Beyond Pretraining: Enhancing LLM Performance with Contextual Data

- Twilio
- Director of Engineering

- @awasthi_ankit
- /in/ankit-awasthi-60aa9412/

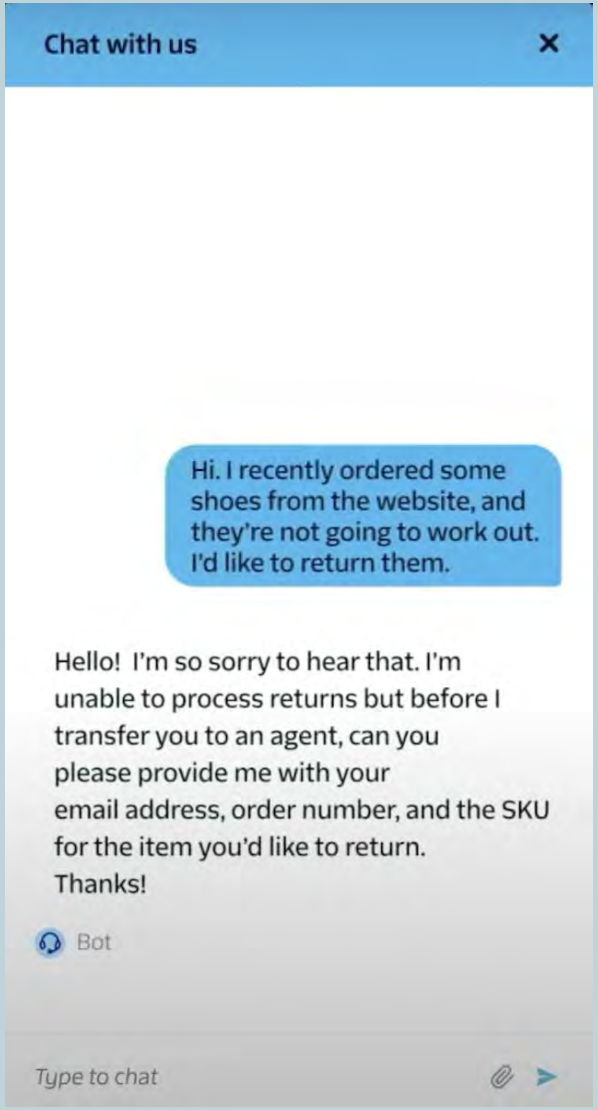
Strengths of LLMs

- Excellent generative capabilities.
- Conversational user interface.
- Remarkable versatility across a large variety of tasks such as summarization, code generation, copywriting.
- Reasoning.

Limitations of LLMs

- LLMs are trained on static training data.
- LLMs often make up information or hallucinate.
- Without contextual information, LLMs aren’t well suited for personalization.

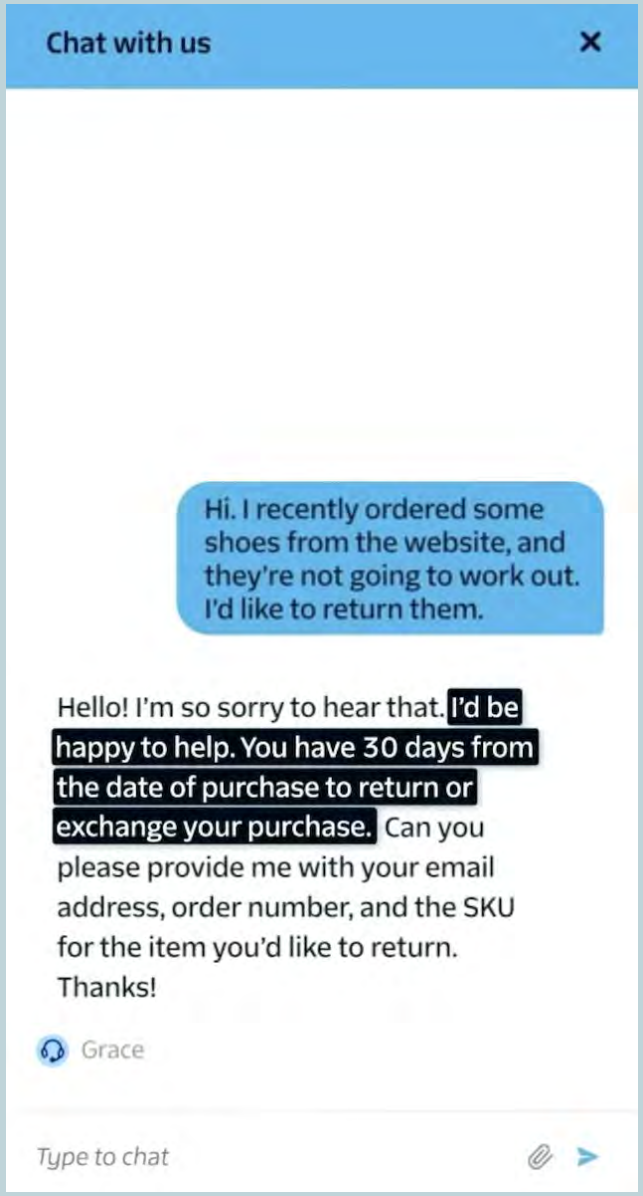
Limitations of LLMs: An Example

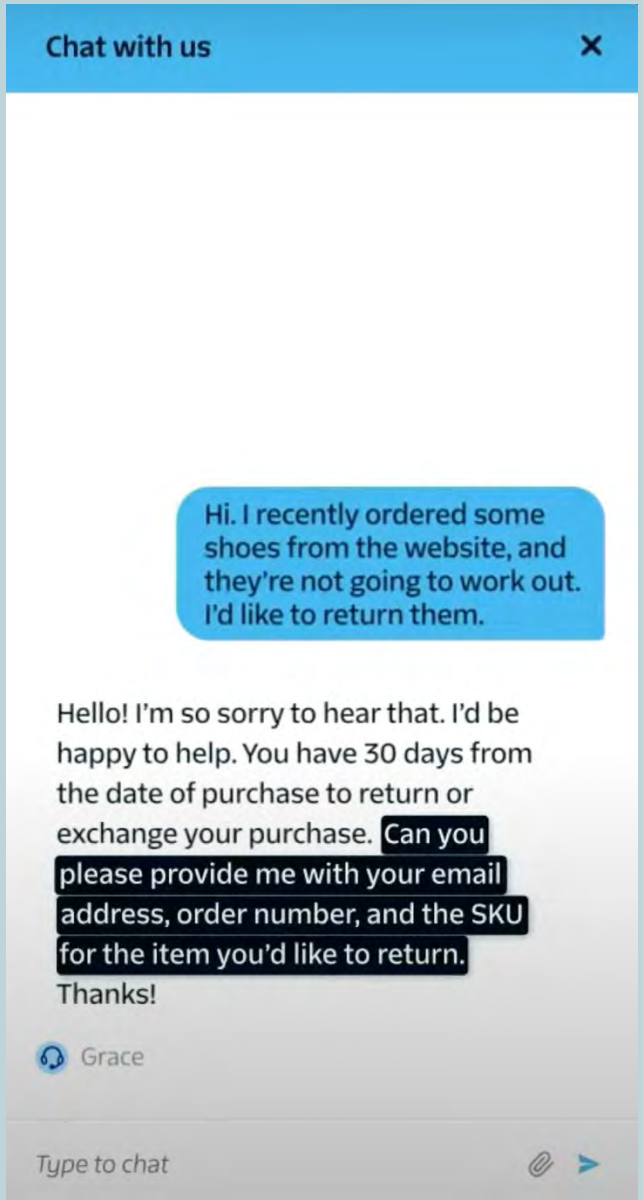


RAG to the rescue

- RAG stands for Retrieval Augmented Generation.
- The basic idea is to retrieve relevant contextual information, use that to augment the prompt to the LLM and then generate the response.
- RAG can provide LLMs
  - Current information.
  - Domain specific and often proprietary knowledge bases.
  - Contextual information for personalization.

LLMs + Knowledge Base (No Personal Context)

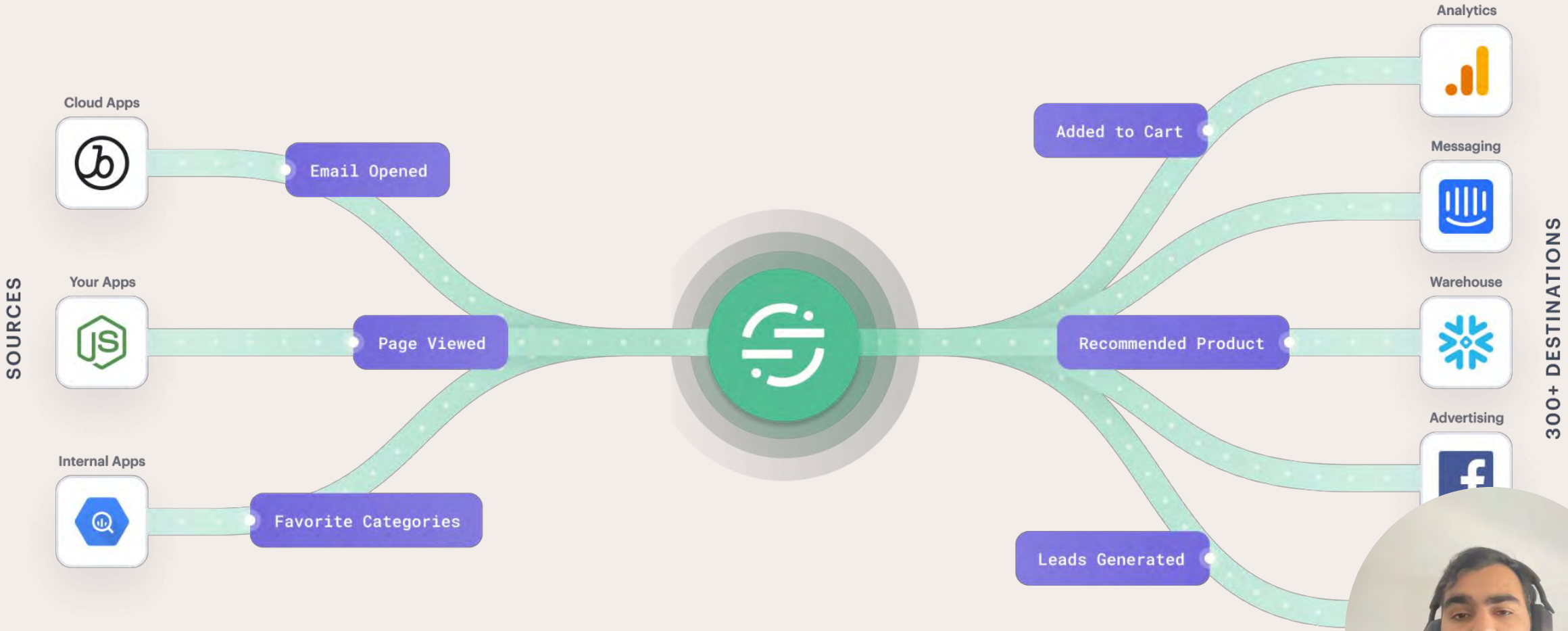




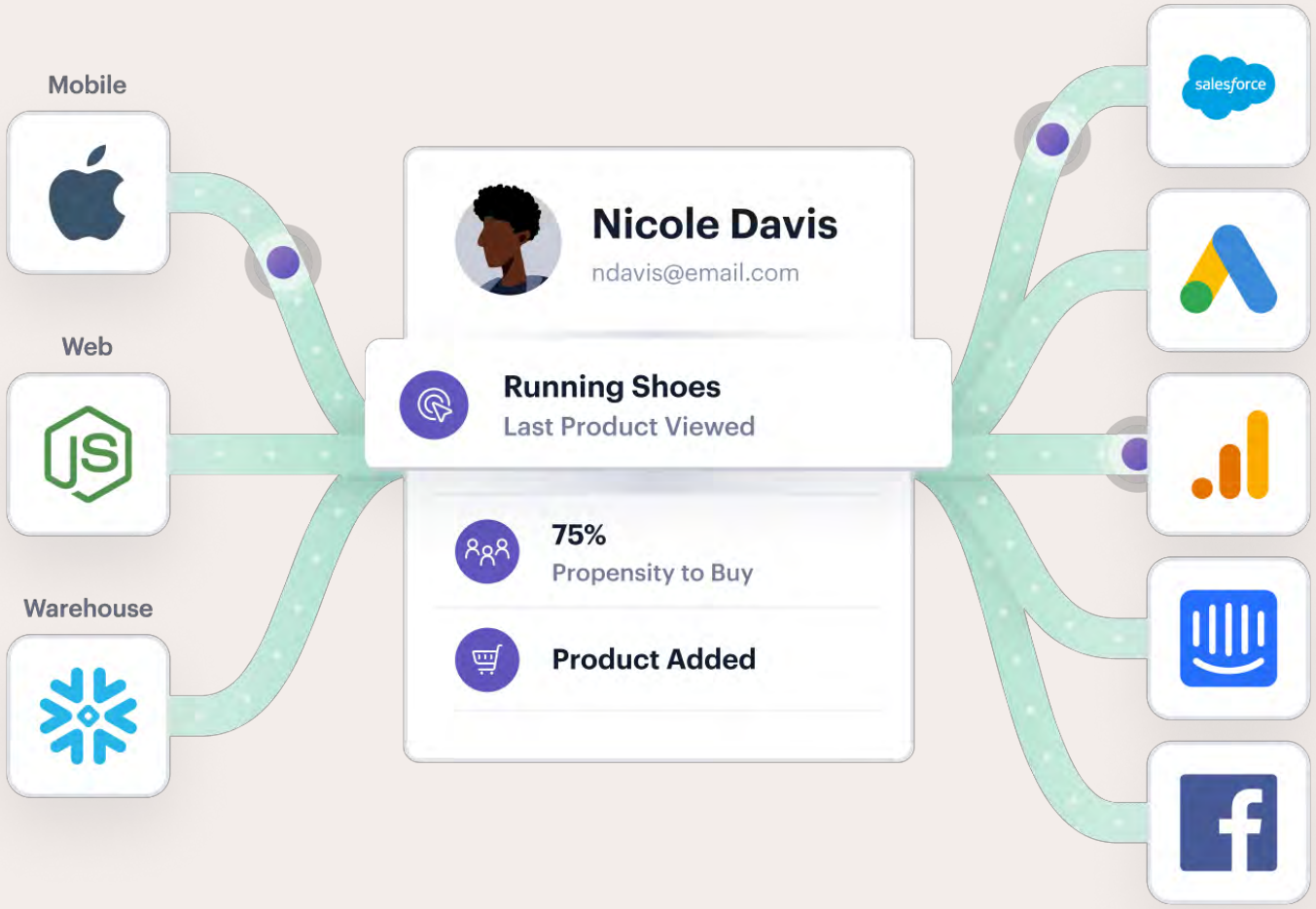
Need contextual information to enable personalization at a scale

Enabling Personalization at scale with Customer Data Platform

What is a Customer Data Platform ?

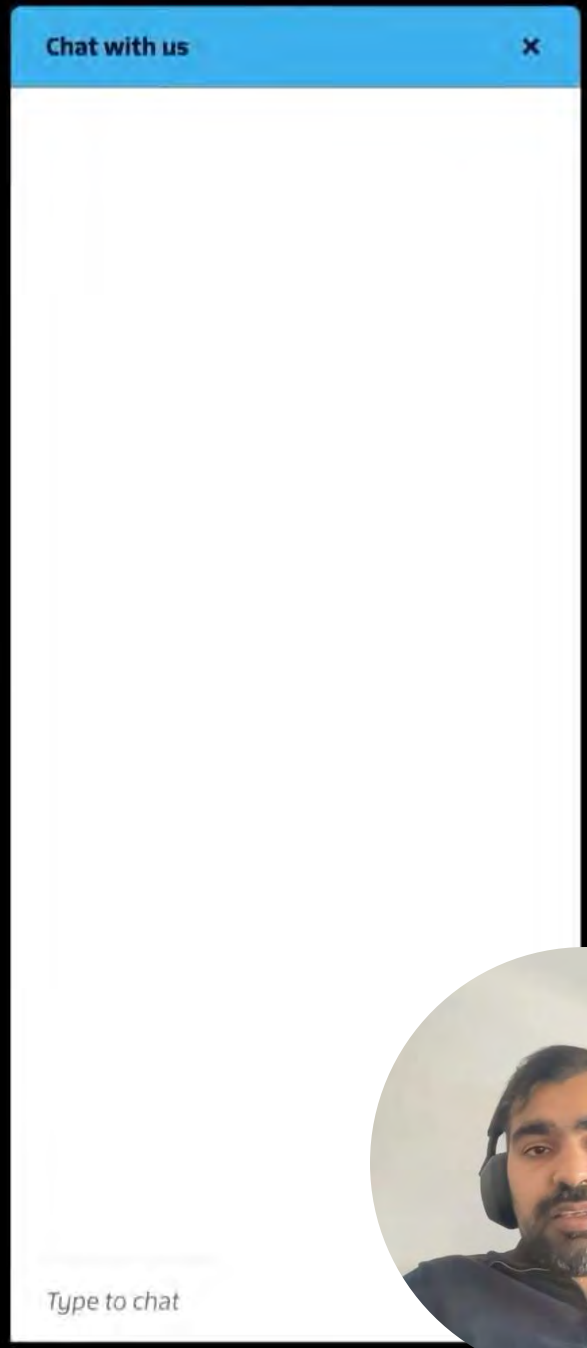


What is a Customer Data Platform ?

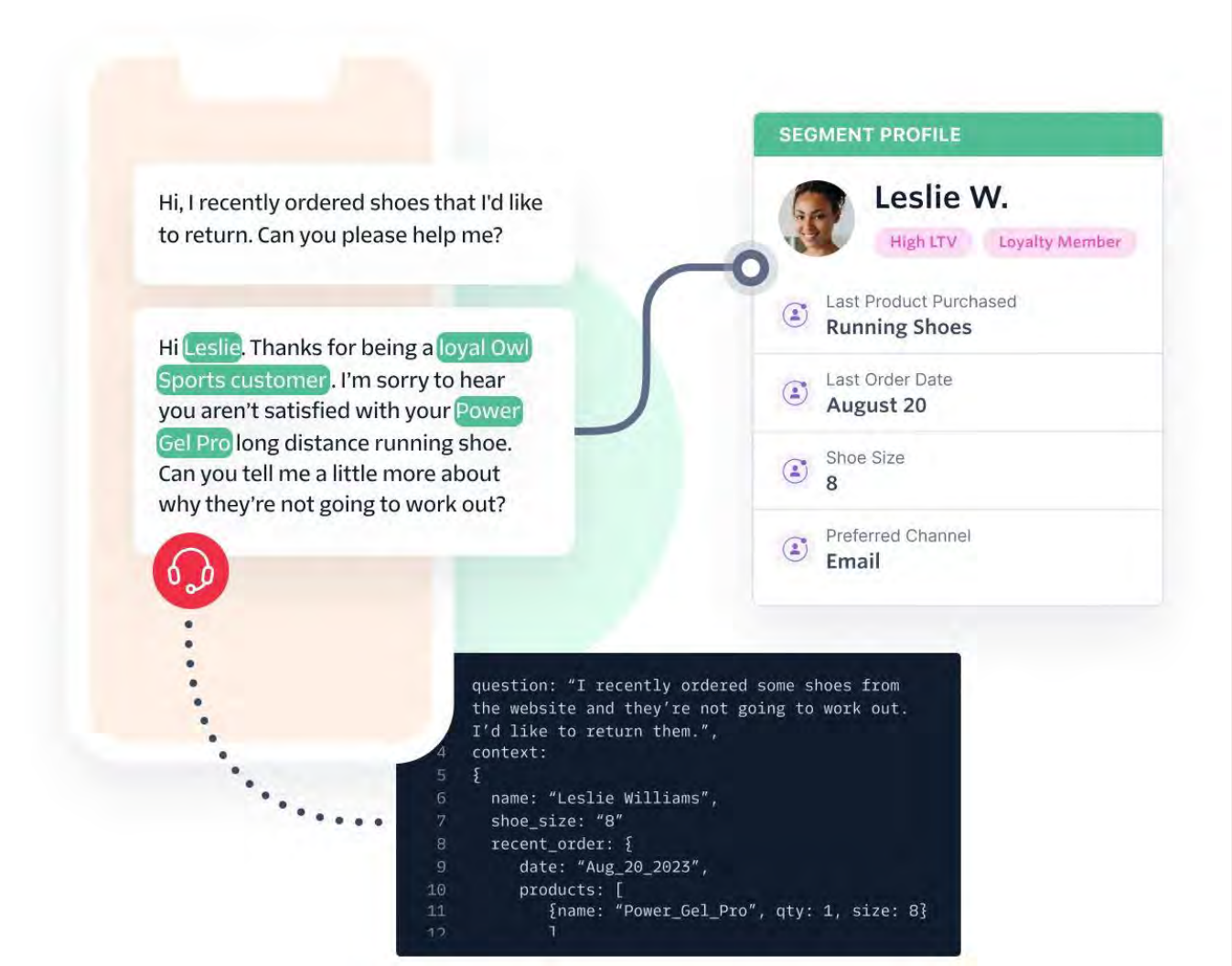


Created Ticket
SMS Sent
Added Item to Cart
Changed color
Viewed Product
SMS Sent

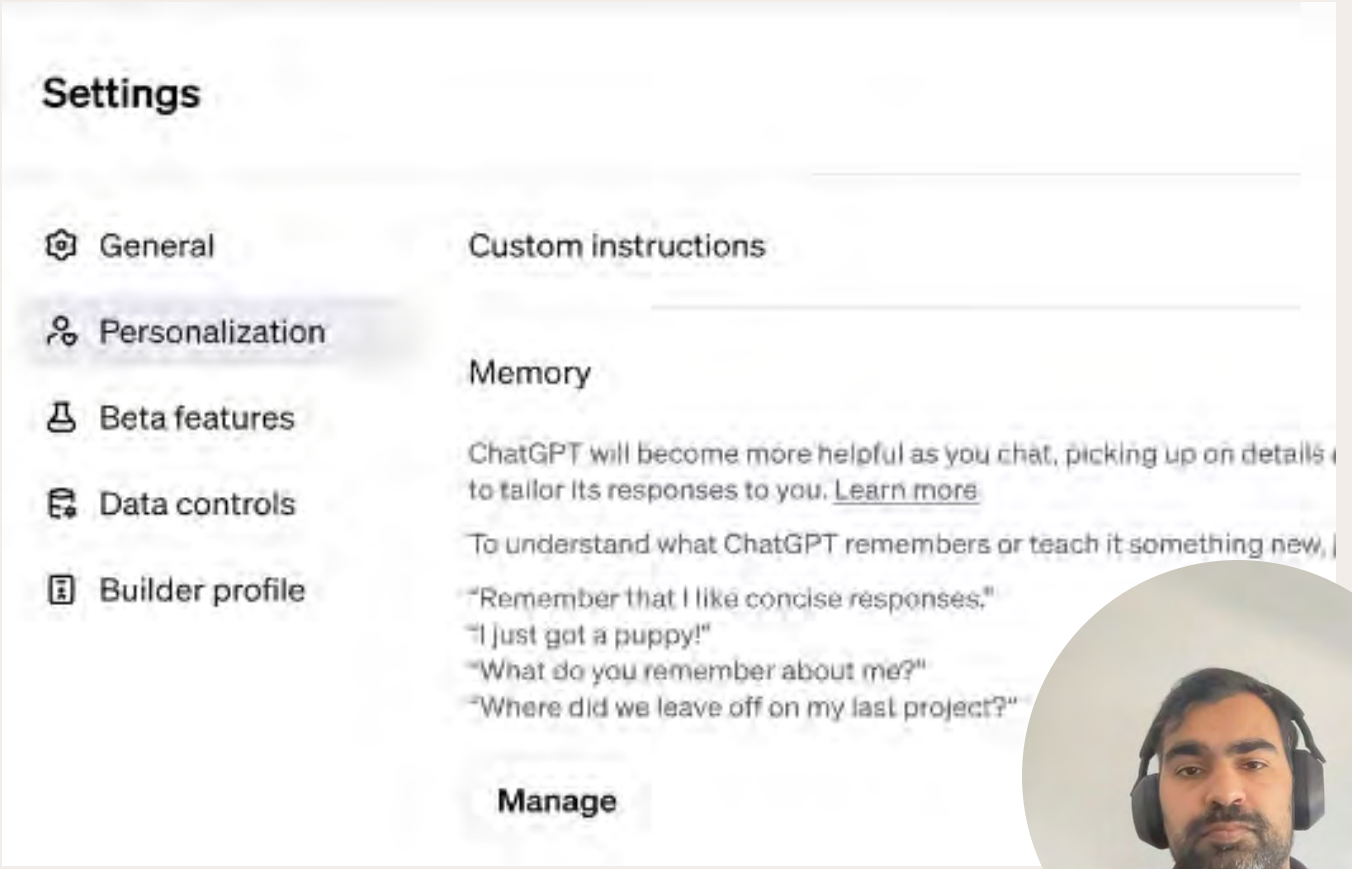
Jeff Lawson
Traits (132)
Events (2114)
Audiences
Identities
Phone Number
Address
Shoe size
Preferred Color
Birthday
Sports
Customer Since
Propensity to buy
Occupation
City
Preferred Channel


Chat with us
Type to chat

CDP as an AI Personalization Engine

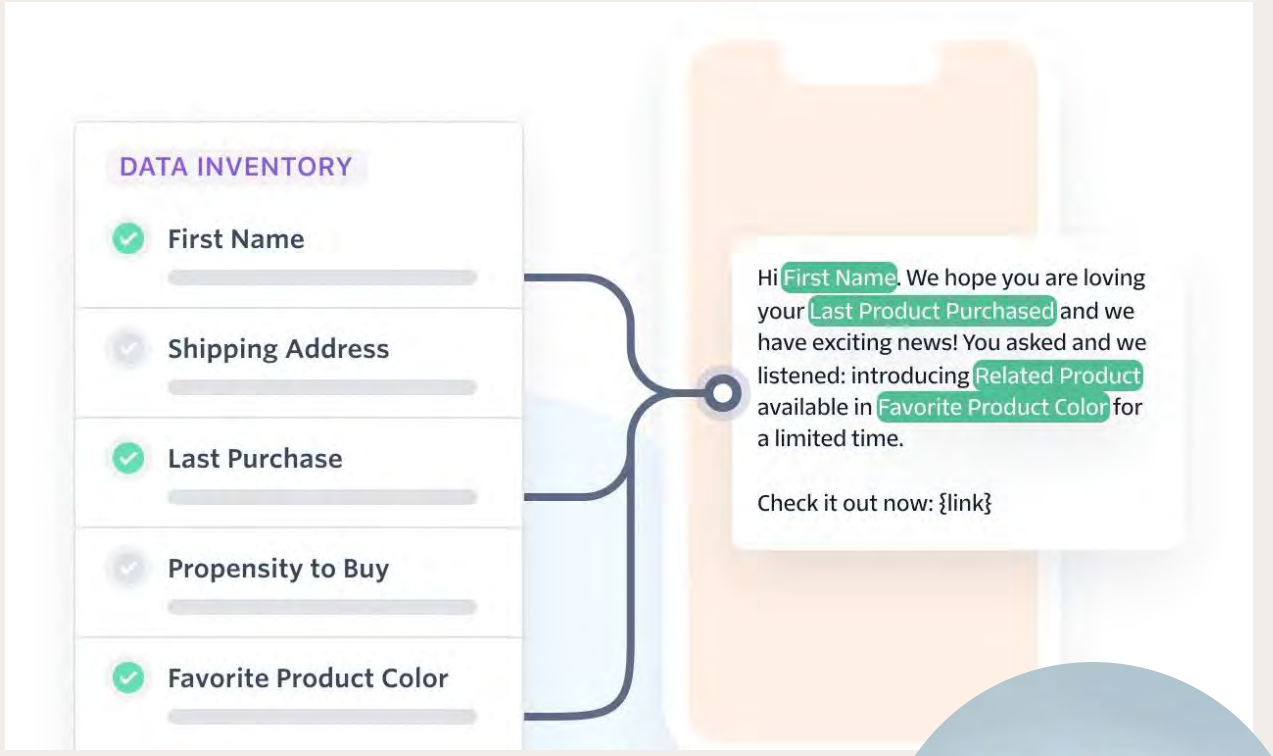


Memory in ChatGPT



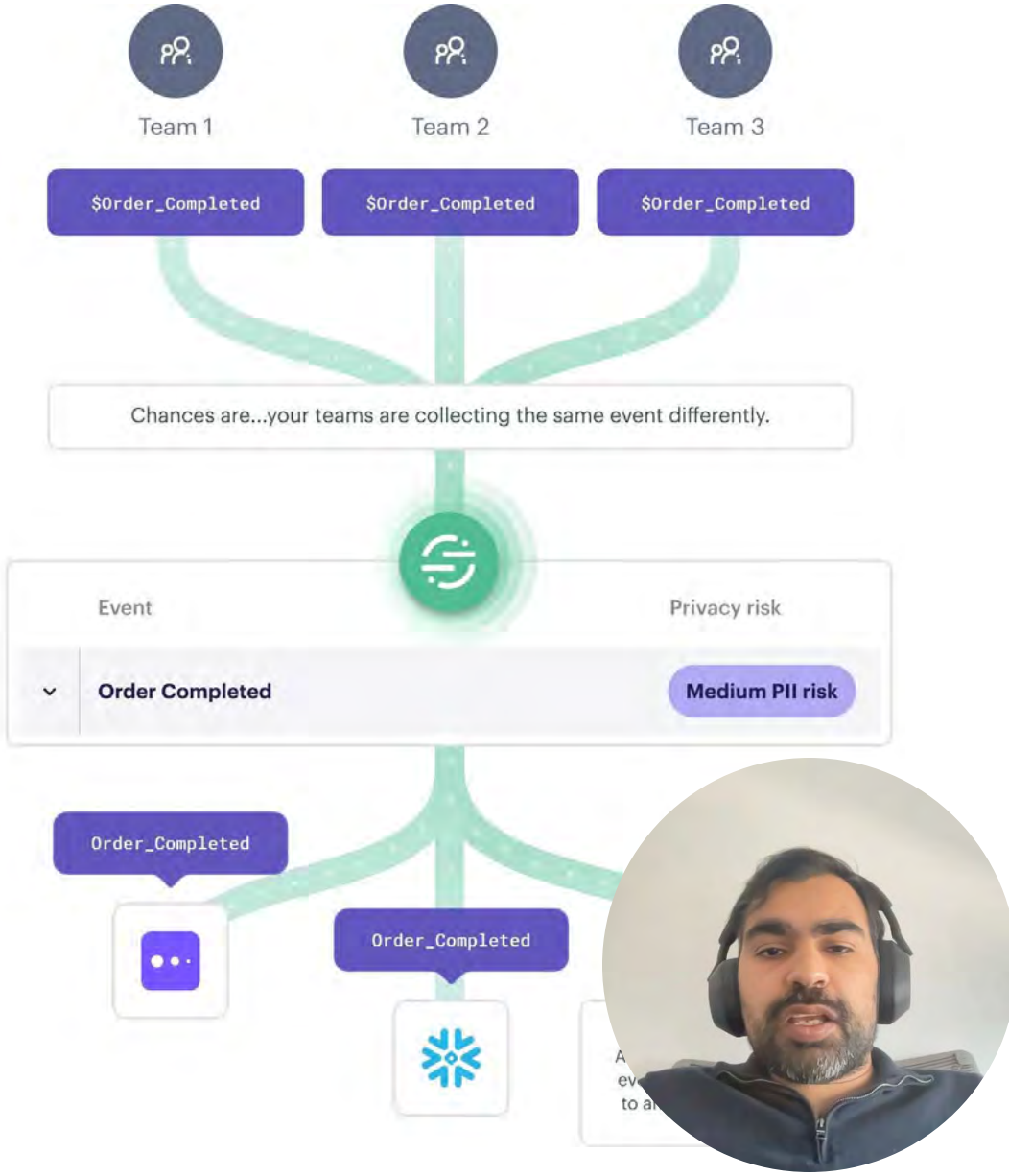
Personalization at Scale

- Truly 1:1 personalized messaging.
- Web personalization.
- Warm start customer support

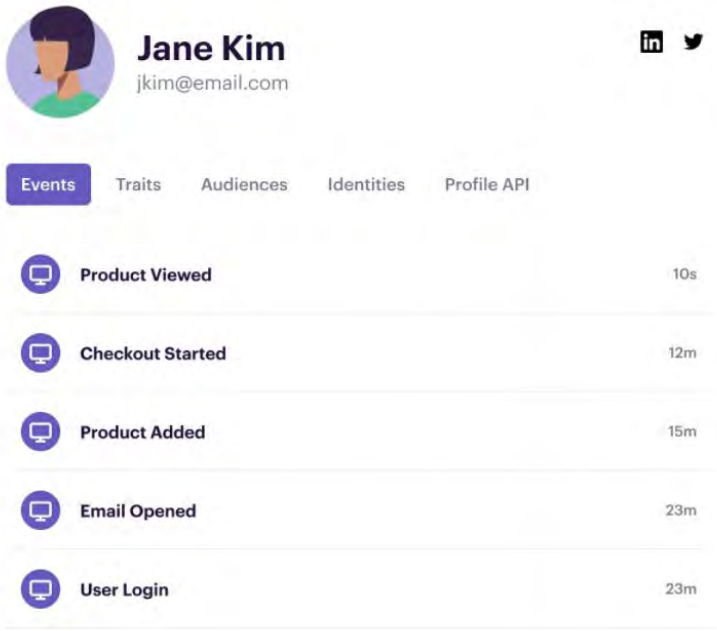


But why do we need a Customer Data Platform ?

- CDPs help break down information silos
- CDPs enable real time data capture and identity resolution across different applications help build a comprehensive view of each user.
- CDPs also provide advanced data governance over sensitive data.



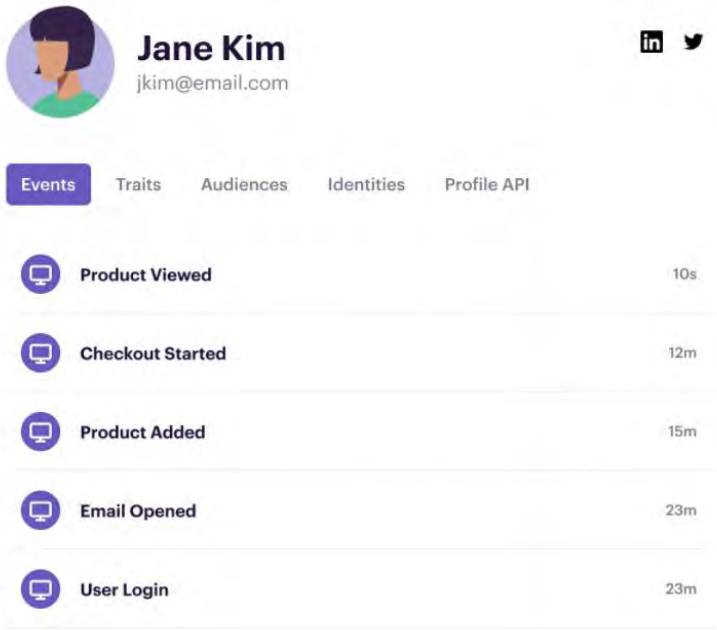
Personalization Engine: Solution 1



- Use all the details from the user profile as part of the prompt

NAME	ROUTE
Get a Profile's Traits	`collections/users/profiles/<identifier>/traits`
Get a Profile's External IDs	`collections/users/profiles/<identifier>/external_ids`
Get a Profile's Events	`collections/users/profiles/<identifier>/events`
Get a Profile's Metadata	`collections/users/profiles/<identifier>/metadata`
Get a Profile's Links	`collections/users/profiles/<identifier>/links`

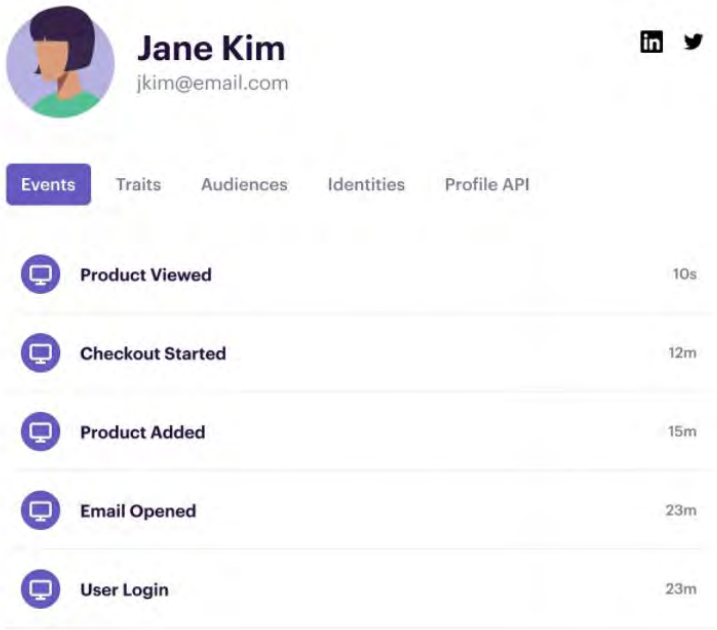
Personalization Engine: Solution 1



- Use all the details from the user profile as part of the prompt
  - Increases the context size dramatically if number of events or traits is high =>Higher latency and costs.

NAME	ROUTE
Get a Profile's Traits	collections/users/profiles/<identifier>/traits
Get a Profile's External IDs	collections/users/profiles/<identifier>/external_ids
Get a Profile's Events	collections/users/profiles/<identifier>/events
Get a Profile's Metadata	collections/users/profiles/<identifier>/metadata
Get a Profile's Links	collections/users/profiles/<identifier>/links

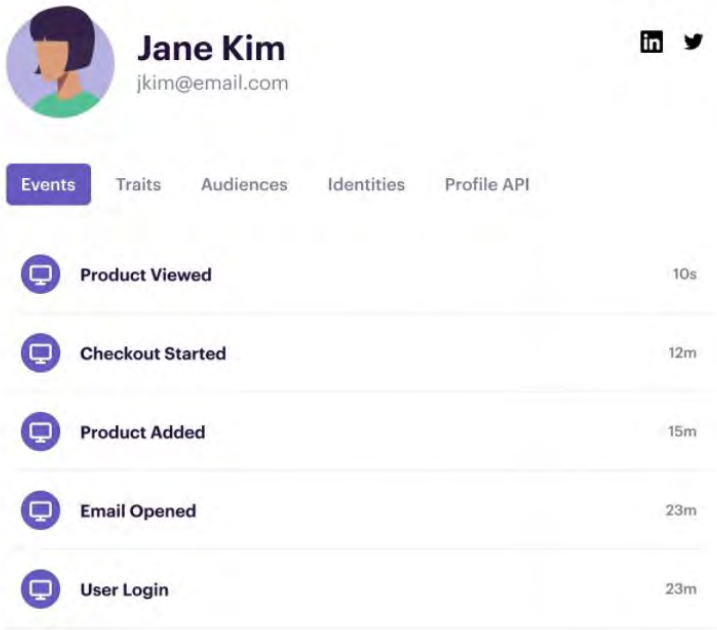
Personalization Engine: Solution 2



- Use only specific events or traits for personalization purposes

NAME	ROUTE
Get a Profile's Traits	`collections/users/profiles/<identifier>/traits`
Get a Profile's External IDs	`collections/users/profiles/<identifier>/external_ids`
Get a Profile's Events	`collections/users/profiles/<identifier>/events`
Get a Profile's Metadata	`collections/users/profiles/<identifier>/metadata`
Get a Profile's Links	`collections/users/profiles/<identifier>/links`

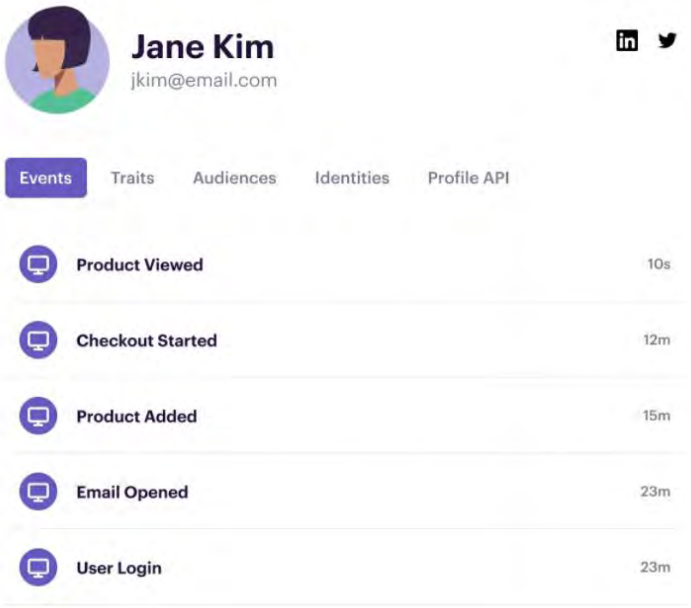
Personalization Engine: Solution 2



Use only specific events or traits for personalization purposes

- Unable to use all the relevant data available in the customer data platform.
- Requires prior knowledge of what is stored in the user profile.

Personalization Engine: Solution 3



Using the application or user context in natural, retrieve a subset of *relevant* traits and events.

- Pros
  - Small number of events or traits that can easily fit in the prompt.
  - Ability to use any data stored in the user profile without having any prior knowledge of what is stored in the profile.

Key Takeaways

- Using LLMs for many use cases becomes a data infrastructure problem.
- CDPs as Personalization engine provide 1:1 context for LLMs in real time.
- CDPs can act Long-term memory for LLMs by deriving struct information and storing them as user profiles traits or ever

Thank you

Ankit Awasthi

@awasthi_ankit

/in/ankit-awasthi-60aa9412/

References

- https://segment.com/solutions/ai/personalization-engine/
- Twilio Signal 2023 presentation
- https://segment.com/docs/