From keywords to conversations: How LLMs are transforming User Research Behavior

Fong Shu Hung 3036073077

TABLE OF CONTENTS

01

How the story began?

Just 2 guys randomly fuzzing around

02

Questions to answer

Methodology
WHAT ARE WE TRYNA DO HERE

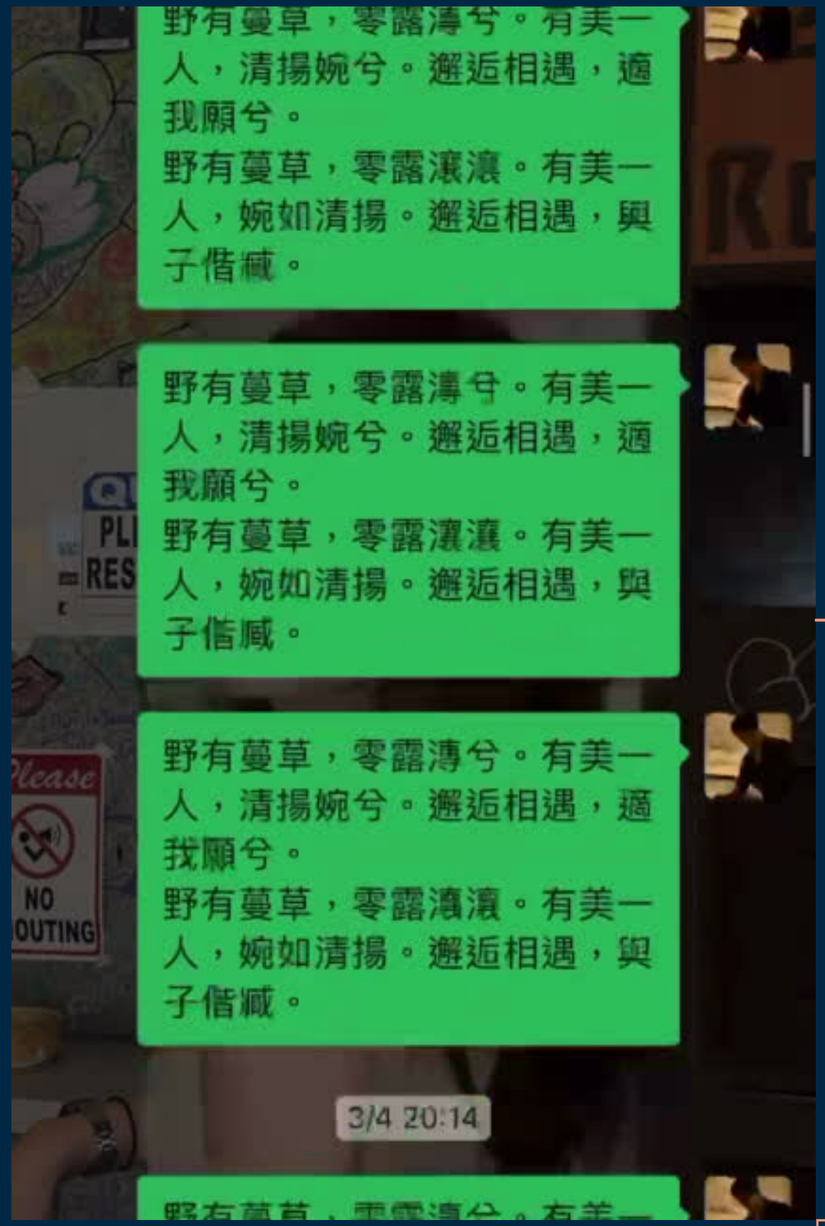
03

Conclusion

Limitations?

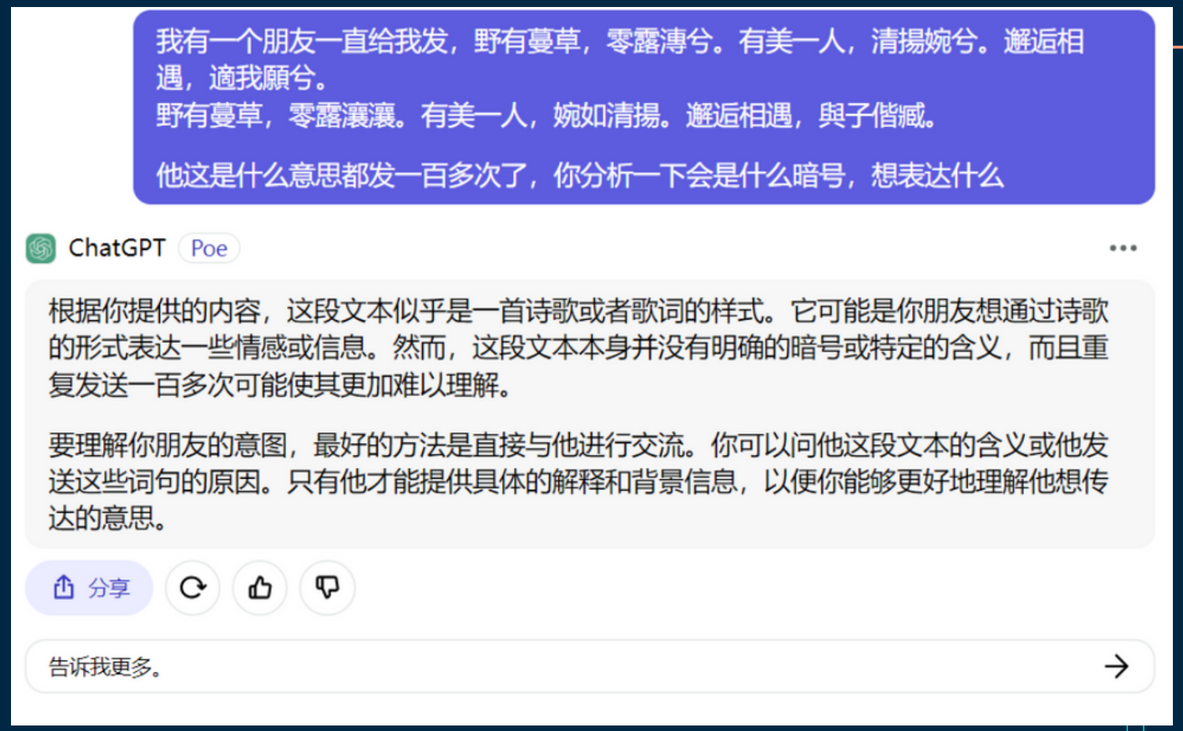
The tale starts with...

Just a crazy madman doing some random shid



The tale continues with...

Something that is counter-intuitive...



WHY AIN’T YOU ASKING GOOGLE FOR THIS???!

COMP2501

You are a professional university professor who is studying in the field of computer science, now you are teaching students R programing, in particular, you are focusing in the topic of "data visualization, data wrangling, string processing and text mining". **You will be provided with questions about R programming based on the above topics asked by the students, and you will be answering all those questions correctly**. Note that the answer you present should be short and concise and accurate, and **should be in the following format:**

First, State the correct answer of the MC question in the first line. Then give an explanation within 2 lines on why each choice is correct or incorrect on a separate line. For example:

A.

A. Correct because orange is a fruit

B. Incorrect because (reason)

C. Incorrect because (reason)

The student may provide his own insights and his mindset while interpreting the question. Please check if the student's understanding is correct, and provide a correct answer for the question. Thank you.

Prompt used for 2501 Midterm

Now, I decided to work on the following field: Which is discovering whether or not Large Language Models are replacing Searching engines gradually (or if there is a trend of this). More specifically, I wish to unveil if there is a potential shift in user behavior and preferences when they want to do some research regarding a question.

Please help me come up with a catchy topic

Claude-3-Opus Poe

Great choice! Exploring the potential shift in user behavior and preferences from search engines to Large Language Models (LLMs) is a timely and relevant research topic. Here are a few catchy titles for your research:

1. "The Language Model Exodus: Are Users Ditching Search Engines for AI Assistants?"
2. "Searching for Answers: The Rise of Large Language Models and the Future of Online Research"
3. "From Keywords to Conversations: How Large Language Models are Transforming User Research Behavior"

Title for this presentation

Just when I realized that we might be overly reliant on LLMs...

Questions arises...

1. Seems that we are using LLMs to tackle a problem instead of relying on searching engines

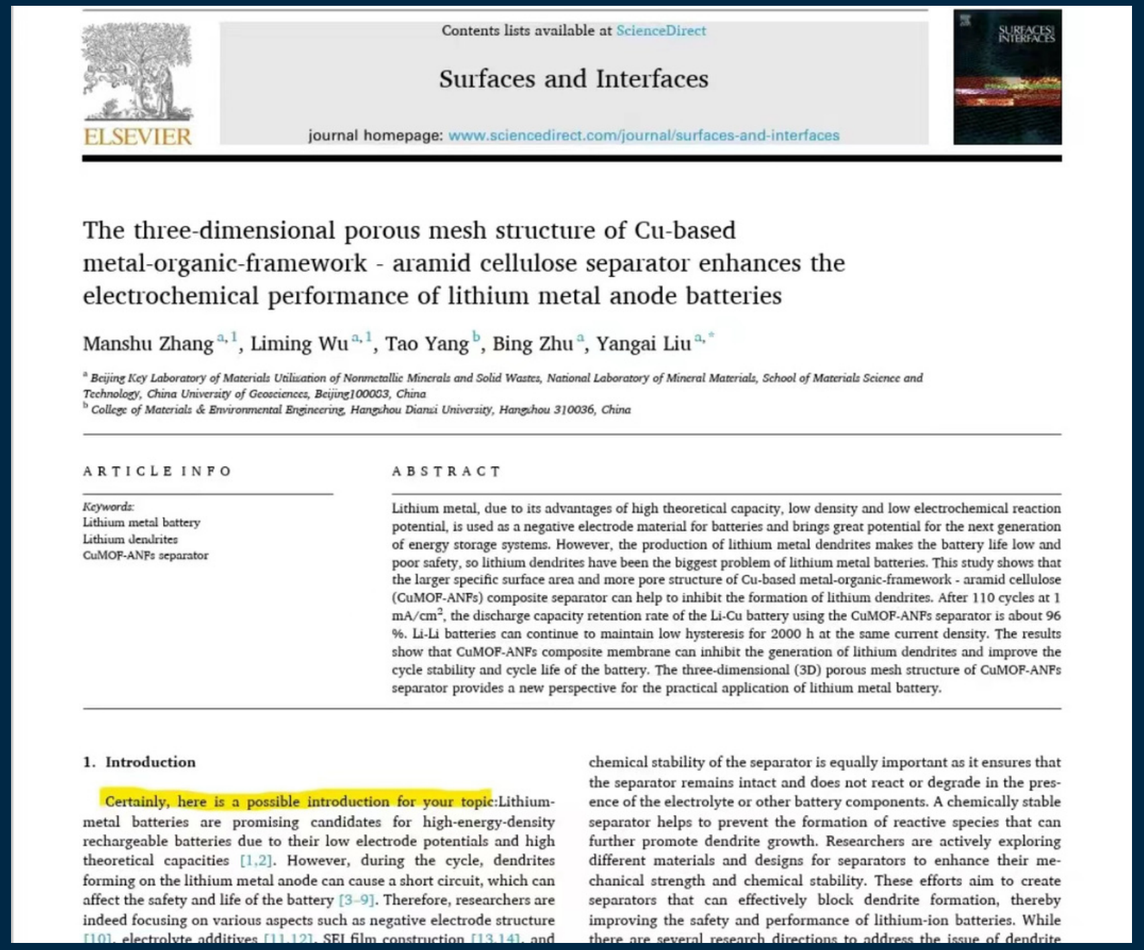
2. Forgot about the existence of searching engines

Problems concerned...

1. Flawed knowledge system

(What generated by GPT **is not 100% compared to the facts on the Internet**)

2. Renewal on how one should / can approach a new thing



Contents lists available at ScienceDirect

Surfaces and Interfaces

journal homepage: www.sciencedirect.com/journal/surfaces-and-interfaces

The three-dimensional porous mesh structure of Cu-based metal-organic-framework - aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries

Manshu Zhang [a,1], Liming Wu [a,1], Tao Yang [b], Bing Zhu [a], Yangai Liu [a,*]

[a] Beijing Key Laboratory of Materials Utilisation of Nonmetallic Minerals and Solid Wastes, National Laboratory of Mineral Materials, School of Materials Science and Technology, China University of Geosciences, Beijing100083, China

[b] College of Materials & Environmental Engineering, Hangzhou Dianzi University, Hangzhou 310036, China

ARTICLE INFO

Keywords:
Lithium metal battery
Lithium dendrites
CuMOF-ANFs separator

ABSTRACT

Lithium metal, due to its advantages of high theoretical capacity, low density and low electrochemical reaction potential, is used as a negative electrode material for batteries and brings great potential for the next generation of energy storage systems. However, the production of lithium metal dendrites makes the battery life low and poor safety, so lithium dendrites have been the biggest problem of lithium metal batteries. This study shows that the larger specific surface area and more pore structure of Cu-based metal-organic-framework - aramid cellulose (CuMOF-ANFs) composite separator can help to inhibit the formation of lithium dendrites. After 110 cycles at 1 mA/cm^2, the discharge capacity retention rate of the Li-Cu battery using the CuMOF-ANFs separator is about 96 %. Li-Li batteries can continue to maintain low hysteresis for 2000 h at the same current density. The results show that CuMOF-ANFs composite membrane can inhibit the generation of lithium dendrites and improve the cycle stability and cycle life of the battery. The three-dimensional (3D) porous mesh structure of CuMOF-ANFs separator provides a new perspective for the practical application of lithium metal battery.

1. Introduction

Certainly, here is a possible introduction for your topic:Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities [1,2]. However, during the cycle, dendrites forming on the lithium metal anode can cause a short circuit, which can affect the safety and life of the battery [3–9]. Therefore, researchers are indeed focusing on various aspects such as negative electrode structure [10], electrolyte additives [11,12], SEI film construction [13,14], and

chemical stability of the separator is equally important as it ensures that the separator remains intact and does not react or degrade in the presence of the electrolyte or other battery components. A chemically stable separator helps to prevent the formation of reactive species that can further promote dendrite growth. Researchers are actively exploring different materials and designs for separators to enhance their mechanical strength and chemical stability. These efforts aim to create separators that can effectively block dendrite formation, thereby improving the safety and performance of lithium-ion batteries. While there are several research directions to address the issue of dendrite

Sci research paper caught using GPT

That’s why I am doing the following research...

01 Is there really a shifting

02 How can I show if there is a shifting?

Methodology

By interview?
Nuh

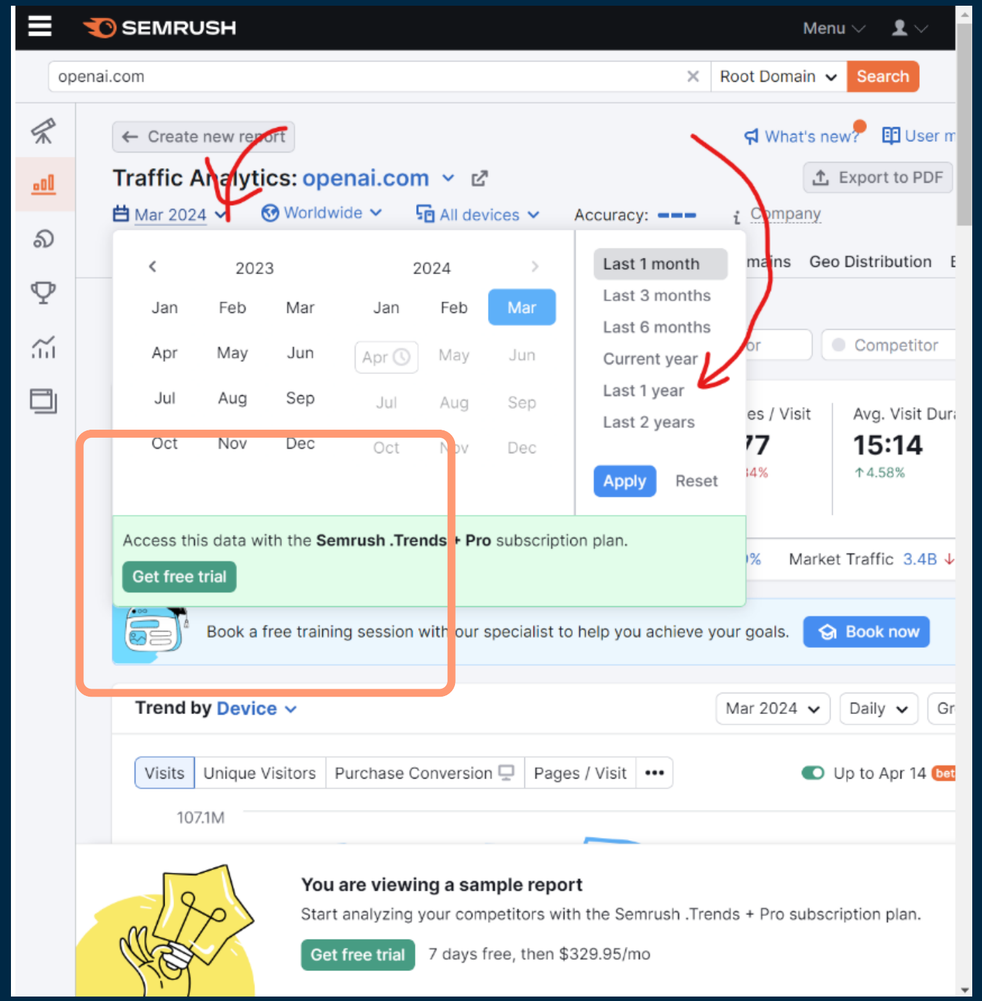
By becoming the CEO of google and get all the secretive data ?
Sounds alright

What remains that I can do is to get the traffic for various website...

- By comparing the traffic of searching engines & LLMs, we can see if there is a trend of changing searching behavior

Website traffic, is still something that is not easy to get...

I literally spent 2 hours tryna purchase it



N as a result my frd’ s visa was frozen
due to “suspecious transections”

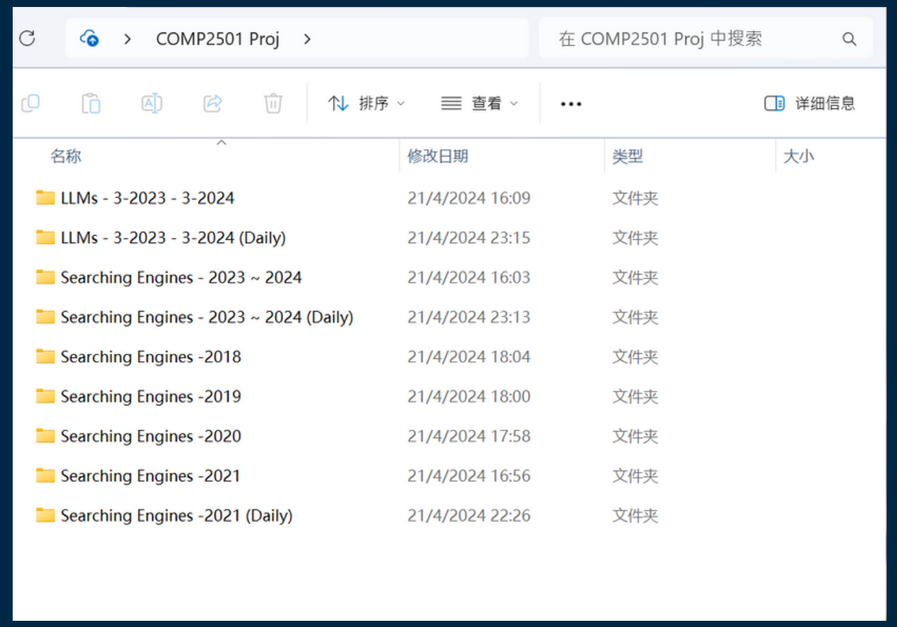
Anyways, we have the followings done

Before starting the visualization process:

1. Showing that before the launch of LLMs (2021) & after (Apr, 2023 - Mar, 2024)

FYI, GPT3.5 is launched on 12/2022

2. Selecting those with highest user-base / market share as representatives



Read in the data (Before Launch of LLMs)

```r
setwd("C:/Users/jimmy/Desktop")

file_path <- "COMP2501 Proj/Searching Engines -2021/Trend

yahoo_data_2021 <- read.csv(file_path)

head(yahoo_data_2021)
```

I choose to consider the total number of visit to a website rather than unique visitors:

Focus more on “Eureka ”- When one faces a problem, would they think of using google / GPT at that moment ?

Thus, comparison of total number of visit is like probability of thinking of LLMs

```
> yahoo_data_2021 <- read.csv(file_path)

> head(yahoo_data_2021)
         X All.Devices    Desktop     Mobile
1 Jan 2021  7266924924 5416701302 1850223622
2 Feb 2021  6143705953 4470732229 1672973724
3 Mar 2021  6627432485 4706000619 1921431866
4 Apr 2021 10590825296 4864757811 5726067485
5 May 2021  9484590882 4582744310 4901846572
6 Jun 2021  7354661436 3413266018 3941395418
```

Same applies for Google, Bing

Read in the data (Before Launch of LLMs)

```r
searching_engines_2021 <- left_join(bing_data_2021, google_data_2021, by = "X") |>
  left_join(yahoo_data_2021, by = "X") |>
  mutate(Total.Devices = All.Devices.x + All.Devices.y + All.Devices) |>
  select(X, Total.Devices)

head(searching_engines_2021)
```

```
> head(searching_engines_2021)
           X Total.Devices
1 Jan 1, 2021    3091248631
2 Jan 2, 2021    3100654906
3 Jan 3, 2021    3137733147
4 Jan 4, 2021    3434057003
5 Jan 5, 2021    3487970924
6 Jan 6, 2021    3464074392
```

No. visits to searching engines each day in 2021

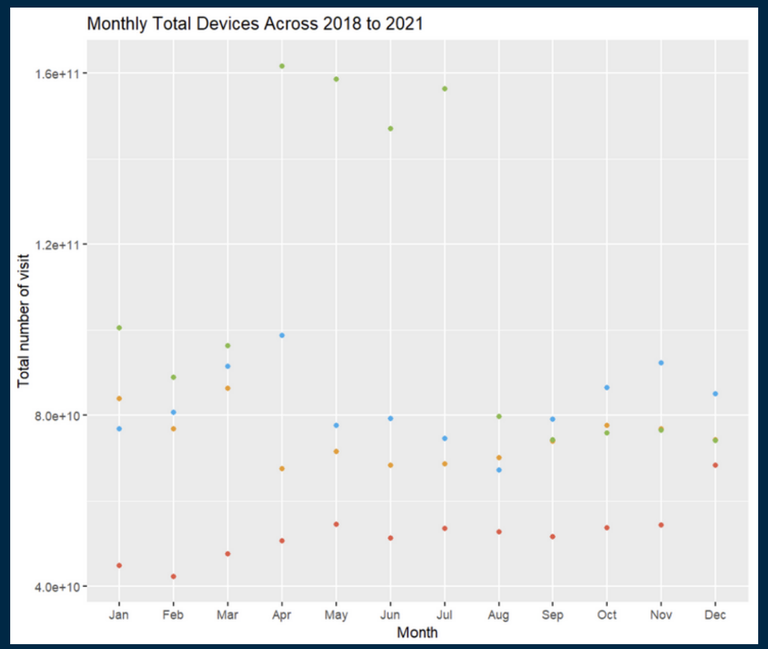
Reshaping the data

As the goal is to show if 2 things are related
-> Linear regression

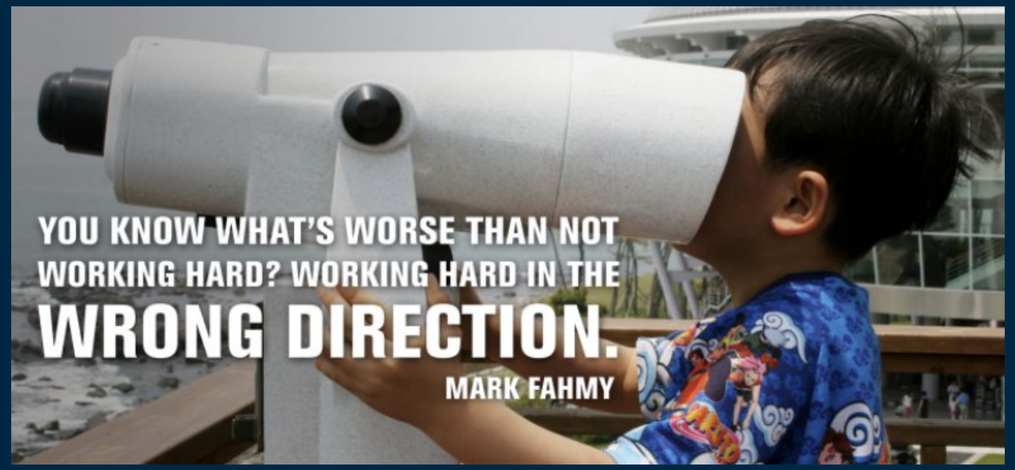
Currently, y (number of visit), x (**DATE**)
While linear regression needs 2 continuous variable
(Date is not continuous data)

```r
library(tidyverse)
library(ggthemes)
library(lubridate)

searching_engines_2021 <- searching_engines_2021 |>
  mutate(X = mdy(X),
         X = format(X, "%Y-%m-%d")) |>
  mutate(X = as.Date(X))

str(searching_engines_2021$X)

reference_date <- as.Date("2021-01-01")

searching_engines_2021 <- searching_engines_2021 |>
  mutate(Days_Since_2021 = as.numeric(X - reference_date))
```

Changing to the number of days passed



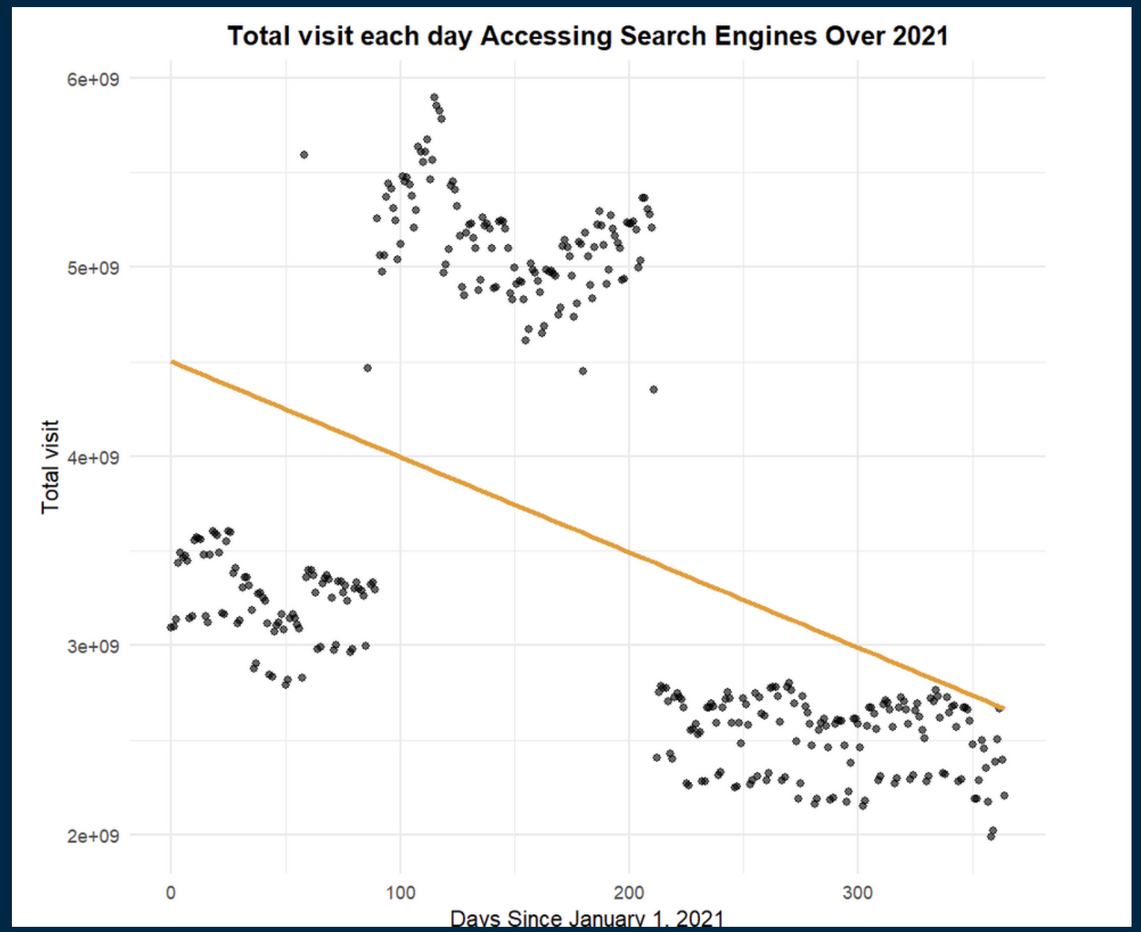
Otherwise, you will get some bs like this



Visualization, 啓動！

```r
library(tidyverse)
library(ggthemes)

plot_for_searching_engines_2021 <- searching_engines_2021 |>
  ggplot(aes(x = Days_Since_2021, y = Total.Devices)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "#ef9b20") +
  labs(
    title = "Total visit each day Accessing Search Engines Over 2021",
    x = "Days Since January 1, 2021",
    y = "Total visit",
    caption = "Data collected throughout 2021"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    plot.caption = element_text(face = "italic")
  )

# Display the plot
print(plot_for_searching_engines_2021)
```

Outliers?



Seems that the trend for using searching engines is decreasing
(Regardless of outliers)

And here we go again... (After Launch of LLMs)

Basically the same things... But

175	Sep 22, 2023	1524182	475919	1048263	69916588	31339120	38577468	n/a	48134	n/a
176	Sep 23, 2023	1494131	411930	1082201	56526531	21000260	35526271	n/a	45375	n/a
177	Sep 24, 2023	1561526	426471	1324119	57692074	22885541	34806533	n/a	39818	n/a
178	Sep 25, 2023	1498885	548462	1150086	77494911	36828533	40666378	n/a	14790	n/a
179	Sep 26, 2023	1548590	616408	932182	81891771	38871699	43020072	n/a	78442	n/a
180	Sep 27, 2023	1972857	614891	1357966	79000404	37640200	41360204	n/a	65004	n/a
181	Sep 28, 2023	1614848	553611	1061237	77312046	36722705	40589341	n/a	57828	n/a
182	Sep 29, 2023	1832770	507422	1325348	68992898	32404116	36588782	n/a	45719	n/a
183	Sep 30, 2023	1544111	419259	1123792	47970242	22434099	32118731	n/a	44082	n/a
184	Oct 1, 2023	1785295	552713	1331091	64124706	23666193	40458513	4393	37653	n/a
185	Oct 2, 2023	1906322	680849	1351739	86822564	36918078	49904486	0606	58506	n/a
186	Oct 3, 2023	2319032	600894	1718138	90206222	38090462	52115760	2157	52326	n/a
187	Oct 4, 2023	1995186	699802	1400064	89297139	37692180	51604959	3847	55865	n/a
188	Oct 5, 2023	2001649	582568	1302769	89903947	37522556	52381391	9505	56785	n/a
189	Oct 6, 2023	1961921	529922	1431999	82022363	33073276	48949087	2292	64918	n/a
190	Oct 7, 2023	1989341	627003	1529299	64599863	22300447	42299416	3913	49325	n/a
191	Oct 8, 2023	2074167	552713	1521454	66937265	23685686	43251579	1686	48926	n/a
192	Oct 9, 2023	2575547	648966	1926581	88358889	37255284	51103605	51039	58506	n/a
193	Oct 10, 2023	2198507	599087	1599420	90260588	38432833	51827755	2157	52326	n/a
194	Oct 11, 2023	2147922	735352	1412570	89014570	38995196	50019374	2090	56362	n/a

The data seems a bit messed up

We are considering OpenAI, Antrophic, POE, this time

And here we go again... (After Launch of LLMs)

```r
LLMs_2024 <- left_join(poe_data_2024, openai_data_2024, by = "X") |>
  left_join(anthropic_data_2024, by = "X") |>
  mutate(All.Devices = ifelse(All.Devices == "n/a", NA, All.Devices)) |>
  mutate(All.Devices = replace(as.numeric(All.Devices), is.na(All.Devices), 0)) |>
  mutate(Total.Devices = All.Devices.x + All.Devices.y + All.Devices) |>
  select(X, Total.Devices)

head(LLMs_2024)
```

The n/a are strings, thus we have to convert them to NA first...

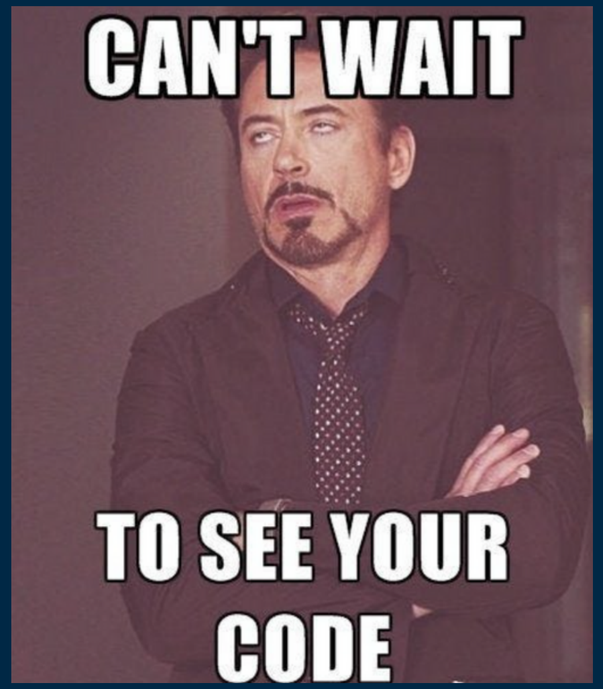
But there are less than 10 rows that are slightly flawed, and since the value is small... It suffices to regard them as 0

```
> head(LLMs_2024)
             X Total.Devices
1 Apr 1, 2023      79685472
2 Apr 2, 2023      81685294
3 Apr 3, 2023      98287501
4 Apr 4, 2023     118234161
5 Apr 5, 2023     105311862
6 Apr 6, 2023     100900884
```

And here we go again... (After Launch of LLMs)

```r
reference_date <- as.Date("2023-04-01")

searching_engines_2024 <- searching_engines_2024 |>
  mutate(Days_Since_2023 = as.numeric(X - reference_date))


LLMs_2024 <- LLMs_2024 |>
  mutate(Days_Since_2023 = as.numeric(X - reference_date))

head(LLMs_2024)
head(searching_engines_2024)
```

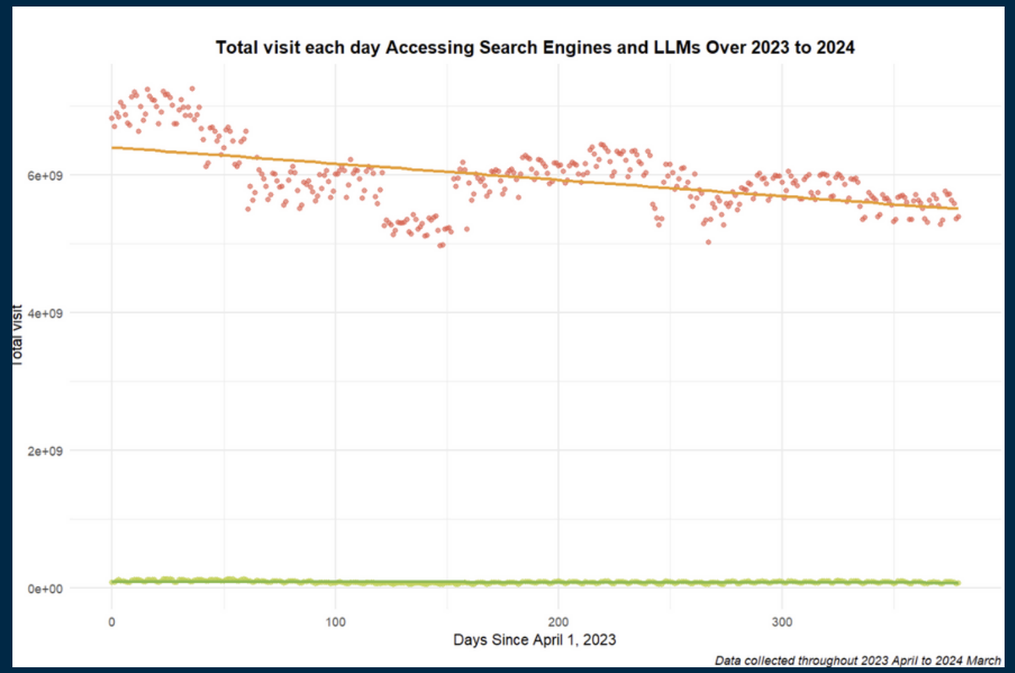
```
> head(LLMs_2024)
           X Total.Devices Days_Since_2023
1 2023-04-01      79685472               0
2 2023-04-02      81685294               1
3 2023-04-03      98287501               2
4 2023-04-04     118234161               3
5 2023-04-05     105311862               4
6 2023-04-06     100900884               5

> head(searching_engines_2024)
           X Total.Devices Days_Since_2023
1 2023-04-01    6821367610               0
2 2023-04-02    6694122482               1
3 2023-04-03    6895712610               2
4 2023-04-04    6834007586               3
5 2023-04-05    7046034097               4
6 2023-04-06    6983816136               5
> View(LLMs_2024)
```



The moment that all of us have been waiting for...

Comparison Plot

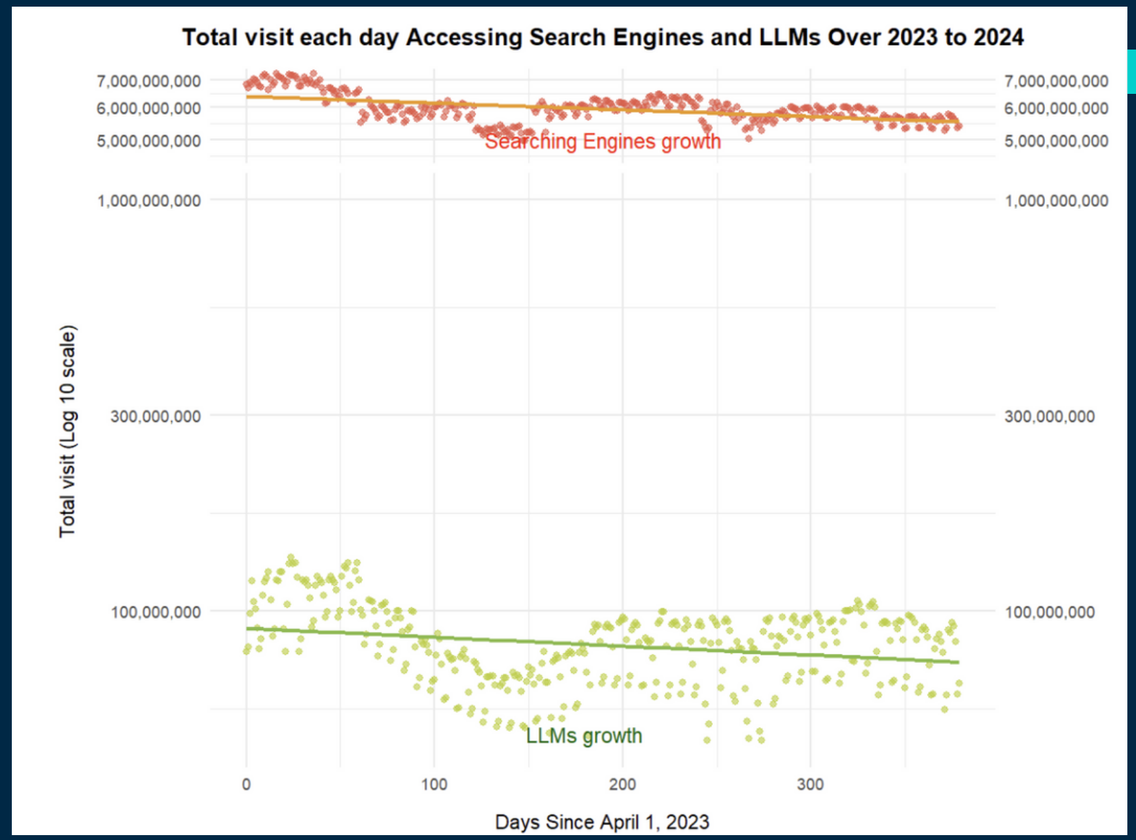
```r
library(tidyverse)
library(ggthemes)

plot_for_searching_engines_and_LLMs_2024 <- ggplot(searching_engines_2024, aes(x = Days_Since_2023, y = Total.Devices)) +
  geom_point(alpha = 0.6, color = "#ea5545") +
  geom_smooth(method = "lm", se = FALSE, color = "#ef9b20") +
  geom_point(data = LLMs_2024, alpha = 0.6, color = "#bdcf32") +
  geom_smooth(data = LLMs_2024, method = "lm", se = FALSE, color = "#87bc45") +
  labs(
    title = "Total visit each day Accessing Search Engines and LLMs Over 2023 to 2024",
    x = "Days Since April 1, 2023",
    y = "Total visit",
    caption = "Data collected throughout 2023 April to 2024 March"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    plot.caption = element_text(face = "italic")
  )


print(plot_for_searching_engines_and_LLMs_2024)
```



Hang on, it didn't go as expected...

Comparison Plot (V2.)

```r
library(tidyverse)
library(ggthemes)
library(ggbreak)

plot_for_searching_engines_and_LLMs_2024 <- ggplot(searching_engines_2024, aes(x = Days_Since_2023, y = Total.Devices)) +
  geom_point(alpha = 0.6, color = "#ea5545") +
  geom_smooth(method = "lm", se = FALSE, color = "#ef9b20") +
  geom_point(data = LLMs_2024, alpha = 0.6, color = "#bdcf32") +
  geom_smooth(data = LLMs_2024, method = "lm", se = FALSE, color = "#87bc45") +
  coord_cartesian(ylim = c(1, 1e+08)) +
  scale_y_break(c(1e+09, 4.5e+09)) +
  scale_y_log10(labels = scales::comma) +
  labs(
    title = "Total visit each day Accessing Search Engines and LLMs Over 2023 to 2024",
    x = "Days Since April 1, 2023",
    y = "Total visit (Log 10 scale)",
    caption = "Data collected throughout 2023 April to 2024 March"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    plot.caption = element_text(face = "italic")
  ) +
  annotate("text", x = 180, y = 5e+07, label = "LLMs growth", color = "darkgreen", size = 4) +
  annotate("text", x = 190, y = 5e+09, label = "Searching Engines growth", color = "Red", size = 4)


print(plot_for_searching_engines_and_LLMs_2024)
```



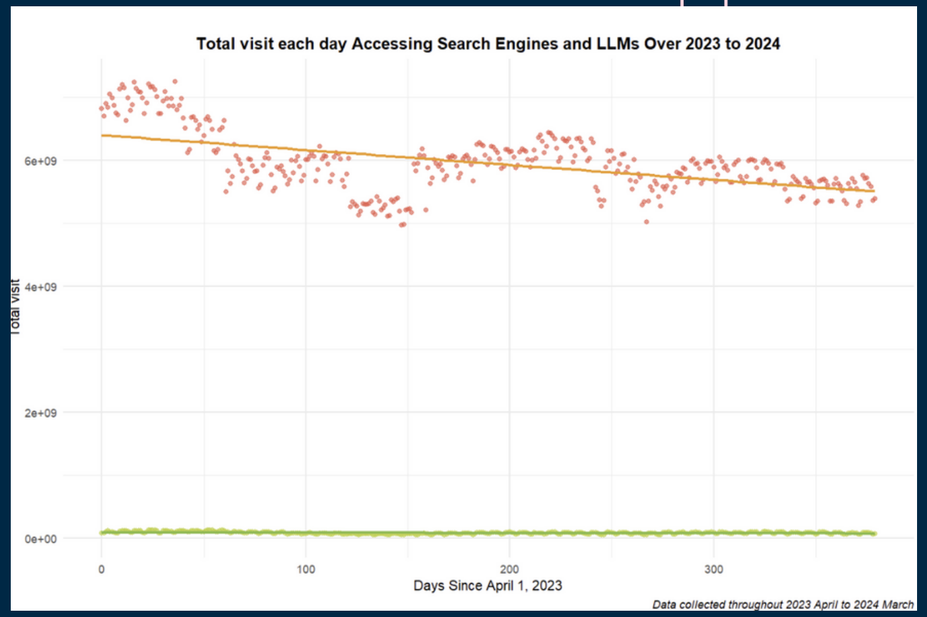
Changes:

1. loonnggggg 10 the y-axis (Longggg 9 plz dun sue me)
2. Break the y-interval that does not have any values
3. zoom in on the interval for LLMs

Logicial deductions?

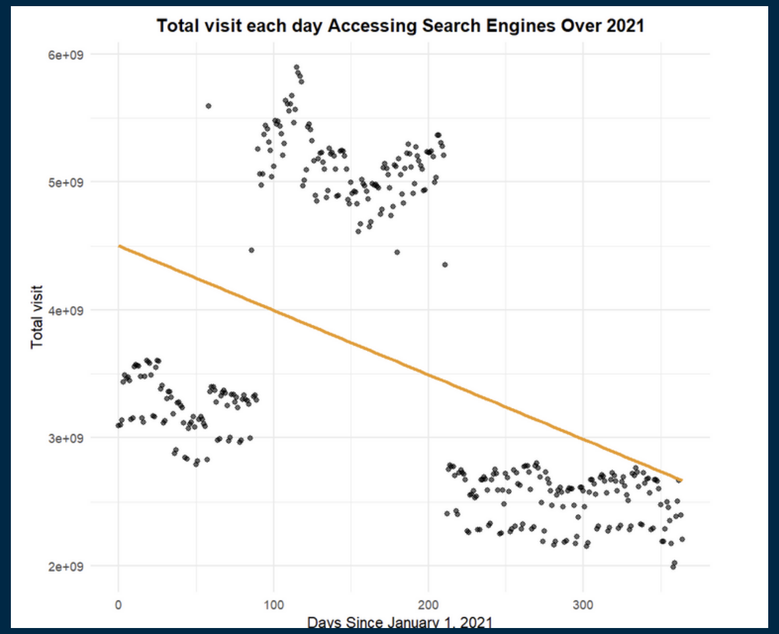
Well I mean, I haven't take the course STAT4600...
So no stat model this time

Use common sense bruh



1. Both regression line for searching engines had been decreasing since 2021
2. While the user base have been increasing

-> The total visits per day is declining, but absolute no of users increased
-> Searching engines are still attracting new users, but engagement for visits are lowered

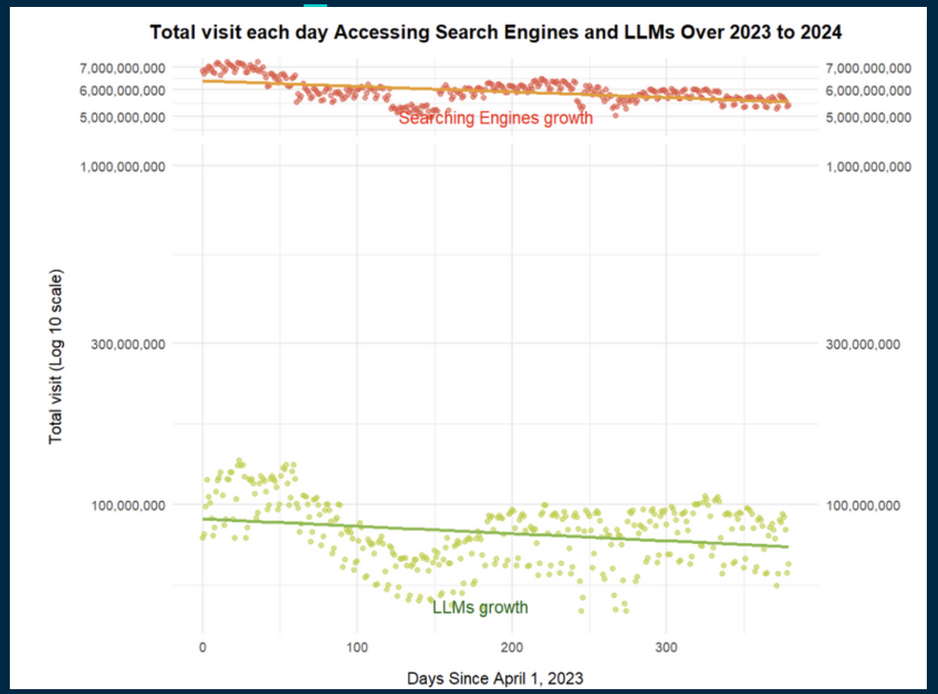


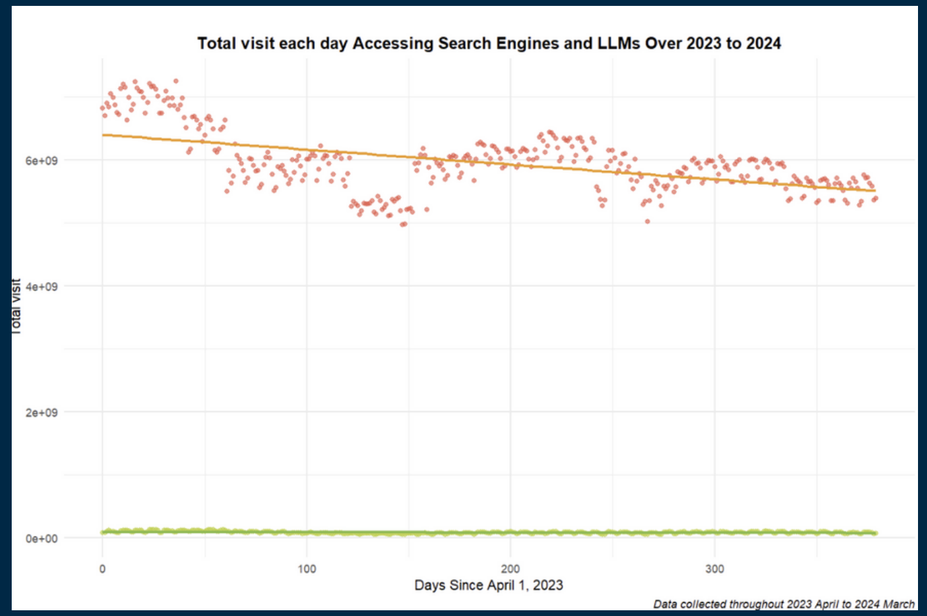
Logicial deductions?

3. The faster Declining trend for search engines suggest LLMs are more resilient / attractive to users

4. Whlie both are declining ??...

-> Users maybe moving towards alternative information platform...
=> LLMs can be one of them





Do you have any questions?

THANKS

next presenter