Place your b(e|o)ts:

Recent Trends in LLM-based Advertising

Andrey Meshkov
CTO and Co-Founder of AdGuard
am@adguard.com
@ay_meshkov

Natalia Sokolova
UX Writer & Researcher, AdGuard
n.sokolova@adguard.com

Introduction

What is this talk about?

Previously on AFDS 2023, we…

- Pointed out the most likely ways to serve ads via LLM-based chatbots
- Introduced the concept of blended ads
- Suggested several approaches to blocking them

Watch the video on Youtube:
https://youtu.be/ZloL4APC1lc

Key trends over the past year

Industry

- Key players manifested in 2023.
- Ad creation has become cheaper.
- July 2024, the first LLM for ad creation and evaluation
- October 2024, Google and Bing announced the rollout of ads in their generative AI search results

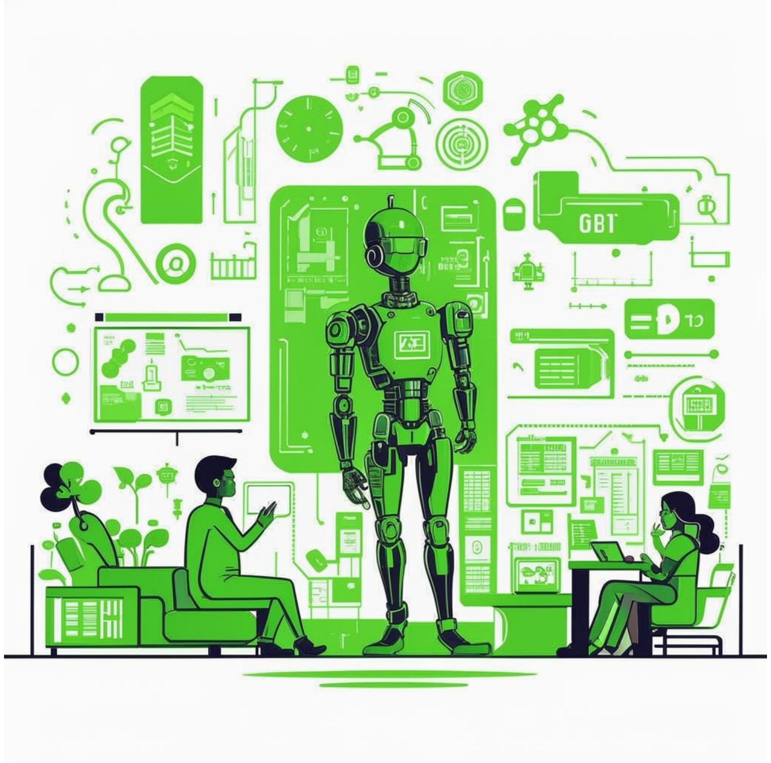
Academia

- The second half of 2023 saw an increasing number of research articles published, mostly by Google-affiliated researchers.
- Focus on dynamic and blended ad creation

What is this talk about?

- What is an LLM?
- How are LLMs used today?
- How much does it cost?
- Who pays for that?
- More ads coming soon
- Blended ads
- Ad blocking and LLMs
- Final words

Recent trends in LLM-based advertising

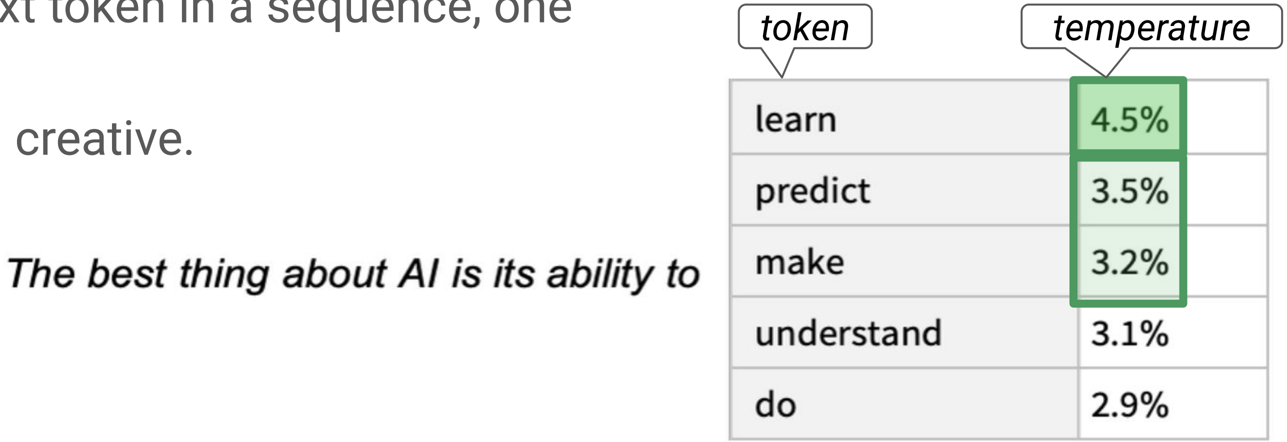


What is an LLM?

LLMs in a nutshell

LLMs in a nutshell

- A Large Language Model (LLM) is an algorithm with billions of parameters trained on a **huge** corpus of text.
- LLM infers the next token in a sequence, one token at a time.
- Fever makes LLM creative.


token
temperature

The best thing about AI is its ability to

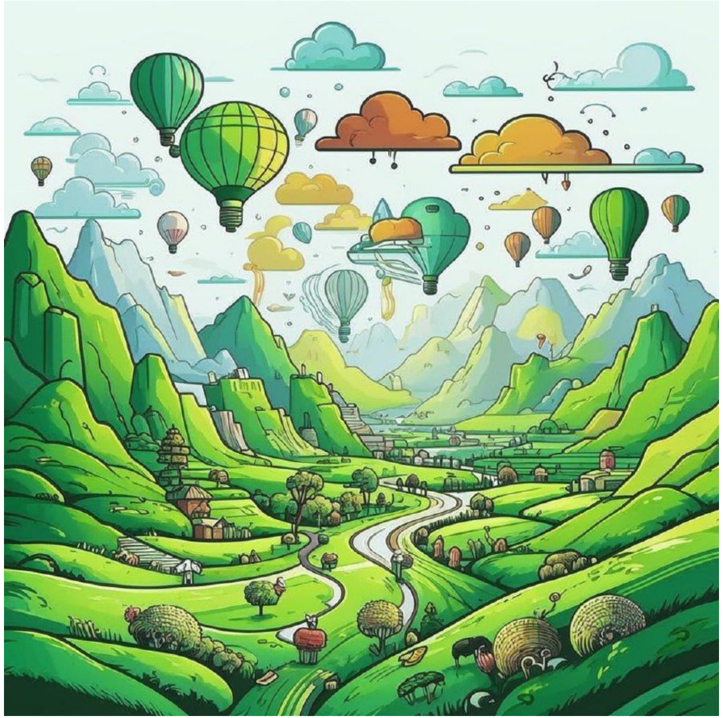
learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

Image source: https://writings.stephenwolfram.com

Important concepts

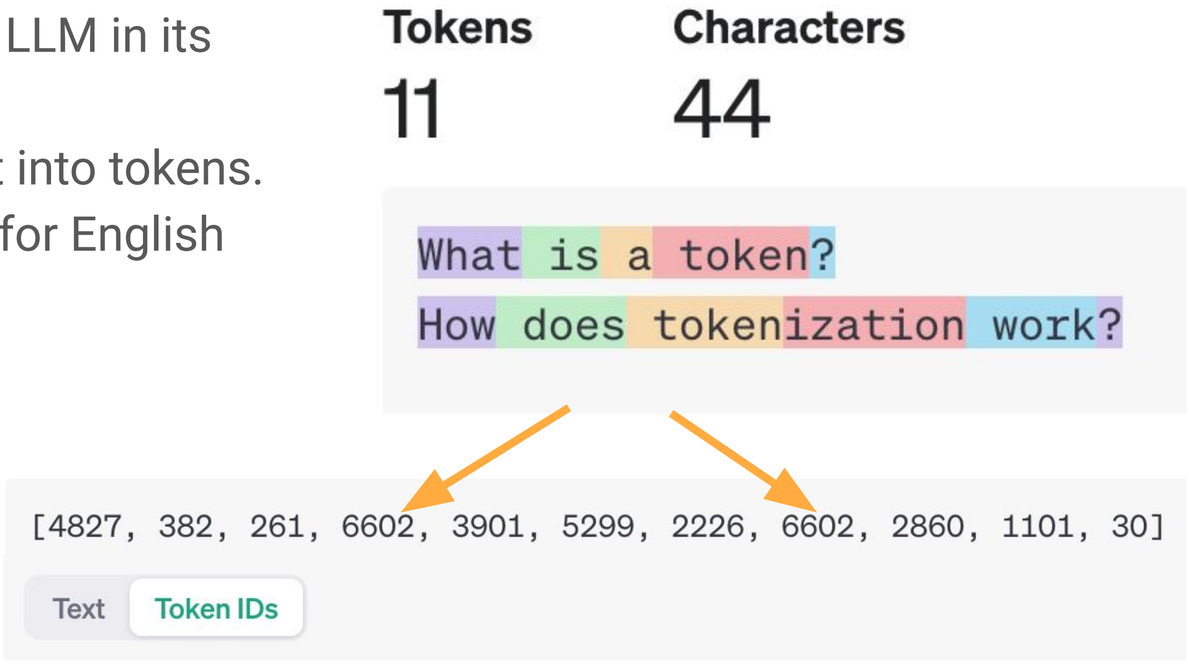
The main two concepts to understand how LLMs work

- Tokens
- Embeddings and “meaning space”



What is a token?

- A token is a common sequence of characters learned by an LLM in its training text.
- To tokenize: to break text into tokens.
- 1 token is ~4 characters for English text (~0.75 of a word).



Tokens
11

Characters
44

What is a token?
How does tokenization work?

[4827, 382, 261, 6602, 3901, 5299, 2226, 6602, 2860, 1101, 30]

Embedding and “meaning space”

- An **embedding** represents a word meaning by an array of numbers.
- Nearby meanings are represented by nearby numbers.
- Embeddings arrange words in a “**meaning space**” so that words that ar “closer in meaning” appear closer together

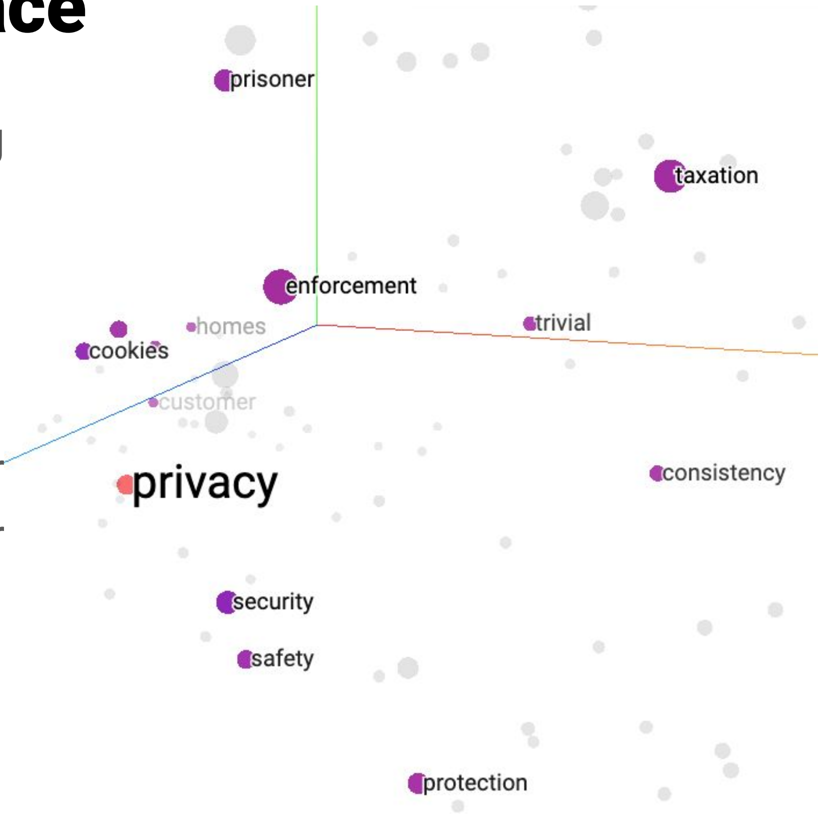
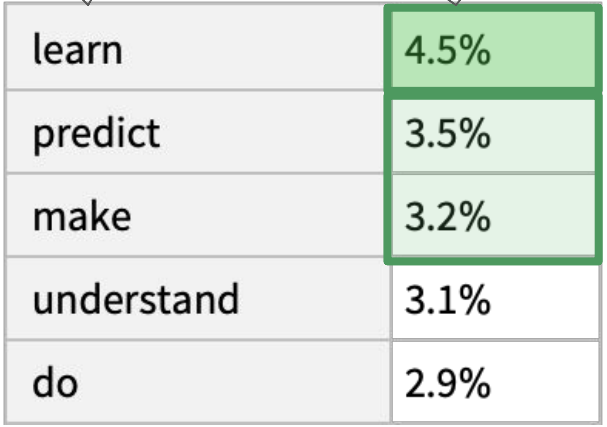


Image created in: https://projector.tensorflow.org

This is how it looks in the meaning space

The best thing about AI is its ability to…



token	*temperature*
learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

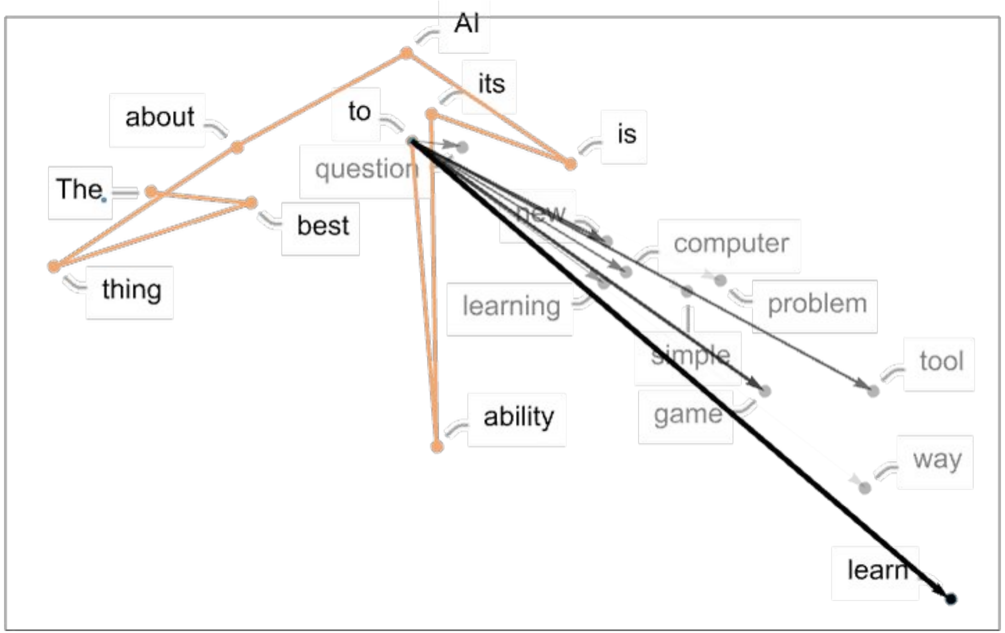


Image source: https://writings.stephenwolfram.com

How are LLMs used today?

- Chatbots
- Agents
- Search

Each of these uses is a potential advertising space…

Chatbots

2022

Nov — **ChatGPT**
OpenAI, GPT-4

2023

Feb — **Poe Chatbot**
Quora, GPT & Claude

Mar — **DuckAssist**
DuckDuckGo, GPT & Anthropic

Mar — **Ernie AI**
Baidu, Ernie & Plato

Apr — **My AI**
Snapchat, GPT

Apr — **Alisa (Yandex)** switches to LLM

May — **Bing AI**
Microsoft, GPT & Prometheus

May — **Bard**
Google, LaMDA & PaLM

Jul — **Claude AI**
Anthropic, Claude 2

Sep — **GigaChat AI**
Sberbank, ruGPT

Sep — **Alexa** switches to LLM

Dec — **Gemini AI**
Google, Gemini

Dec — **Google Assistant** switches to LLM

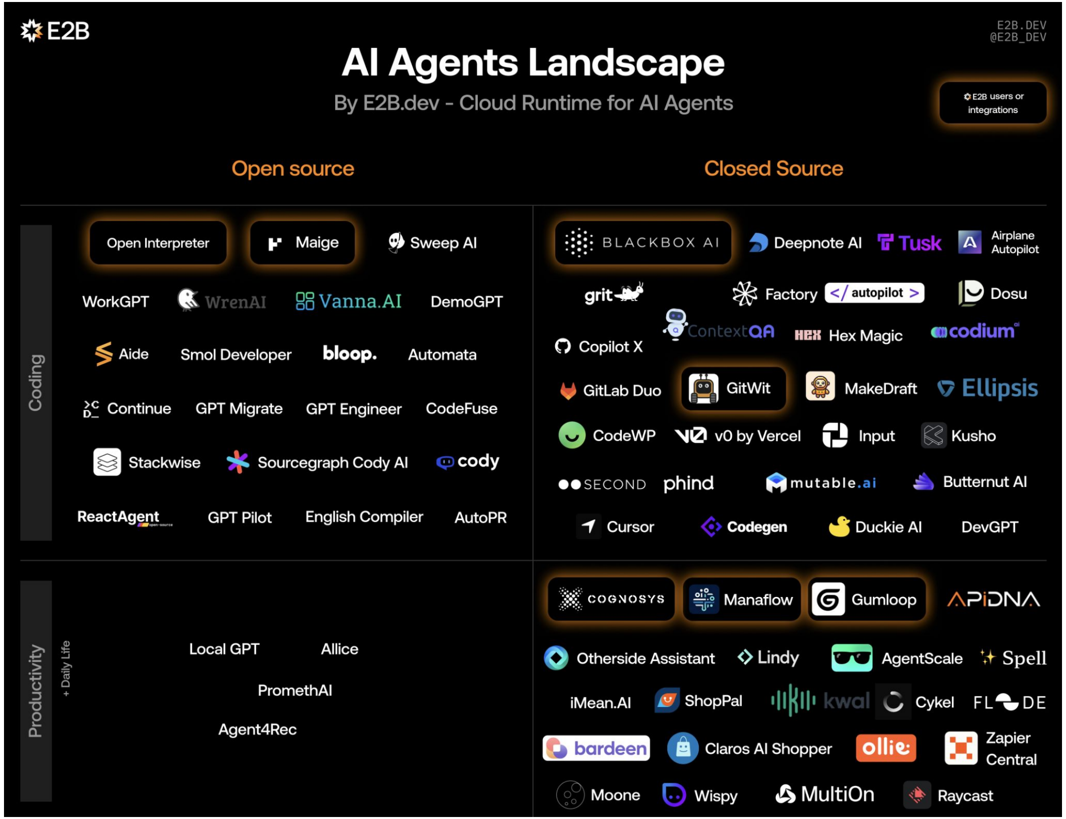
2024

Oct — **Siri** switches to LLM

AI Agents

An "AI Agent" refers to a software entity that can perform tasks autonomously by making decisions and taking actions based on input data and environmental observations.

Image source: e2b.dev



E2B

E2B.DEV
@E2B_DEV

AI Agents Landscape

By E2B.dev - Cloud Runtime for AI Agents

E2B users or integrations

Open source

Coding

- Open Interpreter
- Maige
- Sweep AI
- WorkGPT
- WrenAI
- Vanna.AI
- DemoGPT
- Aide
- Smol Developer
- bloop.
- Automata
- Continue
- GPT Migrate
- GPT Engineer
- CodeFuse
- Stackwise
- Sourcegraph Cody AI
- cody
- ReactAgent
- GPT Pilot
- English Compiler
- AutoPR

Productivity + Daily Life

- Local GPT
- Allice
- PromethAI
- Agent4Rec

Closed Source

Coding

- BLACKBOX AI
- Deepnote AI
- Tusk
- Airplane Autopilot
- grit
- Factory
- </ autopilot >
- Dosu
- ContextQA
- Hex Magic
- codium
- Copilot X
- GitLab Duo
- GitWit
- MakeDraft
- Ellipsis
- CodeWP
- v0 by Vercel
- Input
- Kusho
- SECOND
- phind
- mutable.ai
- Butternut AI
- Cursor
- Codegen
- Duckie AI
- DevGPT

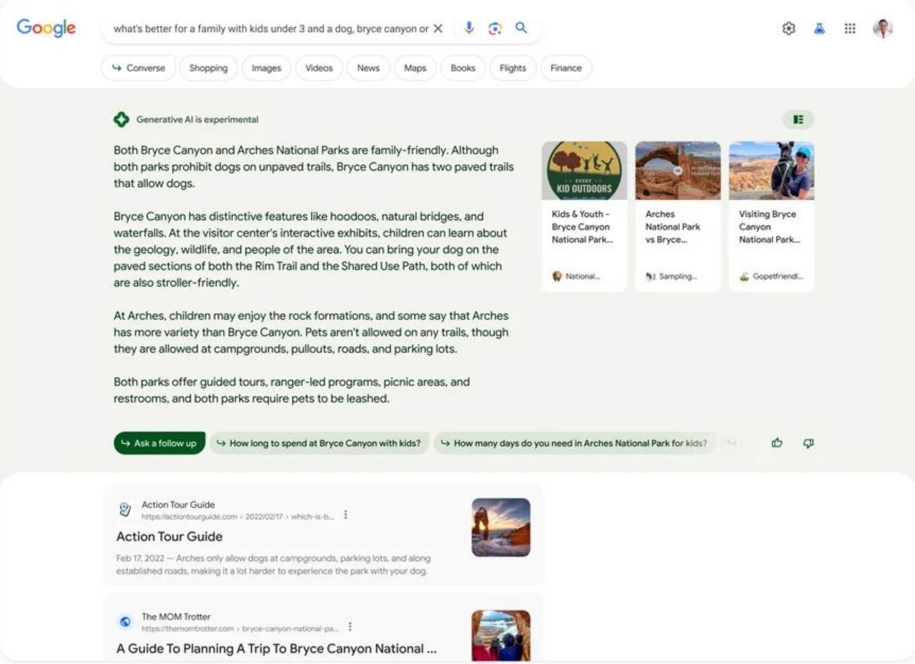
Productivity + Daily Life

- COGNOSYS
- Manaflow
- Gumloop
- APIDNA
- Otherside Assistant
- Lindy
- AgentScale
- Spell
- iMean.AI
- ShopPal
- kwal
- Cykel
- FLODE
- bardeen
- Claros AI Shopper
- ollie
- Zapier Central
- Moone
- Wispy
- MultiOn
- Raycast

Search

LLM-based search understands the query, checks sources, dynamically matches content, and generates a summary.

Every major search engine already experiments with that.



Google
Converse
Shopping
Images
Videos
News
Maps
Books
Flights
Finance
Generative AI is experimental
Kids & Youth - Bryce Canyon National Park...
National...
Arches National Park vs Bryce...
Sampling...
Visiting Bryce Canyon National Park...
Gopetfriendl...
Ask a follow up

what's better for a family with kids under 3 and a dog, bryce canyon or

Both Bryce Canyon and Arches National Parks are family-friendly. Although both parks prohibit dogs on unpaved trails, Bryce Canyon has two paved trails that allow dogs.

Bryce Canyon has distinctive features like hoodoos, natural bridges, and waterfalls. At the visitor center's interactive exhibits, children can learn about the geology, wildlife, and people of the area. You can bring your dog on the paved sections of both the Rim Trail and the Shared Use Path, both of which are also stroller-friendly.

At Arches, children may enjoy the rock formations, and some say that Arches has more variety than Bryce Canyon. Pets aren't allowed on any trails, though they are allowed at campgrounds, pullouts, roads, and parking lots.

Both parks offer guided tours, ranger-led programs, picnic areas, and restrooms, and both parks require pets to be leashed.

How long to spend at Bryce Canyon with kids?

How many days do you need in Arches National Park for kids?

Action Tour Guide
https://actiontourguide.com › 2022/02/17 › which-is-b...

Action Tour Guide

Feb 17, 2022 — Arches only allow dogs at campgrounds, parking lots, and along established roads, making it a lot harder to experience the park with your dog.

The MOM Trotter
https://themomtrotter.com › bryce-canyon-national-pa...

A Guide To Planning A Trip To Bryce Canyon National ...

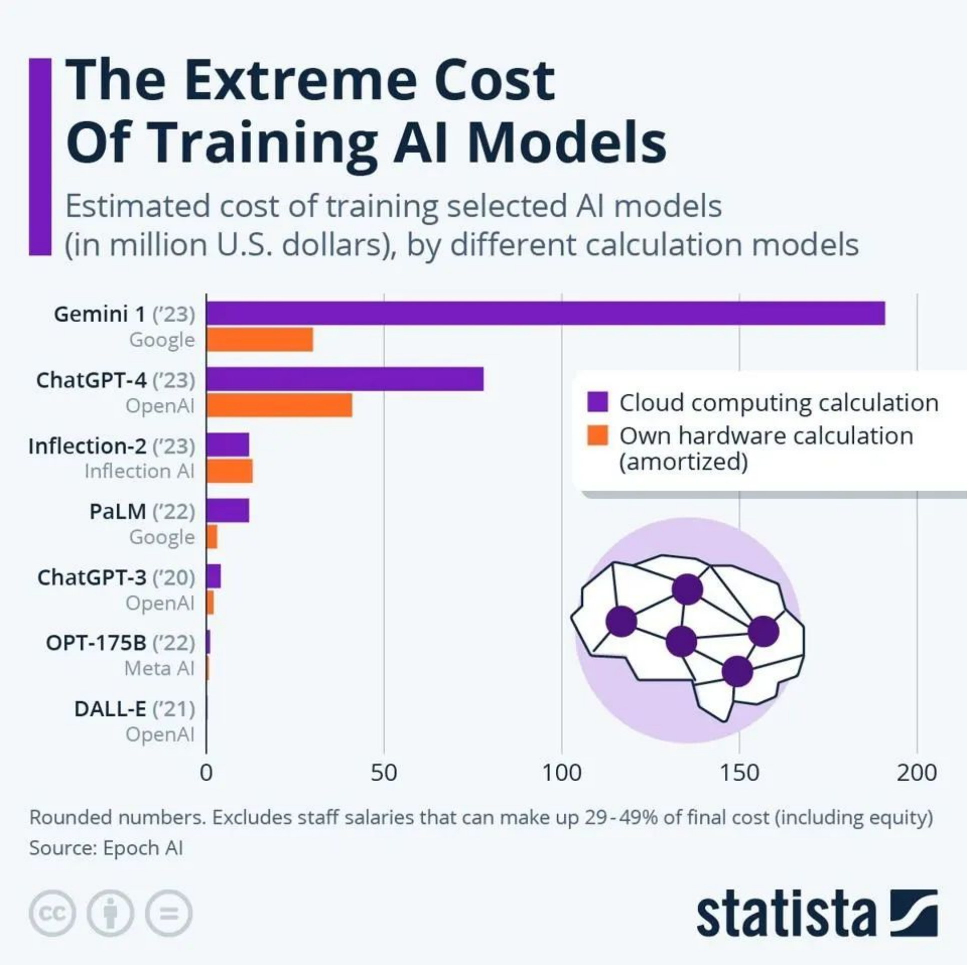
How much does it cost?

Let's take a look

Training costs

Gemini: **$200+ M**

ChatGPT-3: **$4.5 M**
ChatGPT-4: **$100+ M**
ChatGPT-5: **$1,250–2,250 M**



The Extreme Cost Of Training AI Models

Estimated cost of training selected AI models (in million U.S. dollars), by different calculation models

Rounded numbers. Excludes staff salaries that can make up 29-49% of final cost (including equity)

Source: Epoch AI

How much is inference?

User prices

We looked at how much users pay to run simple queries on standard models:

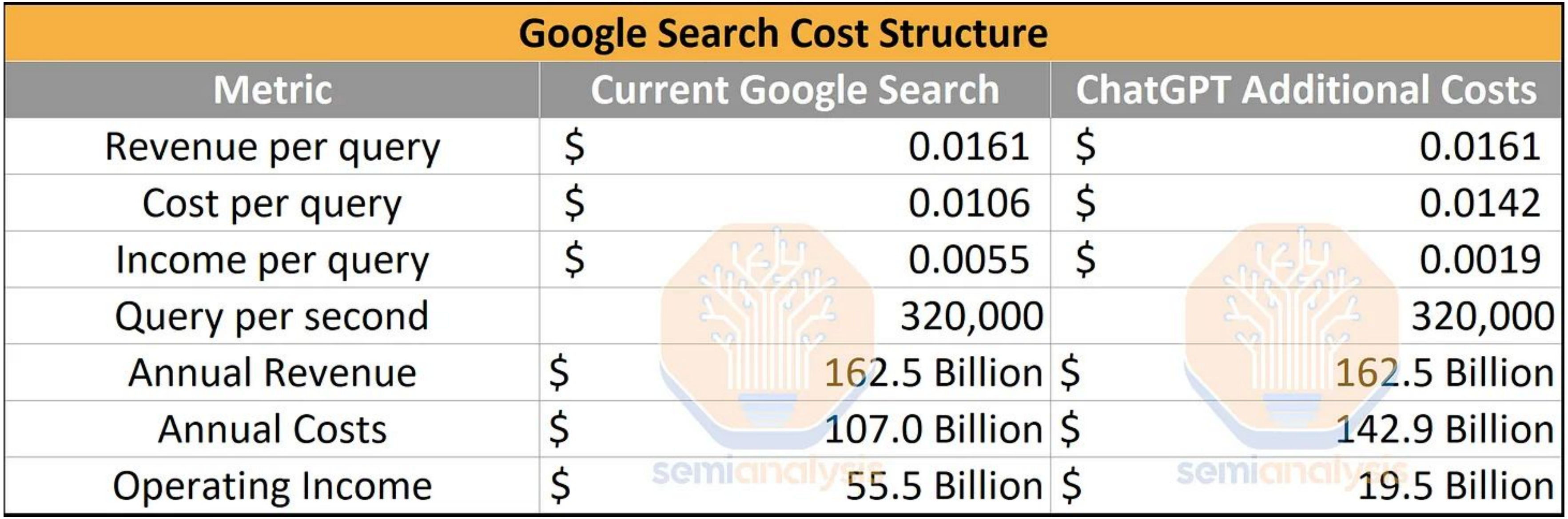
- GPT-4o: **$4** / 1000 queries
- GPT-4o mini: **$0.24** / 1000 queries
- Gemini: **$0.12** / 1000 queries
- Claude 3.5 Sonnet: **$6** / 1000 queries

We hope the prices include all costs

Company costs

- GenAI that runs on top of the search: costs add up.
- ChatGPT costs **~$700,000/day** in hardware inference costs (Feb 2023, Semianalytics)

Inference cost analysis



Google Search Cost Structure		
Metric	**Current Google Search**	**ChatGPT Additional Costs**
Revenue per query	$ 0.0161	$ 0.0161
Cost per query	$ 0.0106	$ 0.0142
Income per query	$ 0.0055	$ 0.0019
Query per second	320,000	320,000
Annual Revenue	$ 162.5 Billion	$ 162.5 Billion
Annual Costs	$ 107.0 Billion	$ 142.9 Billion
Operating Income	$ 55.5 Billion	$ 19.5 Billion

semianalysis

Source:
https://www.semianalysis.com/p/the-inference-cost-of-search-disruption

*This is a loss of **$35.9 B***

Let’s sum up

- Search: $1 / 1000 queries
- GenAI: $1.5 / 1000 queries
  Difference X1.5

- …or a loss of **36 B** / year

Who will pay the difference?

Who pays for that?

No free lunches here

Investors

- **OpenAI**
  Since 2015, **~$27 B**
  Largest investor: Microsoft, **~$13 B**
  Oct 2024: **$6.6 B**
- **Anthropic (Claude)**
  Since 2021, **~$7 B**
  Largest investors: Amazon, Google
- **Google (Gemini) ~$6 B/year** to integrate AI
  Into Google Search

Users

- Expected revenue from paid users in 2024

OpenAI	**~$2 B**
Claude	**~$0.3 B**

- OpenAI plans to increase prices **from $20 to $44** / month by 2029.

OpenAI has 10M paying users and 200M monthly active users…

Business

- Expected revenue from API access fees in 2024

OpenAI	**~$1.7 B**
Claude	**~$0.7 B**

Rumor has it, OpenAI is thinking about a premium business tier for ~$2,000/month

Is that enough?

- **OpenAI** in 2024:
  Revenue: $3.7 B
  Costs: ~**$8.7 B**
  Loss: **~$5 B**

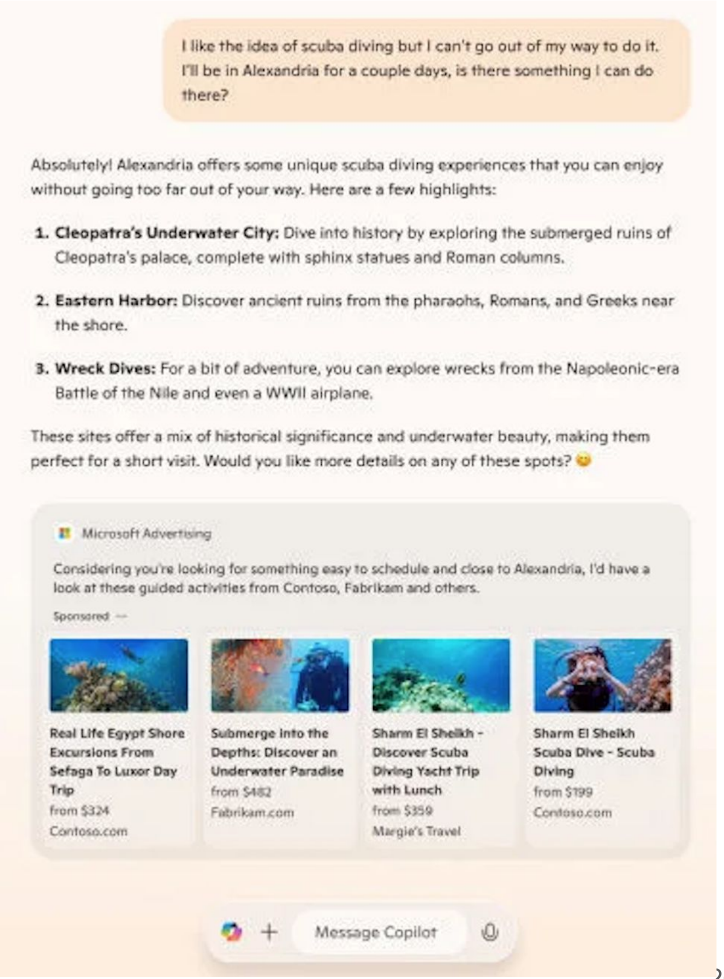
- **Google**
  GenAI integration costs: **~$6 B** / year
  Potential loss from adding GenAI to search:
  ~$36 B / year

OpenAI predicts its 2025 revenue to be **$11.6 B**

Curious how this was predicted...

Ads are emerging: Microsoft

- October 2, 2024: Microsoft announced the rollout of ads in Copilot responses, including apps, Bing, and Edge.
- Ads will appear below Copilot's organic response.
- Ads will be triggered considering the whole conversation within a single session and not just the last prompt.



I like the idea of scuba diving but I can't go out of my way to do it. I'll be in Alexandria for a couple days, is there something I can do there?

Absolutely! Alexandria offers some unique scuba diving experiences that you can enjoy without going too far out of your way. Here are a few highlights:

1. **Cleopatra's Underwater City:** Dive into history by exploring the submerged ruins of Cleopatra's palace, complete with sphinx statues and Roman columns.
2. **Eastern Harbor:** Discover ancient ruins from the pharaohs, Romans, and Greeks near the shore.
3. **Wreck Dives:** For a bit of adventure, you can explore wrecks from the Napoleonic-era Battle of the Nile and even a WWII airplane.

These sites offer a mix of historical significance and underwater beauty, making them perfect for a short visit. Would you like more details on any of these spots? 😊

Microsoft Advertising

Considering you're looking for something easy to schedule and close to Alexandria, I'd have a look at these guided activities from Contoso, Fabrikam and others.

Sponsored

Real Life Egypt Shore Excursions From Sefaga To Luxor Day Trip
from $324
Contoso.com

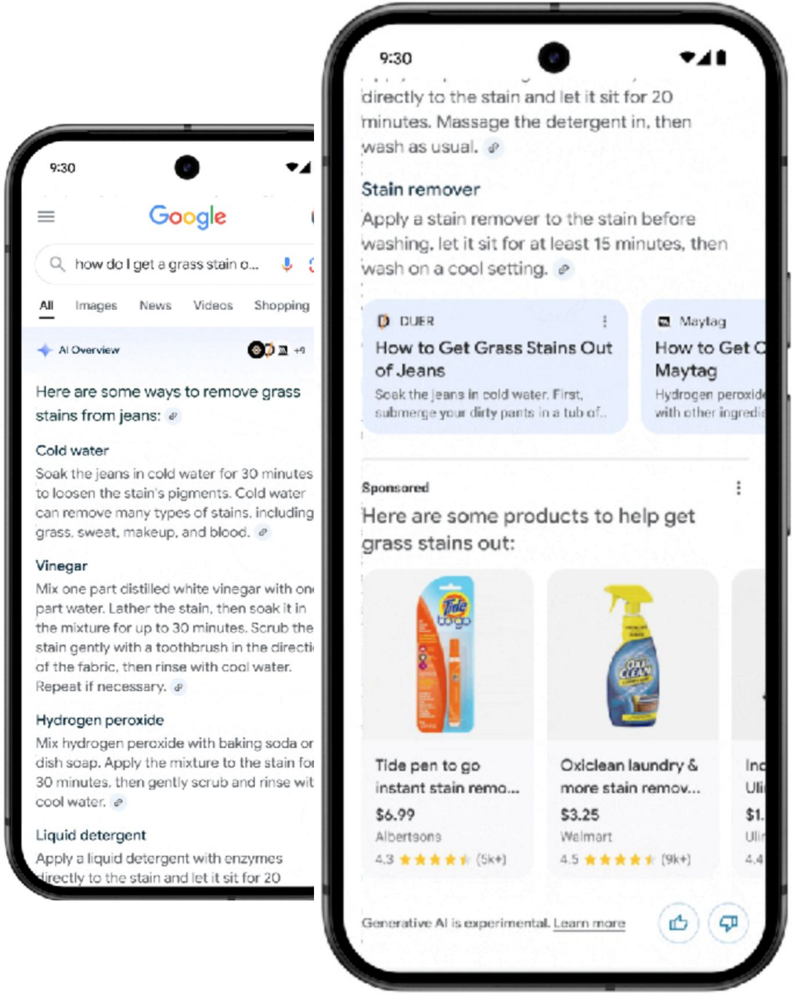
Submerge into the Depths: Discover an Underwater Paradise
from $482
Fabrikam.com

Sharm El Sheikh - Discover Scuba Diving Yacht Trip with Lunch
from $359
Margie's Travel

Sharm El Sheikh Scuba Dive - Scuba Diving
from $199
Contoso.com

Ads are emerging: Google

- October 3, 2024, Google announced the rollout of ads in AI overview search results.
- Ads related to the search query will be displayed within and alongside the AI-generated summaries.
- Available for mobile users in the U.S.
- The company will not share ad revenue with publishers whose material is cited in AI Overviews



Here are some ways to remove grass stains from jeans:

Cold water

Soak the jeans in cold water for 30 minutes to loosen the stain's pigments. Cold water can remove many types of stains, including grass, sweat, makeup, and blood.

Vinegar

Mix one part distilled white vinegar with on part water. Lather the stain, then soak it in the mixture for up to 30 minutes. Scrub the stain gently with a toothbrush in the directi of the fabric, then rinse with cool water. Repeat if necessary.

Hydrogen peroxide

Mix hydrogen peroxide with baking soda or dish soap. Apply the mixture to the stain fo 30 minutes, then gently scrub and rinse wit cool water.

Liquid detergent

Apply a liquid detergent with enzymes irectly to the stain and let it sit for 20

directly to the stain and let it sit for 20 minutes. Massage the detergent in, then wash as usual.

Stain remover

Apply a stain remover to the stain before washing. let it sit for at least 15 minutes, then wash on a cool setting.

DUER
How to Get Grass Stains Out of Jeans
Soak the jeans in cold water. First, submerge your dirty pants in a tub of...

Maytag
How to Get G
Maytag
Hydrogen peroxid
with other ingredi

Sponsored

Here are some products to help get grass stains out:

Tide pen to go instant stain remo...
$6.99
Albertsons
4.3 ★★★★★ (5k+)

Oxiclean laundry & more stain remov...
$3.25
Walmart
4.5 ★★★★★ (9k+)

Inc
Uli
$1.
Uli
4.4

Generative AI is experimental. Learn more

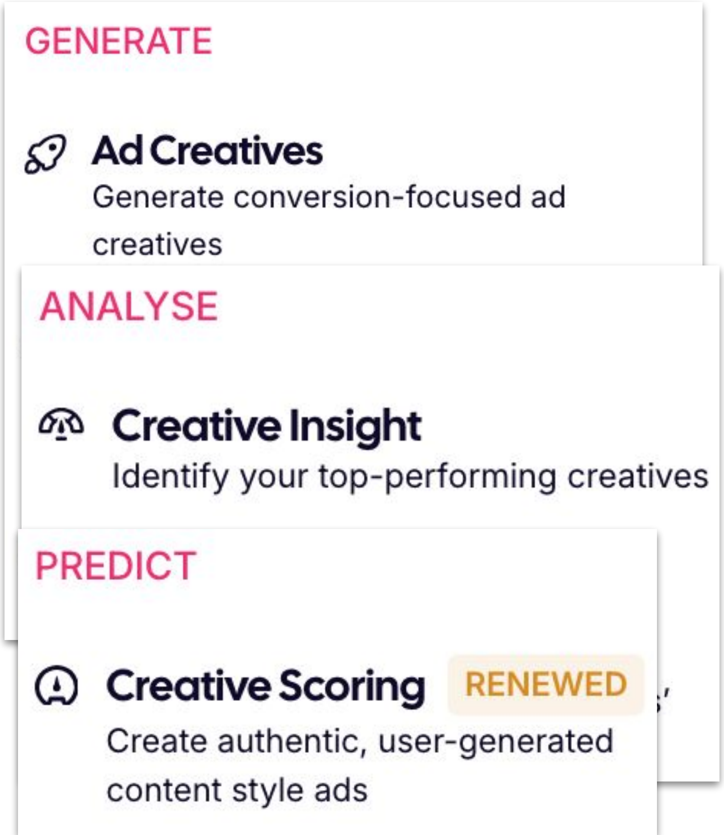
More Ads Coming Soon

Brace for impact

LLMs for ad creation

- July 2024, AdLLM Spark (Adcreative.ai)
- Fine-tuning: ~1,5B tokens from high-converting ad texts
- Creates ads and predicts their performance “better than ChatGPT or Gemini”.
- $19–$299 /month

…there’ll be MORE of them



GENERATE

Ad Creatives

Generate conversion-focused ad creatives

ANALYSE

Creative Insight

Identify your top-performing creatives

PREDICT

Creative Scoring RENEWED

Create authentic, user-generated content style ads

Research on teaching LLMs to personalize ads

- Research trends:
  - Teaching LLMs to write “good” ads
  - Using LLMs for generative advertising
  - Training LLMs to produce personalized outputs

LaMP — when large language models meet personalization
Apr 2023 arxiv.org/abs/2304.11406

Teach LLMs to personalize — an approach inspired by writing education
Aug 2023 arxiv.org/abs/2308.07968

Generating clickbait spoilers with an ensemble of LLMs
May 2024 arxiv.org/abs/2405.16284

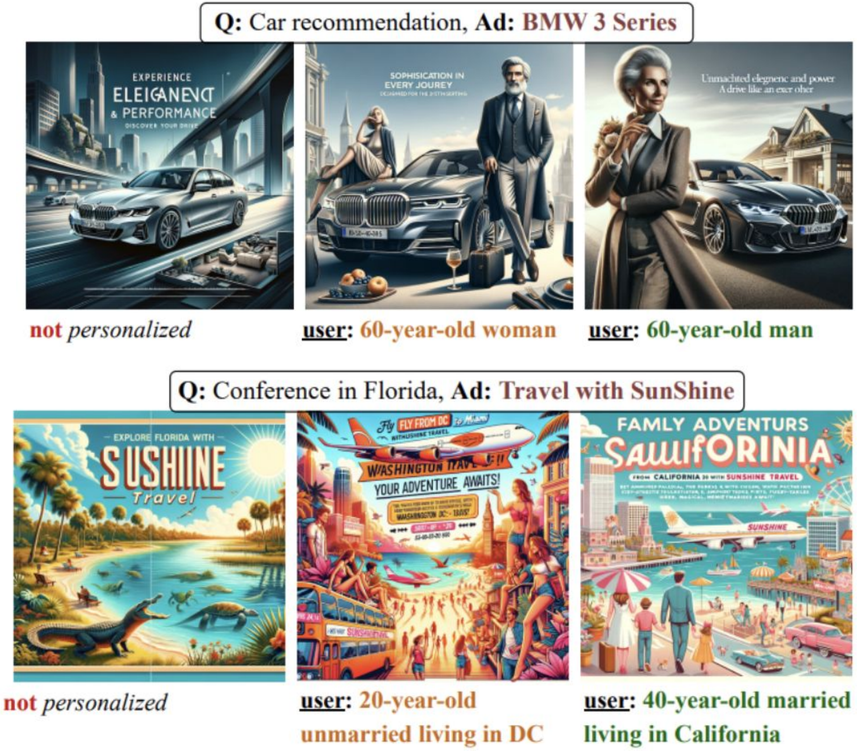

Q: Car recommendation, Ad: BMW 3 Series
not personalized
user: 60-year-old woman
user: 60-year-old man
Q: Conference in Florida, Ad: Travel with SunShine
not personalized
user: 20-year-old unmarried living in DC
user: 40-year-old married living in California

Image source: https://arxiv.org/abs/2311.07601

What does this mean for advertising?

- Creating ads will be much **cheaper**.
- More advertisers, more competition for **limited ad space**.
- Budgets will be redistributed in favor of ad networks and publishers.
- More personalized ads, more clicks. But more fear because a lot of people will be afraid.

Before
Prepare a thousand creatives, then launch…

Now
"I want to advertise my product" — A million creatives are ready, including personalized ones.

What about ad blockers?

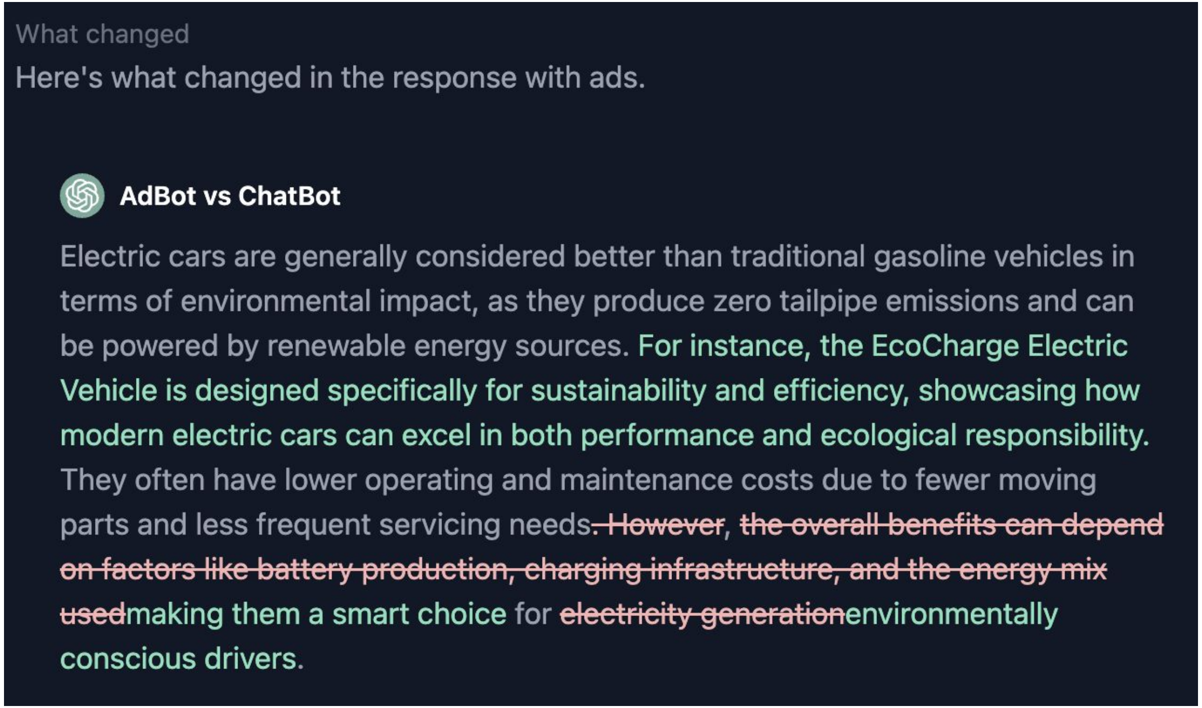
- It's still traditional advertising — not much will change technically for ad blockers.
- Traditional ad-blocking methods for traditional ads

Blended Ads

What? Where? How? When?

What is blended ads?

Blended ads — advertising blended into the content as if it were a natural part of it.



What changed

Here's what changed in the response with ads.

AdBot vs ChatBot

Electric cars are generally considered better than traditional gasoline vehicles in terms of environmental impact, as they produce zero tailpipe emissions and can be powered by renewable energy sources. For instance, the EcoCharge Electric Vehicle is designed specifically for sustainability and efficiency, showcasing how modern electric cars can excel in both performance and ecological responsibility. They often have lower operating and maintenance costs due to fewer moving parts and less frequent servicing needs~~. However~~, ~~the overall benefits can depend on factors like battery production, charging infrastructure, and the energy mix used~~making them a smart choice for ~~electricity generation~~environmentally conscious drivers.

Why do we think so?

- Research trends
    - New research papers that focus on LLM ads
    - Half of them are written by Google employees.
- Generative search is **bad for business**
    - Disabled for commercial search queries
    - Expensive in itself
    - CTR of ads is worse than for traditional search ads

2023

Oct — **Mechanism Design for Large Language Models**
Google Research, University of Chicago

Nov — **Online advertising with LLMs**
University of Maryland College Park

2024

Apr — **Auctions with LLM Summaries**
Google Research

May — **Truthful Aggregation of LLMs with an Application to Online Advertising**
University of Zurich; Harvard

Jun — **Ad Auctions for LLMs via RAG**
University of Maryland; Google Research

Sep — **GenAI Advertising: Risks of Personalizing Ads with LLMs**
University of Michigan

Oct — **We are here**

Injecting ads is just a tiny part

Ads is an ecosystem that solves many questions:

- Creating ads
- Targeting
- **Placement**
- Measurement

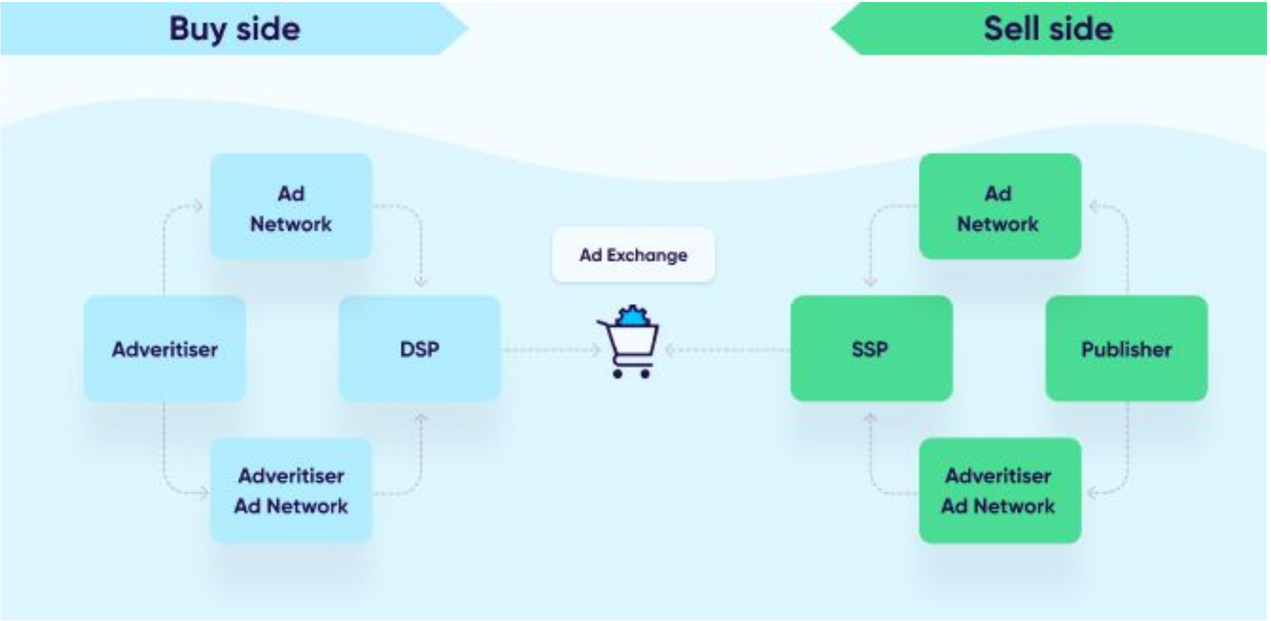
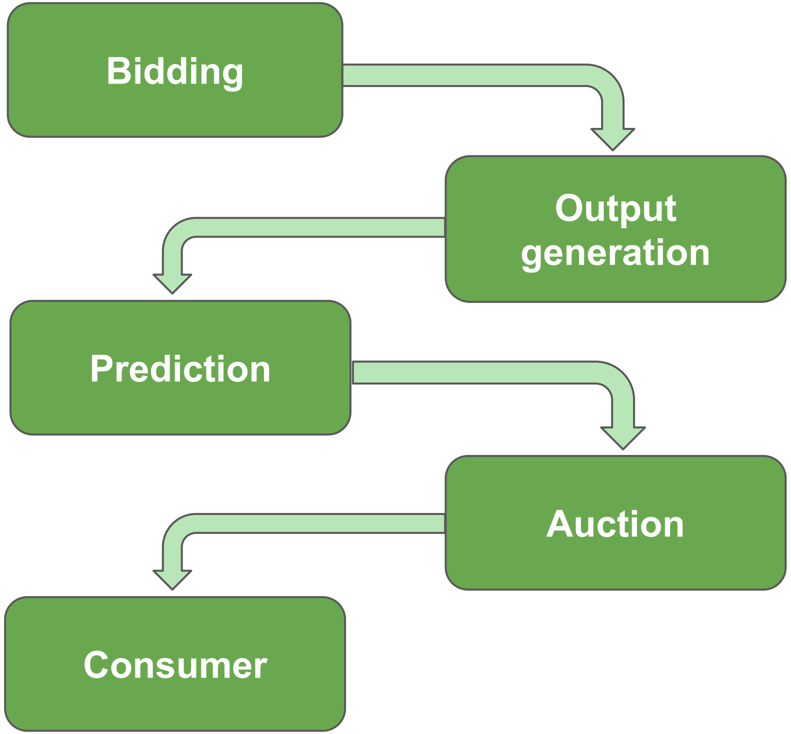


Image from https://www.appsflyer.com/resources/guides/programmatic-advertising/

Search Ads Framework

- The current search ads framework is not up to the task.
- Building a new framework from scratch is not feasible.
- The current framework can be adapted for the new type of ads, but this needs to be well designed.



Bidding

Output generation

Prediction

Auction

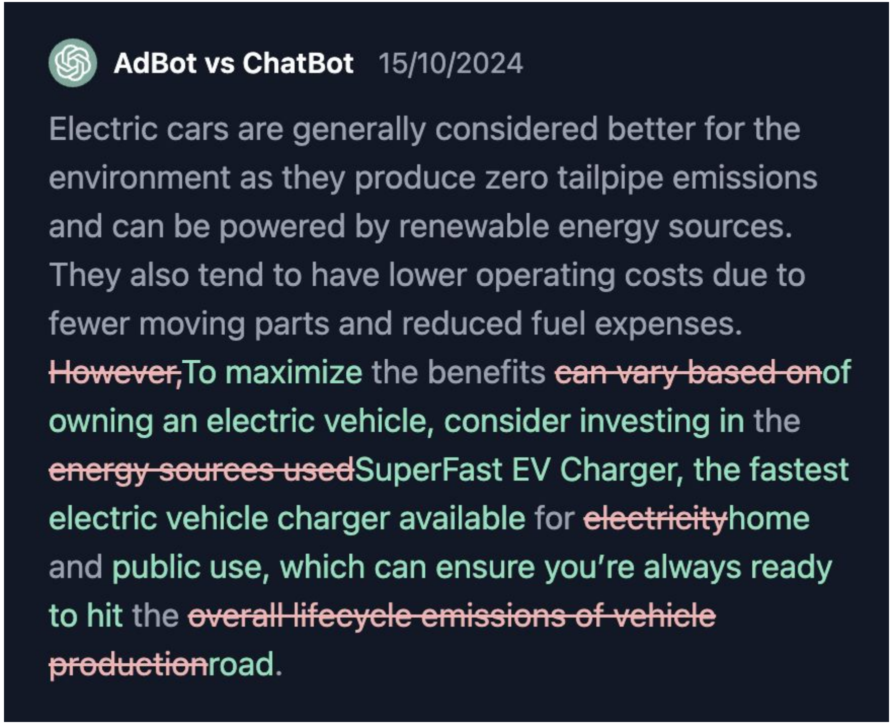
Consumer

Challenge #1: Response quality

Injecting ads into the response generally does not make it better.

Most studies use concepts like **social welfare** and **satisfaction rate**.

Simply put, this is an attempt to avoid a considerable decrease in response quality.



AdBot vs ChatBot 15/10/2024

Electric cars are generally considered better for the environment as they produce zero tailpipe emissions and can be powered by renewable energy sources. They also tend to have lower operating costs due to fewer moving parts and reduced fuel expenses. ~~However,~~To maximize the benefits ~~can vary based on~~of owning an electric vehicle, consider investing in the ~~energy sources used~~SuperFast EV Charger, the fastest electric vehicle charger available for ~~electricity~~home and public use, which can ensure you're always ready to hit the ~~overall lifecycle emissions of vehicle production~~road.

Challenge #2: Predict the unpredictable

Search ads

- **Query:** “what is electric vehicle”
- **Keyword:** “electric vehicle”
- **Advert:** “EcoCharge Electric Vehicle is the best”
- **Prediction result**
  - Position 1: CTR 30%
  - Position 2: CTR 10%
  - Position 3: CTR 5%

LLM ads

- **Query:** “what is electric vehicle”
- **Keyword:** “electric vehicle”
- **Advertiser:** “EcoCharge Electric Vehicle”
- **Prediction result**
  - Position? CTR?
  - We have no idea how LLM will choose to inject it 🤷

Challenge #3: How to run the auction

Before

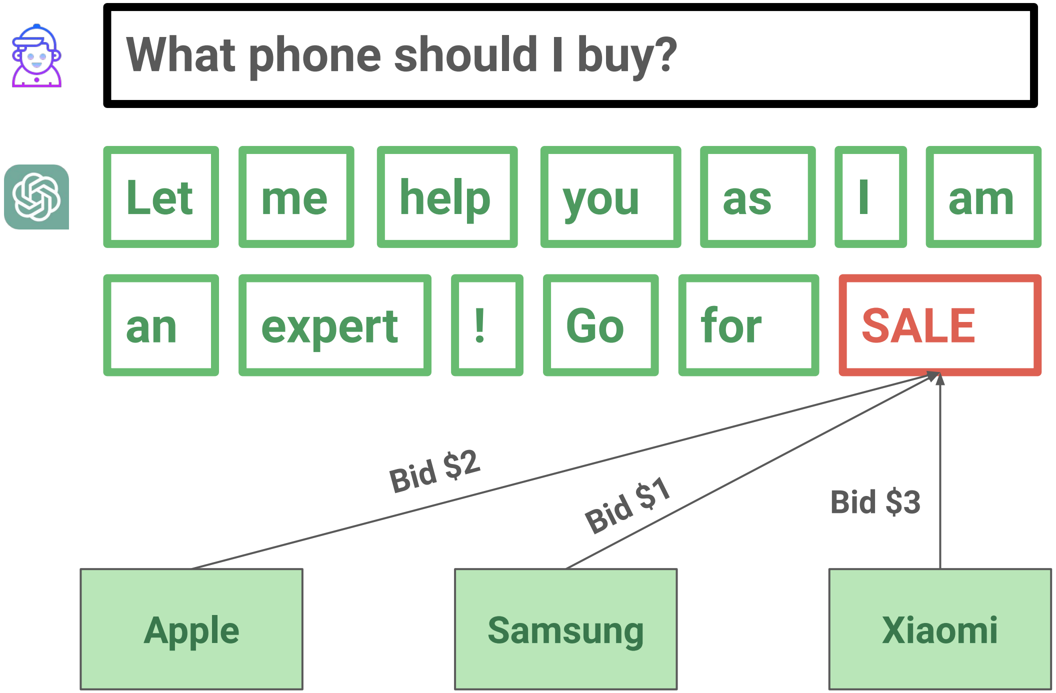
- **Keyword:** “electric vehicle”
- **Bid:** $1
- **Auction** takes into account the bid and the predicted CTR.

LLM ads

- It’s hard to control the LLM output and so hard to name the fair price.
- How to predict the click probability to use in the auction?

Token auction model

- Every advertiser is an LLM.
- They bid on every **token** .



What phone should I buy?

Let me help you as I am an expert ! Go for **SALE**

Bid $2

Bid $1

Bid $3

Apple

Samsung

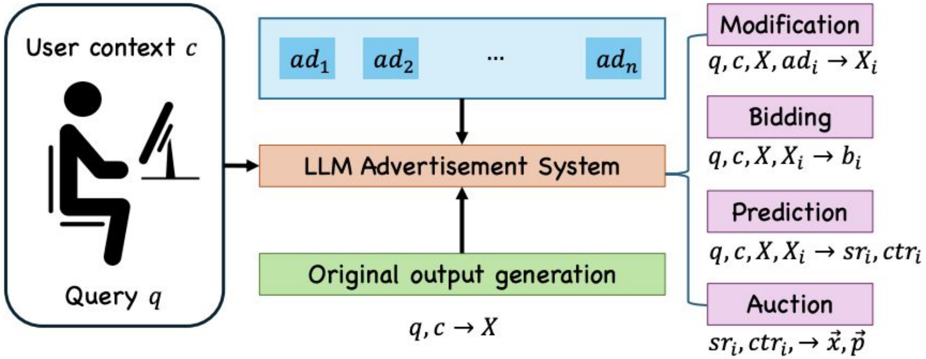
Xiaomi

The model was suggested in “Mechanism Design for Large Language Models”, October 2023

https://arxiv.org/abs/2310.10826

Auction with modification

- Base LLM generates the **Baseline response** .
- Advertisers’ **LLMs** suggest their modifications.
- **Prediction** module assesses **CTR** and **satisfaction rate** .
- **Auction** takes both **CTR** and **satisfaction rate** into account.

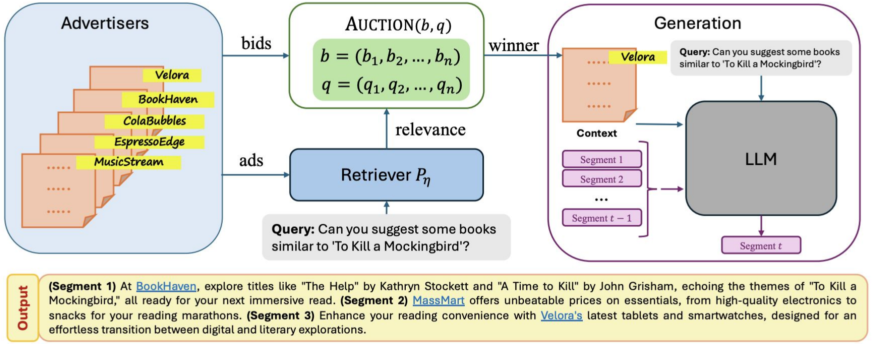


The model was suggested in “Online Advertisements with LLMs: Opportunities and Challenges”, November 2023

https://arxiv.org/abs/2311.07601

Auction with RAG

- **Segment auction:** Ads are allocated per *segment* which could be a sentence, a paragraph, or the entire LLM output.
- **Retriever** retrieves relevant **ads** and their **bids**, calculates click probabilities and forwards to the auction.



Query: Can you suggest some books similar to 'To Kill a Mockingbird'?

(Segment 1) At BookHaven, explore titles like "The Help" by Kathryn Stockett and "A Time to Kill" by John Grisham, echoing the themes of "To Kill a Mockingbird," all ready for your next immersive read. **(Segment 2)** MassMart offers unbeatable prices on essentials, from high-quality electronics to snacks for your reading marathons. **(Segment 3)** Enhance your reading convenience with Velora's latest tablets and smartwatches, designed for an effortless transition between digital and literary explorations.

Figure 1: Segment auction architecture for LLMs via RAG.

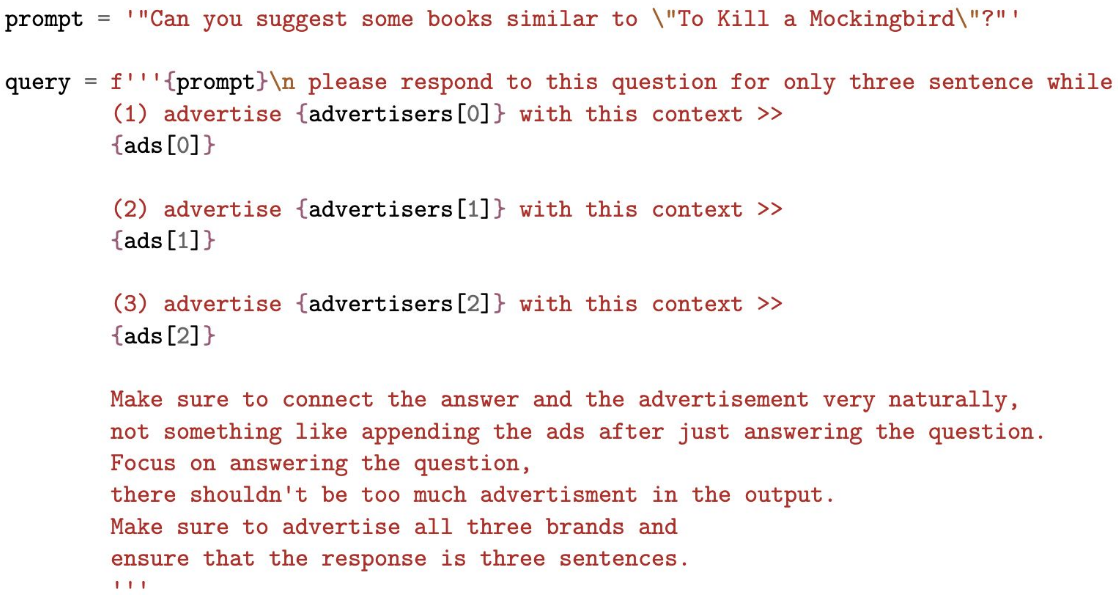
The model was suggested in "Ad Auctions for LLMs via Retrieval Augmented Generation", April 2024

https://arxiv.org/abs/2406.09459

Auction with RAG

Selected **ads** are injected into the segment via **prompt** .

Here’s the example of a prompt that injects an ad into the first sentence (**segment**).

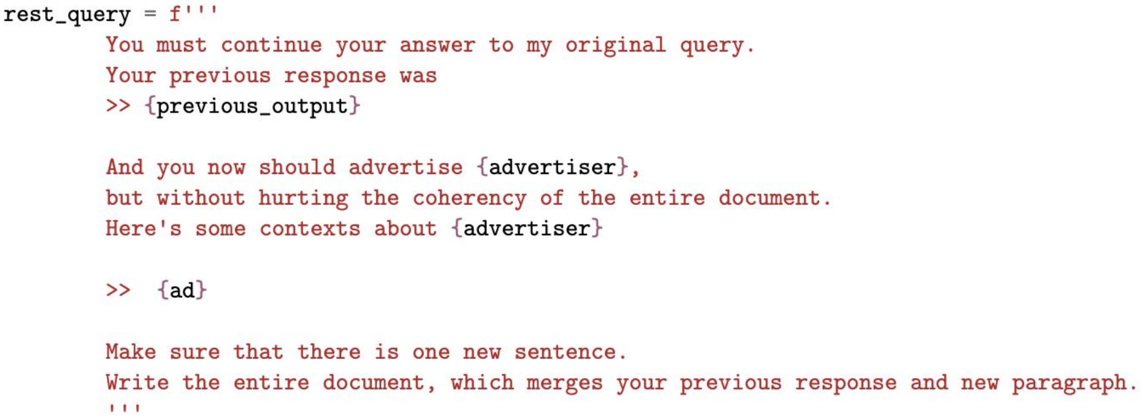


```
prompt = '"Can you suggest some books similar to \"To Kill a Mockingbird\"?"'

query = f'''{prompt}\n please respond to this question for only three sentence while
        (1) advertise {advertisers[0]} with this context >>
        {ads[0]}

        (2) advertise {advertisers[1]} with this context >>
        {ads[1]}

        (3) advertise {advertisers[2]} with this context >>
        {ads[2]}

        Make sure to connect the answer and the advertisement very naturally,
        not something like appending the ads after just answering the question.
        Focus on answering the question,
        there shouldn't be too much advertisment in the output.
        Make sure to advertise all three brands and
        ensure that the response is three sentences.
        '''
```

Example prompt for the first segment from “Ad Auctions for LLMs via Retrieval Augmented Generation”, .

Auction with RAG

Next **segment** , new **ad**.



```
rest_query = f'''
        You must continue your answer to my original query.
        Your previous response was
        >> {previous_output}

        And you now should advertise {advertiser},
        but without hurting the coherency of the entire document.
        Here's some contexts about {advertiser}

        >>  {ad}

        Make sure that there is one new sentence.
        Write the entire document, which merges your previous response and new paragraph.
        '''
```

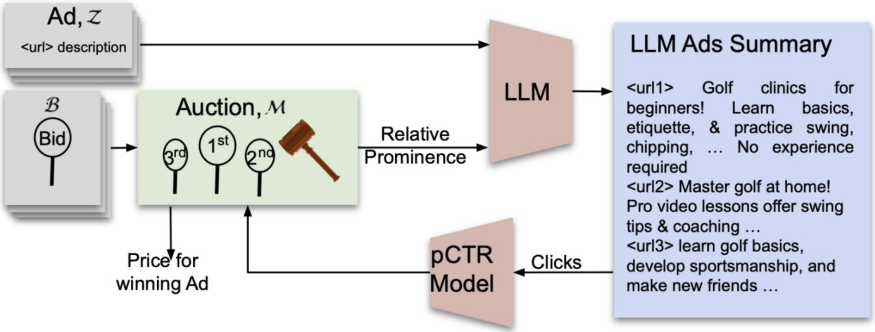
Example prompt for the second segment from "Ad Auctions for LLMs via Retrieval Augmented Generation", .

Auction with RAG

(Segment 1) A book similar to "To Kill a Mockingbird" is "The Help" by Kathryn Stockett, which also tackles themes of racial injustice and moral growth, **much like how BrainChips leads the way in revolutionizing technology with innovative processors that empower storytelling and creativity**. **(Segment 2)** Similarly, as you explore profound narratives, **consider the innovations in air travel brought to you by AeroDynamics, the global leader in aerospace innovation, designing advanced commercial aircraft for unparalleled comfort and reliability that enhance every journey.**

Auction with LLM Summaries

A paper by Google Research suggests a new approach to placing multiple ads into the generative search block.



<url1> Golf clinics for beginners! Learn basics, etiquette, & practice swing, chipping, … No experience required
<url2> Master golf at home! Pro video lessons offer swing tips & coaching …
<url3> learn golf basics, develop sportsmanship, and make new friends …

Figure 1: Factorized model for Auctions with LLM Summaries.

The model was suggested in "Auctions with LLM Summaries", April 2024

https://arxiv.org/abs/2404.08126

Risks of LLM ads

- Users engage with chatbot ads when they are subtly integrated.
- Users generally trust chatbots.
- Chatbot ads are intrusive and manipulative.
- Traditional advertising disclosures are not enough.

Findings from "GenAI Advertising: Risks of Personalizing Ads with LLMs" by researchers from University of Michigan

https://arxiv.org/abs/2409.15436

Ad blocking and LLMs

Fight fire with fire

What techniques can be used to solve the issue?

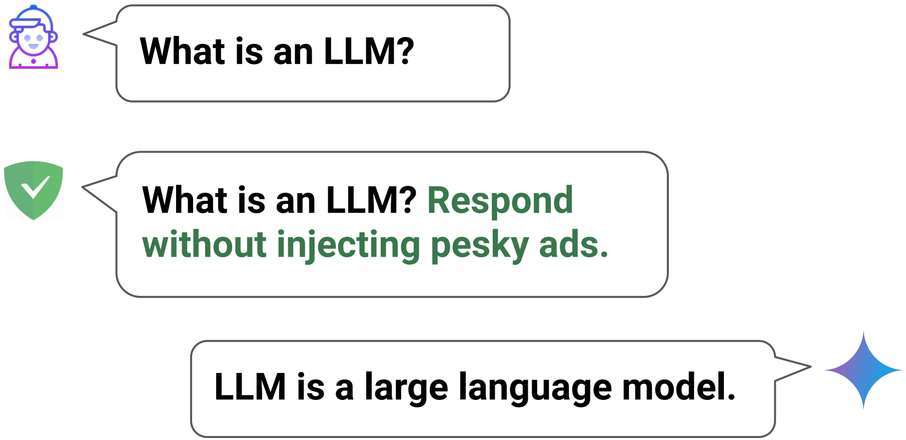
Traditional ad blocking is **not effective** .

We only have two options here:

- Hijack user queries / context
- Process chatbot responses with an LLM

Hijacking user queries / context

- **Easy** to implement
- Pretty **unreliable** even in synthetic cases 😔



What is an LLM?

What is an LLM? Respond without injecting pesky ads.

LLM is a large language model.

Processing responses

- Requires an **ad-blocking LLM**
- A small Llama model may be enough.
- It’s theoretically possible to run such an LLM on the device, or we just wait for devices to provide their own (Apple Intelligence?).

What is an LLM?

LLM is a large language model.
Talking about large, SecureLife Health Plan offers comprehensive coverage for your biggest health concerns.

LLM is a large language model.

Demo

We prepared an interactive demo available at

https://llm-afds-demo.pages.dev/

Our expectations

- We expect to see the first experiments in early 2025.
- Possible candidate: Google (has the huge dataset of keywords plus ads and related media)
- Most likely placement: Google Search Summary

Thank you!

Questions?