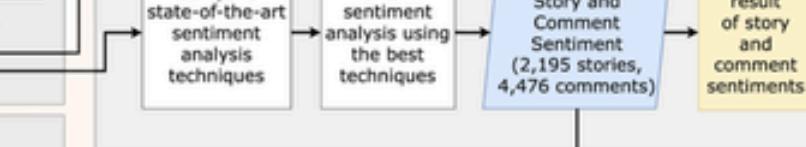
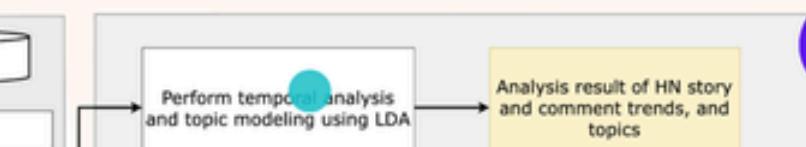
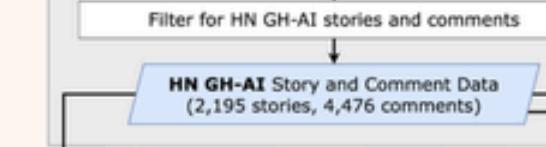
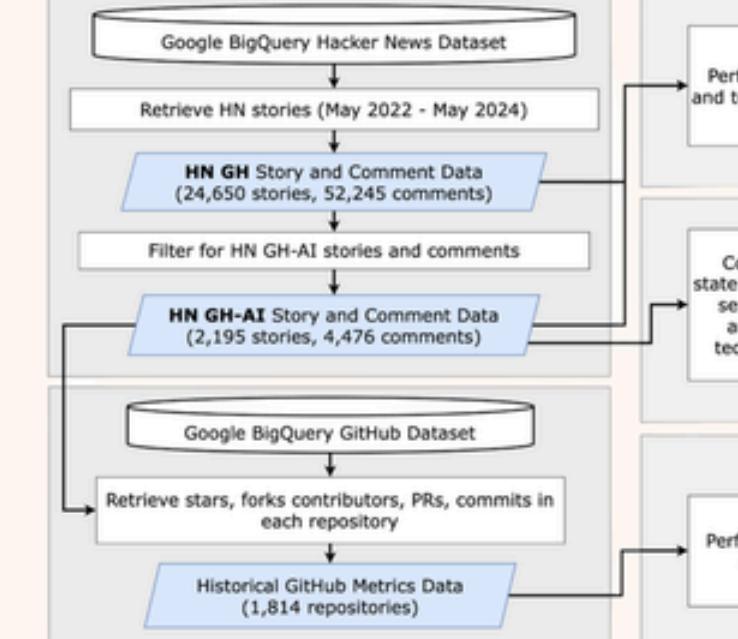


What are the changes of activities in GitHub AI projects after being mentioned in HN?

RQ3

Methodology Overview

Data Collection

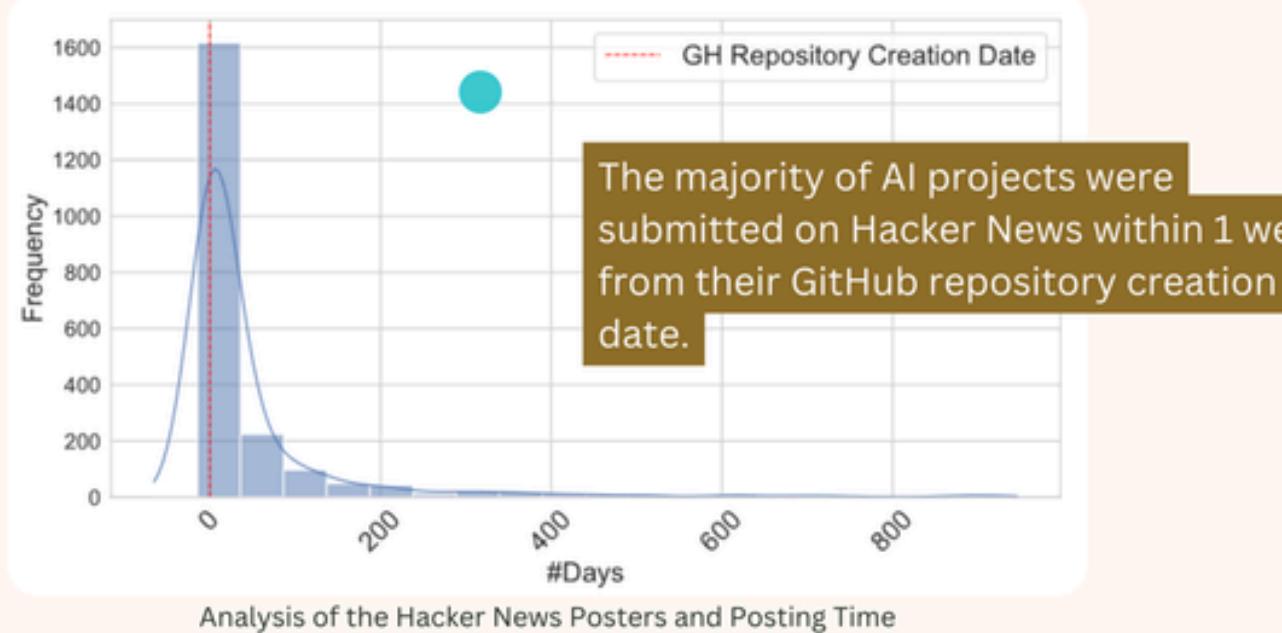


RQ1 What is the spread of AI/LLM projects on HN?

RQ2 What are the social reactions to HN GH-AI stories?

RQ3 What are the changes of activities in GitHub AI projects after being mentioned in HN?

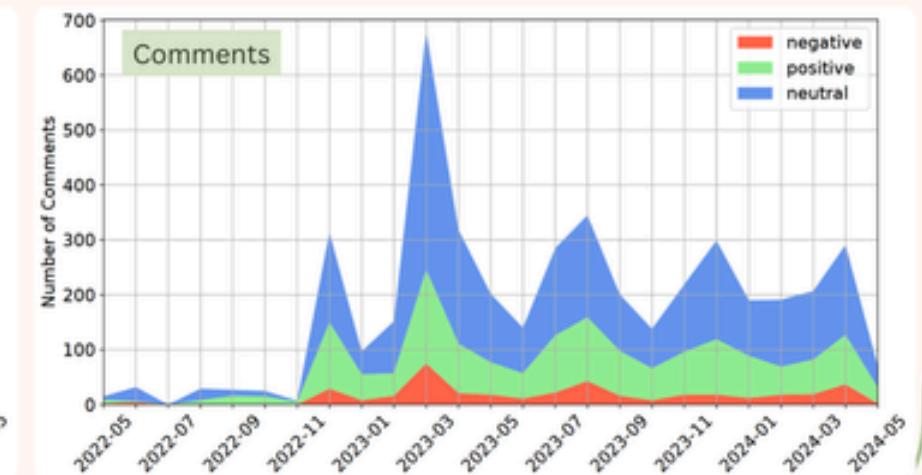
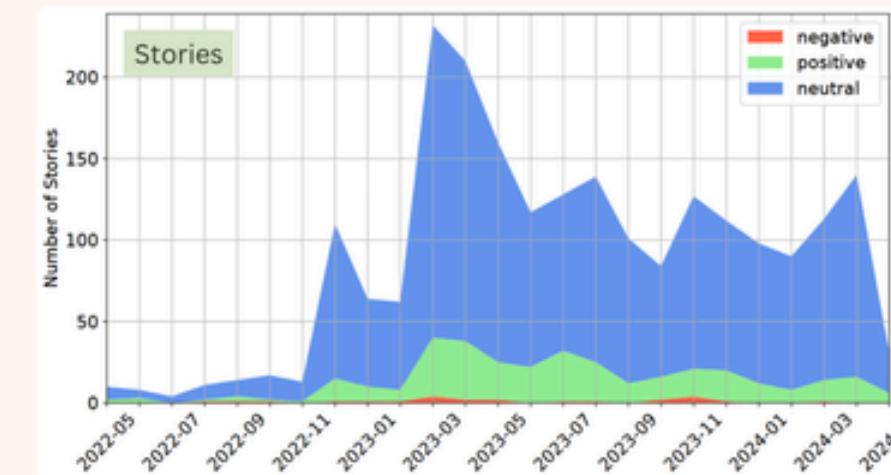
RQ1 What is the spread of AI/LLM projects on HN?



Observation 1: Majority of GitHub AI projects submitted on Hacker News close to the time of their creation on GitHub

RQ2 What are the social reactions to HN GH-AI stories?

2.1 Analysis of HN GH-AI Story and Comment Sentiment



Observation 5: Hacker News community tend to be more positive than negative towards stories related to AI project on GitHub.

RQ3

What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

Causal Effects of Hacker News Submissions:

Answering whether submission on Hacker News yield increased metrics

(1) Validate parallel trend assumption

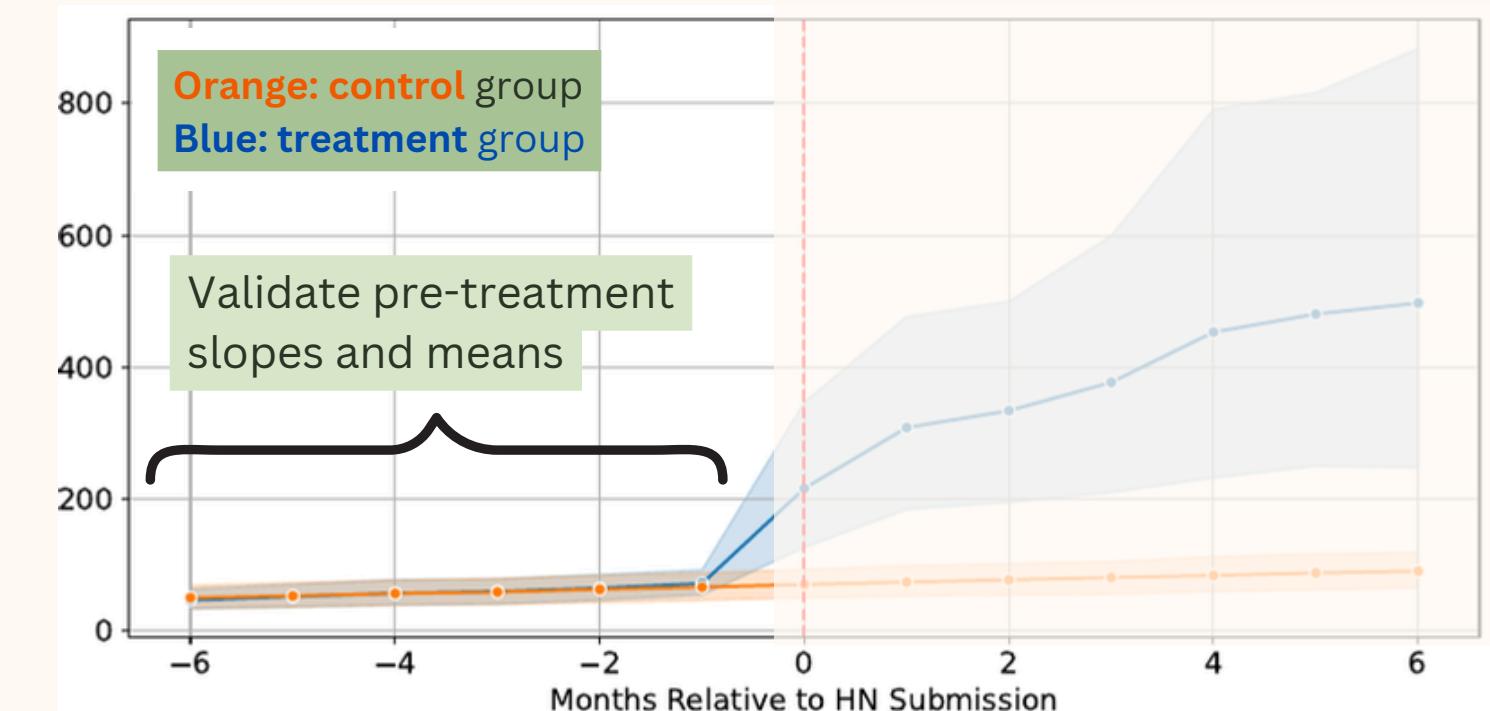
Further verify PTA using **pre-treatment slopes and pre-treatment means (balancing test)**

(2) Perform Difference-in-Difference (DiD)

Aligning HN submission months using month 0 as a **reference point between treatment and control groups**

- Staggered DiD technique

Both results must **agree** to consider the metric's increase as a **strong causal evidence**.



DiD Model

$$Y_{it} = \alpha + \beta_1 \cdot \text{post_treatment}_{it} + \beta_2 \cdot \text{treatment}_{it} + \beta_3 \cdot (\text{post_treatment}_{it} \times \text{treatment}_{it}) + \epsilon_{it}$$

Capture short-term “momentum effects”

Lagged DiD Model

$$Y_{it} = \alpha + \beta_1 \cdot \text{post_treatment}_{it} + \beta_2 \cdot \text{treatment}_{it} + \beta_3 \cdot (\text{post_treatment}_{it} \times \text{treatment}_{it}) + \gamma \cdot Y_{i,t-1} + \epsilon_{it}$$

Takes into account the previous iteration (month)

Isolates only the persistent long-term effect

(Robustness check)

RQ3 What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

Causal Effects of Hacker News Submissions:

(1) Validate Parallel Trend Assumption

Matching final result	
Metric	Number of Project Pairs
Cumulative Commits	35
Cumulative Forks	73
Cumulative PRs	41
Cumulative Stars	43
Monthly Active Contributors	93

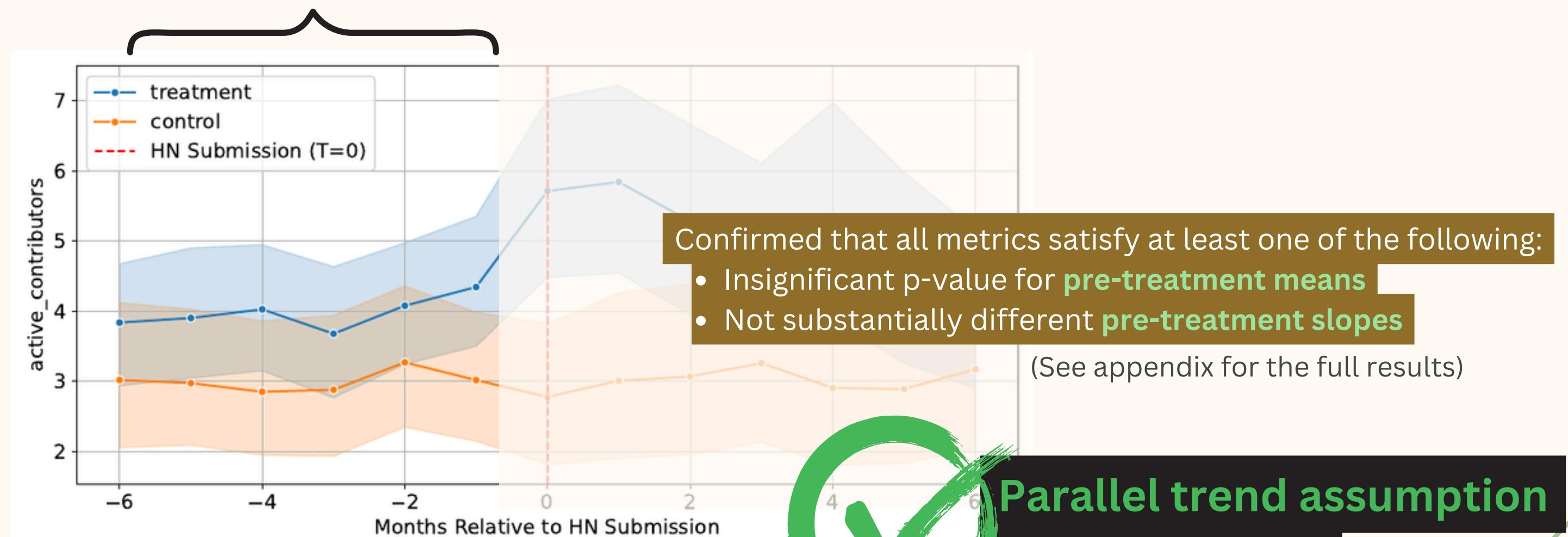
- From a pool of
 - 163 treatment repositories (**HN**)
 - 2,147 candidate control repositories (**non-HN**)

RQ3 What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

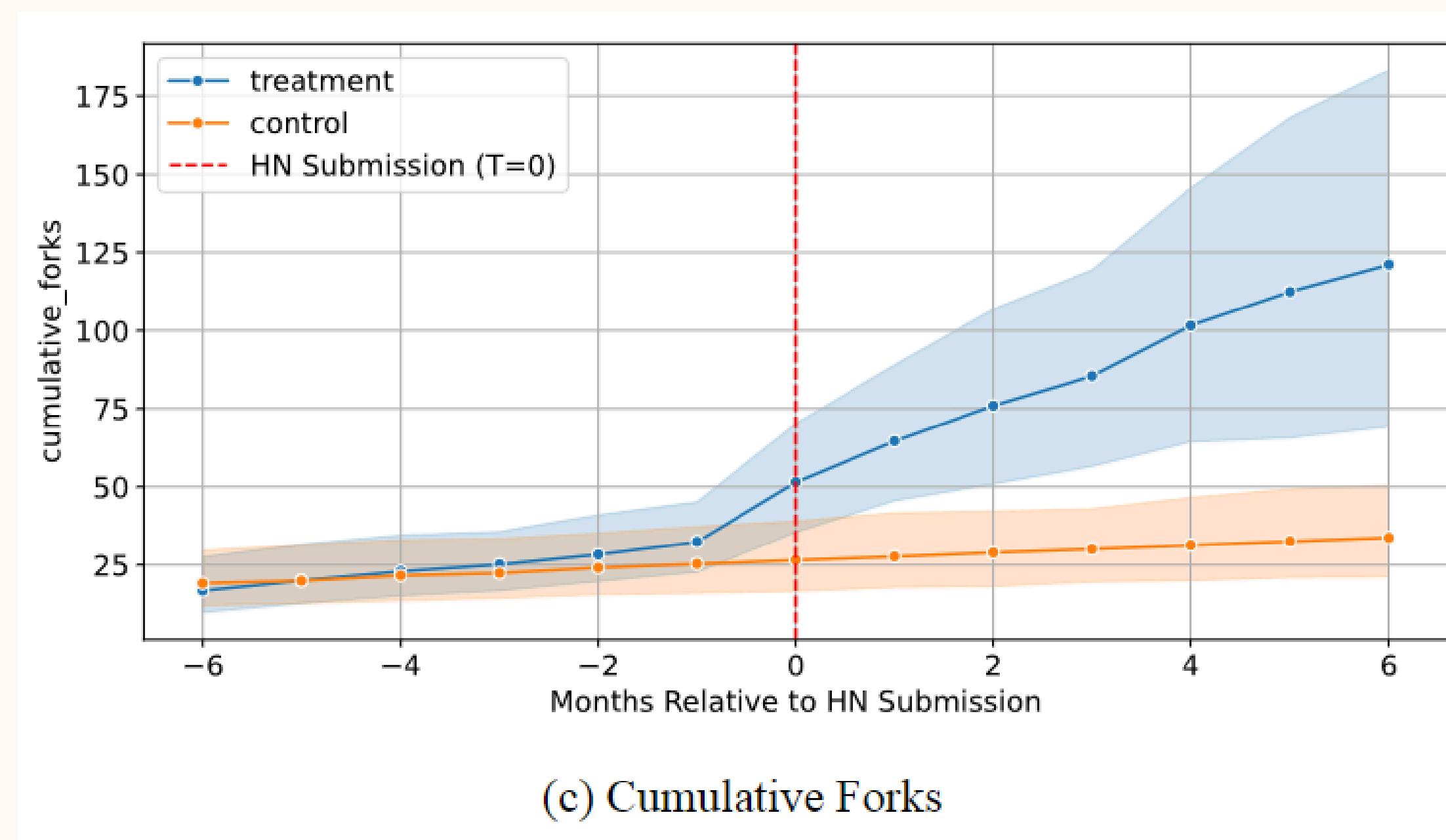
Causal Effects of Hacker News Submissions:

(1) Validate Parallel Trend Assumption

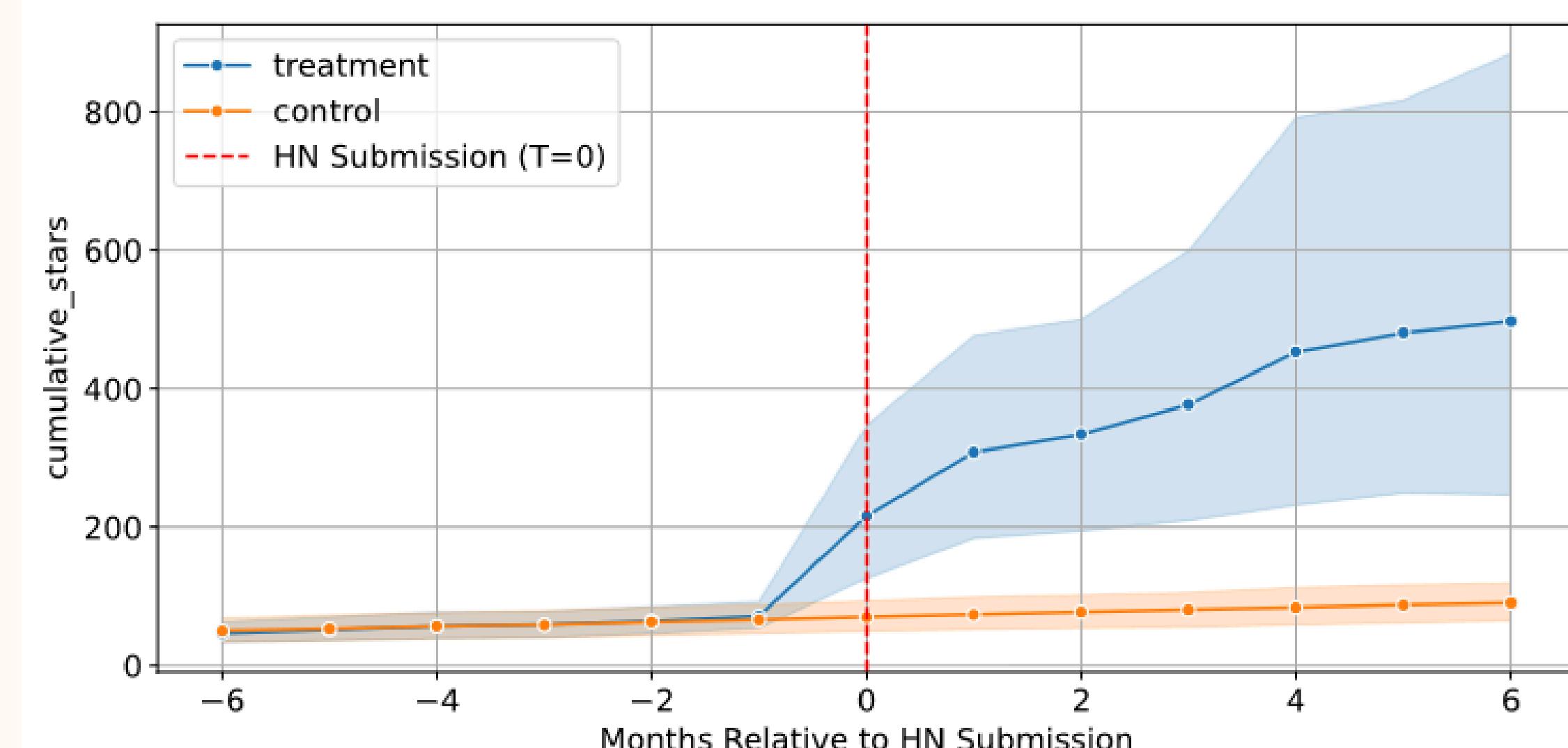
Visualizations further confirmed that pre-treatment metrics are parallel



DiD PTA Relative Metrics

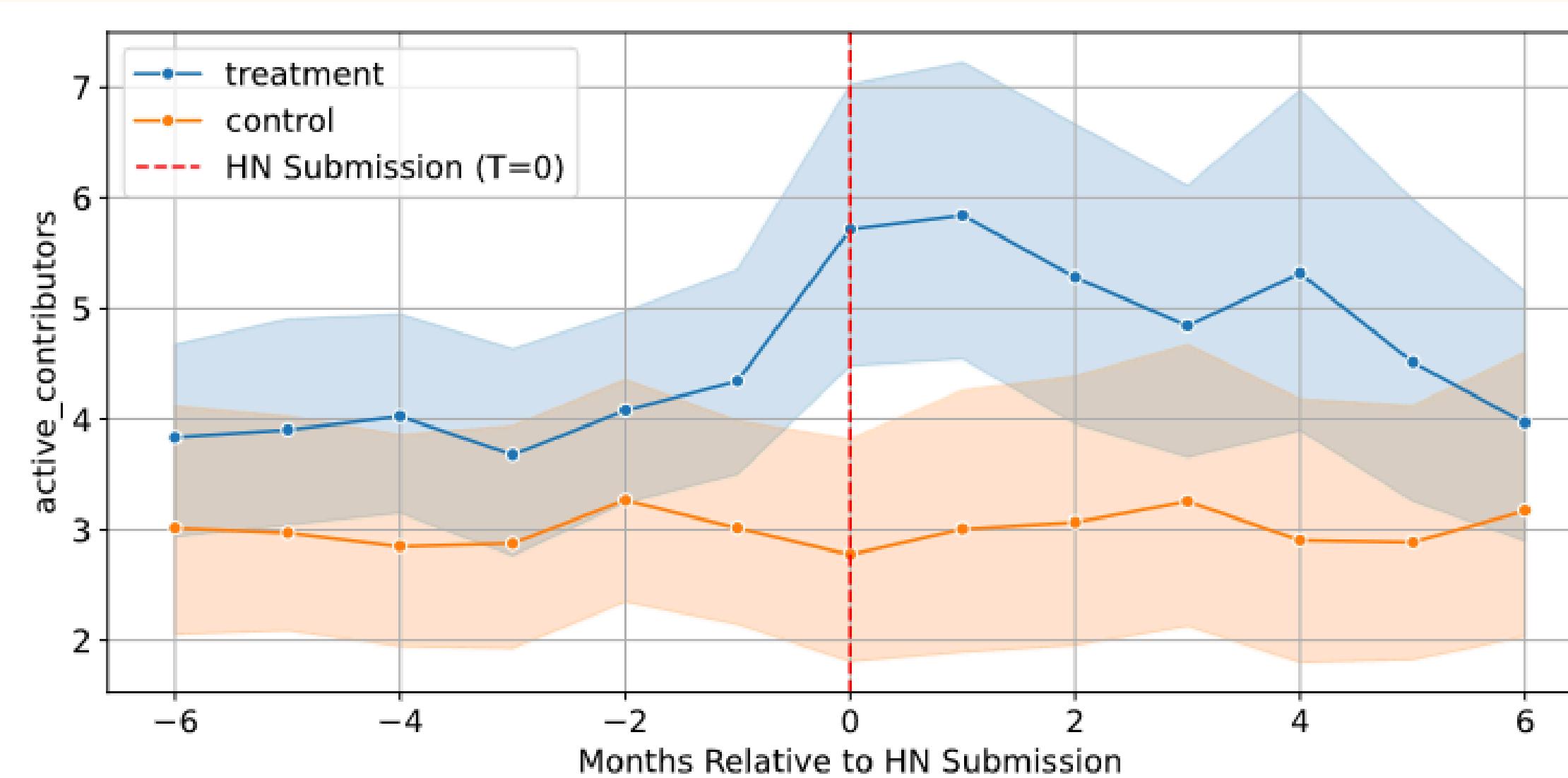


DiD PTA Relative Metrics



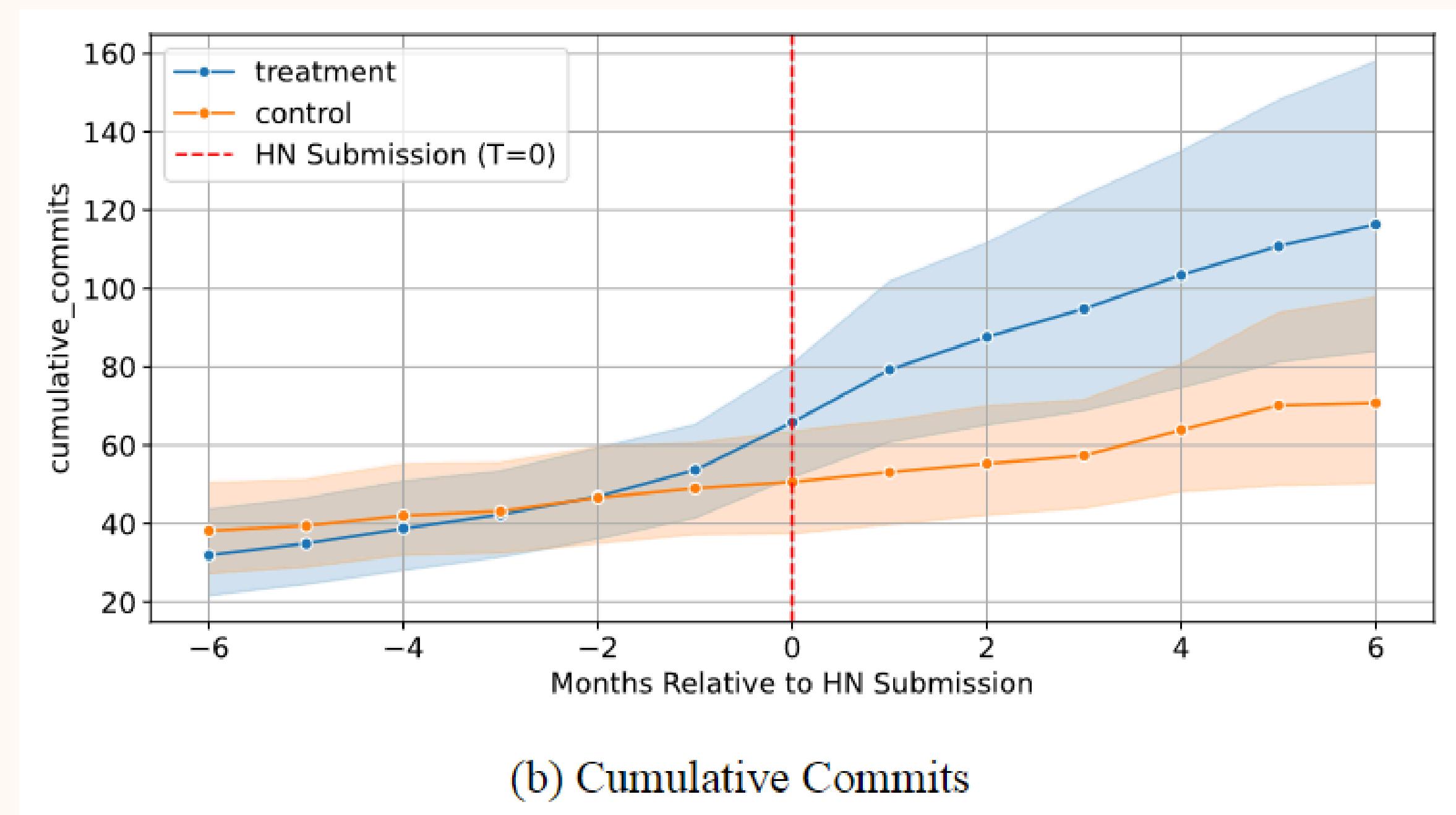
(e) Cumulative Stars

DiD PTA Relative Metrics

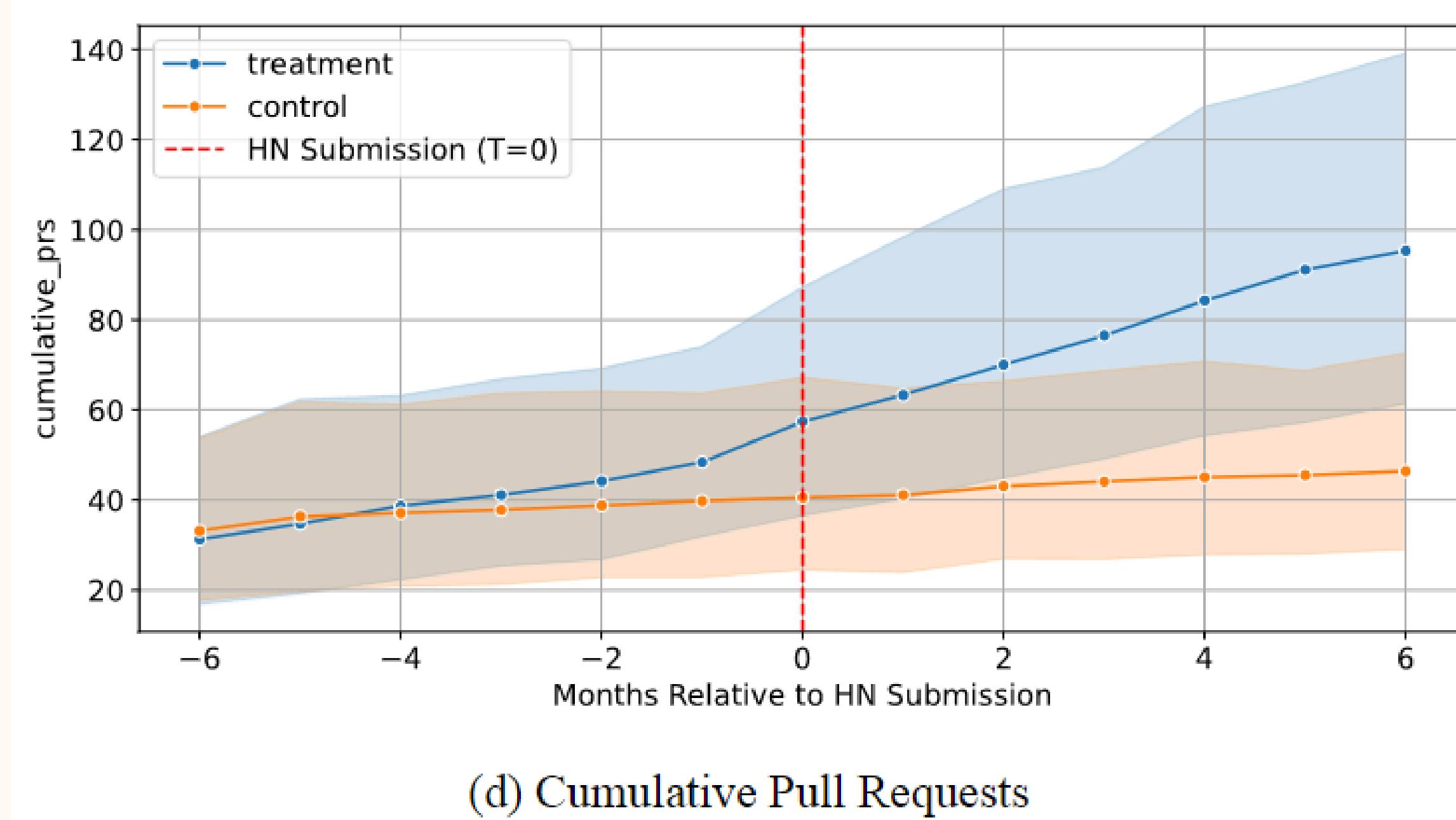


(a) New Monthly Active Contributors

DiD PTA Relative Metrics



DiD PTA Relative Metrics



RQ3 What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

Causal Effects of Hacker News Submissions:

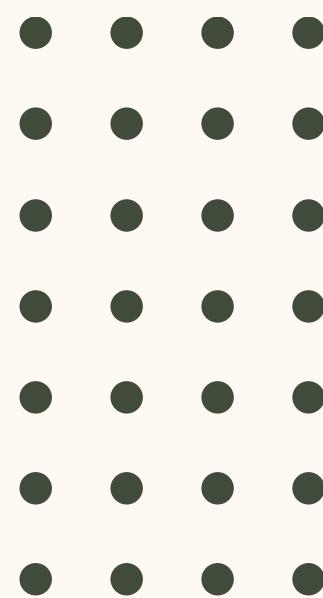
(2) Difference-in-difference results

Metric	DiD	p	Lagged DiD	p	Verdict
Cumulative Stars	299.67	0.0000	30.12	0.0141	✓ Strong Evidence
Cumulative Forks	55.12	0.0000	2.87	0.0415	? Mixed Evidence
Cumulative Commits	35.44	0.0000	0.71	0.6645	? Mixed Evidence
Cumulative PRs	30.56	0.0023	2.65	0.0002	? Mixed Evidence
New Stars	38.35	0.0003	35.71	0.0018	✓ Strong Evidence
New Forks	7.69	0.0000	4.73	0.0009	✓ Strong Evidence
New Commits	3.46	0.1731	1.88	0.5476	✗ No Effect
New PRs	-0.02	0.9862	-0.58	0.4057	✗ No Effect
New Contributors	1.08	0.0161	-0.03	0.8571	? Mixed Evidence

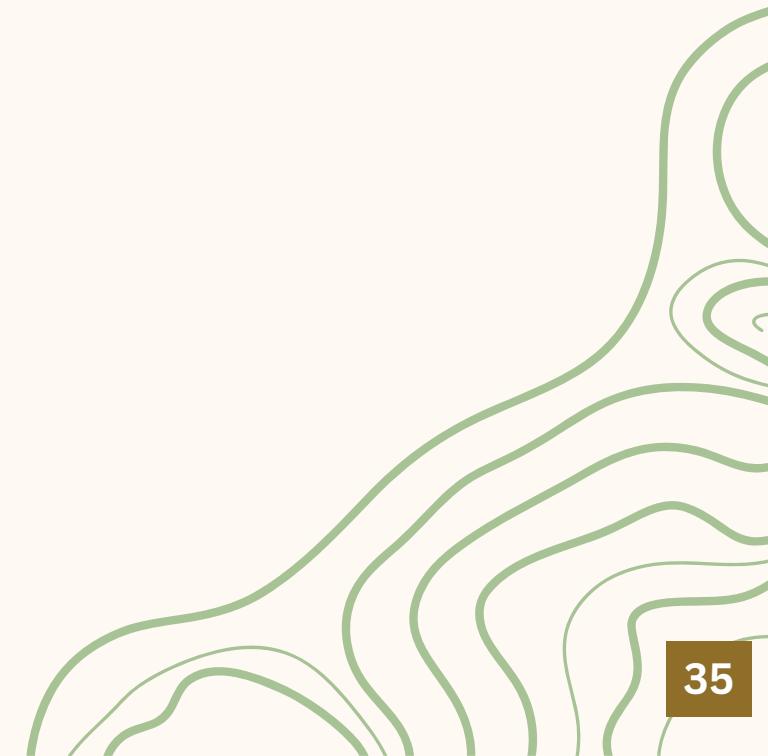
- Visibility gains are caused by Hacker News submission (incl. stars and forks)

- No causal effects on development boost (incl. commits and PRs)

Observation 10: The submission of Hacker News contribute to direct repository stars and forks growth.



Thanks!
Any Questions?



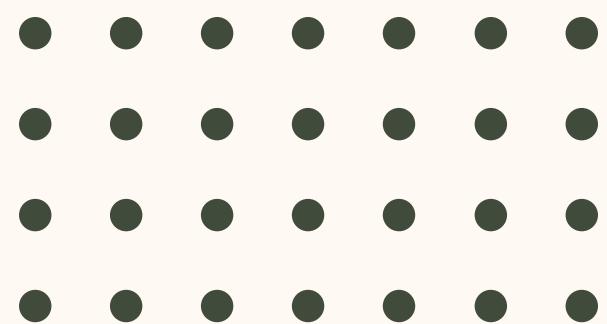
• INTRODUCTION

• BACKGROUND & RELATED WORKS

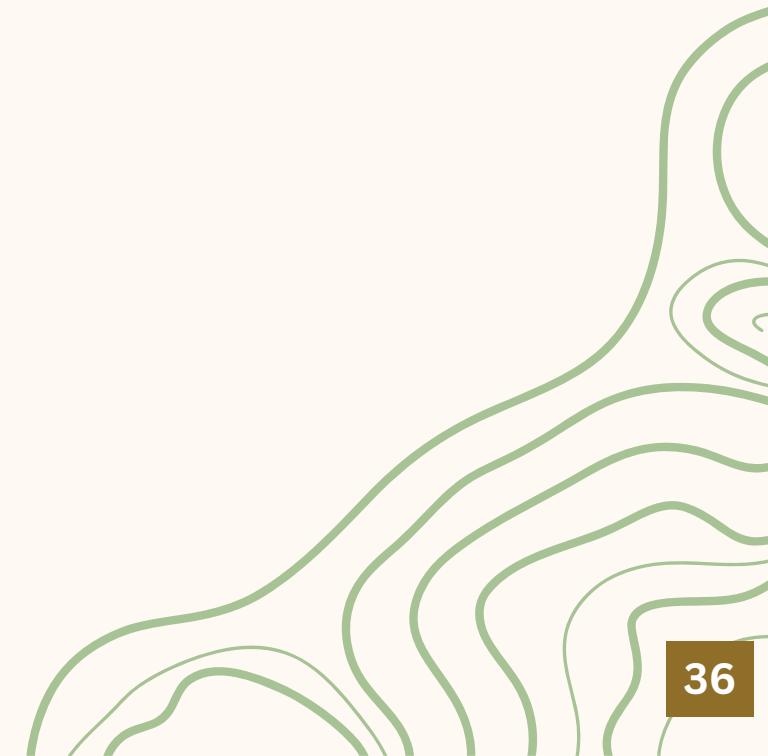
• RQ1 - SPREAD

• RQ2 - REACTION

• RQ3 - ACTIVITIES

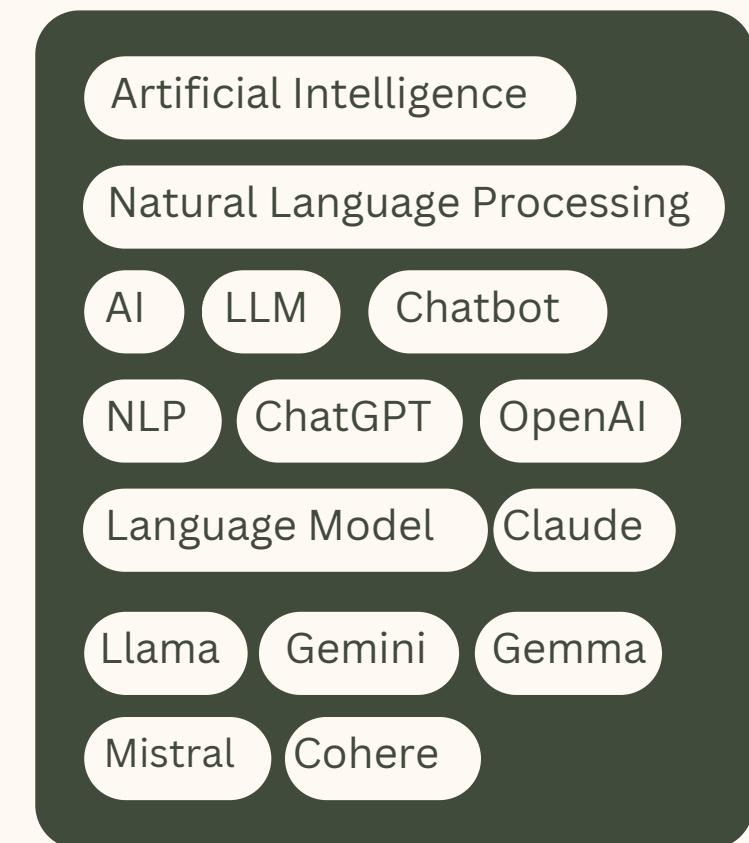


Appendix

An abstract graphic featuring a series of thin, light brown wavy lines that curve and overlap, creating a sense of depth and motion. It serves as a background for the word "Appendix".

Keywords used for data collection

- Major keyword categories
 - (1) frequent AI/LLM-specific terms e.g. “ai”, “nlp”, “openai” [1]
 - (2) additional keywords related to recent LLM models e.g. “claude”, “llama”, “gemini”, etc. [2]

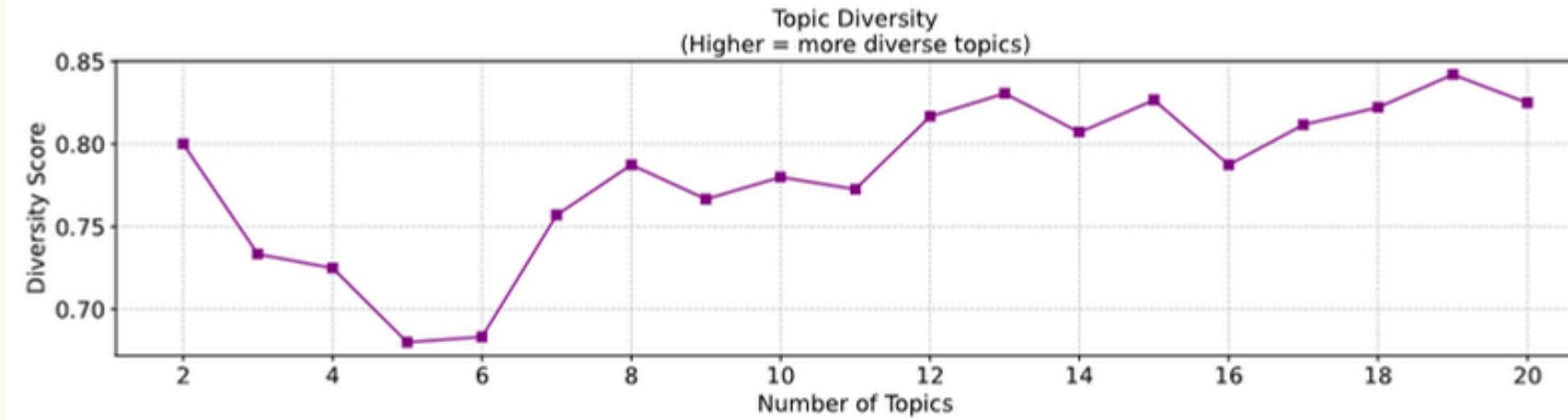
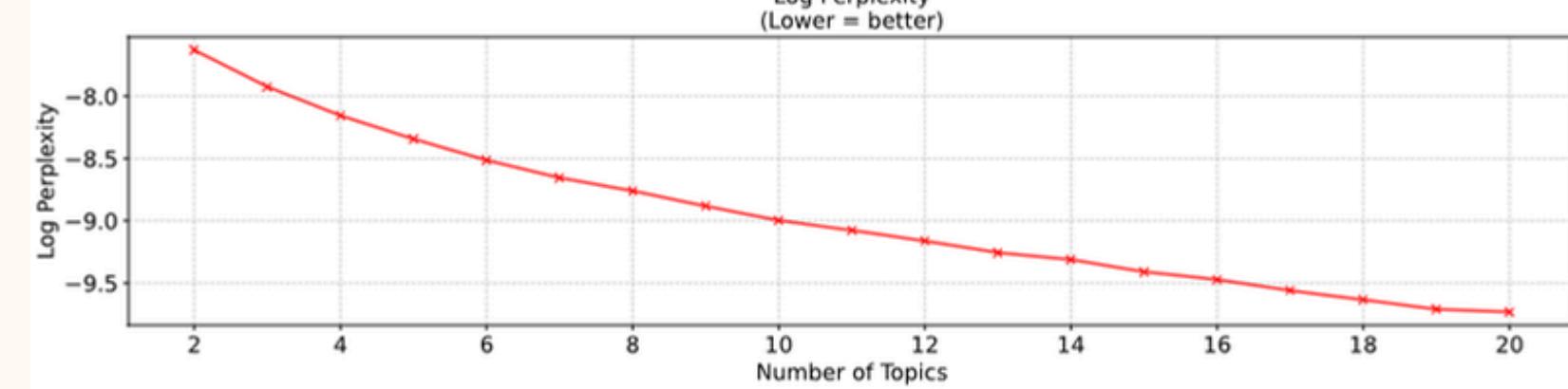
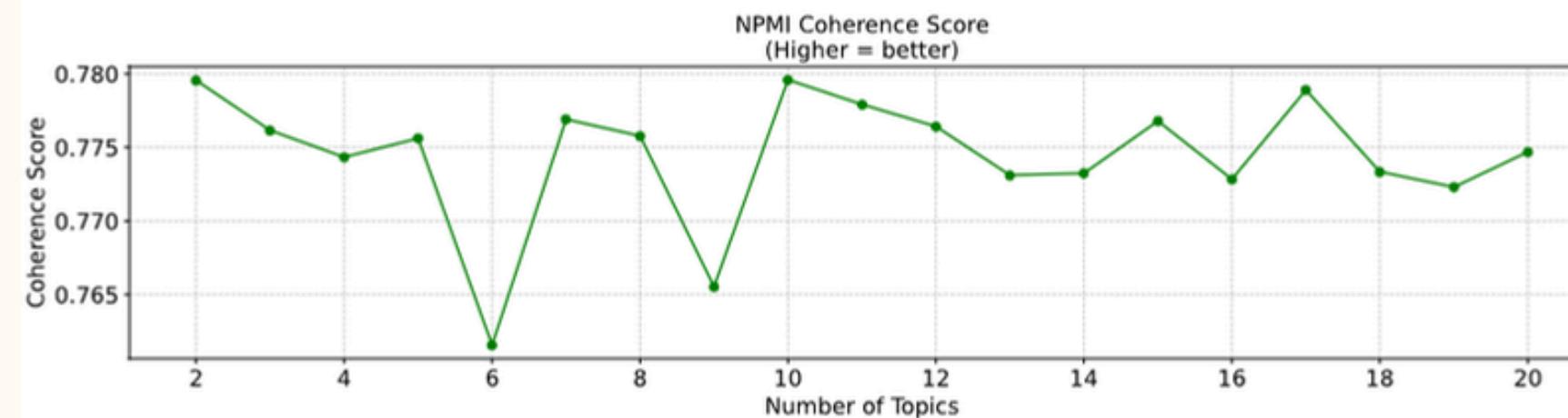
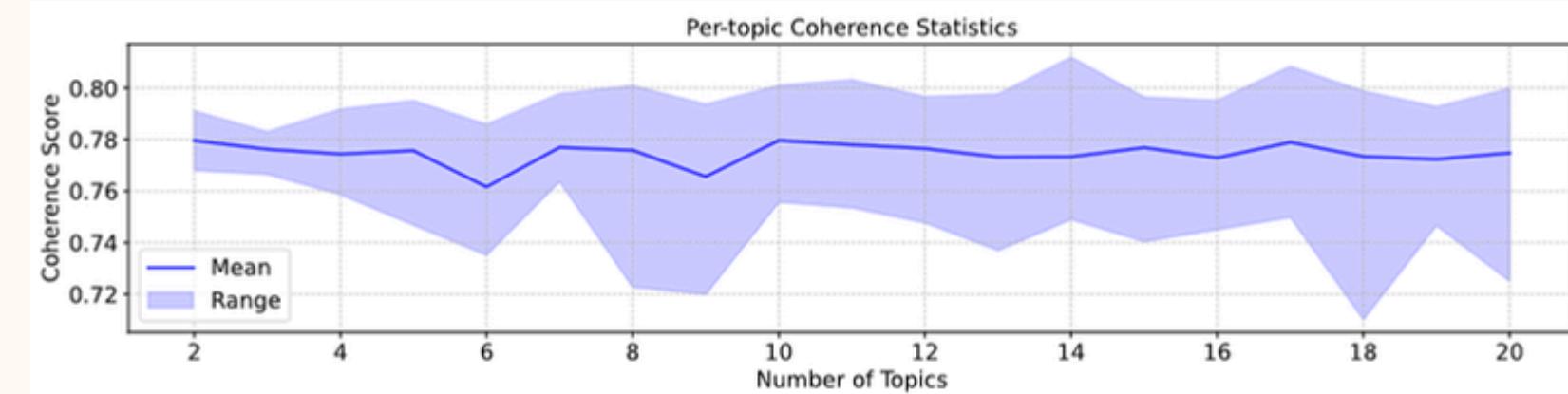
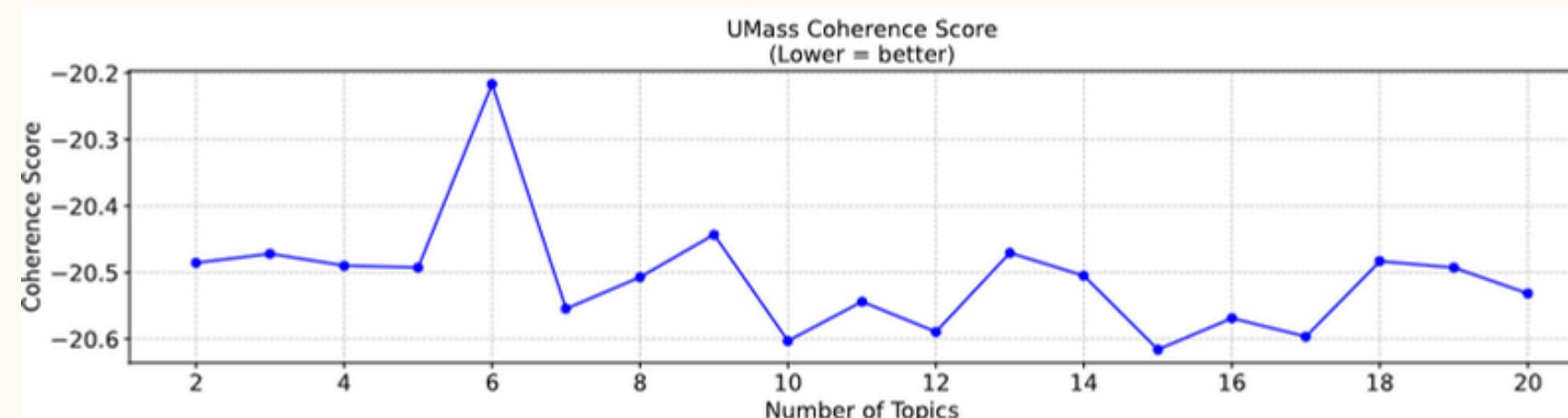


[1] “Machine Learning Glossary,” Google for Developers. Accessed: Nov. 17, 2024. [Online]. Available: <https://developers.google.com/machine-learning/glossary>

[2] S. Minaee et al., “Large Language Models: A Survey,” Feb. 20, 2024, arXiv: arXiv:2402.06196. doi: [10.48550/arXiv.2402.06196](https://doi.org/10.48550/arXiv.2402.06196).

HN GH-AI stories topic number evaluation

For topic modeling using LDA.



LDA additional parameter terminology

- **learning_decay**

- Controls how quickly the learning rate decreases during training.
- 0.7: Balanced decay—adapts to data while stabilizing reasonably fast.

- **iterations**

- Number of passes over the data.
- 30: Moderate; may need more for larger datasets or finer topics.

- **alpha (Document-topic prior)**

- Prior for document-topic distribution i.e. distribution of words per topic
- 'auto': Learns asymmetric topic weights (some topics dominate).

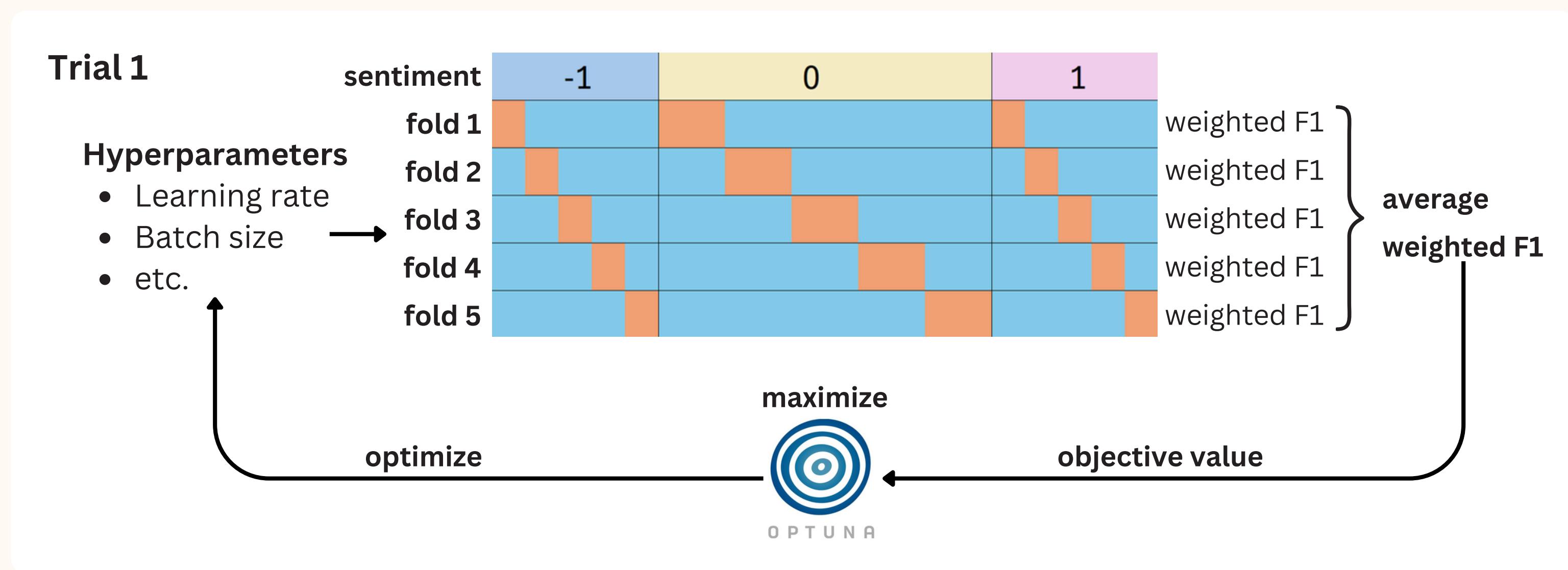
- **eta (Topic-word prior)**

- Prior for topic-word distribution.
- 'auto': Learns asymmetric word weights (key words per topic stand out).

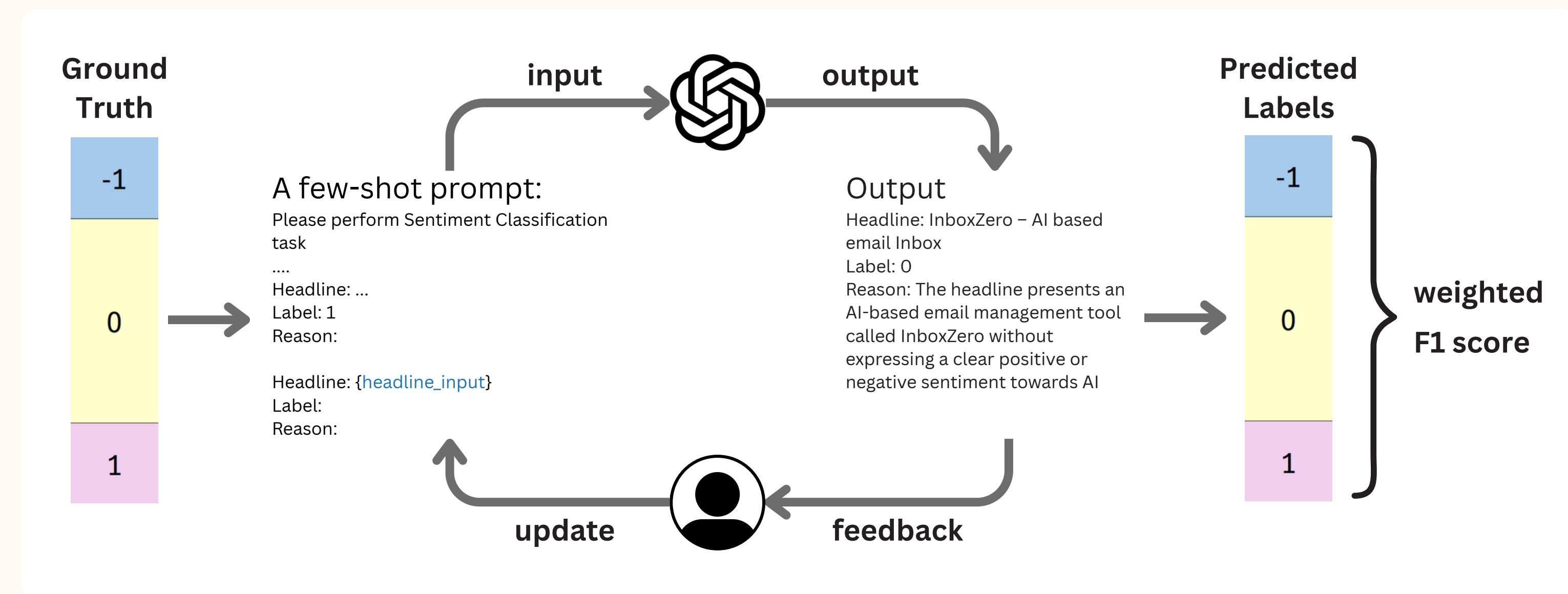
Ground Truth Dataset: Cohen's Kappa coefficient

- **Cohen's kappa agreement** between the two investigators was performed:
 - 0.626 for comments and 0.719 for stories indicating a substantial agreement

Fine-tuning process



Prompting LLM process



RQ2 What are the social reactions to HN GH-AI stories?

1. Analysis of HN GH-AI Story and Comment Sentiment

- Using an **area stack** to plot each sentiment (negative, neutral, positive) of HN GH-AI stories and comments over time.

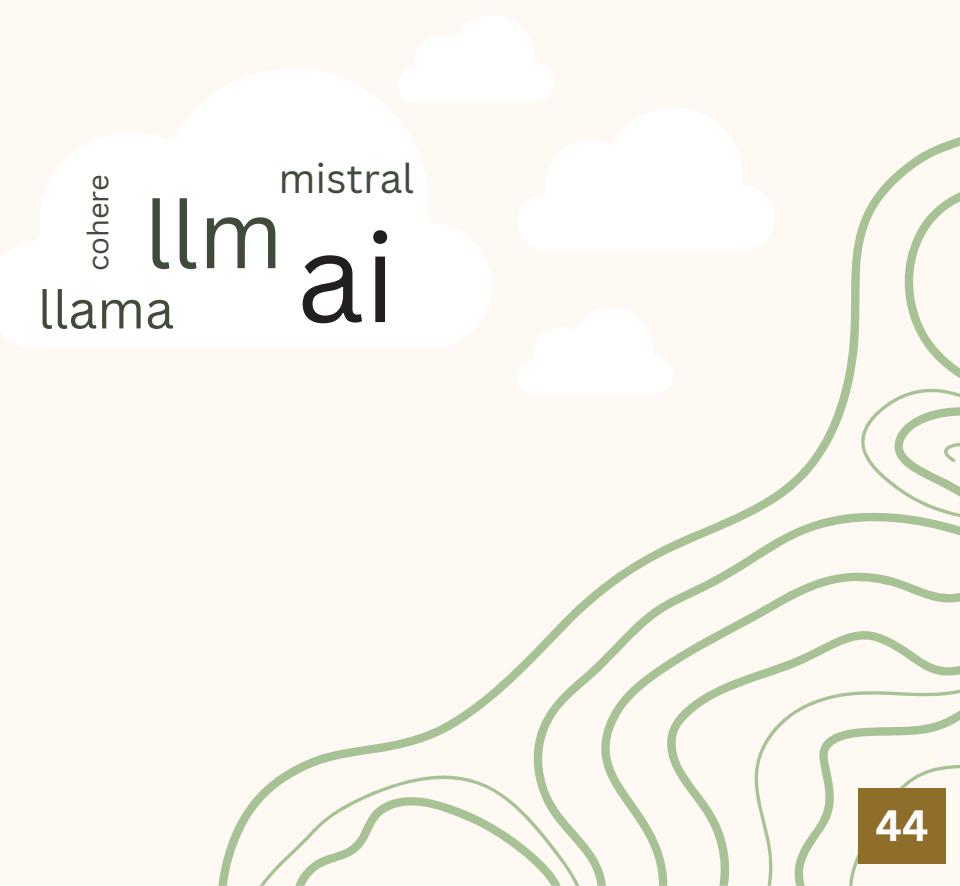
2. Analysis of HN GH-AI Story's Social Reaction

- Relationships were defined:
 - **Negative/Positive** Comment → **Strong** Reaction
 - **Neutral** Comment → **Weak** Reaction
- These reactions were averaged and plotted by month over time using a **heat map**.

RQ2 What are the social reactions to HN GH-AI stories?

3. Analysis of GH Topic Cloud Based on Sentiment

- Sentiment of each repository is defined by its **average comment sentiment**:
 - Average comment sentiment $> 0.5 \rightarrow$ Repository is considered to be **positive**
 - Average comment sentiment $< -0.5 \rightarrow$ Repository is considered to be **negative**
- GitHub topics were preprocessed by **normalizing** topics with typos or equivalent meanings into a unified form.
 - Ex. ‘large-language-model’, ‘large-language-models’, and ‘llms’ \rightarrow ‘llm’
- **Duplicate** topics in the same repository were **removed**.
- **Word clouds** were used to visualize frequency of GH topics in each sentiment (negative or positive).

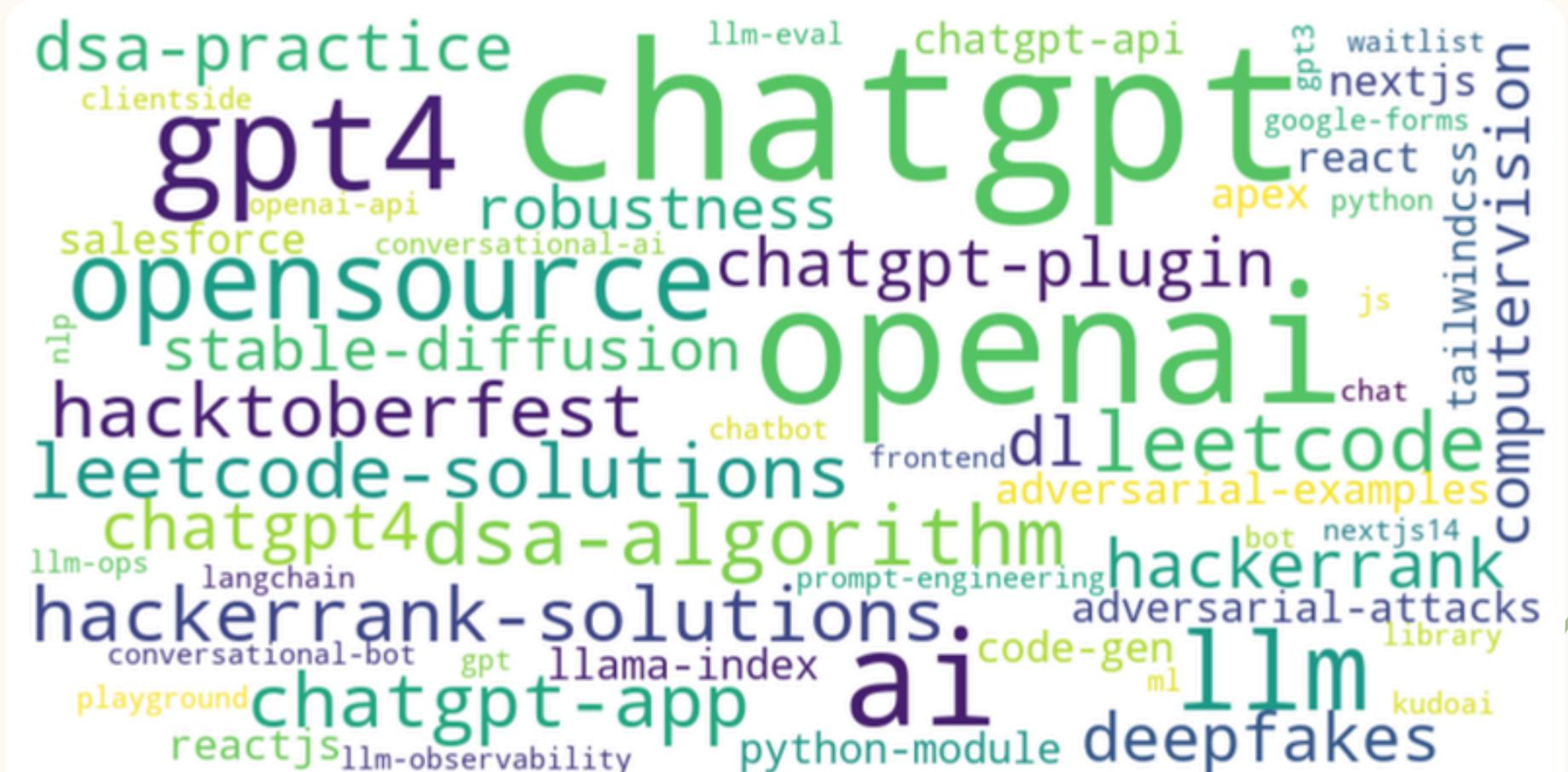


Additional Analysis of GH Topic Cloud Based on Sentiment

Word Cloud in Positive and Negative Repositories (Not Exclusive).



GitHub Topic Cloud in Positive Sentiment Repositories



GitHub Topic Cloud in Negative Sentiment Repositories

Additional Analysis of GH Topic Cloud Based on Sentiment

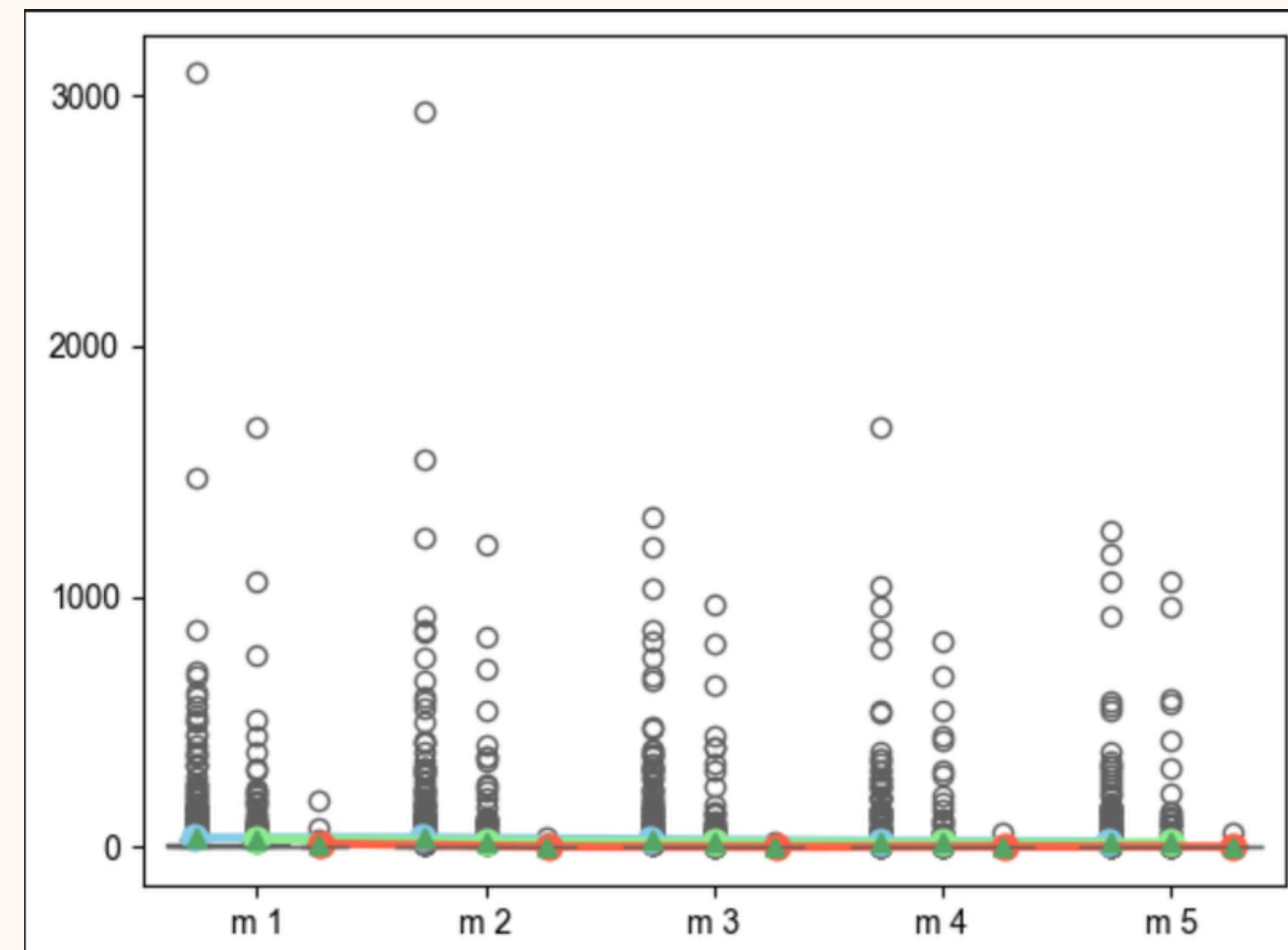
- Comments in **positive** repositories

- Show **enthusiasm** for projects applying AI tools
- Describe the project as ‘**desirable**’, ‘**needed**’, ‘**useful**’, and ‘**full of potential**’

- Comments in **negative** repositories

- Stem from **dissatisfaction** with **LLM** performance
- Concerns about their potential to **replace human workers**

Sentiment-Metric with Outliers



RQ3

What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

GitHub Metrics Changes

Visualizing the Historical metrics dataset

Using our historical GitHub metric dataset

- **Filter outliers using IQR prior to plotting**
 - Due to potential volatility and skewness i.e.
not normally distributed
- **Excluded entries where all metrics were 0**
 - To reduce the chance of metrics from being
skewed towards 0

Then visualize metric changes in each month

DiD terms and definitions

- Y_{it} : The outcome metric (stars, forks, etc.) for repository i at time t
- α : Baseline average outcome for control repositories pre-treatment
- $treatment_{it}$: Binary indicator (1 = repository was submitted to HN, 0 = control repository)
- $post_treatment_{it}$: Binary indicator (1 = time period after HN submission, 0 = before)
- β_1 : Time trend effect (shared by both treatment and control groups)
- β_2 : Pre-existing difference between treatment vs. control groups
- β_3 : Causal effect of HN submission (difference-in-differences estimator)
- ϵ_{it} : Unexplained variation (measurement error, idiosyncratic factors)

DiD Model

$$Y_{it} = \alpha + \beta_1 \cdot post_treatment_{it} + \beta_2 \cdot treatment_{it} + \beta_3 \cdot (post_treatment_{it} \times treatment_{it}) + \epsilon_{it}$$

Captures short-term “momentum effects”

Lagged DiD Model

$$Y_{it} = \alpha + \beta_1 \cdot post_treatment_{it} + \beta_2 \cdot treatment_{it} + \beta_3 \cdot (post_treatment_{it} \times treatment_{it}) + \gamma \cdot Y_{i,t-1} + \epsilon_{it}$$

Takes into account the previous iteration

Isolates only the persistent long-term effect

RQ3

What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

Causal Effects of Hacker News Submissions:

To perform DiD, need to **establish parallel trend assumption (PTA)**

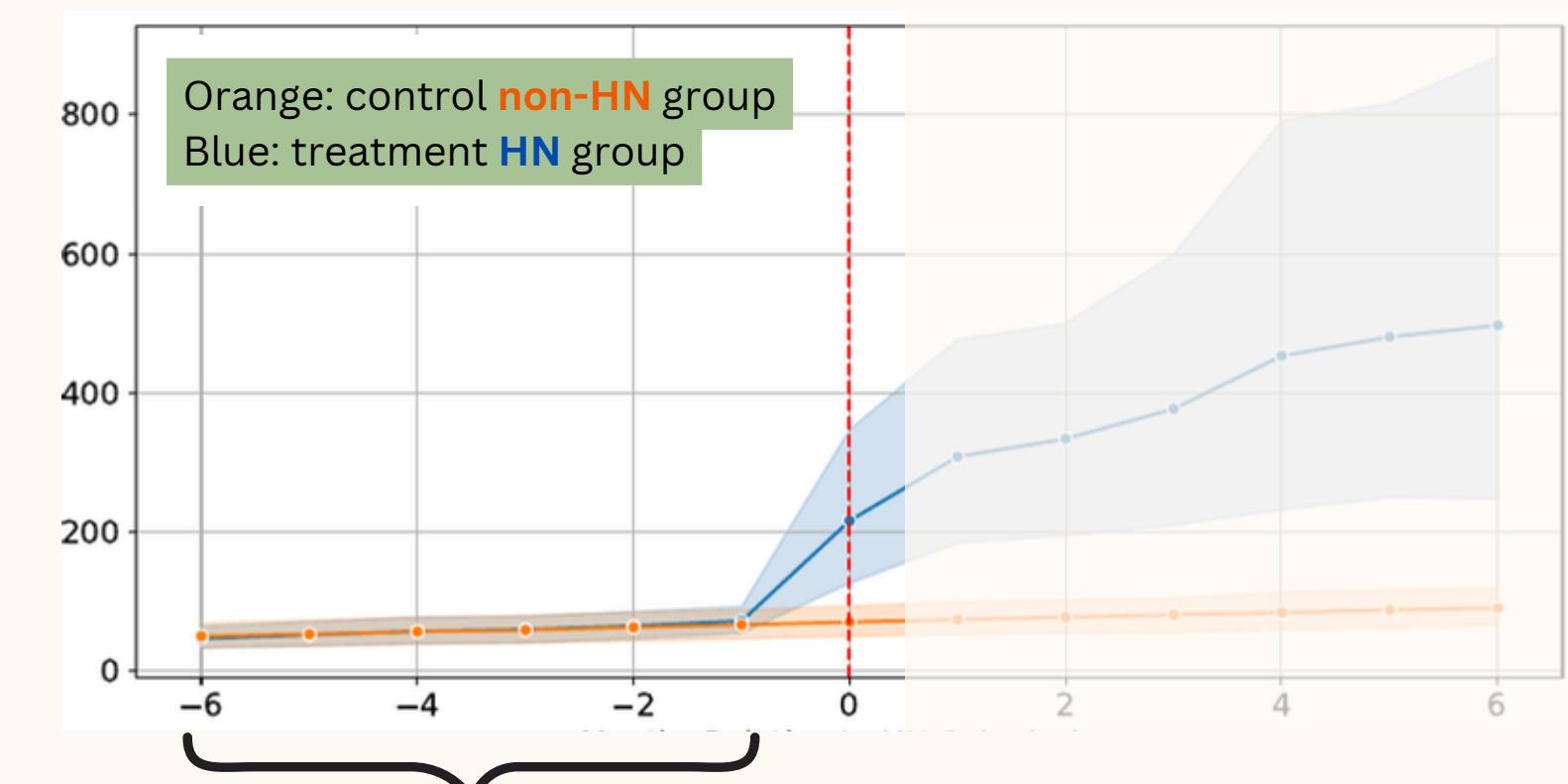
- I.e. metrics prior to **treatment** (HN submission) must have a **similar pre-treatment trajectory**

Issue 1: non-HN GH-AI projects **lack HN submission date**

- **Cannot compare** pre-treatment metrics directly

Issue 2: each treatment repository **can be submitted on HN at different dates**

- **Cannot mark treatment event date** directly for every entries



✓ PTA: both control and treatment are parallel prior to **treatment event: HN submission month**

Need to **match each treatment repository (HN)** with each **control repository (non-HN)**

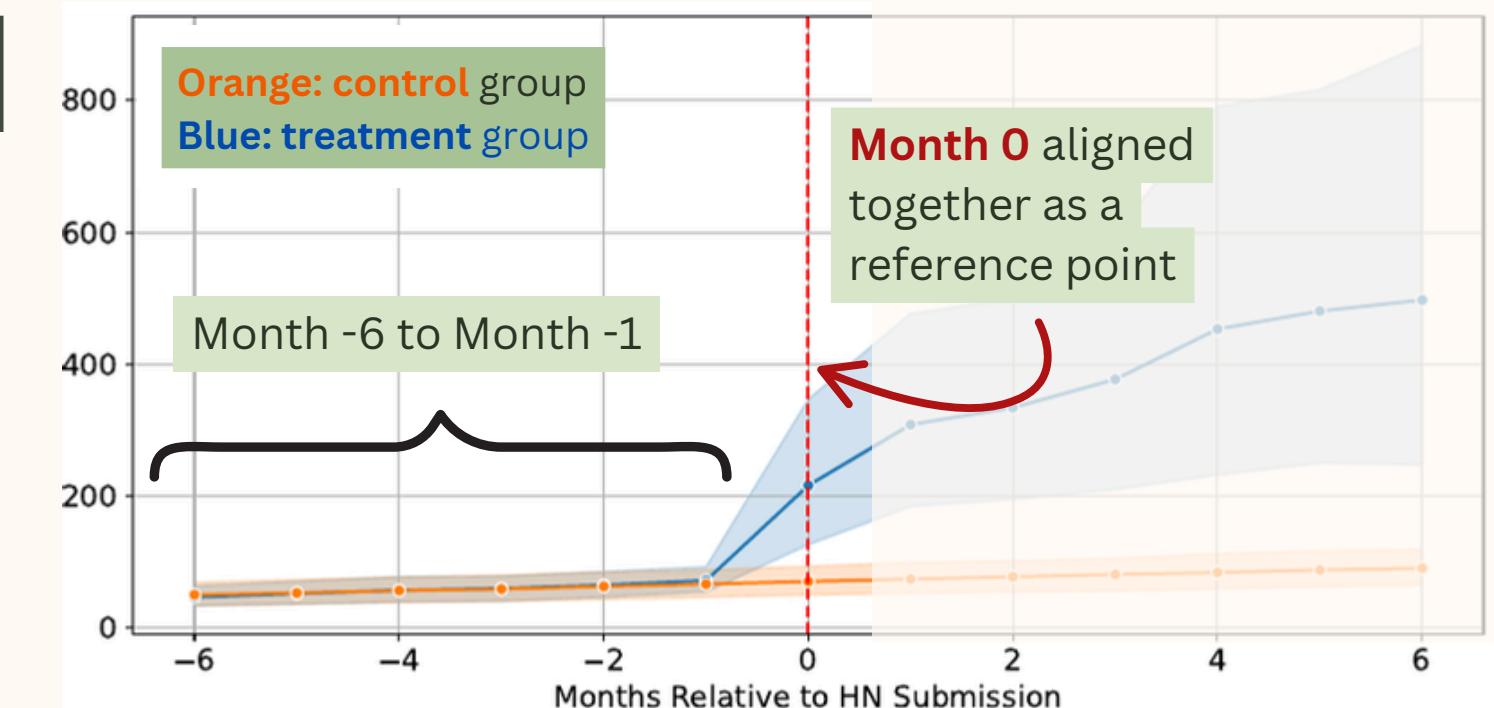
RQ3

What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

Causal Effects of Hacker News Submissions:

(1) Matching Treatment-control repository pair

Similarly, only use treatment HN GH-AI containing **6 months activities** before/after HN submission to ensure sufficient activities.

**1**

Process HN GH-AI dataset into **relative months**, and align all repositories together using **month 0** as a reference point
E.g. “month -6” denotes **6 months before HN submission**

month (month 0)

✓ Can now use HN submission date as a **common treatment event**

3

For each treatment repository, **find control repository** whose **pre-treatment metrics are most similar**

i.e. vectors yield the most similarity based on our criteria

✓ non-HN GH-AI's can now reference its **partner treatment repository's month 0**

Each pair must satisfy the following:

- Cosine similarity (threshold=0.8)
- Euclidean distance (limit=20)

RQ3 What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?

Causal Effects of Hacker News Submissions:

(1) Matching Treatment-control repository pair

Methodology Recap

- Pair each treatment repository (HN) with a control repository (non-HN)
- Most similar cumulative metrics **6 months before HN submission month (month 0)**

Matching final result

Metric	Number of Project Pairs
Cumulative Commits	35
Cumulative Forks	73
Cumulative PRs	41
Cumulative Stars	43
Monthly Active Contributors	93

Matching criteria:

- Cosine similarity (threshold=0.8)
- Euclidean distance (limit=20)

- From a pool of
 - 163 treatment repositories (**HN**)
 - 2,147 candidate control repositories (**non-HN**)

Future works

- **Replicate and extend our analysis to other platforms**
- **Extract HN comments' qualitative feedback and evaluate their values**
to developers
- Look into which aspects are viewed favorably or unfavorably by
implementing aspect-based sentiment analysis
- Extend **Difference-in-Differences (DiD)** to examine **longer-term** effects
- Understand the **intent** behind promoting a project on Hacker News