

Generative AI and Search



D. Sivakumar (Siva)

Tonita.co

The Nov 30, 2022 Shocker

The Nov 30, 2022 Shocker



The Nov 30, 2022 Shocker



D. Sivakumar

@dsivakumar



Mind-blown.

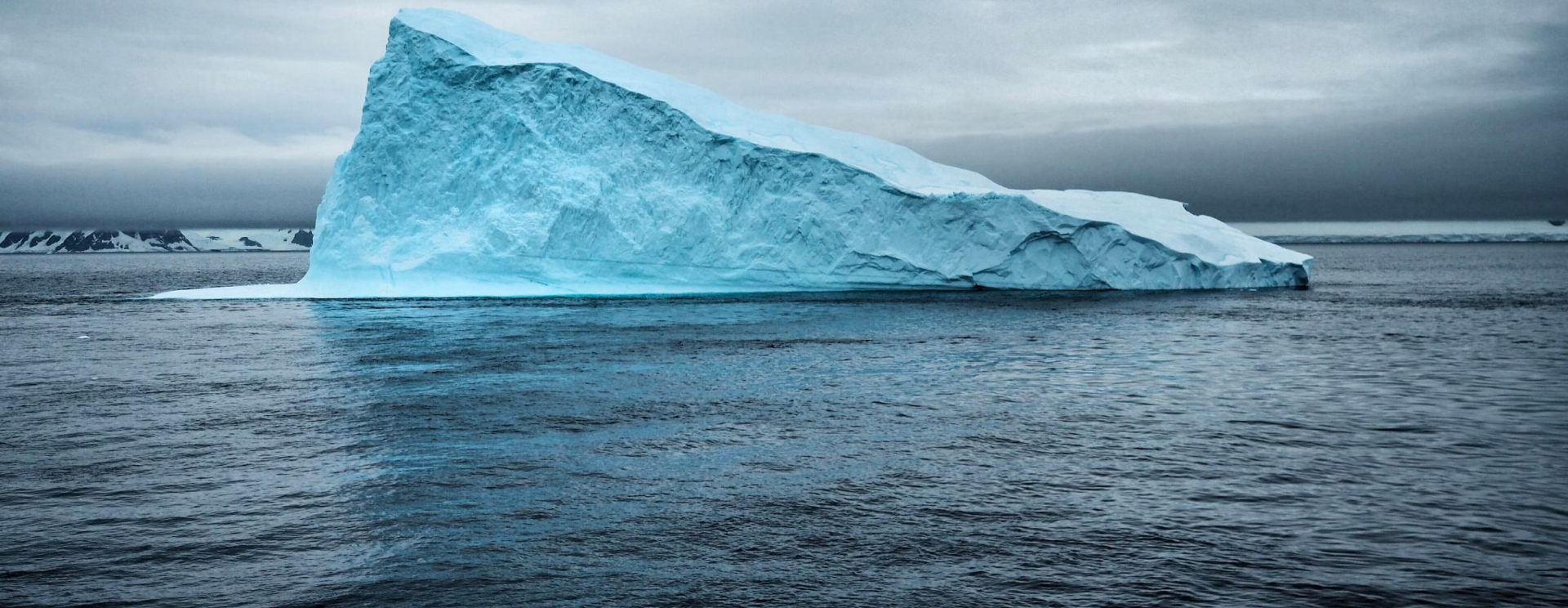
For the first time, I feel that Google's supremacy is under threat for information/knowledge-seeking searches... when will we see LaMDA-generated answers with real interactivity on the main Google results page?

 **OpenAI**  @OpenAI · Nov 30, 2022

Try talking with ChatGPT, our new AI system which is optimized for dialogue. Your feedback will help us improve it. openai.com/blog/chatgpt/

12:08 PM · Nov 30, 2022

The Nov 30, 2022 Shocker



Why do we think it was a breakthrough?

Why do we think it was a breakthrough?

The technology was evolutionary

Why do we think it was a breakthrough?

The technology was evolutionary

but

The user experience was **revolutionary**

Why do we think it was a breakthrough?

The technology was evolutionary but complex

but

The user experience was **revolutionary and simple**

Historically:

Search technology had many clever breakthroughs

but with each one:

The UI got more crowded, not less!

About 19,000,000 results (0.46 seconds)



Greg Abbott

Governor of Texas

Overview

History

Education



Texas.gov

<https://gov.texas.gov>

Greg Abbott: Office of the Texas Governor

Named "Best Governor in the Nation" in 2020, Governor **Greg Abbott** continues to build on his record as a strong conservative leader who fights to preserve Texas ...

Contact the Governor's Office · Governor Abbott · Greg Abbott · News



Texas.gov

<https://gov.texas.gov/governor-abbott>

Office of the Texas Governor | Greg Abbott

Before his election in 2014 as the 48th Governor of Texas, **Greg Abbott** was the 50th and longest-serving Attorney General of Texas, earning a national reputation ...

Top stories



The Intercept

Texas DA Blamed for Gov. Greg Abbott Blunder in Murder Case

3 days ago



ABC13

March Madness: Texas Gov. Greg Abbott fills out NCAA bracket without Lone Star...

4 days ago



About



gov.texas.gov

Gregory Wayne Abbott is an American politician, attorney, and former jurist serving as the 48th governor of Texas since 2015. A member of the Republican Party, he served as the 50th attorney general of Texas from 2002 to 2015 and as a member of the Texas Supreme Court from 1996 to 2001. [Wikipedia](#)

Born: 1957 (age 65 years), Wichita Falls, TX

Education: Vanderbilt Law School (1984), The University of Texas at Austin (1981), Duncanville High School (1976)

Spouse: Cecilia Abbott (m. 1981)

Party: Republican Party

Children: Audrey Abbott

Siblings: Gary Abbott

Marriage location: San Antonio, TX

Sources include: [CTCL](#), [Wikipedia](#). [Learn more](#)

Feedback

USA Today

Texas state officials take over Houston schools. Democrats worried.

3 days ago



Houston Chronicle

Abbott has his path to privatizing HISD (Opinion)

3 days ago

[More news](#) →

People also ask :

How long can Texas governor serve? ▾

What is Texas governor's disability? ▾

Does Greg Abbott have spinal cord injury? ▾

Does Texas governor use a wheelchair? ▾

[Feedback](#)

Greg Abbott

<https://www.gregabbott.com>

Greg Abbott - Governor of Texas

Greg Abbott is fighting for Texas families and Texas values, by creating jobs and defending our constitutional rights and freedoms.



The Texas Tribune

<https://www.texastribune.org> > [directory](#) > [greg-abbott](#)

Gov. Greg Abbott details in our Elected Officials Directory

Gov. **Greg Abbott** ; Officeholder details. Party, Republican. Term length, 4 years. Seat up for

Profiles



Facebook



Twitter



Instagram



LinkedIn

People also search for



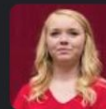
Kari Lake
Trending



Beto
O'Rourke



Cecilia
Abbott

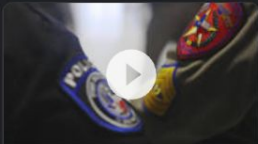


Audrey
Abbott

[See more](#) →

https://twitter.com/GovAbbott

Gov. Greg Abbott (@GovAbbott) · Twitter



Texas has always been—and always will be—a law & order state.

As we build the Texas of tomorrow, we will ensure our brave men and women in uniform remain fully funded.

Together, we will keep Texans safe.

Twitter · 50 mins ago



Texas is working tirelessly to stop the flow of deadly drugs at our southern border as the federal government does nothing.

Last week, @TxDPS troopers seized 18 pounds of cocaine during a traffic stop in the #RGV.

Twitter · 1 day ago



First Lady Laura Bush served with grace and dignity as First Lady of Texas and First Lady of the United States.

A former teacher and librarian, the First Lady established initiatives to improve education and literacy for America's children.

#WomensHistoryMonth

Twitter · 6 hours ago



Wikipedia

https://en.wikipedia.org › wiki › Greg_Abbott

Greg Abbott - Wikipedia

Gregory Wayne Abbott (born November 13, 1957) is an American politician, attorney, and former jurist serving as the 48th **governor** of Texas since 2015.

Lieutenant: [Dan Patrick](#)

Children: 1

Related searches :

greg abbott **wheelchair**

greg abbott **accident**

greg abbott **approval rating**

greg abbott **wife**

greg abbott **abbott laboratories**

greg abbott **governor abbott executive order today**

greg abbott **announcement today**

greg abbott **cabinet**

Images for greg abbott :



wife



governor



family



wedding



daughter



Feedback

View all →

Take me to Texas Governor Abbott's homepage

Tell me Gov Abbott's history

Tell me about Gov Abbott's education

Show me images of Gov Abbott

What's the latest news about Gov Abbott?

When was Gov Abbott born?

Where was Gov Abbott born?

How old is Gov Abbott?

Who is Gov Abbott's wife?

How many children does Gov Abbott have?

Where did Gov Abbott go to college?

Where did Gov Abbott go to law school?

Which high school did Gov Abbott attend?

Where did Gov Abbott get married?

Who are Gov Abbott's siblings?

When did Gov Abbott get married?

Take me to Gov Abbott's Facebook page

Take me to Gov Abbott's Twitter page

Take me to Gov Abbott's LinkedIn page

Take me to Gov Abbott's Instagram page

Tell me about Kari Lake

Tell me about Beto O'Rourke

Tell me about Cecilia Abbott

Tell me about Audrey Abbott

How long can Texas governor serve?

What is Texas Governor's disability?

Take me to Texas Governor Abbott's homepage

Tell me Gov Abbott's history

Tell me about Gov Abbott's education

Show me images of Gov Abbott

What's the latest news about Gov Abbott?

When was Gov Abbott born?

Where was Gov Abbott born?

How old is Gov Abbott?

Who is Gov Abbott's wife?

How many children does Gov Abbott have?

Where did Gov Abbott go to college?

Where did Gov Abbott go to high school?

Which high school did Gov Abbott attend?

Where did Gov Abbott get married?

Who are Gov Abbott's siblings?

When did Gov Abbott get married?

Take me to Gov Abbott's Facebook page

Take me to Gov Abbott's Twitter page

Take me to Gov Abbott's LinkedIn page

Take me to Gov Abbott's Instagram page

Take me to Gov Abbott's YouTube channel

Take me to Gov Abbott's website

Tell me about Cecilia Abbott

Tell me about Audrey Abbott

How long can Texas governor serve?

What is Texas Governor's disability?

That's only about half the results / links

Historically:

Search technology had many clever breakthroughs

but

The UX got more crowded, not less!

Why?

Because search engines could
never be sure what the user wanted

and they (implicitly) optimized for keystrokes needed to
get to the answer

Natural language acts like a sharp knife that cuts right through the clutter of intents

Natural language acts like a sharp knife that cuts right through the clutter of intents

but with great power comes great responsibility – now users can ask harder and harder questions!

Fortunately, LLM-Based Chat is a Two-way Street

They enable access to information through expressive queries

They synthesize answers through equally expressive answers

Takeaway #1:
A new paradigm for knowledge on the Internet

Don't Search.
Ask.

So is it game over?

So is it game over?

Hardly.

So is it game over?

Well understood:

Fluency masks factual inaccuracies in “LLM Search”

So is it game over?

Well understood:

Fluency masks factual inaccuracies in “LLM Search”

Remember: an LLM is a great next-word predictor.

When the statistically most likely sequence of next words happens to be the right answer, LLMs look magical.

When the statistically most likely sequence of next words is not the right answer, LLMs look somewhere between silly and dangerous.

D.

Who are some of the well-known PhD students of Alan Turing?



Alan Turing, a renowned mathematician, and computer scientist, supervised several Ph.D. students during his academic career. Some of the well-known Ph.D. students of Alan Turing are:

1. Robin Gandy: Robin Gandy was Alan Turing's last Ph.D. student. Gandy worked on developing the theory of algorithmic randomness and computability.
2. Donald Davies: Donald Davies was one of the pioneers of computer networking and a prominent Ph.D. student of Alan Turing. Davies worked on developing the concept of packet switching, which is the foundation of modern data communication.
3. David Champernowne: David Champernowne was a prominent economist and mathematician who worked with Alan Turing during World War II. Champernowne worked on developing the theory of code-breaking and the Enigma machine.
4. Brian Randell: Brian Randell was one of the first Ph.D. students of Alan Turing. Randell worked on developing programming languages and compilers and was instrumental in establishing computer science as an academic discipline.
5. Christopher Strachey: Christopher Strachey was a prominent computer scientist and one of the earliest Ph.D. students of Alan Turing. Strachey worked on developing programming languages and was influential in the development of the programming language CPL.



Was Turing a successful PhD adviser?



Yes, Alan Turing was a successful PhD adviser. He supervised several Ph.D. students who went on to become prominent researchers in computer science, mathematics, and other fields. Turing's approach to research was to encourage his students to think independently and to pursue their own ideas, rather than imposing his own views on them. This approach allowed his students to develop their own research interests and contribute to the field in their unique ways.



Turing's Ph.D. students made significant contributions to computer science and other fields, including the development of programming languages, networking, code-breaking, and the theory of algorithmic randomness and computability. Moreover, Turing was known for his commitment to mentoring his students and helping them develop their research skills. He provided his students with a supportive and stimulating research environment, which allowed them to flourish and make significant contributions to their respective fields.

Overall, Turing's success as a PhD adviser is evident in the achievements of his Ph.D. students, many of whom became successful researchers and made significant contributions to their fields.

So is it game over?

Well understood:

Fluency masks factual inaccuracies in “LLM Search”

So is it game over?

Well understood:

Fluency masks factual inaccuracies in “LLM Search”

Less well-understood:

How to build Search systems using LLMs

So is it game over?

Well understood:

Fluency masks factual inaccuracies in “LLM Search”

Less well-understood:

How to build Search systems using LLMs

ChatGPT, Sydney, Bard (+ ofc. Bing, Google) are Search systems.

What is Search?

What is Search?

“A journey of a 1000 miles begins with a single query...”

– “Lao Tzu”

When you think Search, think:

When you think Search, think:

The Google logo is displayed in its characteristic multi-colored font. The letter 'G' is blue, the first 'o' is red, the second 'o' is yellow, the 'g' is blue, the 'l' is green, and the 'e' is red. The letters are rendered in a clean, sans-serif typeface with rounded terminals.

Google

When you think Search, think:

Google



Microsoft Bing

When you think Search, think:

Google



Microsoft Bing



neeva

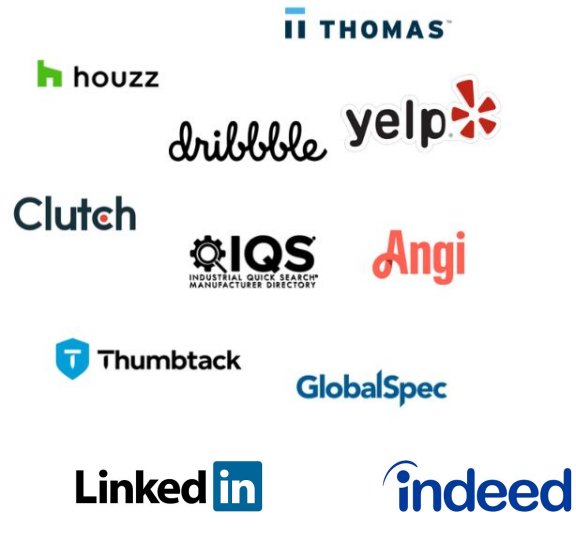


DuckDuckGo

...but don't stop there!

When you think Search, also think:

When you think Search, also think:



Think about the “data types”

Think about the “data types” of:

Search Engines

Search Intents

Think about the “data types” of:

Search Engines

– Web Search

Search Intents

Think about the “data types” of:

Search Engines

- Web Search
- Site Search

Search Intent

Think about the “data types” of:

Search Engines

- Web Search
- Site Search
- Listings Search

Search Intent

Think about the “data types” of:

Search Engines

- Web Search
- Site Search
- Listings Search
- Enterprise Search

Search Intents

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent

Search Intents

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intents

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intents

- Navigational

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intent

- Navigational
 - Informational

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intent

- Navigational
- Informational
- Transactional

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intents

- Navigational
 - Informational
 - Transactional
- Single vs. Multiple outcomes

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intents

- Navigational
 - Informational
 - Transactional
- Single vs. Multiple outcomes
- **Short-lived vs. Long-lived**

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intents

- Navigational
 - Informational
 - Transactional
- Single vs. Multiple outcomes
- Short-lived vs. Long-lived
- Task intent vs. Browse intent

Think about the “data types” of:

Search Engines

- Web Search
 - Site Search
 - Listings Search
 - Enterprise Search
- Crawled vs. Hosted vs. Latent
- Flat vs. Sliced vs. Faceted

Search Intents

- Navigational
 - Informational
 - Transactional
- Single vs. Multiple outcomes
- Short-lived vs. Long-lived
- Task intent vs. Browse intent

Takeaway #2:
Search is messy and multi-dimensional

Takeaway #2:
Search is messy and multi-dimensional

How will Generative AI touch the
broader Search landscape?

Outline for the Rest of the Talk

Organize Search Engines / Search Intents into a simple dichotomy

- **Answers vs. Decisions**
- Spoiler Alert:
Pure LLMs might solve “answer” searches but not “decision” searches
- Identify the fundamental barrier for LLMs to solve the latter

Describe the Tonita Architecture for Decision searches

Hallucinate about how Gen AI might reshape the Web

Let's organize that Engines/Intents taxonomy a bit better...

A. The “interactive Wikipedia” model for answers:

- conversation on a topic – broad or narrow – that can possibly drift as it evolves

B. The “decision-maker” model:

- explore options related to a goal-oriented task, interact with a database of listings, gather knowledge, make one or more choices

Let's organize that Engines/Intents taxonomy a bit better...

- A. The “interactive Wikipedia” model for answers:
 - learn about carbon sequestration, this year's Oscars, diabetes, Victorian era scandals, find out basketball scores...

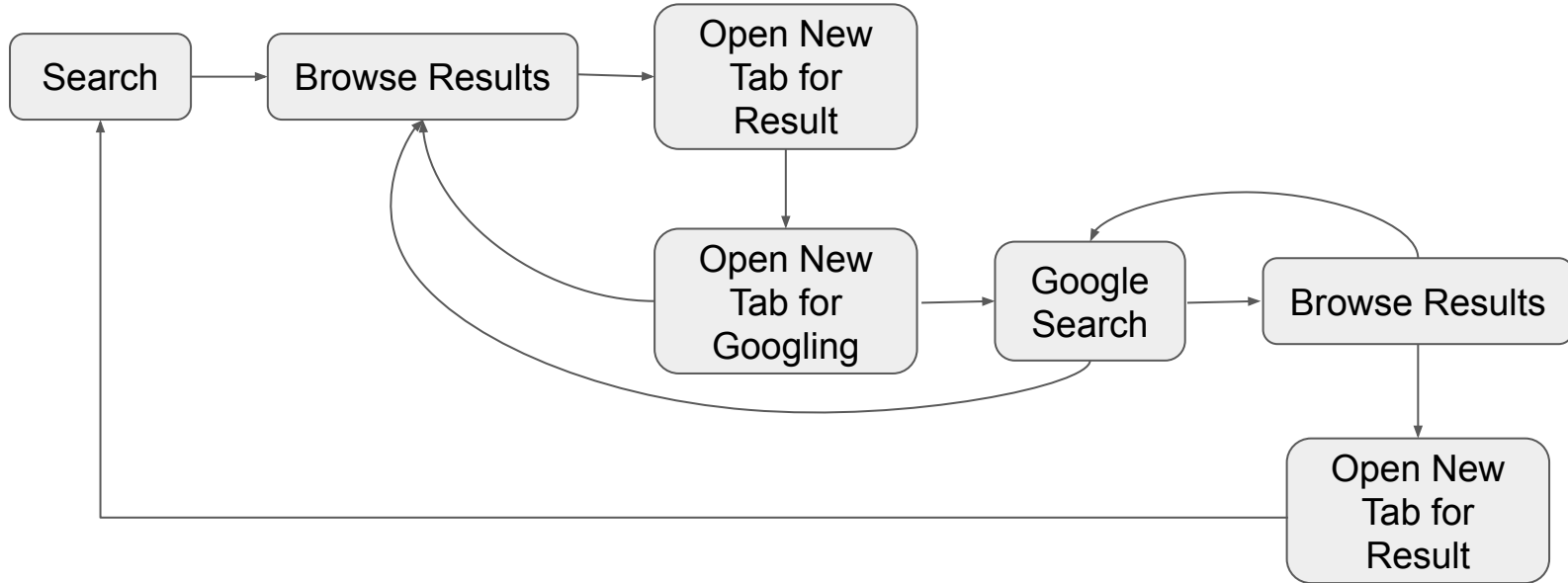
- B. The “decision-maker” model:
 - pick a hair dryer to buy, choose jobs to apply to, find an apartment to rent, get replacement parts for hydraulic press, identify best targets for marketing emails...

The interactive Wikipedia model

Out-of-the-box LLMs do well for “well-represented” topics

- might hallucinate for “tail” topics
- fine-tuning might help significantly:
 - data fine-tuning on specialized corpora, fresh data
 - instruction fine-tuning to encourage or discourage specific “behaviors”
 - **FLAN style:**
fine-tune on a small set of NLP tasks phrased as instructions
 - **InstructGPT style:**
fine-tune with a large set of natural instructions + human feedback (via RL)
 - **Chain-of-thought prompting**

The decision-maker model



The decision-maker model

Out-of-the-box LLMs are largely useless here!

Micro-topics, details matter, often niche vocabulary

Data items have a mix of structured and unstructured info

But there's a more fundamental problem

But there's a more fundamental problem

LLMs can't make "references"

- to specific items in their training data / fine-tuning corpus

("as explained in Section 14.1, *What ails world peace?*, paragraph 2,...")

This inability to "ground" the conversation in real-world entities will likely be a major limiting factor on how much LLMs can solve Search in the "decision-maker" model

An example

Consider the apartment search query

“2br SF apt within walking distance to BART”

Includes structured constraints, geo-spatial constraints,
and assumes a fairly advanced “world model”

+

The actual listings available in the DB at that moment

**What might a powerful
“decision-maker” experience
look like?**

Help me find an apt in SF

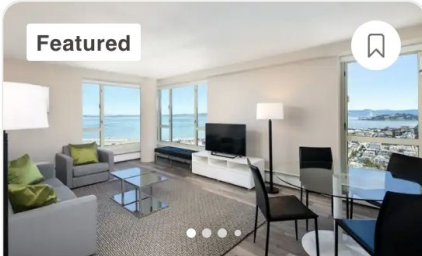
Sure, do you have specific neighborhoods or types of apartments in mind?

yeah, I commute to the Peninsula and Oakland frequently, so closer to the Bay Bridge is good

Sounds good, let me look up some apartments for you

Ok, here are some...

Featured




\$2,999+

Studio - 3 Beds • 1 - 3 Baths 17h ago

1000 Chestnut Street Apart...

Hardwood Floor • Dishwasher • On S...



\$2,150+

Studio - 2 Beds • 1 Bath 2d ago

Nob Hill Place

Fitness Center • On Site Laundry • P...



\$2,77

Studio - 2

The Civic

Assigned

2 br, preferably with a gym

2 br, preferably with a gym

Got it, how about these?



\$4,678+

2 Beds • 2 Baths

5d+ ago

The Civic ✓



\$3,550+

2 Beds • 2 Baths

3d ago

HighPoint

1888 Geneva Ave, San Francisco, ...

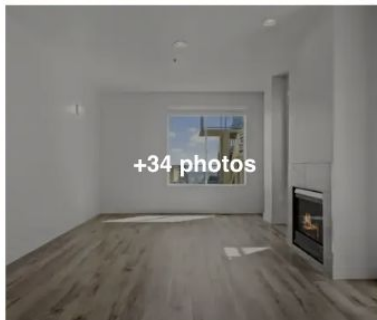


\$3,47

2 Beds •

Lakewo

515 John



\$3,550–\$3,600



HighPoint Apartments

1888 Geneva Ave, San Francisco, CA 94134, USA

Pet Friendly

Excelsior · 1888 Geneva Ave, San Francisco, CA 94134, USA

Amenities for HighPoint

APARTMENT AMENITIES

- + Balcony
- + Dishwasher
- + Hardwood Floor
- + Microwave
- + Satellite
- + Track Lighting
- + Cable
- + Electric Fireplace
- + Large Closets
- + Refrigerator
- + Stainless Steel Appliances
- + Washer and Dryer

BUILDING AMENITIES

- + Cats OK
- + Fitness Center
- + Gaming Stations
- + Multi-use Room
- + Residents Lounge
- + Swimming Pool
- + Dogs OK
- + Free Weights
- + Media Room
- + On-Site Maintenance
- + Storage
- + TV Lounge

Show all 40 amenities

Units, floorplans, amenities, dimensions, details, availability, and prices may be approximate and subject to change. Please contact a community representative for more information.

Policies and fees



2 br, preferably with a gym

Got it, how about these?



\$4,678+

2 Beds • 2 Baths

The Civic ✓

5d+ ago



\$3,550+

2 Beds • 2 Baths

HighPoint

1888 Geneva Ave, San Francisco, ...

3d ago



\$3,47

2 Beds •

Lakewo

515 John

Is Geneva Ave a quiet neighborhood?

Is Geneva Ave a quiet neighborhood?

Yes, Geneva Ave is a quiet residential street; it is also close to McLaren Park, which is very popular among many residents, especially families.

cool - what's the nearest Caltrain station?

Bayshore station is a mile to the East from 1888 Geneva


Is it also close to BART?

The Balboa Park BART station is 2 miles to the West

The Balboa Park BART station is 2 miles to the West

too far! shortlist it, but show me something where I can walk to a BART station

Online tours




\$4,339

2 Beds • 2 Baths

Bennett Lofts ✓

10 mins walk to Bayview BART

2h ago



\$4,200

2 Beds • 2 Baths

5177 Diamond Heights Boulevard #2...

4 mins walk to Mission St BART

5d+ ago



5

2

1

Core Elements

Seamless interleaving of results + conversations

Smooth blending with other standard UI elements

(e.g., scroll, filter, save, share)

Rich Conversations

Question-answering:

about individual results, product categories, “world knowledge”

Adding / undoing constraints, accumulated personalized preferences,
preference elicitation

Bells and Whistles

Longer-lived sessions

Asynchronous sessions with multiple users

Executable actions

Ok, how do we accomplish this?

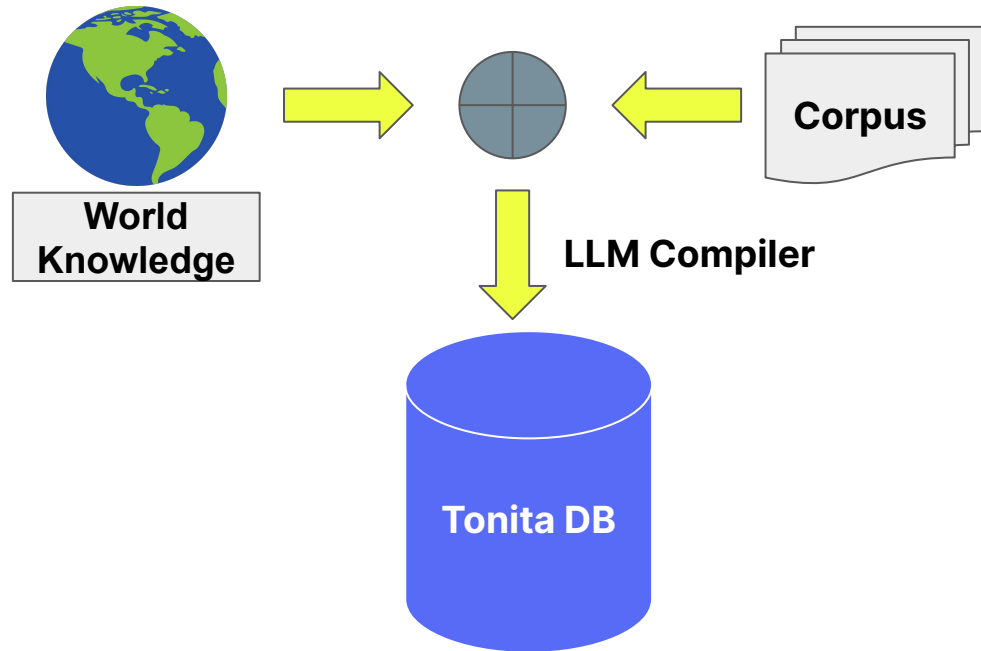
The Tonita Architecture

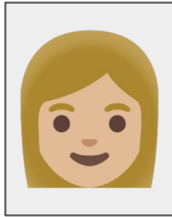
“Everything should be made as simple as possible, but no simpler”

Key Insight: LLMs are Incredible Compilers, Coders!

LLMs
are great at language
transformations,
and can be
trained with
a small number of
high-quality examples.

Challenge: Scaling!



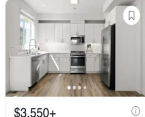


2 br, preferably with a gym

Got it, how about these?



\$4,678+
2 Beds • 2 Baths
5d+ ago
The Civic



\$3,550+
2 Beds • 2 Baths
3d ago
HighPoint
1888 Geneva Ave, San Francisco, ...



\$3,47
2 Beds •
Lakeview
515 John

Is Geneva Ave a quiet neighborhood?

Yes, Geneva Ave is a quiet residential street; it is also close to McLaren Park, which is very popular among many residents, especially families.

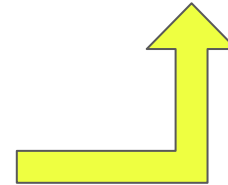
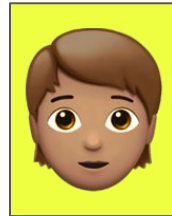
cool - what's the nearest Caltrain station?

Bayshore station is a mile to the East from 1888 Geneva

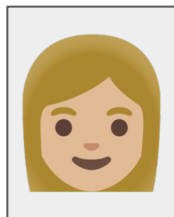
Is it also close to BART?

The Balboa Park BART station is 2 miles to the West

LLMs are great at writing code!
Teach them to write TQL to pull stuff from the compiled DB



Tonita QL

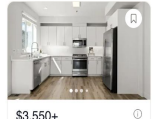


2 br, preferably with a gym

Got it, how about these?



\$4,678+
2 Beds • 2 Baths
5d+ ago
The Civic



\$3,550+
2 Beds • 2 Baths
3d ago
HighPoint
1888 Geneva Ave, San Francisco, ...



\$3,47
2 Beds •
Lakeview
515 John

Is Geneva Ave a quiet neighborhood?

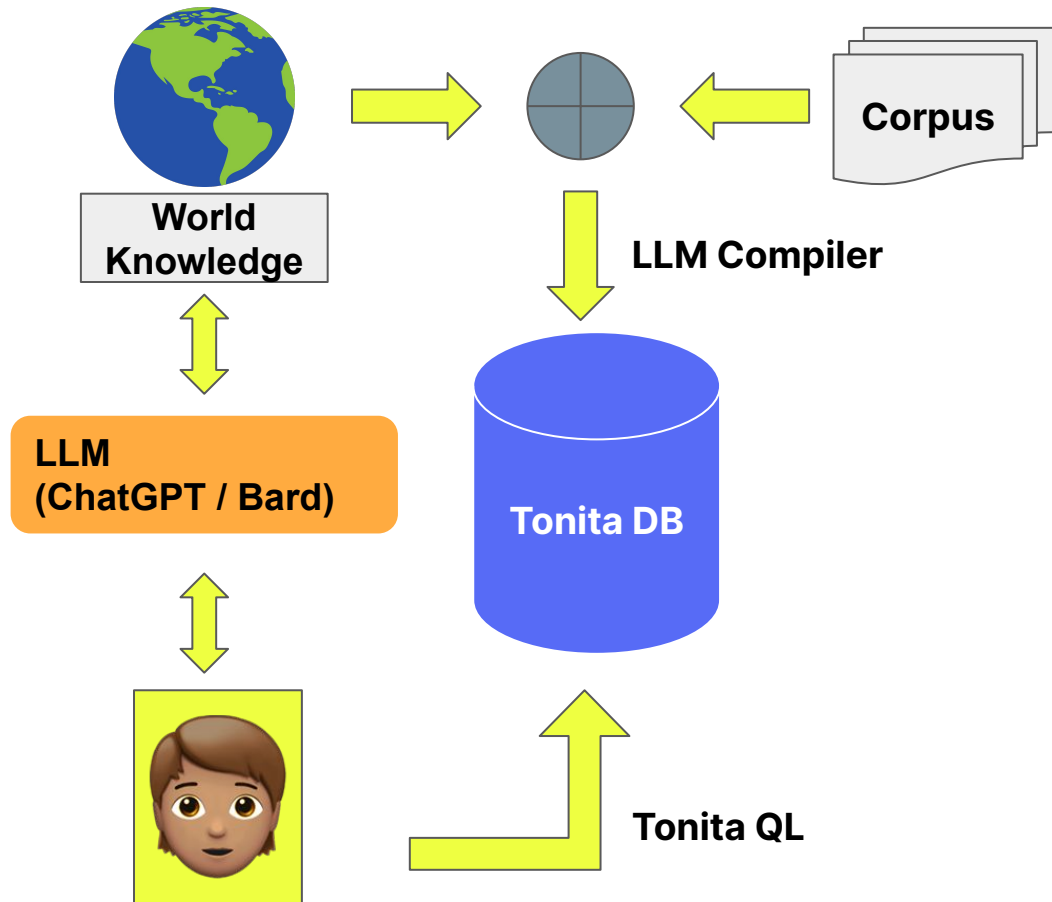
Yes, Geneva Ave is a quiet residential street; it is also close to McLaren Park, which is very popular among many residents, especially families.

cool - what's the nearest Caltrain station?

Bayshore station is a mile to the East from 1888 Geneva

Is it also close to BART?

The Balboa Park BART station is 2 miles to the West



Complement to the ChatGPT “plugin” model

Tonita brings
conversation capabilities
to Searching/browsing an
application corpus

ChatGPT brings
an application into a chat
session through a simple API

Complement to the ChatGPT “plugin” model

Tonita understands the application corpus deeply and uses LLMs to bring language fluency and world knowledge

A Tonita Search for the app is automatically a ChatGPT plugin!

ChatGPT plugins offer a shallow interface to the application corpus

The quality of the experience is only as good as the underlying search engine for the application

Summary

High-level dichotomy in Search:

- Answers vs. Decisions

Generative AI brings major simplification to the Search UX

- great for the “interactive Wikipedia” experience for answers

Described Tonita architecture for “decision-maker” Search

Final Thoughts

Will generative AI reshape the Web?

Will Generative AI reshape the Web?

The Web is awesome at the **presentation** layer and the **transaction** layer

- HTML/CSS/JS are great for these
- None of them is capability-centric

Led to the publish-crawl-index-search paradigm for the middle layer

What if the Web were **capability-centric** and presentation is just icing?

- websites become AI agents with deep knowledge of capabilities
 - compiled from private databases
- search engines evolve into routing / decision-making engines