

Argmax

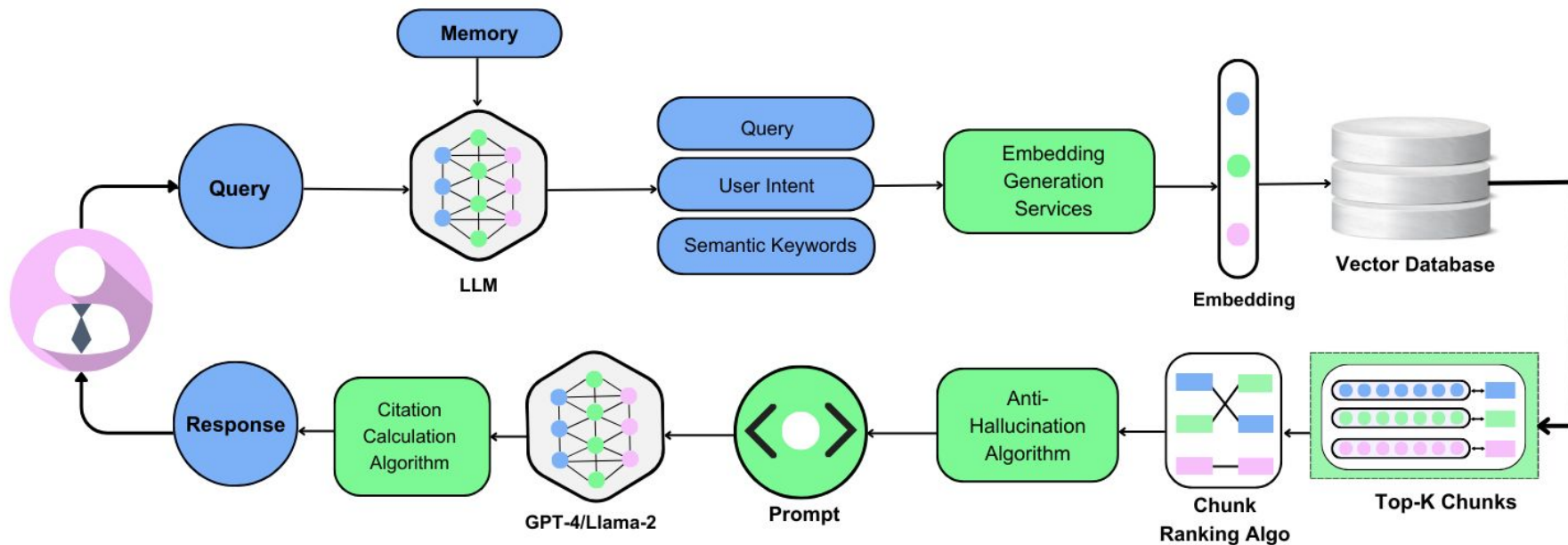


Improving Search w/Persona Based Simulation



Uri Goren
Founder & CEO @Argmax

2024: Year Of The RAG



If 2023 was about LLMs, 2024 will be about Retrieval Augmented Generation (RAG) - the easiest path to democratized Custom GPTs.

Image Source: CustomGPT.ai

Retrieval Augmented Generation in practice

Argmax



Help me improve my chatbot

Use a search engine to improve an LLM

You're using it wrong

You are correct!
Let's see how LLMs can help us improve search



Agenda

- 1 About me
- 2 Better Search results with LLM augmentation
- 3 LLMs for search evaluation
- 4 The hidden state: Simulating user intentions
- 5 Q&A

I'm human – Not AI

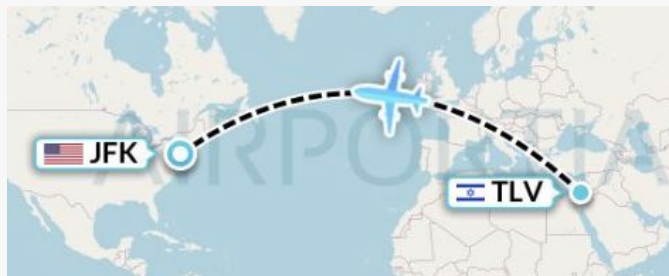
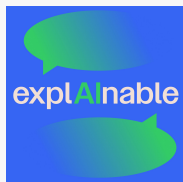
Leading Argmax

Our Expertise:

- NLP
- RecSys

ML Community

- Podcast
- Meetups
- Conferences



Argmax

Argmax



yahoo!



About Argmax

Argmax

Argmax specializes in search and recommendation system

Clients

Artlist

Sports Illustrated

Lightricks

AppsFlyer

VISA

avantis

tango

Unity

Shutterfly

onecut

Digital
Turbine

amdocs

is

SocialKit

Partners

aws

redis

databricks

Deloitte.

HYPERSPACE

AI21labs

JFrog

Argmax

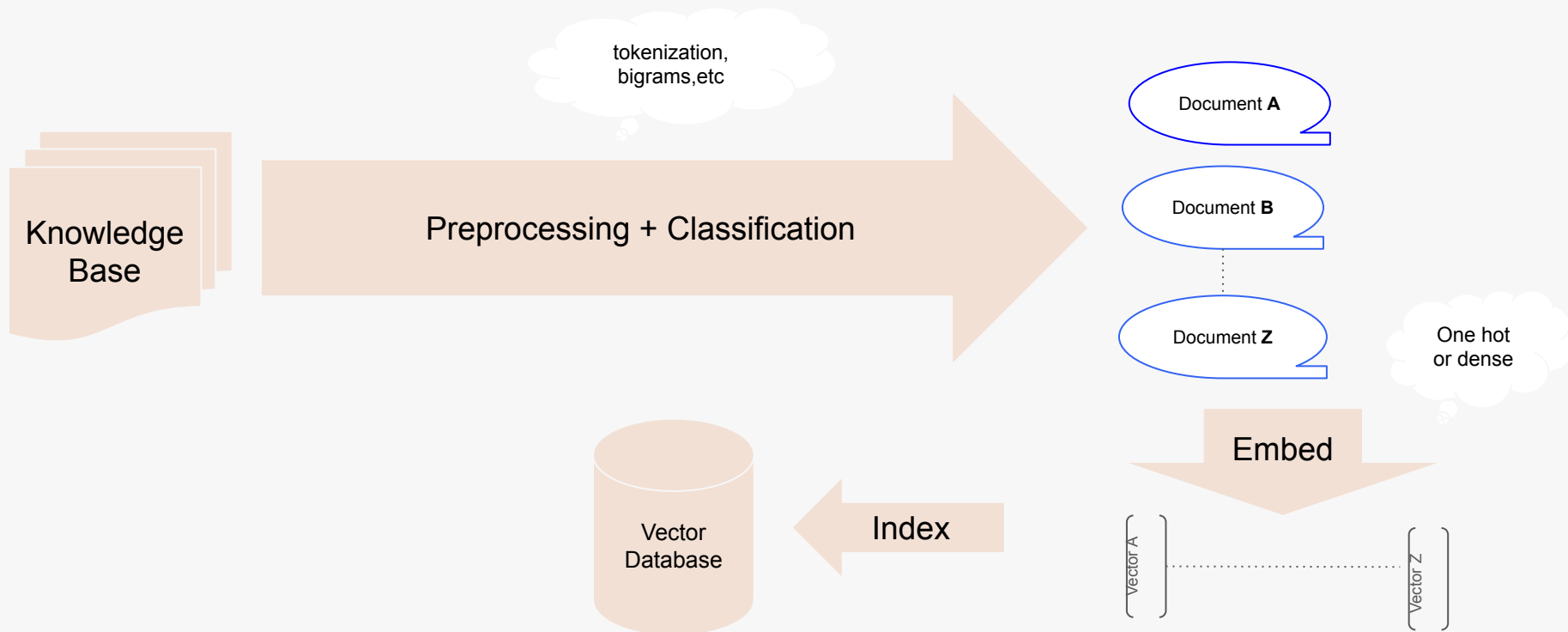
Let's talk about search



"Classical" Indexing Pipeline

Argmax

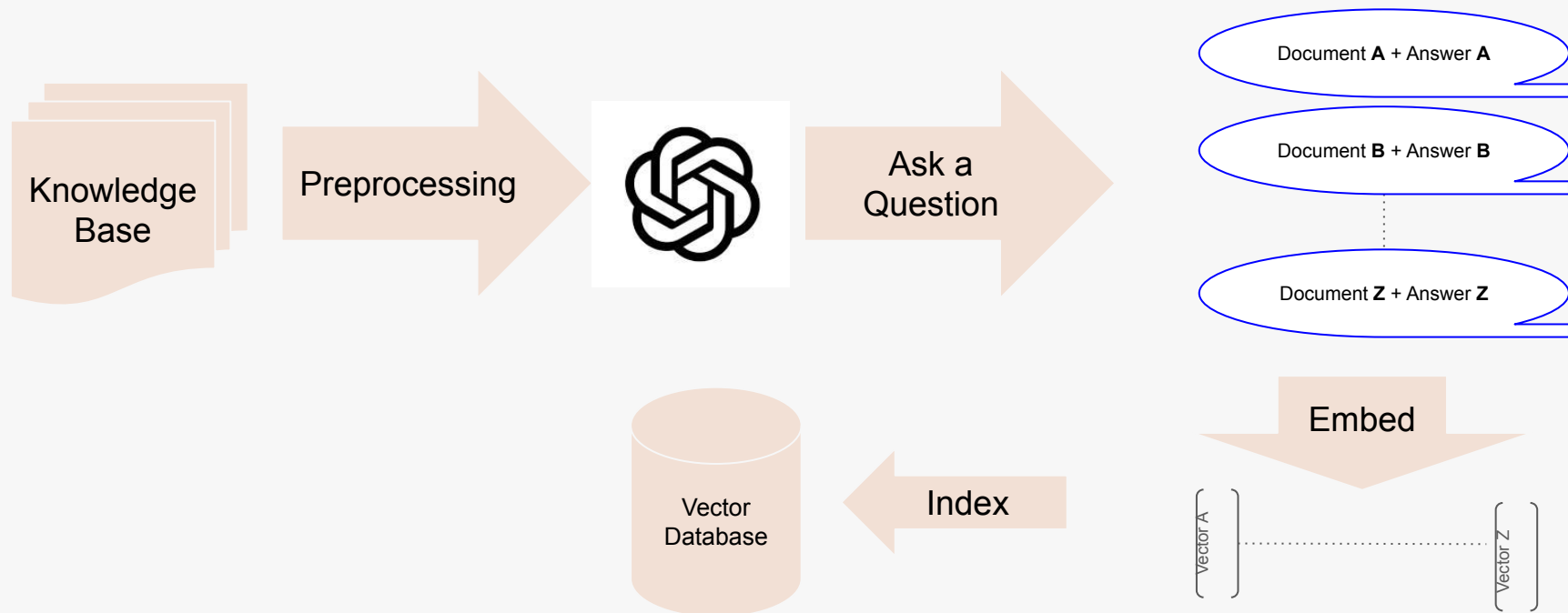
Read documents, Parse them, Index them...



LLM Enriched Indexing Pipeline

Argmax

Read documents, add meta data from an LLM, Index them



Precise Zero-Shot Dense Retrieval without Relevance Labels

Luyu Gao^{*†} Xueguang Ma^{*‡} Jimmy Lin[‡] Jamie Callan[†]

[†]Language Technologies Institute, Carnegie Mellon University

[‡]David R. Cheriton School of Computer Science, University of Waterloo

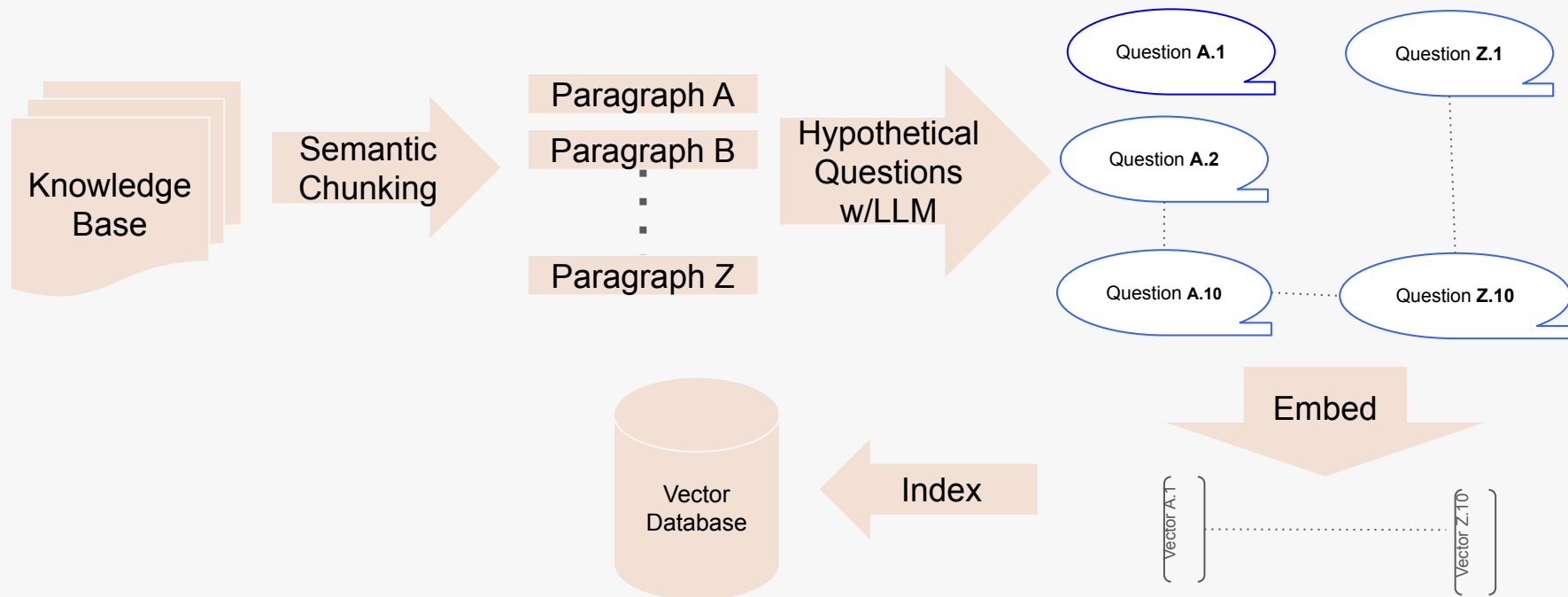
{luyug, callan}@cs.cmu.edu, {x93ma, jimmylin}@uwaterloo.ca

[Link to the full paper](#)

HyDE Indexing Pipeline

Argmax

Read documents, ask hypothetical questions, Index the questions



Doc dialect

Jaundice Symptoms

Jaundice may have no symptoms. Any signs you have may depend on how quickly the condition is getting worse. Well-known symptoms are yellowing of the skin and **jaundice eyes** (also called scleral icterus). But there are others to watch for, including:

<https://www.webmd.com/hepatitis/jaundice-why-happens-adults>



Query dialect

“Yellow eyes”

Using LLMs to bridge the language gap

Argmax

Doc dialect

Jaundice Symptoms

Jaundice may have no symptoms. Any signs you have may depend on how quickly the condition is getting worse. Well-known symptoms are yellowing of the skin and [jaundice eyes](#) (also called scleral icterus). But there are others to watch for, including:

<https://www.webmd.com/hepatitis/jaundice-why-happens-adults>

Ask a
Question



High temperature
Various prompt
Short queries

“Baby has
yellow skin”

“Why are my
eyes yellow”

“Do I have liver
issues”

Argmax

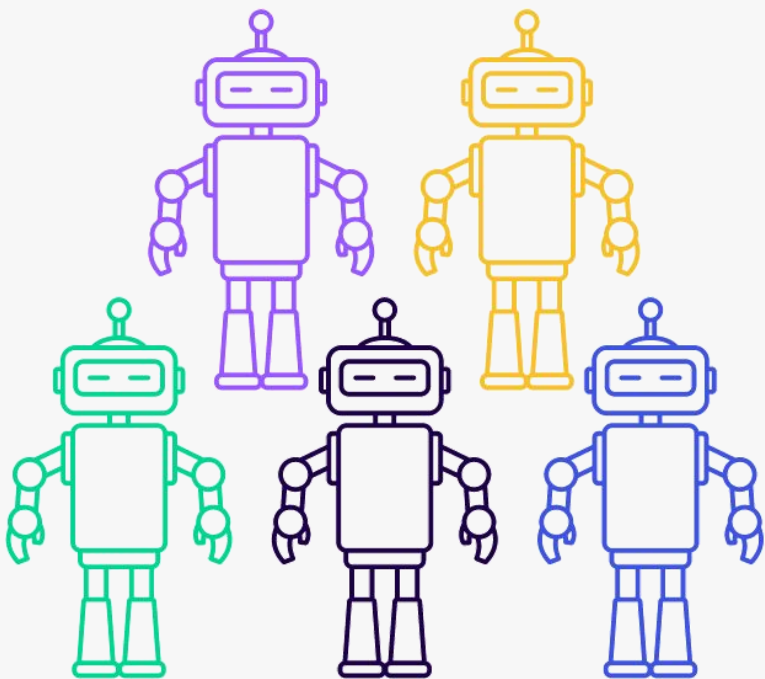
Search Sys Evaluation



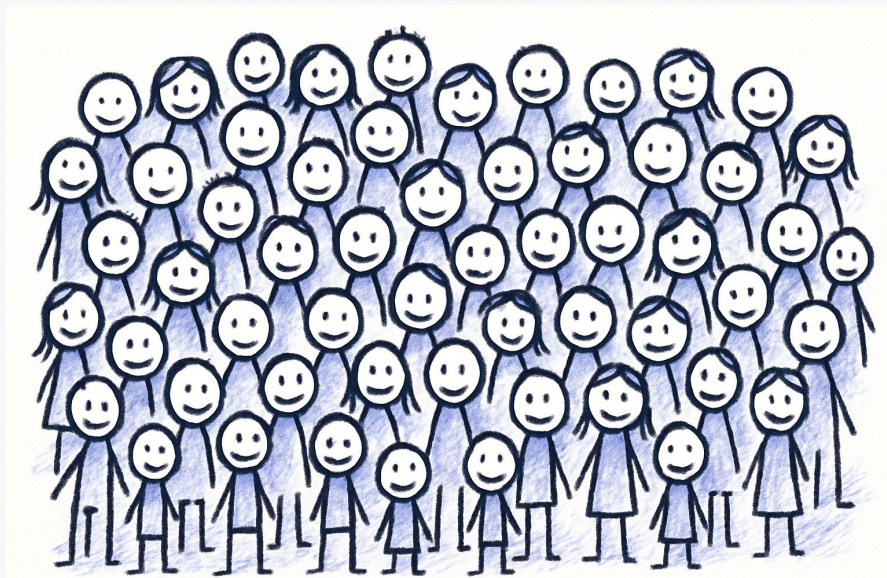
Offline vs Online testing

Argmax

Offline

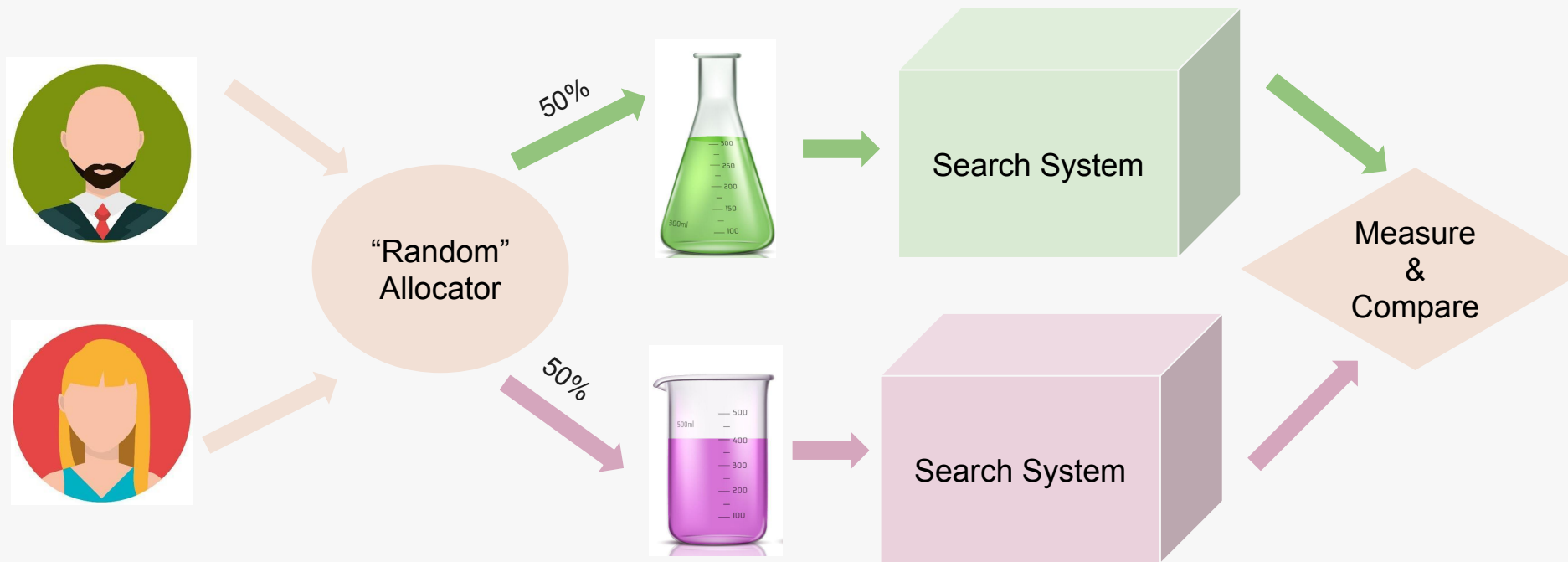


Online



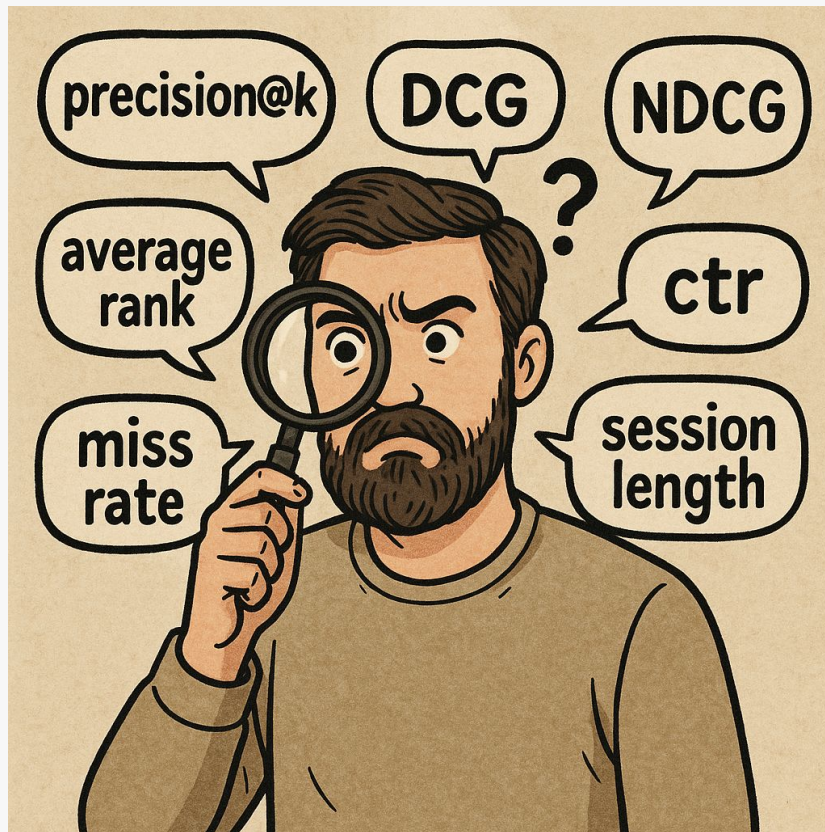
A Quick Refresher on A/B tests

Argmax



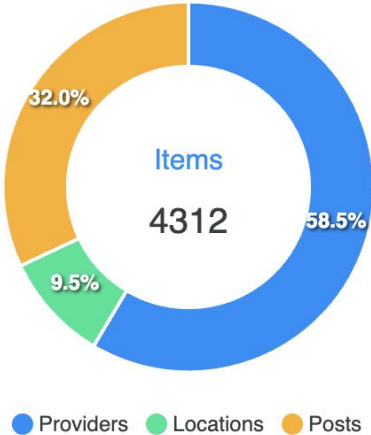
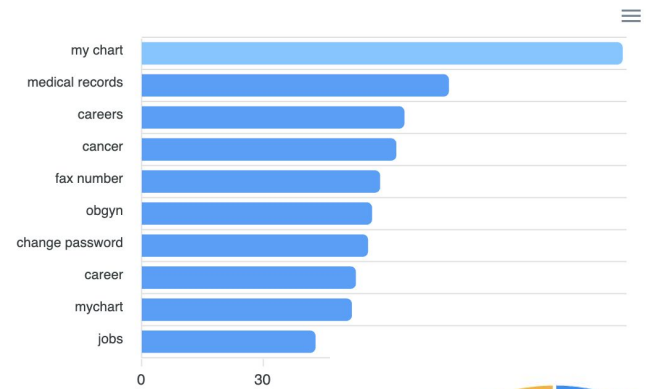
What is better?

Argmax

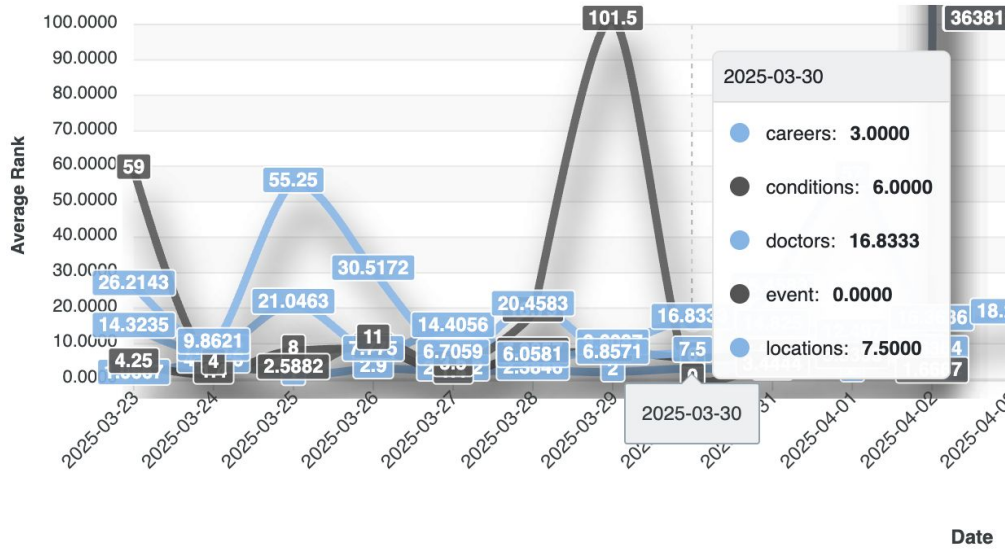


Measuring search is hard

Top Searched Queries

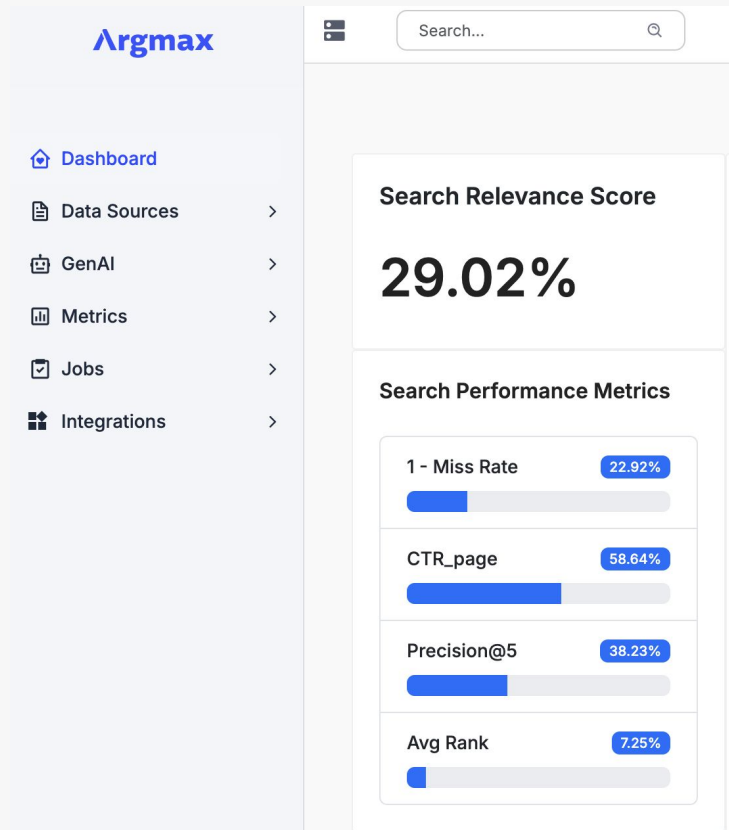


Search Metrics



Measuring search is hard

- It's hard to reason about multiple metrics
- Or combine the metrics into a single KPI



Additional challenges in online testing

Argmax

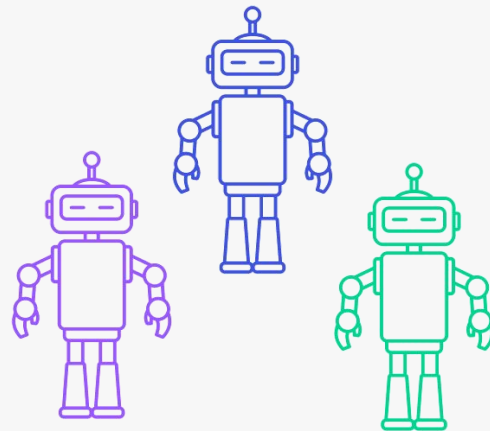
Expensive



Easy to get wrong

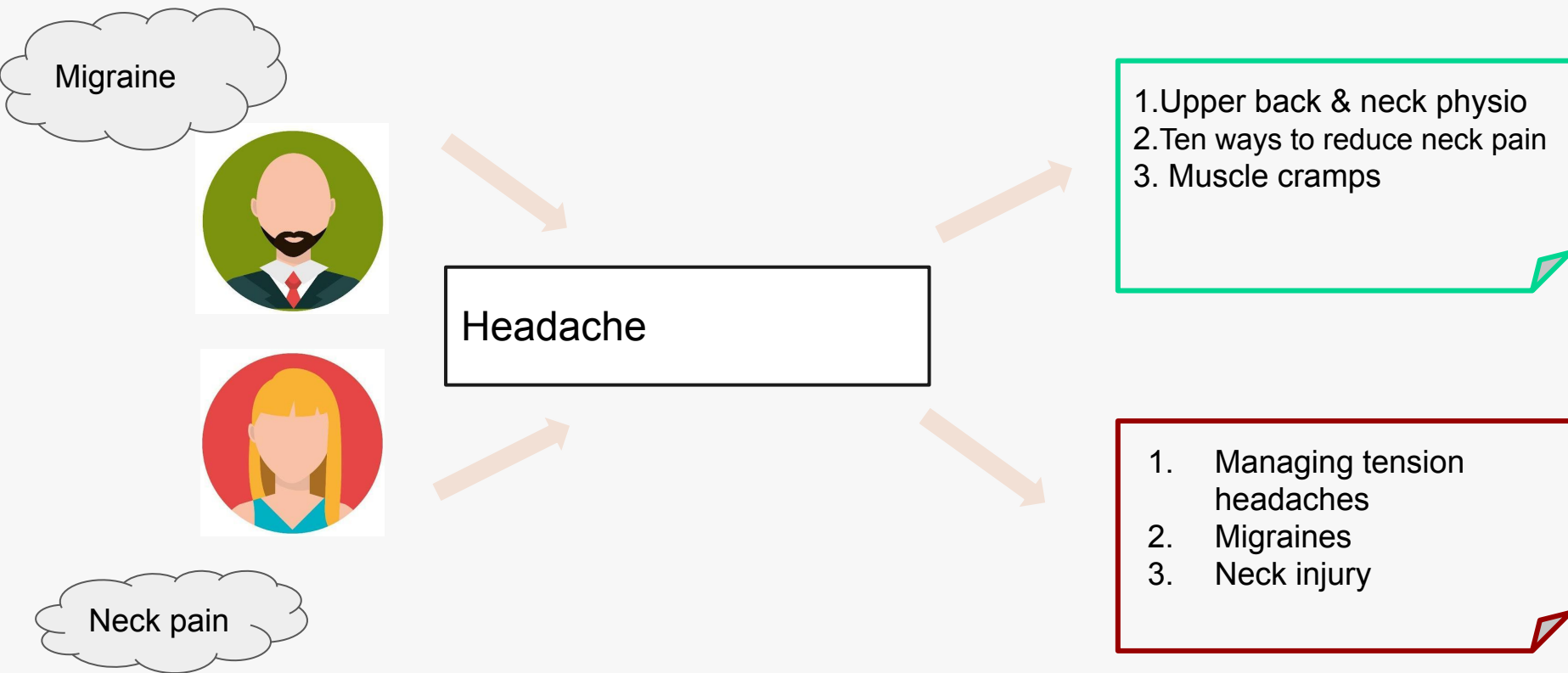


Cannot be automated



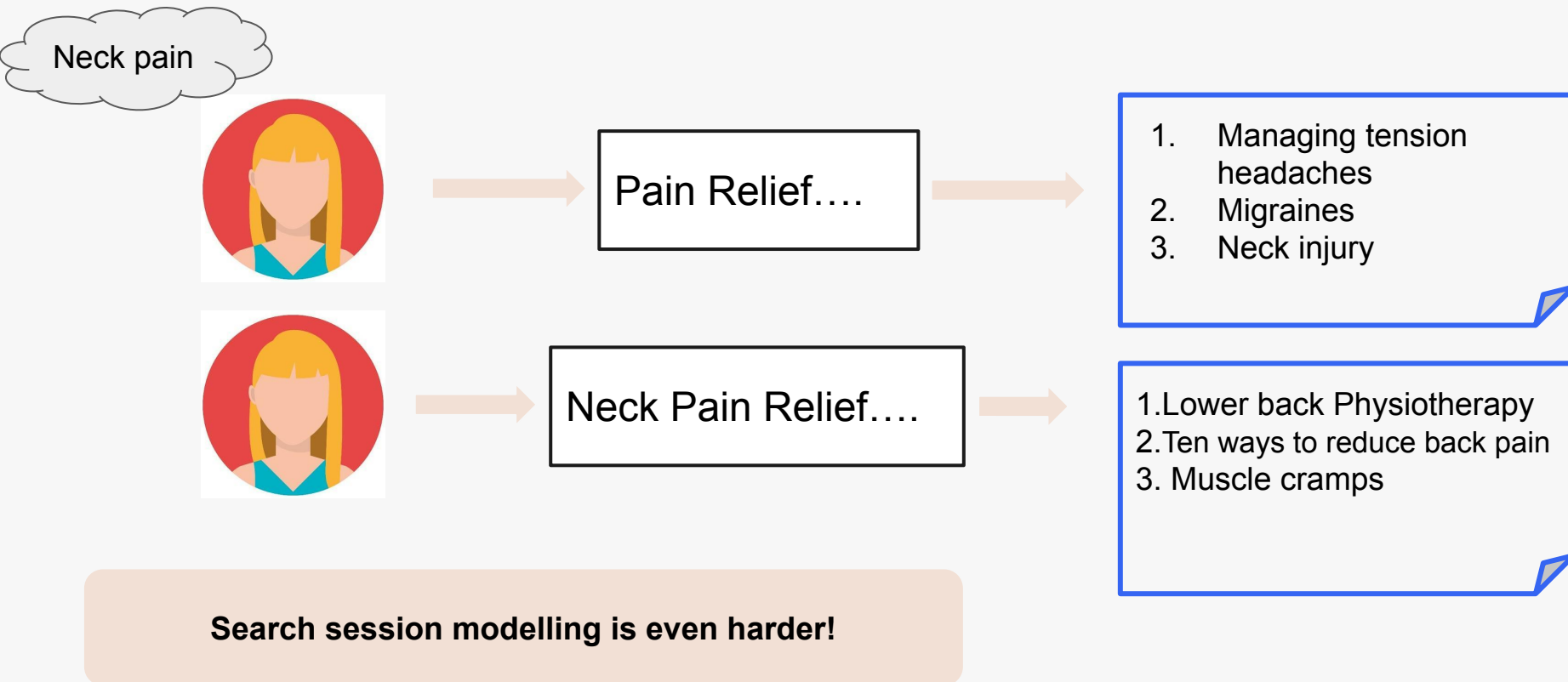
The hidden state – User intentions

Argmax



Measuring search sessions is harder

Argmax



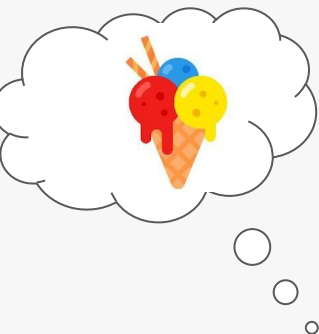
Argmax

LLMs as User Models



Can we model user intentions directly?

Argmax



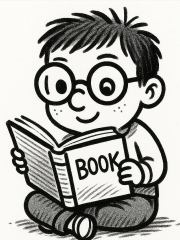
Query



Something sweet...

Catalog





WORD NERD

Question Asking Persona

“Ask a question on the main topic in the article”



VIDEOGRAPHER

Summarizing Persona

“Summarize what you learnt from the video in a sentence”



VISUAL THINKER

Visual Persona

“Describe the main image in 2-3 words”

Persona parameters

Prompt

- **Context:** Give context about what lead the user to search
"You are a 20 year old woman who clicked an ad for \$200 bike"
- **Objective:** Ask questions / Summarize / paraphrase / translate

Modality

- Article title
- Image
- Video
- Sounds

Item

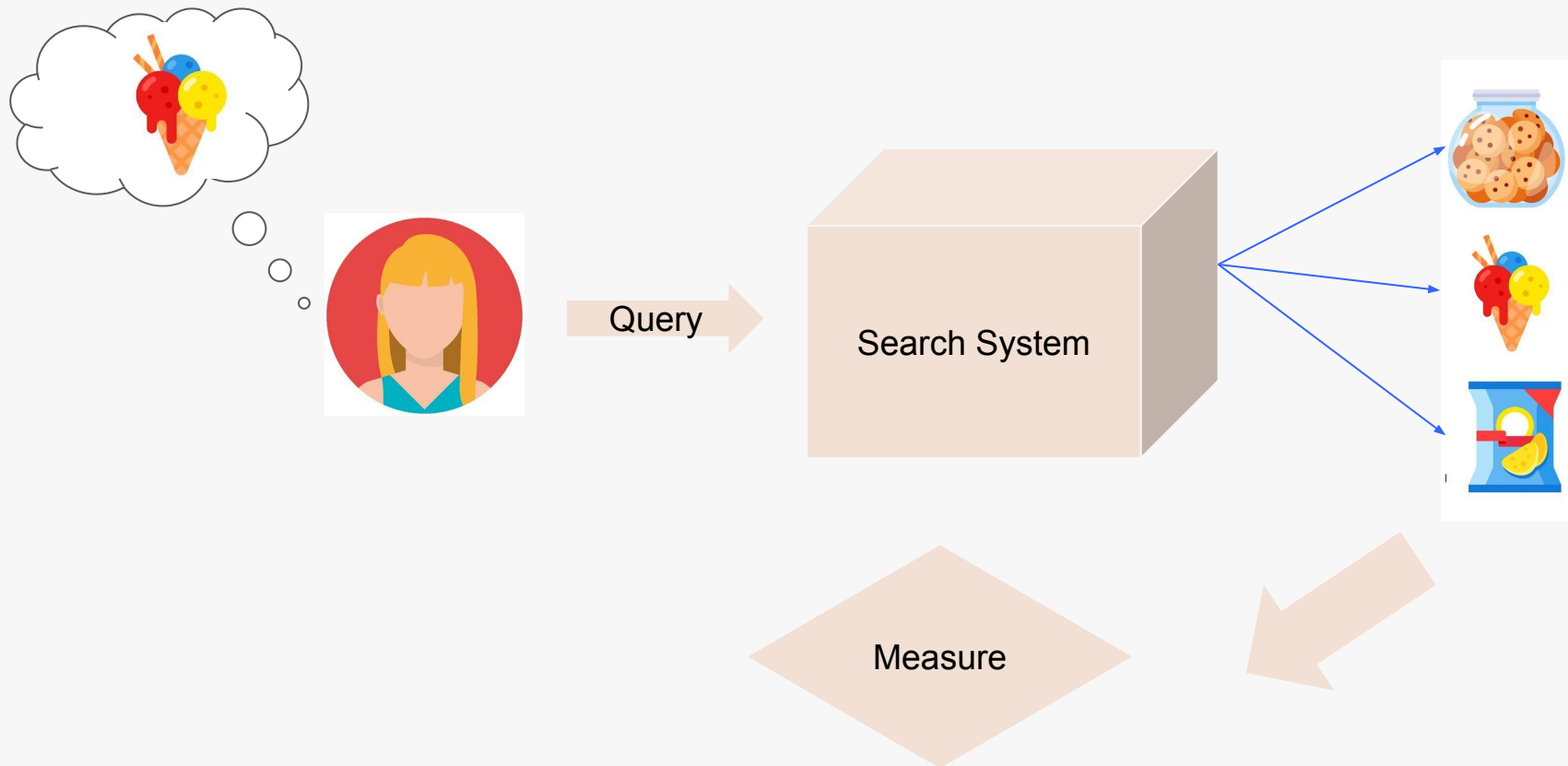
A persona may apply to a small subset of

- Geos
- Categories
- Authors
- Content Tags
- Price ranges
- Languages

A good persona definition should model a real user's intentions

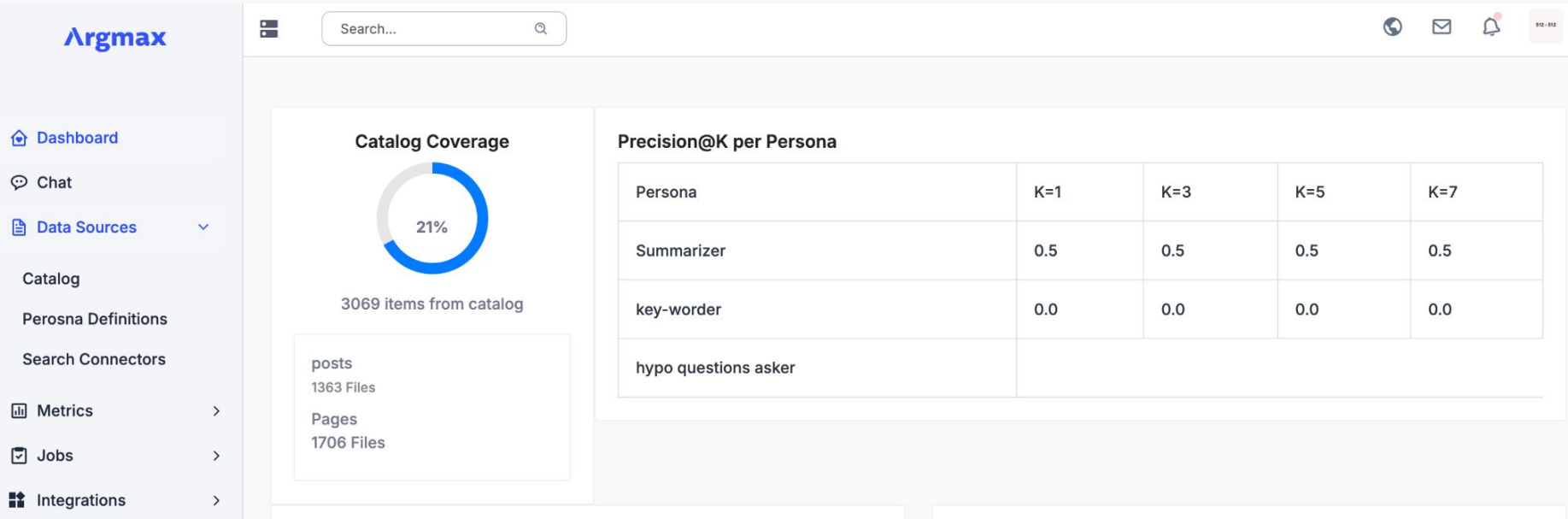
Simulated Persona Evaluation

Argmax



Persona Evaluation is easy to understand

Argmax



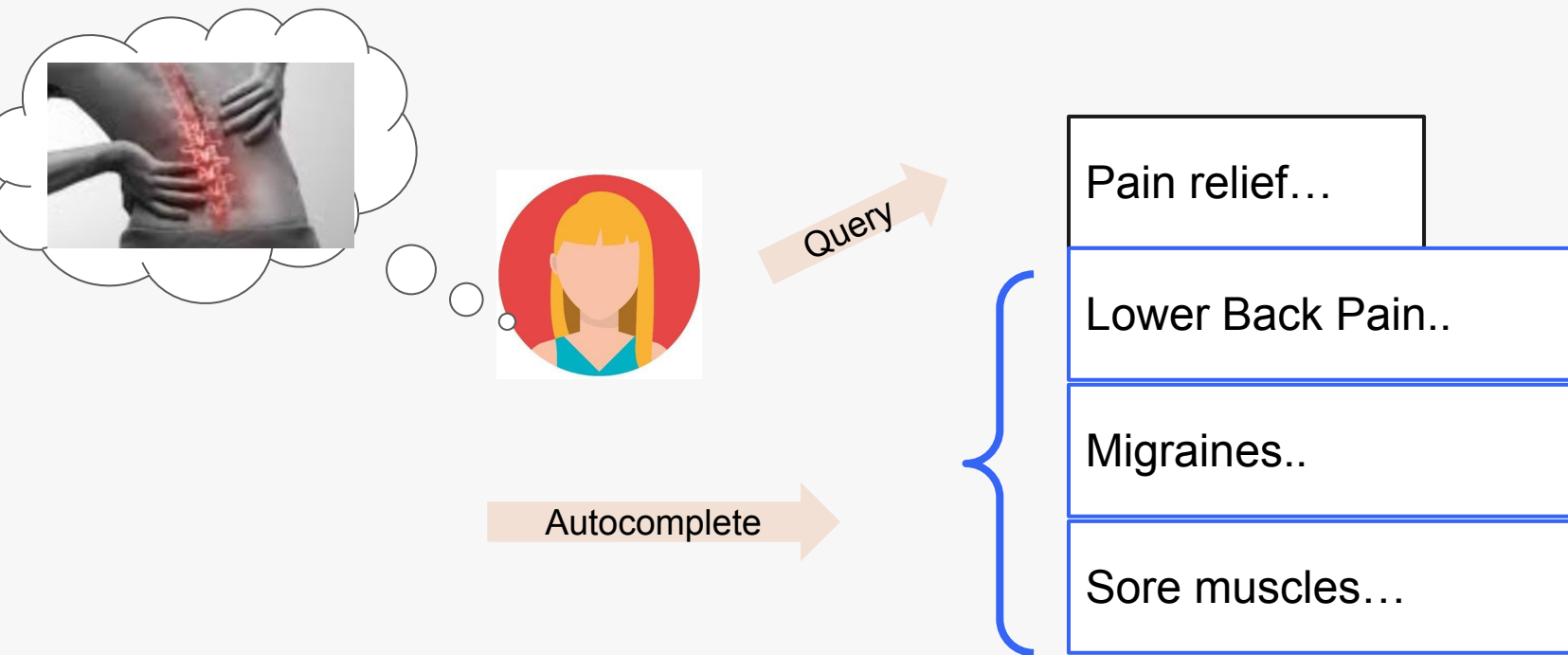
Argmax

Uncovering user Intentions with AI



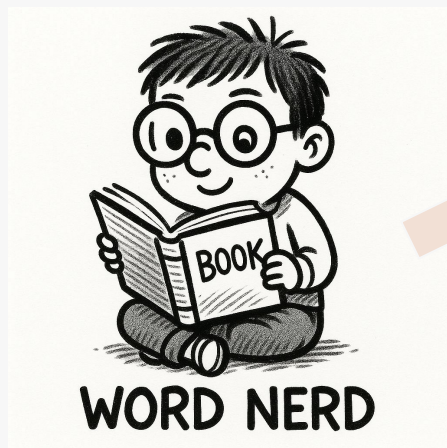
Can we model user intentions directly?

Argmax



Using personas to describe our catalog

Argmax



Lower Back Pain..

Migraines..

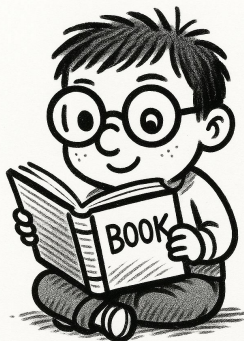


Blue liqui gel

Small green box

AI bridge multiple modalities

Argmax



WORD NERD



VISUAL THINKER



VIDEOGRAPHER

Argmax

Summary



Hypotheticals

Bridge the gap between “user queries” and “doc dialect”

Personas

Adds additional context when evaluating the search sys

Audio & Visual

Search all modalities with AI

Fully Automated

No need to annotate new items or queries.



Thank
You



Argmax