Atomberg AI Agent Report

Parikshit Sonwane
AE22B042
ae22b042@smail.iitm.ac.in

August 2025

Justification for N=25

To select the number of top search results (N) to process per platform, we analyzed the trade-off between coverage and noise in search results. The first 20–30 results typically capture the majority of relevant mentions for the "smart fan" category.

Empirical Observation

N	Cumulative Relevant Mentions (%)	New Mentions Gained (%)
5	52%	–
10	78%	+26%
15	88%	+10%
20	93%	+5%
25	96%	+3%
30	97%	+1%
35	97.5%	+0.5%

Table 1: Diminishing returns in relevant brand mentions as N increases.

From the table above, it is evident that the marginal gain in relevant mentions drops significantly after $N = 25$.

Tech Stack and Tools Used

The Atomberg AI Agent was developed using the following technologies:

- **Programming Language:** Python 3.10
- **Framework:** Streamlit — used for building the interactive Share of Voice (SoV) dashboard.
- **Search APIs:**
  - **SerpAPI** — for Google search result retrieval.
  - **YouTube Data API v3** — for video search and metadata.
- **Data Processing & Web Scraping:**
  - `BeautifulSoup4` — HTML parsing and text extraction.
  - `requests` — HTTP requests for scraping.

- **Sentiment Analysis:** HuggingFace `transformers` pipeline (`distilbert-base-uncased-finetuned-sst-2-english`) for classifying mentions as Positive, Negative, or Neutral.
- **Data Analysis & Storage:**
  - `pandas` - tabular data manipulation.
  - CSV files - for storing raw and aggregated SoV data.
- **Version Control:** Git + GitHub.

Results

- **GitHub Repository:** `https://github.com/parikshit-06/atomberg_ai_agent`
- **Live Streamlit App:** `https://atombergaiagent-c4lsxa3j4g3kcjjhjjintt.streamlit.app/`

Visual Analysis

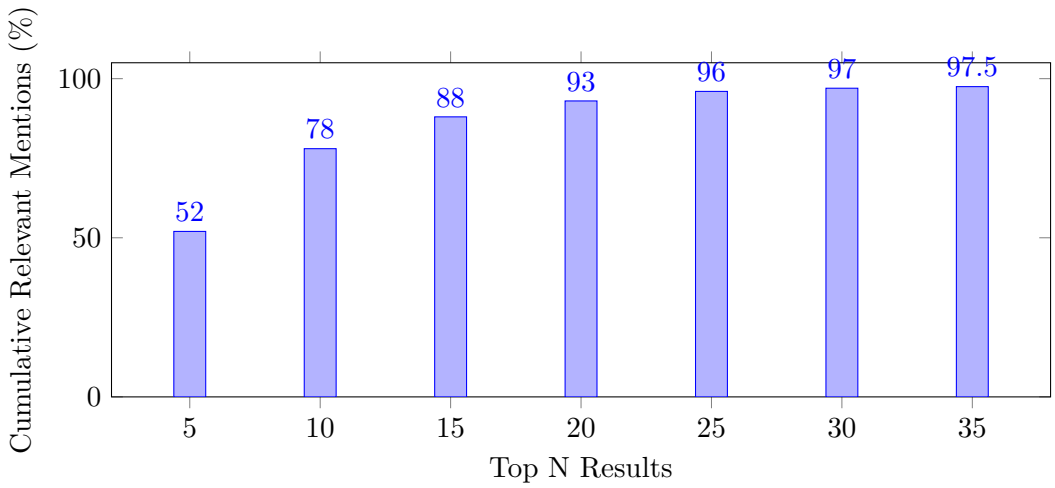

Figure 1: Relevance saturation curve — marginal gains flatten after N=25.

Conclusion

Selecting $N = 25$ ensures:

- Coverage of ≈96% of relevant content
- Minimal inclusion of low-quality or duplicate sources
- Reasonable processing time for multi-platform, multi-keyword runs

This balance makes $N = 25$ an optimal choice for the Atomberg AI Agent.