BlackScan API Cost Analysis

Report Date: February 07, 2026

1. API Services Overview

BlackScan uses two primary external APIs: OpenAI for product label analysis during scans, and Typesense for product search and discovery. This report models costs based on an average user who engages in 3 near-full sessions per day.

OpenAI - Two-Tier Scan Architecture

The app uses a hybrid approach that tries the cheaper model first and falls back to the more expensive vision model only when needed:

Path	Model	When Used	Est. Tokens	Cost / Call
Primary	gpt-4o-mini (text)	OCR succeeds, good quality	~800-1,400	~$0.0003
Fallback	gpt-4o (vision)	OCR fails or confidence < 0.7	~1,400	~$0.006

Blended average cost per scan (assuming ~70% primary path): ~$0.002 per scan.

Typesense - Search & Discovery

Typesense Cloud is priced by cluster size (RAM + CPU), not per query. With ~6,400 products (~few MB), the dataset fits on the smallest tier. Query volume determines whether the cluster needs to scale.

2. Average Session Breakdown

One near-full user session consists of searching, browsing a category, viewing product details, scanning a product, and visiting a company page.

User Action	Typesense Queries	OpenAI Calls
App launch (featured cache)	~0.3 (cached after 1st)	0
Search (typing ~12 chars, debounced)	~6 autocomplete + 1 committed	0
Category browse + 1 Load More	2	0
View 3 product details	~5 (similar products)	0
Scan 1 product	~2 (match passes)	1
Visit 1 company page	1	0
SESSION TOTAL	~17 queries	~1 call

3. Per-User Monthly Cost

Based on 3 sessions/day x 30 days/month:

Metric	Per User / Month
Typesense queries	~1,530
OpenAI scans	~90

OpenAI cost	~$0.19

4. Scaled Monthly Cost Estimates

The following table projects total monthly costs across different user scales. Typesense costs are estimated based on Typesense Cloud cluster tiers appropriate for the query volume.

Active Users	Scans / Mo	OpenAI Cost	Typesense Queries	Est. Typesense
100	9,000	~$19	~153K	~$25
500	45,000	~$95	~765K	~$40
1,000	90,000	~$189	~1.5M	~$50
5,000	450,000	~$945	~7.7M	~$80
10,000	900,000	~$1,890	~15.3M	~$120
50,000	4,500,000	~$9,450	~76.5M	~$300
100,000	9,000,000	~$18,900	~153M	~$600

Combined Total Estimates

Active Users	OpenAI	Typesense	TOTAL / Month	Cost / User
100	$19	~$25	~$44	$0.44
500	$95	~$40	~$135	$0.27
1,000	$189	~$50	~$239	$0.24
5,000	$945	~$80	~$1,025	$0.21
10,000	$1,890	~$120	~$2,010	$0.20
50,000	$9,450	~$300	~$9,750	$0.20
100,000	$18,900	~$600	~$19,500	$0.20

5. Typesense Call Sites (Detailed)

All Typesense queries go through TypesenseClient.swift. Below is every call site in the app with its trigger and volume characteristics.

Flow	perPage	Trigger	Frequency
Search dropdown (autocomplete)	20	Per keystroke (debounced 0.5s)	High
Search grid (committed)	50	User submits search	Medium
Category browse (initial)	50	Tap category chip	Medium
Category browse (pagination)	50	Tap "Load More"	Medium
Scan matches (multi-pass)	150 total	Per scan (1-3 API calls)	Medium
Similar products	30 + 20	Product detail view opens	Medium
Featured products cache	50	App launch (cached)	Low
All featured products	200	"See All" view	Low
Company products	50	Company view opens	Low
Company image fetch	5	Saved company needs image	Low

6. OpenAI Call Flow (Detailed)

The HybridScanService orchestrates a two-tier approach to minimize vision API usage:

- Step 1: OCR extracts text from the camera frame (on-device, free).
- Step 2: OCR text is sent to gpt-4o-mini for classification (~$0.0003).
- Step 3 (conditional): If confidence < 0.7 or OCR failed, the image is sent to gpt-4o vision (~$0.006).
- Step 4: Classification results are used to query Typesense for matching products.

Per-scan cost scenarios:

Scenario	API Calls	Cost
Best case (OCR + mini succeeds)	1x gpt-4o-mini	~$0.0003
Fallback (mini low confidence)	1x gpt-4o-mini + 1x gpt-4o	~$0.0063
OCR failure (direct vision)	1x gpt-4o	~$0.006

7. Cost Safeguards in Place

- 10 scans/day per-user rate limit (UserDefaults + Keychain). Caps worst-case per-user OpenAI spend at ~$0.60/month.
- Search debouncing (0.5s) reduces autocomplete Typesense queries during typing.
- ProductCacheManager caches featured products at app launch, avoiding refetches on tab switches.
- ImageCache (NSCache, 100 images / 50MB) prevents re-downloading product images from retailer CDNs.
- NetworkSecurity.withRetry (max 2 attempts) with exponential backoff handles transient failures without runaway retries.

8. Key Takeaways

- OpenAI dominates costs. OpenAI is 80-95% of the API bill at every scale. Typesense cluster costs are relatively flat.
- Hybrid scan architecture is effective. The hybrid OCR-first approach saves ~3x vs. sending every scan to gpt-4o vision.
- Strong unit economics. Per-user cost drops from ~$0.44 at 100 users to ~$0.20 at scale due to the fixed Typesense cluster cost being amortized.
- Biggest optimization opportunity: scan caching. Caching scan results for popular/repeated products could save 15-25% on OpenAI costs. At 10K users, that is ~$380-470/month.

Assumptions

- 3 near-full sessions per user per day, 30 days/month
- 1 scan per session (3 scans/day, well within 10/day limit)
- 70% of scans resolved by gpt-4o-mini, 30% fall back to gpt-4o vision
- OpenAI pricing: gpt-4o-mini input $0.15/1M tokens, output $0.60/1M; gpt-4o input $2.50/1M tokens, output $10.00/1M
- Typesense Cloud estimates based on published tier pricing for small datasets (<10MB)
- Retry overhead not included (typically <5% of calls)
- Image CDN costs excluded (hosted by retailers, free to consume)