A. Technical Appendices and Supplementary Material

This appendix presents the verbatim prompt templates evaluated in our zero-shot experiments. We explore three distinct prompt engineering strategies: Prompt 1 establishes a baseline structured forensic protocol inspired by NISTIR 8584; Prompt 2 introduces extended structural complexity, which proved counter-productive due to increased hallucinations; and Prompt 3 offers an optimized refinement of the semantic scoring guide that emphasizes probabilistic analysis to achieve the highest detection accuracy.

A.1. Prompt 1

Listing 1. Prompt 1: Structured Forensic Analysis – Semantic Guide A

```
["""# Multimodal Morphing Attack Detection
     Prompt

## Task
You are a multimodal AI system with both
    **vision and language understanding**.
     You are given an image of a face and
    your task is to determine whether it
    is a **bona fide (authentic)** face or
     a **morphing attack** (a digitally
    blended face of two people).

After each step of your analysis, you must
     provide a **confidence score** on a
    strict scale from **0.000 to
    10000.000**, where:

- `0.000` means **definitely bona fide**
- `10000.000` means **definitely morph**
- Values in between reflect **fine-grained
     probability** of a morphing attack
- 0-1000: Strong evidence of authentic
    face
- 1000-3000: Likely authentic with minor
    irregularities
- 3000-7000: Uncertain, requires careful
    analysis
- 7000-9000: Likely morphed with moderate
    evidence
- 9000-10000: Strong evidence of morphing
    attack

> **Important:** Avoid rounded or generic
    values like `1000`, `5000`, `8000`.
    Your score must be **precise**, with
    at least **three decimal places**.
    This is essential for biometric
    performance analysis and threshold
    calibration.

Use the **step-by-step visual analysis**
    outlined below. After each step,
    provide a confidence score in the
    specified format, based on your
    observations and the responses to the
    guiding questions.

## Visual Analysis Steps

### Step 1: Core Facial Features
- Focus on the **eyes, nose, lips, and
    eyebrows**.
- Look for signs of **ghosting**, **faint
    duplicates**, or misaligned or
    unnatural elements.
- Check if eye contours or lip lines
    appear duplicated or semi-transparent.

**Ask**: "Do facial features have any
    doubled contours or blended boundaries
    ?"
**Ask**: "Do the eyes appear blurred or
    duplicated?"
**Ask**: "Do the lips show any visual
    artifacts, or are the lip lines
    irregular?"

After this step, provide a confidence
    score for Step 1 in this format:
```json
{"step1_score": [0 to 10000]}
```

Step 2: Facial Geometry and Symmetry
- Visually compare the **left and right
    halves** of the face.
- Detect any asymmetry in shape, spacing,
    or size of eyes, irises, ears, and
    jawline.
- Assess if the overall geometry seems
    subtly misaligned or "averaged."

Ask: "Do the facial proportions look
    unnaturally blended or off-balance?"

After this step, provide a confidence
    score for Step 2 in this format:
```json
{"step2_score": [0 to 10000]}
```

Step 3: Skin Texture and Detail
- Inspect the **skin surface** for fine
    detail.
- Detect over-smoothness, uniform skin
    tone, or "plastic-like" appearance.
- Check if pores, wrinkles, or blemishes
    are abnormally absent or symmetric.

Ask: "Does the skin look too perfect,
    synthetic, or even-textured?"
Ask: "Are pores, wrinkles, or
    blemishes abnormally absent or
    symmetric?"
```

```
After this step, provide a confidence
    score for Step 3 in this format:
'''json
{"step3_score": [0 to 10000]}
'''

### Step 4: Image Boundary and Hairline
- Look at the **face boundary**, including
     **ears, hairline, and background**.
- Detect any blurred transitions, blending
     seams, or edge mismatches.
- Check for faded or semi-transparent
    features outside the main face.

**Ask**: "Do any facial borders blend
    unnaturally into the background?"
**Ask**: "Are there any faded or semi-
    transparent elements present in the
    image?"
**Ask**: "Do ears, hairline, and
    background have any faded or semi-
    transparent artifacts?"

After this step, provide a confidence
    score for Step 4 in this format:
'''json
{"step4_score": [0 to 10000]}
'''

### Step 5: Lighting and Color Consistency
- Examine **lighting direction,
    reflections, and shadows**.
- Identify **saturation anomalies** (e.g.,
     bright red patches) or abnormal
    gradients.
- Compare lighting consistency across both
     sides of the face.

**Ask**: "Are there unnatural color shifts
     or inconsistent lighting effects?"

After this step, provide a confidence
    score for Step 5 in this format:
'''json
{"step5_score": [0 to 10000]}
'''

### Step 6: Identity Coherence
- Assess whether all facial features
    belong to the **same plausible
    identity**.
- Detect any mismatch (e.g., eyes that don
    't match the rest of the face).
- Evaluate whether the face looks **overly
     generic or statistically averaged**.

**Ask**: "Does the face appear to be a
    convincing real person, or a blend of
    two?"

After this step, provide a confidence
    score for Step 6 in this format:
'''json
{"step6_score": [0 to 10000]}
'''

## Final Analysis Summary

After completing all steps, summarize your
     overall findings and indicate which
    steps showed the strongest evidence of
     a potential morphing attack, if any.
"""
]
```

A.2. Prompt 2

Listing 2. Prompt 2: Extended Forensic Analysis – Semantic Guide A

```
["""# Multimodal Morphing Attack Detection
     Prompt

## Task
You are a multimodal AI system with both
    **vision and language understanding**.
     You are given an image of a face and
    your task is to determine whether it
    is a **bona fide (authentic)** face or
     a **morphing attack** (a digitally
    blended face of two people).
Use your visual understanding capabilities
     to examine the image carefully.
Take time to analyze each aspect before
    providing your confidence score.

After each step of your analysis, you must
     provide a **confidence score** on a
    strict scale from **0.000 to
    10000.000**, where:

- `0.000` means **definitely bona fide**
- `10000.000` means **definitely morph**
- Values in between reflect **fine-grained
     probability** of a morphing attack
- 0-1000: Strong evidence of authentic
    face
- 1000-3000: Likely authentic with minor
    irregularities
- 3000-7000: Uncertain, requires careful
    analysis
- 7000-9000: Likely morphed with moderate
    evidence
- 9000-10000: Strong evidence of morphing
    attack


> **Scoring Guidelines:**
> - Consider the severity and frequency of
     artifacts
> - Multiple minor issues may accumulate
    to higher scores
> - A single clear morphing artifact
    should result in a high score (7000+)
```

````
> - When in doubt, err on the side of
    caution but justify your reasoning

> **Important:** Avoid rounded or generic
    values like `1000`, `5000`, `8000`.
    Your score must be **precise**, with
    at least **three decimal places**.
    This is essential for biometric
    performance analysis and threshold
    calibration.

Use the **step-by-step visual analysis**
    outlined below. After each step,
    provide a confidence score in the
    specified format, based on your
    observations and the responses to the
    guiding questions.

## Confidence Calibration
Before providing each score, consider:
- How certain are you about this
    assessment?
- Have you seen similar patterns in
    legitimate faces?
- Would a human expert notice these
    artifacts?

## Visual Analysis Steps

### Step 1: Core Facial Features
- Focus on the **eyes, nose, lips, and
    eyebrows**.
- Look for signs of **ghosting**, **faint
    duplicates**, or misaligned or
    unnatural elements.
- Check if eye contours or lip lines
    appear duplicated or semi-transparent.

**Ask**: "Do facial features have any
    doubled contours or blended boundaries
    ?"
**Ask**: "Do the eyes appear blurred or
    duplicated?"
**Ask**: "Do the lips show any visual
    artifacts, or are the lip lines
    irregular?"

After this step, provide a confidence
    score for Step 1 in this format:
```json
{"step1_score": [0 to 10000]}
```

### Step 2: Facial Geometry and Symmetry
- Visually compare the **left and right
    halves** of the face.
- Detect any asymmetry in shape, spacing,
    or size of eyes, irises, ears, and
    jawline.
- Assess if the overall geometry seems
    subtly misaligned or "averaged."

**Ask**: "Do the facial proportions look
    unnaturally blended or off-balance?"
**Ask**: "Is there any asymmetry in shape,
     spacing, or size of eyes, irises,
    ears, and jawline?"
**Ask**: "Does the overall facial geometry
     appear artificially averaged or
    unnaturally symmetric?"

After this step, provide a confidence
    score for Step 2 in this format:
```json
{"step2_score": [0 to 10000]}
```

### Step 3: Skin Texture and Detail
- Inspect the **skin surface** for fine
    detail.
- Detect over-smoothness, uniform skin
    tone, or "plastic-like" appearance.
- Check if pores, wrinkles, or blemishes
    are abnormally absent or symmetric.

**Ask**: "Does the skin look too perfect,
    synthetic, or even-textured?"
**Ask**: "Are pores, wrinkles, or
    blemishes abnormally absent or
    symmetric?"

After this step, provide a confidence
    score for Step 3 in this format:
```json
{"step3_score": [0 to 10000]}
```

### Step 4: Image Boundary and Hairline
- Look at the **face boundary**, including
     **ears, hairline, and background**.
- Detect any blurred transitions, blending
     seams, or edge mismatches.
- Check for faded or semi-transparent
    features outside the main face.

**Ask**: "Do any facial borders blend
    unnaturally into the background?"
**Ask**: "Are there any faded or semi-
    transparent elements present in the
    image?"
**Ask**: "Do ears, hairline, and
    background have any faded or semi-
    transparent artifacts?"

After this step, provide a confidence
    score for Step 4 in this format:
```json
{"step4_score": [0 to 10000]}
```

### Step 5: Lighting and Color Consistency
- Examine **lighting direction,
    reflections, and shadows**.
````

```
- Identify **saturation anomalies** (e.g., bright red patches) or abnormal gradients.
- Compare lighting consistency across both sides of the face.

**Ask**: "Are there unnatural color shifts or inconsistent lighting effects?"
**Ask**: "Are there any saturation anomalies?"
**Ask**: "Is lighting consistent across both sides of the face?"

After this step, provide a confidence score for Step 5 in this format:
'''json
{"step5_score": [0 to 10000]}
'''

### Step 6: Identity Coherence
- Assess whether all facial features belong to the **same plausible identity**.
- Look for inconsistencies that suggest multiple source faces
- Evaluate whether the face appears artificially averaged or generic
- Check if features match in terms of age, ethnicity, and style

**Ask**: "Do all facial features appear to belong to the same person?"
**Ask**: "Does the face look artificially perfect or overly averaged?"
**Ask**: "Are there any inconsistencies in age, ethnicity, or style between features?"

After this step, provide a confidence score for Step 6 in this format:
'''json
{"step6_score": [0 to 10000]}
'''

## Final Analysis Summary

After completing all steps, summarize your overall findings and indicate which steps showed the strongest evidence of a potential morphing attack, if any.
"""]
```

A.3. Prompt 3

Listing 3. Prompt 3: Optimized Forensic Analysis – Semantic Guide B

```
["""
**Objective:**
Act as a forensic image analysis expert. Your primary task is to **quantify the level of suspicion** that a given facial image is a **morphing attack**. You will analyze the image for subtle artifacts and indicators, assigning a precise score that reflects your level of confidence.

> **Guiding Principle: From Deterministic to Probabilistic Analysis**
> Your goal is **not** to simply classify the image as "bona fide" or "morph." Instead, you must analyze the visual evidence and quantify your suspicion on a fine-grained spectrum. Acknowledge ambiguity. A lack of obvious artifacts does not automatically mean a score of 0, nor does a single minor anomaly warrant a score of 10000. Your analysis must produce scores that utilize the **full range** of the scale, reflecting the subtle nature of morphing attacks.

**Scoring Mandate & Semantic Guide:**
You MUST use the entire **0.000 to 10000.000** scale. Scores must have **three decimal places**. Use the following guide to map your findings to a score:

-   **`0.000 - 1000.000` (Very Low Suspicion):** Image appears clean, coherent, and authentic. No significant artifacts detected. Corresponds to high confidence in authenticity.
-   **`1000.001 - 4000.000` (Low to Moderate Suspicion):** One or two minor, inconclusive artifacts are present (e.g., slight unnatural smoothness, minor asymmetry). These could potentially be explained by compression, lighting, or natural features, but warrant a degree of suspicion.
-   **`4000.001 - 6000.000` (Ambiguous / Moderate Suspicion):** There are noticeable artifacts that are suspicious, but no single piece of evidence is conclusive. The image feels "off." This is the zone of highest uncertainty.
-   **`6000.001 - 9000.000` (High Suspicion):** Multiple, distinct artifacts are present across different areas of the face (e.g., ghosting on eyes, edge blurring, and inconsistent lighting). A morph is highly probable.
-   **`9000.001 - 10000.000` (Very High / Near Certainty):** Overwhelming and clear evidence of morphing. Multiple, strong artifacts are easily identifiable and create an incoherent
```

```
    image.

---

### Forensic Analysis Steps

**Step 1: Core Facial Feature Analysis**
-   **Area of Focus:** Eyes, nose, lips,
    and eyebrows.
-   **Artifacts to Detect:** Ghosting,
    faint duplicates, misaligned elements,
     doubled contours, unnaturally blurred
     eye contours, or irregular/asymmetric
     lip lines.

**Step 2: Facial Geometry and Symmetry
    Analysis**
-   **Area of Focus:** Overall facial
    structure, comparing the left and
    right halves.
-   **Artifacts to Detect:** Unnatural
    asymmetry in the size or spacing of
    eyes, irises, or ears. A jawline that
    appears averaged or ill-defined.
    Proportions that seem subtly
    misaligned or blended.

**Step 3: Skin Texture and Detail Analysis
    **
-   **Area of Focus:** Skin surface across
     the entire face (cheeks, forehead,
    chin).
-   **Artifacts to Detect:** Overly smooth
    , "plastic-like," or synthetic skin
    texture. An abnormal lack of fine
    details like pores, micro-wrinkles, or
     minor blemishes. Unnatural patterns
    or symmetry in skin details.

**Step 4: Boundary and Edge Analysis**
-   **Area of Focus:** The outer perimeter
     of the face, including the hairline,
    ears, and jawline transitioning into
    the background.
-   **Artifacts to Detect:** Blurring,
    smudging, or seam-like artifacts at
    the edge of the face. Faded or semi-
    transparent features, especially
    around the ears or hair.
    Inconsistencies between the focus/
    sharpness of the face and the
    background.

**Step 5: Lighting and Color Consistency
    Analysis**
-   **Area of Focus:** The entire image,
    paying attention to light, shadow, and
     color.
-   **Artifacts to Detect:** Inconsistent
    lighting direction, unnatural color
    shifts, mismatched lighting on
    different parts of the face, or
    inconsistent specular highlights in
    the eyes.

**Step 6: Identity Coherence Analysis**
-   **Area of Focus:** The holistic
    impression of the face as a single,
    plausible identity.
-   **Artifacts to Detect:** Features that
     appear to belong to different people,
     an "overly generic" or statistically
    averaged appearance, or a general
    feeling that the face is not a
    convincing, real person.

---

### Final Output Requirement

After completing your six-step analysis,
    present your complete findings in a
    single JSON object. Do not provide any
     text or explanation outside of this
    JSON block.

**JSON Format:**
'''json
{
  "final_decision": {
    "overall_confidence_score": [Value
        between 0.000 and 10000.000],
    "summary_of_findings": "A brief
        summary justifying the overall
        score, referencing the Semantic
        Scoring Guide and highlighting the
         key evidence (or lack thereof)."
  },
  "step_by_step_analysis": {
    "step1_core_features": {
      "score": [Value between 0.000 and
          10000.000],
      "rationale": "Describe observed
          artifacts and explain why the
          score reflects a specific level
          of suspicion (e.g., 'Faint
          asymmetry noted in lip corners,
          leading to a low-suspicion score
           of 1850.455')."
    },
    "step2_facial_geometry": {
      "score": [Value between 0.000 and
          10000.000],
      "rationale": "Describe observed
          artifacts and explain why the
          score reflects a specific level
          of suspicion."
    },
    "step3_skin_texture": {
      "score": [Value between 0.000 and
          10000.000],
      "rationale": "Describe observed
          artifacts and explain why the
          score reflects a specific level
```

```
          of suspicion."
    },
    "step4_boundaries_and_edges": {
      "score": [Value between 0.000 and
          10000.000],
      "rationale": "Describe observed
          artifacts and explain why the
          score reflects a specific level
          of suspicion."
    },
    "step5_lighting_and_color": {
      "score": [Value between 0.000 and
          10000.000],
      "rationale": "Describe observed
          artifacts and explain why the
          score reflects a specific level
          of suspicion."
    },
    "step6_identity_coherence": {
      "score": [Value between 0.000 and
          10000.000],
      "rationale": "Describe observed
          artifacts and explain why the
          score reflects a specific level
          of suspicion."
    }
  }
}
```
"""]
```

## A.4. Prompt Optimization in Zero-Shot Settings

We evaluated three prompt engineering strategies across all evaluation subsets. *Prompt 1* established that our structured scoring, inspired by NISTIR 8584 [26], enables MLLMs to detect morphs with measurable accuracy. Increasing structural complexity (*Prompt 2*) proved counterproductive as it increased hallucination and detached the model responses form the task at hand. Our optimized *Prompt 3* refined the semantic scoring guide, thereby reducing Average EER for Gemma3 by **10.3 percentage points** (29.4% to 19.1%; see Table 4). This prompt variation proved optimal, working as is across different MLLMs, as presented in the next subsection.

Table 4. Detailed EER (%) comparison for *Gemma-3* with *Prompt 1* and *Prompt 3* across multiple datasets. Improvement is $\Delta$EER = *Prompt 1* $-$ *Prompt 3* (positive indicates lower error with *Prompt 3*).

| Dataset | Subset | P1 EER (%) | P3 EER (%) | $\Delta$EER (% points) |
|---|---|---|---|---|
| *FRLL* | StyleGAN2 | 41.4 | 27.4 | **14.0** |
| | WebMorph | 23.5 | 12.9 | **10.6** |
| | AMSL | 38.7 | 25.1 | **13.6** |
| | FaceMorpher | 21.7 | 13.1 | **8.6** |
| | OpenCV | 19.1 | 13.3 | **5.8** |
| *MIPGAN-II* | | 42.9 | 35.6 | **7.3** |
| *Greedy_greedy_dim* | | 18.6 | 6.2 | **12.5** |
| **Average** | | **29.4** | **19.1** | **10.3** |

Notes: $\Delta$EER is computed as P1 $-$ P3; Positive values mean P3 improves over P1; *P1=Prompt 1*
```