You are an AI model that analyzes an image and outputs only the estimated age and gender of the main person in the image.
Gender must be either "male" or "female".
Do not include any reasoning, explanation, or extra words.

Analyze the provided image and return only the estimated age and gender of the main person in the format: {age}, {gender}.
Gender must be either "male" or “female".
Example: 61, male

Figure A1. The prompt template used to generate the age and gender of concepts in PerVL-Bench.

Based on the provided age, gender, and persona, generate a realistic and consistent set of attributes for personality traits and preferences.
Preferences should be written in complete sentences and may include information not directly related to the given persona, covering both general likes/dislikes and everyday considerations such as dietary restrictions or allergies.

Age: {age}
Gender: {gender}
Persona: {persona}

Output Format:
- Personality: [a few descriptive words]
- Preferences: [one or more complete sentences describing general likes/dislikes and everyday considerations, which may not be directly related to the persona]

Figure A2. The prompt template used to generate the personality and preferences of concepts in PerVL-Bench.

Based on the provided persona, generate a realistic and consistent set of attributes, including age, gender, personality traits, and preferences.
Preferences should be written in complete sentences and may include information not directly related to the given persona, covering both general likes/dislikes and everyday considerations such as dietary restrictions or allergies.

Persona: {persona}

Output Format:
- Age: [number]
- Gender: [male/female]
- Personality: [a few descriptive words]
- Preferences: [one or more complete sentences describing general likes/dislikes and everyday considerations, which may not be directly related to the persona]

Figure A3. The prompt template used to generate the metadata of the user in PerVL-Bench.

You are a dataset generator that creates factual, memory-like paragraphs for a personalized AI system. Generate one cohesive paragraph per concept based on the given user profile (context only) and concept profiles (name, age, gender, category, personality, preferences).

Rules:
1) Output only concept paragraphs; 1 paragraph per concept.
2) Do not explicitly state profile fields; imply them through behaviors, relationships, or events.
3) Keep the user as the central subject in a neutral tone.
4) Each paragraph must include: history with the user, concrete past events, cautions, memorable facts, and cross-concept references.
5) Include multiple elements shared with various other concepts (e.g., events, preferences, habits, personality traits) spread naturally throughout.
6) Be realistic, logically consistent, and prose only (no lists).

Length: 8–10 sentences per concept.

Generate textual information according to the rules above.

User profile
{metadata of the user}

Concept profiles
{metadata of the concepts}

Output ONLY in the following format and nothing else:
[Textual Information]

<Concept Name>:
<one single paragraph>

<Concept Name>:
<one single paragraph>

Figure A4. The prompt template used to generate user-specific textual information in PerVL-Bench.

You are a dataset generator for evaluating Personalized Multimodal Large Language Models (MLLMs).
Generate QA pairs about behavioral traits, habits, preferences, or cautions of concepts in the input image.

RULES:
1) Inputs:
   - Personalized textual memory for multiple concepts.
   - List of concepts in the image.
2) For every QA pair, choose a trait that applies to at least one concept IN the image and at least one concept NOT in the image.
3) "answer": concept name(s) chosen ONLY from the image list (non-empty).
4) Questions must:
   - Be concise.
   - Include the phrase "in the input image" (or equivalent).
   - Mention only concepts in the image, but use traits that can also apply to concepts not in the image.
5) Use only the provided memory (no hallucination).

Output format (number from 1; add one blank line between items):
1.
question: …
answer: [concept_name_1, …]

Generate up to 6 QA pairs following the above rules described in the system prompt.
Ensure that each QA pair fully satisfies the conditions.
If you cannot generate more QA pairs that meet all conditions, stop early without filling the remaining count.

Concept-specific textual information
{textual information about the concepts}

Concepts in the image
{a list of concepts present in the query image}

Output ONLY in the specified format.

Figure A5. The prompt template used to generate Text-prompt QA in PerVL-Bench.

You are a dataset generator for evaluating Personalized Multimodal Large Language Models (MLLMs).
Create QA pairs where questions mention one or more concept names from the given image list, and answers are free-form text derived only from the provided textual memory.

Rules:
1. Mention only concepts from the image list.
2. When possible, vary concept mentions across QAs:
   - Use a diverse number of concepts (e.g., 1, 2, 3+).
   - Use different concept combinations for different QAs.
3. Frame questions so they can be easily confused with concepts not in the image, based on similarities in the textual memory.
4. Questions must be answerable only from the textual memory, focusing on events, habits, preferences, or shared activities.
5. Answers must be free-form (e.g., "soccer") and based solely on the textual memory.

Output format (number from 1; add one blank line between items):
1.
question: …
answer: …

Generate up to 6 QA pairs following the above rules described in the system prompt.
Ensure that each QA pair fully satisfies the conditions.
If you cannot generate more QA pairs that meet all conditions, stop early without filling the remaining count.

Concept-specific textual information
{textual information about the concepts}

Concepts in the image
{a list of concepts present in the query image}

Output ONLY in the specified format.

Figure A6. The prompt template used to generate Multimodal-prompt QA in PerVL-Bench.

You are a helpful and precise evaluator for a personalized AI system.

Evaluate two responses (A and B) to a given question based on the following criteria:
1. Personalization: Does the response effectively consider the user's provided personalized memory?
2. Helpfulness: Does the response properly address the user's question in a relevant way?

For each criterion, provide a brief one-line comparison of the two responses and select the better response (A, B, or Tie).

- Ensure the comparisons are concise and directly address the criteria.
- If both answers are equally strong or weak in a category, mark it as a Tie.

Inputs:
- Personalized memory: a set of records, each describing one concept in the following format:
  <concept_name>
  concept's information
- Original question: the user's initial natural language query.
- Modified question: a reformulated version of the original question that replaces certain names with visual prompts (e.g., 'red rectangle') to evaluate visual prompting understanding.
- Answer (A): the response produced by Assistant 1.
- Answer (B): the response produced by Assistant 2.

Personalized memory
{textual information about the concepts}

Original question
{original question}

Modified question
{modified question}

Answer (A)
{gt}

Answer (B)
{model's response}

Output Format:
1. Personalization: [Brief comparison of A and B]
2. Helpfulness: [Brief comparison of A and B]
Better Response: [A/B/Tie]

Figure A7. The prompt template used to measure GPT-Score on Multimodal-prompt QA in PerVL-Bench.

You are a question-answering system for Personalized Multimodal LLMs.
Your job is to answer the given question using only the provided concept bank.

Inputs
- Query image: one image to analyze.
- Concept bank: a set of concepts.
  - Each concept is provided in the following format:
    <{name}>
    {concept's image}
    {concept's information}

Task
Answer the given question by comparing the query image with the concept bank and using both the visual and textual information for each concept.

Rules
1) Use only the provided images, names, and textual information. No external knowledge or assumptions.
2) If multiple concepts are relevant, include all of them. If none are relevant, return an empty list.
3) Output only the exact `name` values from the concept bank; preserve spelling and case.
4) Do not guess: if uncertain about a concept, exclude it.
5) Strict format — no explanation, no extra words, no reasoning.

Output Format (must match exactly)
[name_1, name_2, …]
- If none: []

Query image
{query image}

Concepts
{user-specific data}

Instruction
Using the provided concept bank (name, image, and textual information for each concept) and the query image, answer the given question by identifying all relevant concepts.
Return only the names of the concepts from the concept bank that satisfy the question, following the required output format: [name_1, name_2, …]

Question
{question}

Figure A8. The prompt template used to perform inference on Text-prompt QA in PerVL-Bench.

You are a question-answering system for Personalized Multimodal LLMs.
Your job is to answer the given question using only the provided query image and concept bank.

Inputs
- Query image: an image to be analyzed that contains one or more visual prompt(s) (e.g., (red point), (blue rectangle), (green scribble), etc.).
- Concept bank: a set of concepts.
  - Each concept is provided in the following format:
    <{name}>
    {concept's image}
    {concept's information}

Task
Answer the given question by comparing the query image with the concept bank and using both the visual and textual information for each concept.

Rules
1) Use only the provided images, names, and textual information. No external knowledge or assumptions.
2) Do not guess — if uncertain about a concept, exclude it.
3) Strict format — no explanation, no extra words, no reasoning.

Output Format (must match exactly)
answer: …

Query image
{query image}

Concepts
{user-specific data}

Instruction
Using the provided concept bank (name, image, and textual information for each concept) and the query image, answer the given question.

Question
{question}

Figure A9. The prompt template used to perform inference on Multimodal-prompt QA in PerVL-Bench.