Building a Generative Recommender with Chronon

Varant Zanoyan, CEO Zipline Ai

FEATURE STORE SUMMIT 2025

Organized by HOPSWORKS

About Chronon

Chronon


OpenAI


Roku


airbnb


PLAID


stripe

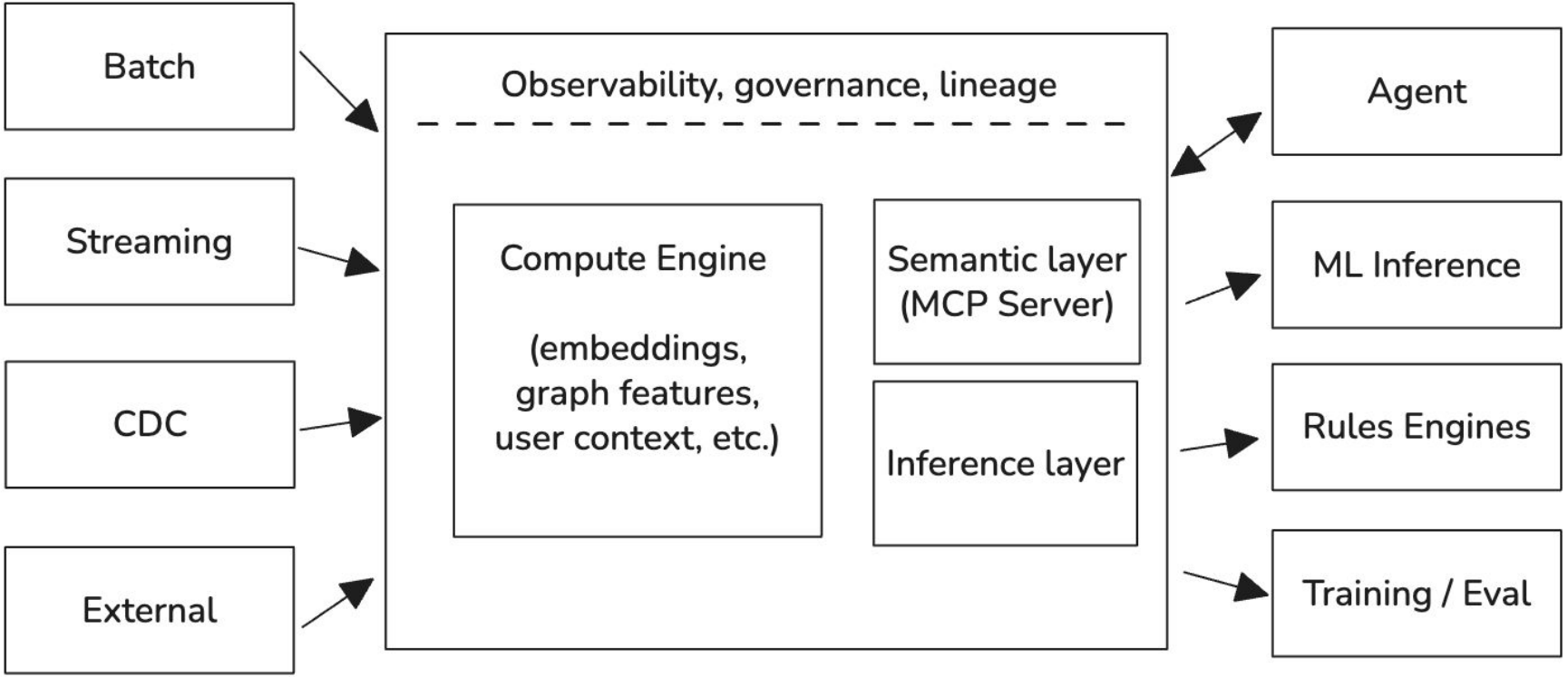

sardine


NETFLIX


Uber


Airwallex

Chronon


Batch
Streaming
CDC
External
Observability, governance, lineage
Compute Engine
(embeddings, graph features, user context, etc.)
Semantic layer (MCP Server)
Inference layer
Agent
ML Inference
Rules Engines
Training / Eval

Recommender Systems

Background

Goal is to recommend items to users

Content (i.e. pinterest, tiktok, instagram)

Listings (Airbnb, Amazon)

Etc.

Generative modeling

Improved metrics like CTR, purchases, etc.

Meta: **12.4%** improvement in engagement.

https://arxiv.org/abs/2402.17152

Generative modeling

Why does it work better?

Understands the temporal nature of events

For ex:

- **clicked on gardening-related items 3x in lifetime**
- **Vs: 3x in past week**

Generative modeling

Mechanism

LLM: attention on prior words to predict next

Recommender: Attention on prior activities

Generative modeling

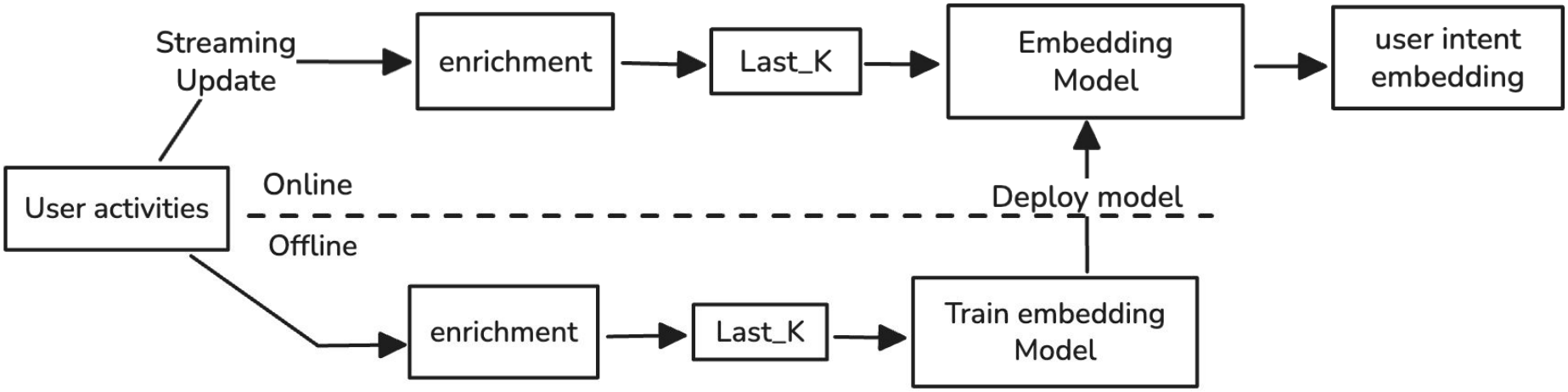
What does it need?

Long sequences + realtime activities

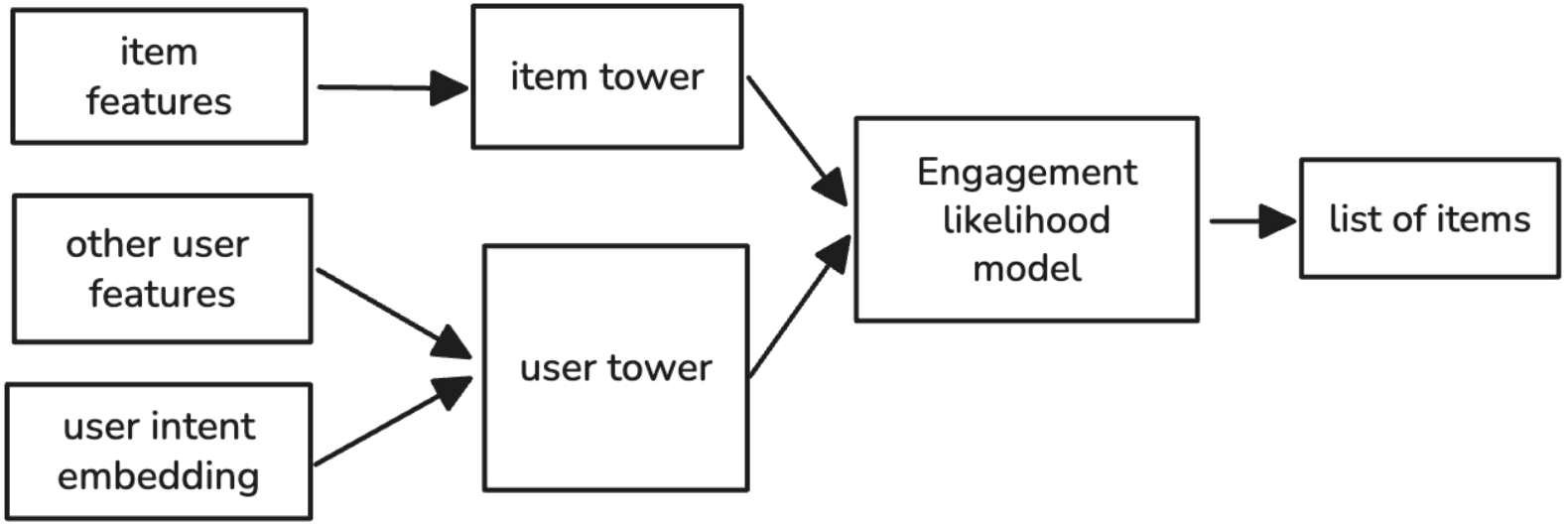
- **Lifetime clicks:** Many DIY projects
- **Recent clicks:** Gardening
- **Next:** DIY Gardening

Generative Models Architecture

User Sequence modeling

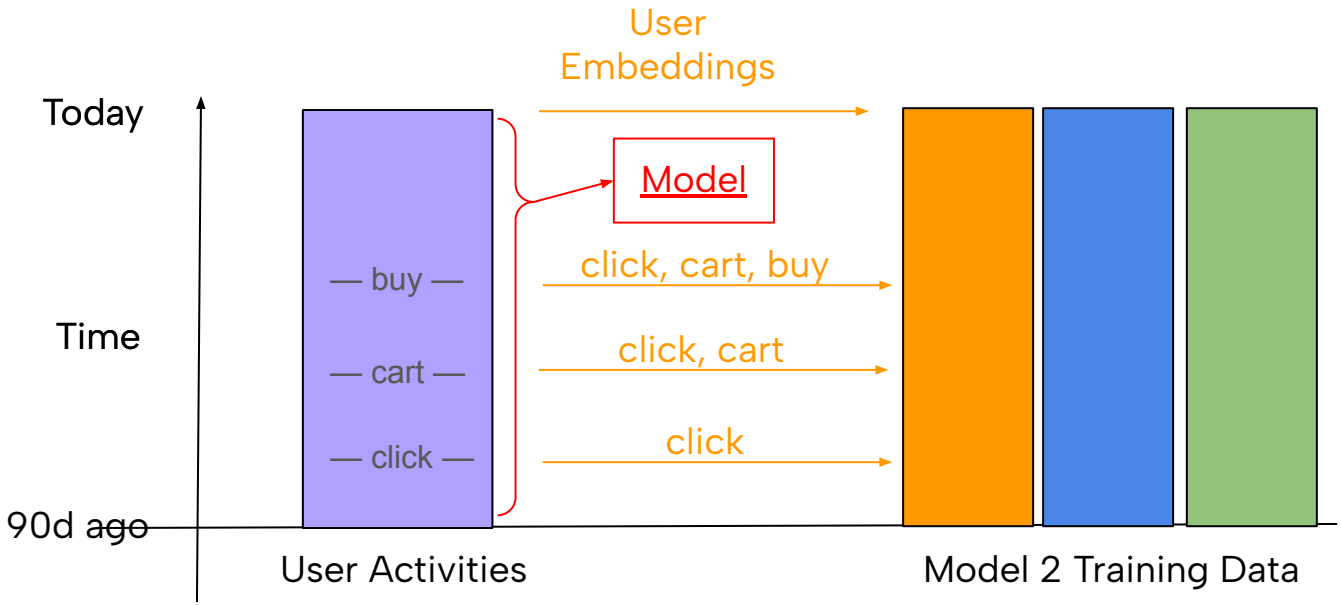


Full pipeline

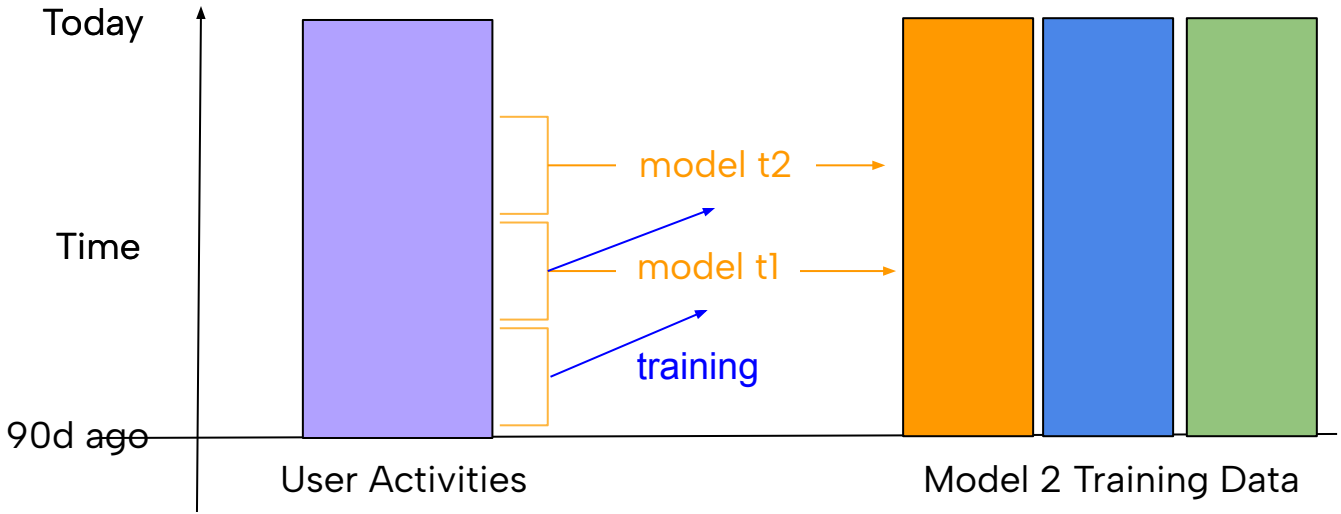


Generative Models Challenges

Model chaining orchestration



Model chaining orchestration



Model chaining orchestration

Chronon Solution

- Treat model as time partitioned
- Ensure that data is not leaked via embedding information

Realtime embeddings generation

Aggregating and embedding user activities

- Preparing user sequence with long history + realtime
- Triggering embeddings generation from user activities at scale
- Maintaining scalable realtime index of user → embedding
    - Batch correction for the index

Realtime embeddings generation

Chronon Solution

- Native API for:
  - Streaming aggregation
  - Model chaining
  - Indexing and serving

Other challenges

- Observability
- Lineage
- Iteration and experimentation

Thank you

varant@zipline.ai