

CDAO Spring 2024: Analytics in the New AI Era

Xin Heng, xin@tote.ai

Head of Data, AI Chief @ Tote AI

Formerly, Chief Data Officer @ Caffeine TV



Table of Contents

Where is New AI's Biggest Impact on Analytics

- The New AI started at ChatGPT's launch in Nov. 2022

The Prospect of the Two Types of Analytics

The New Analytics in Business Growth

How does Tote Approach the New AI

The New Era & Analytics

ChatGPT 4 ▾

• **Answer:** Wolong Nature Reserve.

3. **Question:** This river, also known as the "River of Black Dragons," forms part of the border between Russia and China and is one of the ten longest rivers in the world. What is its name?

• **Answer:** Amur River.

✖ **You**

What is the highest tsunami and what is the cause of it? Heard it was in Alaska

🤖 **ChatGPT**

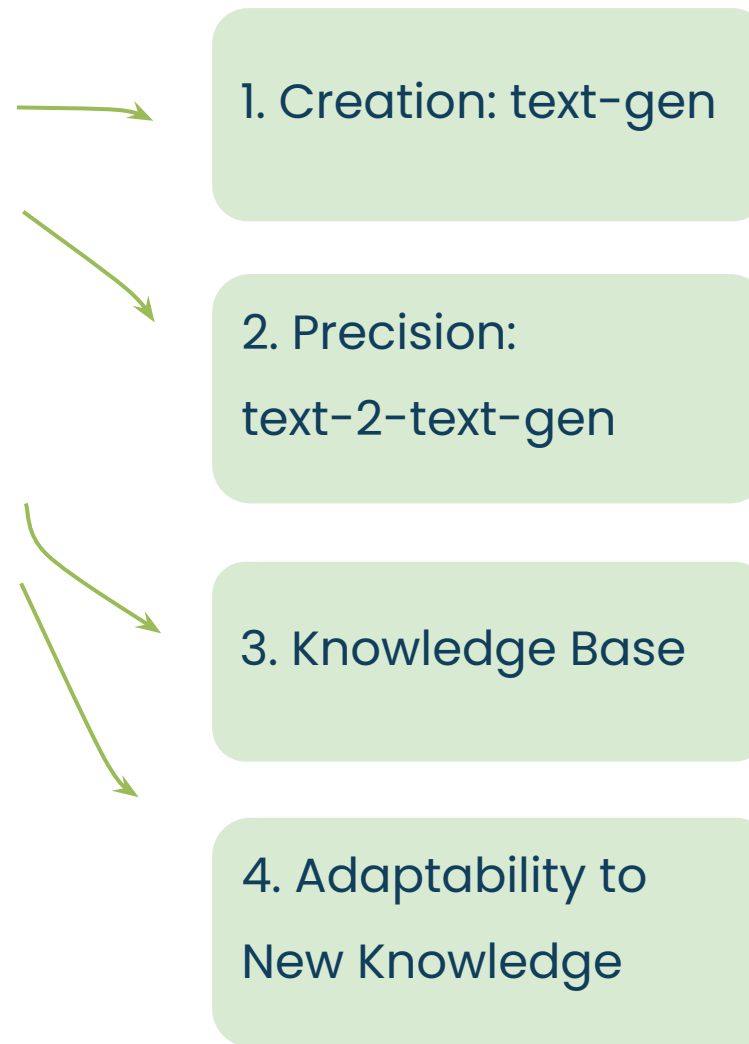
The highest tsunami wave ever recorded occurred in Lituya Bay, Alaska, on July 9, 1958. It reached an astonishing height of approximately 1,720 feet (524 meters) at its peak. This mega-tsunami was caused by a massive landslide triggered by a magnitude 7.8 earthquake along the Fairweather Fault. The landslide displaced a significant volume of water in Lituya Bay, a narrow fjord, resulting in the extraordinarily high wave that swept up the opposite side of the inlet. The tsunami stripped trees and soil down to bedrock along the slopes surrounding the bay, leaving a visible scar that can still be seen today. This event is a stark reminder of the powerful forces at play within the Earth's crust and the potential for sudden, dramatic changes in the landscape.

✖ **You**

synonym of Inescapable

🤖 **ChatGPT**

Unavoidable



The New Analytics

- A. The bot can **translate** a complicated user inquiry into one or multiple SQL queries + present it
- B. It can **adapt** to a particular environment of business and data

→ whoever can do A and B precisely, quickly and cheaply.

OpenAI “saved” the industry: your ROI in AI can be VERY positive

Two Types of Analysts

Data Analysts

They will face challenges

- text-to-sql
- text-to-dash

The average order value for all orders after 2023-04-01?

```
-- sql

select avg(order_value) as aov
from (
  select o.order_id, sum(o.quantity *
od.unit_price) as order_value
  from orders as o
  inner join order_details as od on
o.order_id = od.order_id
  where date(o.trans_time) > '2023-04-01'
  group by o.order_id
) as orders_total
```

- LLM is going to nail **stationary problems**
- Their **terminal state is so well defined** that the bot will get there with training & testing: even though a question can be translated to different sql syntax, the output data has a clear definition of correct or incorrect.

Data Scientists

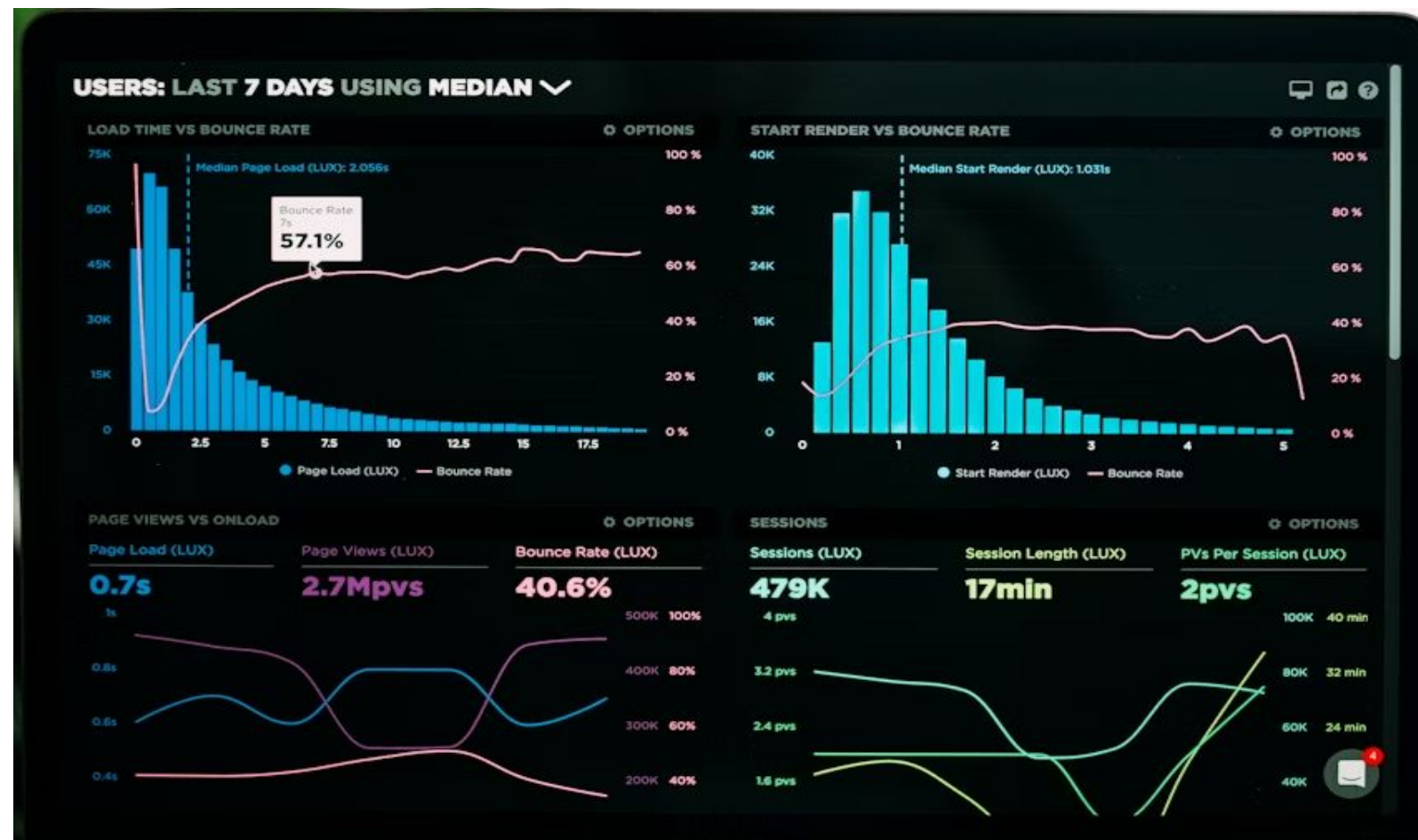
They will stay for a long while

- I believe LLM is not (or never) going to be good at non-stationary problems or as bad as humans
- The terminal state of a data science problem is ill-defined. 10 economists can come up with 10 reasonable thoughts
 - Example: what can explain my app's lack of user growth in Q1?

Data Analysts Must Evolve

The average order value for all orders after 2023-04-01?

```
-- sql
select avg(order_value) as aov
from (
  select o.order_id, sum(o.quantity *
od.unit_price) as order_value
  from orders as o
  inner join order_details as od on
o.order_id = od.order_id
  where date(o.trans_time) > '2023-04-01'
  group by o.order_id
) as orders_total
```



(Or say how an enterprise structures a analytics group must evolve)

Data analysts can also become an “Analytics Engineer”

- Expert user of text-to-sql, text-to-dash, ...
- Knows dbt, airflow, and other analytics tech stack
- Knows SQL optimization, and how to instruct / fine-tune the sql agent

Data analysts should become a data scientist

- Deep understanding of the economics of the business
- Knows statistics/economics

New Analytics and Product & Business Growth

Popular Areas

- Gen-AI can analyze (or **predict**) user behavior & improve product offerings.

- Gen-AI can completely upgrade customer service experiences, or generally **human-machine** interactions

- Gen-AI can reduce human errors

- Gen-AI can automate data collection, processing and reporting

The Impact

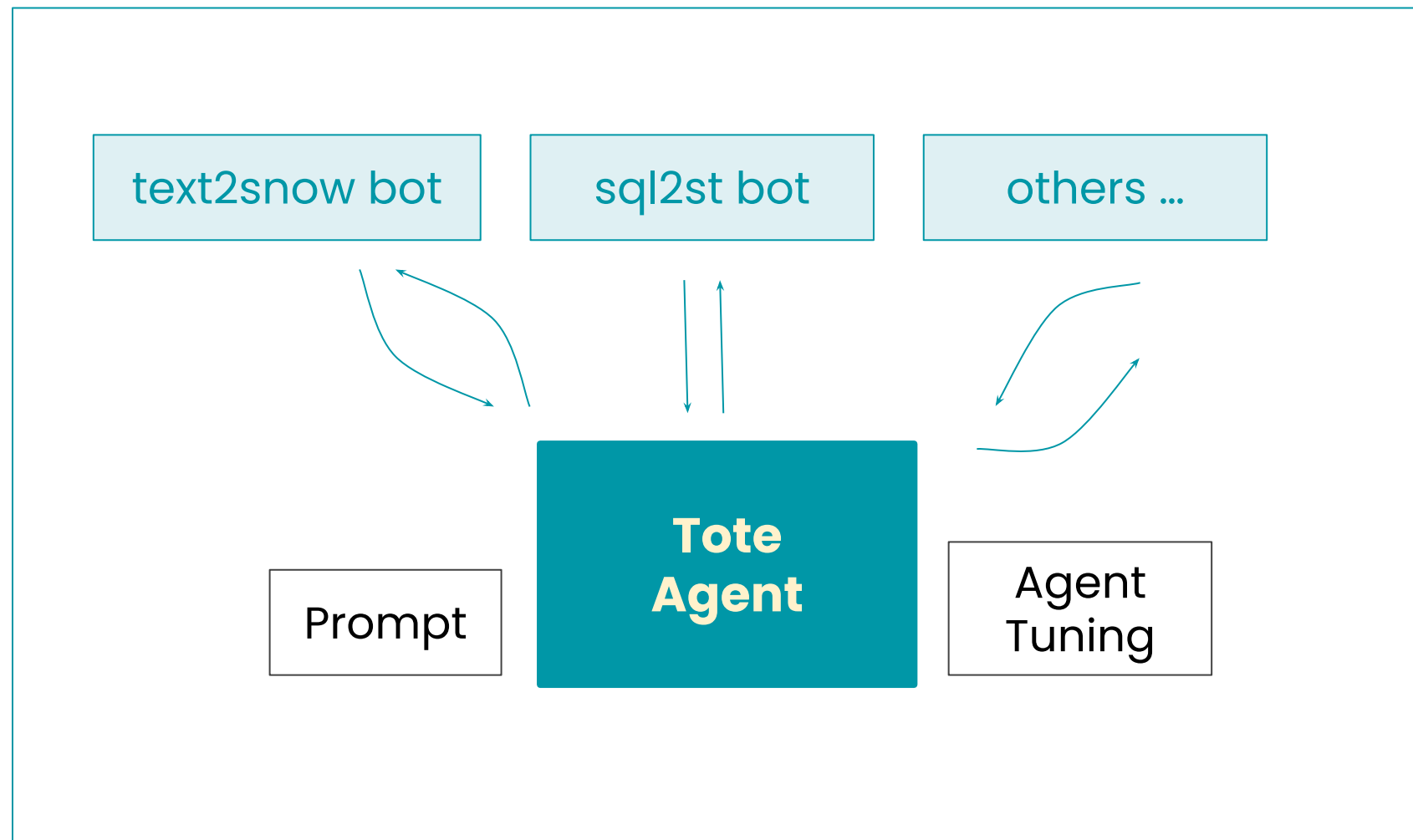
- Up the efficiency of product or marketing workflow (1st-order)
- Difficult to do much better than current approaches (2nd-order effect)

- Big impact (1st order effect)

- Big impact on efficiency & reducing fatigue
- Less trivial in reducing logic errors (2nd-order)

- Big impact: Analyst + Engineer = Analytics Engineer w/ half pay \$

How does Tote Approach “New AI”



The fundamental abstraction: **AGENT**

- Agent is Jarvis to Iron Man
- Features: multi-turn, multi-hop, multi-agent, and more
- **Bots** focus on doing one task very well. It has its own dedicated training (or dedicated developers)
 - Ex: “sql2st” converts sql into proper presentation in streamlit
- Bots are **modularized**, so are the agents. Their I/O interfaces are well-defined.

In 1 Year ...

text-to-sql, text-to-dash, ... many tools inside your data platform

Workflow improvement for product, marketing teams

- But can you really trust a PM self-drive it? 🤔

Anything else crazy you think Gen-AI can do

- No chance! 😄

Thank You



Thank You!
