



ICSME
2025
Auckland, New Zealand

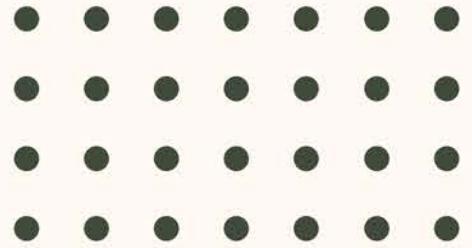


Mahidol University
Faculty of Information
and Communication Technology



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. Choetkiertikul,
T. Sunetnanta

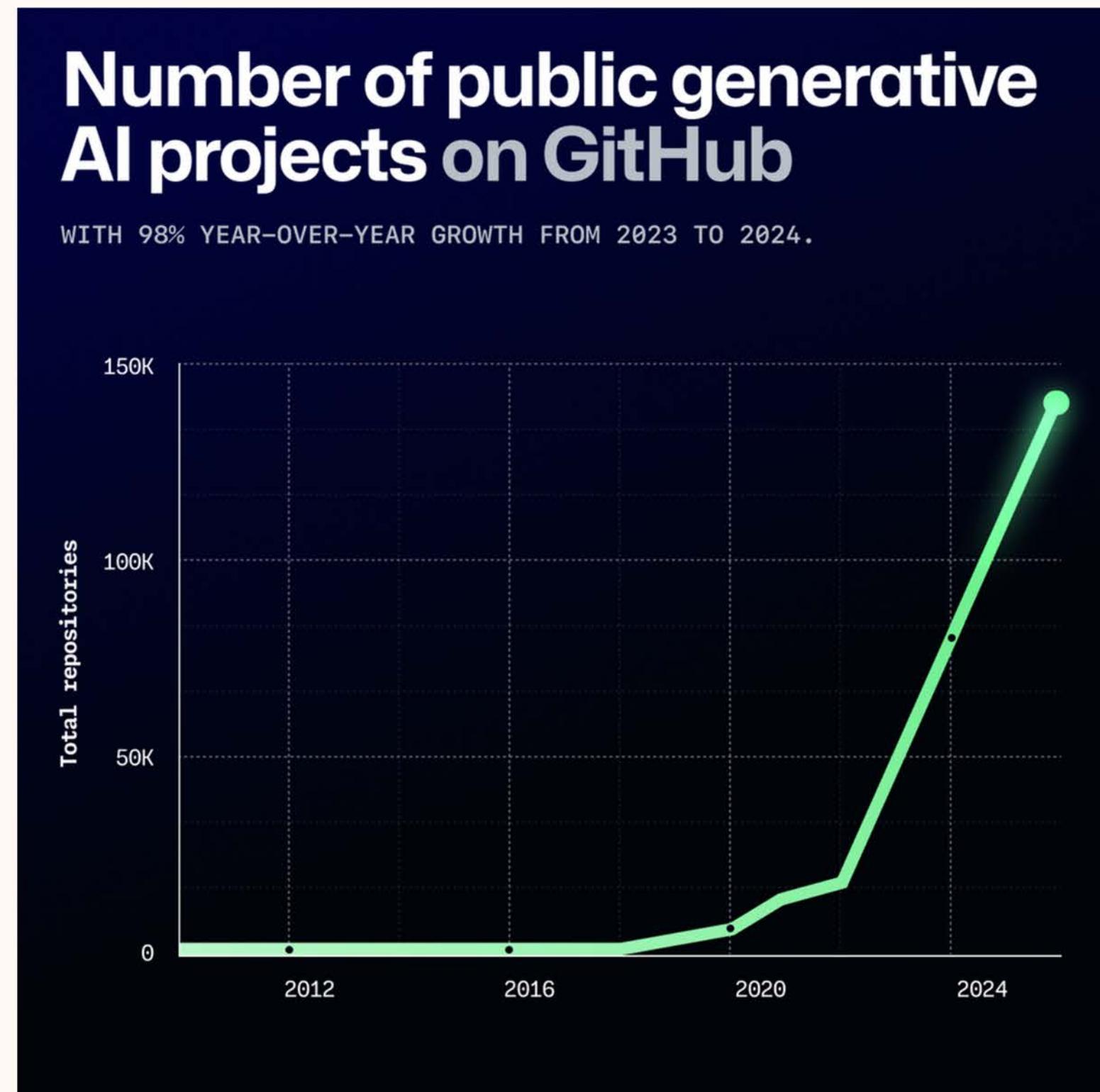
R.G. Kula K. Matsumoto

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

Why HackerNews? [1]

A story provides a **summary** of the project, its **key features**, and a **link** to OpenLIT's GitHub repository.

- **No user feed personalization**

- Top posts and comments are the same for all users, minimizing biases

- **Comments highly opinionated**

- Able to gather clear sentiment insights and statistics

- **The demographics is more fixated on technology topics**

- Compared to Reddit which is a generalized platform filled with various sub-communities

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

- ▲ The Rectangular Cows of UK Art (2018) (artuk.org)
100 points by 4 hours ago | hide | 41 comments
- ▲ Launch HN: Fresco (YC F24) – AI Copilot for Construction Superintendents
28 points by 2 hours ago | hide | 4 comments
- ▲ PHP 8.4 (php.net)
150 points by ours ago | hide | 117 comments
- ▲ Show HN: Bike route planner that follows almost only official bike trails (trailimap.com)
125 points by 's ago | hide | 76 comments
- ▲ Show HN: Yami – An Open Source Music Player with Spotal Integration (github.com/dever-m)
35 points by ours ago | hide | 18 comments
- ▲ Hyrum's Law in Golang (abenezer.org)
192 points by 10 hours ago | hide | 125 comments
- ▲ A binary tree of all Pythagorean triples (richardt.io)
80 points by urs ago | hide | 9 comments
- ▲ Everyone is capable of, and can benefit from, mathematical thinking (quantamagazine.org)
394 points by 15 hours ago | hide | 248 comments

Same feed for every users

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

▲ Show HN: OpenLIT – Open-Source LLM Observability with OpenTelemetry (github.com/openlit)
62 points by aman_041 10 months ago | hide | past | favorite | 22 comments

Hey HN, we're super excited to share something we've been working on: OpenLIT. After an en now proudly announcing our first stable release!

What's OpenLIT? Simply put, OpenLIT is an open-source tool designed to make monitoring straightforward. It's built on OpenTelemetry, aiming to reduce the complexity that comes with

Beyond Basic Text Generation: OpenVision, DALL·E, and OpenAI Audio. Esse not stopping here; more updates and m

1 hour ago | prev | next [-]

So basically you took the idea of Jazelle extensions that This is amazing, great work!

reply

Y Combinator

Make something people want.

Apply to YC

Hacker News



[1] "Hacker News FAQ." Accessed: Nov. 19, 2024. [Online]. Available: <https://news.ycombinator.com/newsfaq.html>

[2] "Y Combinator." Accessed: Nov. 19, 2024. [Online]. Available: <https://www.ycombinator.com/>

HackerNews Anatomy

The screenshot shows a Hacker News story page for a project called "Yami". The story title is "Show HN: Yami - An Open Source Music Player with Spotal Integration". The GitHub repository URL is provided as a comment: "github.com/dever-m". A red box highlights the URL, and a blue box highlights the "Story title". Below the story, there is a comment from a user named "lars ago" (depth 1) which includes a link to PyPI: "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/>". A green box highlights this comment. A reply to this comment (depth 2) from "L-M" says: "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." Another reply (depth 3) from "L-M" says: "Copy the an directory of audiofiles to /tmp, browse there, it works. Very strange." A green box highlights this comment.

Story URL
(GitHub repository URL in this case)

Show HN: Yami - An Open Source Music Player with Spotal Integration github.com/dever-m

Story title

add comment

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 1)

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 2)

if possible can you create an issue about this in github with more details(screen shots)

Comment (depth 3)

▲ 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!

▲ 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 for LLMs like Qwen, Qwen+, and Qwen-4.

Vision
not st

*Key

- *In
Supp
*Alig
comm
*Wide
Open

*Gett
https

We ge
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 [-]

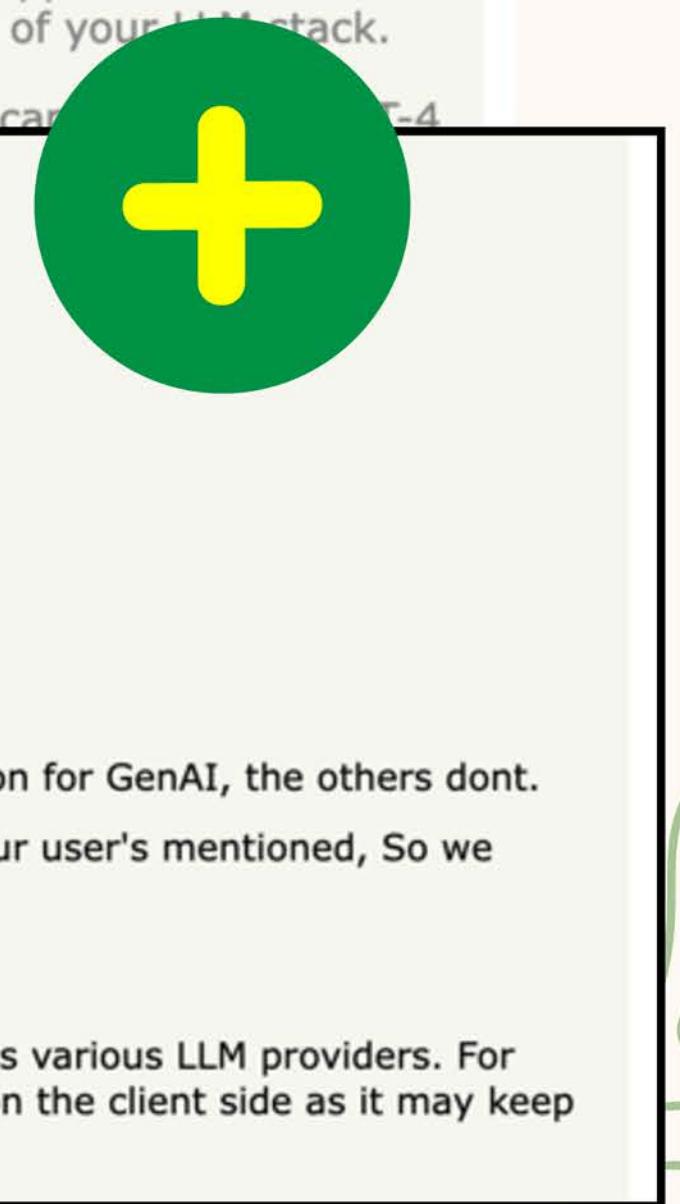
Yup that's correct, Love as more tools come up! We though tried to follow the actual OpenTelemetry's Semantic Convention for GenAI, the others don't.

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



Background & Related Works

Related Works about Social Platforms, Hacker News, and GitHub Studies

Studies about AI/LLM on other platforms

- Study of discussions related to deep learning across platforms such as Github and Stack Overflow (Han et al., 2020)[1]

Studies based on HackerNews data

- Study of popularity dynamic in Reddit and Hacker News (Stoddard, 2021)[2]
- Study of Reddit and Hacker News as modern news aggregators for developers (Aniche et Al, 2018)[3]
- Study of Hacker News's effectiveness in software engineering research (Titus B et al, 2015)[4]

Key Takeaway

HackerNews is a highly efficient and reliable source of high quality information

The trend of AI/LLM is different on each platform so these platforms may need to be studied separately.

[1] Han J., Shihab E., Wan Z., Deng S., Xia X. (2020) "What Do Programmers Discuss about Deep Learning Frameworks",

[2] Stoddard G. (2021) "Popularity Dynamics and Intrinsic Quality in Reddit and Hacker News", AAAI'21

[3] Aniche et al. (2018) "How Modern News Aggregators Help Development Communities Shape and Share Knowledge", ICSE '18

[4] Titus B et al. (2015) "I heart hacker news", FSE'15

Used in RQ2

Sentiment Analysis

- A technique used to determine **emotional** tone or **attitude** expressed in **text**
- This work uses sentiment analysis to determine **HN community's reception towards AI GitHub project.**

Text	Sentiment towards AI
This chatbot is the best I have tried.	1
ChatGPT is getting worse every year.	-1
Have a nice day, everyone.	0

Examples of text and its sentiment towards AI

Used in RQ2

Related Works about Sentiment Analysis in Software Engineering

Studies about Sentiment Analysis

1. Exploration of multiple **sentiment analysis models** to software engineering (SE) dataset
(Lin et al., 2018) [1]

2. Study of **fine-tuned pretrained transformer models** for sentiment analysis on SE dataset
(Zhang et al., 2020) [2]

3. Investigation of **LLMs** for sentiment analysis on SE dataset (Zhang et al., 2025) [3]

Key Takeaway

These studies suggest we should **experiment with multiple approaches** for sentiment analysis.

We explored **fine-tuned pretrained transformer models** and **LLM**.

Used in RQ3

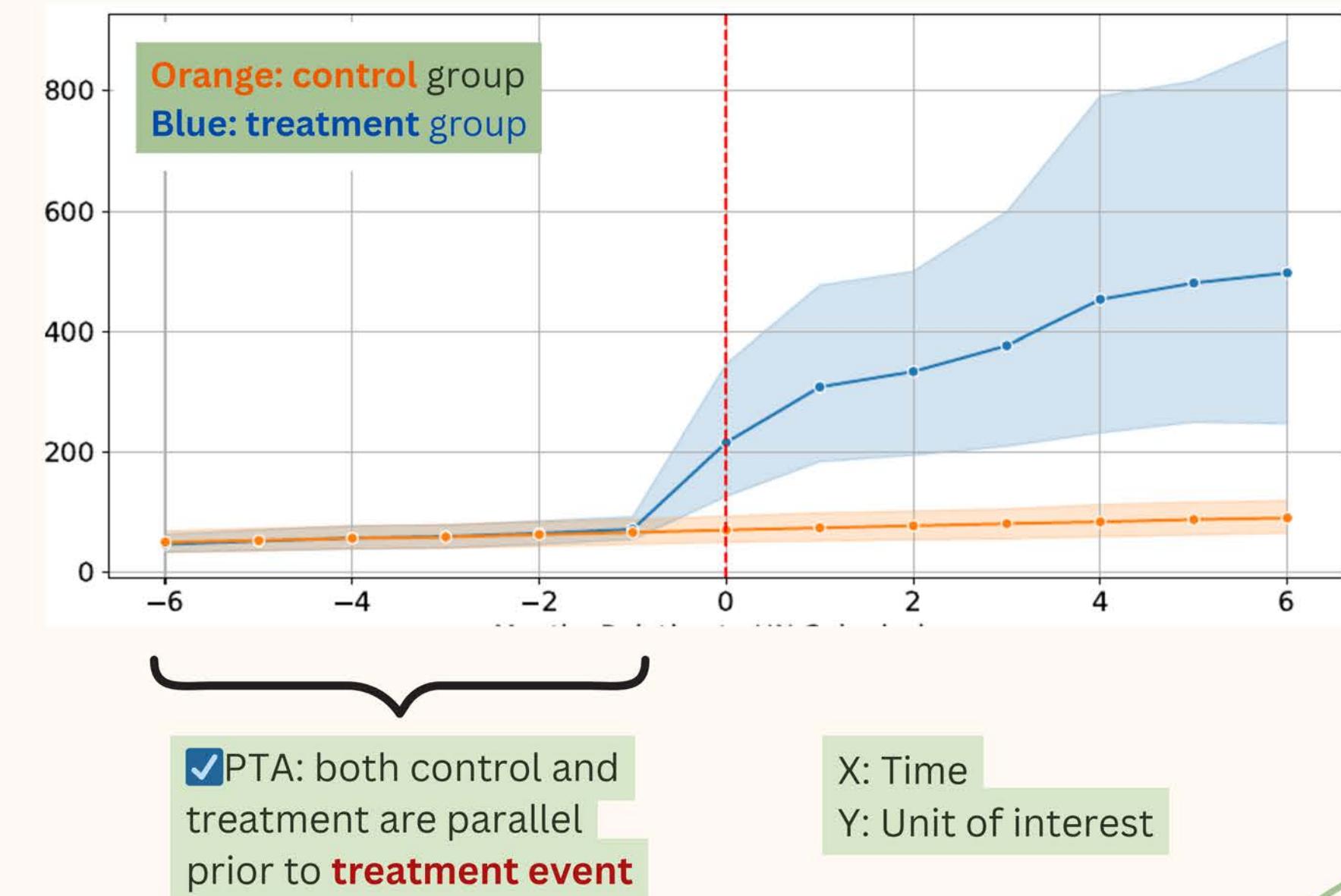
Causal Inference

- Techniques to estimate the true effect of a treatment while accounting external factors
- Answering whether “correlation implies causation”
- E.g. propensity score matching (PSM), and DiD

Difference-in-differences (DiD)

- Compare changes before and after an intervention between the Treatment group & Control group
- Must obey the Parallel Trend Assumption (PTA)

Treatment occurs on the red dashed line
e.g. covid-19 occurs, policy change occurs



Note: parallel does not mean perfectly identical trends,
only means the same expected trajectory

• INTRODUCTION

• BACKGROUND & RELATED WORKS

• RQ1 - SPREAD

• RQ2 - REACTION

• RQ3 - ACTIVITIES

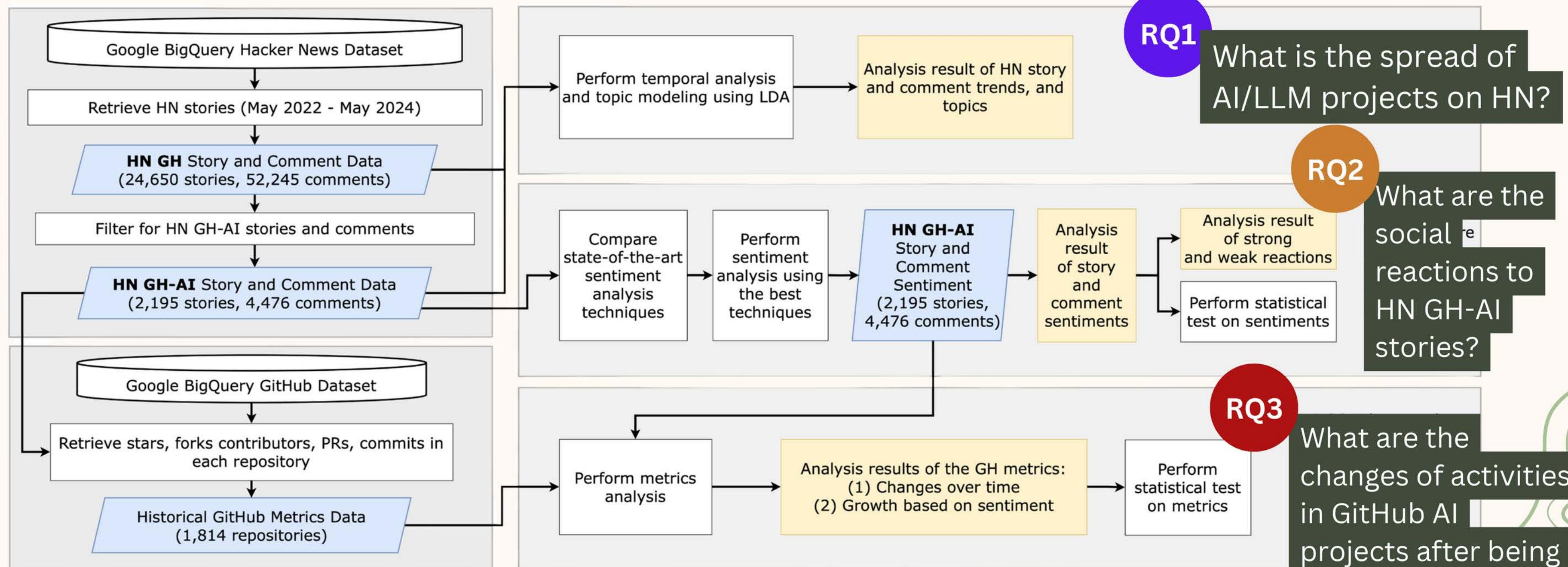


Methodology

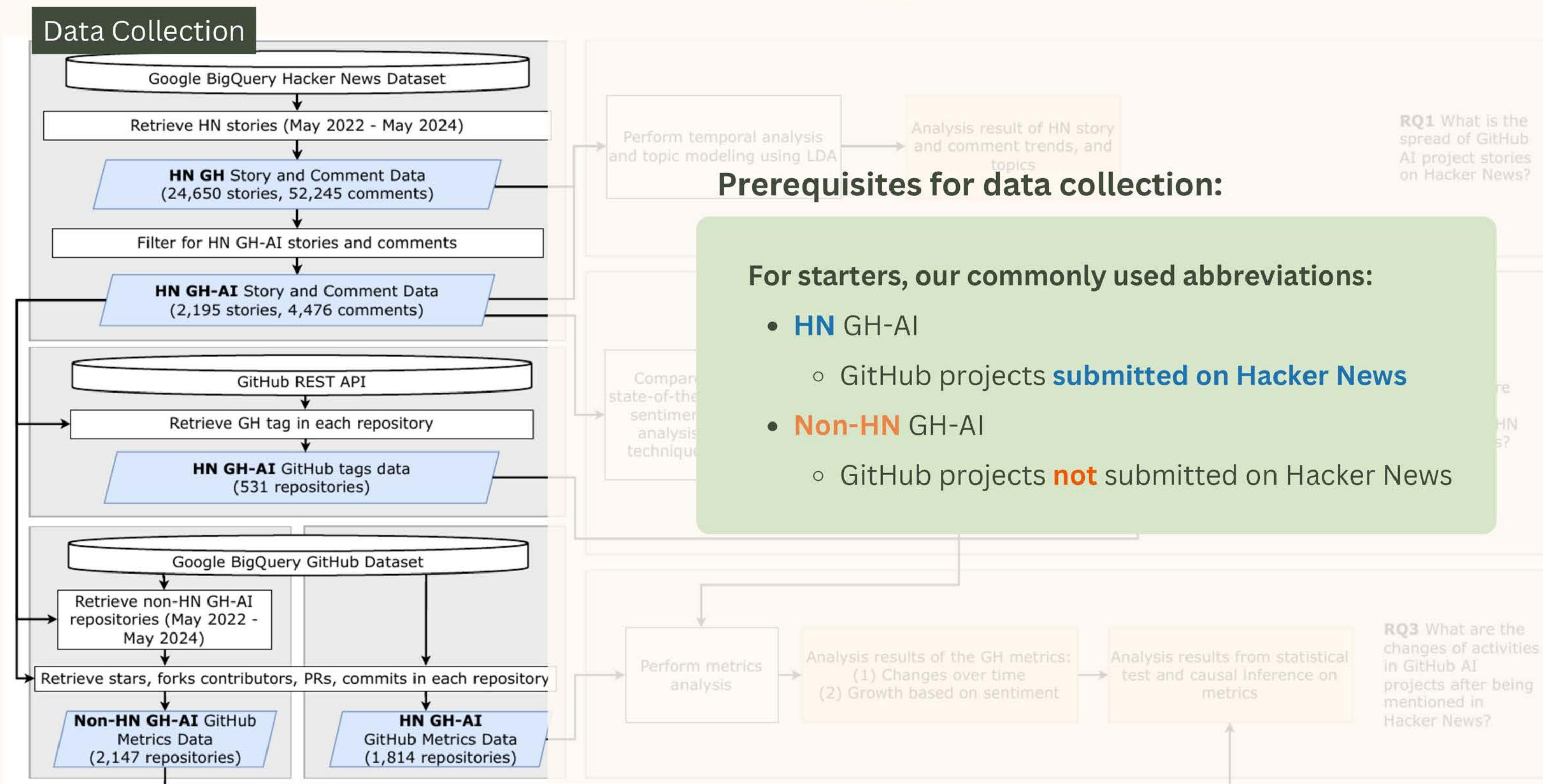
The background features abstract wavy patterns in gold and green on the right side, and a grid of dots on the left side.

Methodology Overview

Data Collection



Methodology Overview

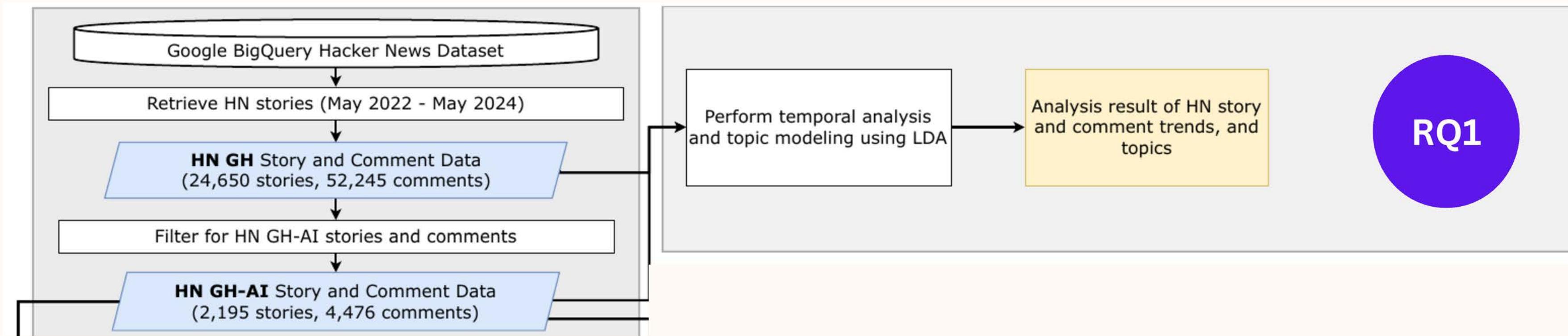




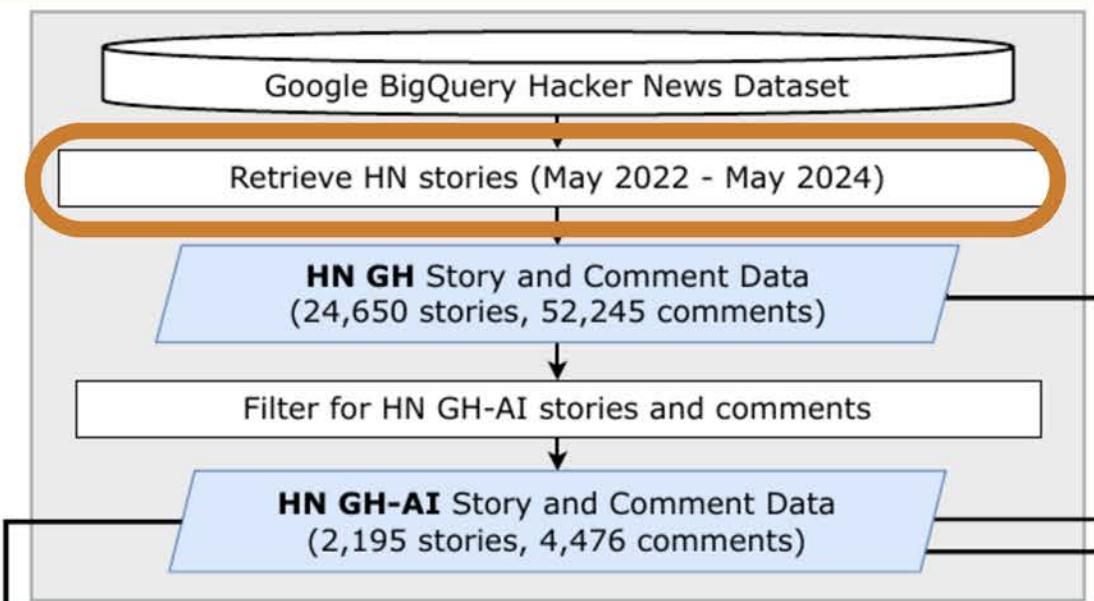
Results

An abstract graphic featuring a series of thin, light brown wavy lines that curve and flow across the slide, creating a sense of motion and depth.

Methodology Overview



Data Collection: Collecting HackerNews Data



Our BigQuery story **script** accomplishes the following:

- Retrieved from May 2022 to May 2024
 - From **6 months prior to ChatGPT release** (Nov 30th 2022) to near-present
- Filtering by checking story URL for **GitHub repository URLs**

Story Example

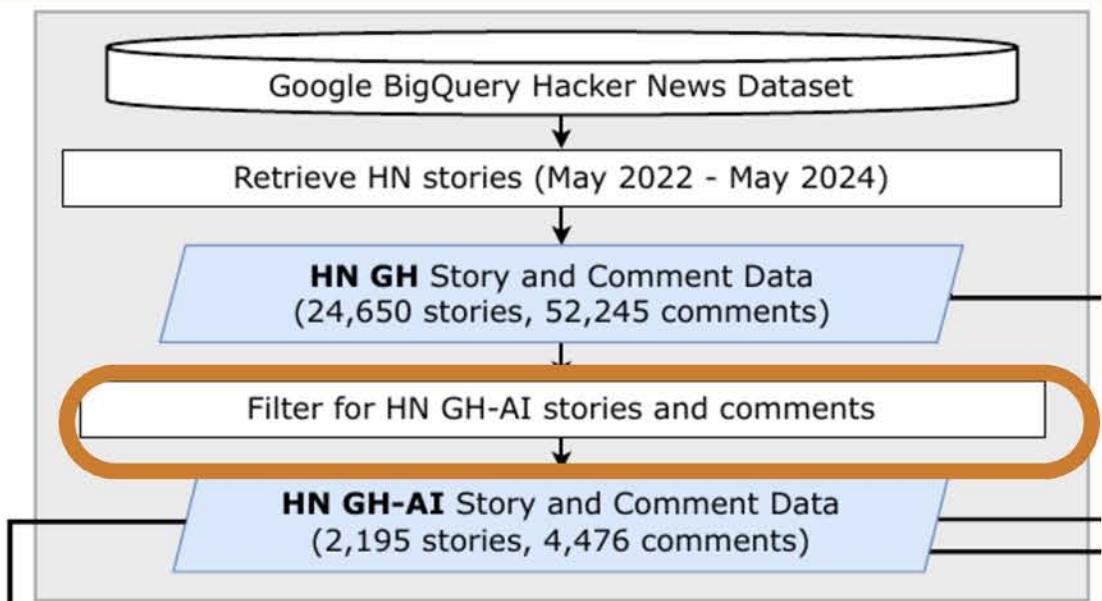
A screenshot of a Hacker News story page. The story title is "Text generator: an Obsidian plugin that generates text using GPT-3 (OpenAI) (github.com/nhaouari)". The story has 31 points and was posted on Oct 19, 2022. A callout bubble highlights the GitHub repository URL in the story text, with the text "Contains GitHub repository URL". Another callout bubble highlights the word "openai" in the title, with the text "Contains keyword 'openai' in the title".

- Filter for stories with title containing our predefined **keywords related to AI/LLM**
 - e.g. "llm", "openai", "nlp", "chatgpt", etc. in the story title

We will have 2 story datasets:

- **HN GH Stories**
- **HN GH-AI Stories**

Data Collection: Collecting HackerNews Data



- Fetch comment details for each story

- Because story dataset contains only comment IDs in each story
- Crucial details include text, created datetime

- Criteria:

- Excludes “dead”, empty comments
- Be a direct reply to the story discussion (rather than a reply to another comment)

Comment Example



on May 11, 2022 | parent | context | favorite | on: AI-engineered enzyme eats entire plastic container...

I've always sort of thought that, given the amount of plastics that's in our oceans, together with a bunch of hungry microorganisms, odds are eventually one will evolve that eats plastics and it will *flourish*. Prima facie this may even sound like a good development.

Except now you're faced with a world where plastics rot, not just the plastics we want to go away, but the plastics we want to stay around. We use plastics in a lot of places where this would be a Bad Thing, such as electrical insulators.

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

Temporal Trend Analysis

- Analyze the **frequency of stories and comments** posted over time
- Analyze the frequency of each title **keyword** in stories **over time**

Additional analyses:

- #Days before HN submission** distributions
- Relationship between **HN poster** and **GitHub repository owner** based on **HN About section**

About section on Hacker News

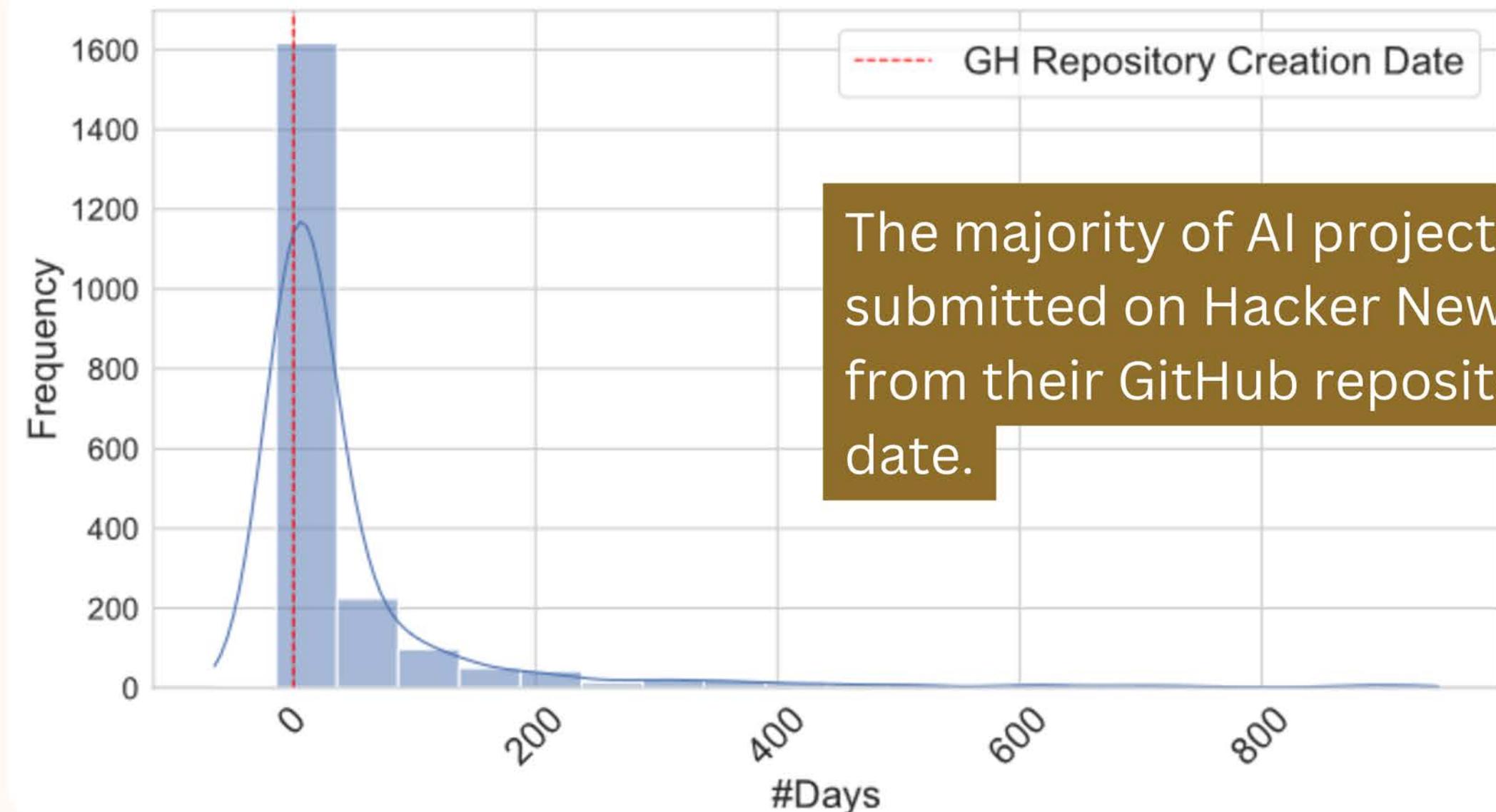
```
user:      muehuu
created:   9 months ago
karma:     1
about:    my profile might contain github username
```

Topic Modeling

Extract and interpret topics from the story titles using **LDA model**

- 1 **Preprocess story titles**
 - Lemmatization, Stemming
 - Remove frequent words (“AI”, “Show HN”) to emphasize prominent themes
- 2 **Optimize parameters and evaluate using t-SNE**
 - Number of topics** (by analyzing coherence scores)
 - Learning rate** and **decay, coherence score, etc.**
- 3 **Train LDA model based on the optimal parameters**
 - For further evaluation of the LDA model’s performance
 - Visualize & evaluate** topic clusters using **pyLDAvis** to **cross-check** for any overlapping themes



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

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

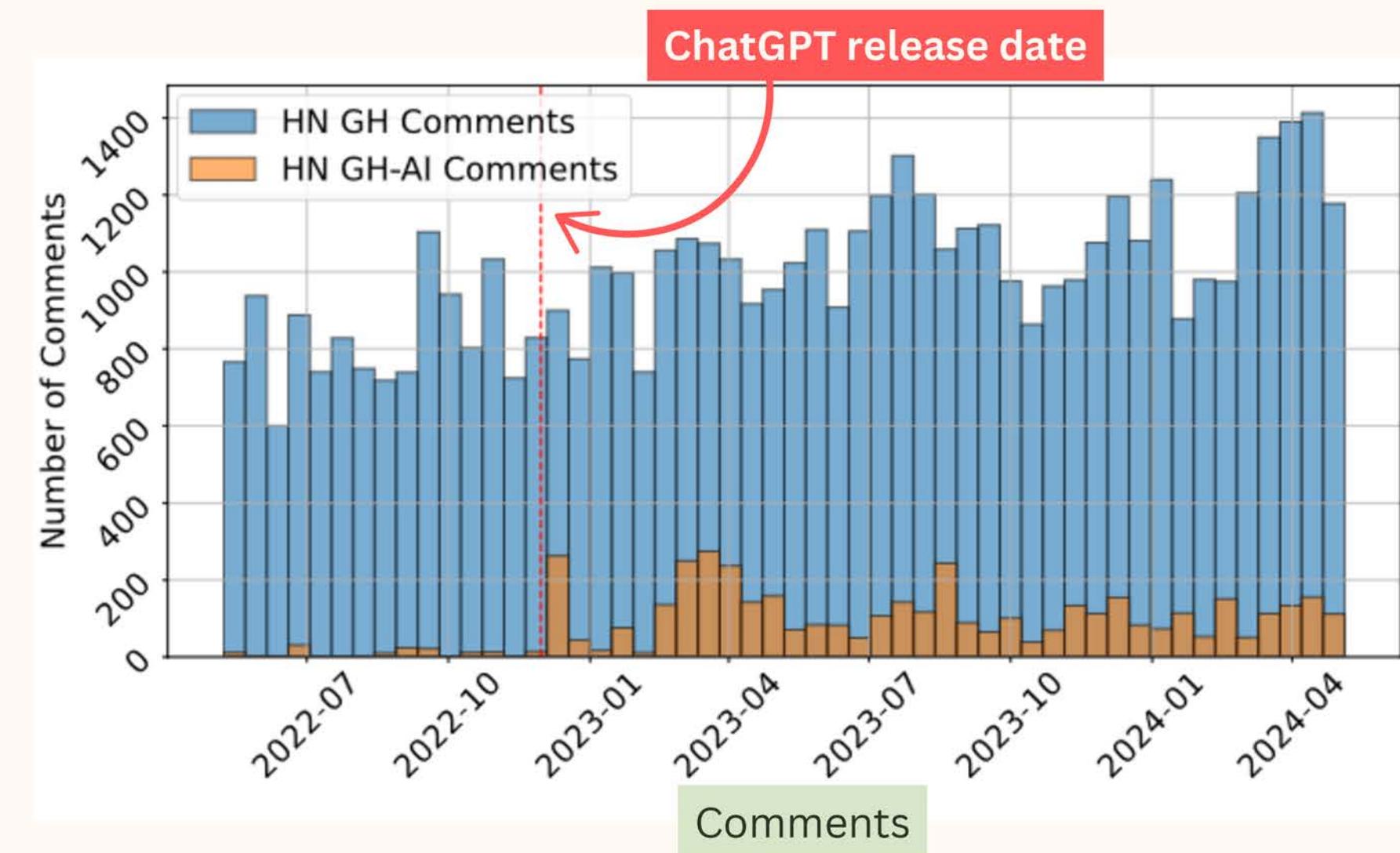
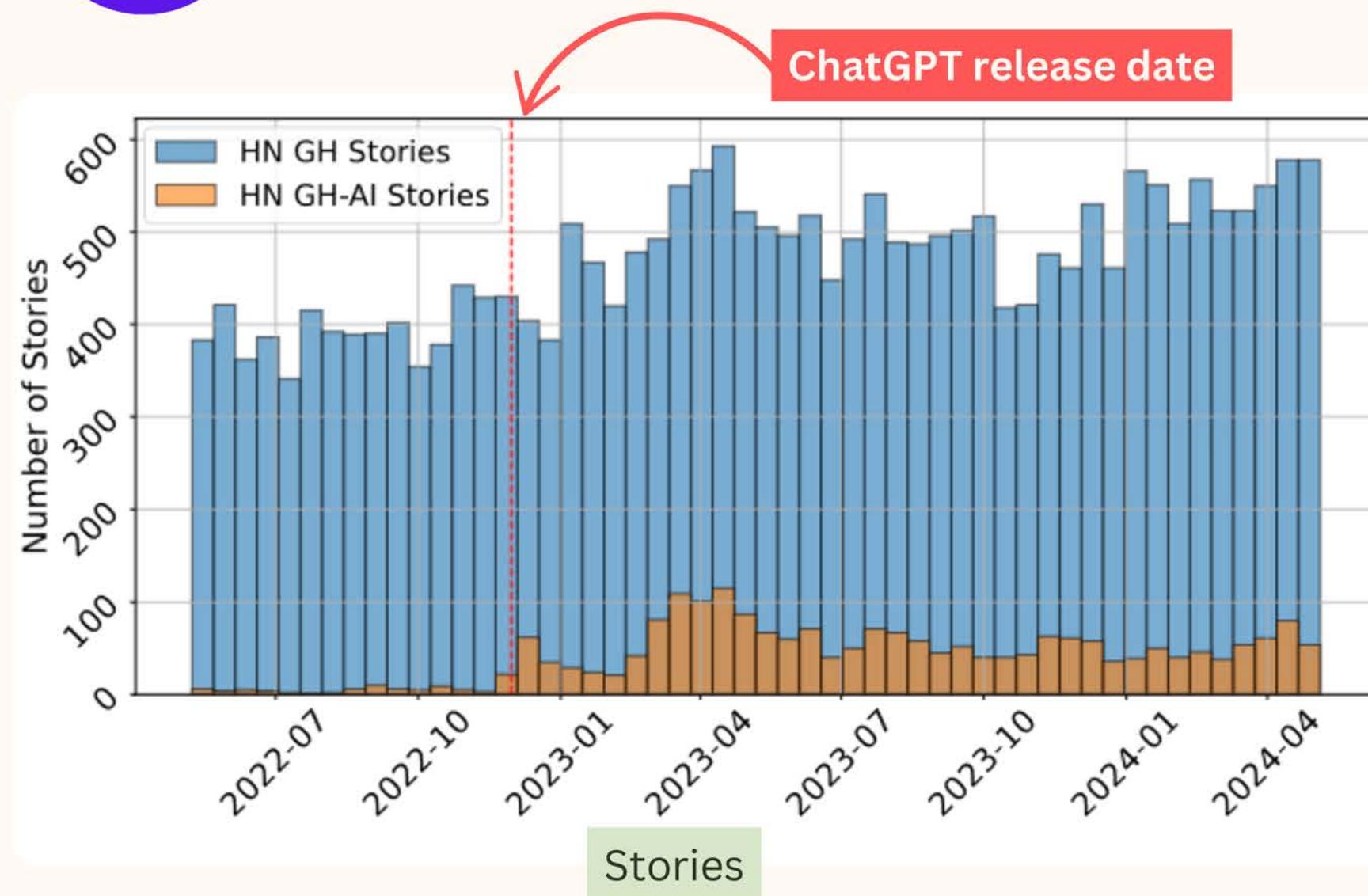
Time to HN submission vs. GH repo creation date

- 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

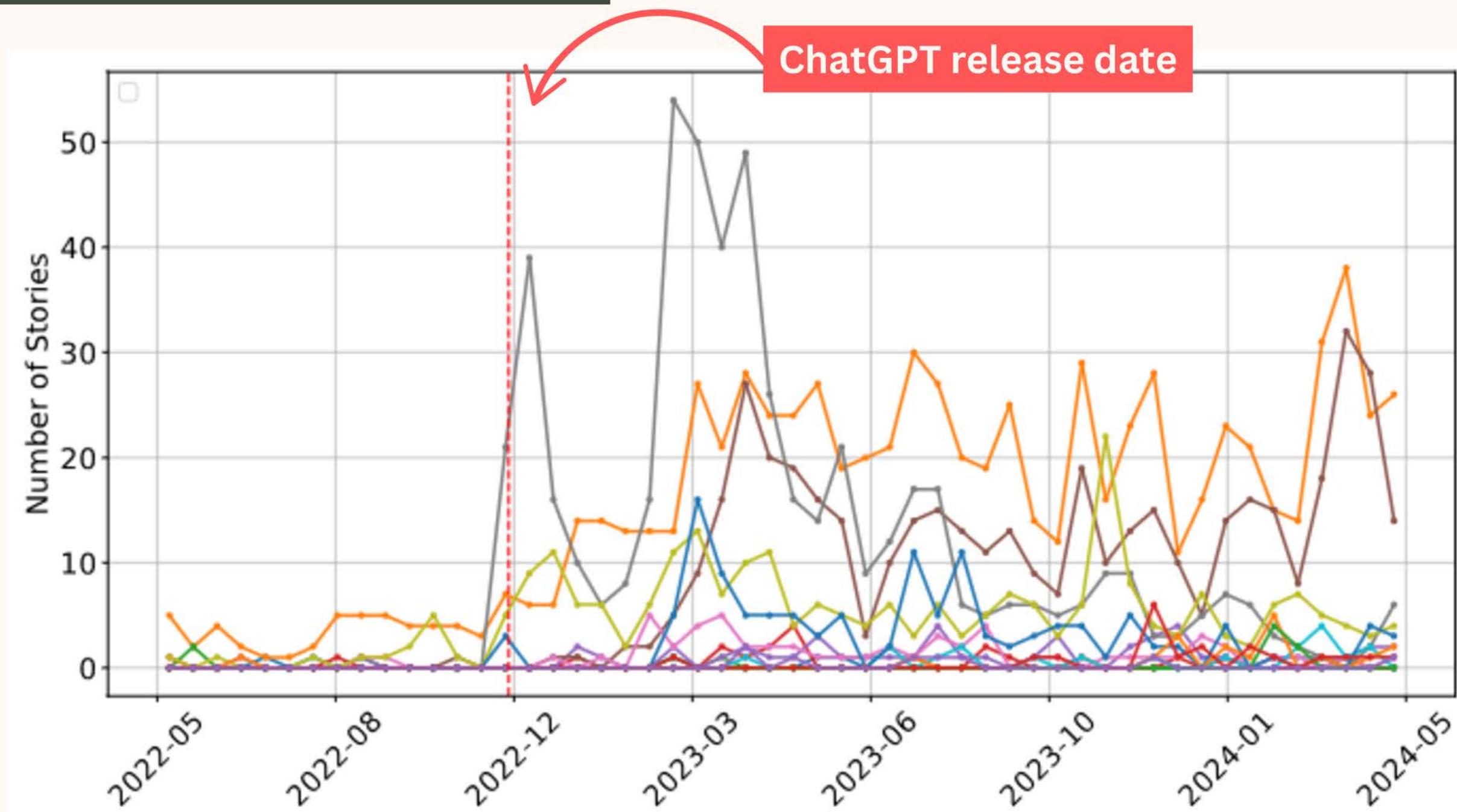
Observation 2: At least 19% of the posters of HN GH-AI stories were contributors to the GitHub project themselves, implying self-promotion

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?**

- Language Model
- LLM
- Chatbot
- ChatGPT
- Artificial Intelligence
- AI
- Natural Language Processing
- NLP
- Gemma
- Mistral
- Cohere
- OpenAI
- Claude
- Llama
- Gemini



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

• INTRODUCTION

• BACKGROUND & RELATED WORKS

• RQ1 - SPREAD

• RQ2 - REACTION

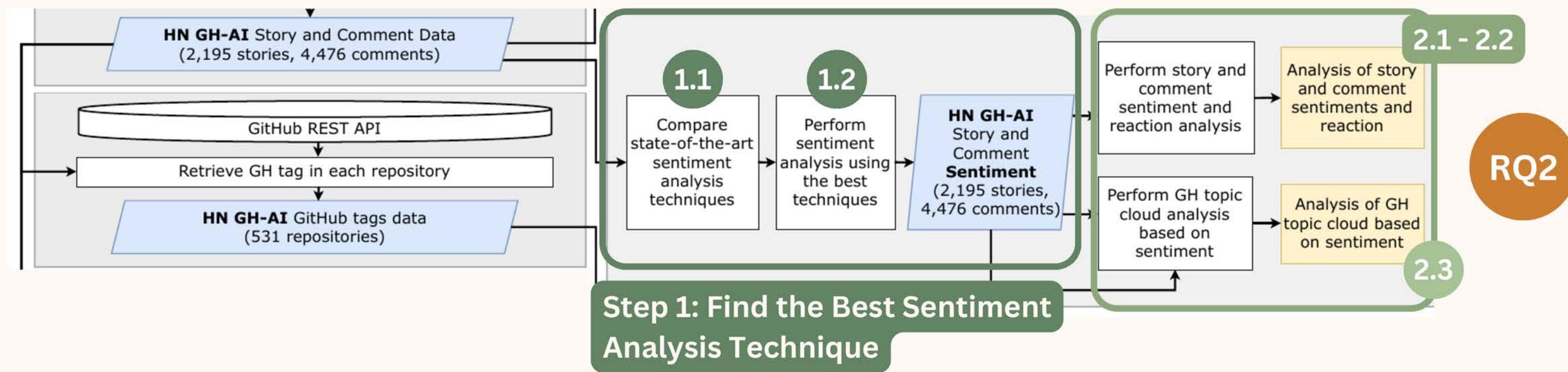
• RQ3 - ACTIVITIES



Methodology & Results for RQ2

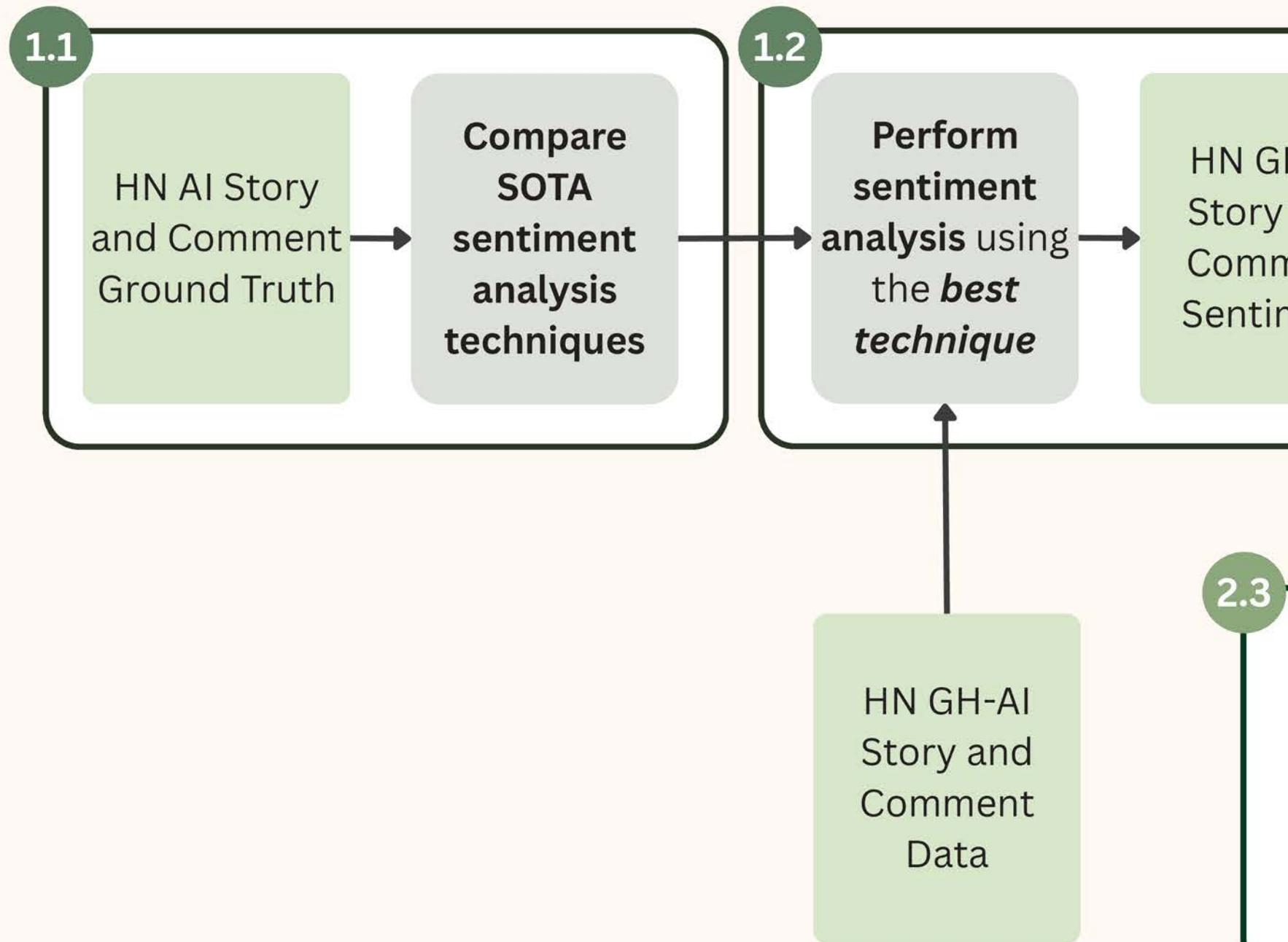


Methodology Overview

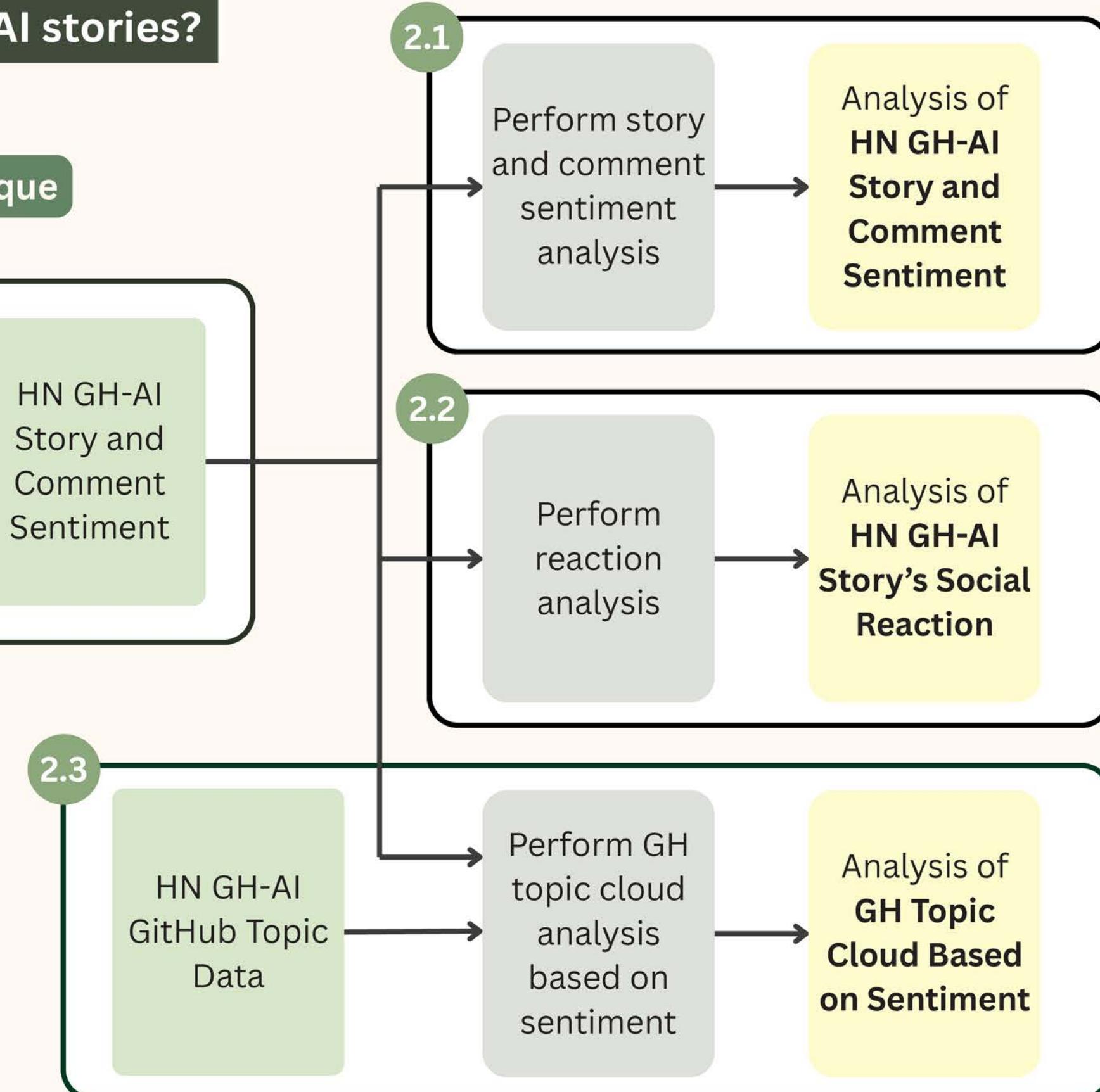


RQ2 What are the social reactions to HN GH-AI stories?

Step 1: Find the Best Sentiment Analysis Technique

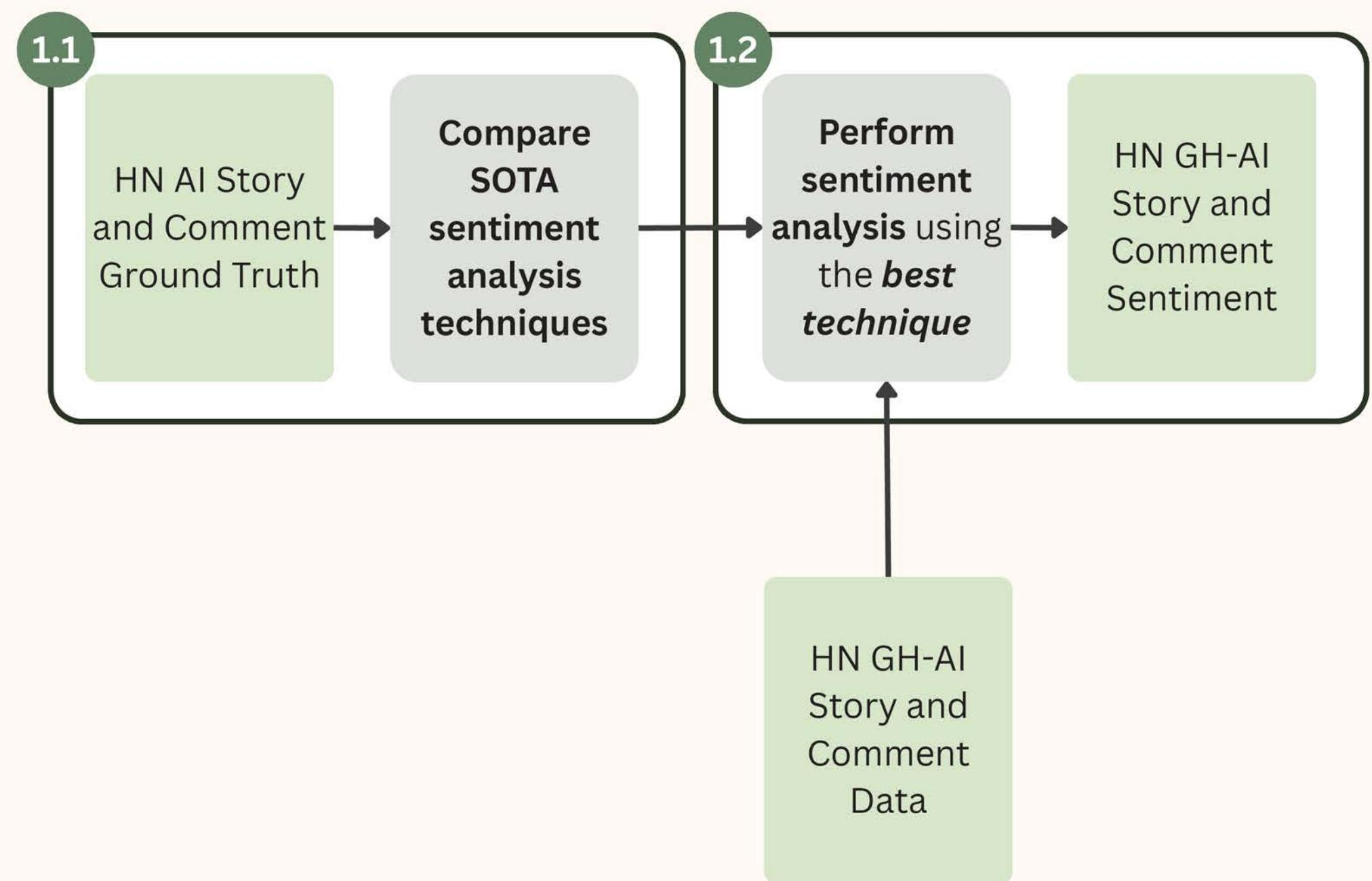


Step 2: Analysis to Answer RQ2



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

Step 1: Find the Best Sentiment Analysis Technique



Data Creation: Constructing Ground Truth

The **ground truth dataset** was constructed for

Step 1: Find the Best Sentiment Analysis Technique

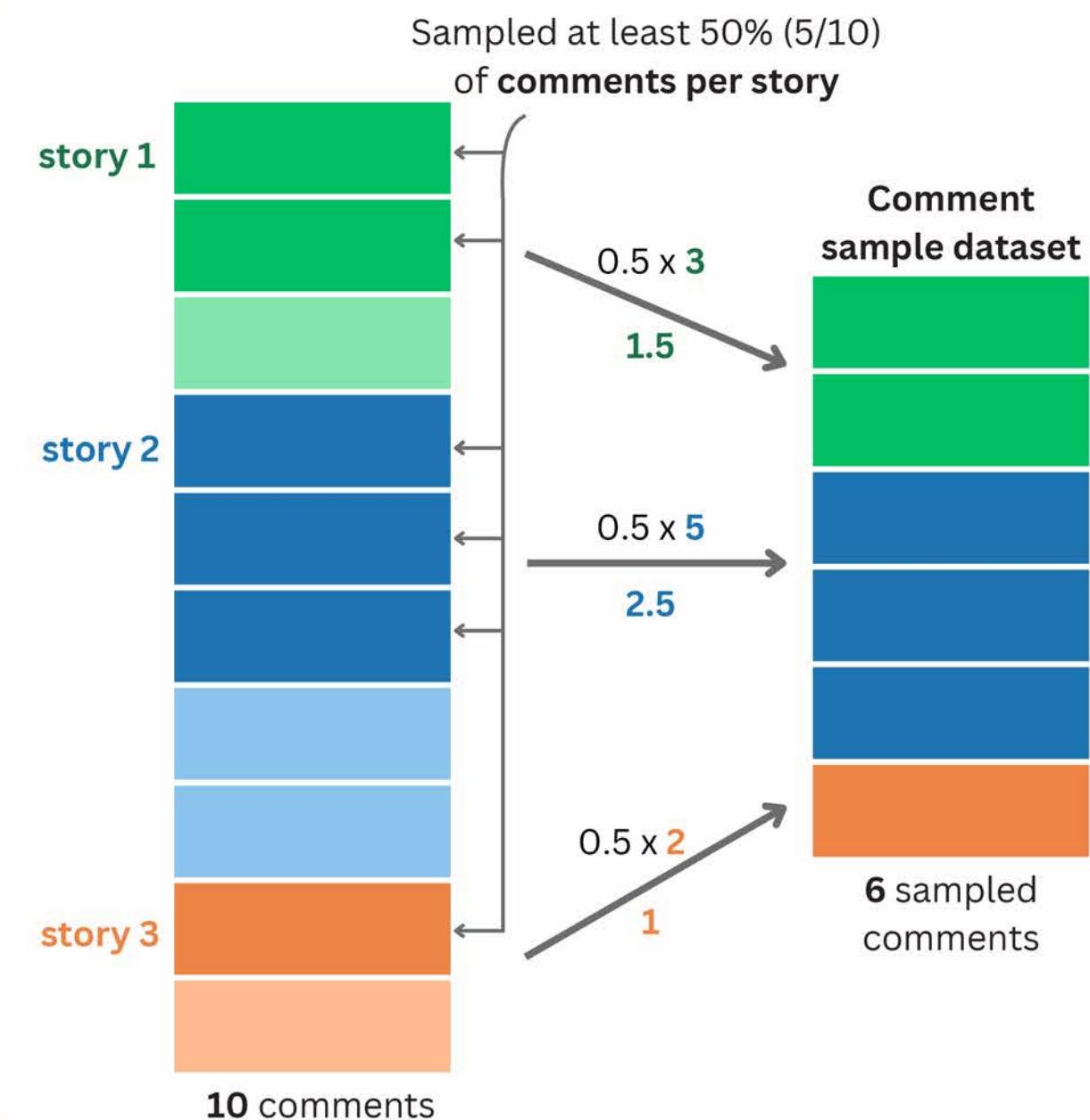
1. **385 stories** were **uniformly sampled** from
HN AI stories with a score ≥ 20

2. **385 comments** were **stratified sampling** from
218 uniformly selected HN AI stories

385 is the minimum number of samples needed to have a
95% confidence level and **5% margin of error**.

Stratified Sampling Process

Stratified sampling **5** comments from **3** stories.



Data Creation: Constructing Ground Truth

3. Two human investigators labeled all sampled stories and comments **independently**



4. Cohen's kappa agreement between the two investigators was performed.



5. Investigators resolved disagreements and generated **consensus labels**

Output: Story & Comment Ground Truth

Labeling Process

Labeling 3 stories and 6 comments

Comment sample dataset

1	1	1
1	0	0
1	1	1
1	0	0
0	1	1
-1	-1	-1



6 sampled comments

Consensus labels

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

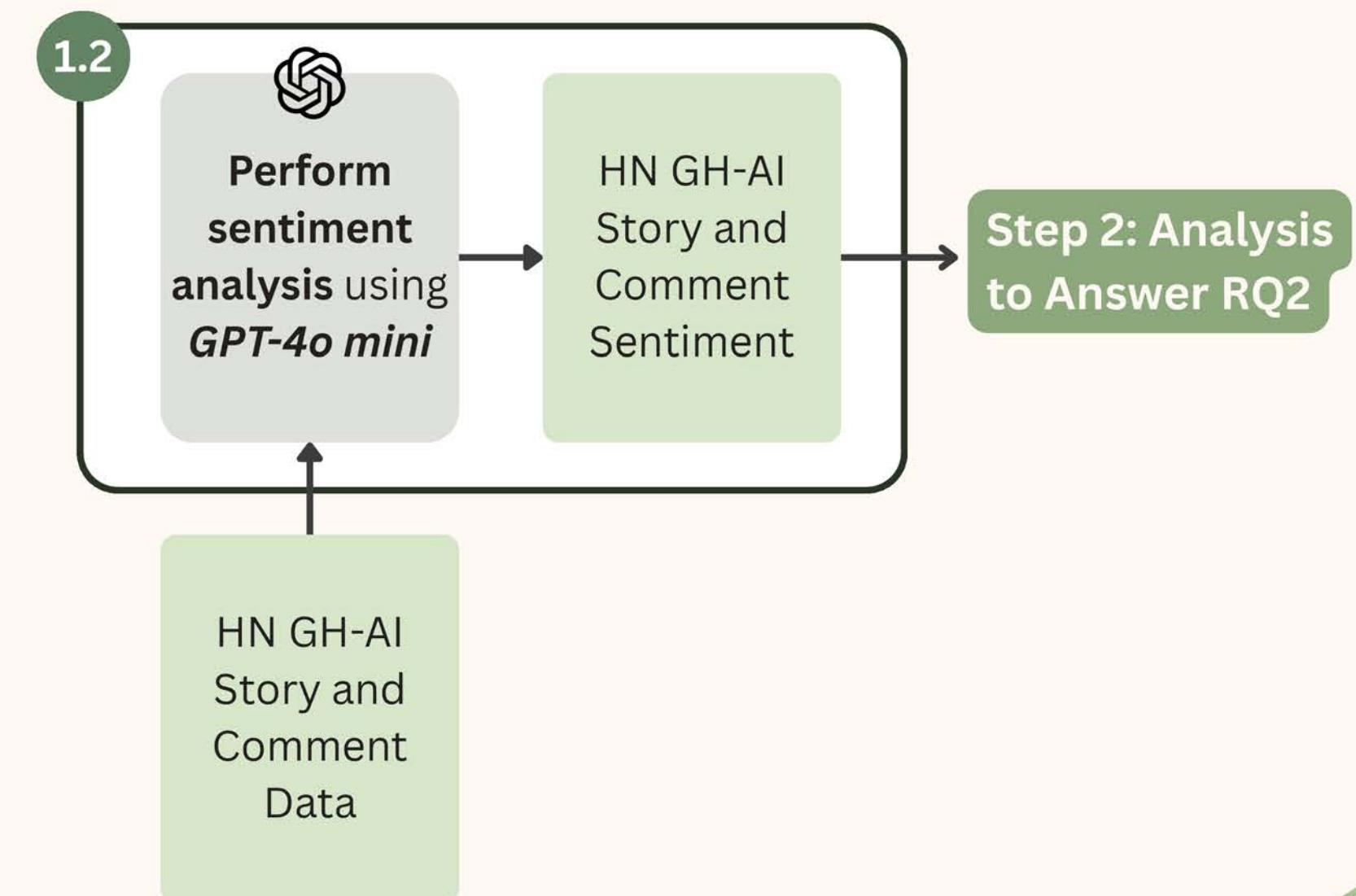
Select to perform
sentiment
analysis

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

1.2 Perform Sentiment Analysis

- **HN GH-AI stories and comments** that had HTML tags removed was feed into the prompt to **GPT-4o mini**

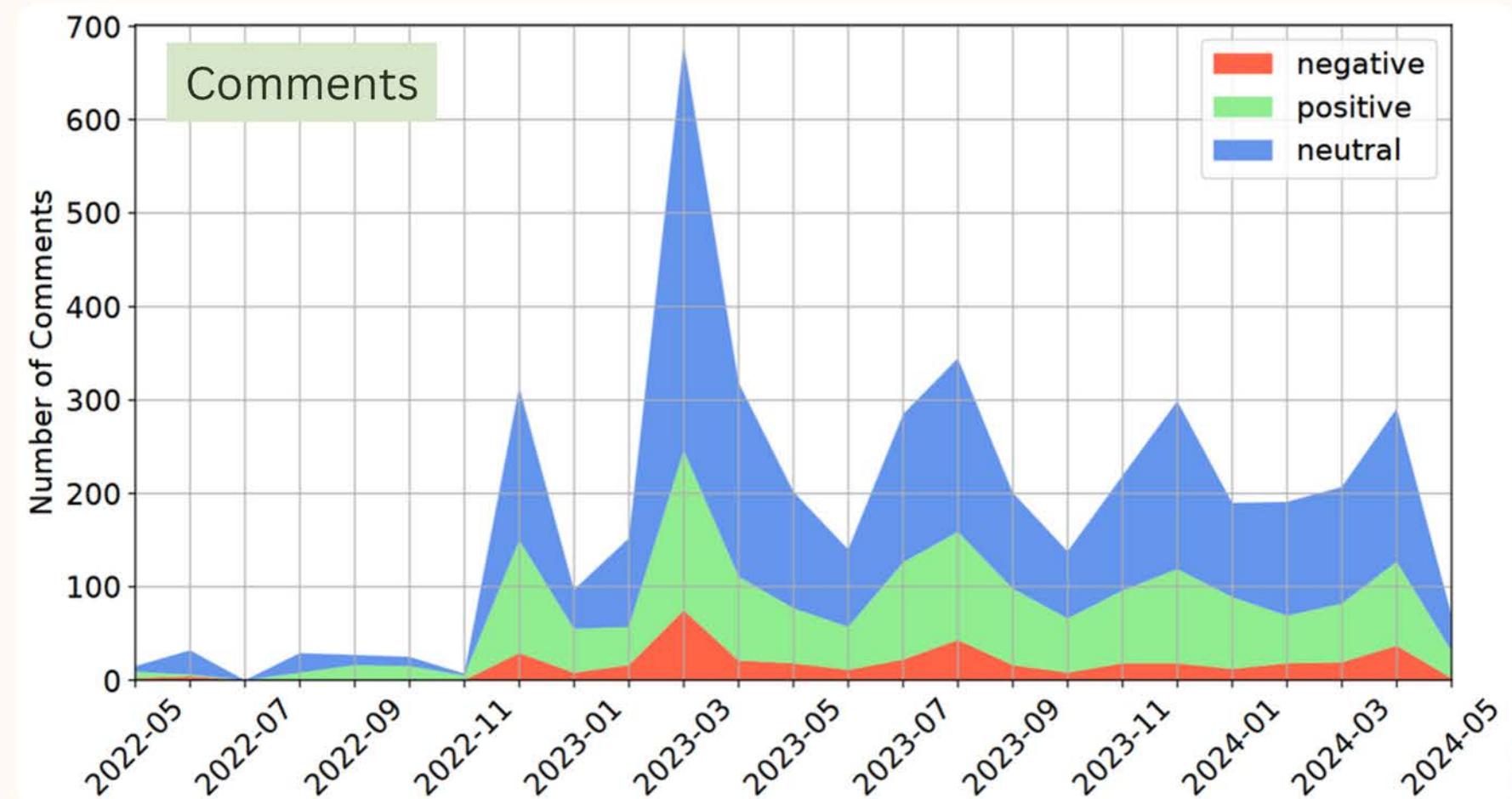
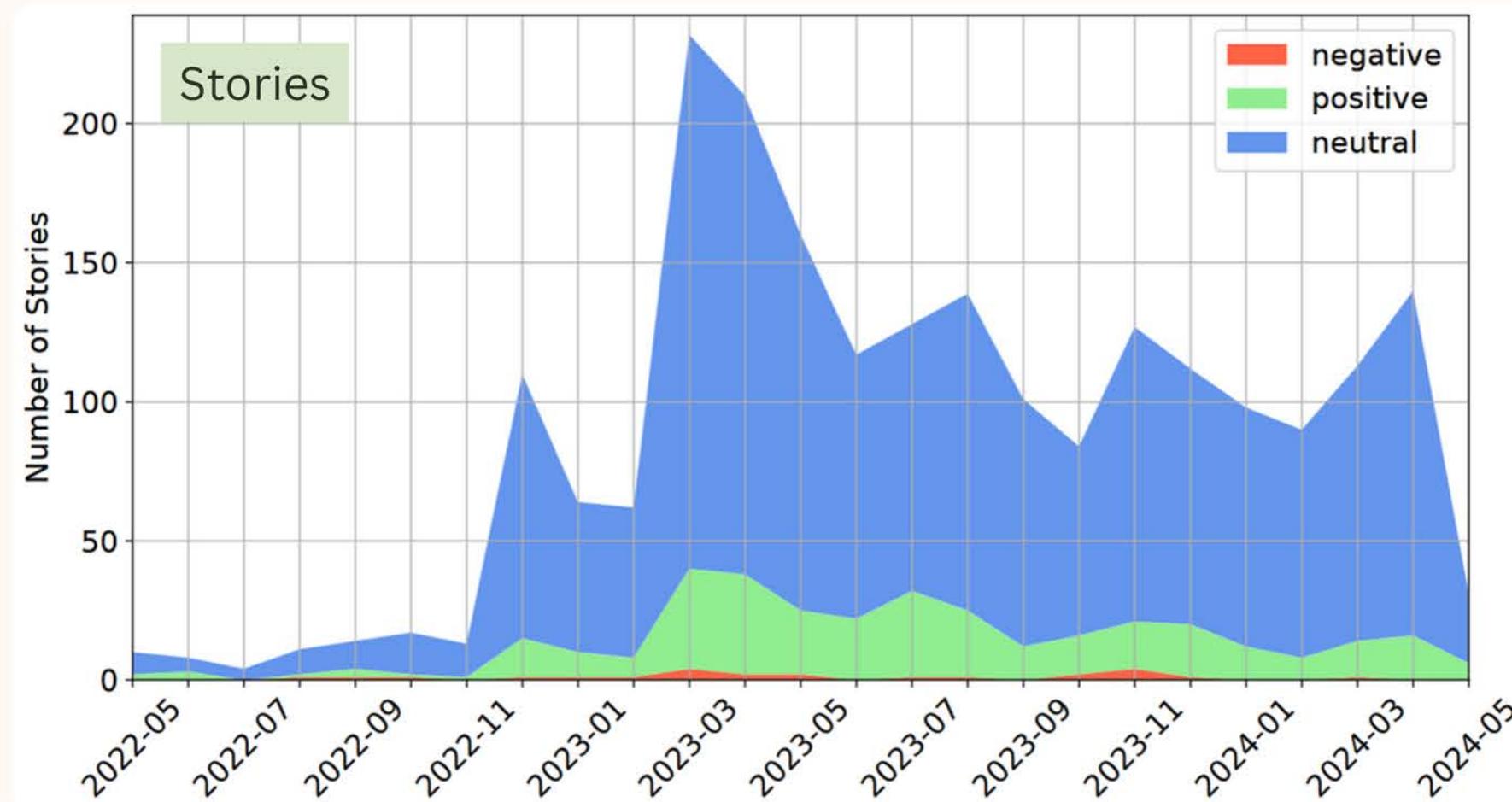
- The output is **HN GH-AI story and comment sentiment dataset**



RQ2

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

2.1 Analysis of HN GH-AI Story and Comment Sentiment

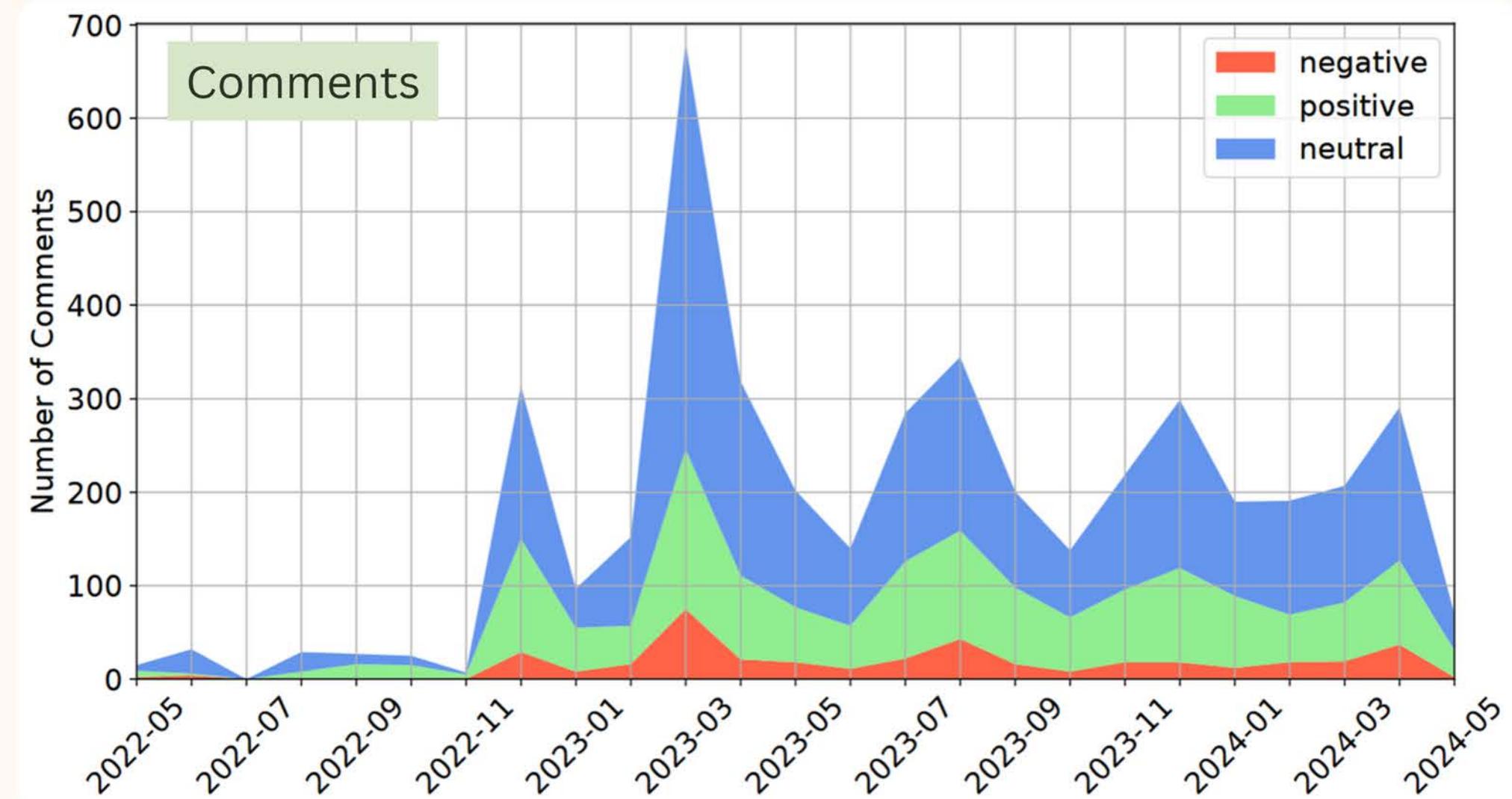


Observation 5: Hacker News community tend to be **more positive than negative** towards stories related to AI project on GitHub.

RQ2

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

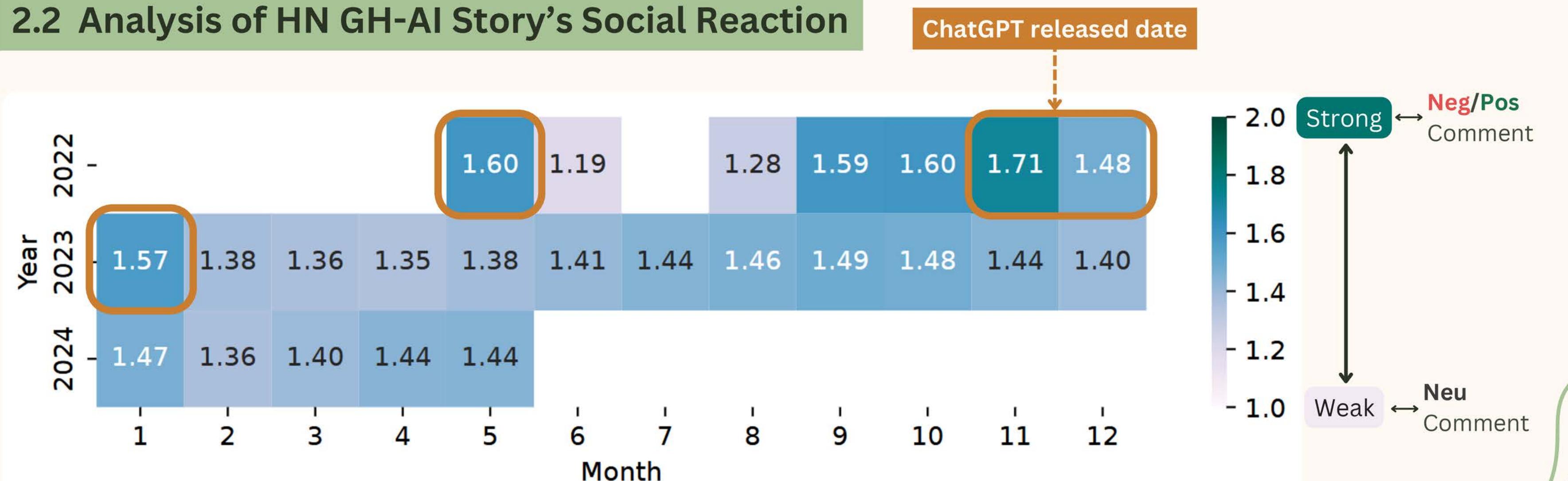
2.1 Analysis of HN GH-AI Story and Comment Sentiment



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?

2.2 Analysis of HN GH-AI Story's Social Reaction



Observation 6: Hacker News discussions show a few strong reactions in May 2022 and around ChatGPT's release from Sept 2022 to Jan 2023.

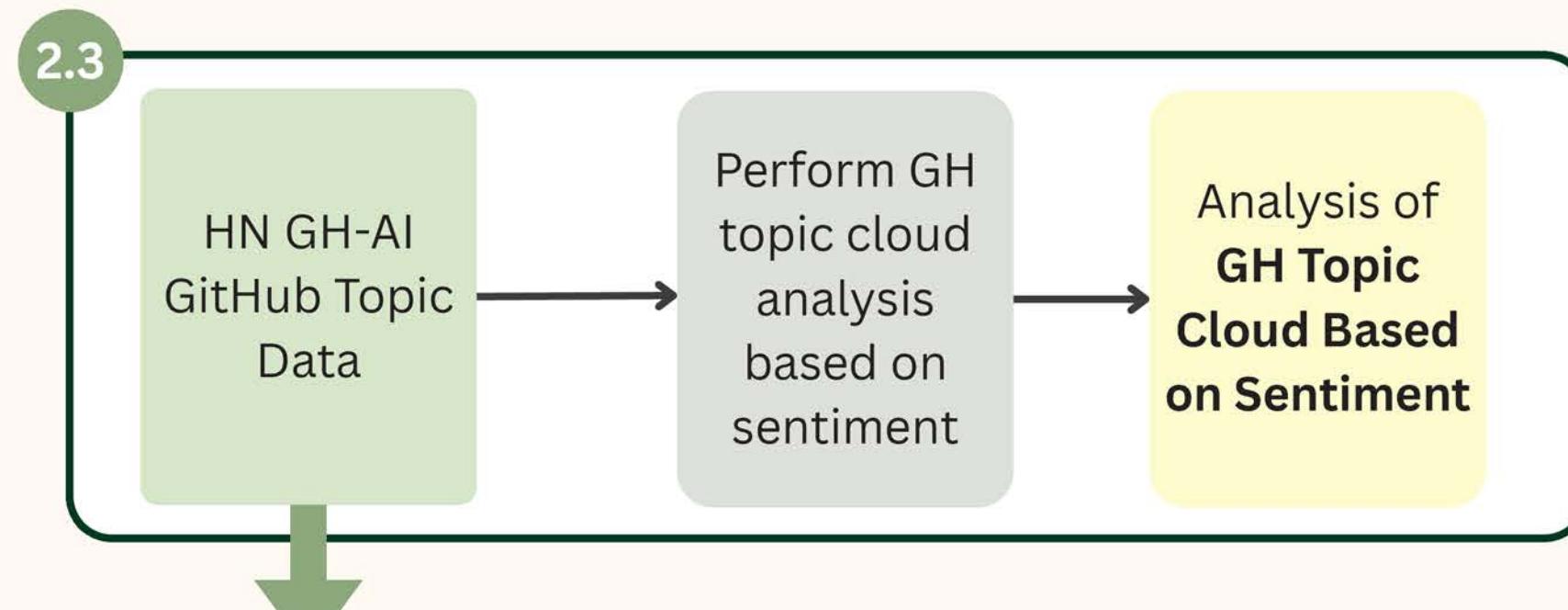
RQ2 What are the social reactions to HN GH-AI stories?

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>

Data Collection: Collecting GitHub Topics data

2.3 Analysis of GH Topic Cloud Based on Sentiment

- GitHub topics of each repository in HN GH-AI stories were retrieved from GitHub REST API.



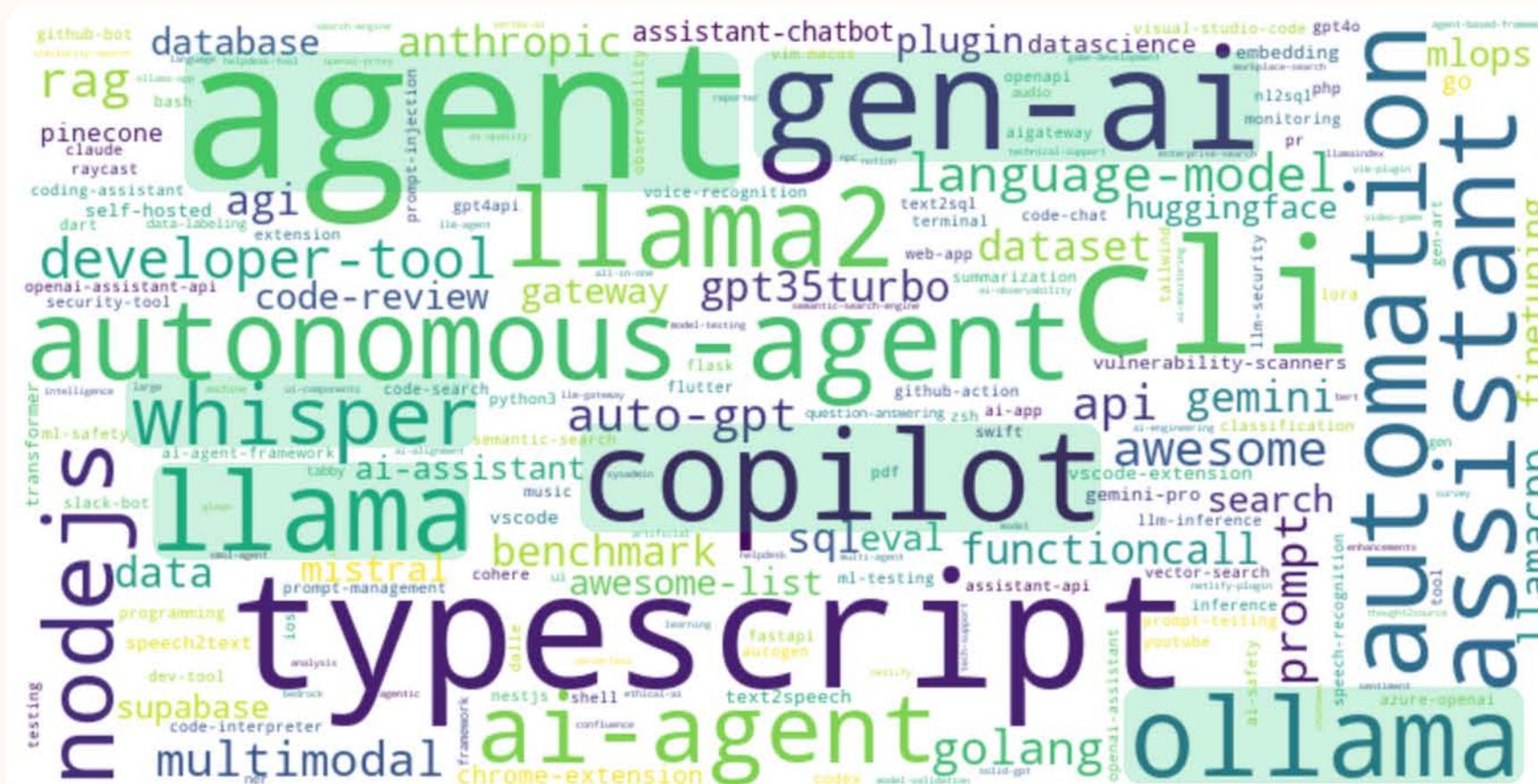
- 531 valid repositories with GitHub topics were identified.



An example of GitHub topics from ggml-org/llama.cpp repository

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

2.3 Analysis of GH Topic Cloud Based on Sentiment



GitHub Topic Cloud Exclusive to Positive Sentiment Repositories



GitHub Topic Cloud Exclusive to Negative Sentiment Repositories

Observation 7: GitHub topics exclusive to positive-sentiment projects include AI tools such as ‘llama’ and ‘whisper’. Negative-sentiment projects are about impersonation or replacing humans.

• INTRODUCTION

• BACKGROUND & RELATED WORKS

• RQ1 - SPREAD

• RQ2 - REACTION

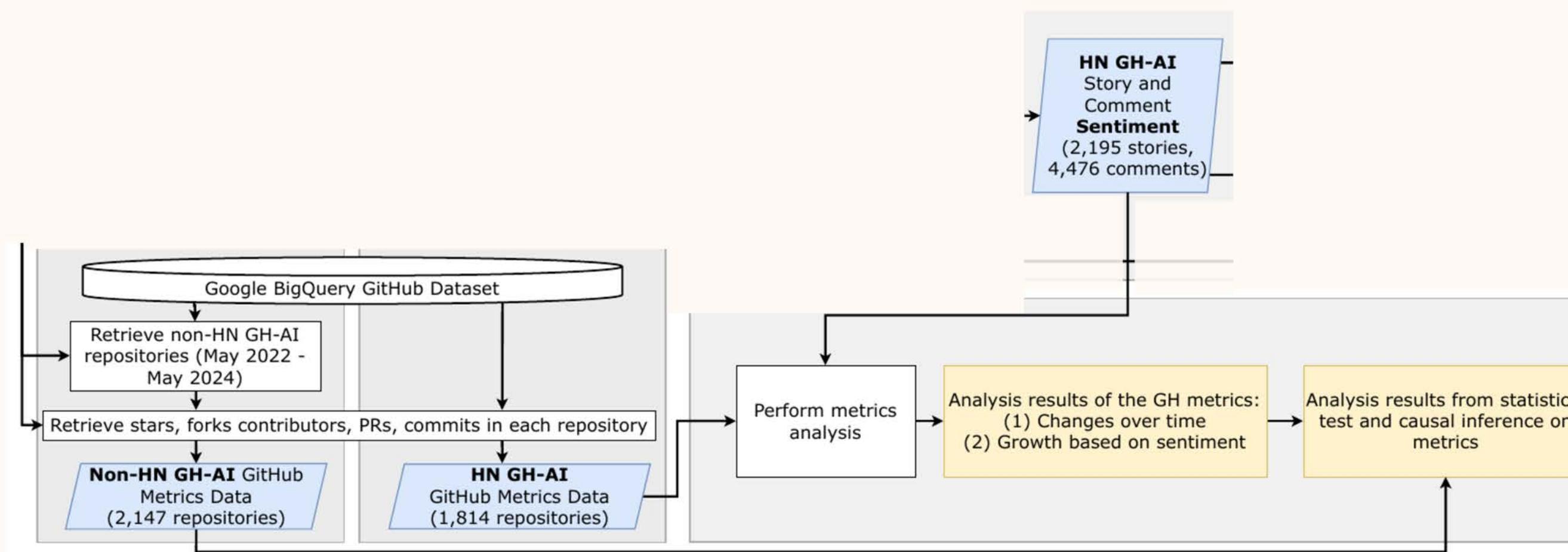
• RQ3 - ACTIVITIES



Methodology & Results for RQ3

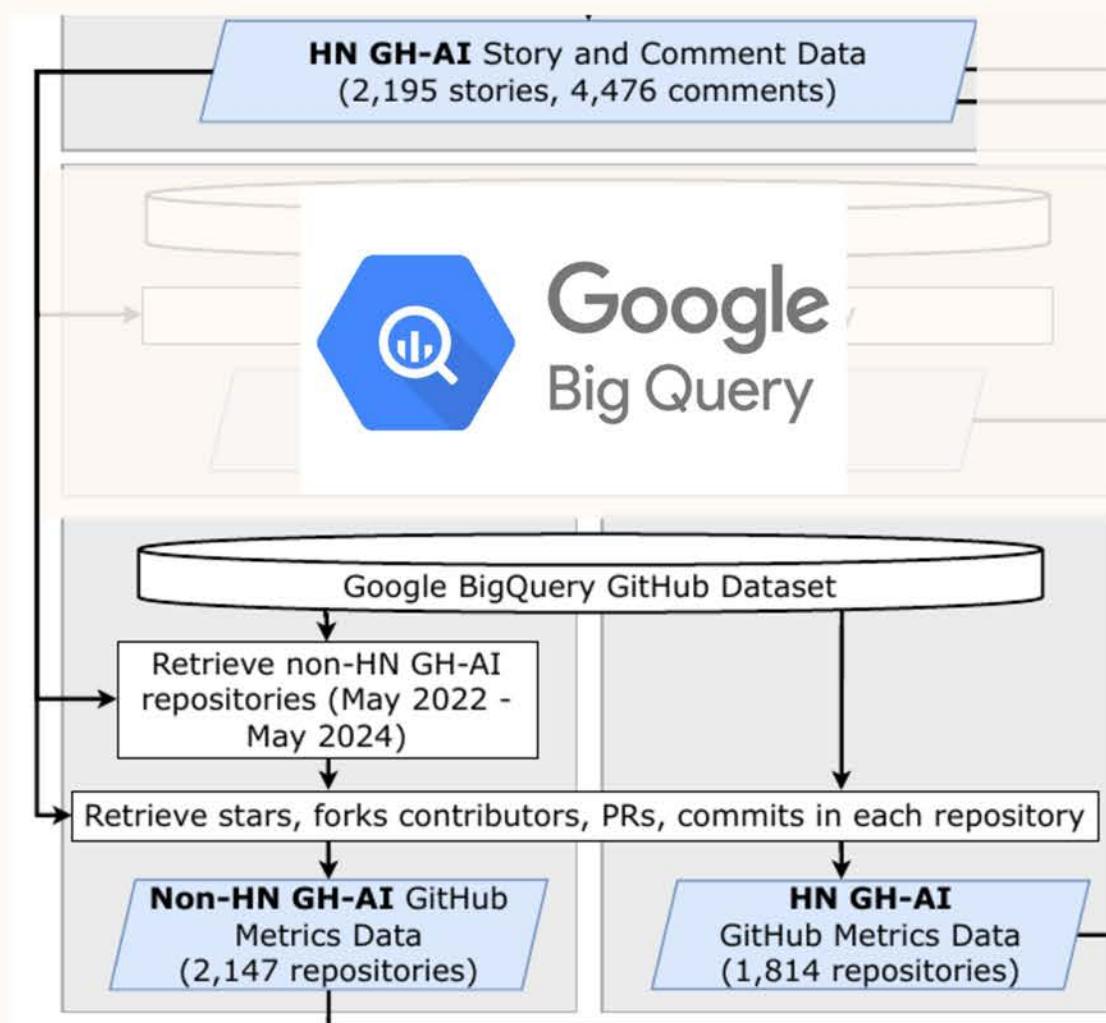


Methodology Overview



RQ3

Data Collection: Collecting Historical GitHub Metrics Data



For non-HN GH-AI (For RQ3):

- Collect a **list of AI projects** from GHArchive (BigQuery dataset)
- **Filter out overlaps with HN** projects using URLs from the HN GH-AI dataset

For HN GH-AI:

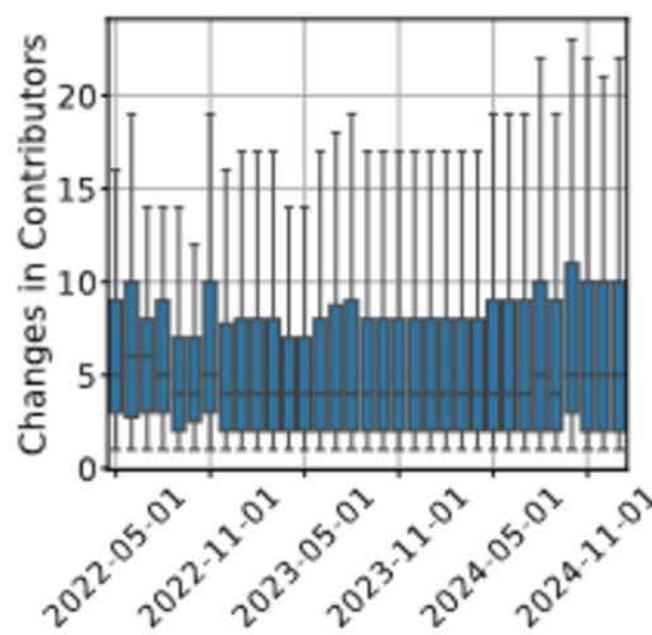
- Already contain **GitHub repository URLs** in the story dataset previously.

Then retrieve monthly metrics for both non-HN GH-AI and HN GH-AI repository URLs

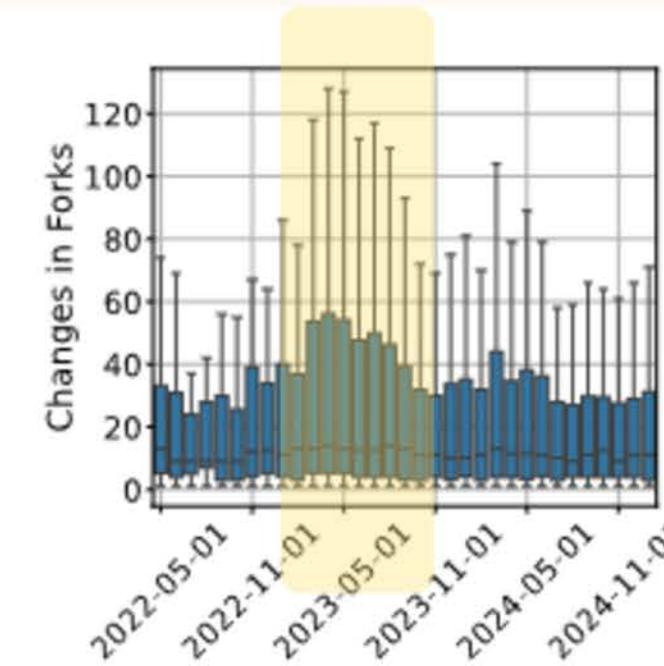
- **Accumulate** on each day in each month to obtain **monthly metrics**

RQ3

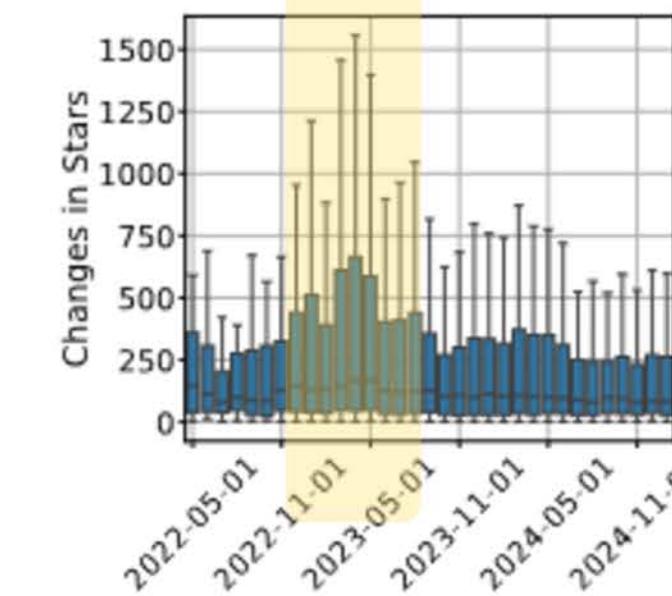
What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?



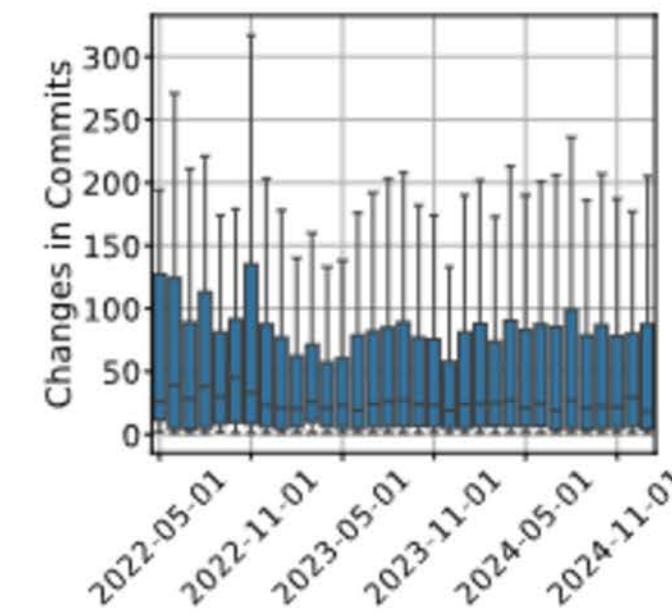
(a) Contributors



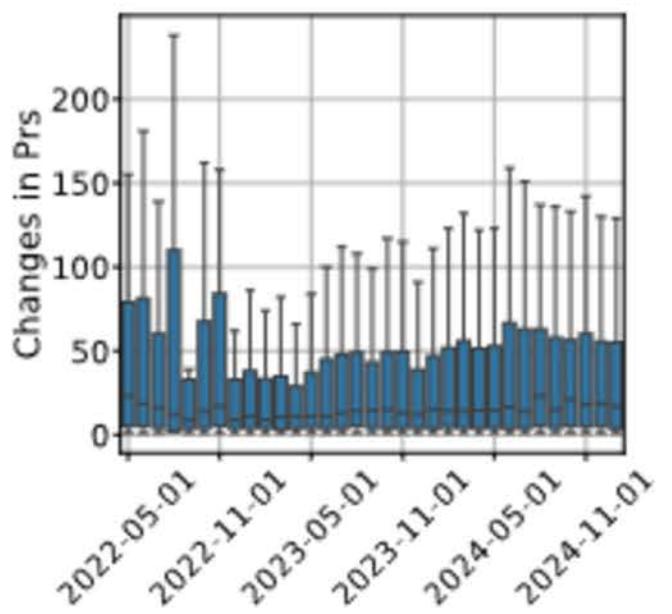
(b) Forks



(c) Stars

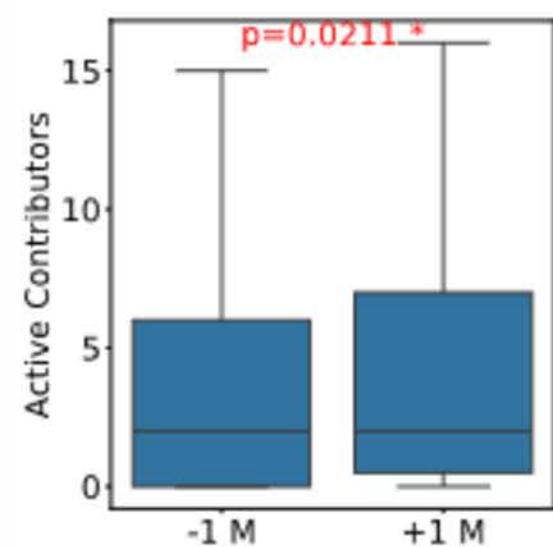


(d) Commits

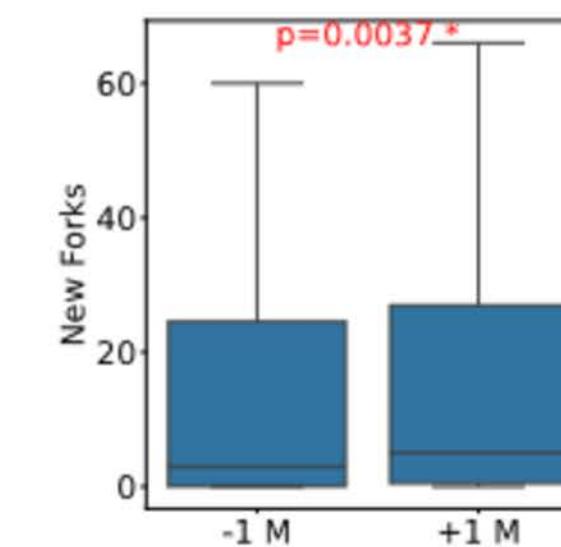


(e) Pull Requests

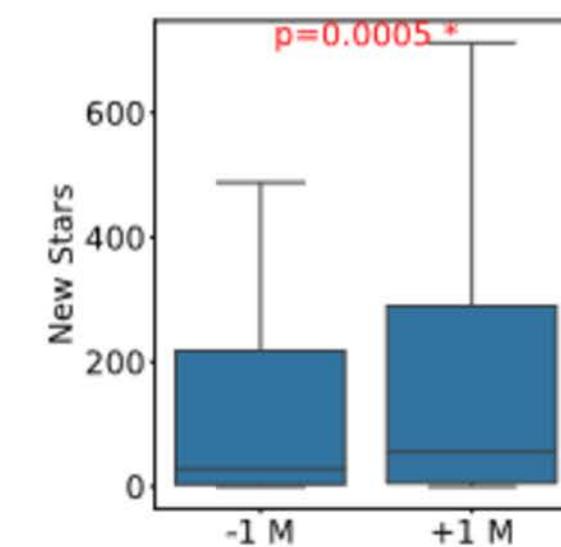
Observation 7: GitHub AI projects on Hacker News received an **increase in activity on stars and forks shortly after the release of ChatGPT.**

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

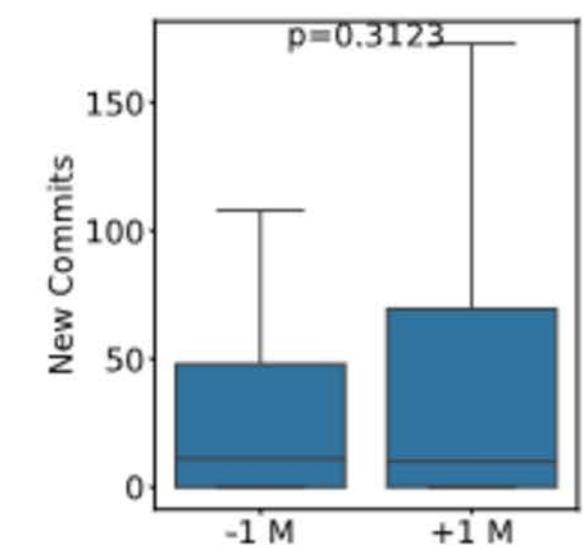
(a) Contributors



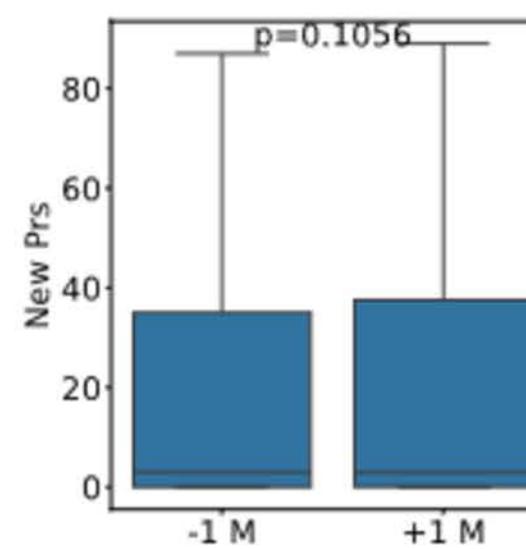
(b) Forks



(c) Stars



(d) Commits



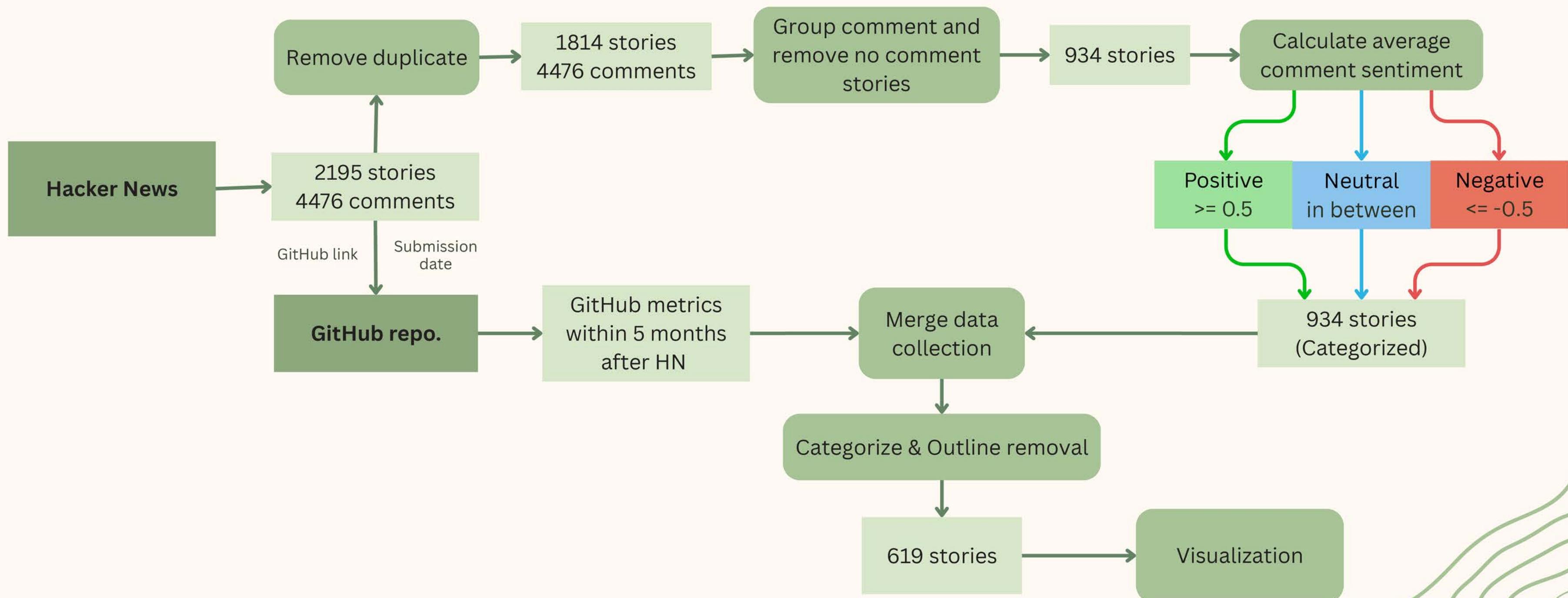
(e) Pull Requests

RQ3

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

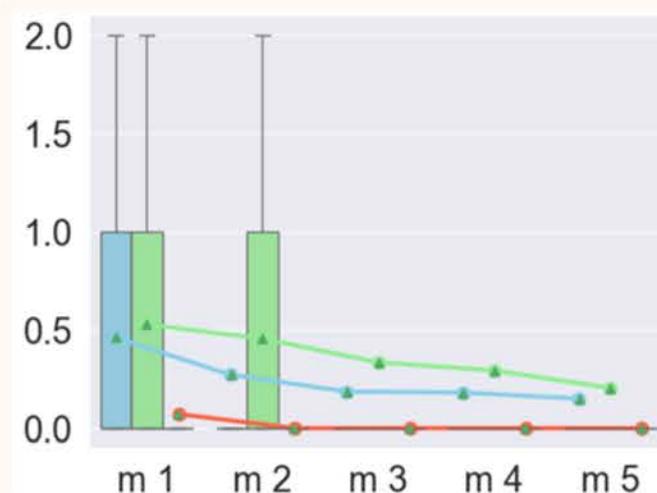
Potential Impact of Hacker News Exposure

Metric changes based on sentiment on Hacker News

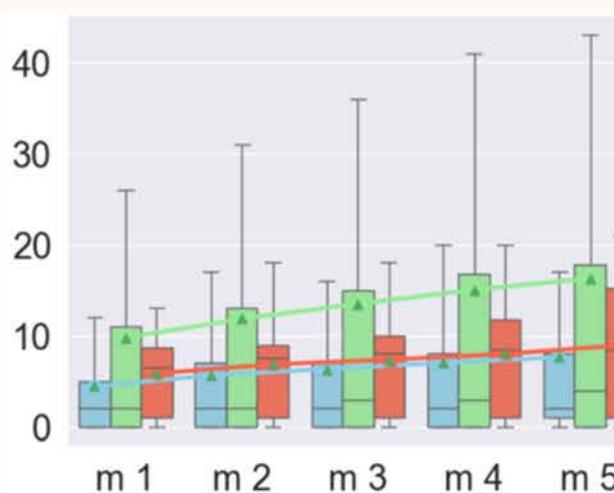


RQ3

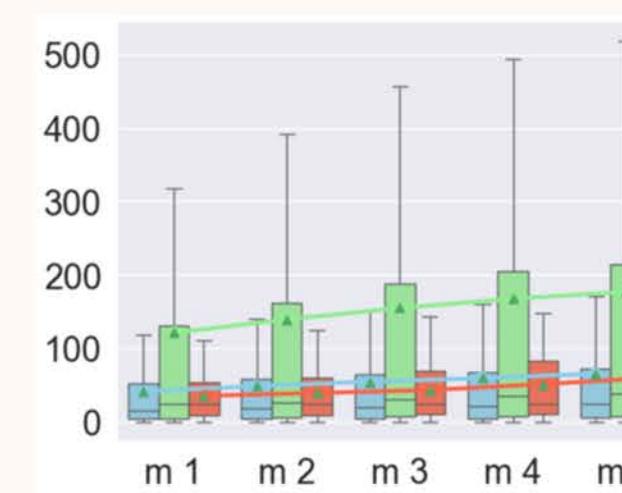
What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?



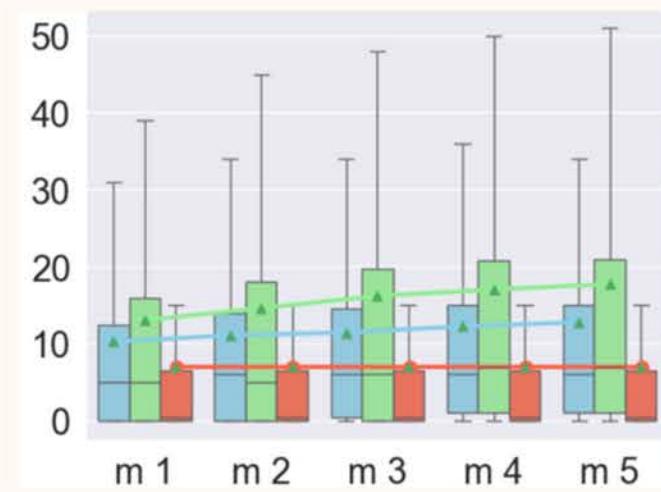
Contributors changes



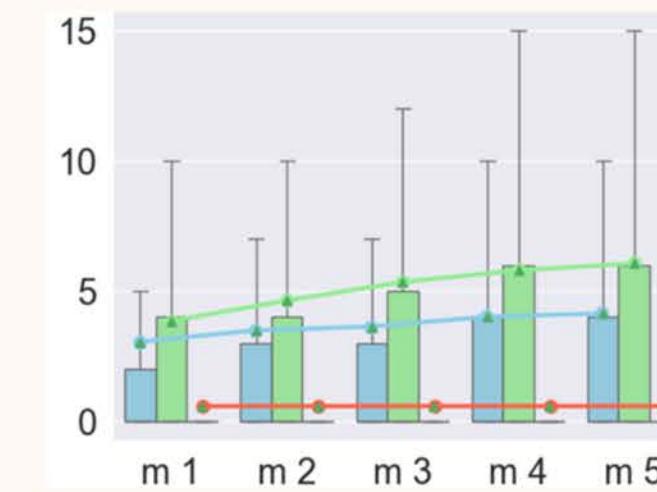
Forks Growth



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

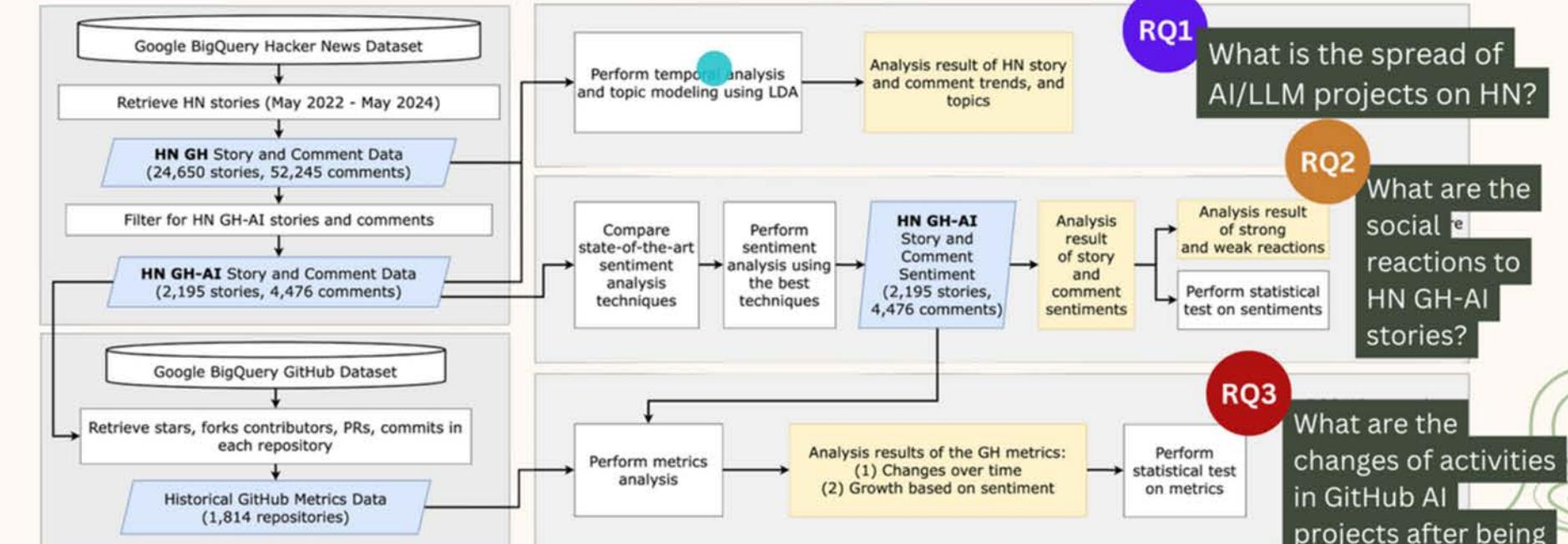


What are the **changes of activities** in GitHub AI projects after being mentioned in HN?

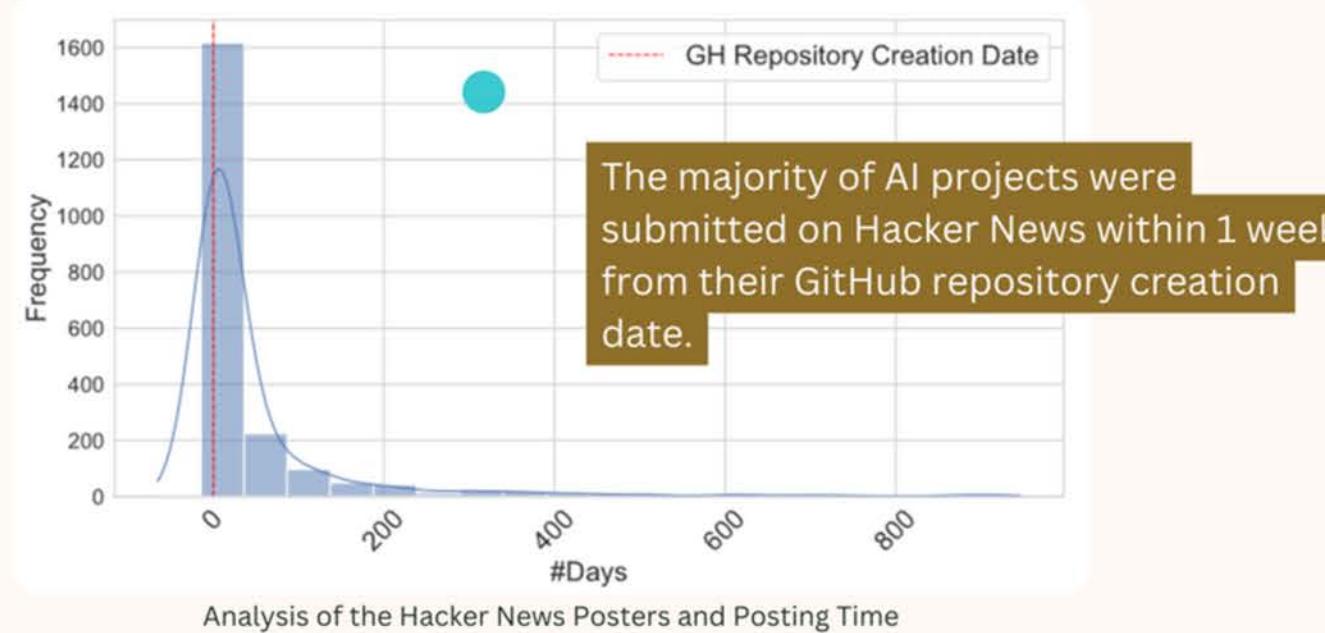
RQ3

Methodology Overview

Data Collection



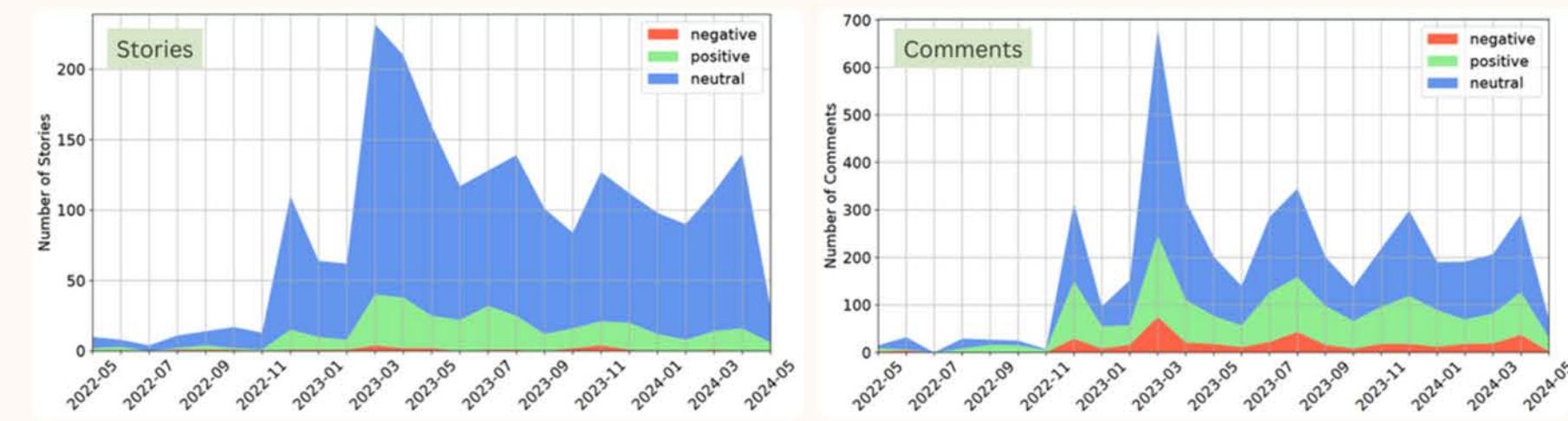
RQ1 What is the spread of AI/LLM projects on HN?



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

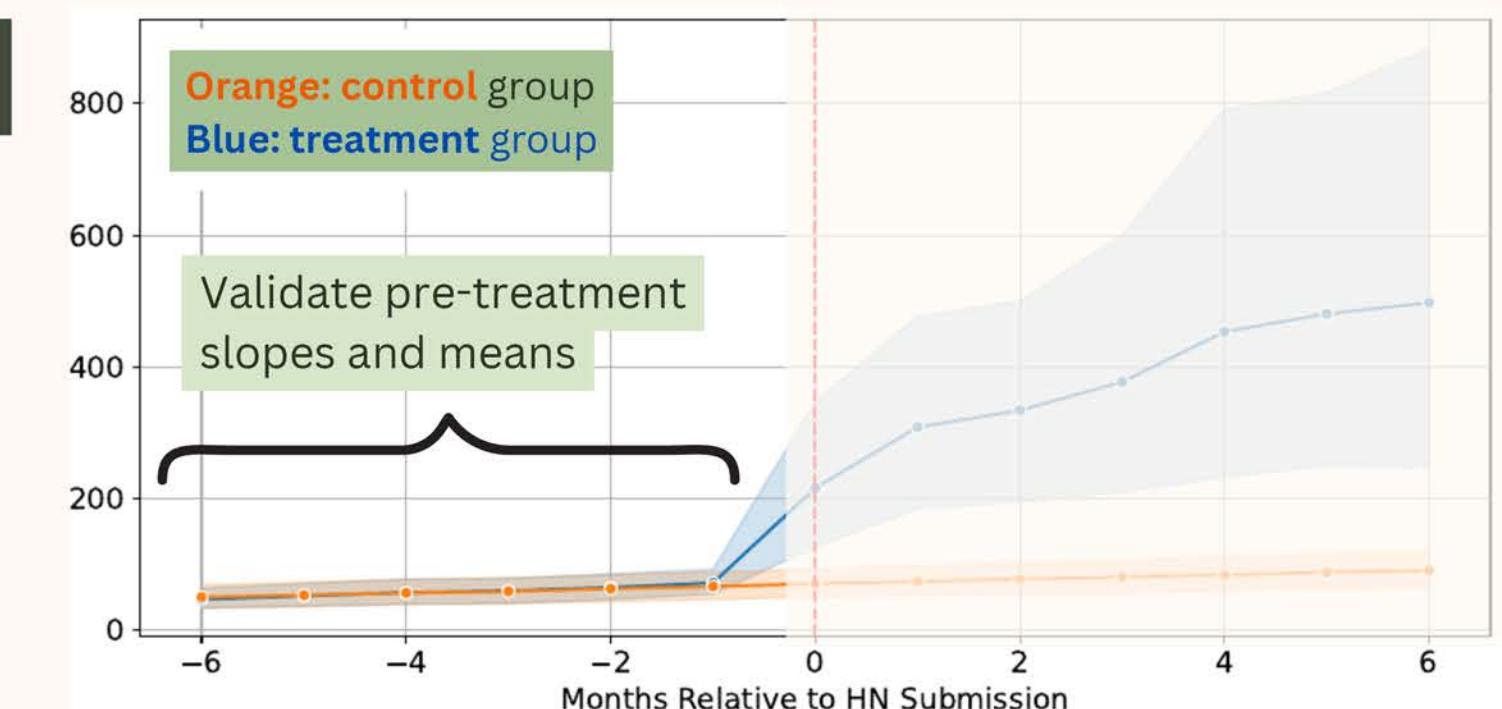
Further verify PTA using **pre-treatment slopes and pre-treatment means (balancing test)**

(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

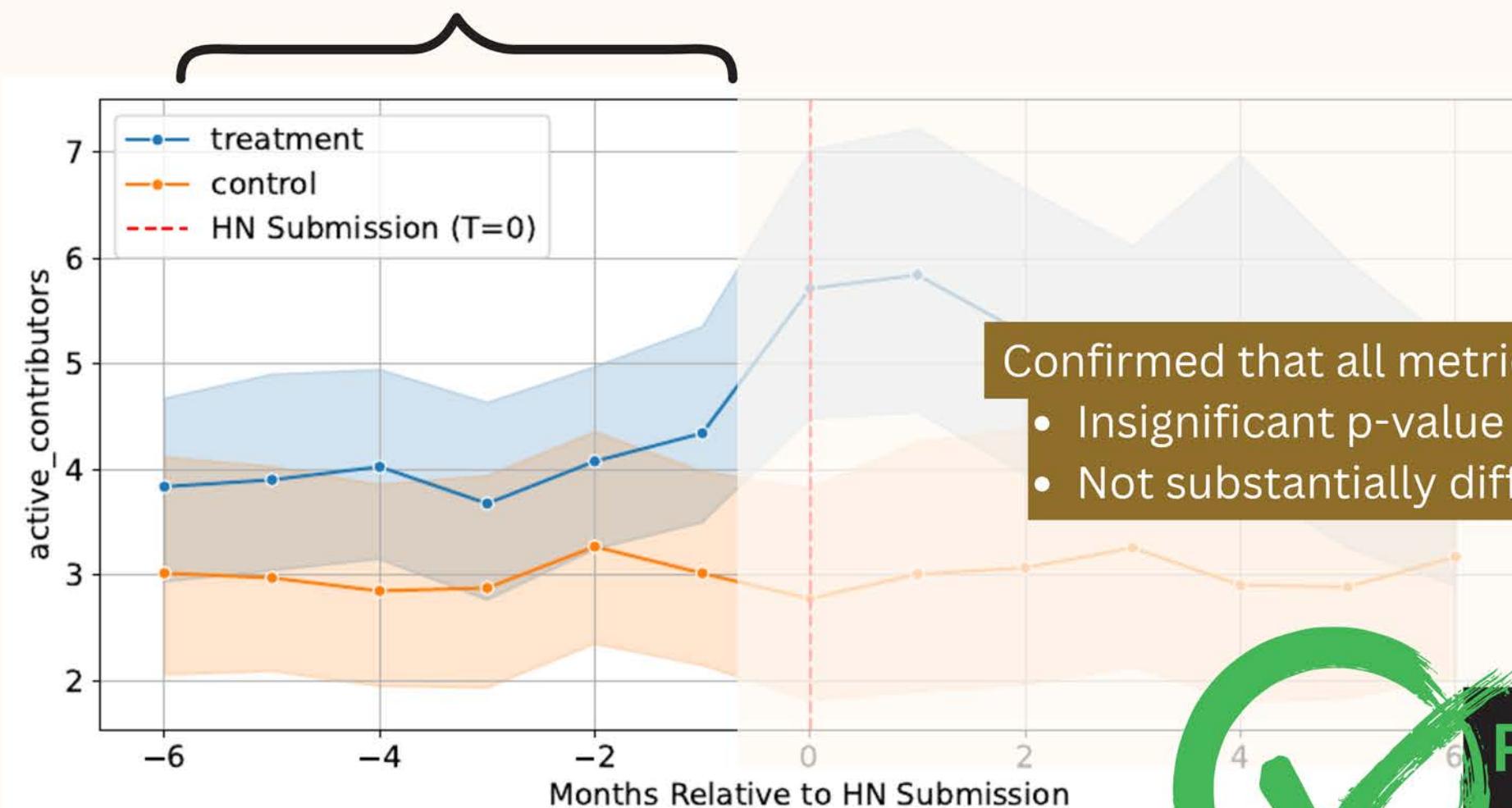
Metric	Number of Project Pairs
Cumulative Commits	35
Cumulative Forks	73
Cumulative PRs	41
Cumulative Stars	43
Monthly Active Contributors	93

- 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



Confirmed that all metrics satisfy at least one of the following:

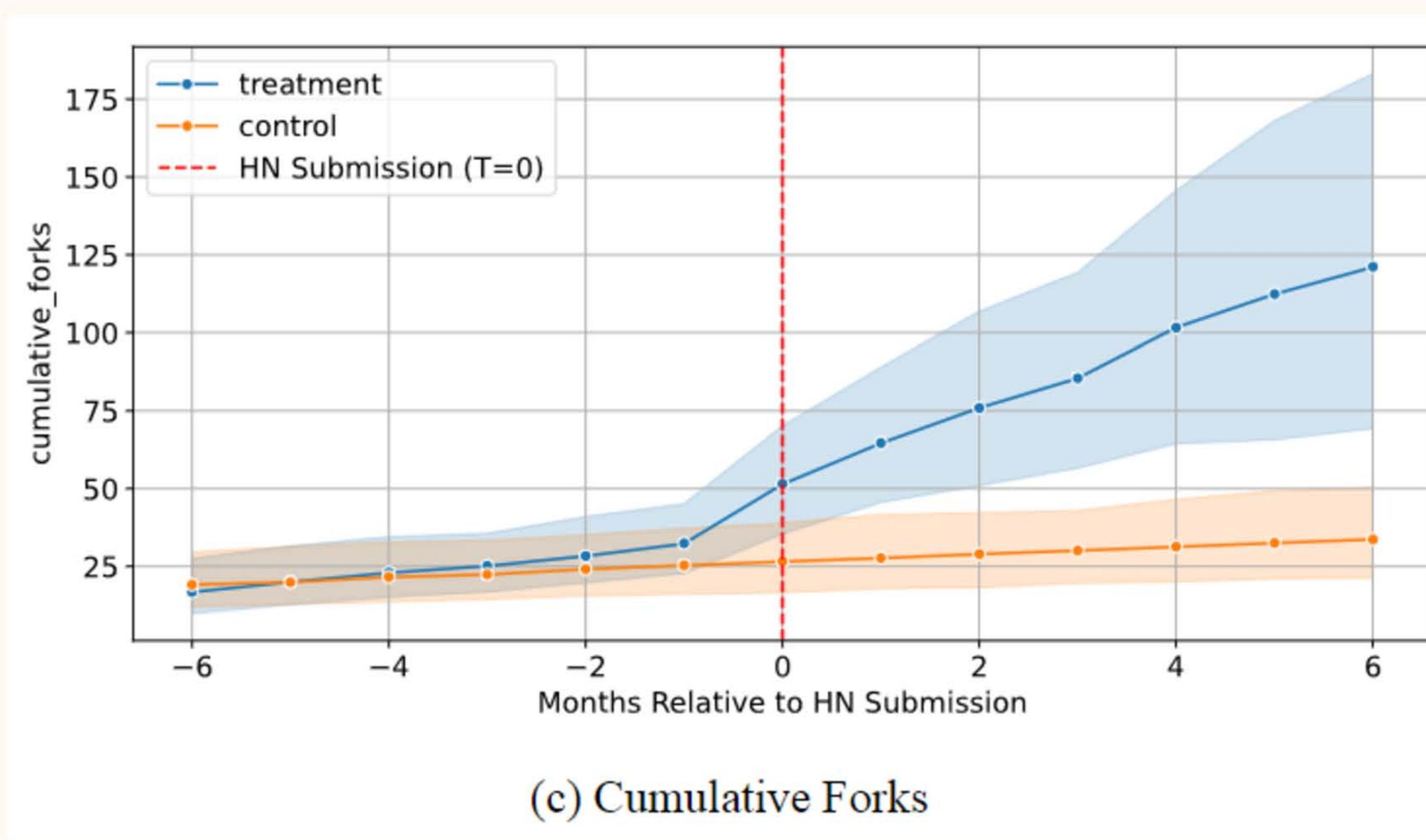
- Insignificant p-value for **pre-treatment means**
- Not substantially different **pre-treatment slopes**

(See appendix for the full results)

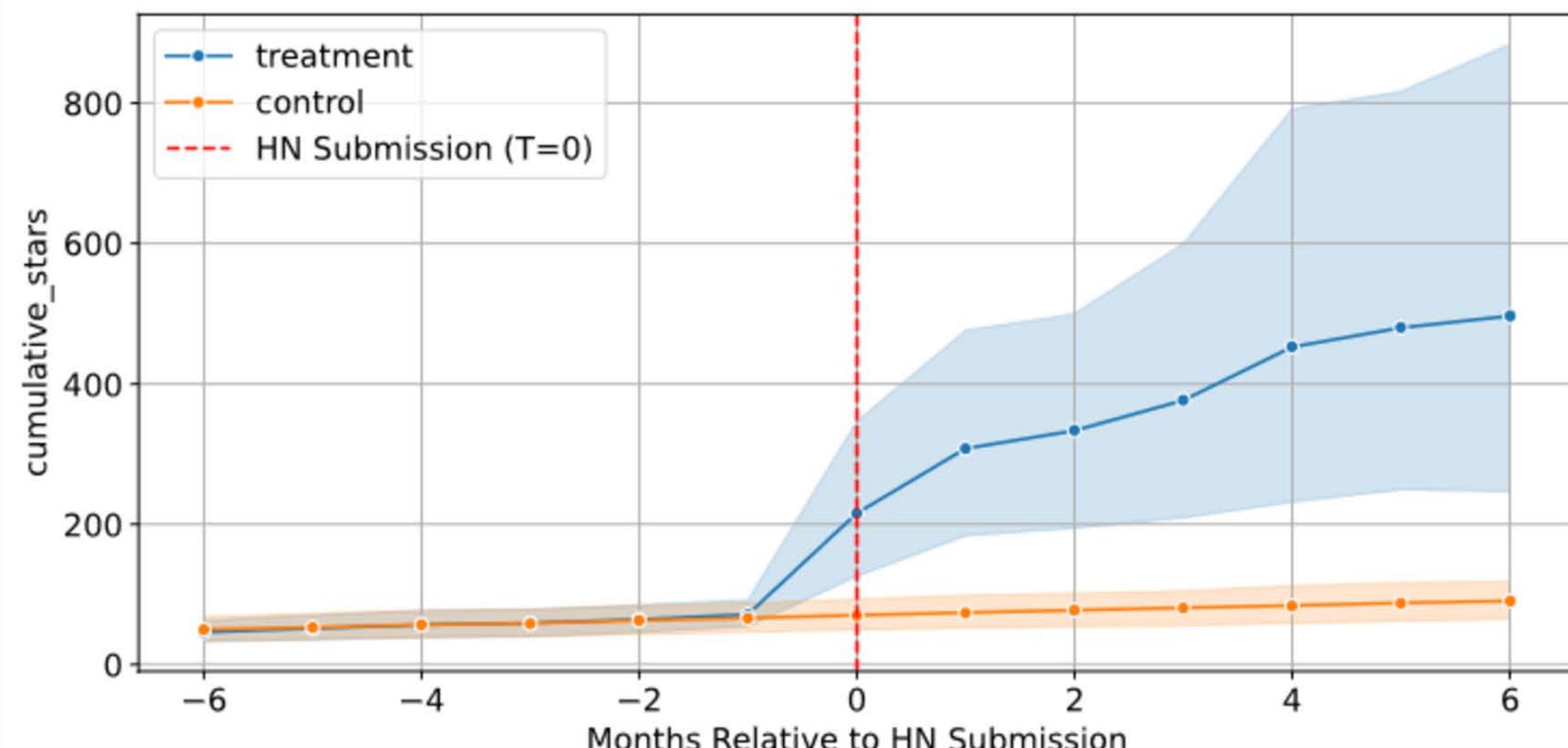


**Parallel trend assumption
(PTA) validated**

DiD PTA Relative Metrics

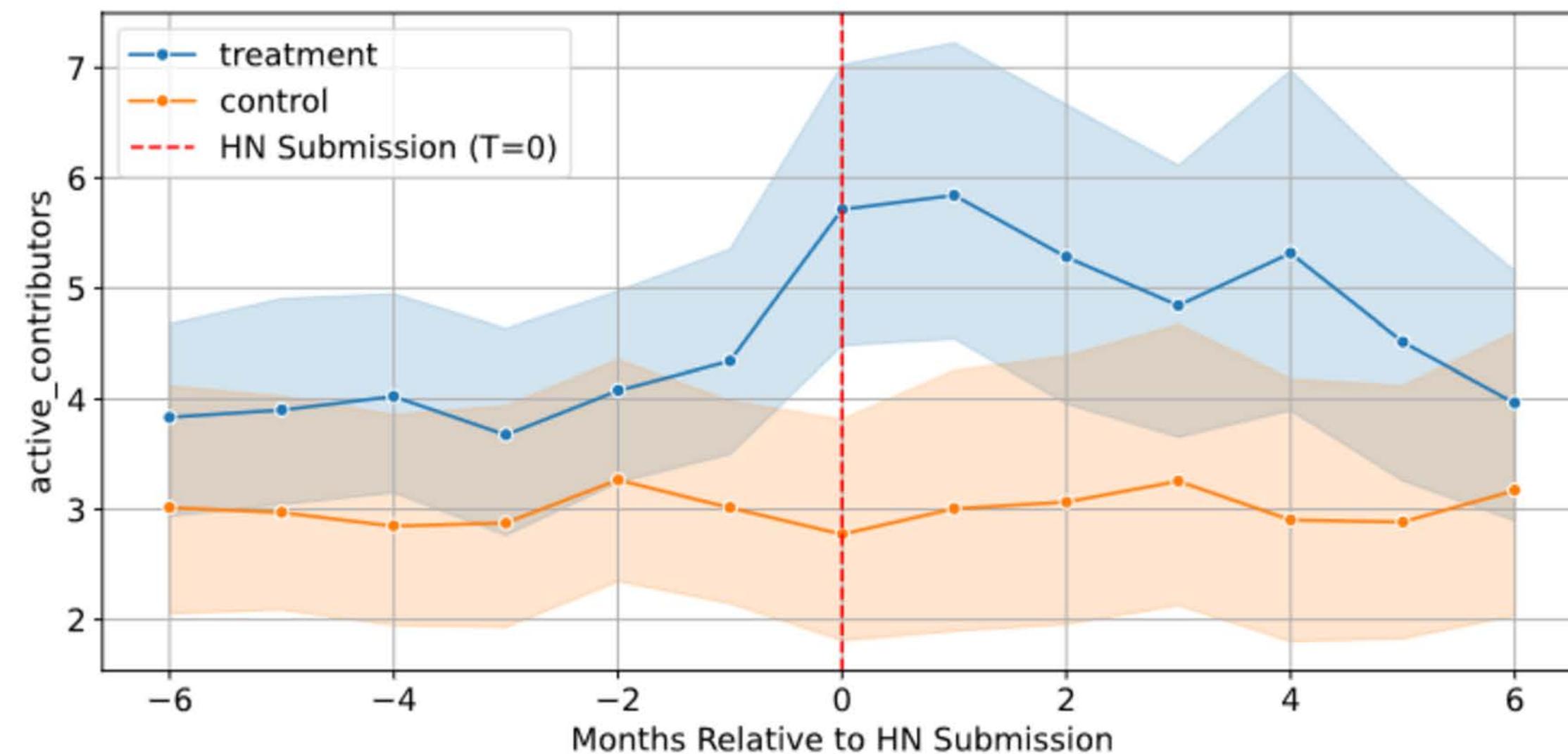


DiD PTA Relative Metrics



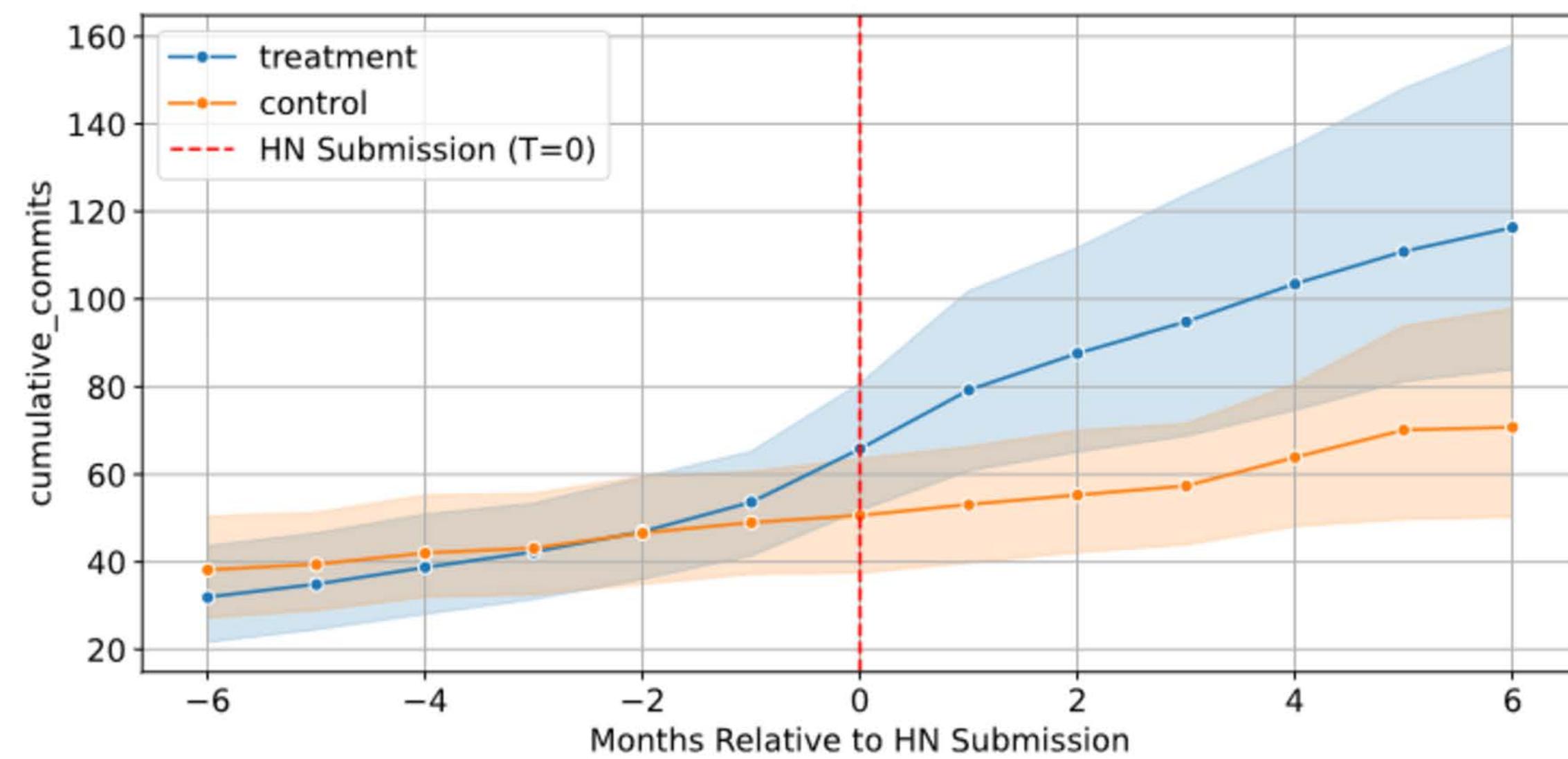
(e) Cumulative Stars

DiD PTA Relative Metrics



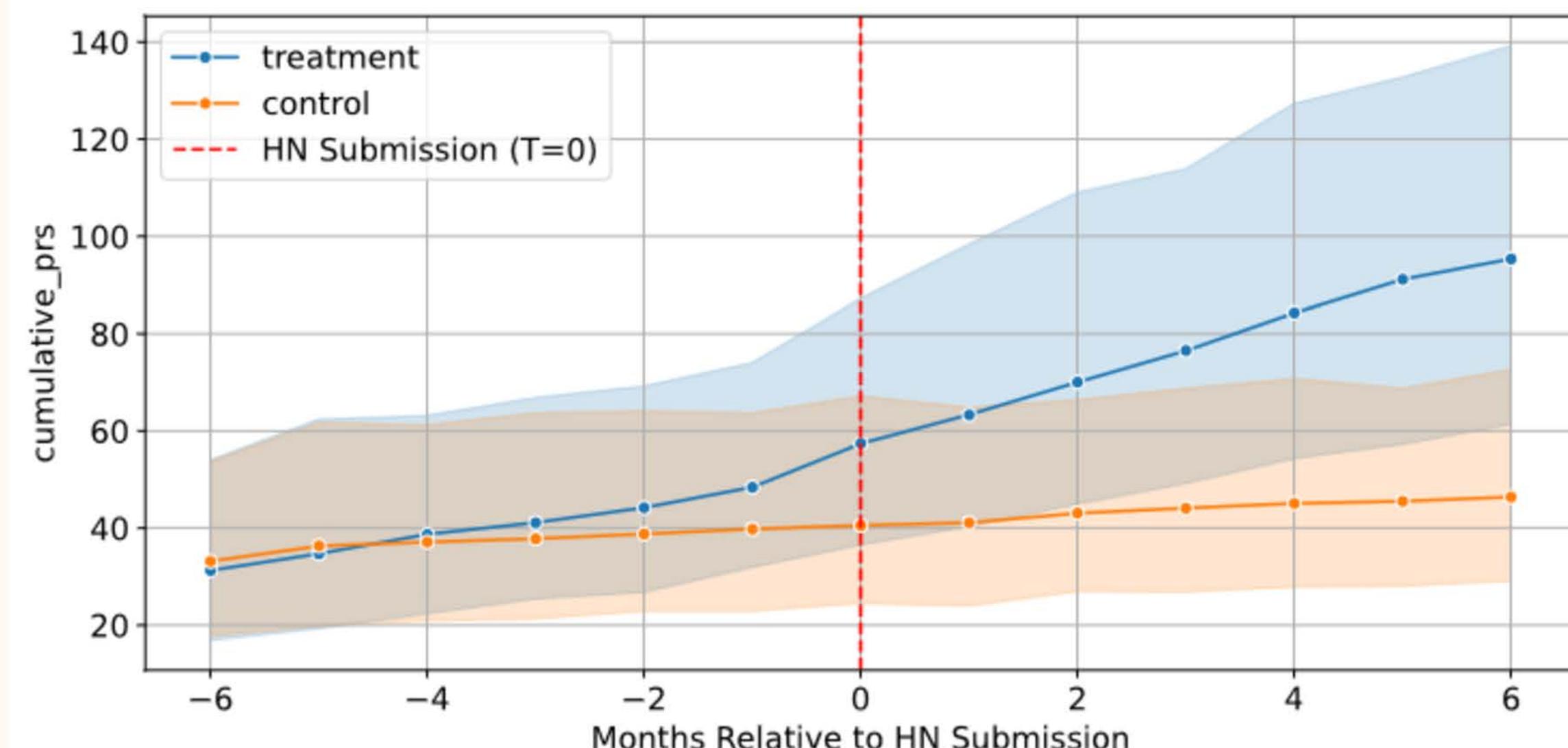
(a) New Monthly Active Contributors

DiD PTA Relative Metrics



(b) Cumulative Commits

DiD PTA Relative Metrics



(d) Cumulative Pull Requests

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.

• INTRODUCTION

• BACKGROUND & RELATED WORKS

• RQ1 - SPREAD

• RQ2 - REACTION

• RQ3 - ACTIVITIES



Conclusion

An abstract graphic element featuring several thin, light brown wavy lines that radiate from the bottom left towards the center, creating a sunburst or fan-like effect.

Observations

RQ1

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

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

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

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

RQ2

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

6: Hacker News discussions show **a few strong reactions** in May 2022 and around ChatGPT's release from Sept 2022 to Jan 2023.

7: **GitHub topics** exclusive to **positive** repository includes AI tools such as '**llama**' and '**whisper**', whereas tools such as '**chatgpt4**' and '**'conversational-ai'**' were more frequently linked to **negative** sentiment.

RQ3

8: GitHub AI projects on Hacker News received an **increment of activities on stars and forks** shortly after the release of ChatGPT.

9: **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

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

Implications

Implication for SE researchers

- This study introduces **Hacker News** as another valuable source of information for empirical software engineering research.



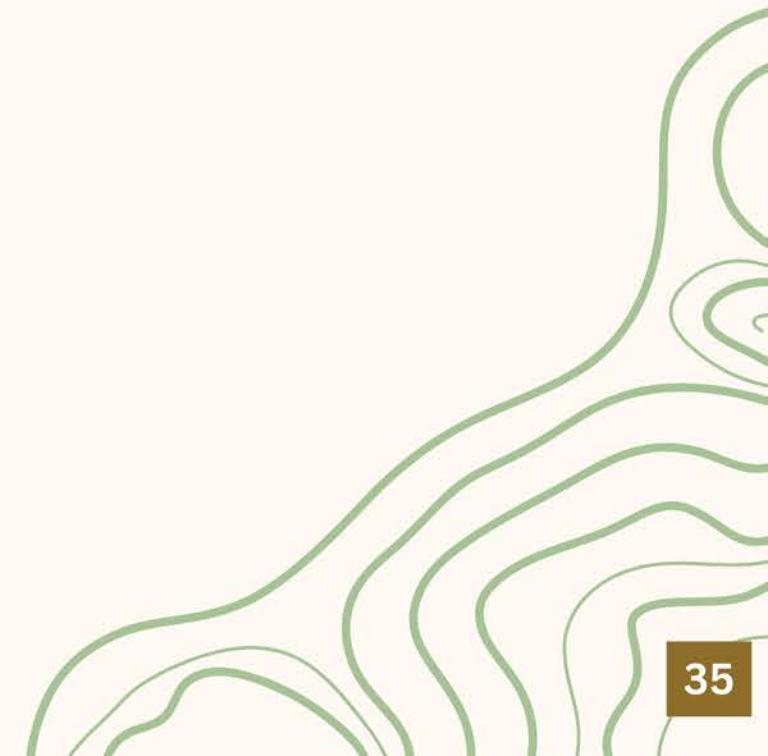
Implication for AI developers

- This study shows that **Hacker News community** is quite supportive of AI/LLM projects which can serve as an **effective mean for AI developers to promote their projects** to gain attraction and further increase their project popularity.





Thanks!
Any Questions?



• INTRODUCTION

• BACKGROUND & RELATED WORKS

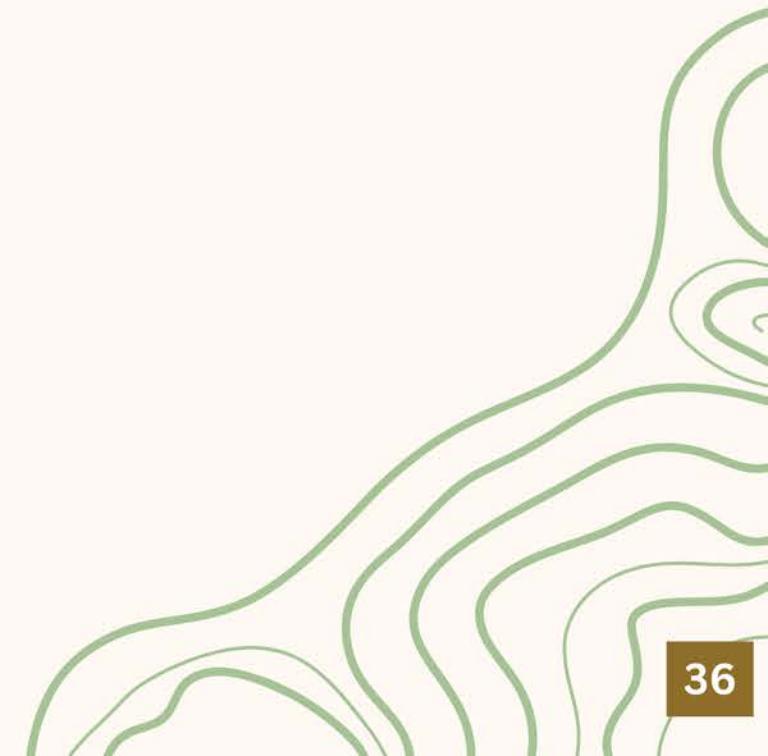
• RQ1 - SPREAD

• RQ2 - REACTION

• RQ3 - ACTIVITIES

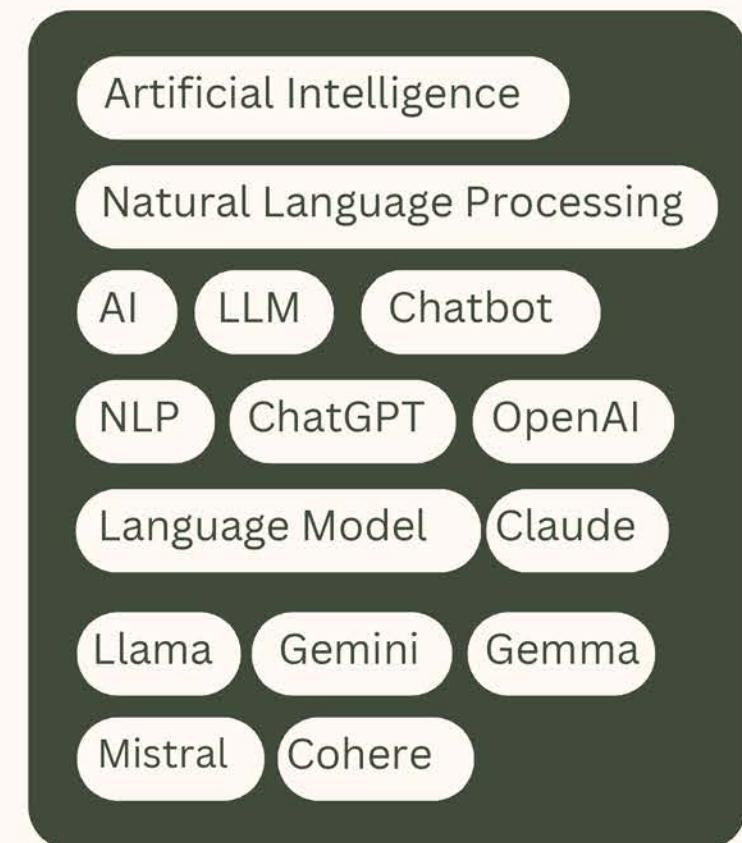


Appendix

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]

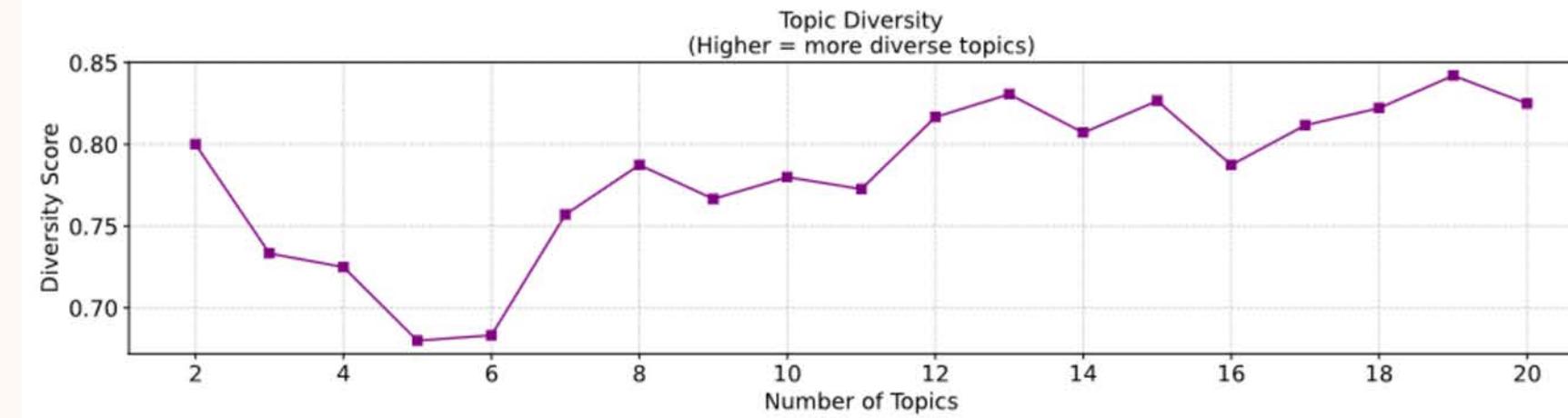
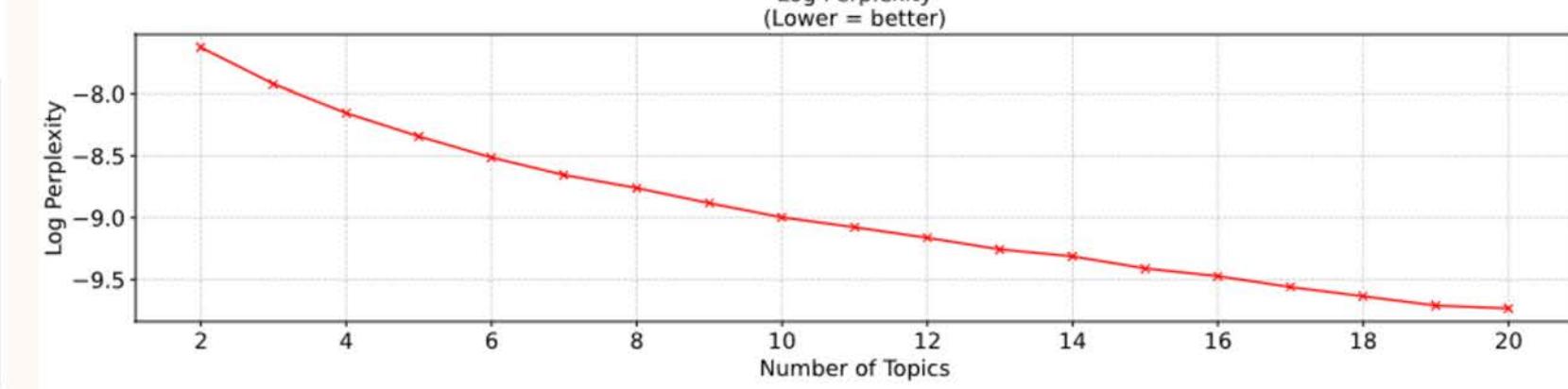
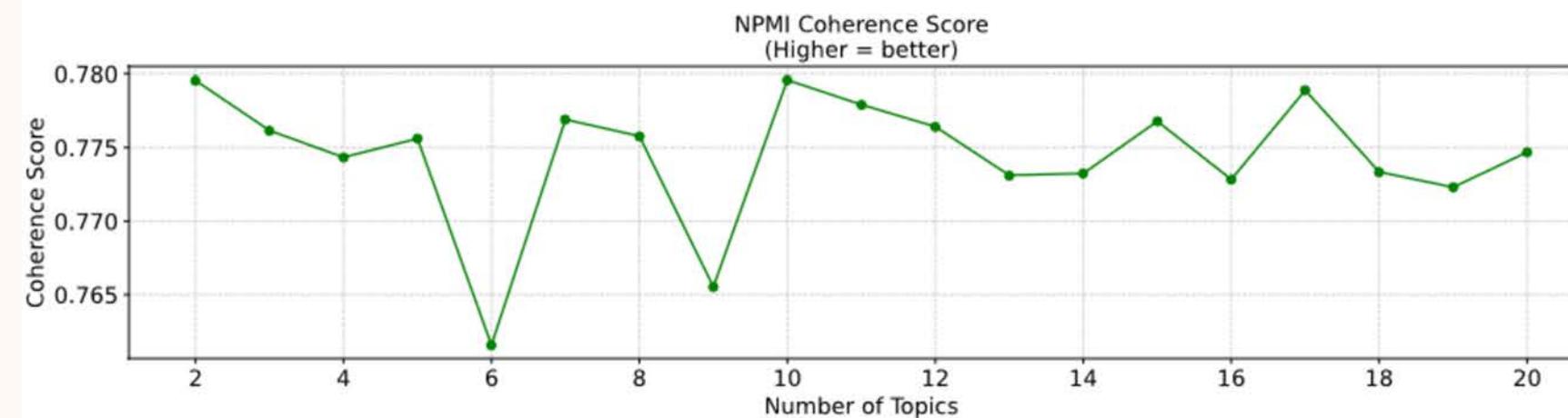
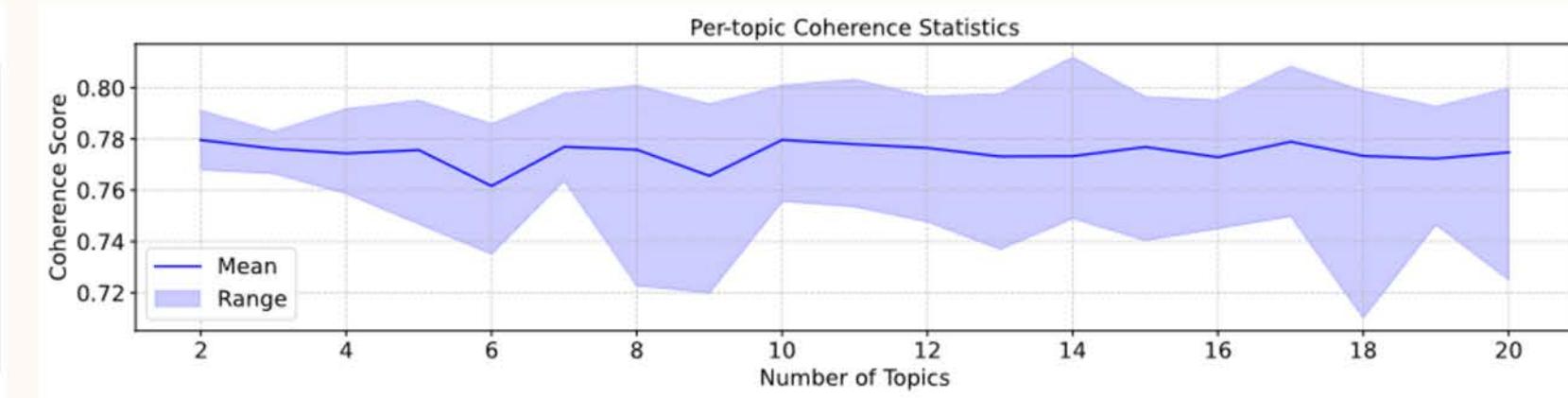
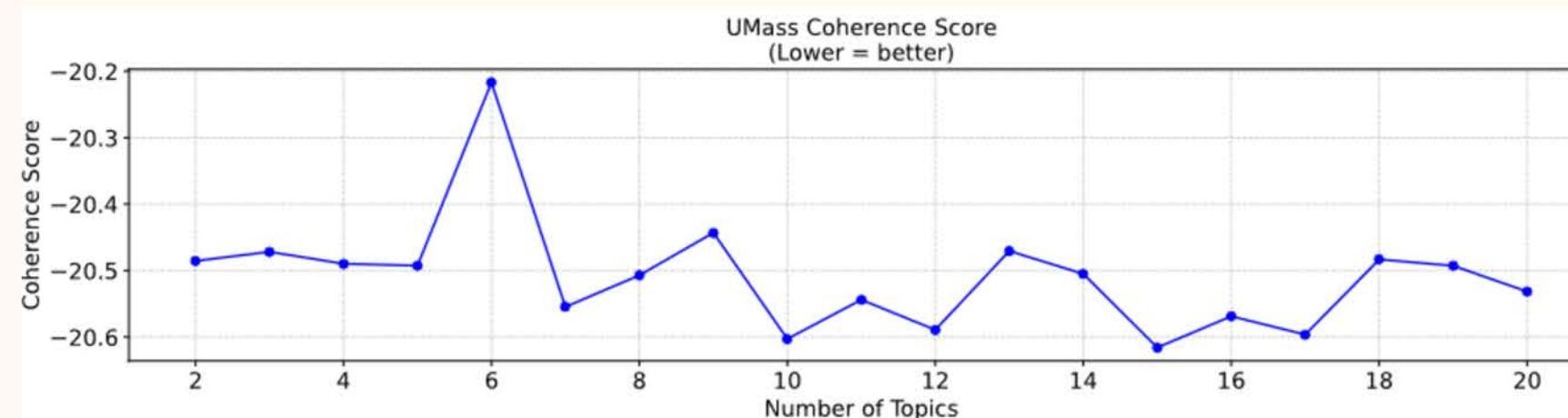


[1] “Machine Learning Glossary,” Google for Developers. Accessed: Nov. 17, 2024. [Online]. Available: <https://developers.google.com/machine-learning/glossary>

[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).

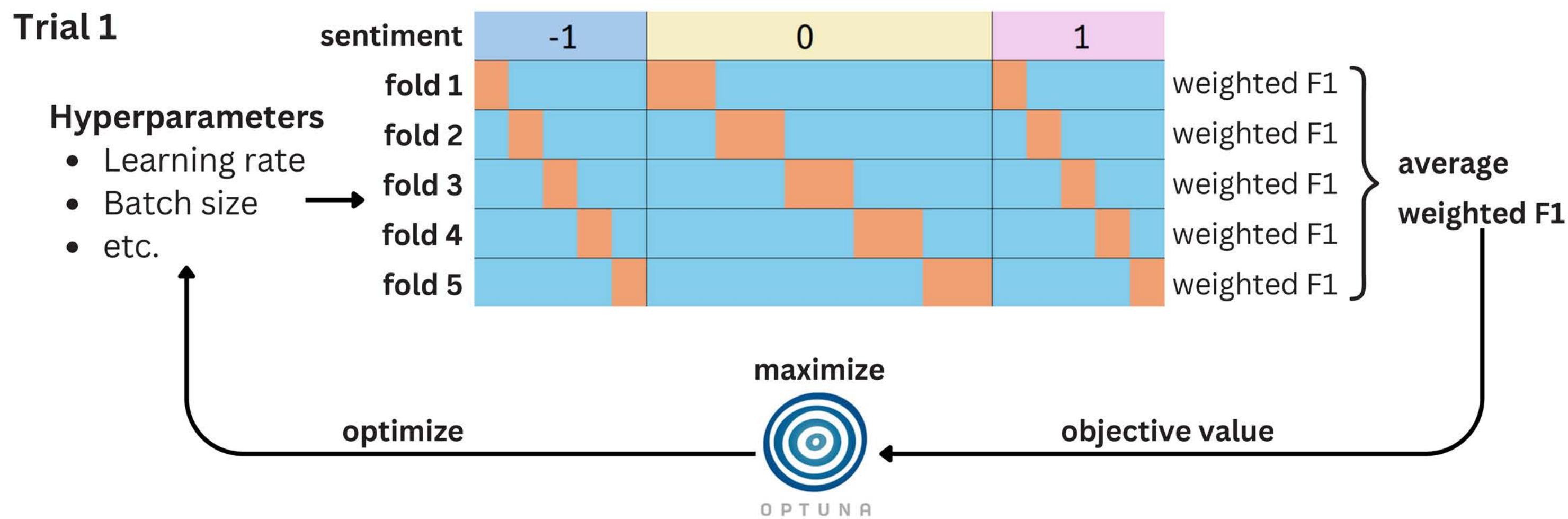
- **eta (Topic-word prior)**

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

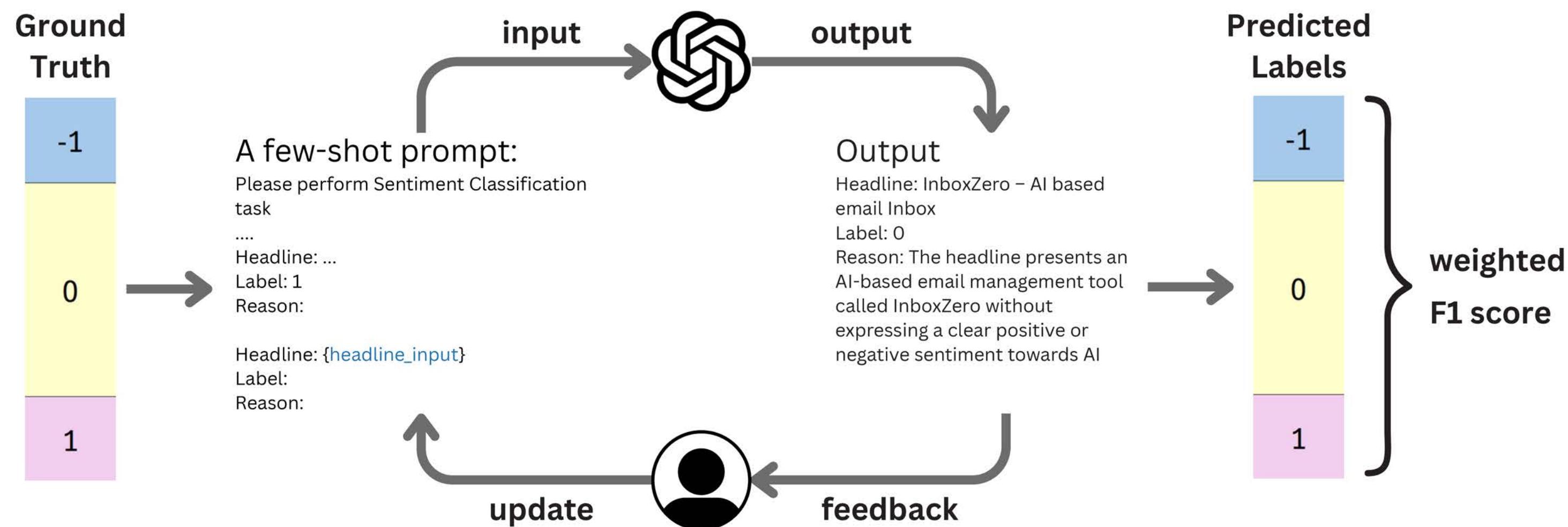
Ground Truth Dataset: Cohen's Kappa coefficient

- **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

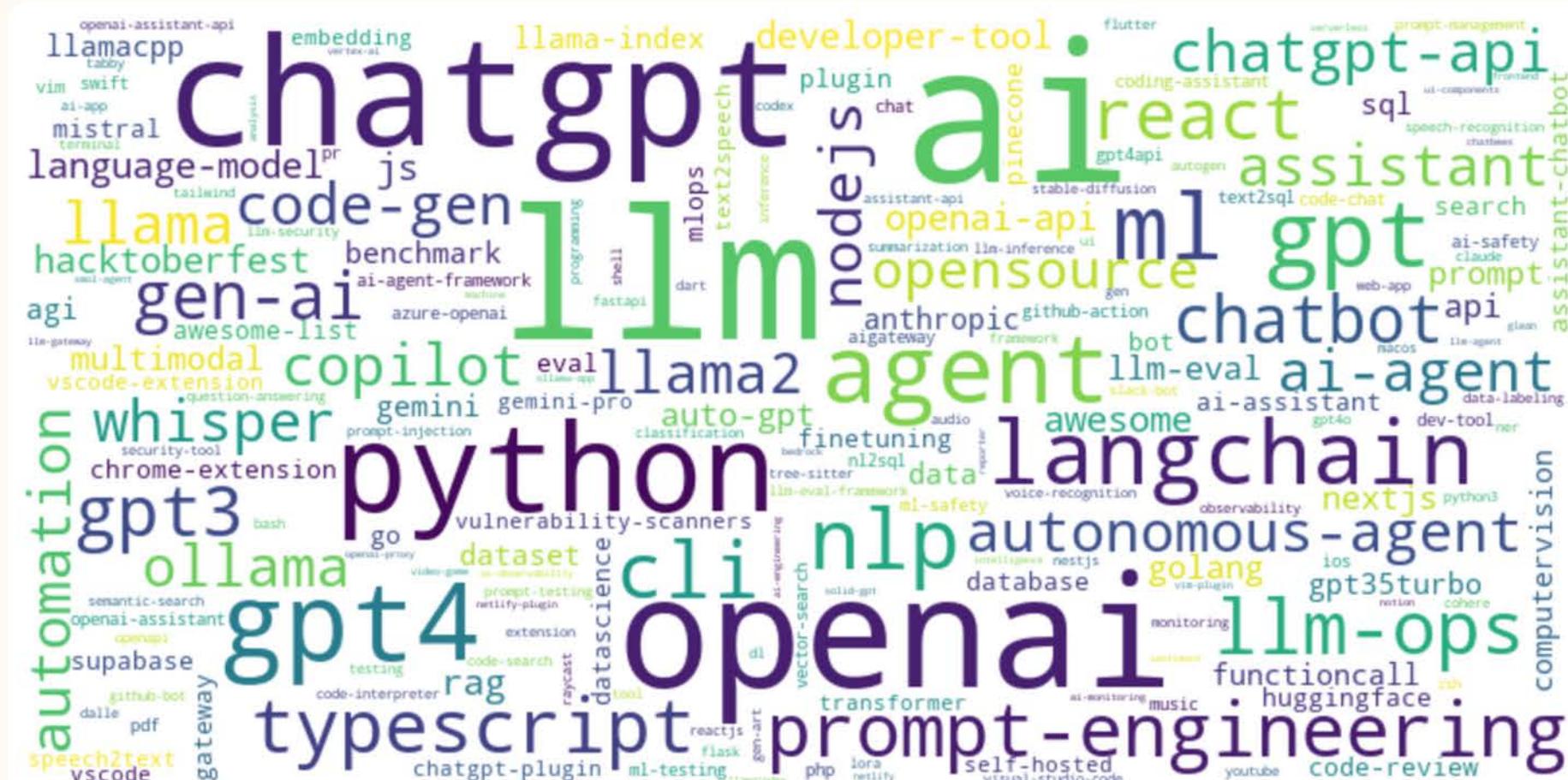
- Sentiment of each repository is defined by its **average comment sentiment**:
 - Average comment sentiment $> 0.5 \rightarrow$ Repository is considered to be **positive**
 - Average comment sentiment $< -0.5 \rightarrow$ Repository is considered to be **negative**
- GitHub topics were preprocessed 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).



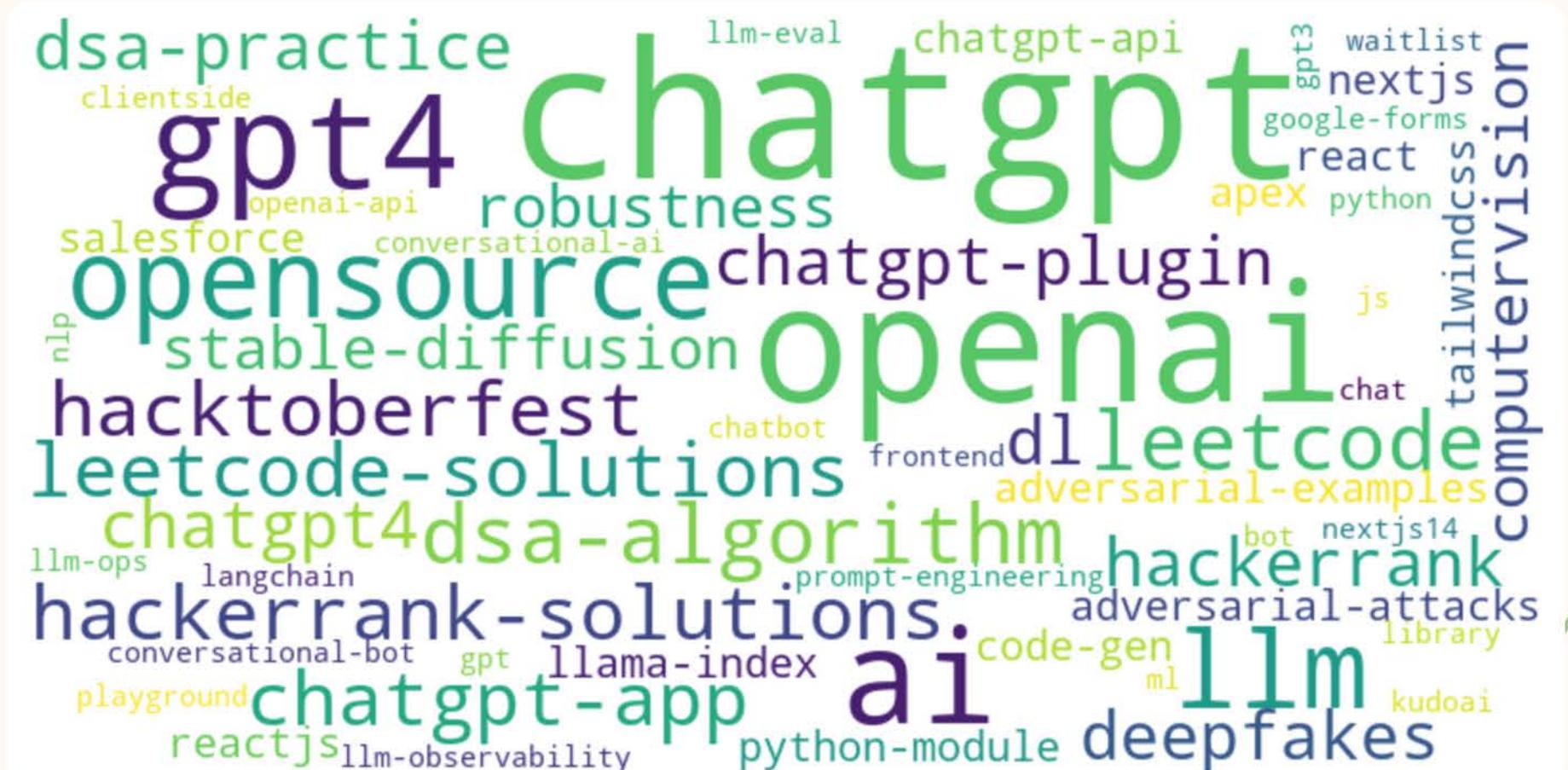
cohere
llama
llm
mistral
ai

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

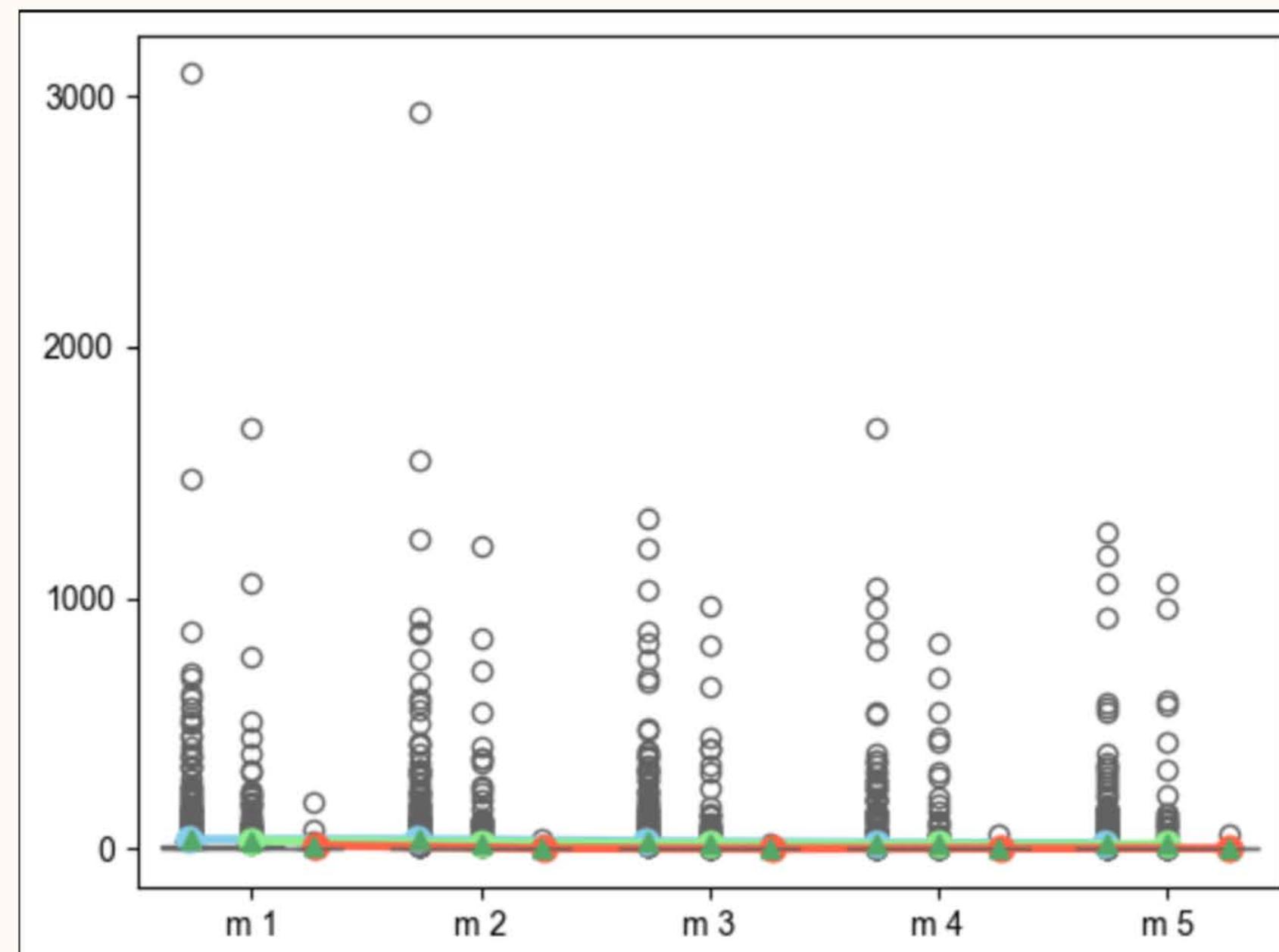
- Comments in **positive** repositories

- 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)**

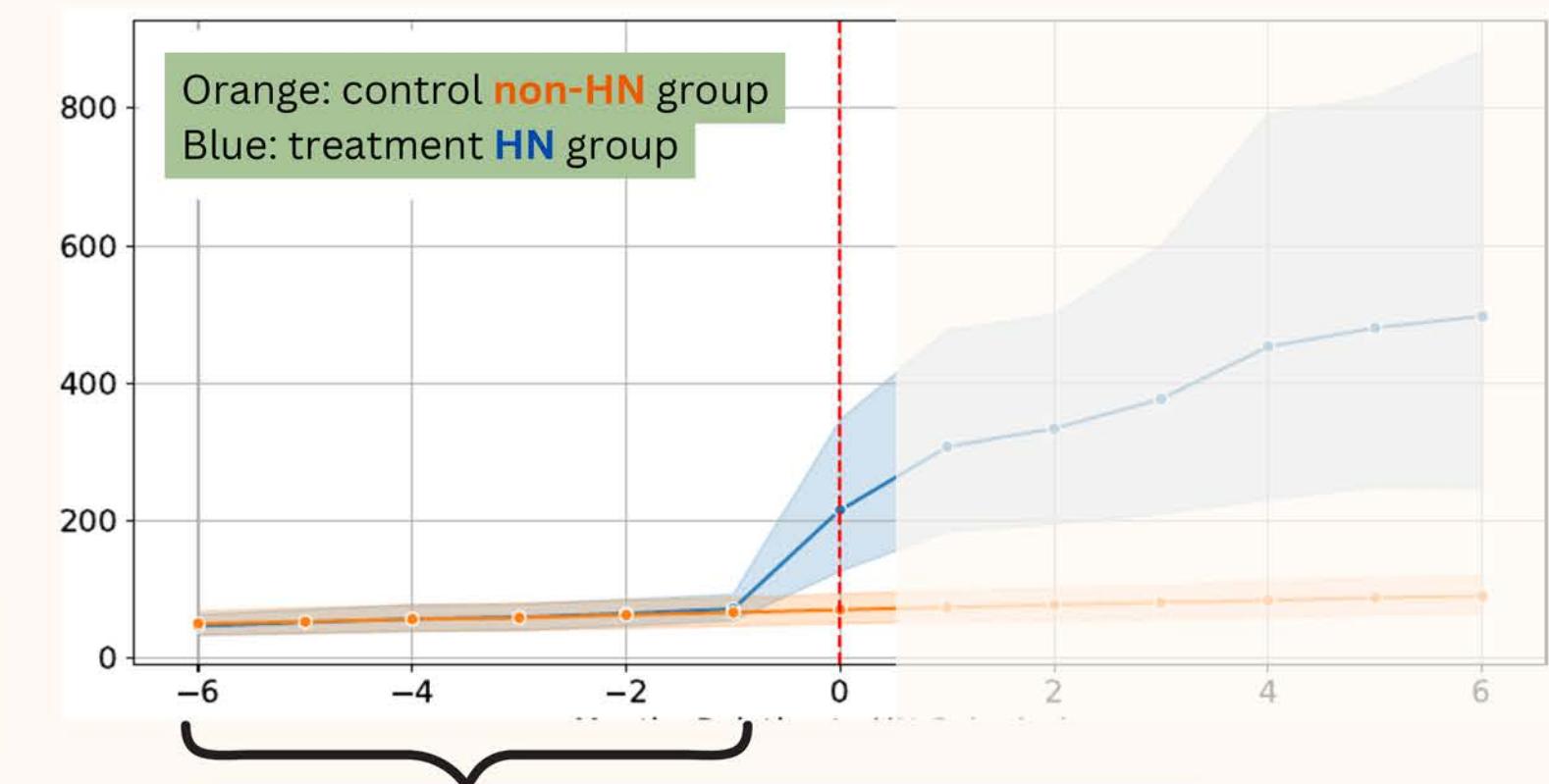
- I.e. metrics prior to **treatment** (HN submission) must have a **similar pre-treatment trajectory**

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



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

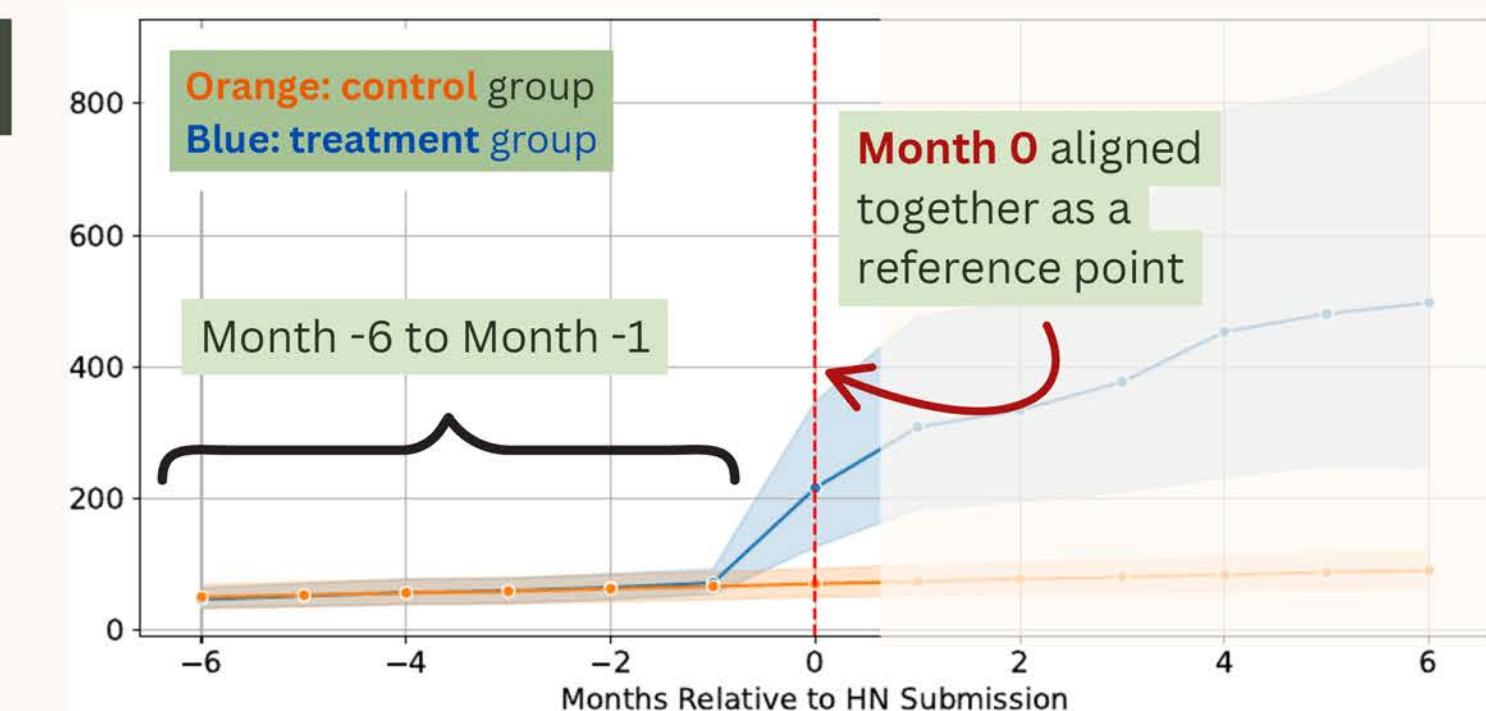
RQ3

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**

2

Convert each HN GH-AI repository metrics from **month -6 to month -1** into a vector

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

• INTRODUCTION

• BACKGROUND & RELATED WORKS

• RQ1 - SPREAD

• RQ2 - REACTION

• RQ3 - ACTIVITIES

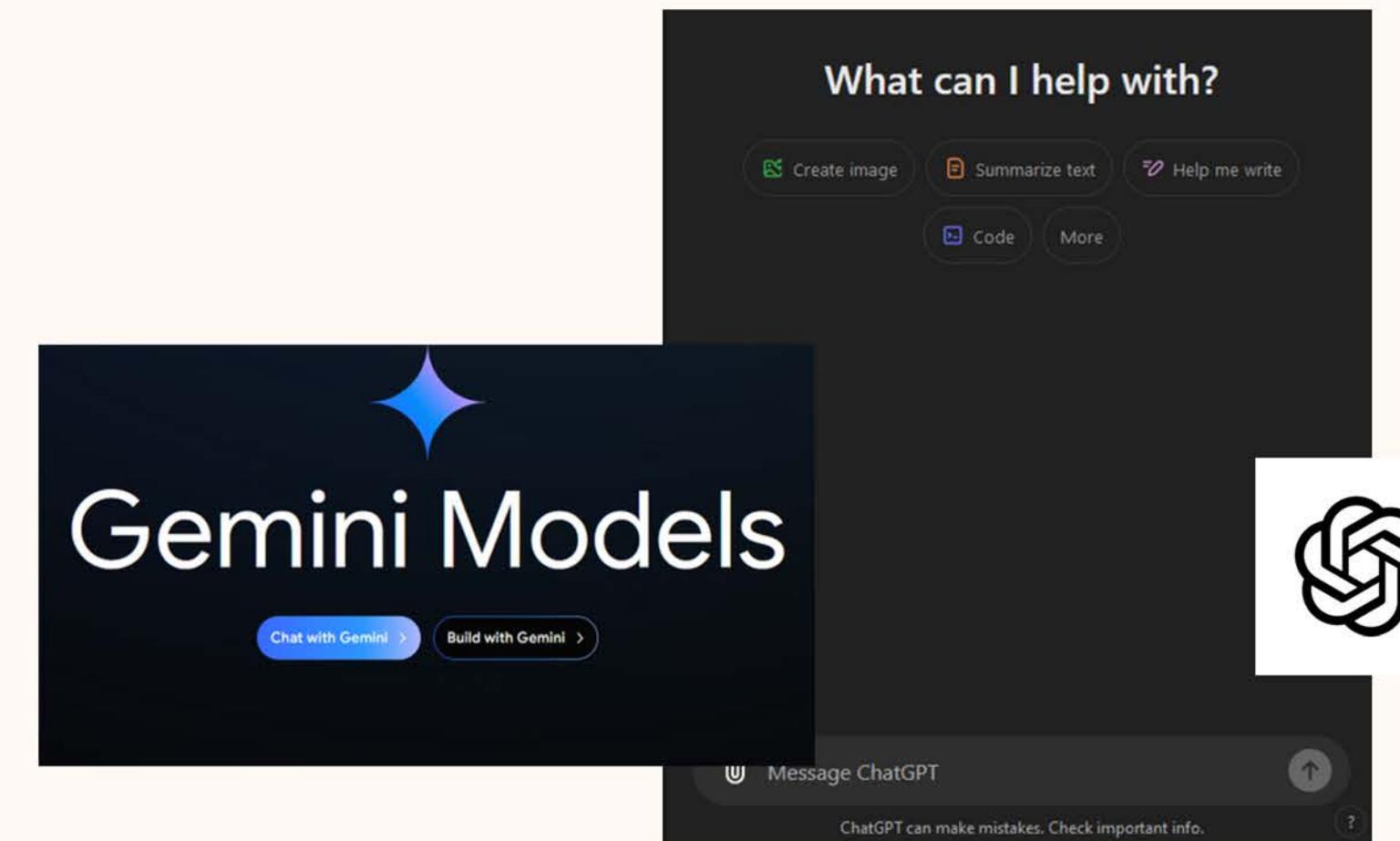


Introduction



Introduction - AI and LLM

- Among these is generative AI such as Large Language Model (LLM) - comprehend and generate human language texts [2]
 - e.g. OpenAI's ChatGPT [3] and Google's Gemini [4]



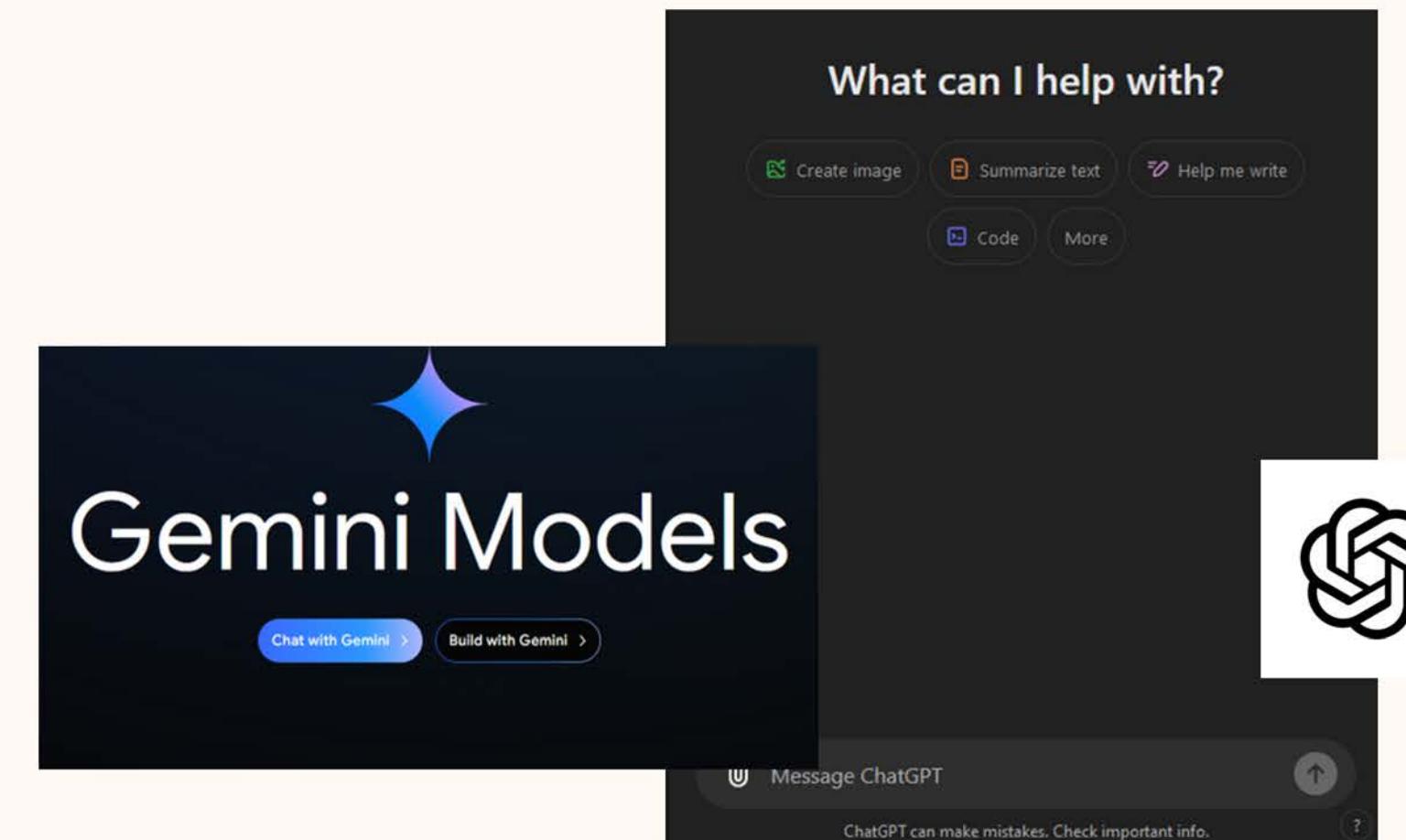
[2] "What is a large language model (LLM)?" Accessed: Nov. 19, 2024. [Online]. Available: <https://www.cloudflare.com/learning/ai/what-is-large-language-model/>

[3] "ChatGPT." Accessed: Aug. 02, 2024. [Online]. Available: <https://openai.com/chatgpt/>

[4] "Gemini - chat to supercharge your ideas," Gemini. Accessed: Nov. 19, 2024. [Online]. Available: <https://gemini.google.com>

Introduction - AI and LLM

- Artificial Intelligence (AI) - Simulation of human intelligence to learn and think [1]
 - i.e., to accomplish tasks like humans, including speech recognition, classification, sentiment analysis, etc.
- Among these is generative AI such as Large Language Model (LLM) - comprehend and generate human language texts [2]
 - e.g. OpenAI's ChatGPT [3] and Google's Gemini [4]



[1] "What Is Artificial Intelligence (AI)? | Google Cloud." Accessed: Nov. 19, 2024. [Online]. Available: <https://cloud.google.com/learn/what-is-artificial-intelligence>

[2] "What is a large language model (LLM)?" Accessed: Nov. 19, 2024. [Online]. Available: <https://www.cloudflare.com/learning/ai/what-is-large-language-model/>

[3] "ChatGPT." Accessed: Aug. 02, 2024. [Online]. Available: <https://openai.com/chatgpt/>

[4] "Gemini - chat to supercharge your ideas," Gemini. Accessed: Nov. 19, 2024. [Online]. Available: <https://gemini.google.com>

AI projects examples

- AudioCraft : a library for audio processing and generation with deep learning.
- BibiGPT : AI-driven video and audio summarizer
- PrivateGPT : a private and secure version of the GPT language model.[3]

AudioCraft

audiocraft_docs passing audiocraft_linter passing

y for deep learning research
aft contains inference and
-the-art AI generative models
AudioGen and MusicGen.

BibiGPT: one-Click AI Summary for Audio/Video & Chat with Learning Content

.9, PyTorch 2.1.0. To install
following:

```
have torch installed
already have PyTorch
torch==2.1.0'
llowing before trying
etuptools wheel
f the following
```

PrivateGPT

GITHUB TRENDING #1 Repository Of The Day

tests passing Documentation check it PrivateGPT 534 online

X Follow @ZylonPrivateGPT

PRIVATE GPT

Mode: RAG Search BASIC
LLM: ollama | Model: llama3.5
What is the main innovation of the Transformer model?

Get contextualized answers from selected files.
Upload File(s)

File name: 1706.03762v7.pdf
Sources: 1. 1706.03762v7.pdf (page 2)
2. 1706.03762v7.pdf (page 1)

Selected for Query or Deletion:
All files
Delete selected file
Delete ALL files
Retry Undo Clear Submit
Type a message...

PrivateGPT is a production-ready AI project that allows you to ask questions about your documents using the power of Large Language Models (LLMs), even in scenarios without

内容一键总结 & 对
轻松学习哔哩哔哩 |

[1] "Gartner Poll Finds 55% of Organizations are in Piloting or Production Mode with Generative AI," Gartner. Accessed: Nov. 21, 2024. [Online]. Available: <https://www.gartner.com/en/newsroom/press-releases/2023-10-03-gartner-poll-finds-55-percent-of-organizations-are-in-piloting-or-production-mode-with-generative-ai>

[2] "PrivateGPT". Accessed April. 28 2025 [Online]. Available: <https://github.com/zylon-ai/private-gpt>

[3] "BibiGPT: one-Click AI Summary for Audio/Video & Chat with Learning Content". Accessed Apr. 28 2025 [Online]. Available: <https://github.com/JimmyLv/BibiGPT-v1>

Background & Related works outline

- Social platforms, Hacker News and Software Engineering
- Sentiment Analysis & LLMs
- Causal Inference, Diff-in-Diff, Parallel Trend Assumptions (PTA)

GitHub & GitHub API

- Platform to host projects as well as offering version control, collaboration, and project management
 - Provides GitHub API to retrieve repository data, including metrics e.g. no. stars/forks [1]

Why GitHub?

- Popular among developer community and filled with various software development projects and open-source projects [3]
- The repositories on GitHub are well documented and GCP's BigQuery GHArchive allows repository data to be retrieved efficiently

The screenshot shows the GitHub REST API documentation landing page. It features a dark header with the title 'GitHub REST API documentation' and a sub-header 'Create integrations, retrieve data, and automate your workflows with the GitHub REST API.' Below the header are two buttons: 'Overview' (highlighted in green) and 'Quickstart'. The main content area is divided into several sections: 'Start here' (with a 'View all' link), 'Popular' (listing 'About the REST API', 'Rate limits for the REST API', 'Getting started with the REST API', and 'Troubleshooting the REST API'), and 'What's new' (listing 'Upcoming changes to data retention for Events API, Atom feed, /timeline and /dashboard-feed features' with a 'November 6' date). There is also a note about 'View an organization's REST API activity with API insights in public preview'.

GitHub REST API for retrieving repository metrics [1]

The screenshot shows the GitHub profile for the 'Snowflake AI Research' repository. At the top, it displays the license as 'apache2.0' and a 'Follow @snowflakedb' button. The repository name is 'Snowflake AI Research'. On the right side, there are sections for 'No releases published', 'Packages' (no packages published), 'Contributors' (10 contributors shown with icons), and 'Languages' (Python 98.5%, Other 1.5%). The left side shows the 'Latest News' section with two items: '[07/23/2024] Snowflake Teams Up with Meta to Host and Optimize New Flagship Model Family in Snowflake Cortex AI' and '[04/24/2024] Snowflake Arctic: The Best LLM for Enterprise AI — Efficiently Intelligent, Truly Open'. Below that is the 'Overview' section, which provides a brief description of the repository's purpose: 'The Snowflake AI Research team is conducting open, foundational research to advance the field of AI while making enterprise AI easy, efficient, and trusted. This repo contains several artifacts to help efficiently train and inference models using the Snowflake Data Cloud. We launched this in April of 2023, and have since trained over 100 billion parameters from scratch.'

Example of AI GitHub open-source project [2]

[1] "GitHub REST API documentation" Accessed: Nov. 22, 2024. [Online]. Available: <https://docs.github.com/en/rest?apiVersion=2022-11-28>

[2] "Snowflake AI Research" Accessed: Nov. 22, 2024. [Online]. Available: <https://github.com/Snowflake-Labs/snowflake-arctic>

[3] K. D. Staff GitHub, "Octoverse: The state of open source and rise of AI in 2023," The GitHub Blog. Accessed: Nov. 21, 2024. [Online]. Available: <https://github.blog/news-insights/research/the-state-of-open-source-and-ai/>

Related Works about Social Platforms, Hacker News, and GitHub Studies

Studies about AI/LLM on other platforms

- Study of discussions related to deep learning across platforms such as Github and Stack Overflow (Han et al., 2020)
 - Han J., Shihab E., Wan Z., Deng S., Xia X. (2020) “What Do Programmers Discuss about Deep Learning Frameworks”,

Studies based on HackerNews data

- Study of popularity dynamic in Reddit and Hacker News
 - Stoddard G. (2021) “Popularity Dynamics and Intrinsic Quality in Reddit and Hacker News”, AAAI’21
- Study of Reddit and Hacker News as modern news aggregators for developers
 - Aniche et al. (2018) “How Modern News Aggregators Help Development Communities Shape and Share Knowledge”, ICSE ’18
- Study of Hacker News’s effectiveness in software engineering research
 - Titus B et al. (2015) “I heart hacker news”, FSE’15

Key Takeaway

HackerNews is a highly efficient and reliable source of high quality information

The trend of AI/LLM is different on each platform so these platforms may need to be studied separately.

Sentiment Analysis

- A technique used to determine **emotional** tone or **attitude** expressed in **text**

Studies about Sentiment Analysis

- Exploration of multiple **sentiment analysis models** to software engineering (SE) dataset
 - Lin et al. (2018) “Sentiment Analysis for Software Engineering: How Far Can We Go?”
- Study of **fine-tuned pretrained transformer models** for sentiment analysis on SE dataset
 - Zhang et al. (2020) “Sentiment Analysis for Software Engineering: How Far Can Pre-trained Transformer Models Go?”
- Investigation of **LLMs** for sentiment analysis on SE dataset
 - Zhang et al. (2025) “Revisiting Sentiment Analysis for Software Engineering in the Era of Large Language Models”

Key Takeaway

These studies suggest we should **experiment with multiple approaches** for sentiment analysis

We explored **fine-tuning pretrained transformer models** and **LLM**.

Topic Modeling

- A technique with natural language processing (NLP) used to uncover themes or topics within a collection of documents with no requirement of supervision.

Latent Dirichlet Allocation (LDA)

- Generative probabilistic model used for topic modeling to identify themes or topics and their distribution within a collection of documents along with the distribution of words in each topic



Related Works about Topic Modelling, LDA

Studies about Topic Modelling

- Study of discussions related to deep learning across platforms such as Github and Stack Overflow
 - Han J., Shihab E., Wan Z., Deng S., Xia X. (2020) “What Do Programmers Discuss about Deep Learning Frameworks”,

Studies about LDA

- Study of popularity dynamic in Reddit and Hacker News
 - Stoddard G. (2021) “Popularity Dynamics and Intrinsic Quality in Reddit and Hacker News”, AAAI’21
- a

aa
aa
aa

Key Takeaway

a

a



Used in RQ3

Related Works about causal analysis and DiD

Studies implementing DiD for causal analysis related to Software Engineering

- Study of GitHub repositories resilience under Covid-19 (Lu et al., 2023) [1]

Studies discussing approaches to designing DiD

- Designing DiD for Health policy research (C. Wing et al., 2018) [2]

Key Takeaway

The paper demonstrates the usage of **difference in difference** for causal analysis on **GitHub repository metrics**. They also suggested methods to **address insufficient control** data using ML.

The paper recommends researchers to perform **robustness checks, sensitivity analyses** to validate the method's assumptions. This also notes the potential to combine **quasi-experimental techniques** in DiD.

[1] X. Lu, W. Ai, Y. Wang, and Q. Mei, "Team Resilience under Shock: An Empirical Analysis of GitHub Repositories during Early COVID-19 Pandemic," Jan. 29, 2023, arXiv: arXiv:2301.12326. doi: [10.48550/arXiv.2301.12326](https://doi.org/10.48550/arXiv.2301.12326).

[2] C. Wing, K. Simon, and R. A. Bello-Gomez, "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research," Annual Review of Public Health, vol. 39, no. 1, pp. 453–469, Apr. 2018, doi: [10.1146/annurev-publhealth-040617-013507](https://doi.org/10.1146/annurev-publhealth-040617-013507).

Research gap

- Lacks information on AI/LLM trends on the Hacker News platform
- No studies into the effects of Hacker News submission on FOSS projects, especially now that AI projects are gaining popularity

By addressing this gap, there will be a more complete **insight into AI trends on HackerNews** which can be valuable for IT practitioners and enthusiasts.

Related Works

Studies based on HackerNews data

- Study of popularity dynamic in Reddit and Hacker News
 - Stoddard G. (2021) "Popularity Dynamics and Intrinsic Quality in Reddit and Hacker News", AAAI'21
- Study of Reddit and Hacker News as modern news aggregators for developers
 - Aniche et al. (2018) "How Modern News Aggregators Help Development Communities Shape and Share Knowledge", ICSE '18
- Study of Hacker News's effectiveness in software engineering research
 - Titus B et al. (2015) "I heart hacker news", FSE'15

HackerNews is a highly efficient and reliable source of high quality information

Studies about AI/LLM on other platforms

- Study of discussions related to deep learning across platforms such as Github and Stack Overflow
- Han J., Shihab E., Wan Z., Deng S., Xia X. (2020) "What Do Programmers Discuss about Deep Learning Frameworks",

The trend of AI/LLM is different on each platform so these platforms may need to be studied separately.

Sentiment Analysis in Software Engineering data

- Study of applying Sentiment Analysis models to software engineering related texts
 - Lin et al (2018) "Sentiment Analysis for Software Engineering: How Far Can We Go?",
- Study of pre-trained transformer models for sentiment analysis
 - Zhang et al. (2020) "Sentiment Analysis for Software Engineering: How Far Can Pre-trained Transformer Models Go?"
- Investigation on general capability of LLMs in sentiment analysis
 - Zhang et al. (2024) "Sentiment Analysis in the Era of Large Language Models: A Reality Check"
- Investigation on the prompt sensitivity of bigger language models (BLMs)
 - Zhang et al. (2025) "Revisiting Sentiment Analysis for Software Engineering in the Era of Large Language Models"

The models provided insufficient results regarding software engineer thus further fine-tuning may be required for improvement

Research gap

- However, existing research lacks information on AI/LLM trends on the HackerNews platform which presents an opportunity to study how AI technologies are used and received by the HackerNews community.
-

By addressing this gap, there will be a more complete insight into AI trends on HackerNews which can be valuable for IT practitioners and enthusiasts.

RQ3**What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?****Statistical Analysis of HN submission**

Determine whether the metric **increases are significant** after Hacker News submission, or are only due to **random noise**

Using **Wilcoxon signed-rank test**

- Non-parametric rank test
- Between metric changes **1 month before** and **1 month after** HN submission month (month 0)

To ensure **sufficient project activities**

- Only select HN GH-AI projects containing metrics **6 months before and after HN submission**

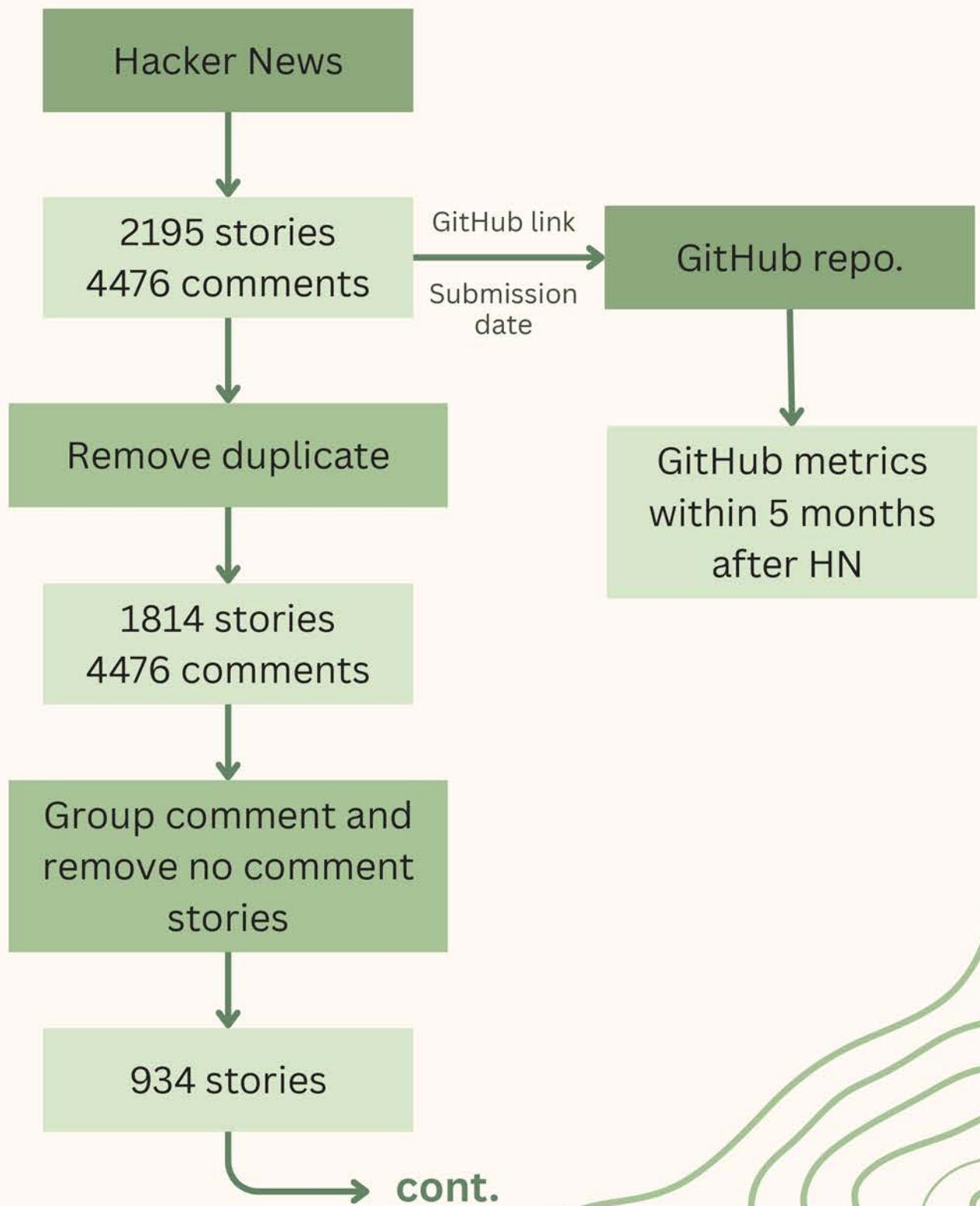
RQ3

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

Potential Impact of Hacker News Exposure

Metric changes based on sentiment on Hacker News

1. Sample **2195 stories** and **4476 comments** with their sentiment from Hacker News
2. Collect metrics data of GitHub projects that are linked to the sampled stories with the time of 5 months after the date of the project's Hacker News submission
3. Remove stories with duplicate GitHub project, group comments by their story, and remove stories with no comment



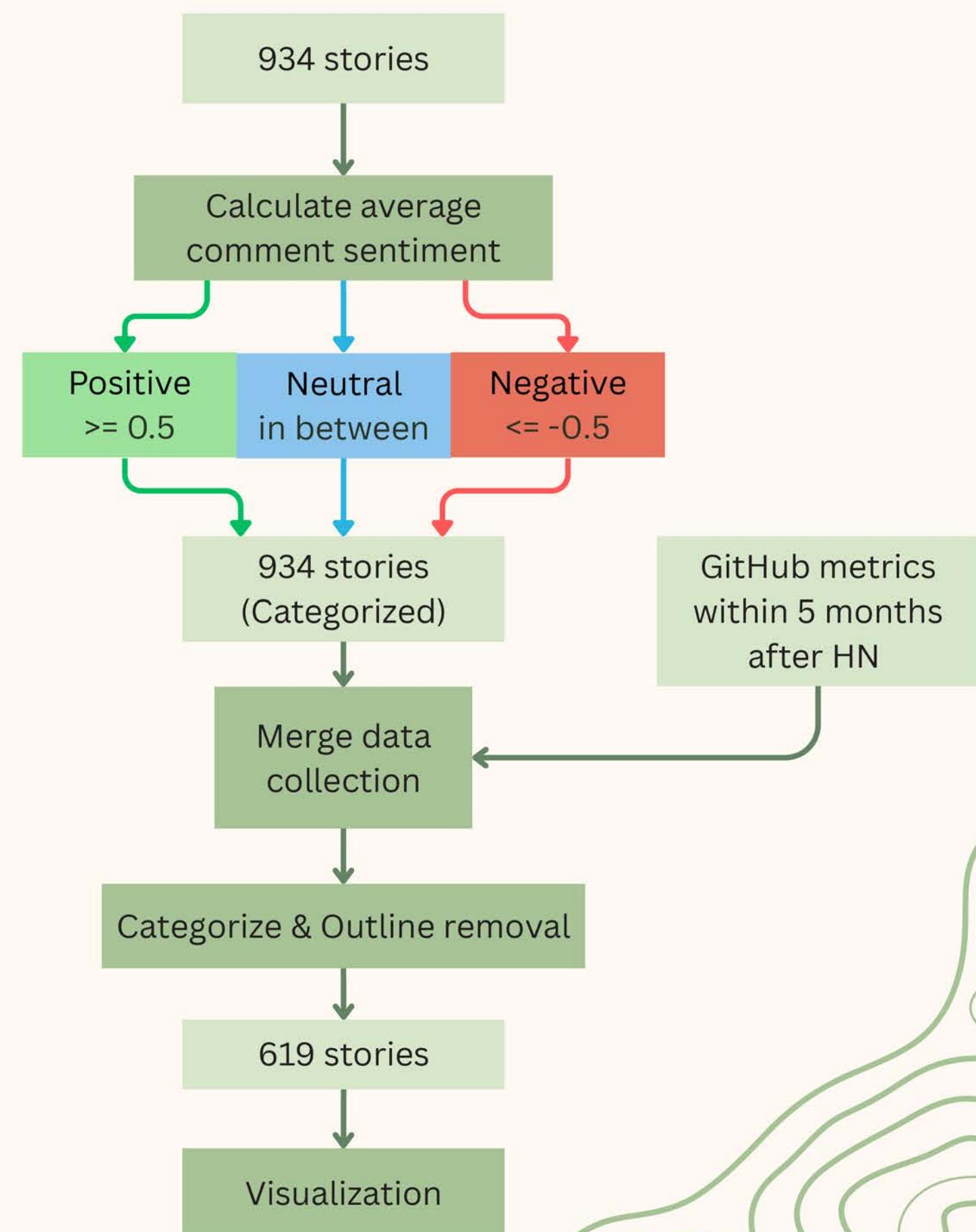
RQ3

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

Potential Impact of Hacker News Exposure (cont.)

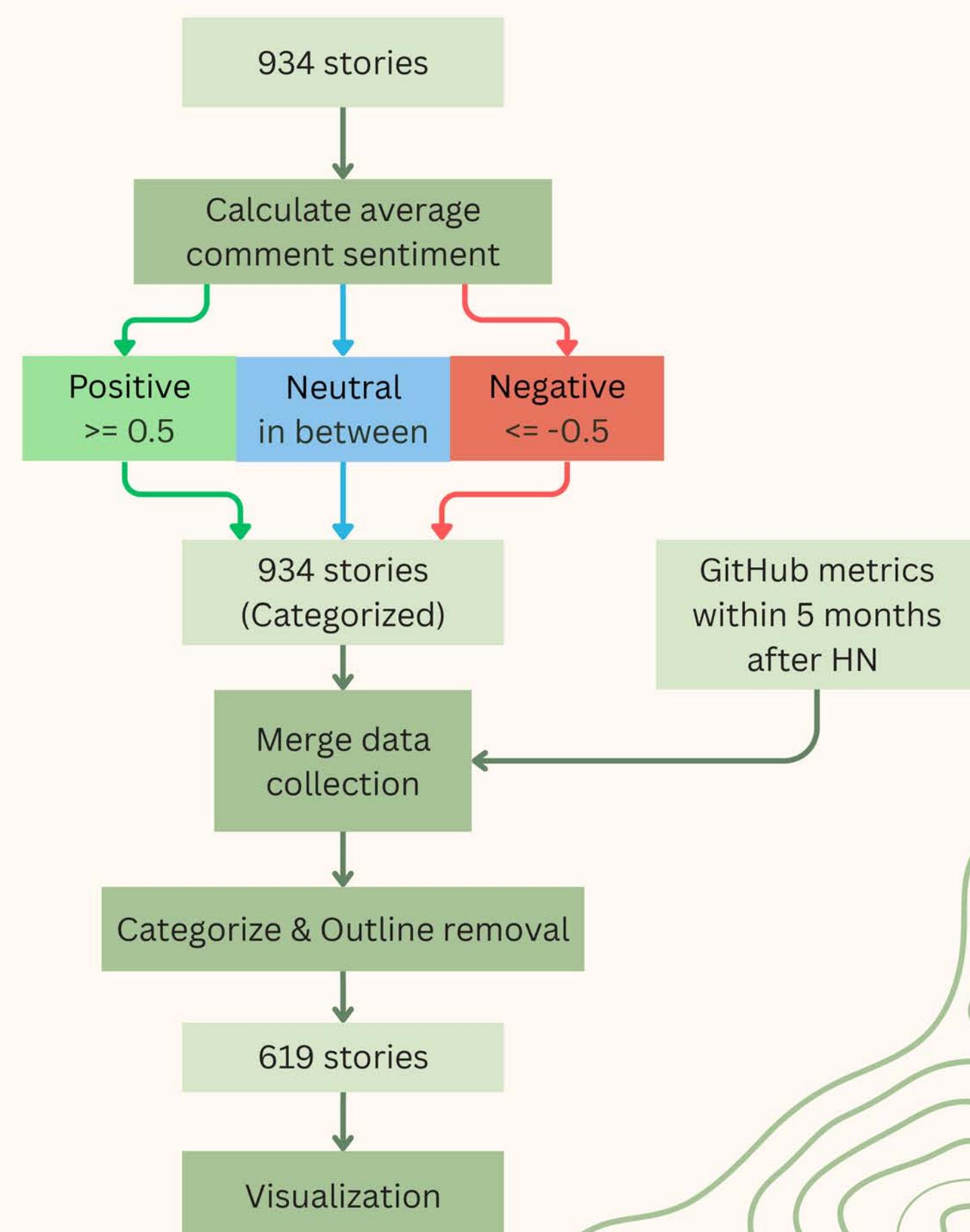
Metric changes based on sentiment on Hacker News

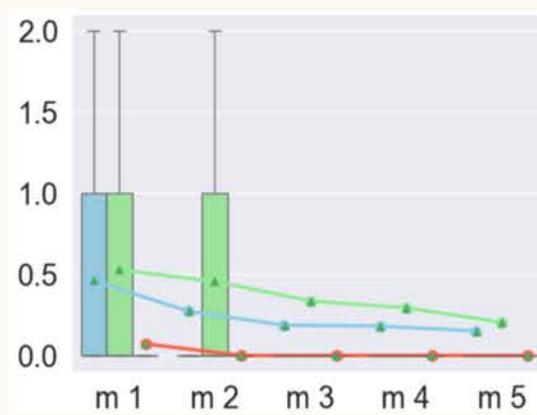
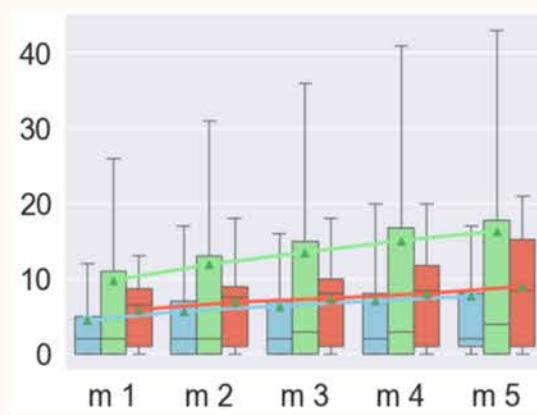
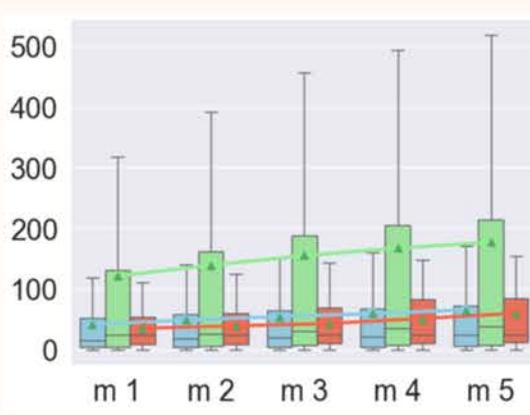
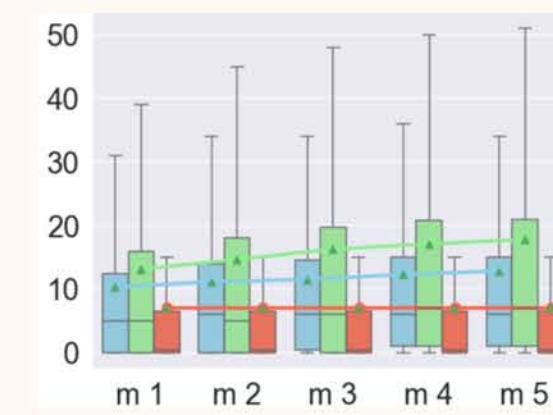
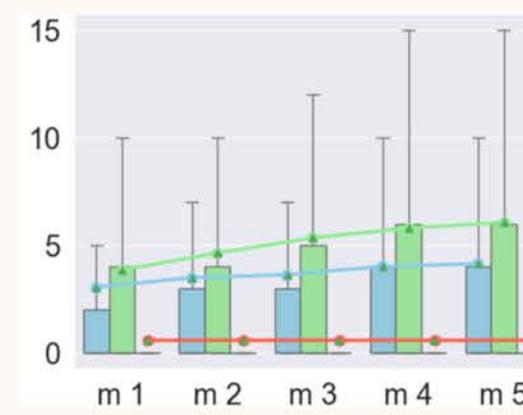
4. Calculate average comment sentiment of each story and categorize sentiment based on the following:
(≥ 0.5 : Positive , ≤ -0.5 : Negative, Else : Neutral)
5. Merge Hacker News story data collection with metrics data collection
6. categorize metrics data into groups based on their Hacker News sentiment and clear the outliers of each group
7. Visualize result using box plot for medians and point plot for means



RQ3

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



RQ3**What are the changes of activities in GitHub AI projects after being mentioned in Hacker News?****Metric changes based on sentiment on Hacker News****Contributors changes****Forks Growth****Stars Growth****Commits Growth****PRs Growth**

- Accumulative amount of each GitHub metric each month (except contributors that is non-cumulative) with each line show the progressive mean. The colors represent each sentiment group as color shown below:

■ : AI projects with **Neutral** sentiment on Hacker News

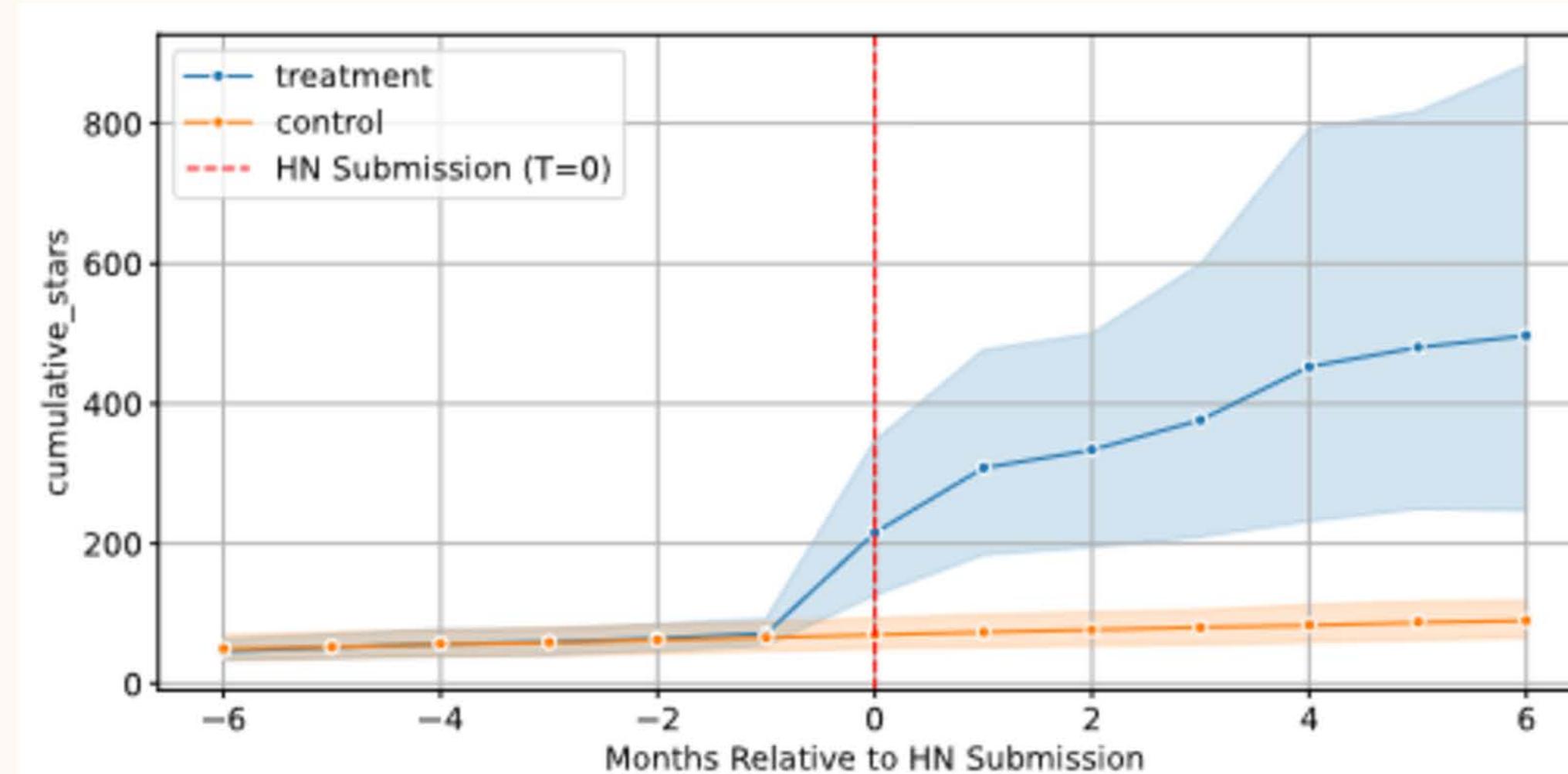
■ : AI projects with **Positive** sentiment on Hacker News

■ : AI projects with **Negative** sentiment on Hacker News

- The results show that **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

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

Example monthly **stars** metric (See appendix for all metric visualizations)