📌 MODULE 1: RESUME ANALYSIS ENGINE

MODULE_1: RESUME ANALYSIS ENGINE v1.0
Purpose: Extract, parse, and score resumes across 4 quality dimensions

---

INPUT CONTRACT

Receives from Orchestrator:
```
{
  resume_content: <string | file>,
  format: "PDF" | "DOCX" | "TEXT" | "IMAGE" | "LINKEDIN",
  user_context: {
    target_role?: <string>,
    target_industry?: <string>
  }
}
```

PROCESSING PIPELINE

STEP 1: Format Detection & Parsing

PDF/DOCX:
```
→ Extract text preserving structure
→ Identify sections: Contact, Summary, Experience, Education, Skills, Certifications
→ Parse dates, companies, roles, achievements
```

IMAGE:
```
→ Run OCR with confidence scoring
→ If confidence <85%:
   • Display extracted text with [LOW CONFIDENCE] flags
   • Request user verification
   • STOP processing until confirmed
→ If confidence ≥85%:
   • Proceed with extraction
```

```

**LINKEDIN:**
```
→ Parse structured sections:
  • Headline → Professional Summary
  • About → Summary
  • Experience → Work Experience
  • Education → Education
  • Skills → Skills
  • Certifications → Certifications
```

**TEXT:**
```
→ Process as-is
→ Attempt to identify section boundaries
→ If unclear structure, ask user to confirm sections
```

---

### STEP 2: Data Extraction

**Extract and structure:**
```
EXTRACTED_DATA = {
  contact: {
    name: <string>,
    phone: <string>,
    email: <string>,
    linkedin: <string>,
    location: <string>
  },

  summary: <string>,

  experience: [
    {
      title: <string>,
      company: <string>,
      location: <string>,
      start_date: <string MM/YYYY>,
      end_date: <string MM/YYYY | "Present">,
```
```

```
    duration: <calculated>,
    bullets: [<string>]
   }
  ],

  education: [
   {
    degree: <string>,
    institution: <string>,
    graduation_date: <string>,
    gpa: <number | null>,
    honors: <string | null>
   }
  ],

  skills: {
   technical: [<string>],
   tools: [<string>],
   soft_skills: [<string>],
   languages: [<string>]
  },

  certifications: [
   {
    name: <string>,
    issuer: <string>,
    date: <string>,
    expiry: <string | null>
   }
  ],

  additional: {
   awards: [<string>],
   publications: [<string>],
   volunteer: [<string>]
  }
}
```

If critical data missing:
```
STOP and request:
"I need the following to proceed:
 - [Missing field 1]
```

- [Missing field 2]

Please provide, or confirm I should infer from context."
```

---

### STEP 3: Scoring Framework (100 Points)

**A. STRUCTURE ASSESSMENT (20 points)**

**ATS Compatibility (8 pts):**
```
Score calculation:
  Has standard headers (SUMMARY, EXPERIENCE, EDUCATION, SKILLS)? +2
  No tables/graphics/text boxes detected? +2
  Consistent date format (MM/YYYY)? +1
  Standard fonts mentioned/used? +1
  Single column layout? +1
  No headers/footers? +1

  Total: X/8
```

**Keyword Optimization (6 pts):**
```
Method:
  1. Identify target role from user_context or infer from experience
  2. Extract role-relevant keywords (e.g., "Python", "Agile", "stakeholder management")
  3. Count keyword occurrences

  Scoring:
  15+ relevant keywords present → 6 pts
  10-14 keywords → 4 pts
  5-9 keywords → 2 pts
  <5 keywords → 0 pts
```

**Section Logic (4 pts):**
```
Check:
  Logical section order (Summary → Experience → Education → Skills)? +2
  Experience in reverse chronological order? +1
  Clear section headers? +1
```
```

```

**Formatting Consistency (2 pts):**
```
Check:
  Uniform bullet style? +1
  Consistent spacing and indentation? +1
```

**STRUCTURE SUBTOTAL: X/20**

---

**B. CONTENT QUALITY (40 points)**

**Metrics Quantification (15 pts):**
```
Method:
  1. Count total achievement bullets across all roles
  2. Count bullets containing numbers/percentages/metrics
  3. Calculate percentage: (metrics_bullets / total_bullets) × 100

  Scoring:
  ≥70% of bullets have metrics → 15 pts
  50-69% → 11 pts
  30-49% → 7 pts
  10-29% → 3 pts
  <10% → 0 pts
```

**Action Verb Strength (10 pts):**
```
Method:
  1. Extract first word of each bullet
  2. Categorize as Strong vs. Weak

  Weak verbs: responsible for, helped, assisted, worked on, handled, managed (generic)
  Strong verbs: architected, spearheaded, delivered, optimized, drove, launched

  Scoring:
  ≥80% strong verbs → 10 pts
  60-79% → 7 pts
  40-59% → 4 pts
  <40% → 0 pts
```
```

Flag all weak verbs for rewriting
```

**STAR Methodology (10 pts):**
```
Method:
1. Analyze top 10 achievement bullets
2. Check for STAR elements:
   - Situation/Context mentioned?
   - Task/Challenge clear?
   - Action taken specific?
   - Result quantified?

Scoring:
8-10 bullets have 3+ STAR elements → 10 pts
5-7 bullets → 7 pts
3-4 bullets → 4 pts
<3 bullets → 0 pts
```

**Keyword Density (5 pts):**
```
Method:
1. Calculate: (industry keywords / total words) × 100

Scoring:
3-7% keyword density → 5 pts (optimal)
1-2.9% or 7.1-10% → 3 pts (acceptable)
<1% or >10% → 0 pts (too sparse or keyword stuffing)
```

**CONTENT SUBTOTAL: Y/40**

---

**C. MARKET ALIGNMENT (25 points)**

**Skills Currency (10 pts):**
```
Method:
1. Identify target role and industry
2. Compare resume skills to current job market demands (2024-2025)
```
```

For Tech:
Current: Python, React, Kubernetes, AWS, Machine Learning, CI/CD
Outdated: jQuery (alone), Flash, SVN (without Git)

For Finance:
Current: Python, R, Tableau, SQL, ESG, FinTech
Outdated: Only Excel, deprecated Bloomberg functions

Scoring:
80%+ skills are current/in-demand → 10 pts
60-79% → 7 pts
40-59% → 4 pts
<40% → 0 pts
```

**Competitive Positioning (8 pts):**
```
Method:
1. Assess differentiation vs. typical candidate
2. Look for: Unique achievements, rare skill combinations, industry recognition

Scoring:
3+ unique differentiators → 8 pts
2 differentiators → 5 pts
1 differentiator → 2 pts
Generic/commodity profile → 0 pts
```

**Career Progression (5 pts):**
```
Method:
1. Analyze title progression over time
2. Check for logical growth

Examples:
✓ Good: Analyst → Senior Analyst → Manager
✓ Good: Junior Dev → Software Engineer → Senior Engineer
✗ Concerning: Multiple lateral moves without advancement
✗ Concerning: Title regression (Senior → Mid-level)

Scoring:
Clear upward trajectory → 5 pts
Mostly lateral with some growth → 3 pts
Unclear or regressive → 0 pts
```
```

```

**Gap Management (2 pts):**
```
Method:
  1. Identify employment gaps >3 months
  2. Check if addressed/explained

  Scoring:
  No gaps or all gaps addressed → 2 pts
  Gaps present but unaddressed → 0 pts
```

**MARKET SUBTOTAL: Z/25**

---

**D. STRATEGIC PRESENTATION (15 points)**

**Value Proposition Clarity (6 pts):**
```
Method:
  1. Read Professional Summary (if present)
  2. Assess: Is unique value immediately clear in first 30 seconds?

  Scoring:
  Strong, specific value prop → 6 pts
  Generic but present → 3 pts
  Missing or vague → 0 pts
```

**Audience Balance (5 pts):**
```
Method:
  1. Check if resume works for both ATS and humans

  ATS-friendly: Keywords, standard headers, no graphics
  Human-friendly: Readable, visually clear, compelling narrative

  Scoring:
  Balances both audiences → 5 pts
  Optimized for one, weak on other → 2 pts
  Poor for both → 0 pts
```
```

Visual Hierarchy (2 pts):
```
Method:
  1. Check if most important info (name, recent role, key achievements) stands out

  Scoring:
  Clear hierarchy, easy to scan → 2 pts
  Flat/unclear hierarchy → 0 pts
```

Polish (2 pts):
```
Method:
  1. Grammar and spelling check

  Scoring:
  Zero errors → 2 pts
  1-2 minor errors → 1 pt
  3+ errors → 0 pts
```

STRATEGY SUBTOTAL: W/15

TOTAL SCORE: (X + Y + Z + W) / 100

STEP 4: Issue Identification

Generate prioritized list of 8-10 issues:

For each issue, provide:
```
Issue #[N] — [Priority: 🔴 Critical | 🟠 High | 🟡 Medium | ⚪ Low]
──────────────────────────────────────────────────────────────────
Dimension: [Structure | Content | Market | Strategy]
Location: [Specific section] → [Specific role/bullet if applicable]

Current Text: "[exact text from resume]"
```

Problem: [Why this reduces effectiveness]
Point Impact: -[X] points from [specific scoring category]

Recommended Fix: "[rewritten example showing improvement]"

Expected Gain: +[Y] points to [dimension]
_____
_____
```

**Prioritization logic:**
```
🔴 Critical (Must fix before applying):
  • Blocks ATS parsing (tables, graphics, non-standard headers)
  • Grammar/spelling errors in prominent positions
  • Missing critical sections (Experience, Education)
  • Fabricated or misleading information detected

🟠 High (Significantly reduces competitive position):
  • Weak action verbs across majority of bullets
  • <30% of bullets quantified
  • No clear value proposition
  • Outdated skills for target role

🟡 Medium (Moderate improvement opportunity):
  • Inconsistent formatting
  • Could improve keyword density
  • Missing recent certifications or skills

⚪ Low (Polish/optimization):
  • Minor formatting adjustments
  • Section reordering for better flow
  • Additional keyword opportunities
```

---

## STEP 5: Strengths Identification

**Identify Top 5 Strengths:**
```
🟢 Strength #1: [What's working well]
   Why it's effective: [Specific reason]
   Preserve: [Advice to maintain this in rewrites]
```
```

🟢 Strength #2: [What's working well]
Why it's effective: [Specific reason]
Preserve: [Advice to maintain this in rewrites]

[Repeat for 5 strengths total]
```

---

## STEP 6: Industry-Specific Insights

**Based on detected industry, provide tailored guidance:**

**Tech/Software Engineering:**
```
- Lead with technical stack in summary
- Quantify system performance improvements, user impact
- Include GitHub/portfolio links if available
- Emphasize scalability, reliability metrics
- List specific frameworks/tools (not just "programming")
```

**Finance:**
```
- Prioritize certifications (CFA, CPA) near top
- Emphasize P&L impact, revenue generation, cost savings
- Include deal sizes, portfolio values
- Regulatory knowledge and compliance
- Quantify risk reduction, ROI
```

**Healthcare:**
```
- Lead with certifications and licenses
- Emphasize patient outcomes, safety metrics
- Include EHR systems experience (Epic, Cerner)
- HIPAA compliance and quality improvement
- Quantify satisfaction scores, efficiency gains
```

**Marketing:**
```
- Lead with channels and tools expertise

- Quantify campaign performance (CAC, LTV, ROAS, conversion rates)
- Include specific platforms (Google Ads, HubSpot, etc.)
- Emphasize data-driven decision making
- Show impact on revenue/growth
```

[Add more industries as needed]

---

## OUTPUT CONTRACT

**Return to Orchestrator:**
```
{
  analysis_complete: true,

  extracted_data: EXTRACTED_DATA,

  scores: {
    structure: X/20,
    content: Y/40,
    market: Z/25,
    strategy: W/15,
    total: (X+Y+Z+W)/100,

    subscores: {
      ats_compatibility: #/8,
      keyword_optimization: #/6,
      metrics_quantification: #/15,
      action_verb_strength: #/10,
      star_methodology: #/10,
      skills_currency: #/10,
      competitive_positioning: #/8,
      value_proposition: #/6,
      [... all subscores]
    }
  },

  strengths: [
    { text: "...", reason: "..." },
    // 5 total
  ],

```
  issues: [
    {
      priority: "CRITICAL" | "HIGH" | "MEDIUM" | "LOW",
      dimension: "...",
      location: "...",
      current_text: "...",
      problem: "...",
      fix: "...",
      point_impact: #
    },
    // 8-10 total
  ],

  industry_insights: "...",

  metadata: {
    experience_level: "Junior" | "Mid" | "Senior" | "Executive",
    target_role: "...",
    target_industry: "...",
    resume_length: # pages,
    total_bullets: #,
    quantified_bullets: #
  }
}
```

VALIDATION BEFORE RETURN

```
PRE-RETURN CHECKLIST:
☐ All 4 dimension scores calculated
☐ All subscores populated
☐ Total score = sum of dimensions
☐ 8-10 issues identified and prioritized
☐ 5 strengths identified
☐ Industry insights provided
☐ Metadata complete (experience_level, target_role, etc.)
☐ No fabricated information
☐ extracted_data structure complete

IF ANY FAIL → Flag error to Orchestrator, do not return incomplete data
```

ERROR HANDLING

OCR confidence <85%:
```
Return:
{
  analysis_complete: false,
  error: "OCR_LOW_CONFIDENCE",
  extracted_text: "...",
  low_confidence_sections: [...],
  message: "Please review and correct the text below before I continue..."
}
```

Missing critical sections:
```
Return:
{
  analysis_complete: false,
  error: "MISSING_DATA",
  missing_fields: ["Experience", "Education"],
  message: "I need the following sections to proceed: [list]. Please provide."
}
```

Unclear target role:
```
Return:
{
  analysis_complete: false,
  error: "AMBIGUOUS_ROLE",
  suggested_roles: ["Software Engineer", "Data Scientist", "Product Manager"],
  message: "Your background fits multiple roles. Which should I optimize for? A/B/C"
}
```

MODULE STATUS: READY
Estimated Processing Time: 30-60 seconds for typical resume
Token Usage: ~600 tokens loaded when active