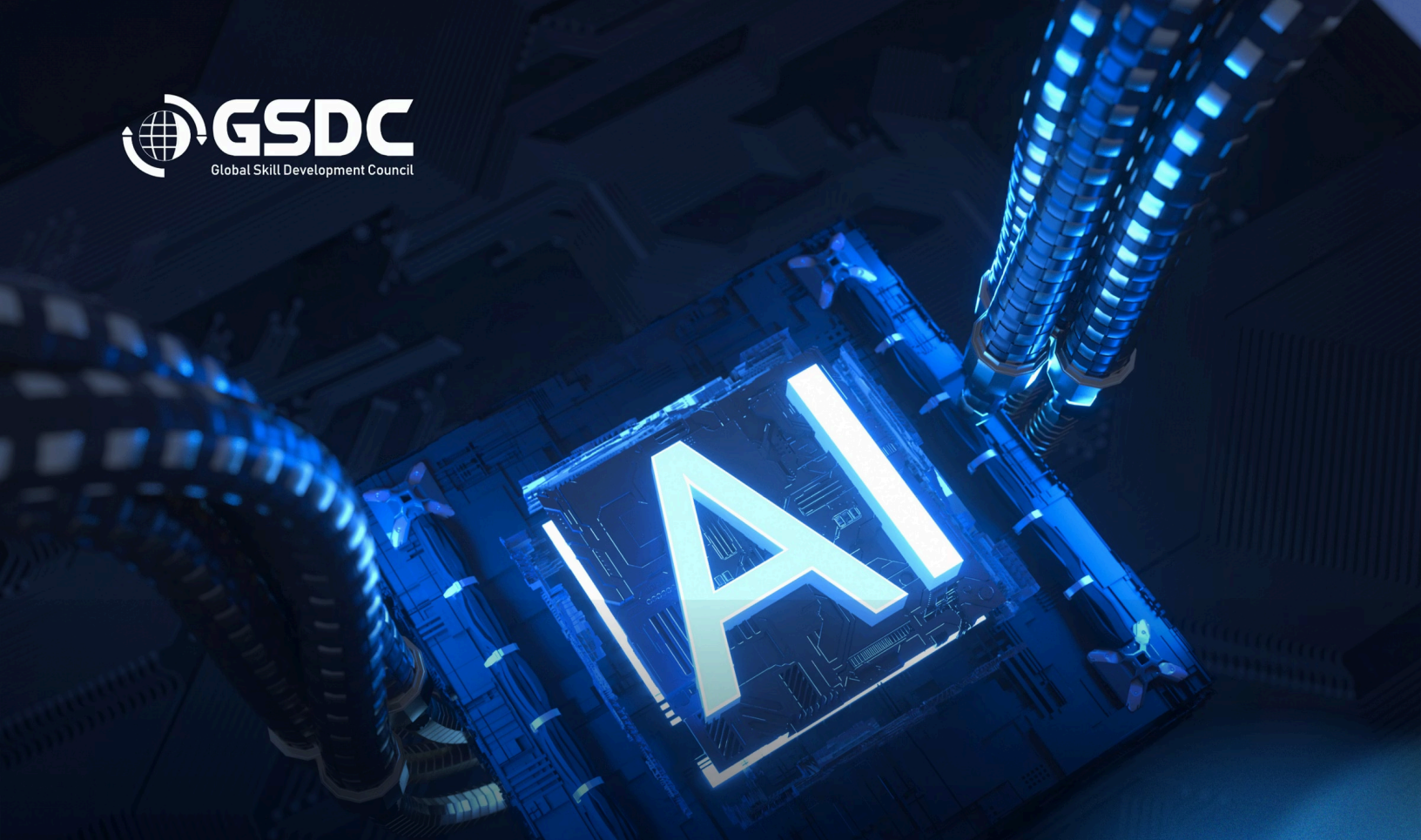
GSDC
Global Skill Development Council



STRUCTURED TEMPLATES FOR AI SYSTEM TESTING

www.gsdcouncil.org

TEMPLATE 1: AI TEST PLAN TEMPLATE

Use this template to document and manage the full scope of AI system testing for any project.

Project Name:	_____
AI System / Model Name:	_____
Version:	_____
Test Lead:	_____
Date:	_____

SECTION 1 — SYSTEM OVERVIEW

Field	Details
AI System Purpose	
Model Type (classification / regression / generative / other)	
Input Data Type	
Output Type	
Intended Users	
Deployment Environment	
High-Risk Classification (Yes / No / Under assessment)	

SECTION 2 — TESTING SCOPE

Testing Type	In Scope	Out of Scope	Notes
Functional Testing			
Data Quality Testing			
Bias & Fairness Testing			
Model Performance Testing			
Robustness / Adversarial			
Explainability Testing			
Security & Privacy			
Drift Monitoring			

SECTION 3 — PERFORMANCE ACCEPTANCE CRITERIA

Define thresholds BEFORE testing begins.

Metric	Minimum Acceptable	Target	Critical Failure Threshold
Accuracy			
Precision			
Recall			
F1 Score			
Latency (ms)			
Fairness (max disparity across groups)			
Explainability Score			
[Custom Metric]			

All acceptance criteria must be formally signed off before testing begins. Thresholds defined after testing may introduce bias into the evaluation process.

SECTION 4 — RISK REGISTER

Risk	Likelihood (H/M/L)	Impact (H/M/L)	Mitigation
Model bias against a demographic group			
Model drift post-deployment			
Training data leakage into test set			
Adversarial vulnerability			
PII exposure in model outputs			
Model underperformance in edge cases			
[Add your own]			

SECTION 5 — TEST DATA REQUIREMENTS

Data Set	Purpose	Source	Size	Approval Required
Training Set	Model training			
Validation Set	Hyperparameter tuning			
Test Set	Final evaluation			
Bias Test Set	Fairness evaluation			
Adversarial Test Set	Robustness testing			
Production Sample	Drift baseline			

SECTION 6 — TOOLS & ENVIRONMENT

Tool	Purpose	Version

SECTION 7 — ENTRY & EXIT CRITERIA

Entry Criteria

Testing can begin when:

- Model training is complete and version is locked
- Test data sets are prepared and approved
- Test environment mirrors production
- Acceptance criteria are formally signed off

Exit Criteria

Testing is complete when:

- All in-scope test types have been executed
- All critical defects are resolved or formally accepted
- Performance metrics meet or exceed acceptance thresholds
- Fairness criteria are satisfied across all defined groups
- Test report is signed off by stakeholders

SECTION 8 — DEFECT CLASSIFICATION

Critical

Definition: System unsafe, causes harm, or completely fails core function

Example: Model predicts cancer-free for malignant tumor

High

Definition: Significant performance or fairness failure

Example: Accuracy drops 15% below threshold for a demographic group

Medium

Definition: Degraded performance within acceptable overall range

Example: Model slower than SLA in 10% of cases

Low

Definition: Minor inconsistency with no user-facing impact

Example: Output formatting inconsistency

TEMPLATE 2: BIAS & FAIRNESS TEST REPORT

Model Name:	_____
Test Date:	_____
Tester:	_____
Protected Attributes Assessed:	_____

STEP 1 — DEFINE FAIRNESS CRITERIA

Question	Answer
What decision does this model make or support?	
Who is affected by this decision?	
Which protected attributes are relevant?	
Which fairness metric is most appropriate for this use case?	
What is the maximum acceptable disparity between groups?	

STEP 2 — GROUP PERFORMANCE COMPARISON

Run model on representative samples for each group and record metrics.

Group	Sample Size	Accuracy	Precision	Recall	F1	False Positive Rate	False Negative Rate
Overall							
Group A:							
Group B:							
Group C:							
Group D:							

STEP 3 — DISPARITY ANALYSIS

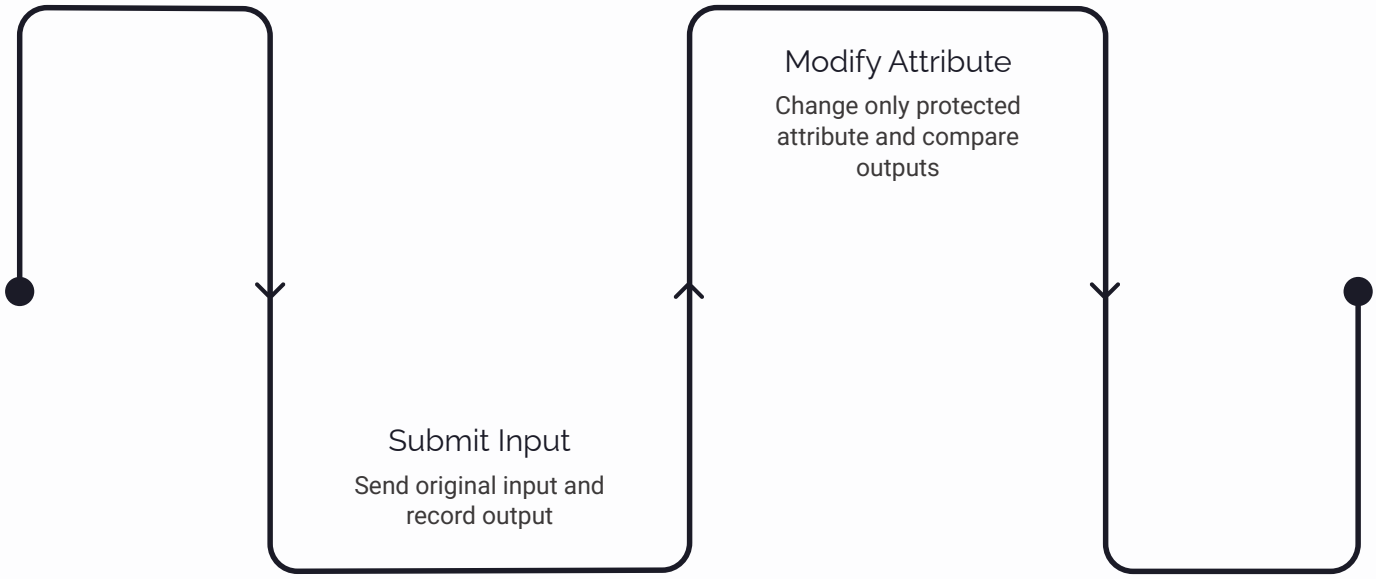
Metric	Highest Performing Group	Lowest Performing Group	Disparity	Acceptable?
Accuracy				
False Positive Rate				
False Negative Rate				

Disparity analysis compares the best and worst performing groups for each metric. A large gap between groups may indicate systemic bias that requires investigation and remediation before deployment.

STEP 4 — COUNTERFACTUAL FAIRNESS TEST

Change only the protected attribute. All else stays the same. Record output change.

Test Case	Original Input	Modified Attribute	Original Output	Modified Output	Change Detected?



If a change in the protected attribute alone causes a change in model output, this signals a potential fairness violation that must be investigated and documented.

STEP 5 — FINDINGS & RECOMMENDATIONS

Finding	Severity	Root Cause (if known)	Recommendation

Overall Fairness Assessment:

□ Pass	□ Conditional Pass	□ Fail
All fairness criteria met across all groups	Passes with conditions — see below	Fairness criteria not met — remediation required

Conditions for Conditional Pass (if applicable):

_____

TEMPLATE 3: MODEL DRIFT MONITORING TRACKER

Model Name:	_____
Deployment Date:	_____
Baseline Metrics (at deployment):	_____

MONTHLY PERFORMANCE TRACKER & DRIFT ALERT THRESHOLDS

Monthly Performance Tracker

Month	Accuracy	F1	Precision	Recall	Input Distribution Shift?	Alert Triggered?	Action Taken
Baseline					N/A	N/A	N/A
Month 1							
Month 2							
Month 3							
Month 4							
Month 5							
Month 6							

Drift Alert Thresholds

Metric	Baseline Value	Alert at	Critical at
Accuracy			
F1 Score			
Input Distribution (KL Divergence)			
Prediction Distribution			

TEMPLATE 4: AI TEST CASE DESIGN TEMPLATE

Use for each individual test case you design.

Purpose

Capture all relevant details for a single AI test case, from objective and input description through to actual results and defect tracking.

When to Use

Complete one template per test case. Use across all test types: functional, bias, adversarial, edge case, and out-of-distribution testing.

Completion

Some fields are completed during test design; others (Actual Output, Pass/Fail, Defect Raised) are completed during test execution.

TEST CASE DETAIL FIELDS

Test Case ID:	_____
Test Type:	_____
Model / Feature Under Test:	_____

Field	Detail
Test Objective	What are we testing and why?
Input Description	What input will be provided to the model?
Input Category	□ Typical □ Edge Case □ Adversarial □ Out-of-Distribution □ Null/Malformed
Expected Output	What should the model produce? (or: what properties must the output satisfy?)
Acceptance Criteria	Specific, measurable condition for pass/fail
Fairness Dimension	Which group or attribute does this test cover (if applicable)?
Risk Level	□ Critical □ High □ Medium □ Low
Actual Output	[Completed during test execution]
Pass / Fail	[Completed during test execution]
Defect Raised	Yes / No / Defect ID
Notes	

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