SECTION A - Architecture Solutions

1. Intent Extraction Strategy

1.1 Pipeline:

Tokenization/NER: Use lightweight NLP (spaCy) for basic entity recognition (product types, colors, sizes, prices, locations).

Rule-Based Matching: Regex patterns for explicit filters (e.g., /\bunder ₹?(\d+)/i → price_max=2000).

ML Classifiers: Fine-tuned BERT for ambiguous terms (e.g., "running" classified as "sport" using product taxonomy).

Attribute Mapping: Dictionary lookup for canonical attributes (e.g., "size" → shoe_size, "red" → color).

1.2 Fallback:

Unclassified terms default to full-text search (Elasticsearch). UI displays "Search instead for: running shoes" with partial filters applied.

2. Flexible Schema for Attributes

2.1 Data Model: Hybrid approach (relational core + document extensions):

Relational: Fixed fields (ID, SKU, category, price) in PostgreSQL.

Document (NoSQL): Dynamic attributes (e.g., {"energy_rating": "A++"}) stored in Elasticsearch or as PostgreSQL JSONB.

Why? Merchandisers add attributes via admin UI; no schema migrations. New attributes auto-indexed in search engines.