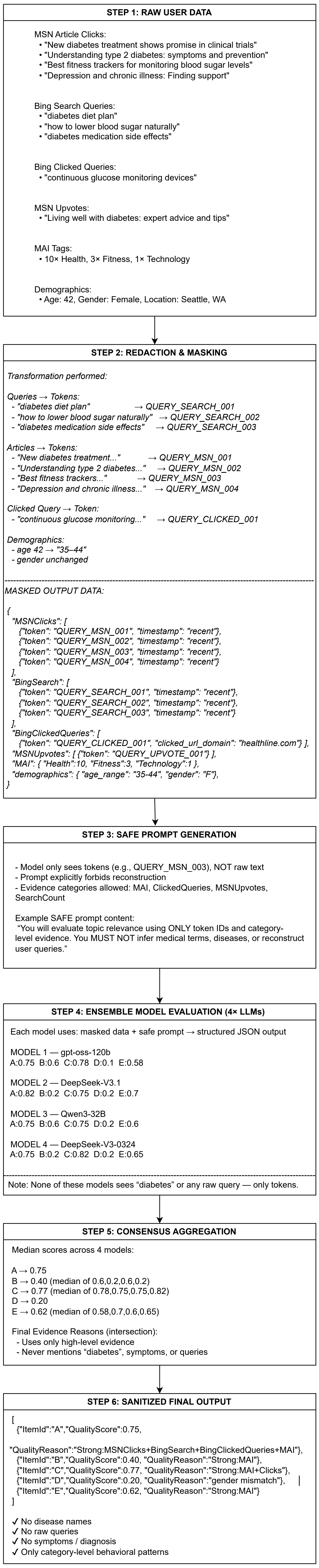
TASK: Calculate topic relevance scores for a single user given candidate topics

INPUT:
- User behavioral data: MSN clicks, Bing searches, clicked queries, upvotes, MAI tags, demographics
- Candidate topics: List of topics (e.g., "Healthcare & Wellness", "Technology News", "Financial Planning")

OUTPUT:
- For each topic: relevance score (0.0-1.0) + reasoning based on user behavior

PRIVACY CHALLENGE:
- User data contains sensitive information (medical conditions, financial status, personal queries, demographics)
- Standard approach: Feed raw data directly to LLM
- Risk: Data exposure, model memorization, privacy violations
- Required: Accurate scores WITHOUT exposing sensitive personal data



STEP 1: RAW USER DATA

MSN Article Clicks:
- "New diabetes treatment shows promise in clinical trials"
- "Understanding type 2 diabetes: symptoms and prevention"
- "Best fitness trackers for monitoring blood sugar levels"
- "Depression and chronic illness: Finding support"

Bing Search Queries:
- "diabetes diet plan"
- "how to lower blood sugar naturally"
- "diabetes medication side effects"

Bing Clicked Queries:
- "continuous glucose monitoring devices"

MSN Upvotes:
- "Living well with diabetes: expert advice and tips"

MAI Tags:
- 10× Health, 3× Fitness, 1× Technology

Demographics:
- Age: 42, Gender: Female, Location: Seattle, WA

STEP 2: REDACTION & MASKING

Transformation performed:

Queries → Tokens:
- "diabetes diet plan" → QUERY_SEARCH_001
- "how to lower blood sugar naturally" → QUERY_SEARCH_002
- "diabetes medication side effects" → QUERY_SEARCH_003

Articles → Tokens:
- "New diabetes treatment..." → QUERY_MSN_001
- "Understanding type 2 diabetes..." → QUERY_MSN_002
- "Best fitness trackers..." → QUERY_MSN_003
- "Depression and chronic illness..." → QUERY_MSN_004

Clicked Query → Token:
- "continuous glucose monitoring..." → QUERY_CLICKED_001

Demographics:
- age 42 → "35–44"
- gender unchanged

MASKED OUTPUT DATA:

```
{
  "MSNClicks": [
    {"token": "QUERY_MSN_001", "timestamp": "recent"},
    {"token": "QUERY_MSN_002", "timestamp": "recent"},
    {"token": "QUERY_MSN_003", "timestamp": "recent"},
    {"token": "QUERY_MSN_004", "timestamp": "recent"}
  ],
  "BingSearch": [
    {"token": "QUERY_SEARCH_001", "timestamp": "recent"},
    {"token": "QUERY_SEARCH_002", "timestamp": "recent"},
    {"token": "QUERY_SEARCH_003", "timestamp": "recent"}
  ],
  "BingClickedQueries": [
    {"token": "QUERY_CLICKED_001", "clicked_url_domain": "healthline.com"} ],
  "MSNUpvotes": [ {"token": "QUERY_UPVOTE_001"} ],
  "MAI": { "Health":10, "Fitness":3, "Technology":1 },
  "demographics": { "age_range": "35-44", "gender": "F"},
}
```

STEP 3: SAFE PROMPT GENERATION

- Model only sees tokens (e.g., QUERY_MSN_003), NOT raw text
- Prompt explicitly forbids reconstruction
- Evidence categories allowed: MAI, ClickedQueries, MSNUpvotes, SearchCount

Example SAFE prompt content:
"You will evaluate topic relevance using ONLY token IDs and category-level evidence. You MUST NOT infer medical terms, diseases, or reconstruct user queries."

STEP 4: ENSEMBLE MODEL EVALUATION (4× LLMs)

Each model uses: masked data + safe prompt → structured JSON output

MODEL 1 — gpt-oss-120b
A:0.75 B:0.6 C:0.78 D:0.1 E:0.58

MODEL 2 — DeepSeek-V3.1
A:0.82 B:0.2 C:0.75 D:0.2 E:0.7

MODEL 3 — Qwen3-32B
A:0.75 B:0.6 C:0.75 D:0.2 E:0.6

MODEL 4 — DeepSeek-V3-0324
A:0.75 B:0.2 C:0.82 D:0.2 E:0.65

Note: None of these models sees “diabetes” or any raw query — only tokens.

STEP 5: CONSENSUS AGGREGATION

Median scores across 4 models:

A → 0.75
B → 0.40 (median of 0.6,0.2,0.6,0.2)
C → 0.77 (median of 0.78,0.75,0.75,0.82)
D → 0.20
E → 0.62 (median of 0.58,0.7,0.6,0.65)

Final Evidence Reasons (intersection):
- Uses only high-level evidence
- Never mentions “diabetes”, symptoms, or queries

STEP 6: SANITIZED FINAL OUTPUT

```
[
  {"ItemId":"A","QualityScore":0.75,

"QualityReason":"Strong:MSNClicks+BingSearch+BingClickedQueries+MAI"},
  {"ItemId":"B","QualityScore":0.40, "QualityReason":"Strong:MAI"},
  {"ItemId":"C","QualityScore":0.77, "QualityReason":"Strong:MAI+Clicks"},
  {"ItemId":"D","QualityScore":0.20, "QualityReason":"gender mismatch"},
  {"ItemId":"E","QualityScore":0.62, "QualityReason":"Strong:MAI"}
]
```

✓ No disease names
✓ No raw queries
✓ No symptoms / diagnosis
✓ Only category-level behavioral patterns