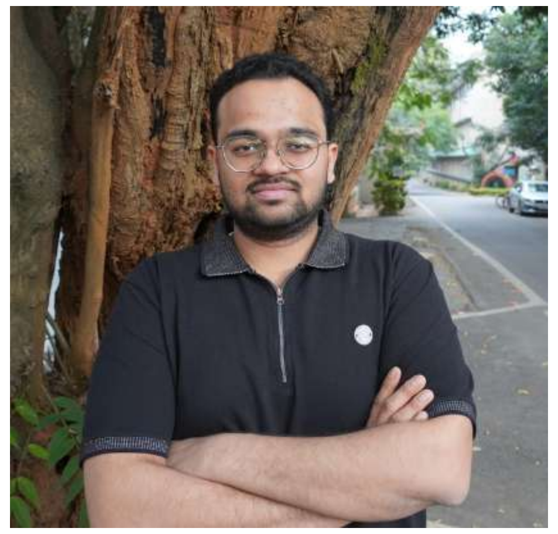
Diversity in Similarity Search

A Math-Econ Perspective

Accepted in ICLR 2026

Nirjhar Das
IISc Bangalore

Siddharth Barman
IISc Bangalore



Shivam Gupta
IIT Bhilai

Kiran Shiragur
Microsoft Research

Diversity in Search

Visualisation on Amazon Dataset

Diversity in Search

Visualisation on Amazon Dataset

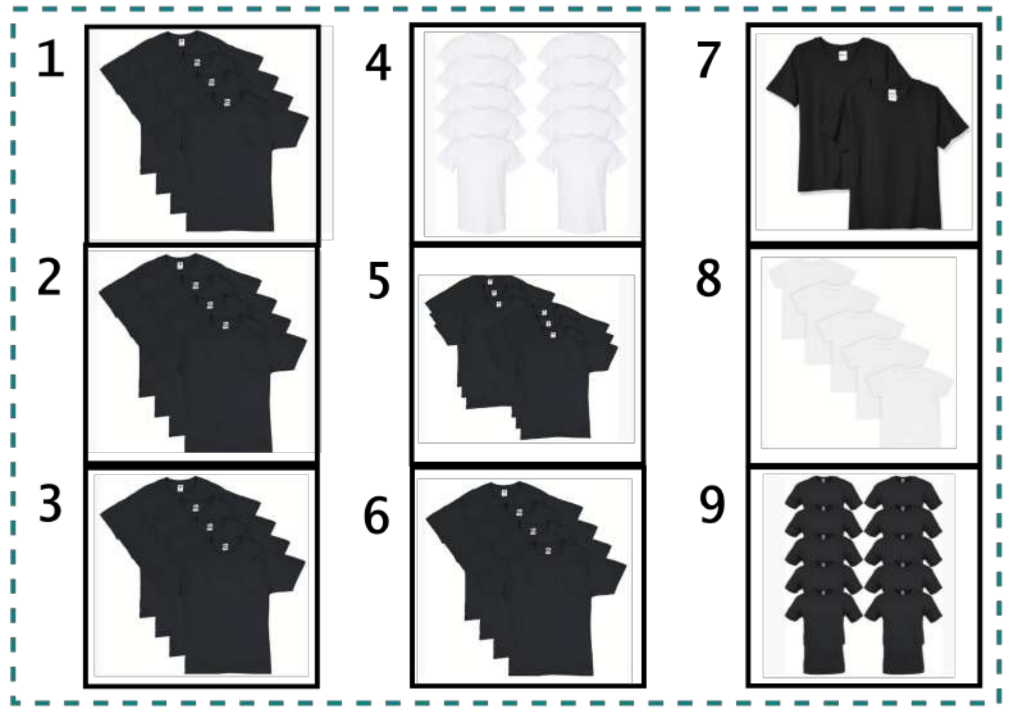
Query = “Shirts”

Diversity in Search

Visualisation on Amazon Dataset

Standard Search

Query = “Shirts”



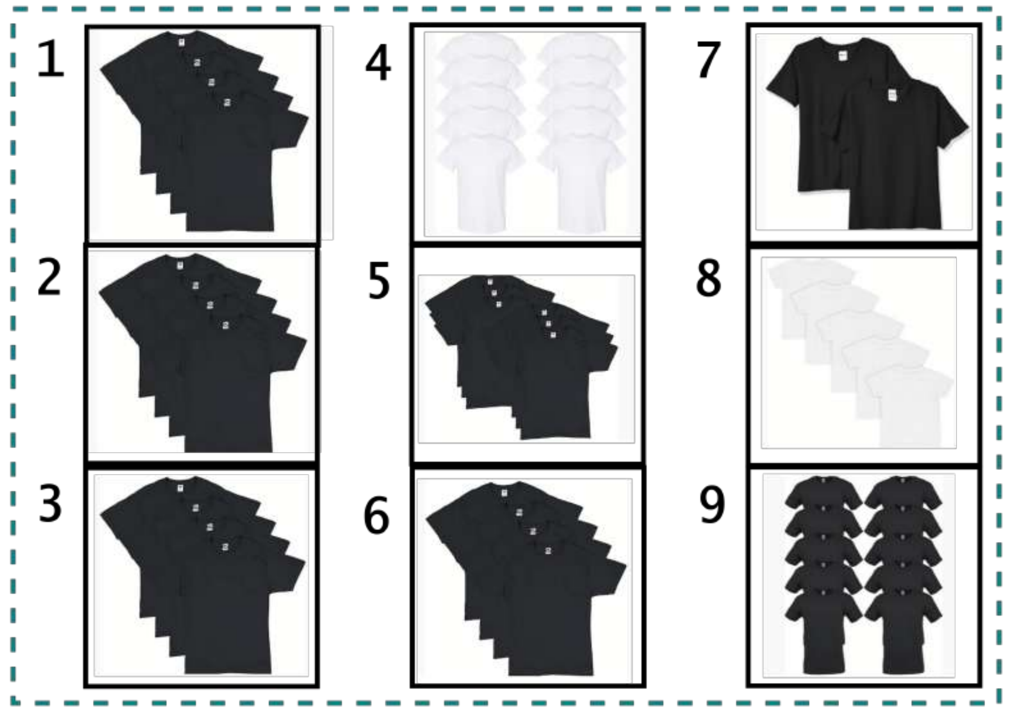
Diversity in Search

Visualisation on Amazon Dataset

Standard Search

Constrained Search

Query = “Shirts”



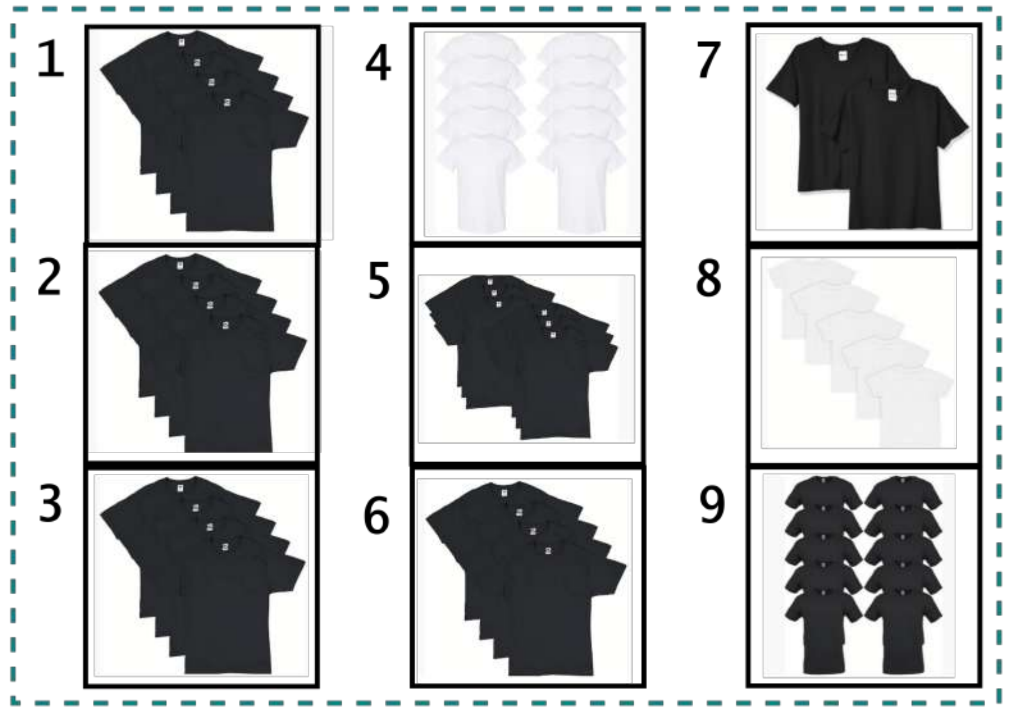
Diversity in Search

Visualisation on Amazon Dataset

Standard Search

Constrained Search

Query = "Shirts"



Query = "Blue Shirts"

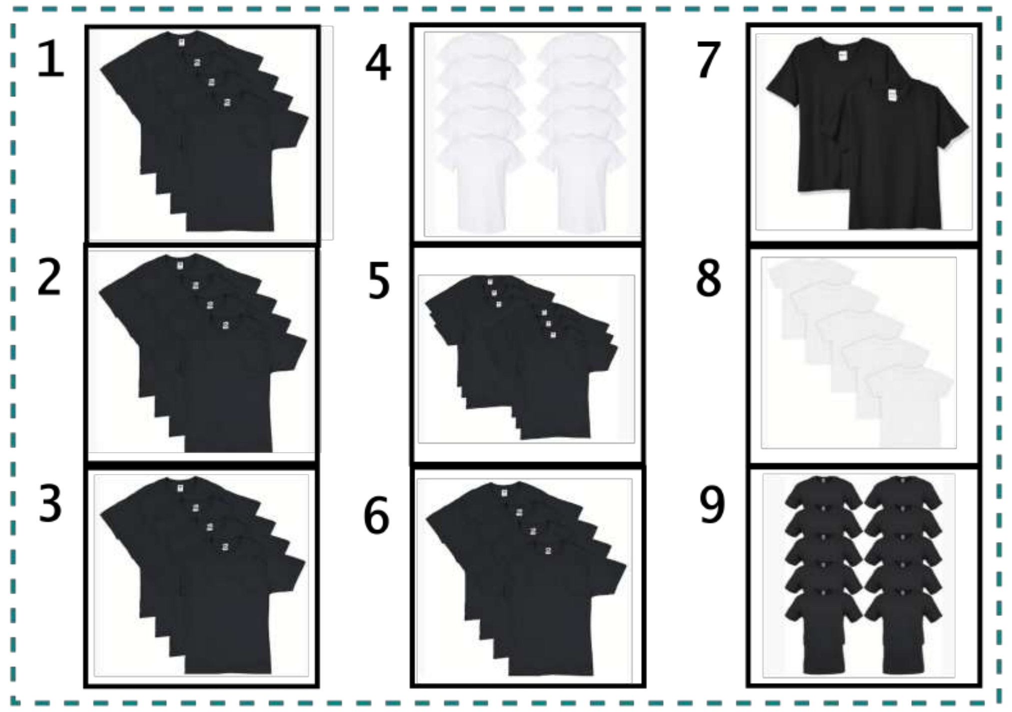
Diversity in Search

Visualisation on Amazon Dataset

Standard Search

Constrained Search

Query = “Shirts”



Query = “Blue Shirts”

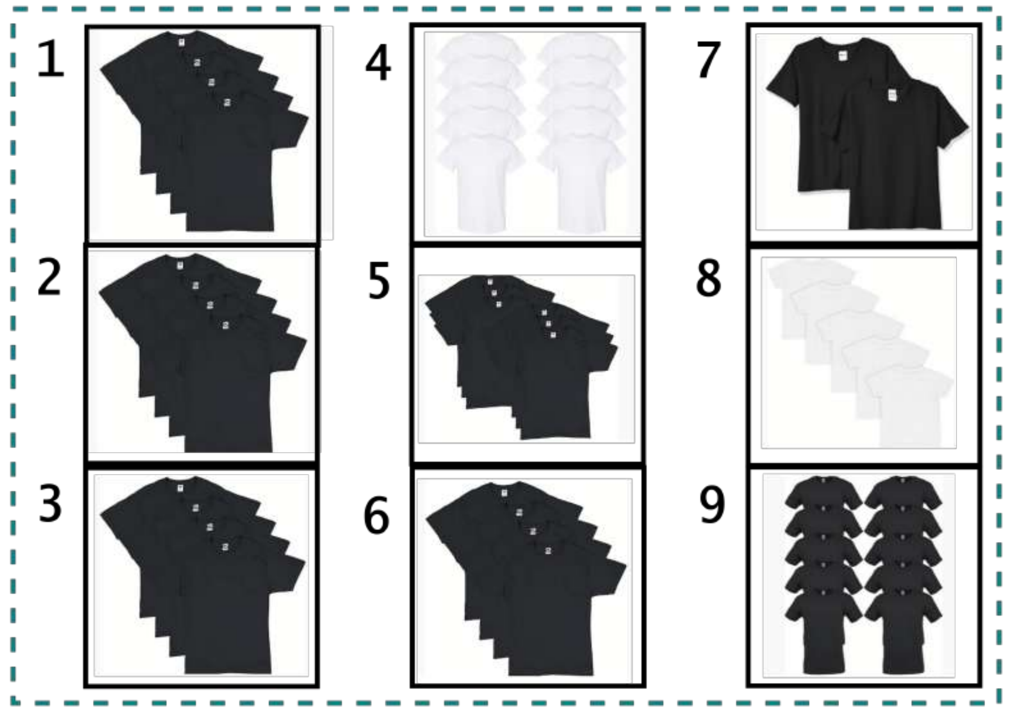
Diversity in Search

Visualisation on Amazon Dataset

Standard Search

Constrained Search

Query = “Shirts”



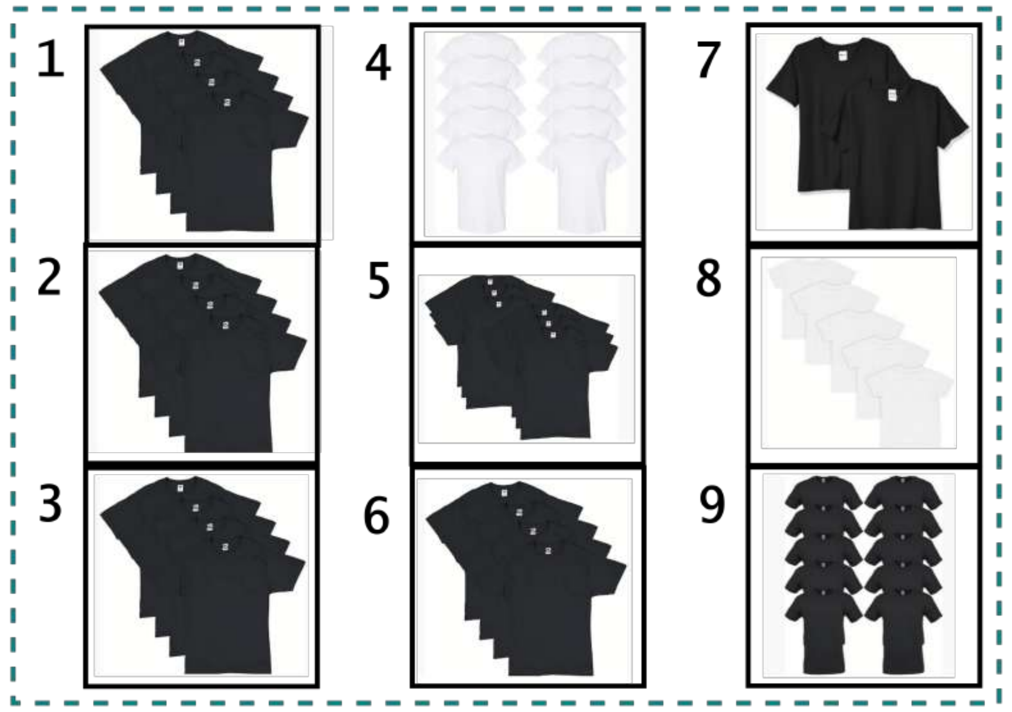
Query = “Blue Shirts”

Diversity in Search

Visualisation on Amazon Dataset

Standard Search | Constrained Search | Nash Search (Ours)

Query = “Shirts”



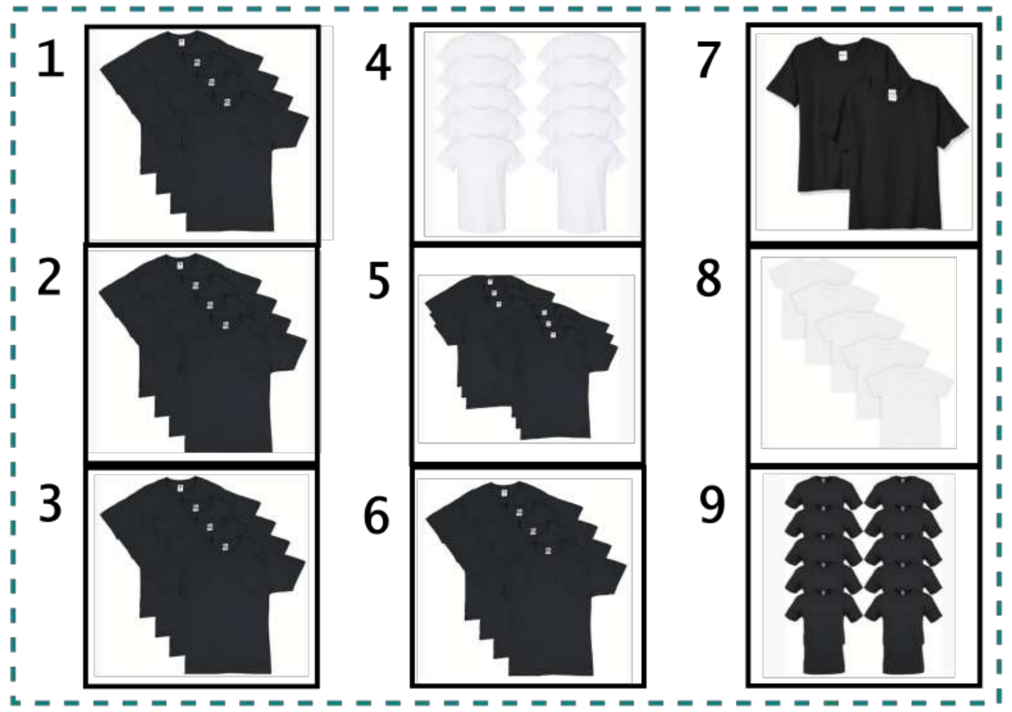
Query = “Blue Shirts”

Diversity in Search

Visualisation on Amazon Dataset

Standard Search | Constrained Search | Nash Search (Ours)

Query = “Shirts”



Query = “Blue Shirts”

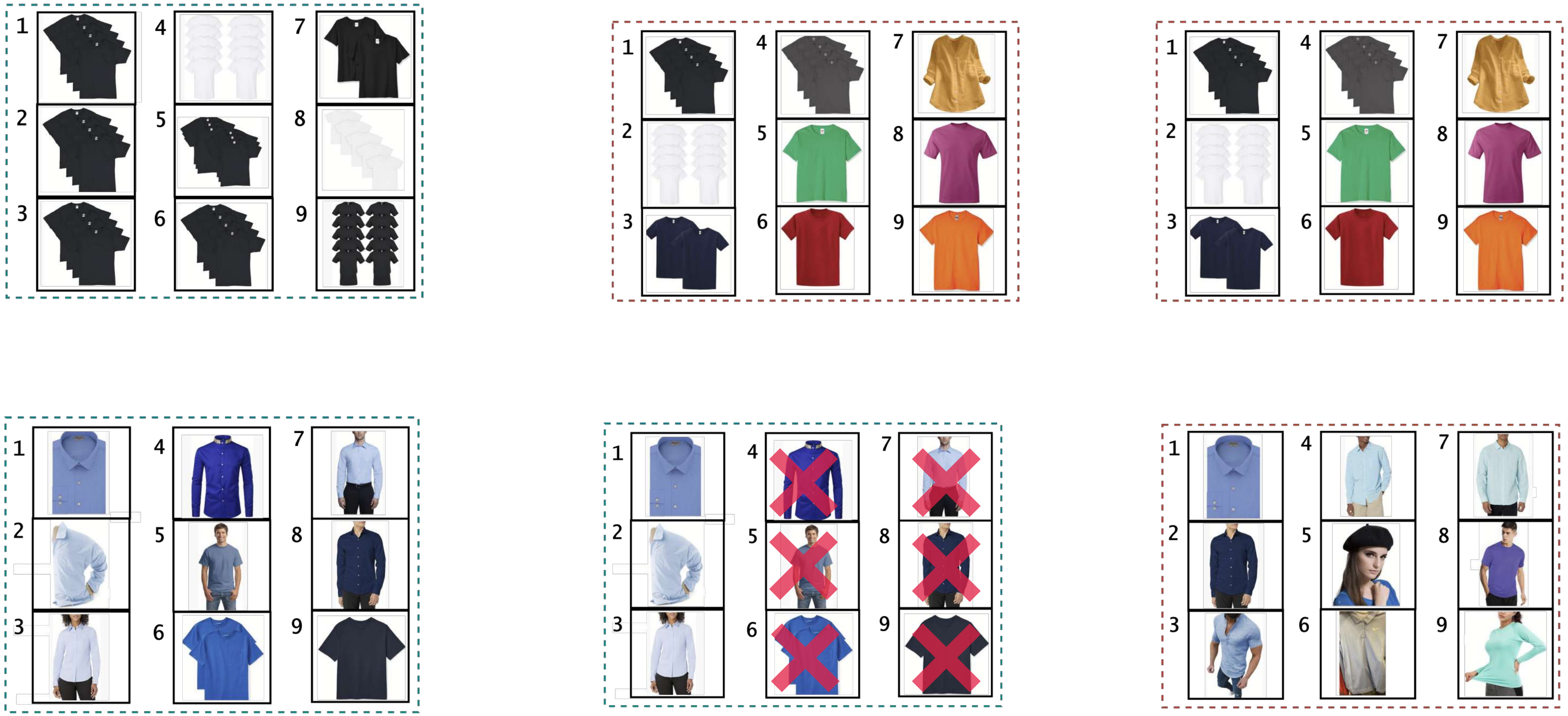
Diversity in Search

Visualisation on Amazon Dataset

Standard Search | Constrained Search | Nash Search (Ours)

Query = “Shirts”

Query = “Blue Shirts”



Diversity in Search

Motivation

Diversity in Search

Motivation

- Search

Diversity in Search

Motivation

- Search
  - A large collection of items

Diversity in Search

Motivation

- Search
  - A large collection of items
  - A query arrives

Diversity in Search

Motivation

- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query

Diversity in Search

Motivation

- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query
- Applications:

Diversity in Search

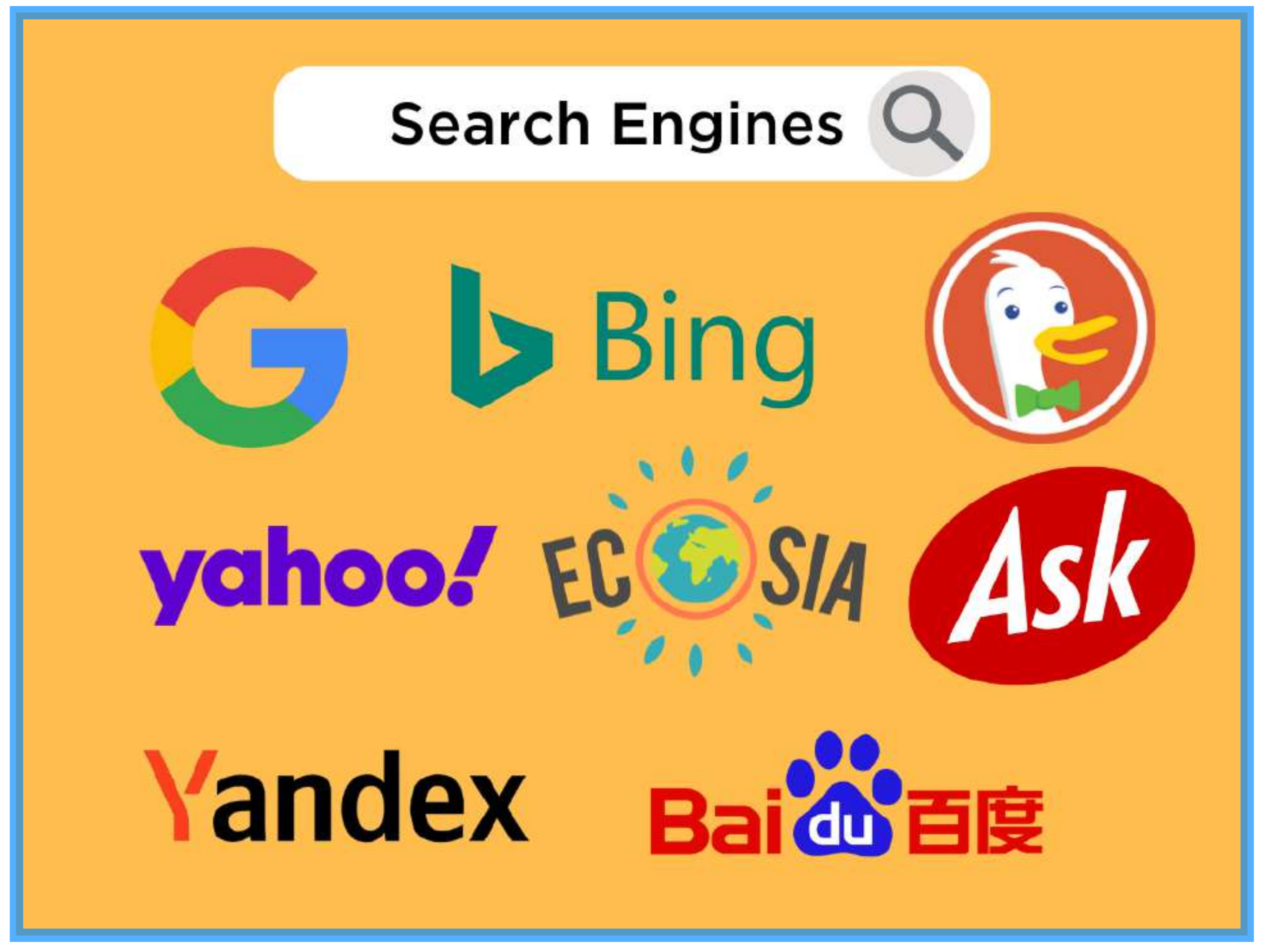
Motivation

- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query
- Applications:
  - Recommender Systems

Diversity in Search

Motivation

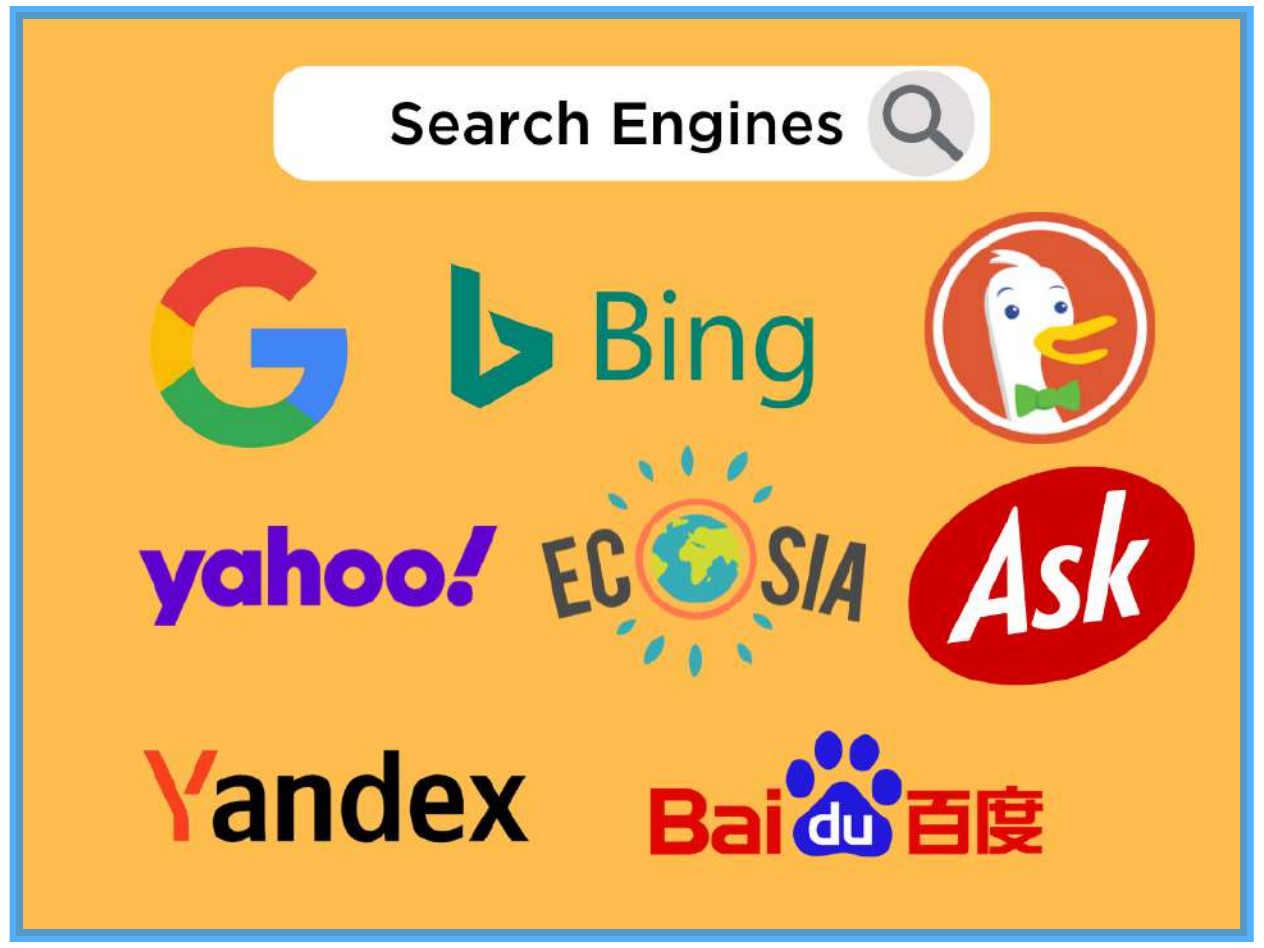
- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query
- Applications:
  - Recommender Systems
  - Search Engines

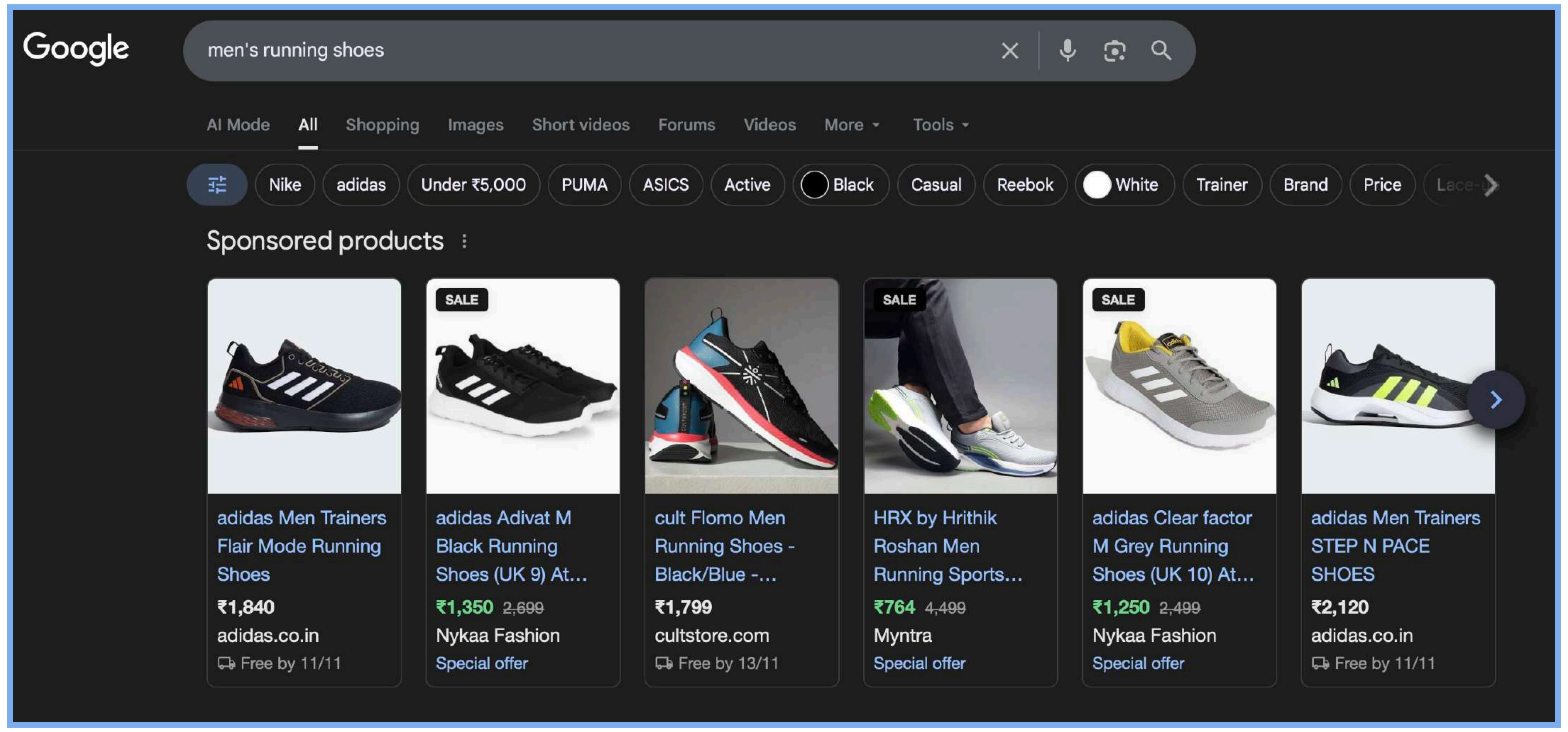


Diversity in Search

Motivation

- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query
- Applications:
  - Recommender Systems
  - Search Engines
  - Sponsored Advertisements

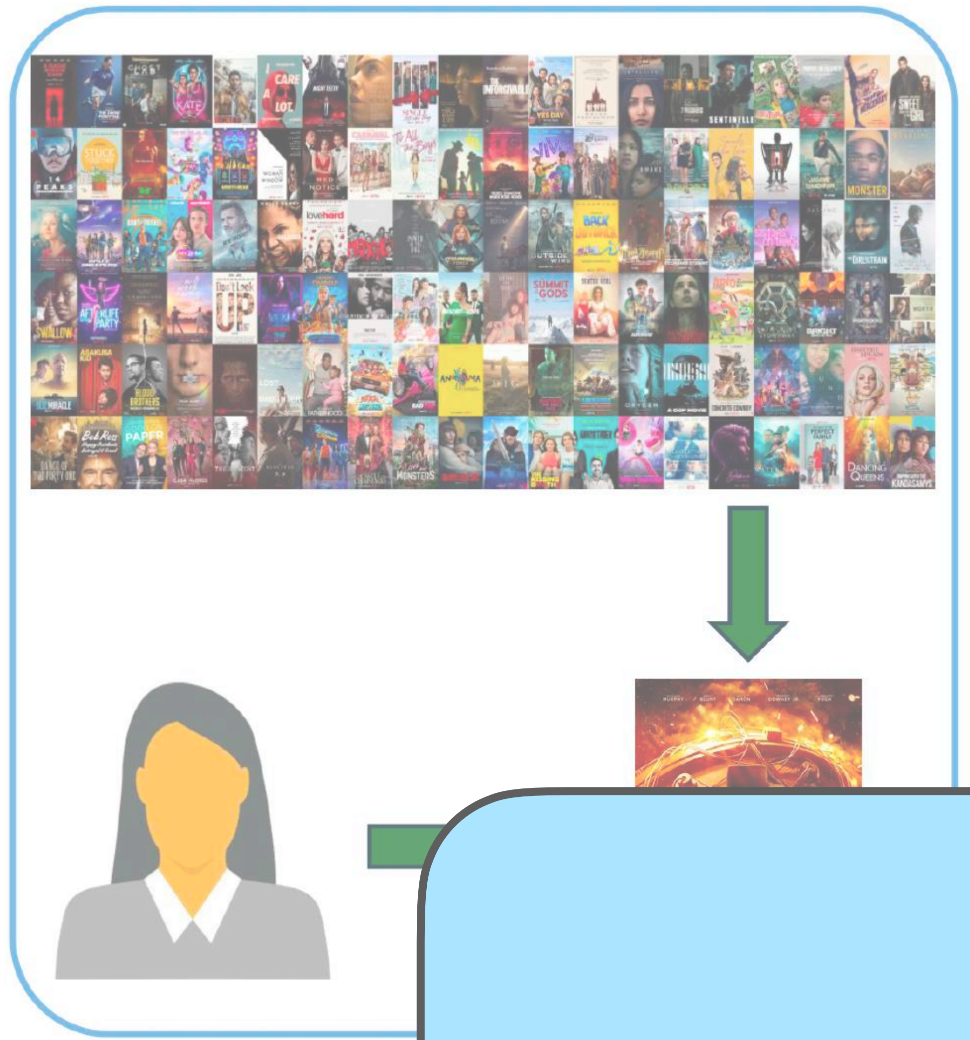


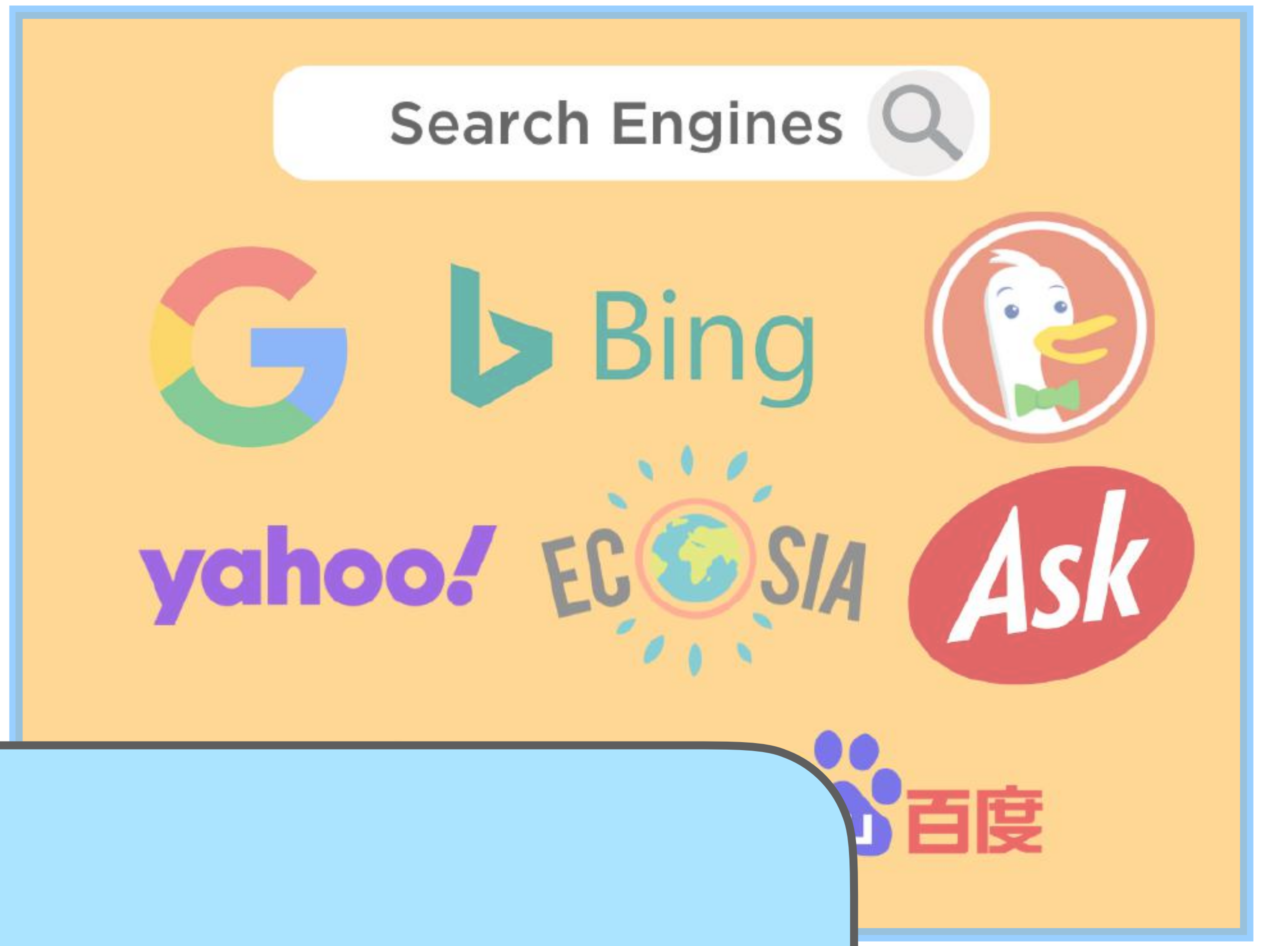


Diversity in Search

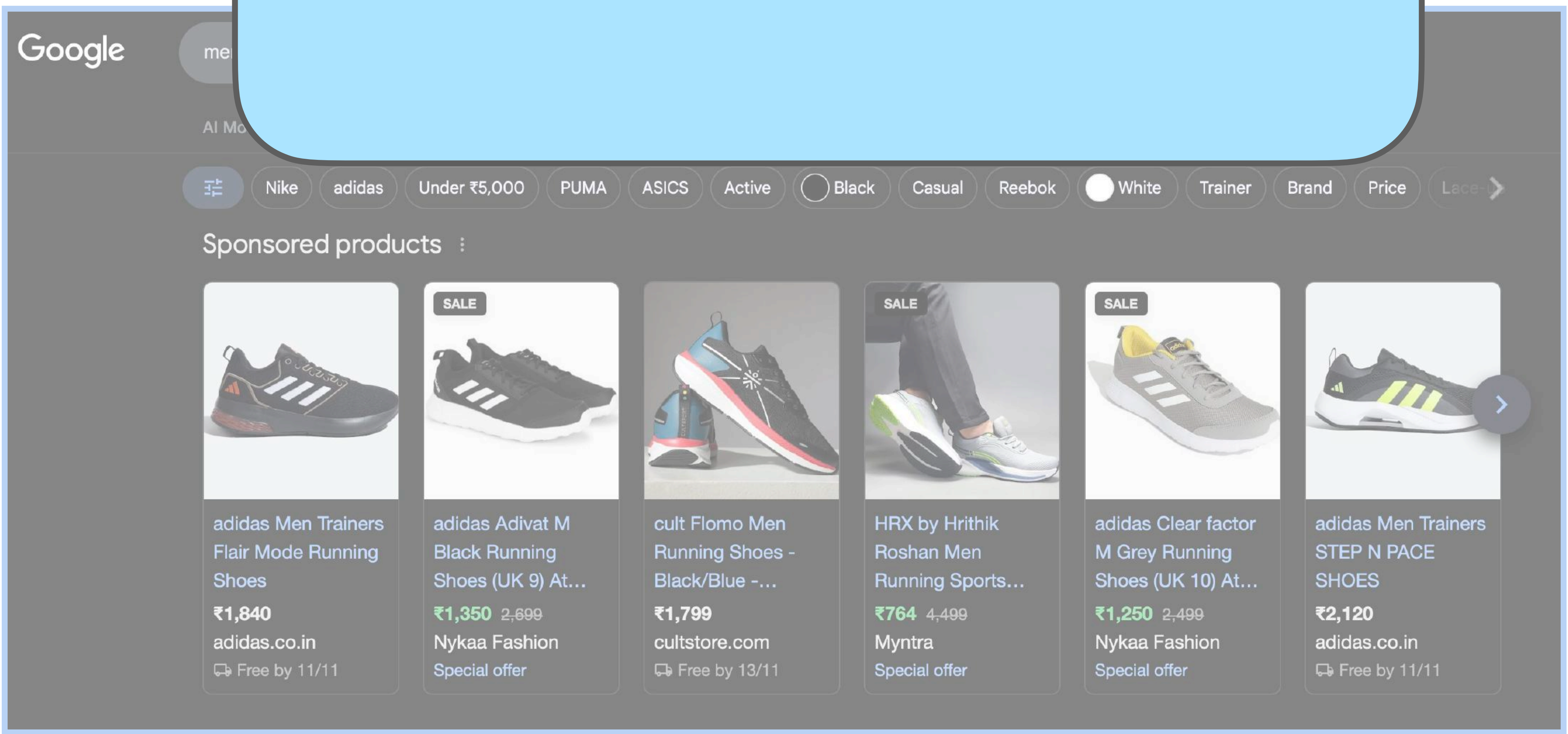
Motivation

- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query
- Applications:
  - Recommender Systems
  - Search Engines
  - Sponsored Advertisements





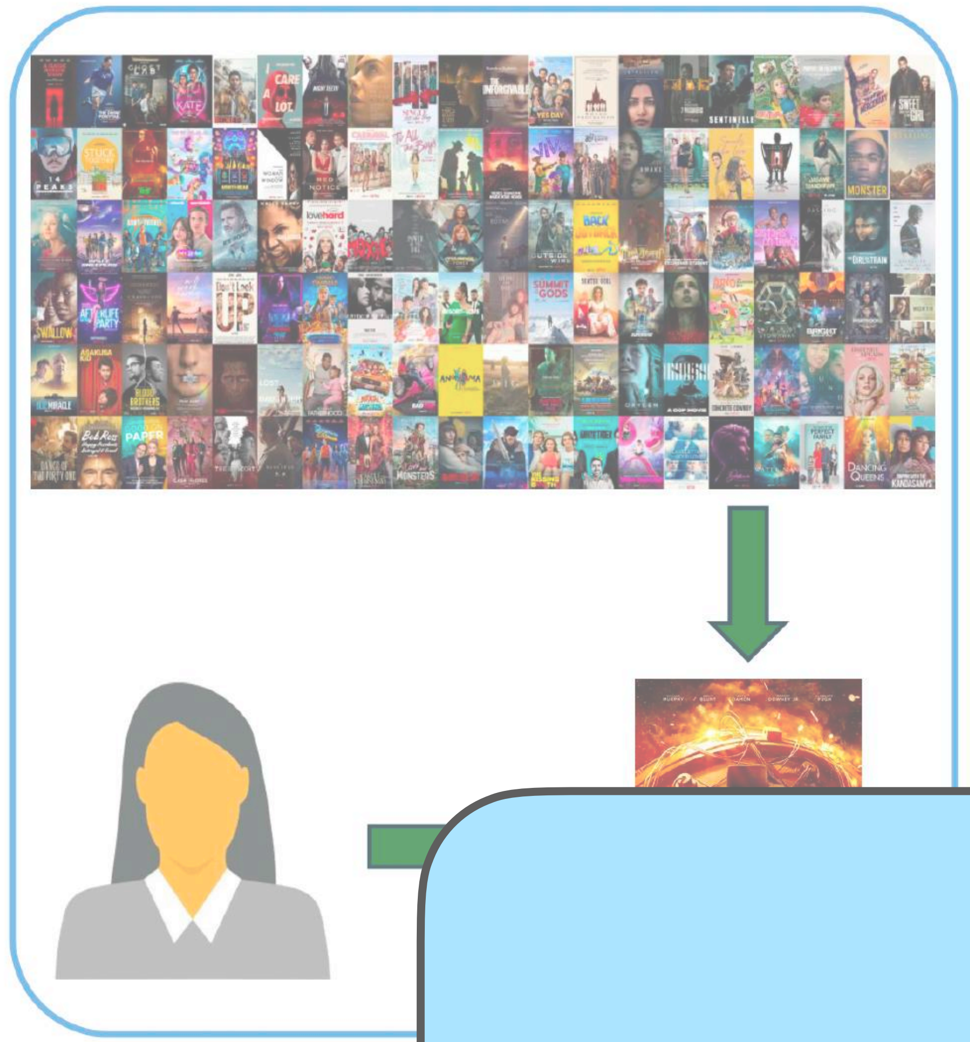
Key Challenge:
Retrieve the small set very fast

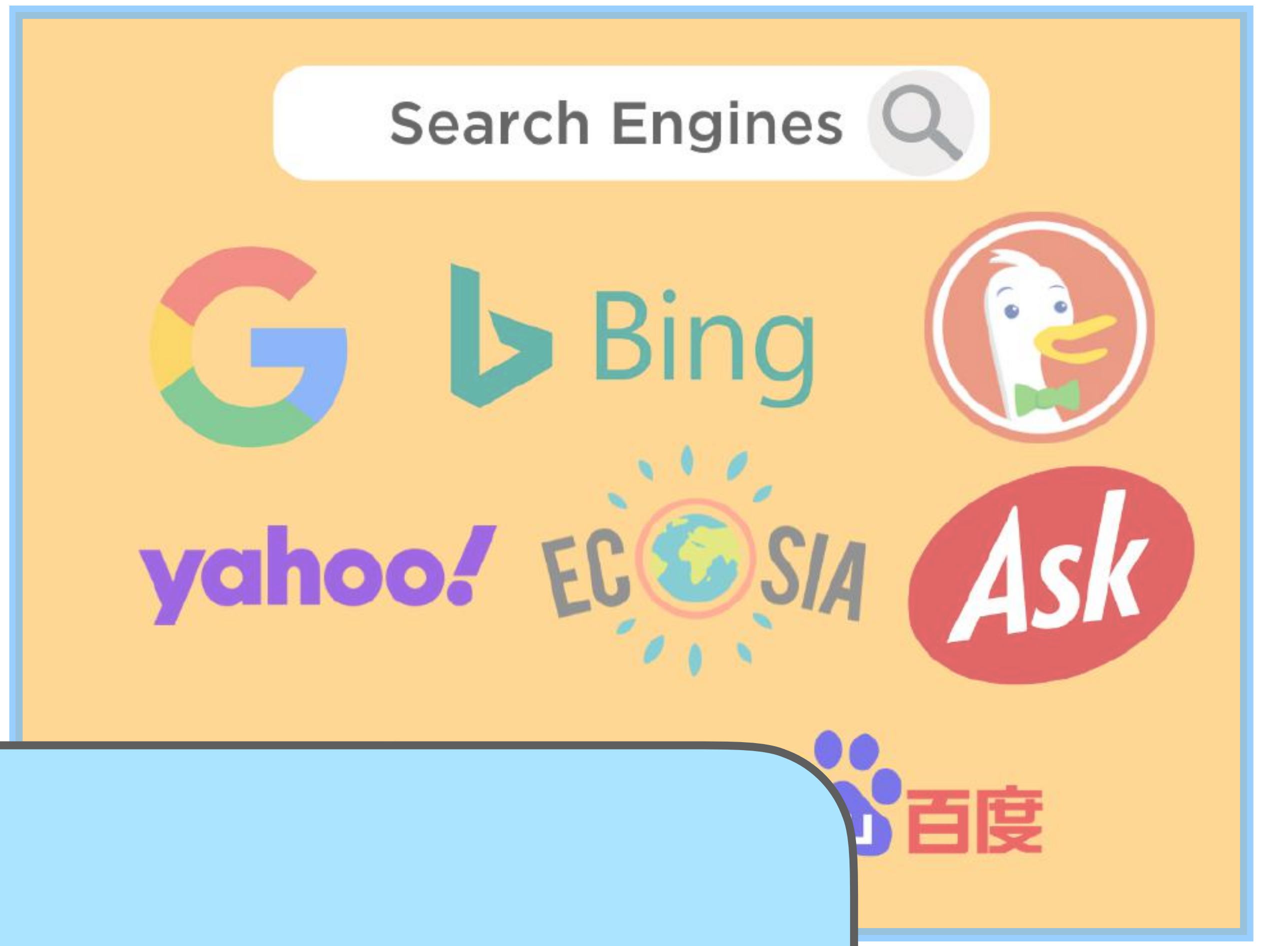


Diversity in Search

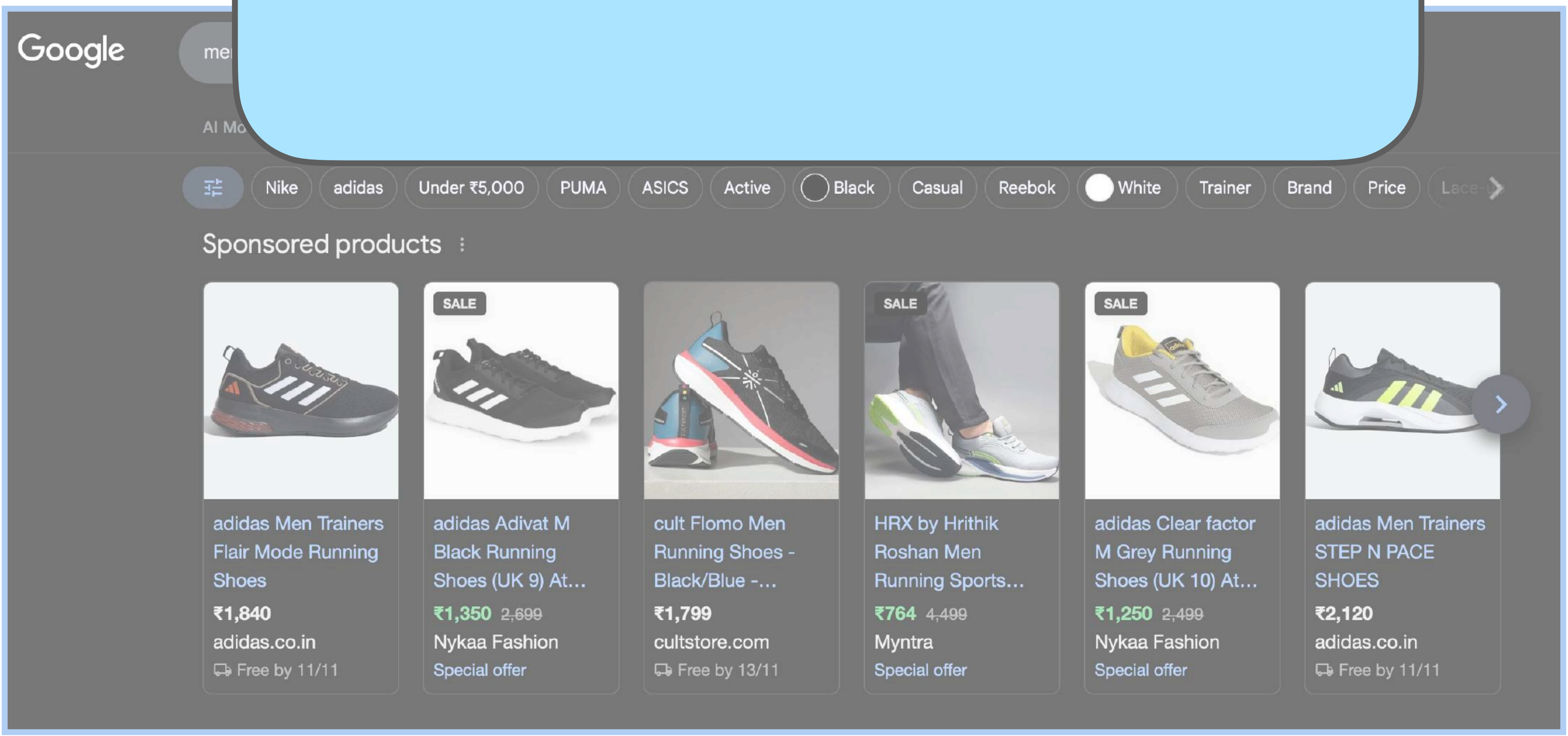
Motivation

- Search
  - A large collection of items
  - A query arrives
  - Return a small set of items that are "relevant" to the query
- Applications:
  - Recommender Systems
  - Search Engines
  - Sponsored Advertisements





Modern Requirement:
Diversity across attributes



Diversity in Search

Attribute Diversity

Diversity in Search

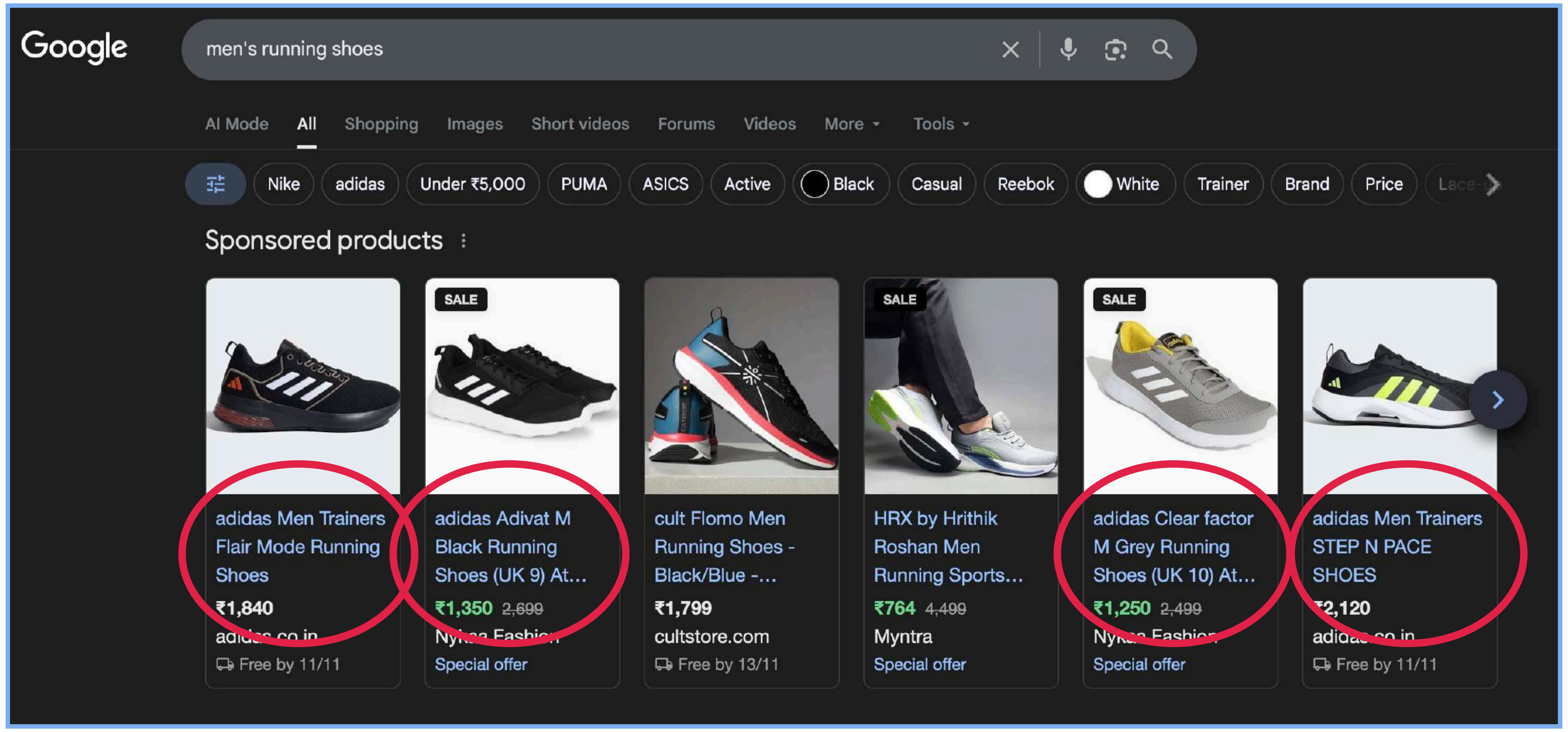
Attribute Diversity

- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads

Diversity in Search

Attribute Diversity

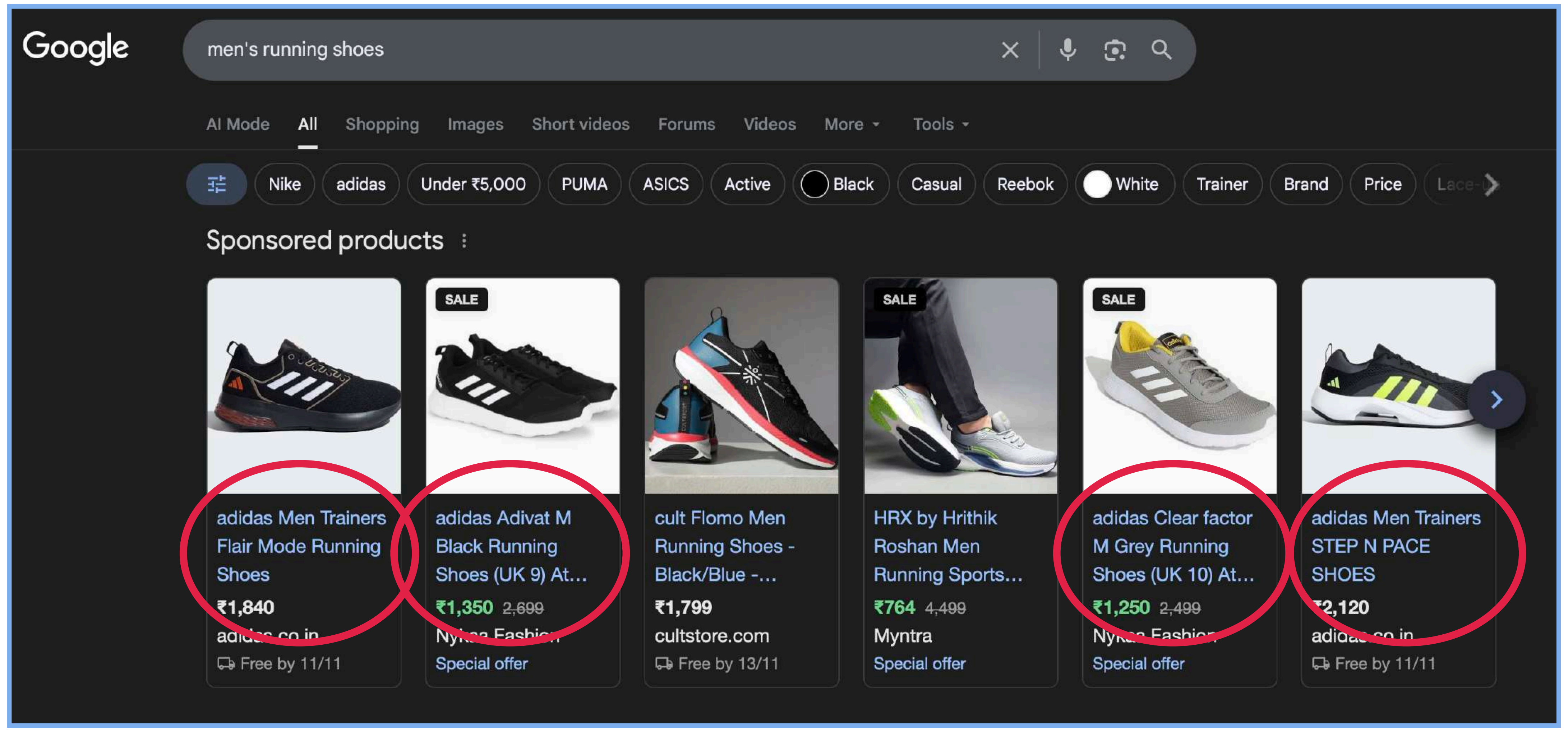
- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads



Diversity in Search

Attribute Diversity

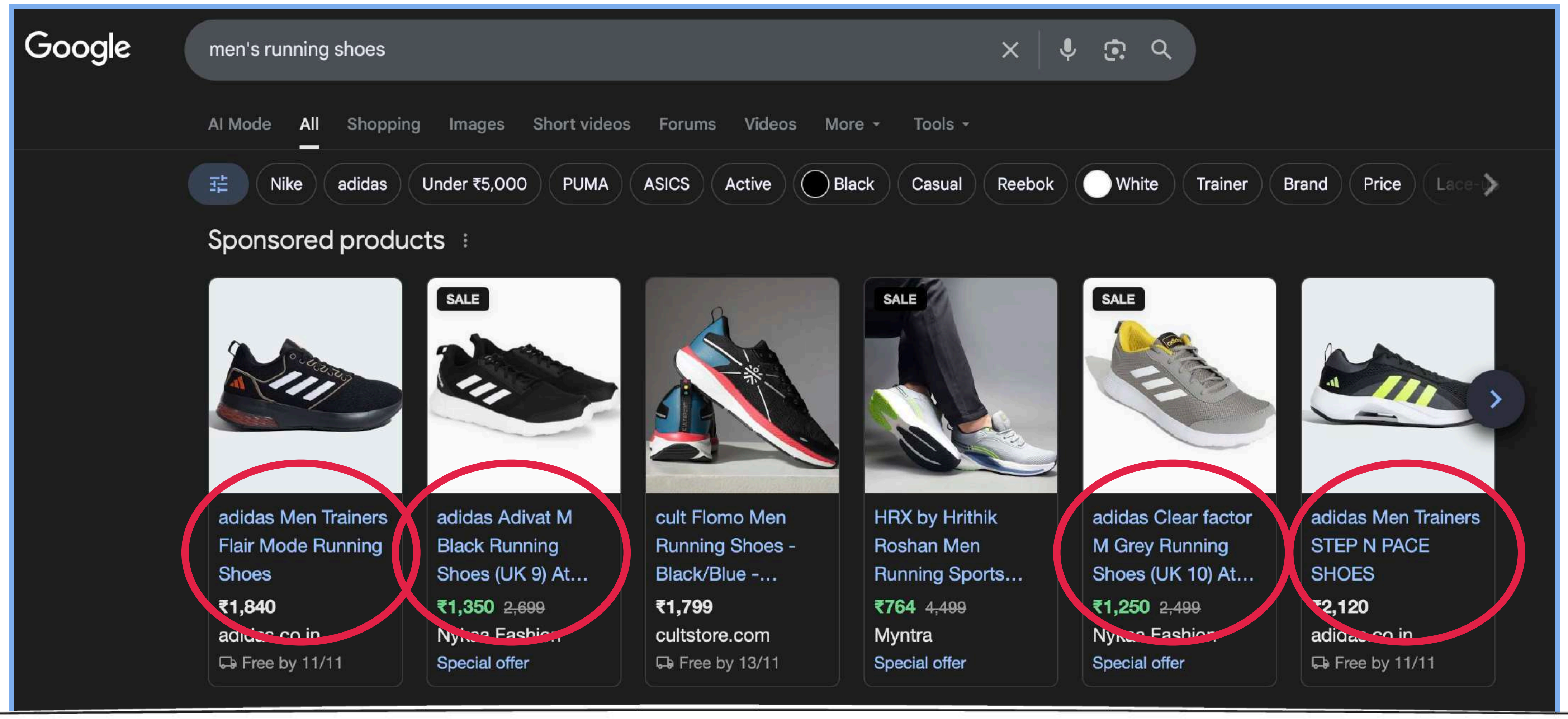
- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads

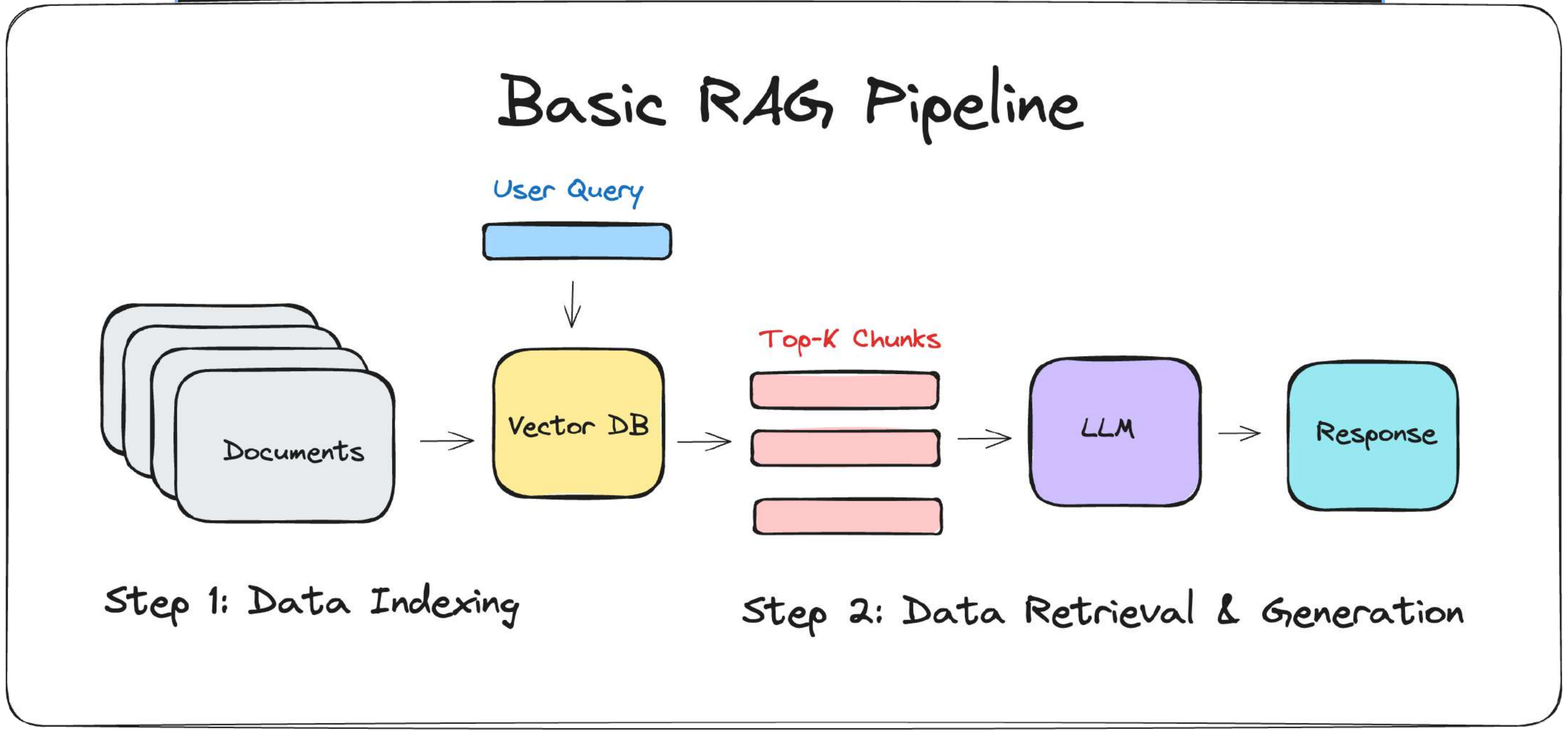


Diversity in Search

Attribute Diversity

- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads
- Retrieval-augmented Generation
  - synthesize information from various documents
  - e.g. query = “Mamdani’s policy”, response should balance different political stances

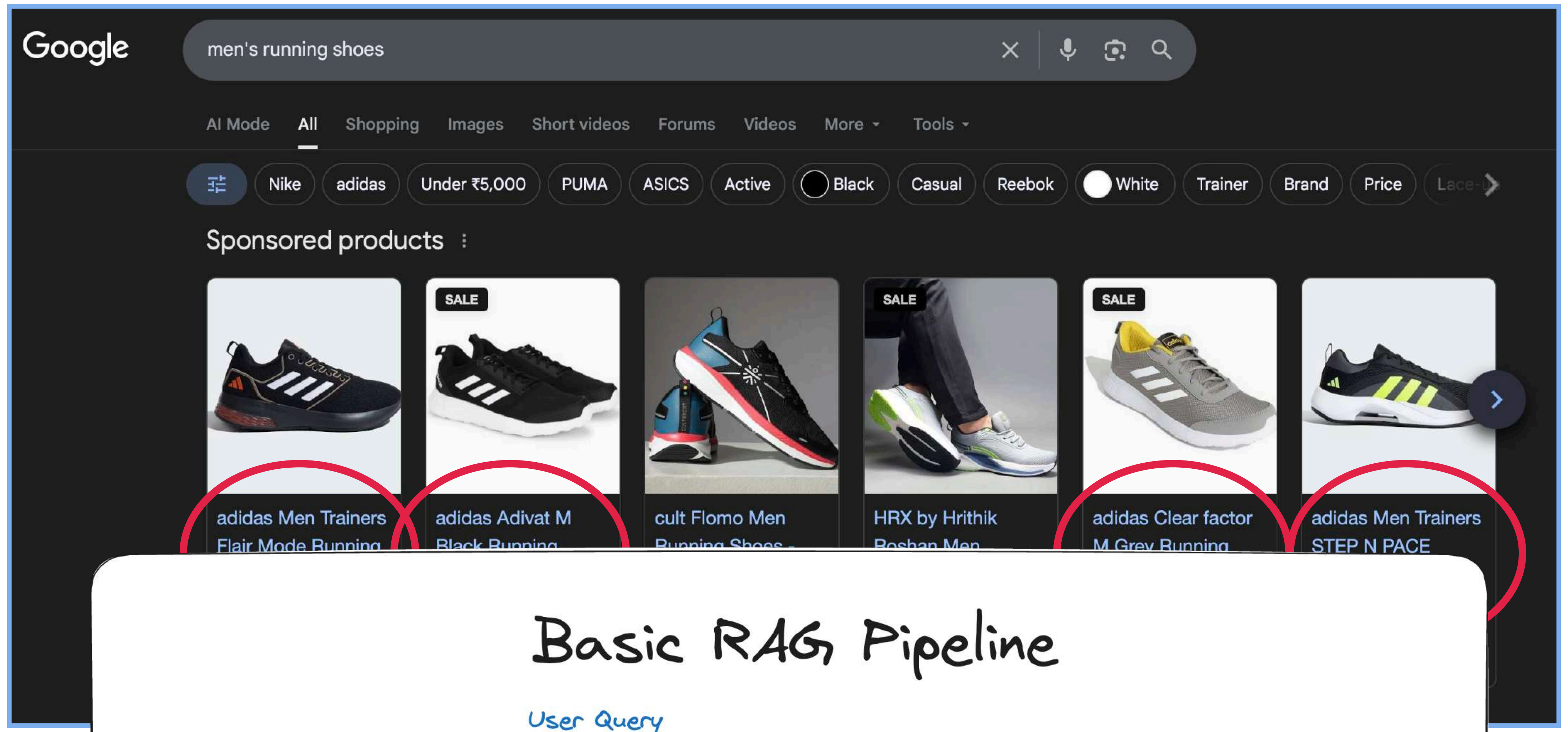


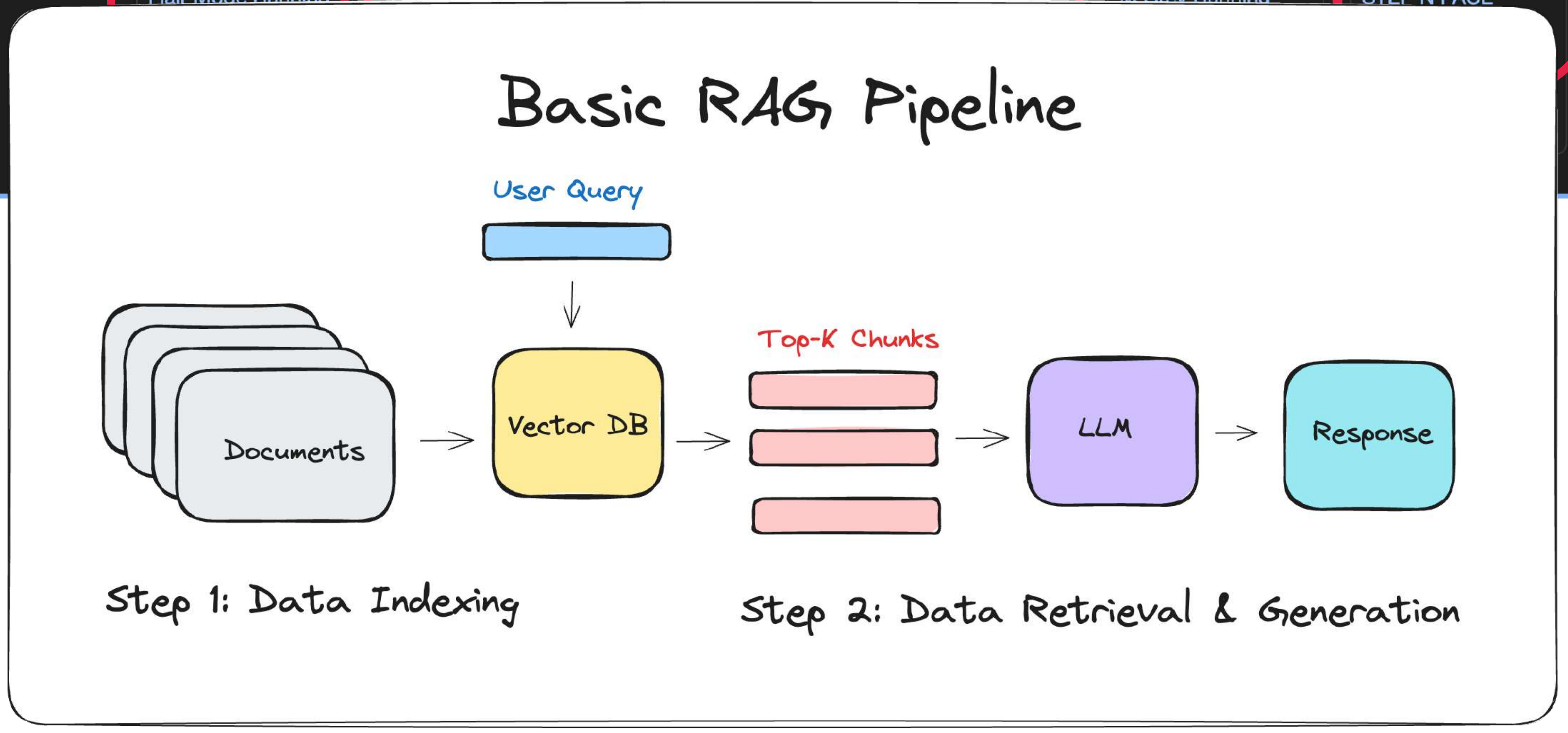


Diversity in Search

Attribute Diversity

- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads
- Retrieval-augmented Generation
  - synthesize information from various documents
  - e.g. query = “Mamdani’s policy”, response should balance different political stances

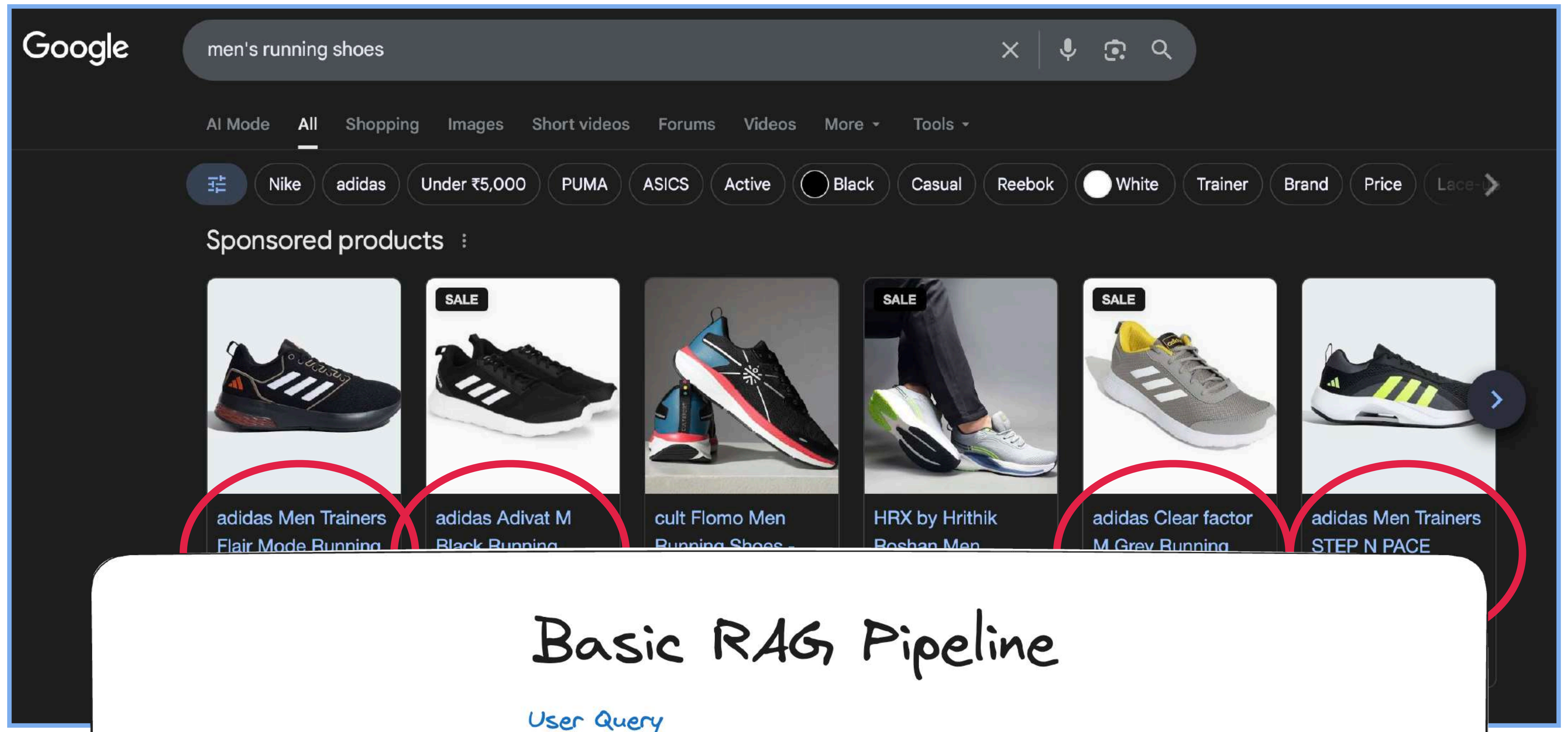


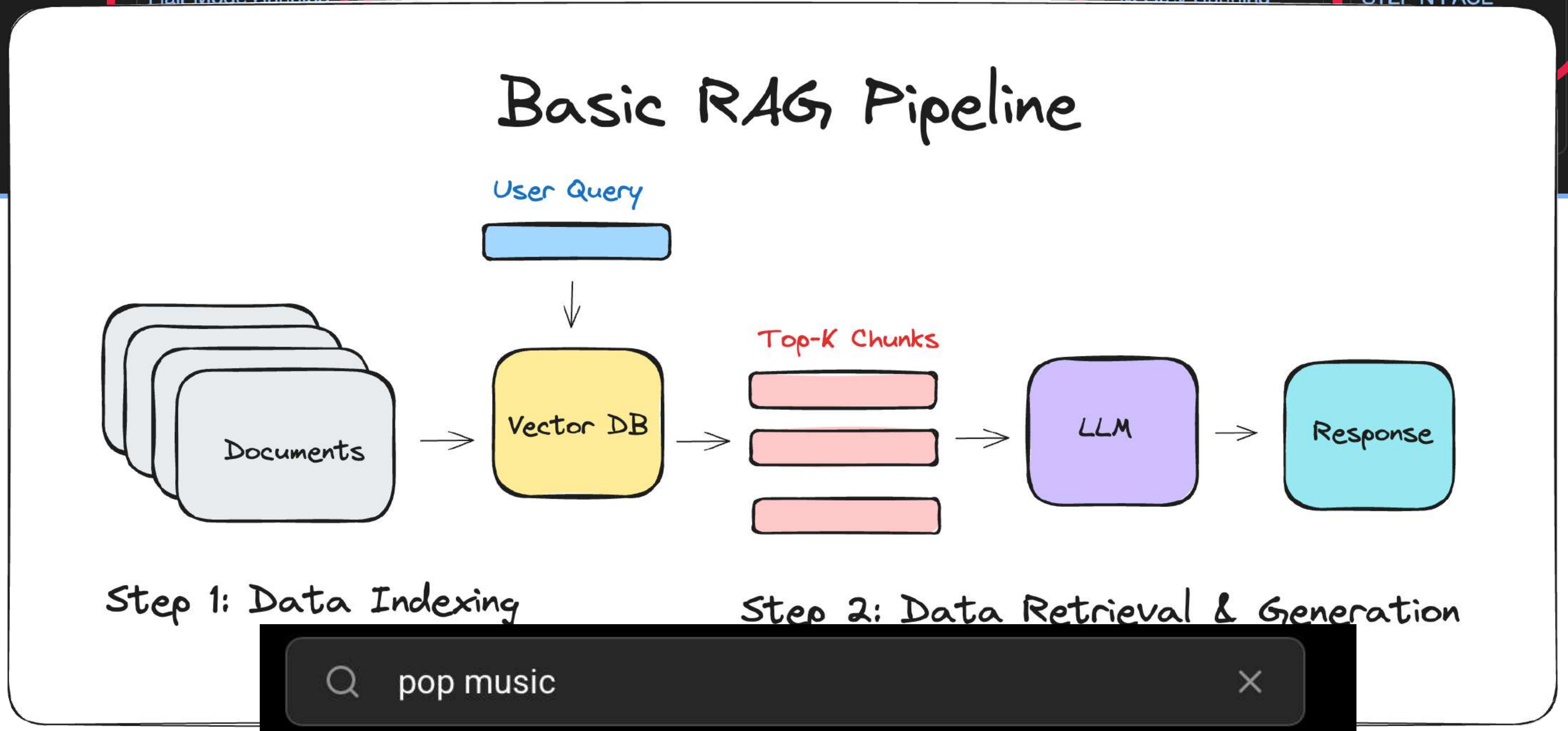


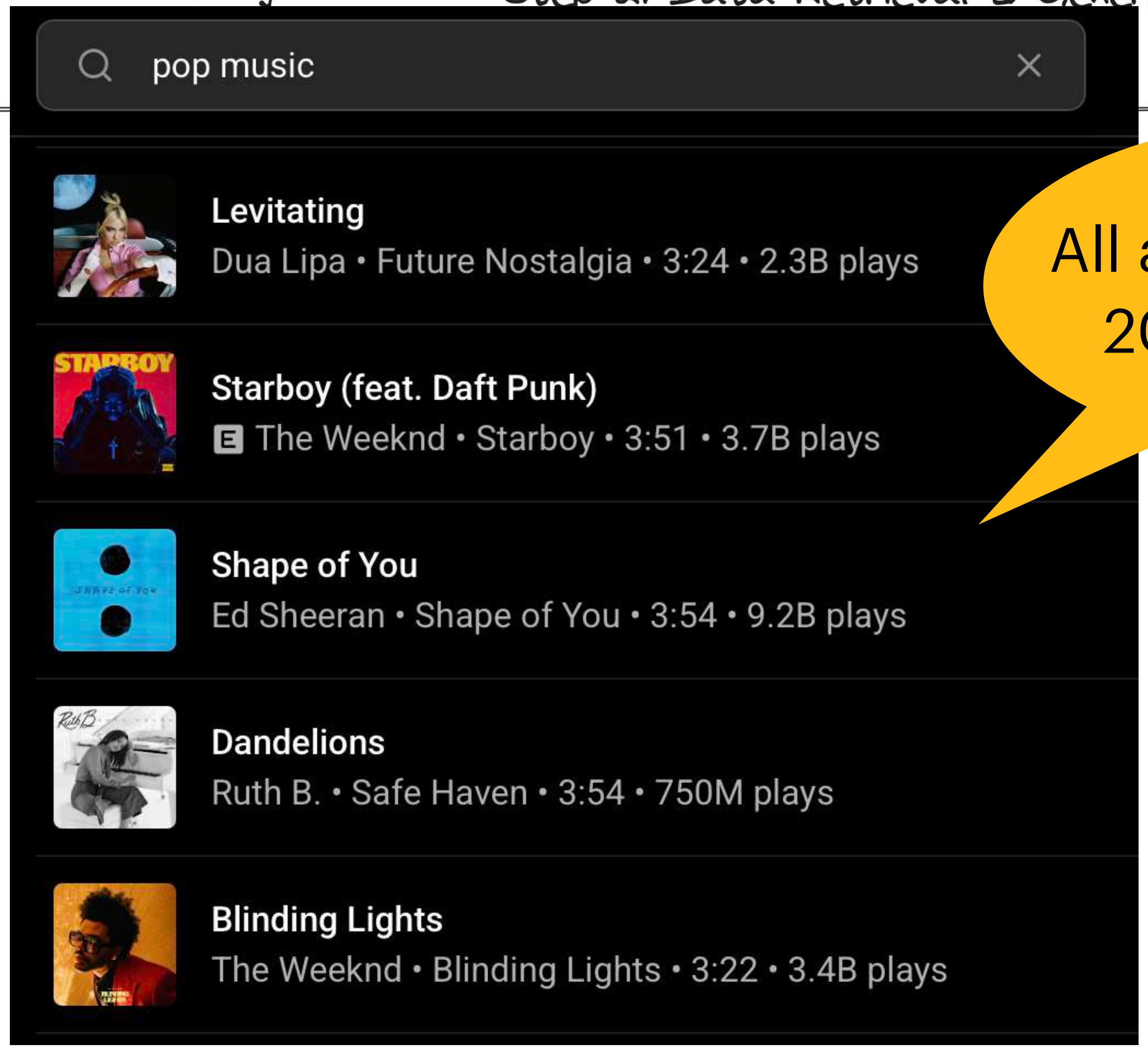
Diversity in Search

Attribute Diversity

- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads
- Retrieval-augmented Generation
  - synthesize information from various documents
  - e.g. query = “Mamdani’s policy”, response should balance different political stances
- Music Recommendation
  - a few famous artists
  - many small artists





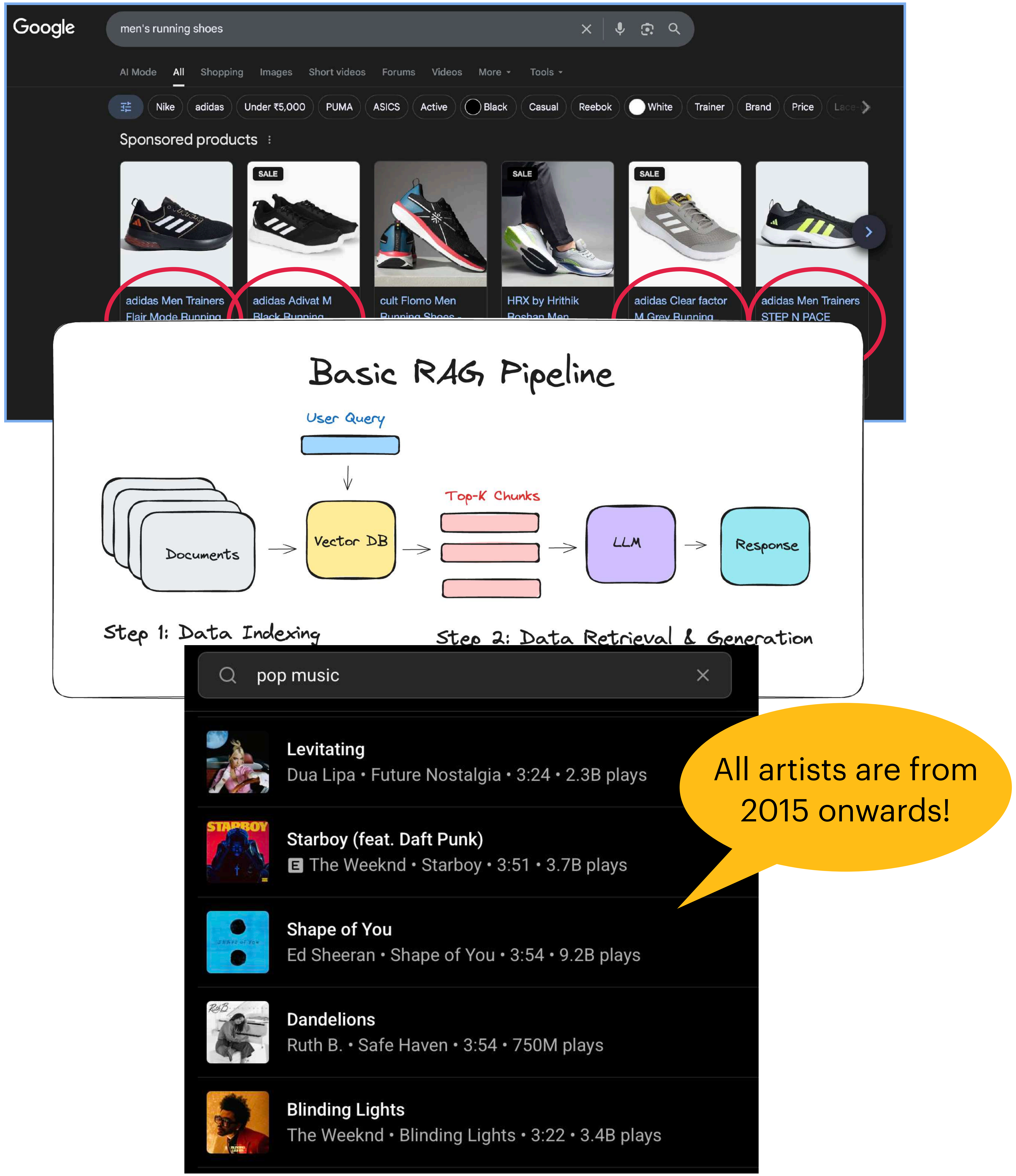


All artists are from 2015 onwards!

Diversity in Search

Attribute Diversity

- Sponsored ads
  - a few big sellers with many ads
  - many small sellers with few ads
- Retrieval-augmented Generation
  - synthesize information from various documents
  - e.g. query = “Mamdani’s policy”, response should balance different political stances
- Music Recommendation
  - a few famous artists
  - many small artists
- **Diversity can improve user experience**



Approximate Nearest Neighbor Search

Formal Setup

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\text{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, return $S \subseteq P$, $|S| = k$, containing k most similar vectors to q.

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S|=k} \sum_{v \in S} \mathrm{sim}(q, v)$$

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S| = k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S|=k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.
- Algorithms & Implementations: DiskANN [SDKKS19], [IX24], [GKSW25]

Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S|=k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.
- Workhorse: Recommender System

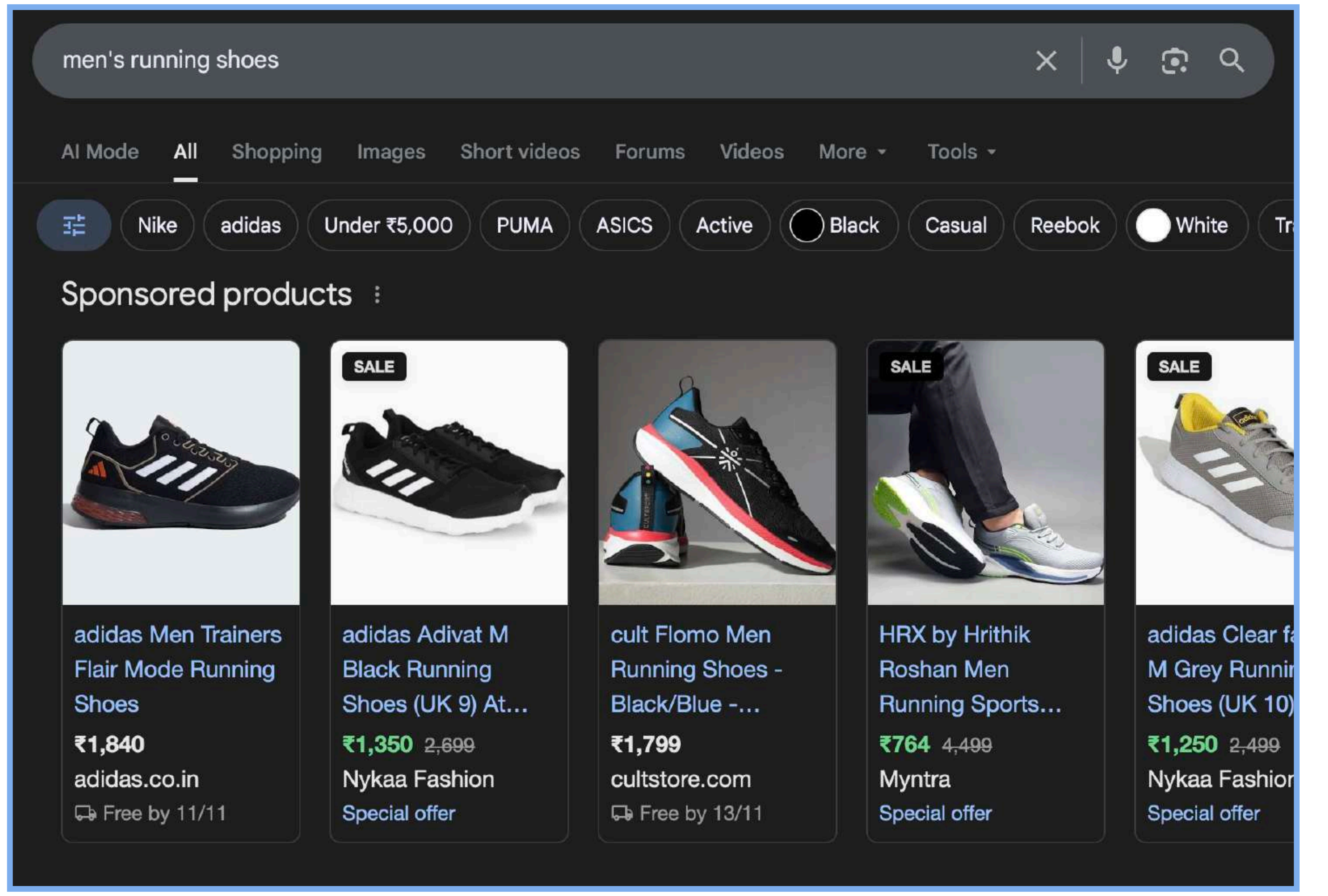
Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S| = k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.
- Workhorse: Recommender System, Display ads



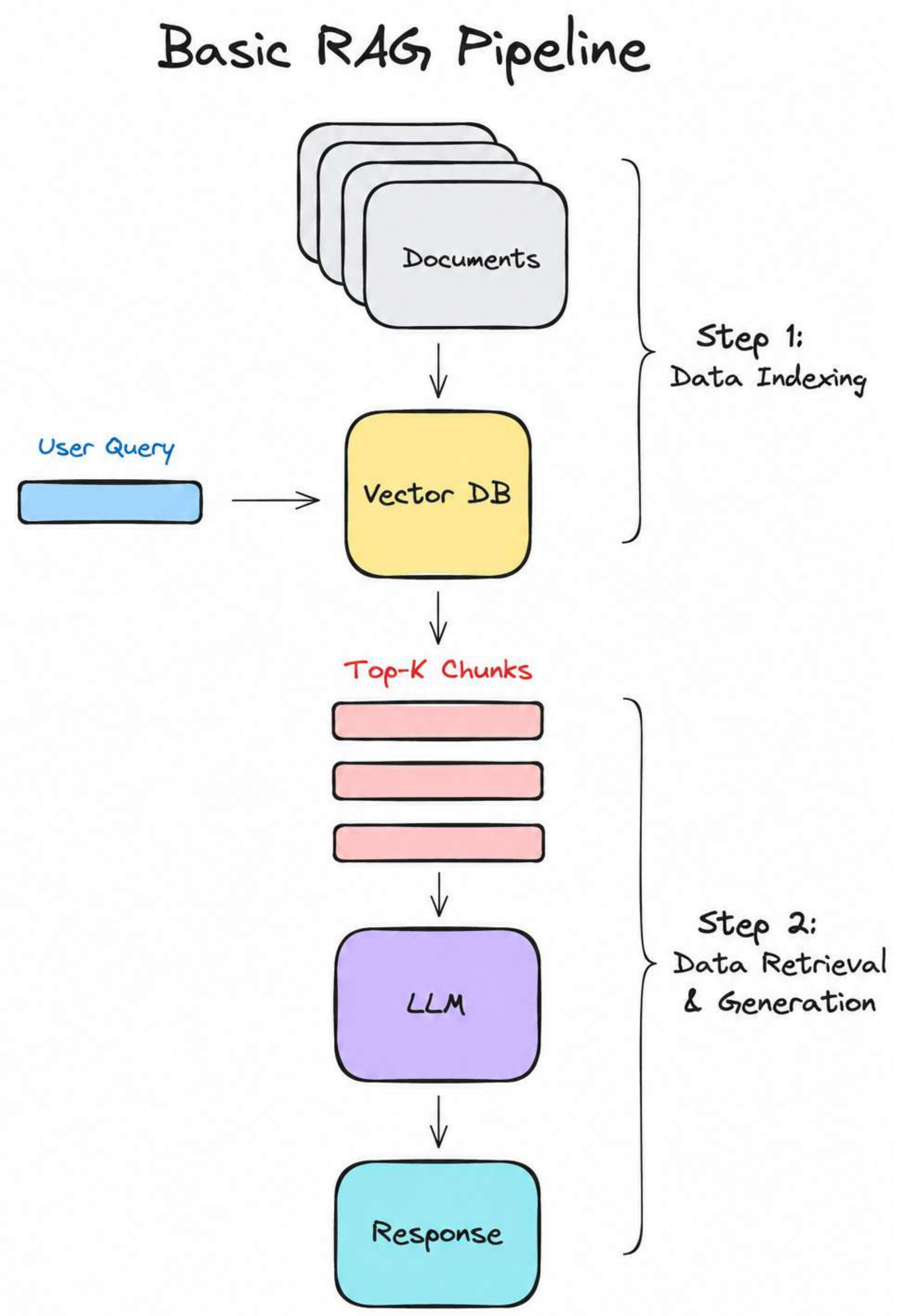
Approximate Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S|=k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.
- Workhorse: Recommender System, Display ads, RAG



Diverse Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S| = k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.
- Workhorse: Recommender System, Display ads, RAG

Diversity across Attributes

- Set of attributes $\{1, 2, \ldots, c\}$
- Every vector $v \in P$ is associated with $\mathrm{atb}(v) \in [c]$

Diverse Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- Given any query $q \in \mathbb{R}^d$, alternatively, solve

$$\max_{S \subseteq P, |S| = k} \sum_{v \in S} \mathrm{sim}(q, v)$$

- Approximate Nearest Neighbor (ANN) Search — similarities are approximately close to the optimal ones.
- Workhorse: Recommender System, Display ads, RAG

Diversity across Attributes

- Set of attributes $\{1, 2, \ldots, c\}$
- Every vector $v \in P$ is associated with $\mathrm{atb}(v) \in [c]$

Desideratum: Retrieved set S should be attribute-diverse

Diverse Nearest Neighbor Search

Formal Setup

- A set of vectors $P \subseteq \mathbb{R}^d$ with $|P| = n$.
- A similarity metric $\mathrm{sim} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Integer $k \in \mathbb{N}$.
- **Constrained ANN** [AIKMRSX25]: Given $q \in \mathbb{R}^d$, solve

$$\max_{S \subseteq P, |S| = k} \sum_{v \in S} \mathrm{sim}(q, v)$$
$$\text{s.t.} \quad |\{v \in S : \mathrm{atb}(v) = \ell\}| \leq k' \text{ for each } \ell \in [c]$$

  - At most k' vectors from each attribute $\ell \in [c]$.

Diversity across Attributes

- Set of attributes $\{1, 2, \ldots, c\}$
- Every vector $v \in P$ is associated with $\mathrm{atb}(v) \in [c]$

Desideratum: Retrieved set S should be attribute-diverse

Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25]: At most k' vectors from each attribute $\ell \in [c]$.

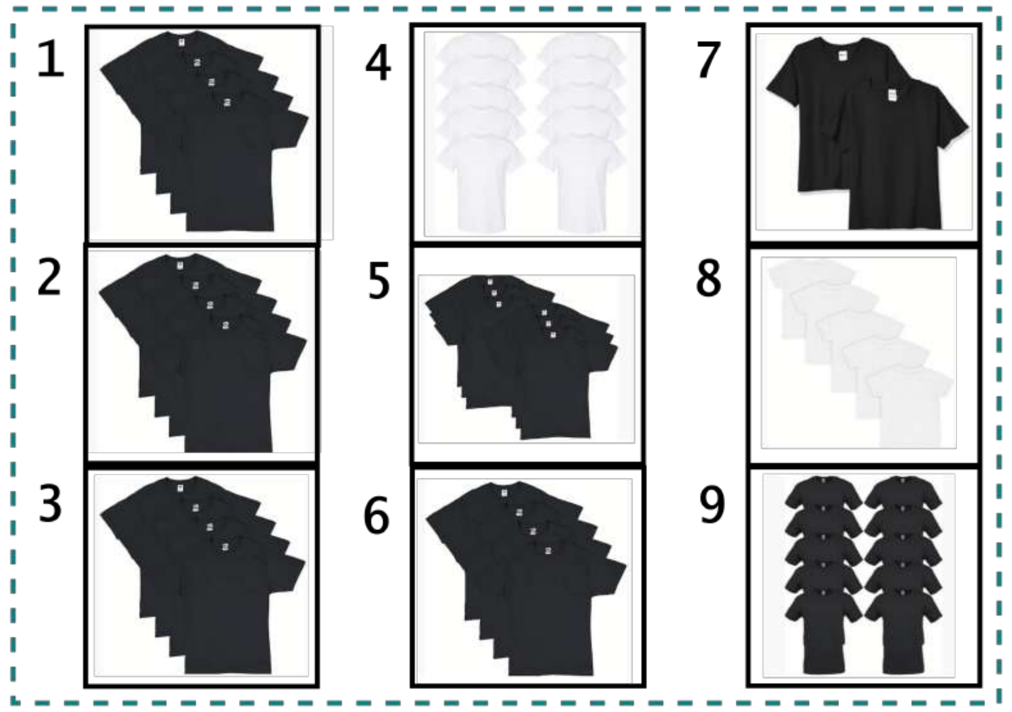
Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25]: At most k' vectors from each attribute $\ell \in [c]$.

Standard Search

Constrained Search

Query = “Shirts”



Query = “Blue Shirts”

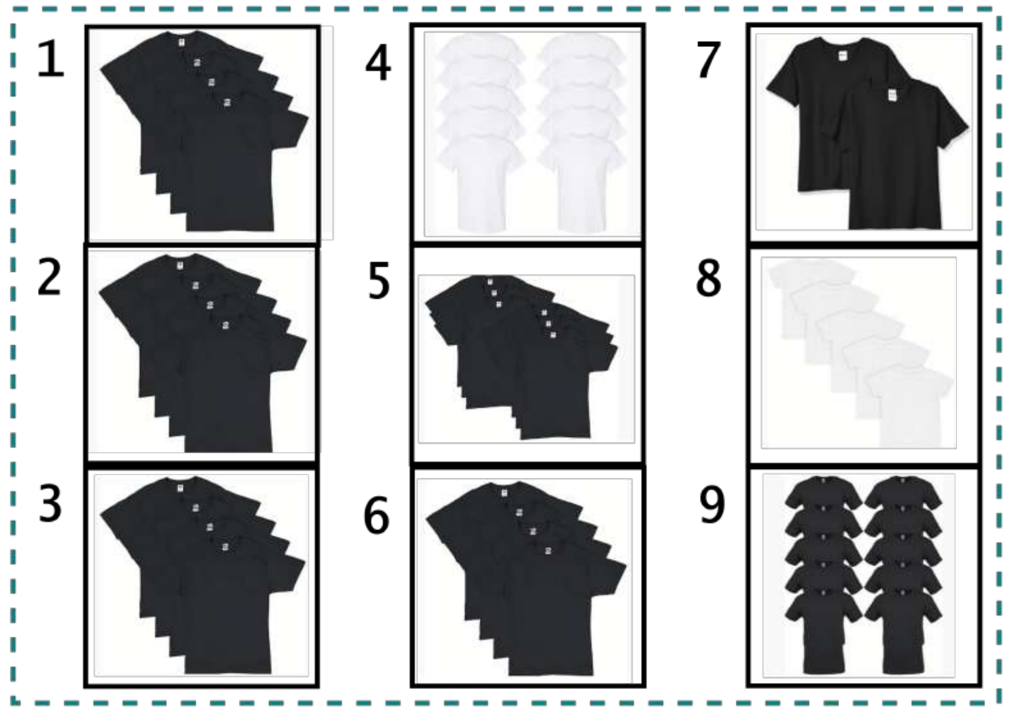
Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25]: At most k' vectors from each attribute $\ell \in [c]$.

Standard Search

Constrained Search

Query = "Shirts"



Query = "Blue Shirts"

Parameter k'

- How to choose?
- What fairness guarantees for a given k'?
- Relevance vs Diversity

A Mathematical Economics Detour

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

A Mathematical Economics Detour

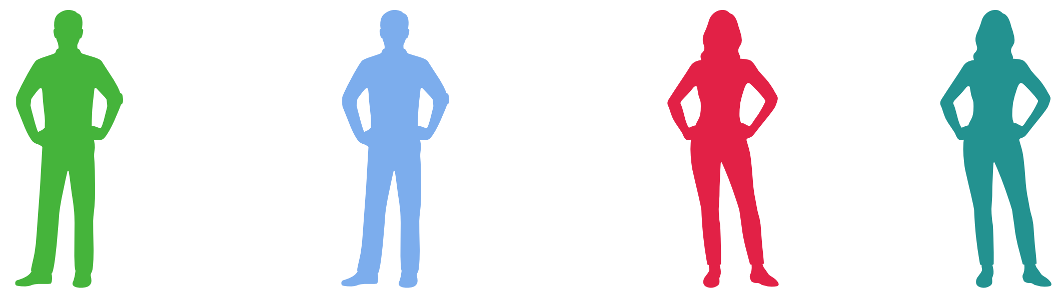
- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

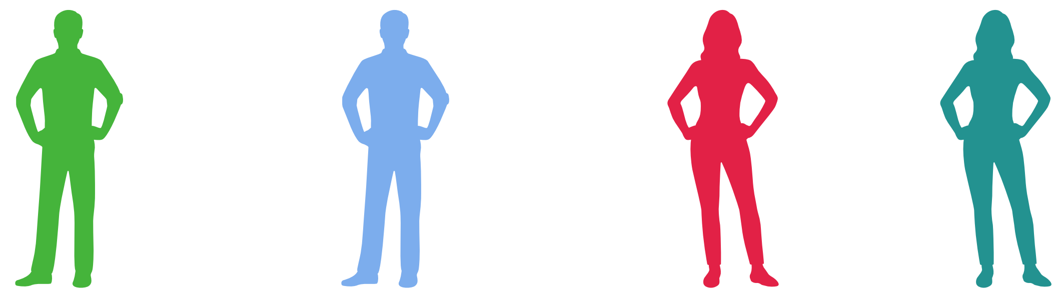


Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.



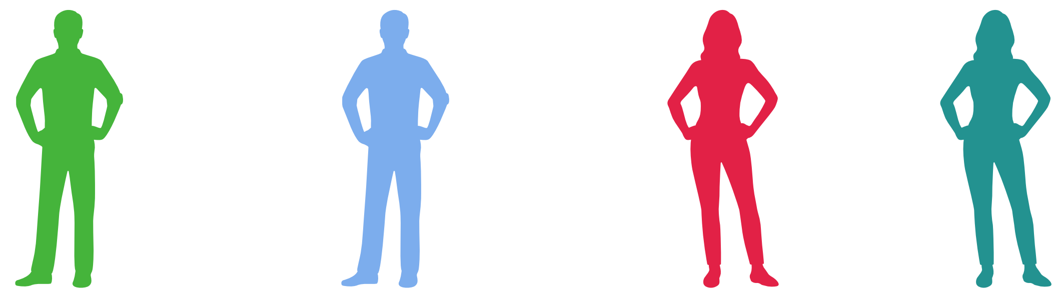
Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

Which option is better?

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \rightarrow \mathbb{R}_+$.



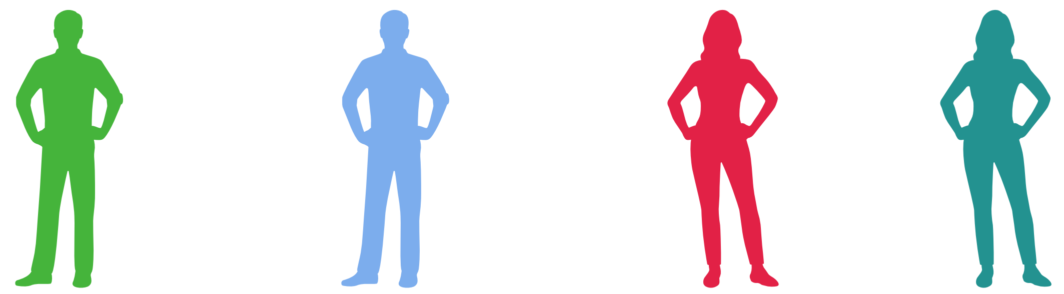
Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

Axiomatic Basis



Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

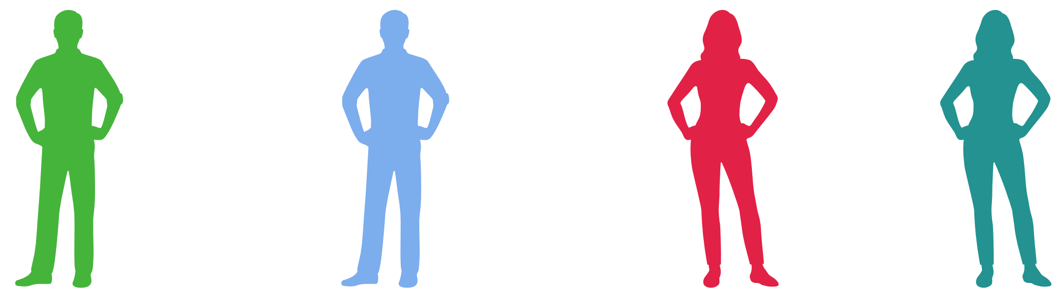
A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

Axiomatic Basis

- Pareto Efficiency



Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

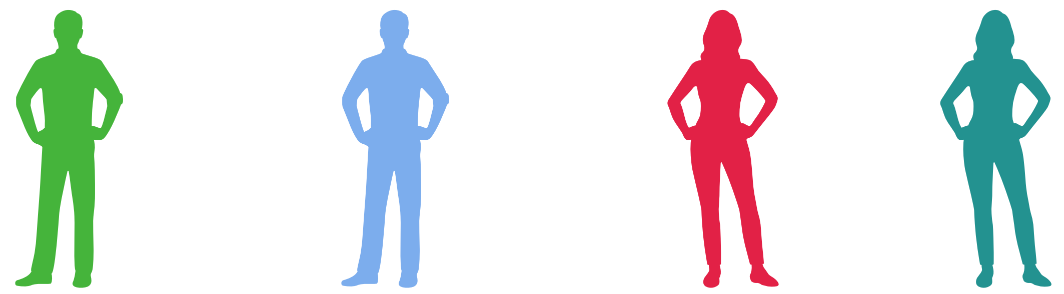
A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

Axiomatic Basis

- **Pareto Efficiency**
- **Pigou-Dalton Principle**



Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

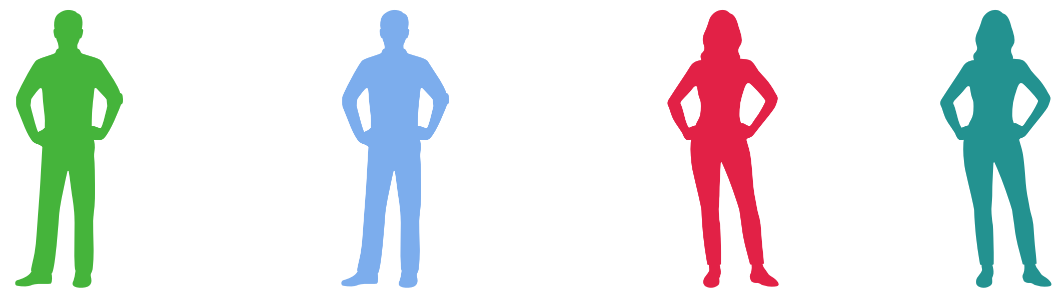
A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

Axiomatic Basis

- **Pareto Efficiency**
- **Pigou-Dalton Principle**
- **Symmetry, Independece of Irrelevant Agent**



Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

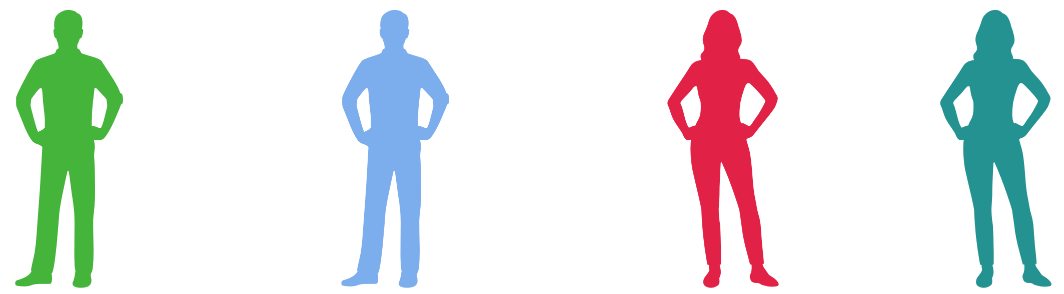
A Mathematical Economics Detour

- Algorithms should be viewed as economic policies.
- Algorithm's decision induces utilities on the participating agents.
- For a decision S, the utilities of m agents are $u_1(S), u_2(S), \ldots, u_m(S)$.
- How do we compare decisions S and S'?

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

Axiomatic Basis

- **Pareto Efficiency**
- **Pigou-Dalton Principle**
- **Symmetry, Independece of Irrelevant Agent**
- **Scale-free**



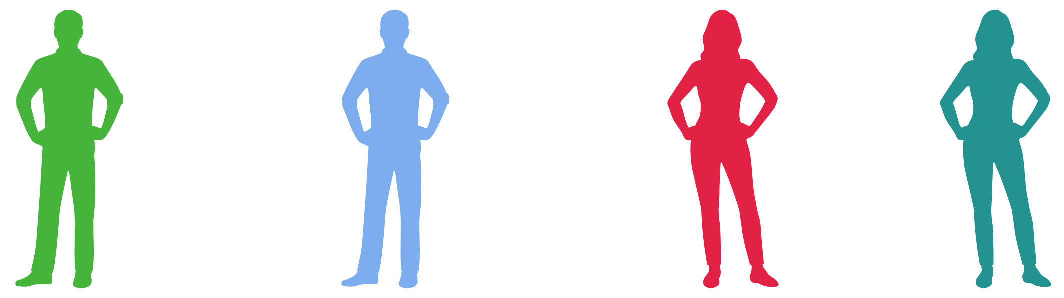
Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

A Mathematical Economics Detour

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+.$

Axiomatic Basis

- Pareto Efficiency
- Pigou-Dalton Principle
- Symmetry, Independece of Irrelevant Agent
- Scale-free



Option 1	100	20	10	0
Option 2	45	25	25	32
Option 3	30	30	30	30

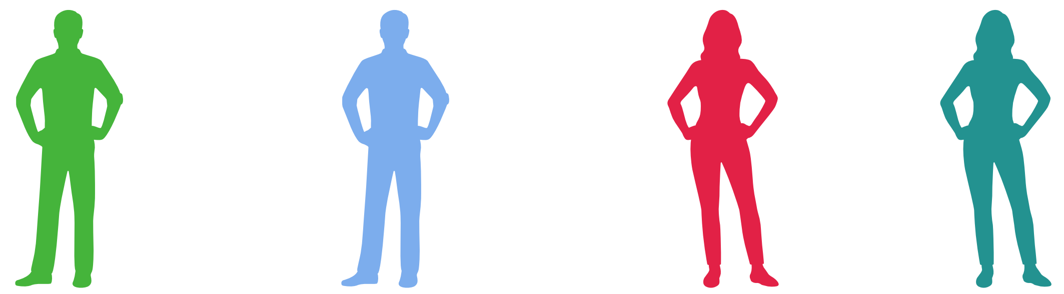
A Mathematical Economics Detour

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

- Social Welfare - $\mathrm{SW}(u_1, \ldots, u_m) = \frac{1}{m} \sum_{i=1}^{m} u_i$.

Axiomatic Basis

- Pareto Efficiency
- Pigou-Dalton Principle
- Symmetry, Independece of Irrelevant Agent
- Scale-free



					SW
Option 1	100	20	10	0	32.5
Option 2	45	25	25	32	31.75
Option 3	30	30	30	30	30

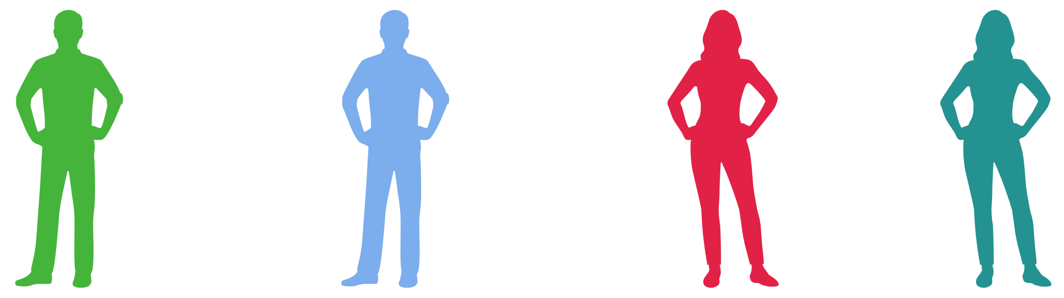
A Mathematical Economics Detour

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

Axiomatic Basis

- **Pareto Efficiency**
- **Pigou-Dalton Principle**
- **Symmetry, Independece of Irrelevant Agent**
- **Scale-free**

- Social Welfare - $\mathrm{SW}(u_1, \ldots, u_m) = \frac{1}{m}\sum_{i=1}^{m} u_i$.
- Nash Social Welfare - $\mathrm{NSW}(u_1, \ldots, u_m) = \left(\prod_{i=1}^{m} u_i\right)^{\frac{1}{m}}$.



					SW	NSW
Option 1	100	20	10	0	32.5	0
Option 2	45	25	25	32	31.75	30.8
Option 3	30	30	30	30	30	30

A Mathematical Economics Detour

Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

- Social Welfare - $\mathrm{SW}(u_1, \ldots, u_m) = \frac{1}{m} \sum_{i=1}^{m} u_i$.
- Nash Social Welfare - $\mathrm{NSW}(u_1, \ldots, u_m) = \left(\prod_{i=1}^{m} u_i \right)^{\frac{1}{m}}$.
- Egalitarian Welfare - $\mathrm{EgW}(u_1, \ldots, u_m) = \min_{i \in [m]} u_i$.

Axiomatic Basis

- Pareto Efficiency
- Pigou-Dalton Principle
- Symmetry, Independece of Irrelevant Agent
- Scale-free

					SW	NSW	EgW
Option 1	100	20	10	0	32.5	0	0
Option 2	45	25	25	32	31.75	30.8	25
Option 3	30	30	30	30	30	30	30

A Mathematical Economics Detour

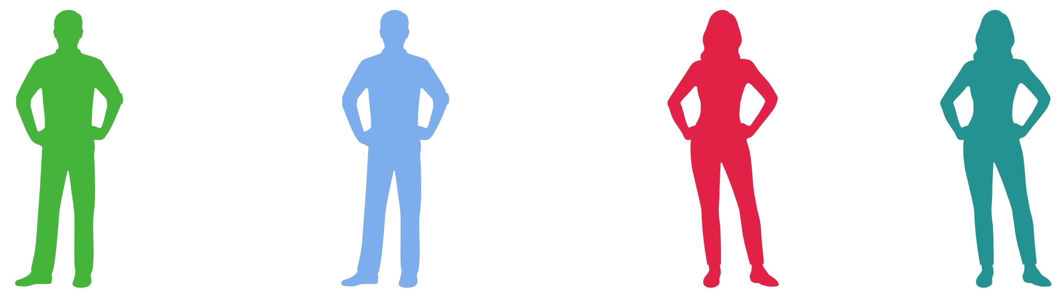
Welfare Functions $f : \mathbb{R}_+^m \to \mathbb{R}_+$.

- Social Welfare - $\mathrm{SW}(u_1, \ldots, u_m) = \frac{1}{m}\sum_{i=1}^{m} u_i$.
- Nash Social Welfare - $\mathrm{NSW}(u_1, \ldots, u_m) = \left(\prod_{i=1}^{m} u_i\right)^{\frac{1}{m}}$.
- Egalitarian Welfare - $\mathrm{EgW}(u_1, \ldots, u_m) = \min_{i \in [m]} u_i$.

Theorem: *NSW is the unique welfare function satisfying all the axioms.*

Axiomatic Basis

- **Pareto Efficiency**
- **Pigou-Dalton Principle**
- **Symmetry, Independece of Irrelevant Agent**
- **Scale-free**



					SW	NSW	EgW
Option 1	100	20	10	0	32.5	0	0
Option 2	45	25	25	32	31.75	30.8	25
Option 3	30	30	30	30	30	30	30

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1, \ldots, c\}$ as an agent participating in the process.

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.
- In other words, utility of agent ℓ for set S: $u_\ell(S) = \sum_{v \in S \,:\, \mathrm{atb}(v) = \ell} \mathrm{sim}(q, v)$

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.
- In other words, utility of agent ℓ for set S: $u_\ell(S) = \sum_{v \in S \,:\, \mathrm{atb}(v) = \ell} \mathrm{sim}(q, v)$
- Standard nearest neighbor search problem: $\max_{S \subseteq P,\ |S|=k} \frac{1}{c} \sum_{\ell=1}^{c} u_\ell(S)$

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.
- In other words, utility of agent ℓ for set S: $u_\ell(S) = \sum_{v \in S\,:\,\mathrm{atb}(v) = \ell} \mathrm{sim}(q, v)$
- Standard nearest neighbor search problem: $\max_{S \subseteq P,\ |S|=k} \frac{1}{c} \sum_{\ell=1}^{c} u_\ell(S)$

Social Welfare Maximization

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.
- In other words, utility of agent ℓ for set S: $u_\ell(S) = \sum_{v \in S \,:\, \mathrm{atb}(v) = \ell} \mathrm{sim}(q, v)$
- Standard nearest neighbor search problem: $\max_{S \subseteq P,\ |S|=k} \frac{1}{c} \sum_{\ell=1}^{c} u_\ell(S)$

Social Welfare Maximization

- Our formulation Nash Nearest Neighbor Search: $\max_{S \subseteq P,\ |S|=k} \left(\prod_{\ell=1}^{c} u_\ell(S) \right)^{\frac{1}{c}}$

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1,\ldots,c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.
- In other words, utility of agent ℓ for set S: $u_\ell(S) = \sum_{v \in S \,:\, \mathrm{atb}(v) = \ell} \mathrm{sim}(q, v)$
- Standard nearest neighbor search problem: $\max_{S \subseteq P,\ |S|=k} \frac{1}{c} \sum_{\ell=1}^{c} u_\ell(S)$

Social Welfare Maximization

- Our formulation Nash Nearest Neighbor Search: $\max_{S \subseteq P,\ |S|=k} \left(\prod_{\ell=1}^{c} u_\ell(S) \right)^{\frac{1}{c}}$

Nash Welfare Maximization

Our Approach to Diversity

Key Idea: View every attribute $\ell \in \{1, \ldots, c\}$ as an agent participating in the process.

- Agents derive utilities from returned set $S \subseteq P$, $|S| = k$.
- Each vector $v \in S$ with attribute $\ell \in [c]$ contributes $\mathrm{sim}(q, v)$ to utility of ℓ.
- In other words, utility of agent ℓ for set S: $u_\ell(S) = \sum_{v \in S \,:\, \mathrm{atb}(v) = \ell} \mathrm{sim}(q, v)$
- Standard nearest neighbor search problem: $\max_{S \subseteq P,\ |S|=k} \frac{1}{c} \sum_{\ell=1}^{c} u_\ell(S)$

Social Welfare Maximization

- Our formulation Nash Nearest Neighbor Search: $\max_{S \subseteq P,\ |S|=k} \left(\prod_{\ell=1}^{c} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$

Nash Welfare Maximization

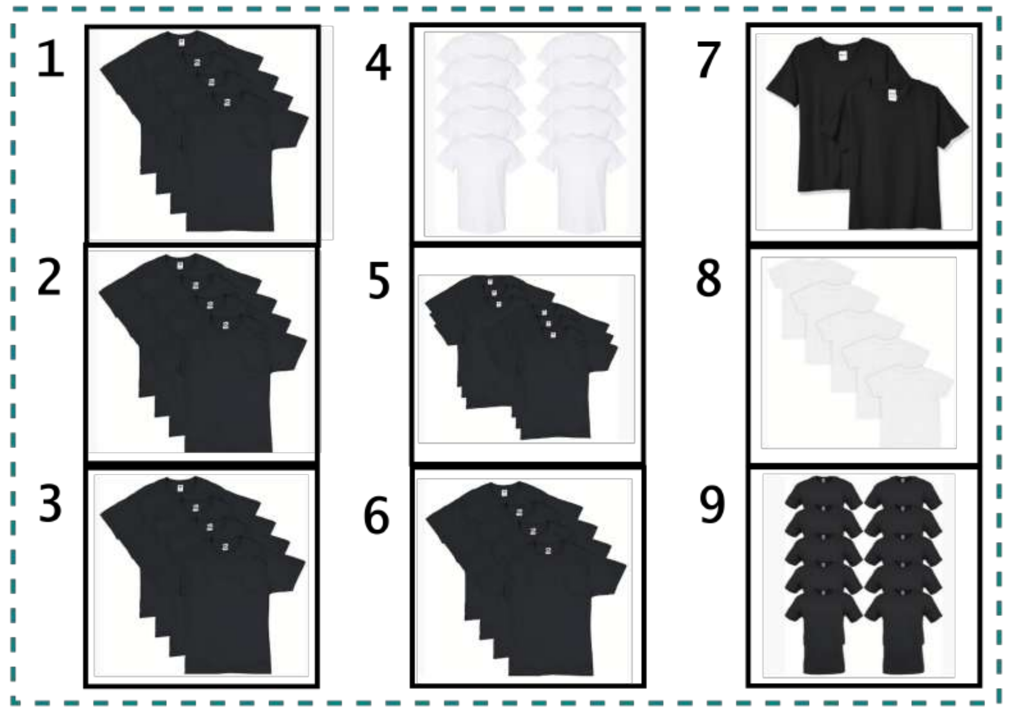
Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25] vs Nash ANN (Ours)

Standard Search

Constrained Search

Query = “Shirts”



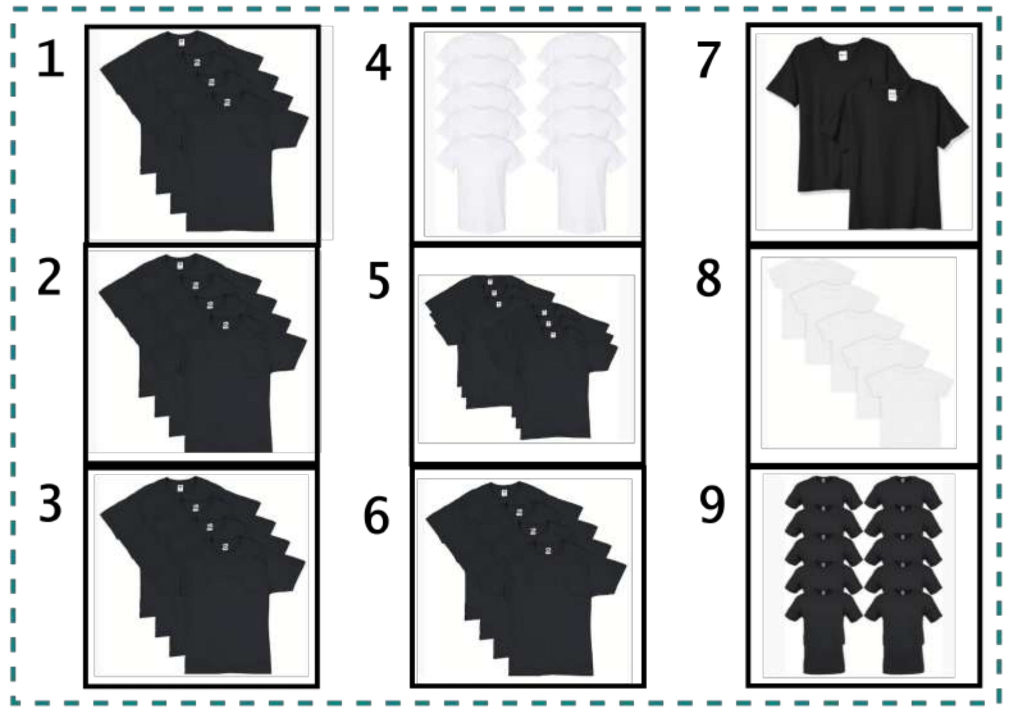
Query = “Blue Shirts”

Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25] vs Nash ANN (Ours)

Standard Search | Constrained Search | Nash Search (Ours)

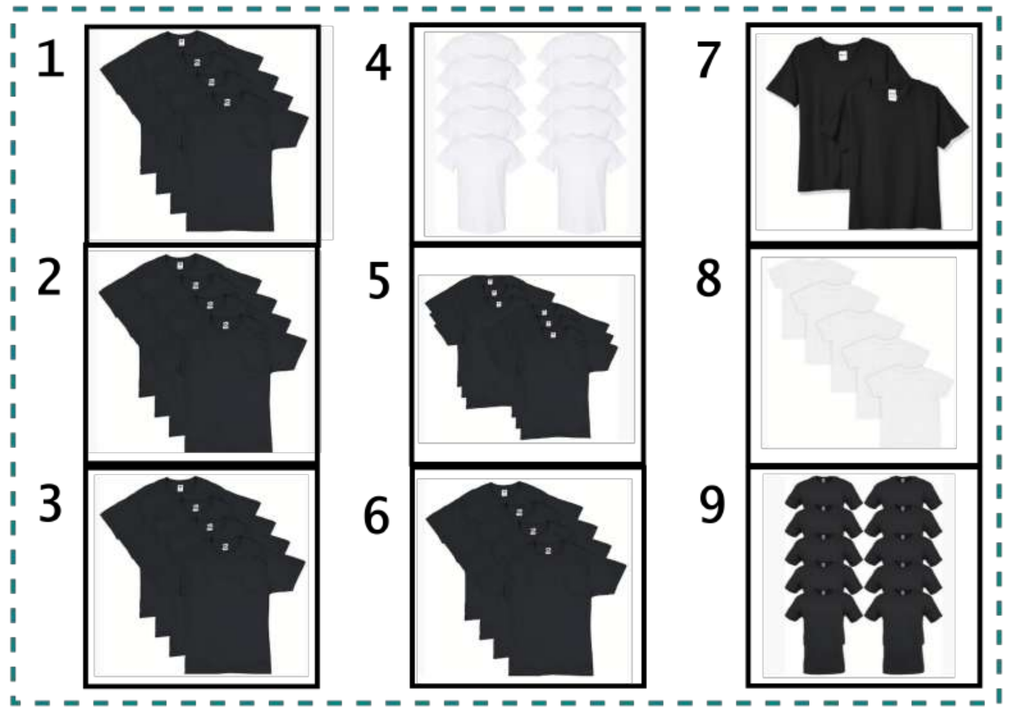
Query = “Shirts”



Query = “Blue Shirts”

Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25] vs Nash ANN (Ours)

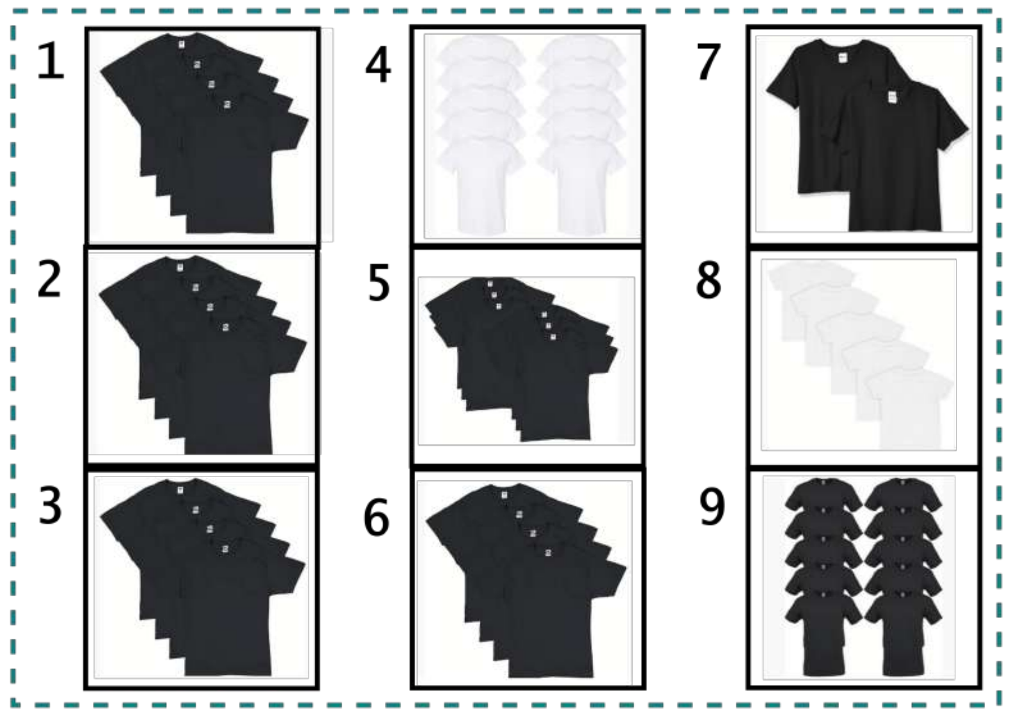
	Standard Search	Constrained Search	Nash Search (Ours)
Query = "Shirts"			
Query = "Blue Shirts"			

Diverse Nearest Neighbor Search

Constrained ANN [AIKMRSX25] vs Nash ANN (Ours)

Standard Search | Constrained Search | Nash Search (Ours)

Query = “Shirts”



Query = “Blue Shirts”

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.
3. Initialize $\mathsf{ALG} = \emptyset$.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.
3. Initialize $\mathsf{ALG} = \emptyset$.
4. While $|\mathsf{ALG}| < k$ do:

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.
3. Initialize $\mathsf{ALG} = \emptyset$.
4. While $|\mathsf{ALG}| < k$ do:
    1. Select $\widehat{\ell} \in [c]$ such that the next most similar vector $\widehat{v}$ in $\widehat{D}_{\widehat{\ell}}$ gives highest marginal increase to Nash Welfare.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.
3. Initialize $\mathsf{ALG} = \emptyset$.
4. While $|\mathsf{ALG}| < k$ do:
   1. Select $\widehat{\ell} \in [c]$ such that the next most similar vector $\widehat{v}$ in $\widehat{D}_{\widehat{\ell}}$ gives highest marginal increase to Nash Welfare.
   2. Update $\mathsf{ALG} \leftarrow \mathsf{ALG} \cup \{\widehat{v}\}$. Remove $\widehat{v}$ from $\widehat{D}_{\widehat{\ell}}$.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.
3. Initialize $\mathsf{ALG} = \emptyset$.
4. While $|\mathsf{ALG}| < k$ do:
   1. Select $\widehat{\ell} \in [c]$ such that the next most similar vector $\widehat{v}$ in $\widehat{D}_{\widehat{\ell}}$ gives highest marginal increase to Nash Welfare.
   2. Update $\mathsf{ALG} \leftarrow \mathsf{ALG} \cup \{\widehat{v}\}$. Remove $\widehat{v}$ from $\widehat{D}_{\widehat{\ell}}$.
5. Return ALG.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Algorithm in the Single-Attribute Setting

1. Input the query $q \in \mathbb{R}^d$.
2. For every $\ell \in [c]$, fetch $\widehat{D}_\ell$ containing k most similar vectors with attribute ℓ.
3. Initialize ALG $= \emptyset$.
4. While $|\text{ALG}| < k$ do:
   1. Select $\widehat{\ell} \in [c]$ such that the next most similar vector $\hat{v}$ in $\widehat{D}_{\widehat{\ell}}$ gives highest marginal increase to Nash Welfare.
   2. Update ALG $\leftarrow$ ALG $\cup \{\hat{v}\}$. Remove $\hat{v}$ from $\widehat{D}_{\widehat{\ell}}$.
5. Return ALG.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \text{atb}(v) \ni \ell}} \text{sim}(q, v)$$

Theorem: ALG $\in \text{argmax}_{S \subseteq P,\ |S|=k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$.

Nash Nearest Neighbor Search

Main Results

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Main Results

- **Theorem 1**: For single-attribute setting ($|\mathrm{atb}(v)| = 1 \, \forall \; v \in P$), the greedy algorithm finds exact optimal solution using an exact nearest neighbor search oracle.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Main Results

- **Theorem 1**: For single-attribute setting ($|\text{atb}(v)| = 1 \, \forall \, v \in P$), the greedy algorithm finds exact optimal solution using an exact nearest neighbor search oracle.
    - **Corollary**: With α-approximate nearest neighbor search oracle, an α-approximate solution can be obtained.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \text{atb}(v) \ni \ell}} \text{sim}(q, v)$$

Nash Nearest Neighbor Search

Main Results

- **Theorem 1**: For single-attribute setting ($|\mathrm{atb}(v)| = 1 \,\forall\, v \in P$), the greedy algorithm finds exact optimal solution using an exact nearest neighbor search oracle.
  - **Corollary**: With α-approximate nearest neighbor search oracle, an α-approximate solution can be obtained.
- Generalizes to multi-attribute setting: $\mathrm{atb}(v) \subseteq [c]$ for every $v \in P$.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

Main Results

- **Theorem 1**: For single-attribute setting ($|\text{atb}(v)| = 1 \,\forall\, v \in P$), the greedy algorithm finds exact optimal solution using an exact nearest neighbor search oracle.
  - **Corollary**: With α-approximate nearest neighbor search oracle, an α-approximate solution can be obtained.
- Generalizes to multi-attribute setting: $\text{atb}(v) \subseteq [c]$ for every $v \in P$.
- **Theorem 2**: For multi-attribute setting ($|\text{atb}(v)| \geq 1 \,\forall\, v \in P$), the problem is NP-Hard.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \text{atb}(v) \ni \ell}} \text{sim}(q, v)$$

Nash Nearest Neighbor Search

Main Results

- **Theorem 1**: For single-attribute setting ($|\text{atb}(v)| = 1 \,\forall\, v \in P$), the greedy algorithm finds exact optimal solution using an exact nearest neighbor search oracle.
  - **Corollary**: With α-approximate nearest neighbor search oracle, an α -approximate solution can be obtained.
- Generalizes to multi-attribute setting: $\text{atb}(v) \subseteq [c]$ for every $v \in P$.
- **Theorem 2**: For multi-attribute setting ($|\text{atb}(v)| \geq 1 \,\forall\, v \in P$), the problem is NP-Hard.
- **Theorem 3**: 4-approximation algorithm in multi-attribute setting.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \text{atb}(v) \ni \ell}} \text{sim}(q, v)$$

Nash Nearest Neighbor Search

Main Results

- **Theorem 1**: For single-attribute setting ($|\mathrm{atb}(v)| = 1\ \forall\ v \in P$), the greedy algorithm finds exact optimal solution using an exact nearest neighbor search oracle.
  - **Corollary**: With α-approximate nearest neighbor search oracle, an α-approximate solution can be obtained.
- Generalizes to multi-attribute setting: $\mathrm{atb}(v) \subseteq [c]$ for every $v \in P$.
- **Theorem 2**: For multi-attribute setting ($|\mathrm{atb}(v)| \geq 1\ \forall\ v \in P$), the problem is NP-Hard.
- **Theorem 3**: 4-approximation algorithm in multi-attribute setting.

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Key Features

- Works on top of any ANN implementation
- Fast principled heuristic versions

Nash Nearest Neighbor Search

Experimental Results

Dataset	Dataset Size	Query Vectors	Dimension	Attributes
Amazon	92,092	8,956	768	product color
Arxiv	2,00,000	50,000	1536	year, paper category
SIFT-1M	10,00,000	10,000	128	synthetic
Deep-1B	99,90,000	10,000	96	synthetic

Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} (u_\ell(S) + \eta)^{\frac{1}{c}}$$

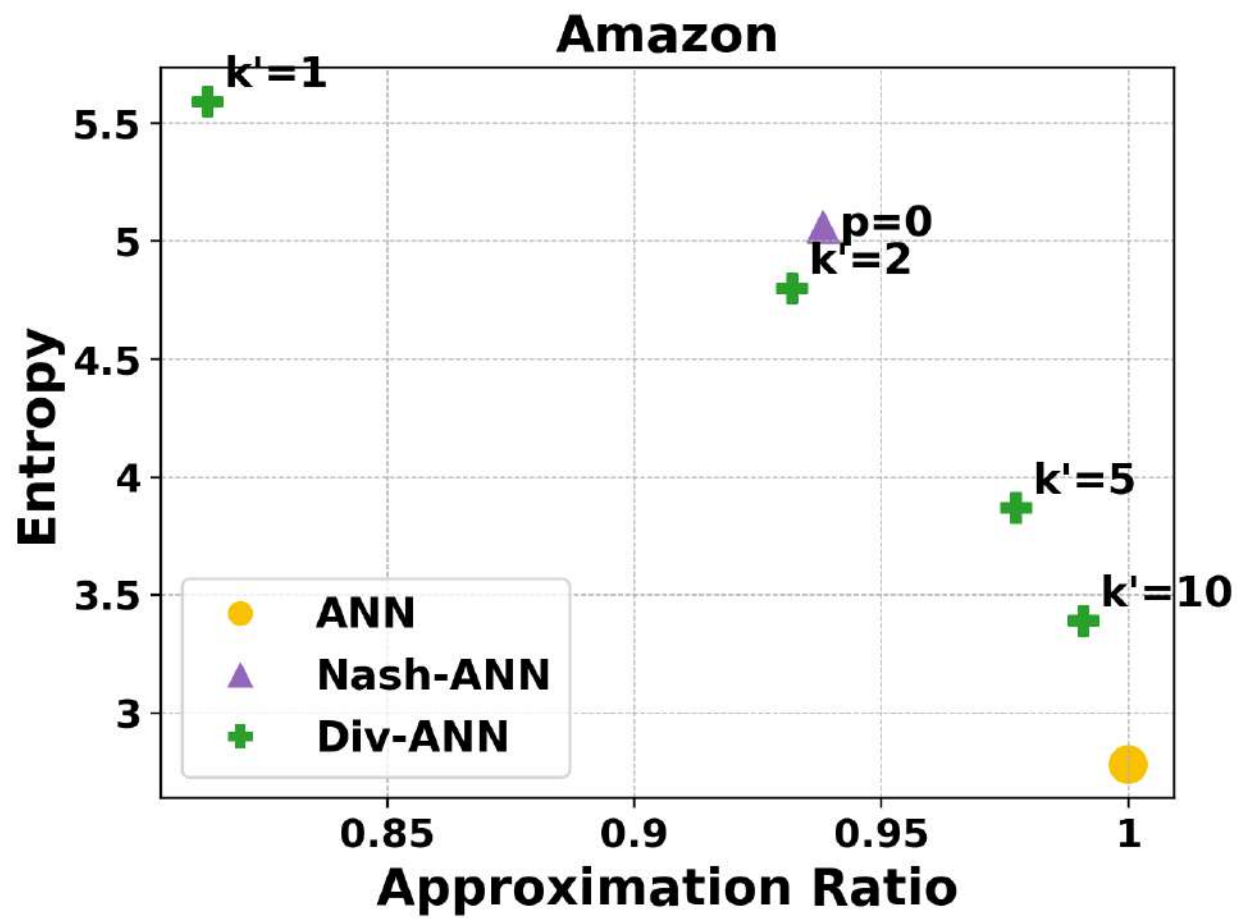
where

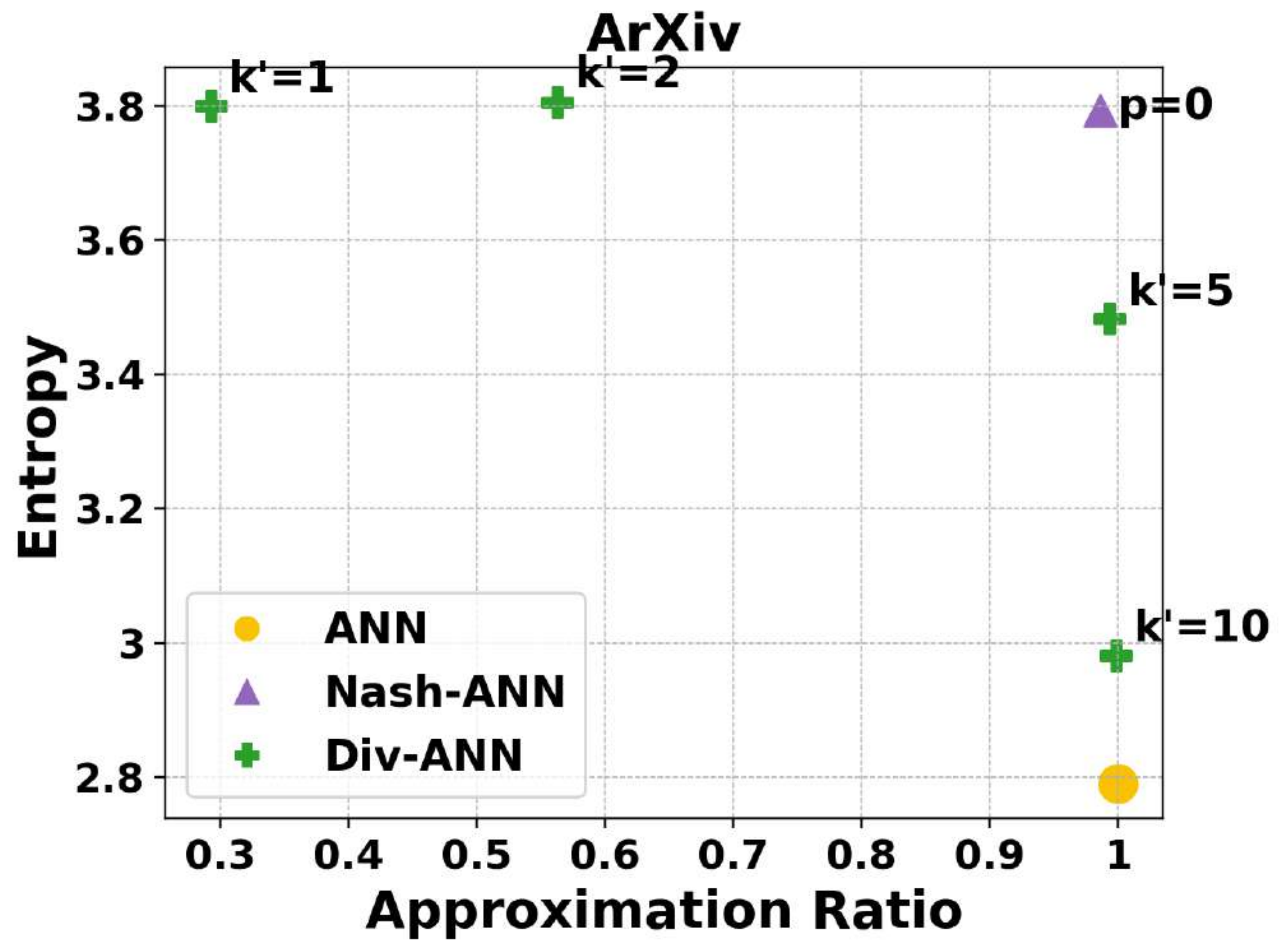
$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Nash Nearest Neighbor Search

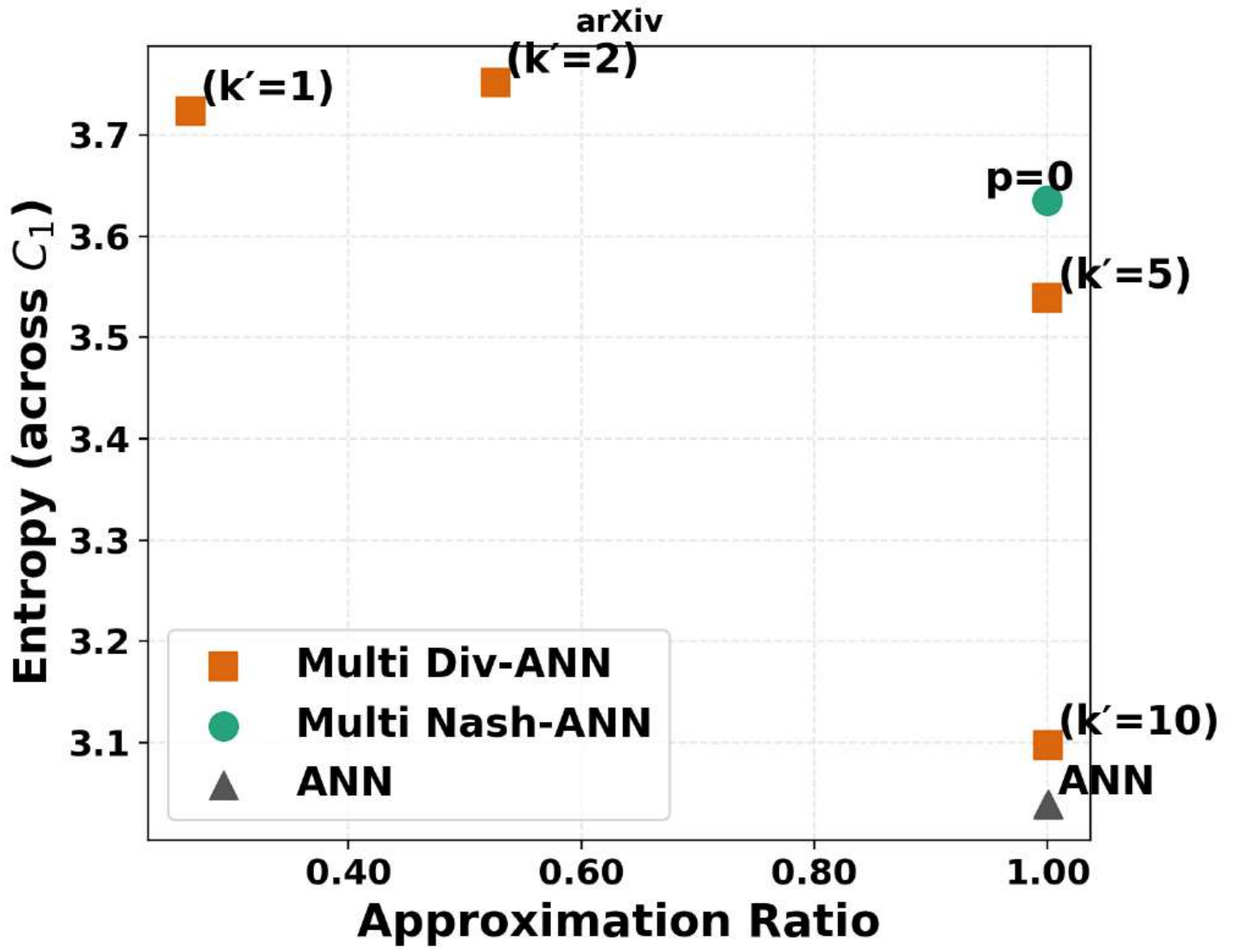
Experimental Results

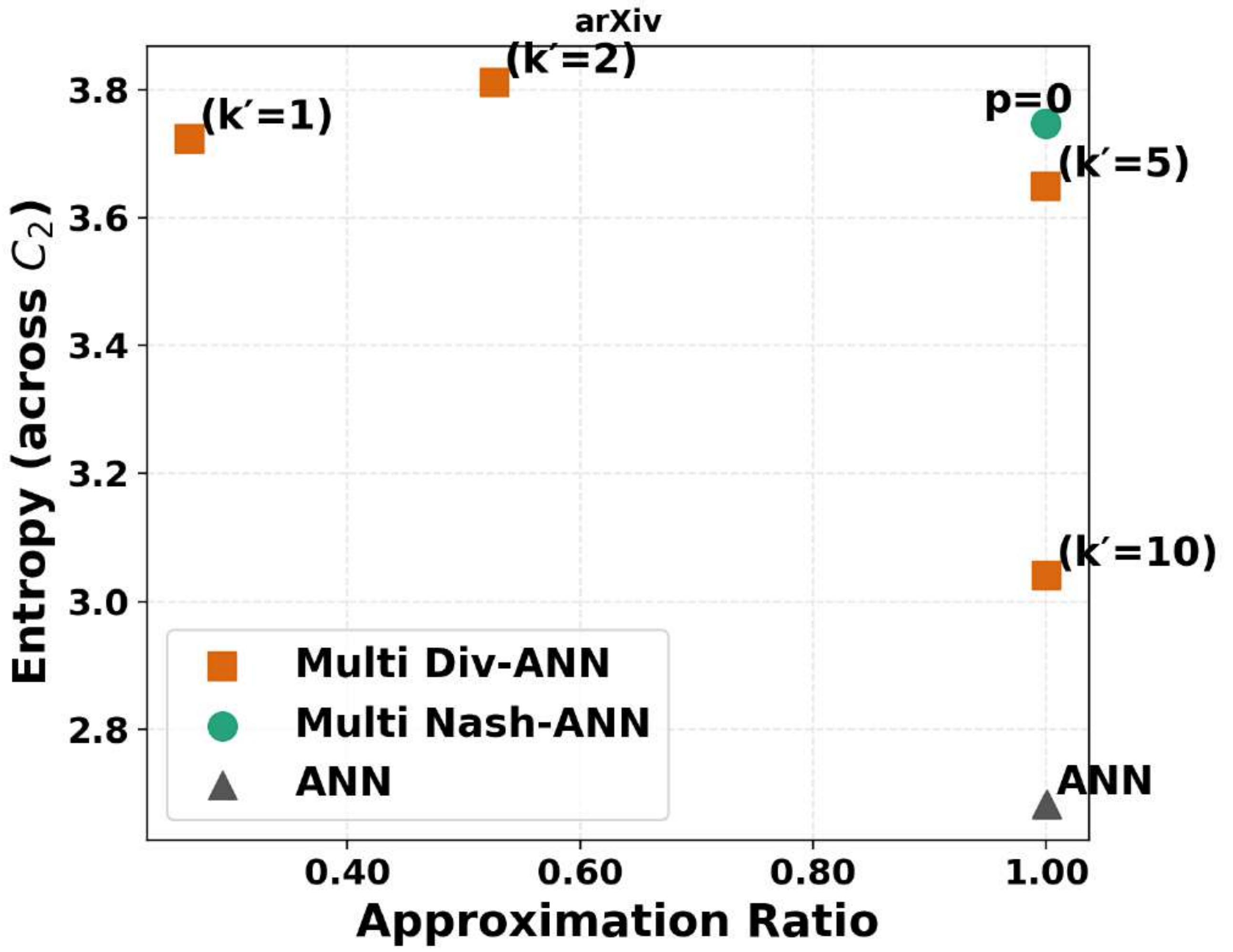
Single-Attribute





Multi-Attribute





Given $q \in \mathbb{R}^d$, find

$$\max_{S \subseteq P,\ |S| = k} \prod_{\ell \in [c]} \left(u_\ell(S) + \eta\right)^{\frac{1}{c}}$$

where

$$u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$$

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \text{atb}(v) \ni \ell}} \text{sim}(q, v)\,.$
- Social Welfare: $\text{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)\,.$
- Nash Social Welfare: $\text{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}.$
- Egalitarian Welfare: $\text{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta\,.$

$$\text{EgW}(S) \leq \text{NSW}(S) \leq \text{SW}(S)$$

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

$$\mathrm{EgW}(S) \leq \mathrm{NSW}(S) \leq \mathrm{SW}(S)$$

p-Mean Welfare

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

$$\mathrm{EgW}(S) \leq \mathrm{NSW}(S) \leq \mathrm{SW}(S)$$

p-Mean Welfare

For $p \in (-\infty, 1]$, define

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)\,.$
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)\,.$
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}.$
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta\,.$

$$\mathrm{EgW}(S) \le \mathrm{NSW}(S) \le \mathrm{SW}(S)$$

p-Mean Welfare

For $p \in (-\infty, 1]$, define

$$M_p(S) := \left(\frac{1}{c} \sum_{\ell=1}^{c} (u_\ell(S) + \eta)^p \right)^{\frac{1}{p}}$$

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

$$\mathrm{EgW}(S) \leq \mathrm{NSW}(S) \leq \mathrm{SW}(S)$$

p-Mean Welfare

For $p \in (-\infty, 1]$, define

$$M_p(S) := \left(\frac{1}{c} \sum_{\ell=1}^{c} (u_\ell(S) + \eta)^p \right)^{\frac{1}{p}}$$

- $p = 1$: Social Welfare

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

$$\mathrm{EgW}(S) \leq \mathrm{NSW}(S) \leq \mathrm{SW}(S)$$

p-Mean Welfare

For $p \in (-\infty, 1]$, define

$$M_p(S) := \left(\frac{1}{c} \sum_{\ell=1}^{c} (u_\ell(S) + \eta)^p \right)^{\frac{1}{p}}$$

- $p = 1$: Social Welfare
- $p \to 0$: Nash Social Welfare

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k$, we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

$$\mathrm{EgW}(S) \leq \mathrm{NSW}(S) \leq \mathrm{SW}(S)$$

p-Mean Welfare

For $p \in (-\infty, 1]$, define

$$M_p(S) := \left(\frac{1}{c} \sum_{\ell=1}^{c} (u_\ell(S) + \eta)^p \right)^{\frac{1}{p}}$$

- $p = 1$: Social Welfare
- $p \to 0$: Nash Social Welfare
- $p \to -\infty$: Egalitarian Welfare

Generalization of Nash NNS

p-Mean Welfare

- For $S \subseteq P, |S| = k,$ we have $u_\ell(S) = \sum_{\substack{v \in S \\ \mathrm{atb}(v) \ni \ell}} \mathrm{sim}(q, v)$.
- Social Welfare: $\mathrm{SW}(S) = \frac{1}{c} \sum_{\ell \in [c]} (u_\ell(S) + \eta)$.
- Nash Social Welfare: $\mathrm{NSW}(S) = \left(\prod_{\ell \in [c]} (u_\ell(S) + \eta) \right)^{\frac{1}{c}}$.
- Egalitarian Welfare: $\mathrm{EgW}(S) = \min_{\ell \in [c]} u_\ell(S) + \eta$.

$$\mathrm{EgW}(S) \leq \mathrm{NSW}(S) \leq \mathrm{SW}(S)$$

p-Mean Welfare

For $p \in (-\infty, 1]$, define

$$M_p(S) := \left(\frac{1}{c} \sum_{\ell=1}^{c} (u_\ell(S) + \eta)^p \right)^{\frac{1}{p}}$$

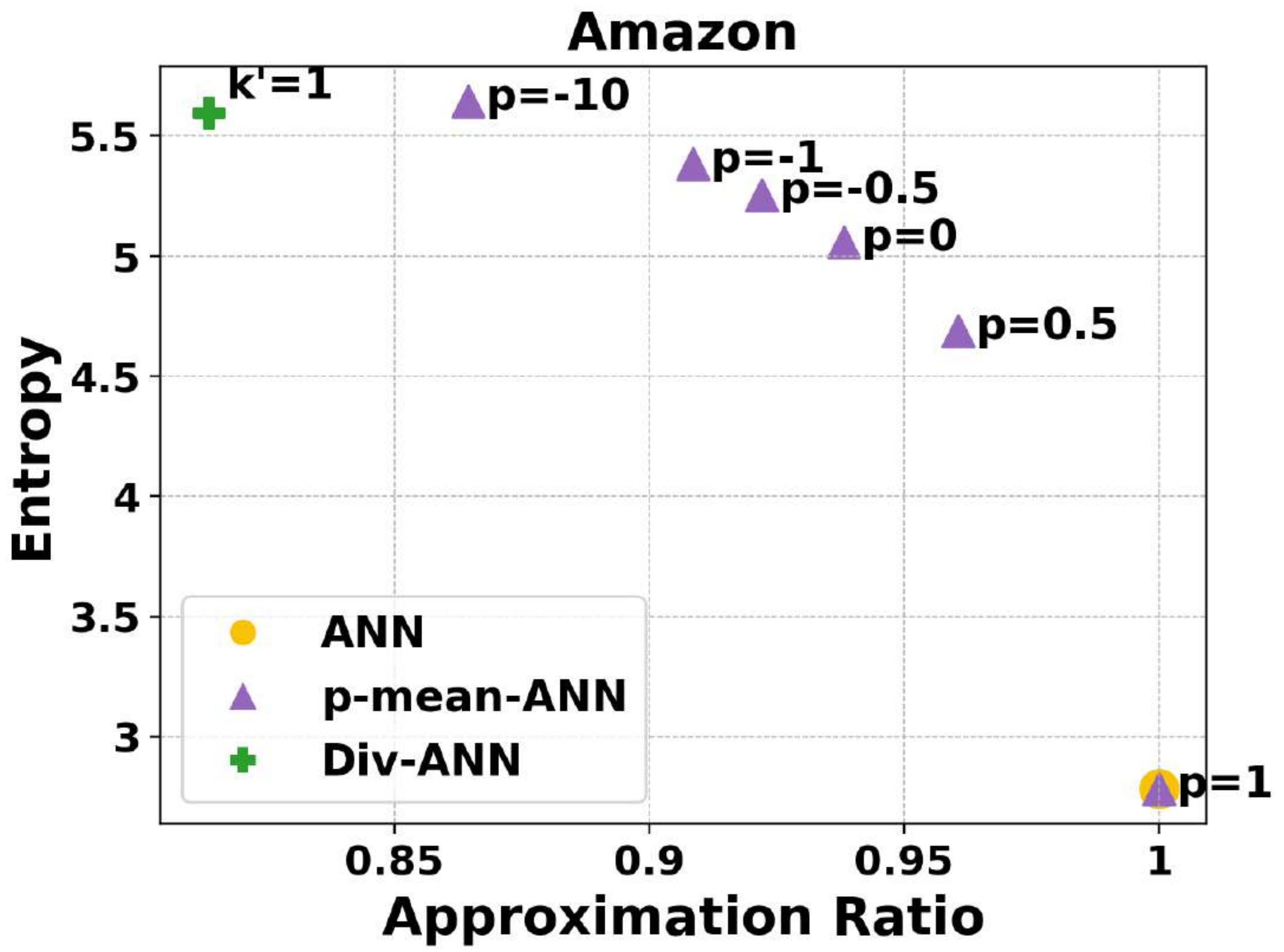
- $p = 1$: Social Welfare
- $p \to 0$: Nash Social Welfare
- $p \to -\infty$: Egalitarian Welfare

New Objective: $\max_{S \subseteq P, |S| = k} M_p(S)$

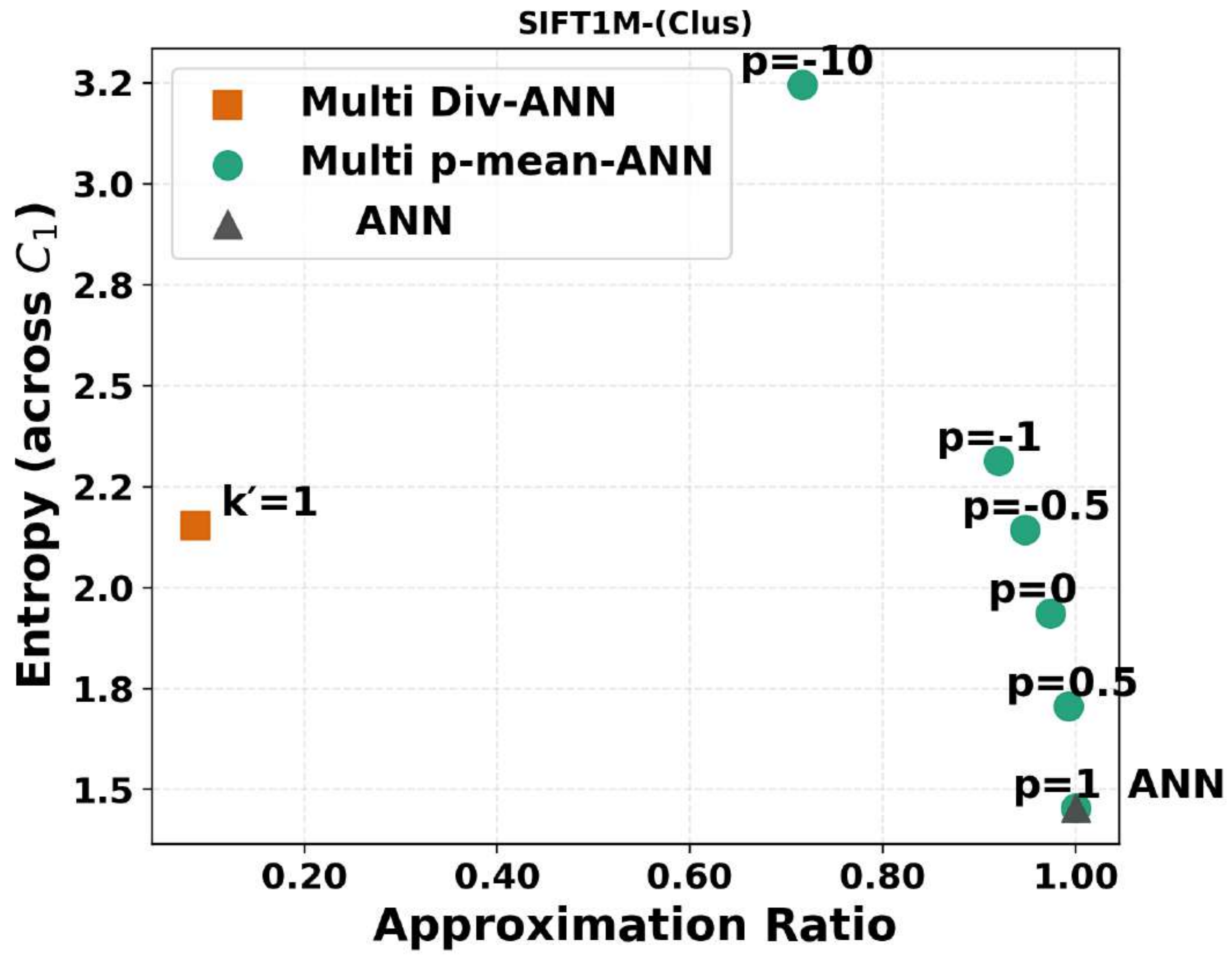
Generalization of Nash NNS

p-Mean Welfare

Single-Attribute



Multi-Attribute



p-Mean Welfare

For $p \in (-\infty, 1]$, define

$$M_p(S) := \left(\frac{1}{c} \sum_{\ell=1}^{c} (u_\ell(S) + \eta)^p \right)^{\frac{1}{p}}$$

- $p = 1$: Social Welfare
- $p \to 0$: Nash Social Welfare
- $p \to -\infty$: Egalitarian Welfare

New Objective: $\max_{S \subseteq P, |S| = k} M_p(S)$

Fast Heuristic Implementation

Metric	Algorithm	p = -10	p = -1	p = -0.5	p = 0	p = 0.5	p = 1
Approx. Ratio	p-Mean-ANN	0.749±0.051	0.810±0.045	0.812±0.043	0.846±0.036	0.932±0.028	1.000±0.000
	p-FetchUnion-ANN	0.979±0.014	0.980±0.013	0.980±0.013	0.981±0.012	0.983±0.011	1.000±0.000
	ANN	1.000±0.000					
	Constrained (k' = 1)	0.315±0.021					
Entropy	p-Mean-ANN	4.285±0.012	4.293±0.002	4.293±0.001	4.197±0.045	3.506±0.275	0.892±0.663
	p-FetchUnion-ANN	2.235±0.802	2.238±0.802	2.239±0.802	2.239±0.802	2.231±0.800	0.892±0.663
	ANN	0.892±0.663					
	Constraiend (k' = 1)	4.289±0.053					

Metric	Algorithm	p = -10	p = -1	p = -0.5	p = 0	p = 0.5	p = 1
Query per Second	p-Mean-ANN	120.86	115.78	107.01	135.98	122.59	122.59
	p-FetchUnion-ANN	1324.53	1324.62	1337.28	1442.03	1443.38	1327.03
Latency (μs)	p-Mean-ANN	264566.00	276129.00	298804.00	230318.00	235144.00	260800.00
	p-FetchUnion-ANN	24133.80	24134.00	23907.00	22170.20	22149.30	28990.40
99.9th percentile of Latency	p-Mean-ANN	484601.00	513036.00	478821.00	477925.00	482777.00	479132.00
	p-FetchUnion-ANN	52943.40	53474.70	54283.40	56128.70	53082.20	24088.70

Conclusion

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.
- Smooth tradeoff between relevance and diversity via p-mean Welfare.

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.
- Smooth tradeoff between relevance and diversity via p-mean Welfare.
- Heuristic that can be implemented on top of any existing ANN pipeline.

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.
- Smooth tradeoff between relevance and diversity via p-mean Welfare.
- Heuristic that can be implemented on top of any existing ANN pipeline.
- Future work:

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.
- Smooth tradeoff between relevance and diversity via p-mean Welfare.
- Heuristic that can be implemented on top of any existing ANN pipeline.
- Future work:
  - A custom data structure and search algorithm for Nash Nearest Neighbour Search.

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.
- Smooth tradeoff between relevance and diversity via p-mean Welfare.
- Heuristic that can be implemented on top of any existing ANN pipeline.
- Future work:
  - A custom data structure and search algorithm for Nash Nearest Neighbour Search.
  - Generalization to the setting where attributes are not explicitly available.

Conclusion

- Principled approach to Diversity in Nearest Neighbor Search via welfare economics perspective.
- Nash Welfare strikes a balance between relevance and diversity, both in theory and practice.
- Smooth tradeoff between relevance and diversity via p-mean Welfare.
- Heuristic that can be implemented on top of any existing ANN pipeline.
- Future work:
  - A custom data structure and search algorithm for Nash Nearest Neighbour Search.
  - Generalization to the setting where attributes are not explicitly available.

Thank You!

https://arxiv.org/abs/2602.08742