

ANB Audience Classification

**Mapping + NLP + AI
Fallback Architecture**

What is ANB Audience Classification?

ANB = Audience Needs Buckets

It is a structured marketing segmentation framework that groups users into **high-level behavioral / intent-based categories** based on what they are searching for, buying, or reading.

Rather than segmenting only by *demographics* (age/gender) or *firmographics* (industry/company size), ANB focuses on:

- What the user is trying to accomplish
- Based on their search behavior (intent)
- Enriched with context (season, demographics, etc.)

ANB buckets are **goal-driven**.

Example ANB categories:

ANB Category (Bucket)	Examples of Search Intent
Beauty & Grooming	"best anti-aging cream for men 40+"
Health & Wellness	"vitamin D supplements winter"
Home Improvement	"kitchen cabinet refacing cost"
Personal Finance	"best high yield savings account"
Tech & Gadgets	"best smartwatch for fitness tracking"
Travel & Leisure	"cheap flights for Christmas vacation"
Education / Self Improvement	"data engineering course online"

Instead of segmenting based on consumer characteristics, ANB segments based on **consumer motivation**.

Why ANB Is Powerful

Traditional segmentation answers:

"Who is this person?"

ANB answers:

"What are they trying to do right now?"

That difference is massive.

- Demographics: *static*
- ANB search intent: changes weekly or even daily

Intent → Action → Conversion

This is the **closest proxy to real buying intent** without needing cookies or personal identity.

🔥 Where ANB Classification Is Used

Use Case	How ANB adds value
Digital advertising	Helps platforms automatically route prospects into the right ad campaign
Search optimization (SEO/SEM)	Predicts commercial vs informational intent
Personalization	Shows the right offer/product to the right user
Predictive modeling	Enables look-alike audiences based on behavior

In your project, ANB classification is applied to:

- search content + demographics + seasonal context
- enriched data → segmentation → activation

ANB Audience Classification – Mapping + NLP + AI Fallback

This project converts raw user search behavior into standardized ANB (Audience Needs Bucket) marketing segments using a deterministic-first workflow that minimizes LLM usage, improves latency, and enforces governance.

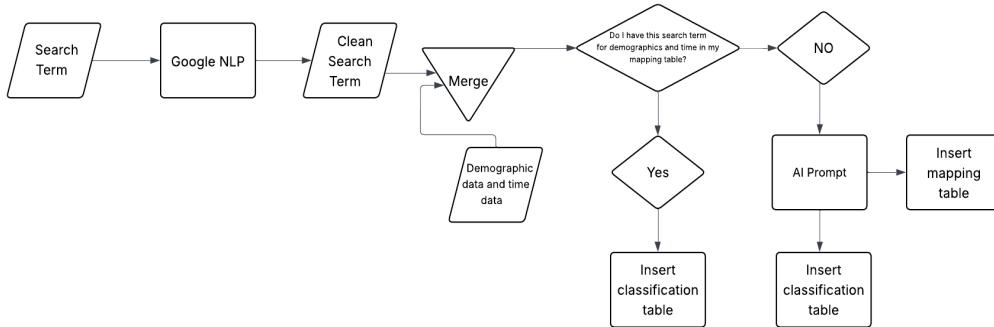
Architecture Overview

Bronze → NLP Enrichment → Deterministic Mapping → AI Fallback → Silver Segments → Feedback to Mapping

End-to-End Flow

1. Ingest raw search events into BigQuery Bronze (search_term, age_range, gender, season, region, event_ts).

2. 2. Enrich text using Google Cloud Natural Language API (Entities, Categories, Syntax, Sentiment, Salience Scores).
3. 3. Normalize search signals into features (lemmatization, entity type extraction, noun chunking, salience filtering).
4. 4. Deterministic mapping lookup: match canonical term or aliases from mapping.search_term_to_anb.
5. - If found → assign ANB, write to Silver layer.
6. - If not found → route to Vertex AI Gemini batch classification.
7. 5. Apply Quality Control (confidence threshold, human review sampling).
8. 6. Approved AI-labeled terms are inserted into mapping with lineage.
9. 7. Next identical search resolves 100% locally with no LLM call.
10. 8. Serve ANB segments to downstream marketing, analytics, and activation systems.



Natural Language Processing (NLP) Enrichment Layer

Google Cloud NLP transforms raw search text into structured signals used for classification and mapping (before AI fallback).

- Entity Extraction: detects real-world objects (e.g., 'rolex', 'mortgage refinance', 'tesla model y').
- Entity Types: PERSON, LOCATION, ORGANIZATION, CONSUMER_GOOD, EVENT, WORK_OF_ART, OTHER.
- Category Classification: assigns IAB-taxonomy-adjacent topics with confidence score.
- Part-of-Speech + Lemmatization: normalizes 'running shoes' → 'run shoe'.

- Syntax Parsing: identifies noun phrases likely to represent intent units.
- Sentiment + Magnitude: optional signal (high urgency vs neutral research).
- Salience Score: filters dominant versus incidental entities.

Deterministic Mapping Table (Governed Knowledge Base)

Schema:

```
CREATE TABLE mapping.search_term_to_anb (
    canonical_term STRING,
    term_aliases ARRAY<STRING>,
    anb_segment STRING,
    source STRING,
    confidence FLOAT,
    created_by STRING,
    created_at TIMESTAMP
);
```

Vertex AI Fallback (Only When No Mapping Exists)

Airflow Batch Classification Operator:

```
create_batch_prediction_job = CreateBatchPredictionJobOperator(
    task_id="create_batch_prediction_job",
    job_display_name="nav_unmapped_searches_batch_prediction",
    model_name="publishers/google/models/gemini-1.5-pro-002",
    bigquery_source="bq://project.dataset.staging_unmapped_searches",
    bigquery_destination_prefix="bq://project.dataset.staging_unmapped_ai_responses",
    region="us-east1",
    project_id="project",
    gcp_conn_id="cda_project"
)
```

AI Prompt Used for Classification

You are a marketing taxonomy classifier.

Input: search term + age range + gender + season + region + NLP enrichment.

Task: assign exactly one primary ANB category and optional secondary tag.

Return JSON:

```
{
  "canonical_term": "...",
  "anb_category": "...",
  "confidence": 0.0-1.0,
  "justification": "...",
  "related_entities": []
}
```

Rules:

- If intent contains purchase signals → prioritize Commerce buckets.
- If intent expresses lifestyle, identity, aspiration → map to Identity/Interest buckets.
- No open-ended categories allowed.
- Confidence < 0.6 requires HUMAN REVIEW flag.

Benefits Architecture

- 90%+ searches resolved via deterministic mapping (not LLM) → cost efficient.
- Continuous learning through self-healing mapping expansion.
- Full label lineage and auditability.
- Low-latency real-time classification possible once mapped.
- NLP ensures structured intent signals rather than blind text embedding reliance.
- Governance layer prevents LLM label drift.