

Can we Retrieve Everything All at Once?

ARM: An Alignment-Oriented LLM-based Retrieval Method

Peter Baile Chen (MIT), Yi Zhang (AWS AI), Michael Cafarella (MIT), Dan Roth (UPenn)



An optimal retrieval system should consider both *information alignment* and *structural alignment*



What is the **highest eligible free rate for K-12 students** in the schools in the **most populous county in California**?

Step 1: Information alignment

Unduplicated Student Poverty – Free or Reduced Price Meals Data 2023–2024					
School Name	School Type	Educational Option Type	Enrollment (K-12)	Free Meal Count (K-12)	Percent (%) Eligible Free (K-12)
Envision Academy	K-12 Schools (Public)	Traditional	223	154	69.1%
Community School	Elementary Schools	Traditional	186	155	83.3%
Yu Ming Charter	Elementary Schools	Traditional	865	257	29.7%
Urban Montessori	Elementary Schools	Traditional	349	106	30.4%
Alameda County Juvenile Court School	Juvenile Court School		44	44	100.0%
Alameda County County Community	County Community	County Community	58	40	69.0%

California Counties by Population (2024)

by Kristen Carney

Updated:

The **most populous counties in California** are [Los Angeles County](#) with 9,663,345, [San Diego County](#) with 3,269,973, [Orange County](#) with 3,135,755, [Riverside County](#) at 2,492,442, and [San Bernardino County](#) with 2,195,611. You can get the most current Census population data for all [58 California counties](#), sorted by **most populous counties in California** to the least, according to the 2023 Population Estimates Program.

California Department of Education School and District Data Directory Export Educational Data Management Division December 9, 2024			
Record Type	CDS Code	School	Funding Type
District	01407900000000	No Data	No Data
District	01100170000000	No Data	No Data
School	01100170130419	Alameda County Community	No Data
School	01100170130401	Alameda County Juvenile Hall/Court	No Data
School	01100170130427	Alameda County Opportunity	No Data

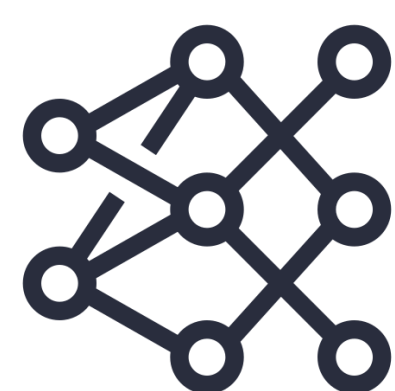
List of School Districts in California

Fiscal Year 2024–25
California Department of Education
Legend: County District School (CDS) code; N/A = Not a component of another district because district [EC Section 35515]

CDS Code	County	District
01611190000000	Alameda	Alameda Unified
01611270000000	Alameda	Albany City Unified
01611430000000	Alameda	Berkeley Unified
01611500000000	Alameda	Castro Valley Unified
01750930000000	Alameda	Dublin Unified
01611680000000	Alameda	Emery Unified

Step 2: Structural alignment

Step 3: Self- verification and aggregation



Vote for each object based on

- Relevance between user query and object content
- Compatibility within objects

Return the top-*k* objects with the highest votes

End-to-end performance

	Bird	OTT-QA	2WikiMultiHop
<i>Model for answer generation: Llama-3.1-8B-Instruct</i>			
Standard RAG	17.5	39.8	31.7
Agentic RAG	4.7	26.0	33.8
ARM_{Llama-3.1-8B}	20.6	44.0	45.0
<i>Model for answer generation: GPT-4o-mini</i>			
Standard RAG	29.8	45.9	41.7
Agentic RAG	26.3	41.2	52.9
ARM_{Llama-3.1-8B}	31.8	49.1	53.6

Monetary cost

Cost of Agentic RAG relative to ARM	Llama-3.1-8B-Instruct	GPT-4o-mini
Bird	5.36x more expensive	3.07x more expensive
OTT-QA	4.07x more expensive	3.38x more expensive
2WikiMultiHop	1.44x more expensive	1.31x more expensive