PUBLICATION

EXAM-TM40H-POST

POST-TEST — SL 4H: AI ENGINEER

Maven Smart System (MSS) — USAREUR-AF

HEADQUARTERS
UNITED STATES ARMY EUROPE AND AFRICA
(USAREUR-AF)
Wiesbaden, Germany

26 MARCH 2026

DRAFT — UNOFFICIAL — NOT FOR OPERATIONAL USE

POST-TEST — SL 4H: AI ENGINEER

MAVEN SMART SYSTEM (MSS) — USAREUR-AF

Field	Detail
Course	SL 4H: AI Engineer
Form	Post-Test
Level	SL 4H (Specialist)
Audience	AI/ML specialists; prerequisite: SL 1+20+30 + Python + prompt engineering familiarity
Time Allowed	45 minutes
Passing Score	70% (46/66)

INSTRUCTIONS

This assessment evaluates mastery of course learning objectives. A passing score of 70% is required to receive credit. Complete independently without reference to training materials.

SECTION 1 — MULTIPLE CHOICE

Circle the letter of the best answer. (2 points each)

1. When completing an AIP Authorization Checklist for a proposed AI workflow, the first category checked must be:

A. Model selection and performance thresholds B. Compute resource requirements and cost estimate C. Whether the use case falls in a prohibited category (lethal autonomous targeting, unchecked personnel evaluation, classification without review) D. Whether the workflow will integrate with an existing Ontology Object Type

2. In AIP Logic, the "Draft → Review → Publish" workflow enforces:

A. That all AIP workflows are reviewed by the MSS program office before deployment B. That pipeline dependencies are validated before the AIP workflow is activated C. That the workflow code is reviewed by a security engineer before reaching production D. That no AIP Logic output is released to operational users without passing through an authorized human review step

3. A human-in-the-loop (HITL) requirement for an AIP Logic workflow is classified as:

A. A best practice — recommended but configurable based on time constraints B. NON-NEGOTIABLE — every AIP Logic workflow must include a human review gate before outputs are acted upon or distributed C. Required only for workflows processing SECRET or above data D. Required only when the AIP workflow modifies Ontology Objects directly

4. You are building an AIP Logic workflow that processes incoming OSINT document summaries. Each document averages 8,000 tokens but your LLM context window is 4,096 tokens. The correct approach is:

A. Truncate each document to 4,096 tokens before sending to the LLM B. Implement a chunking strategy that splits each document into overlapping segments fitting within the context window C. Switch to a model with a larger context window without further configuration D. Pre-summarize documents using a pipeline before feeding them to AIP Logic

5. In a RAG architecture for an MSS operational workflow, the correct sequence of operations is:

A. Prompt → LLM → Retrieval → Context injection → Output → Write to Ontology B. Write to Ontology → Retrieval → LLM inference → Output C. LLM inference → Retrieval of supporting documents → Output rewrite → Write to Ontology D. Retrieval → Context injection into prompt → LLM inference → Output review → Write to Ontology

6. An AIP Logic workflow writes a summarized lesson learned to an Ontology Object upon completion. Per SL 4H standards, this write action must:

A. Occur automatically after the LLM inference step to minimize latency B. Require data steward approval for every individual write operation C. Be performed by a separate pipeline, not within the AIP Logic workflow D. Be gated behind a human review step — the output is written only after a reviewer approves it

7. When red-teaming an AIP Logic workflow against the USAREUR-AF AI Output Validation Framework, you should test for:

A. Model inference latency under peak user load B. Prompt injection vulnerabilities, output hallucinations, classification boundary violations, and role-constraint bypass attempts C. Whether the workflow can be run by users without Editor access D. Whether the LLM's output matches expected results across 100 standard test cases

8. You are building an AIP Logic workflow to extract key entities (unit names, grid coordinates, equipment types) from field reports. The LLM is producing inconsistently formatted outputs. The correct iterative improvement approach per SL 4H is:

A. Revise the prompt to explicitly specify the output schema, add few-shot examples, test against a sample set, score outputs, and iterate until the format is consistent B. Fine-tune the LLM on correctly-formatted examples to lock in the output format C. Add a post-processing script that normalizes any output format into the target schema D. Switch to a rules-based entity extraction system instead of an LLM

9. Military terminology (DTG, DODAAC, MTOE, OPORD) in AIP Logic prompts is important because:

A. It reduces token count by using abbreviations B. It is required by Army CIO policy for all AI prompts on Army networks C. Correct terminology in the prompt improves the model's ability to correctly process and extract operationally relevant entities from military documents D. It prevents the LLM from generating responses in non-military language

10. An Agent Studio agent is configured with multiple tools including a database query tool and an email send tool. The correct authorization control configuration is:

A. Define explicit tool-use authorization rules for each tool, specifying which tools can be used under which conditions and requiring human approval for irreversible actions (e.g., sending email) B. Grant the agent access to all tools by default and rely on the agent's own judgment to use them appropriately C. Allow the agent to use any tool it determines is relevant without human oversight D. Disable the email tool — agents may not send external communications under any circumstances

11. When deploying an AIP Logic workflow to production, the monitoring configuration must include:

A. A notification to the MSS program office every time the workflow runs B. Output quality metrics, alert thresholds for anomalous outputs, and a defined rollback procedure if output quality degrades C. A weekly manual review of all outputs stored in the Ontology D. Real-time display of LLM inference tokens consumed per run

12. An AIP Logic workflow that processes maintenance records produces a structured JSON output. The field `action_required` contains values like "yes", "YES", "Yes", "1", and "true" for the same condition. This is an example of:

A. A prompt injection vulnerability B. A data type mismatch in the target Ontology property C. Inconsistent output schema — the prompt must be revised to enforce a strict output format for this field D. Normal LLM variability — post-processing normalization is always required for JSON outputs

13. Per USAREUR-AF AI policy, which of the following use cases is PROHIBITED for AIP Logic automation?

A. Automatically assigning a performance evaluation score to a Soldier based on AI analysis of their records without human review B. Generating a draft equipment readiness summary for analyst review C. Summarizing incoming maintenance work orders into a daily digest for the G4 NCOIC D. Extracting unit designations and grid references from OSINT reports for analyst review

14. A Python transform that extracts Ontology data for AIP Logic context must handle pagination when:

A. The result set exceeds the single-page record limit and additional pages must be fetched to retrieve all records B. The Ontology dataset contains more than 100 columns C. The transform runs in an incremental (@incremental) mode D. The Ontology Object Type has more than 10 Link Types

15. After a production AIP Logic workflow produces an output that is later found to be incorrect, the rollback procedure requires:

A. Deleting the affected output records from the Ontology and rerunning the workflow manually B. Contacting the Palantir support team to restore the previous model version C. Using the documented rollback plan — reverting to the previous workflow version, notifying affected users, and investigating root cause before re-deploying D. Disabling the workflow and rebuilding it from scratch

16. Per ADP 3-13, the doctrinal principle governing AI/ML integration into operational workflows is:

A. AI systems should fully automate repetitive staff tasks to free human capacity for higher-order analysis B. AI enables speed; humans provide judgment — no AI/ML output replaces commander decision authority C. AI systems are authorized for autonomous action when latency requirements exceed human reaction time D. AI outputs are considered equivalent to human analysis once validated during initial testing

17. FM 2-0 describes the intelligence cycle as Processing, Exploitation, and Dissemination (PED). When mapped to an AI/ML pipeline on MSS, the "Exploitation" phase corresponds to:

A. Data ingestion and ETL transforms that structure raw feeds into datasets B. Model training, inference, prompt-based reasoning, and pattern detection under analyst supervision C. Delivering formatted outputs to Workshop dashboards and Contour reports D. Storing processed data in the Ontology for downstream retrieval

SECTION 2 — SHORT ANSWER

Answer in 2–5 sentences. (6 points each)

SA-1. You are asked to build an AIP Logic workflow that reads incoming field reports, extracts key entities (unit, location, equipment status), and writes the results to an Ontology Object for analyst review. Describe the complete architecture of this workflow including the RAG retrieval step, context injection, human review gate, and Ontology write step.

SA-2. A new AI engineer on your team argues that the human-in-the-loop requirement slows operations and proposes removing it for "low-risk" workflows. Write your response, citing SL 4H policy and the operational risk of removing the review gate.

SA-3. Describe the red-teaming process for an AIP Logic workflow under the USAREUR-AF AI Output Validation Framework. What specific failure modes do you test for, and what is the minimum requirement to pass before production deployment?

SA-4. An Agent Studio agent with access to an Ontology query tool and a Workshop form submission tool is producing unexpected results — it is submitting Workshop forms without the user explicitly requesting form submission. Describe the root cause and the correct fix using SL 4H tool-use authorization controls.

SA-5. Describe the complete AIP Authorization Checklist process for a proposed workflow that summarizes incoming OSINT reports and writes draft summaries to an analyst review queue. Identify any prohibited categories that must be checked, and describe what happens if a use case fails the checklist.

SA-6. Describe how AIP Logic AI engineering capability supports two WFF functions. For each, identify the WFF track (SL 4A through SL 4F) and give a concrete example of an AIP workflow that supports decision-making in that function.

SCORING SUMMARY

Section	Questions	Points Each	Total Points
Multiple Choice	17	2	34
Short Answer	6	6	36
Total	—	—	**70**

Passing: 49/70 (70%) — Post-test only. Pre-test is diagnostic.

ANSWER KEY — INSTRUCTOR USE ONLY

Do not distribute to students.

Multiple Choice: 1. C — Prohibited category check is the first and most critical gate on the AIP Authorization Checklist. 2. D — Draft→Review→Publish enforces human review before operational distribution. 3. B — HITL is NON-NEGOTIABLE — not configurable, not conditional on classification level. 4. B — Chunking with overlap is the correct strategy for documents exceeding context window. 5. D — Retrieval → context injection → LLM → review → write is the correct RAG sequence. 6. D — Ontology write must be gated behind human review, not automatic after inference. 7. B — Red-teaming tests prompt injection, hallucinations, classification boundary violations, role bypass. 8. A — Prompt revision with explicit schema, few-shot examples, test/score/iterate is the correct approach. 9. C — Military terminology in prompts improves entity extraction accuracy for military documents. 10. A — Explicit tool-use authorization rules with human approval for irreversible actions is correct. 11. B — Output quality metrics, alert thresholds, and rollback procedure are required monitoring elements. 12. C — Inconsistent output format requires prompt revision to enforce strict schema, not normalization. 13. A — Automated personnel evaluation scoring without human review is a prohibited use case. 14. A — Pagination handling

is required when result sets exceed the single-page record limit. 15. C — Rollback plan: revert workflow version, notify users, investigate root cause before re-deploying. 16. B — AI enables speed; humans provide judgment — no AI/ML output replaces commander decision authority (ADP 3-13). 17. B — Exploitation maps to model training, inference, prompt-based reasoning, and pattern detection under analyst supervision.

Short Answer Guidance:

SA-1. Full credit: (1) incoming field reports retrieved from Ontology or dataset; (2) RAG: retrieve relevant reference documents (unit lists, equipment codes) from corpus to enrich context; (3) context injection: field report + retrieved reference context assembled into prompt; (4) LLM inference extracts entities in structured JSON; (5) output routed to analyst review queue (Draft status in Ontology); (6) analyst reviews, edits if needed, approves → status changes to Reviewed; (7) approved record written to Ontology Object. All six steps must be present for full credit; HITL gate is required.

SA-2. Full credit: response must cite — HITL is NON-NEGOTIABLE per SL 4H policy (not "low-risk" configurable); operational risk: incorrect AI output without review could corrupt Ontology data, generate false operational reports, or trigger incorrect downstream actions; the time cost of review is deliberate — it ensures accountability and accuracy in operational data; the correct solution for speed is to optimize the review workflow, not remove it. Partial credit (3 pts) for citing policy without operational risk argument.

SA-3. Full credit: red-teaming tests: prompt injection (inject instructions in simulated user input or retrieved documents); hallucination probing (inputs designed to elicit fabricated facts); classification boundary violations (inputs that test whether the workflow handles CUI/SECRET content correctly); role-constraint bypass (attempt to make the workflow act outside its defined scope); minimum to pass before production: all critical failure modes resolved, no prompt injection vulnerabilities, output schema consistent across 100+ test cases, HITL gate confirmed functional. Partial credit (3 pts) for listing test types without minimum threshold requirement.

SA-4. Full credit: root cause — agent's tool-use authorization for the Workshop form submission tool does not require explicit user request or confirmation; agent is autonomously deciding to submit forms as part of its planning process; correct fix — configure the form submission tool with an authorization rule requiring explicit user confirmation before execution; all irreversible Actions (form submission, Ontology writes, external sends) must require human-in-the-loop approval in Agent Studio configuration. Partial credit (3 pts) for identifying root cause without specific fix.

SA-5. Full credit: AIP Authorization Checklist steps — (1) check prohibited categories: autonomous lethal action (N/A), unchecked personnel evaluation (N/A — this is OSINT, not personnel), automated official release without review (applies — check: is there a human review gate? Yes — passes); (2) verify HITL gate exists (draft summaries to analyst queue = HITL present — passes); (3) verify classification handling (OSINT at what level? — confirm appropriate network and handling); (4) confirm use case complies with Army CIO Memo (April 2024); if use case fails any checklist item, it must be redesigned or formally waived through the command AI governance process — not deployed as-is. Full credit requires checking prohibited categories AND describing failure consequence.

SA-6. Full credit: any two WFF tracks correctly identified with an AIP example — SL 4A (Intelligence): AIP Logic workflow extracts entities from OSINT reports to populate an intelligence Object Type for analyst review; SL 4B (Fires): AIP workflow processes battle damage assessment reports and classifies target status for fires coordination; SL 4C (Movement & Maneuver): AIP workflow summarizes route reconnaissance reports for maneuver planners; SL 4D (Sustainment): AIP workflow processes incoming supply requests and flags priority shortfalls for G4 review; SL 4E (Protection): AIP workflow analyzes threat reports and categorizes force protection advisories; SL 4F (Mission Command): AIP workflow drafts SITREP summaries from unit reports for commander review before distribution. Each response must identify the correct SL 4 letter (A–F) and provide a concrete AIP workflow example with HITL noted for full credit.

USAREUR-AF Operational Data Team TM-40H Post-Test | Version 1.0 | March 2026