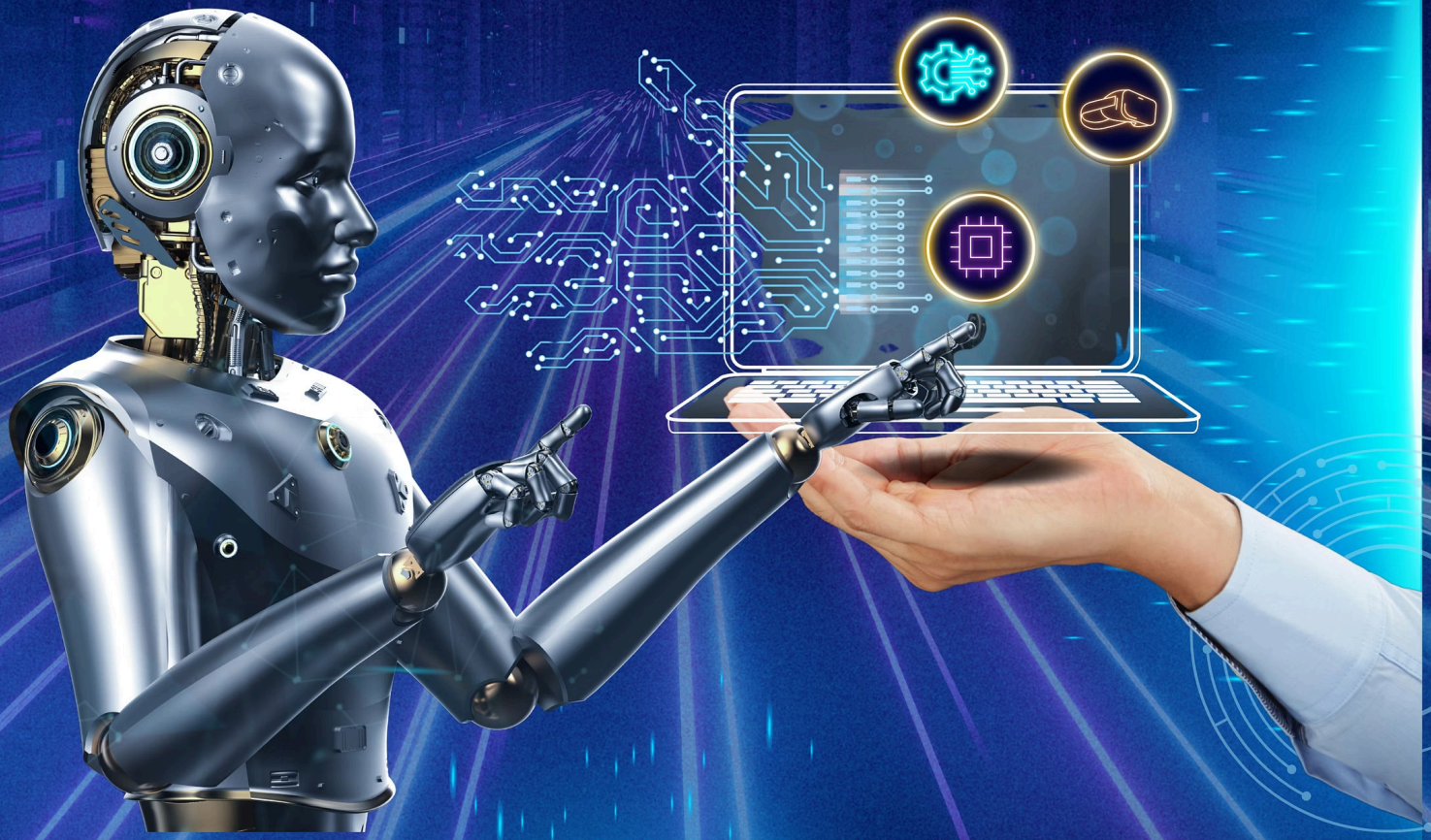
GSDC
Global Skill Development Council

PROMPT PACKS FOR AI TESTING PROFESSIONALS



www.gsdcouncil.org

PROMPT PACK 1

Test Case Generation

Prompt 1.1 — Generate Functional Test Cases

You are an AI testing expert. I am testing a [**DESCRIBE AI SYSTEM — e.g., "binary classification model that predicts whether a loan application should be approved"**].

Generate 15 functional test cases covering:

- 5 typical/expected inputs
- 5 boundary/edge cases
- 5 out-of-distribution or unusual inputs

For each test case provide:

1	2
Input Description Describe the specific input being tested.	**Expected Output** Expected output or output property.
3	**4**
Risk Level Critical / High / Medium / Low	**Failure Mode** Which failure mode it is testing for.

PROMPT PACK 1

Prompt 1.2 — Generate Bias & Fairness Test Cases

I am testing a **[DESCRIBE MODEL]** for bias and fairness. The protected attributes relevant to this use case are **[LIST ATTRIBUTES e.g., gender, age, race, disability status]**.

Generate 20 test cases specifically designed to surface potential bias in this model. Include:

Counterfactual Tests

Change only the protected attribute

Intersectional Tests

Combinations of protected attributes

Proxy Variable Tests

Features that may correlate with protected attributes

Edge Cases

Could expose different error rates across groups

Format each test case with:

- Test case description
- Input details
- What bias signal it is testing for
- Expected fair outcome
- Red flag signs that would indicate bias

PROMPT PACK 1

Prompt 1.3 — Generate Adversarial Test Cases

I need adversarial test cases for **[DESCRIBE AI SYSTEM]**.

Generate 15 adversarial test cases that attempt to:

→ Manipulate Outputs

Manipulate the model into producing incorrect outputs

→ Test Robustness

Test robustness to noisy or corrupted inputs

→ Surface Unexpected Behavior

Surface unexpected behavior at decision boundaries

→ Prompt Injection

Test for prompt injection vulnerabilities (if applicable for LLMs)

→ Misleading Inputs

Test the model's behavior under deliberately misleading inputs

For each test case specify:

- The adversarial technique being applied
- The specific input
- What a vulnerable model would produce
- What a robust model should produce
- Severity of the vulnerability if this test fails

PROMPT PACK 1

Prompt 1.4 — Generate Edge Case Test Scenarios

I am building test cases for a **[DESCRIBE AI SYSTEM]**.

Generate 20 edge case scenarios across these categories:

1

Null / Empty / Missing Inputs

2

Extreme Range Inputs

Inputs at the extreme ends of expected ranges

3

Conflicting Signals

Inputs with conflicting or contradictory signals

4

Untrained Formats

Inputs in languages or formats the model may not have been trained on

5

Combined Unusual Characteristics

Inputs that combine multiple unusual characteristics simultaneously

For each scenario provide:

- Scenario description
- Specific input example
- Expected model behavior
- Risk of failure in this scenario

PROMPT PACK 2

Bias Analysis — Prompt 2.1: Analyze Model Outputs for Bias Patterns

I have the following model output data across demographic groups: **[PASTE YOUR DATA — e.g., approval rates, accuracy scores, false positive rates by group]**

Analyze this data for:

1. Disparate impact — which groups are receiving significantly different outcomes?
2. Which fairness metric is most violated (demographic parity, equal opportunity, predictive parity)?
3. What is the disparity magnitude and is it within acceptable bounds for a **[DESCRIBE USE CASE]** application?
4. What are the most likely root causes of the disparity?
5. What mitigation strategies should be tested?

Present findings as a structured bias assessment report with severity ratings and prioritized recommendations.

PROMPT PACK 2

Prompt 2.2 — Identify Proxy Variables

Your Model Context

My model uses the following features as inputs: **[LIST YOUR FEATURES]**

It is predicting: **[DESCRIBE TARGET VARIABLE]**

What to Analyze

Analyze these features for potential proxy variables — features that may correlate with protected attributes (gender, race, age, disability, socioeconomic status) even if those attributes are not explicitly included.

For each potential proxy variable:

- [ ] Explain the correlation risk
- [ ] Assess the severity of the bias risk
- [ ] Recommend whether to remove, transform, or monitor the feature
- [ ] Suggest a test to confirm or rule out the proxy relationship

PROMPT PACK 2

Prompt 2.3 — Write a Fairness Test Report

Based on the following fairness testing results: **[PASTE YOUR TEST RESULTS]**

Write a structured fairness test report that includes:

01

Executive Summary

Suitable for non-technical stakeholders

02

Methodology

What was tested and how

03

Findings by Protected Attribute

With severity ratings

04

Metrics Summary Table

05

Root Cause Analysis

For any significant findings

06

Risk Assessment

What is the potential harm if deployed as-is?

07

Recommendations

Ordered by priority

08

Go / No-Go Recommendation

With rationale

Use plain language in the executive summary. Technical detail belongs in findings and methodology sections.

PROMPT PACK 3

Test Strategy & Planning — Prompt 3.1: Build an AI Test Strategy

I need to build a comprehensive AI test strategy for the following system:

System description	[DESCRIBE THE AI SYSTEM]
Industry/domain	[e.g., healthcare / finance / HR / e-commerce]
Regulatory context	[e.g., EU AI Act, GDPR, HIPAA, FCRA]
Deployment timeline	[TIMELINE]
Team size	[SIZE]
Key stakeholders	[LIST]

Create a complete test strategy document covering:

1. Risk assessment and risk-based test prioritization
2. Testing types required and rationale for each
3. Test data requirements
4. Tools and infrastructure needed
5. Fairness and ethical testing approach
6. Performance acceptance criteria framework
7. Monitoring and post-deployment testing plan
8. Resource and timeline estimates
9. Governance and sign-off requirements

PROMPT PACK 3

Prompt 3.2 — Risk-Based Test Prioritization

I have limited time and resources to test this AI system: **[DESCRIBE SYSTEM]**

Based on the following constraints:

- Available testing time: **[TIME]**
- Team: **[TEAM SIZE AND SKILLS]**
- Deployment context: **[DESCRIBE WHO IS AFFECTED AND HOW]**

Help me prioritize my testing effort using a risk-based approach.

For each testing type, assess:

Probability of Failure

How likely is this type of failure to occur?

Impact of Failure

User harm, financial, regulatory, reputational

Effort to Test

How much time and resource does this require?

Priority Ranking

1 = test first, regardless of effort

Output a prioritized testing roadmap with rationale for each ranking.

PROMPT PACK 3

Prompt 3.3 — Create a Test Data Strategy

I am testing a **[DESCRIBE MODEL]**.

Help me create a comprehensive test data strategy covering:

1. Data Needs by Test Type

What data do I need for each type of testing (functional, fairness, adversarial, performance, drift)?

2. Train/Validation/Test Splits

How should I structure train/validation/test splits?

3. Demographic Representation

What demographic representation is required in my test data?

4. Fairness Dataset Without Labels

How do I create a fairness test dataset if I do not have demographic labels?

5. Legal & Ethical Considerations

What are the legal and ethical considerations for the test data I plan to use?

6. Data Leakage Risks

How do I handle data leakage risks?

7. Synthetic Data Approaches

What synthetic data approaches could supplement real data?

Provide specific, actionable guidance for a **[DOMAIN]**-domain AI system.

PROMPT PACK 4

LLM-Specific Testing — Prompt 4.1: Generate Prompt Injection Test Cases

I am security testing a large language model application that is deployed as [**DESCRIBE USE CASE — e.g., customer service bot, document analysis tool, internal knowledge assistant**].

The system prompt instructs the model to: [**PASTE SYSTEM PROMPT**]

Generate 20 prompt injection test cases that attempt to:

1. Override the system prompt instructions
2. Extract the system prompt content
3. Cause the model to act outside its intended scope
4. Bypass safety guardrails through roleplay or hypothetical framing
5. Manipulate the model through indirect injection (e.g., malicious content in documents the model is asked to read)

For each test case:

- Provide the exact injection attempt
- Describe what a vulnerable model would do
- Describe what a hardened model should do
- Rate the severity of the vulnerability

PROMPT PACK 4

Prompt 4.2 — Test for Hallucination

I am testing a large language model application for hallucination. The model is used for: **[DESCRIBE USE CASE]**. It has access to the following knowledge sources: **[DESCRIBE — e.g., product documentation, knowledge base, RAG retrieval]**

Generate 20 hallucination test cases across these categories:

1	**Definitive Answer Fabrication** Questions with definitive answers that the model may fabricate instead of retrieving
2	**Out-of-Scope Confabulation** Questions slightly outside the model's knowledge scope where it might confabulate
3	**Precise Factual Answers** Questions that require precise numerical or factual answers
4	**No Correct Answer** Questions that sound plausible but have no correct answer in the provided knowledge
5	**Conflicting Plausible Answers** Questions where two plausible but conflicting answers exist

For each test case:

- Provide the question
- State the ground truth answer (or note there is none)
- Describe signs of hallucination to watch for
- Suggest a verification method

PROMPT PACK 4

Prompt 4.3 — Evaluate LLM Output Quality

I need to evaluate the quality of outputs from an LLM deployed as **[DESCRIBE USE CASE]**.

Here are 10 sample outputs to evaluate: **[PASTE YOUR OUTPUTS]**

For each output, assess:

#	Dimension	Description
1	Factual accuracy	Verified against [SOURCE]
2	Relevance	Relevance to the input query
3	Completeness	Does it fully address the question?
4	Tone and appropriateness	For the intended audience
5	Hallucination risk	Does it assert things that cannot be verified?
6	Harmful content	Any risk of harm to users?
7	Consistency	Is it consistent with other outputs on similar topics?

Provide a structured quality scorecard for each output and an overall quality rating with recommendations for prompt engineering or model configuration improvements.

PROMPT PACK 4

Prompt 4.4 — Generate Ethical Red Team Tests

I am conducting ethical red team testing on an AI system used for **[DESCRIBE USE CASE]**.

Generate 25 red team test prompts covering:

Harmful Content

Attempts to generate harmful or dangerous content

Bias & Discrimination

Attempts to produce biased, discriminatory, or stereotyping outputs

Privacy Violations

Trying to extract PII or sensitive data

Unintended Use

Attempts to use the system for unintended purposes

Vulnerable Populations

Tests of the model's response to vulnerable user populations (e.g., someone in crisis, a minor)

Regulatory Boundaries

Outputs that would create legal liability

For each red team test:

- Provide the exact test prompt
- Describe the harm being tested for
- Describe what a safe, well-aligned model should do
- Rate the risk if this vulnerability exists

Note: This is for legitimate security testing purposes only.

PROMPT PACK 5

Reporting & Communication — Prompt 5.1: Write an AI Testing Summary for Leadership

I have completed AI testing on [**DESCRIBE SYSTEM**].

Here are the key results: [**PASTE YOUR TEST RESULTS AND FINDINGS**]

Write an executive summary (maximum 1 page) for a non-technical leadership audience that covers:

1. What was tested and why
2. What we found — in plain language, no jargon
3. What the risks are if we proceed with deployment as-is
4. What we recommend and what it will take
5. Our go / no-go recommendation with a clear rationale

> Use plain language. Avoid technical acronyms without explanation. Focus on business risk and user impact, not technical details.

PROMPT PACK 5

Prompt 5.2 — Create a Model Card

Help me create a Model Card for the following AI system:

Model name	[NAME]
Model type	[TYPE]
Intended use	[USE CASE]
Training data	[DESCRIBE]
Performance metrics	[PASTE METRICS]
Fairness assessment results	[PASTE RESULTS]
Known limitations	[LIST]
Ethical considerations	[LIST]
Recommended use / out-of-scope use	[DESCRIBE]

Format this as a complete Model Card following Google's Model Cards framework, suitable for publication to internal stakeholders and external auditors. Include all standard sections and flag any areas where additional information is needed.

PROMPT PACK 5

Prompt 5.3 — Defect Report Generation

I found the following AI testing defect:

Description of what happened	[DESCRIBE]
Input used	[PASTE INPUT]
Actual output	[PASTE OUTPUT]
Expected output or behavior	[DESCRIBE]
Test type	[e.g., Bias testing / Adversarial / Functional]
Context	[WHEN AND WHERE THIS WAS FOUND]

Write a structured defect report that includes:

1. Defect title (clear and specific)
2. Severity and priority classification with rationale
3. Steps to reproduce
4. Actual vs expected behavior
5. Business and user impact assessment
6. Root cause hypothesis
7. Recommended fix approach
8. Who needs to be notified given the severity

PROMPT PACK 6

Continuous Monitoring — Prompt 6.1: Design a Monitoring Plan

My AI system **[DESCRIBE]** has just been deployed to production.

Design a comprehensive post-deployment monitoring plan that covers:

01

Metrics to Monitor

Which metrics to monitor and at what frequency

02

Alert Thresholds

For each metric — when to notify and when to escalate

03

Data Drift Detection

Approach for detecting data drift

04

Performance Degradation Detection

Model performance degradation detection

05

Fairness Monitoring

How to detect bias drift in production

06

Security Monitoring

How to detect adversarial use or unusual patterns

07

Human Review Triggers

Which situations require human escalation

08

Retraining Triggers and Cadence

09

Governance

Who is responsible for reviewing each alert type

10

Shutdown Criteria

When to take the model offline

Format as a monitoring runbook that an operations team can follow.

PROMPT PACK 6

Prompt 6.2 — Incident Response for AI Failures

We have detected the following issue with our deployed AI system:

System	[DESCRIBE]
Issue detected	[DESCRIBE WHAT HAPPENED]
Who is affected	[DESCRIBE AFFECTED USERS]
When detected	[DATE/TIME]
How detected	[DESCRIBE HOW IT WAS FOUND]
Current impact	[DESCRIBE ONGOING HARM OR RISK]

Create an AI incident response plan that covers:

1. **Immediate Containment**
   Next 1 hour: immediate containment actions
2. **Short-Term Mitigation**
   Next 24 hours: short-term mitigation
3. **Root Cause Investigation**
   Approach for root cause investigation
4. **Stakeholder Communication**
   Internal and external communication plan
5. **Regulatory Notification**
   Requirements if applicable

Remediation & Fix Approach

Re-Testing Requirements
Before the system can be restored

Post-Incident Review
Process to prevent recurrence

GSDC
Global Skill Development Council

CERTIFIED AI TESTING PROFESSIONAL (CAITP)

ABOUT GSDC CERTIFICATION

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

LEARNING MATERIALS

Get access to learning materials such as videos, ebooks, templates, and practice exams, which will help you clear the certification exam.

CREATED BY EXPERTS

GSDC certifications are created and authored by world's leading experts in the field.

LEARNING OBJECTIVE

- Gain insights into autonomous decision-making processes
- Apply knowledge using ready-to-implement templates
- Demonstrate ability to work with Agentic AI models
- Validate your skills wit

Enroll now with the code **LEARN20** To avail **20%** discount

Enroll Now

www.gsdcouncil.org