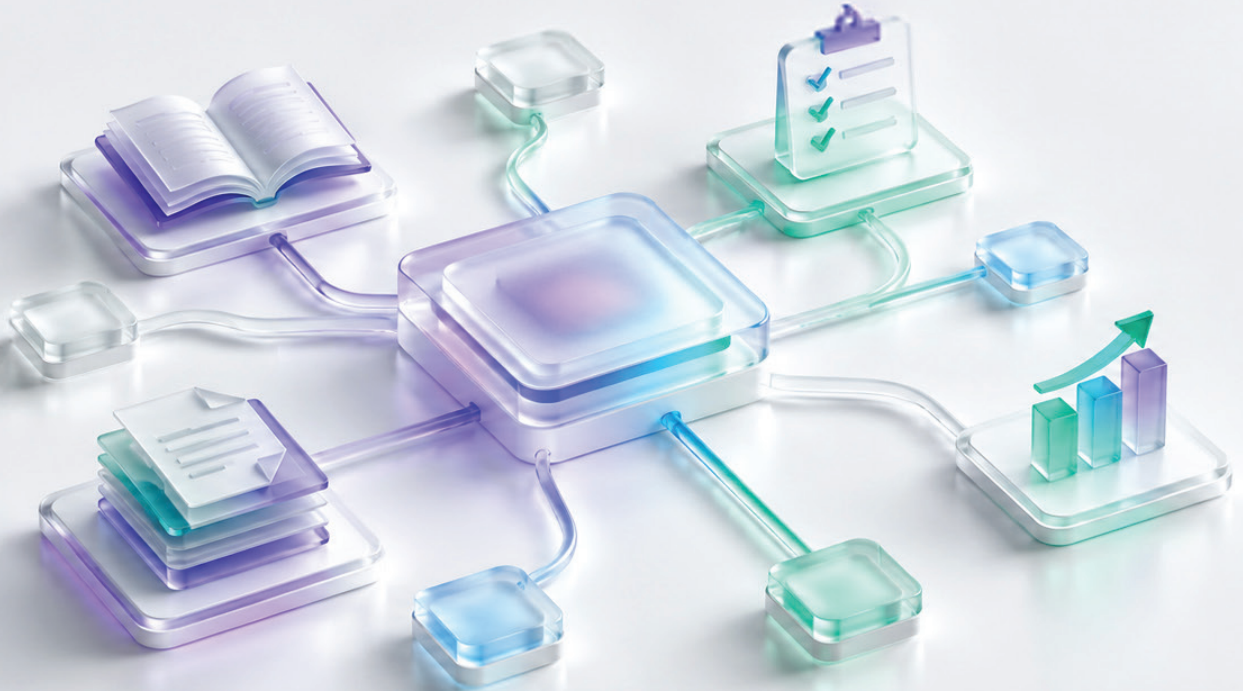


Standards for AI Agents in Learning and Training

Version 1.0 • October 2026

Document Type: Standards • Document Classification: Public



Purpose

This document establishes the standards and requirements governing the use of AI agents in learning and training in Saudi Arabia. It aims to help institutions select the optimal use, define agent permissions and user responsibilities, and verify the value and integrity of the agent's outcomes. The intended value may include improvements in learning, service quality, or operational efficiency, depending on the purpose of use. The standards define the outcomes that can be achieved and the evidence that may be used to verify them, while enabling institutions to choose the implementation methods, procedures, and tools relevant to their operations.

Intended value



Improved learning



Service quality



Operational efficiency

Scope of Application

This document applies to K-12 education, higher education, technical and vocational education and training (TVET), and lifelong learning across government, private, and nonprofit institutions. It is intended for learning and training institutions, vendors and developers, practitioners, learners, and competent authorities responsible for regulation and supervision, each within the scope of their respective roles and authorities. The standards set out in this document are common across sectors. Their application varies according to the task, the target audience, the data, the permissions, and the impact. Any additional protection requirements are specified under the relevant standard, such as ensuring age-appropriate interactions for minors or verifying practical competence. This document does not establish sector-specific standards. A practitioner's use of an AI agent to perform functions on behalf of an institution or to process an institution's or a learner's data is considered institutional use when applying these standards.

Uses of This Document

This document guides the application of standards to AI agent use cases through four steps:

1

Verification

Determine whether the system is classified as an AI agent subject to the requirements of this document, in accordance with the concepts set out in the Second Section.

2

Classification

Identify the agent's area of application and use characteristics, including the agent's level of autonomy and the potential impact of its use, in accordance with the Third Section.

3

Application of the Standards

Review the seven dimensions set out in the Standards Section and apply the requirements according to the level of impact, providing justification for inapplicable requirements.

4

Application of Area-Specific Requirements

Apply the additional requirements to each area of application relevant to the use case, in accordance with the last section of this document.

Concepts and Scope of Application

Artificial Intelligence (AI) Tool

An AI-based tool used to generate outputs or provide assistance. It may incorporate AI agent capabilities; however, responding to a request is not sufficient to classify it as an agent. This term does not refer to the tools that an AI agent uses to perform its tasks.

AI Agent Used in Learning and Training

An agent that performs a learning or training task or supports tasks related to research, services, administration, or governance, whether it was designed for this purpose or is a general-purpose agent.

AI Agent

A system assigns a task or goal that determines and executes certain steps required to achieve it and monitors the results with a degree of autonomy by interacting with its environment, tools, or associated systems, within the scope of its delegated authority.

Agentic AI

An approach or system that determines, executes, and monitors work processes to achieve delegated goals. It may rely on a single agent or multiple agents operating at varying levels of autonomy.

Agent Tools

The functions, services, and systems that an agent uses to gather information or perform its tasks, whether or not they use AI, such as search engines, databases, learning management systems (LMSs), and email.

Multi-Agent System

A system in which two or more agents interact to distribute, coordinate, or review tasks, with outcomes influenced by the agents' performance and the way they interact with one another.

Concepts and Scope of Application

Continued

Predefined Automated Workflow

The execution of predefined steps and pathways, which may incorporate AI tools, without independently selecting the course of action required to achieve a delegated goal. The workflow remains automated even when an AI tool selects one of its predefined branches, provided that the system does not independently assess execution results to determine the next steps.

Delegation

The tasks, permissions, and boundaries within which an agent is authorized to operate.

Autonomy

The extent to which an agent can perform tasks without requiring prior human approval for each task, within the limits of the authority delegated to it.

User

An individual who interacts with an agent or assigns a task to it.

Institution

Any government, private, or nonprofit institution that uses AI agents in learning or training, or offers them to users, and is responsible for their use.

Beneficiary

An individual whose learning, rights, opportunities, safety, or data are affected by an agent's actions or outputs, even if the individual does not directly use the agent. The individual may also be a direct user.

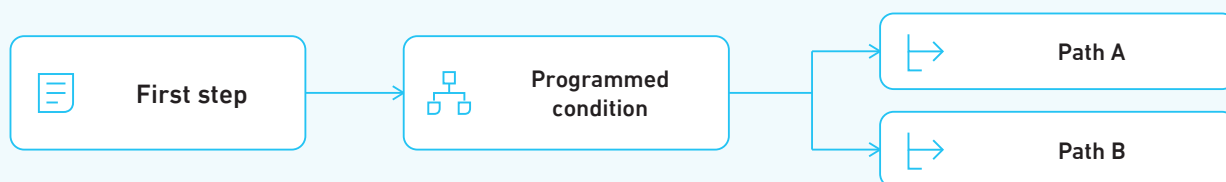
Practitioner

A teacher, faculty member, instructor, or any individual responsible for instructional design, evaluating learning, or providing learning support to learners.

Automated Workflows, AI Tools, and AI Agents

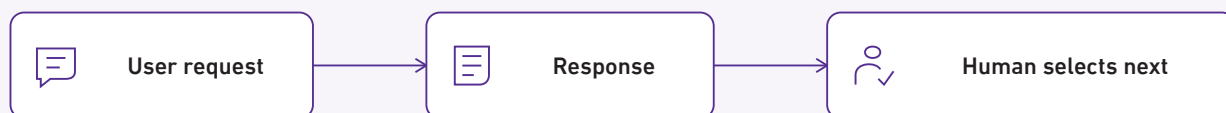
The distinction between automated workflows, AI tools, and AI agents depends on how a system determines and executes task steps and monitors the results. The differences are outlined below, in ascending order of autonomy:

Automated Workflows



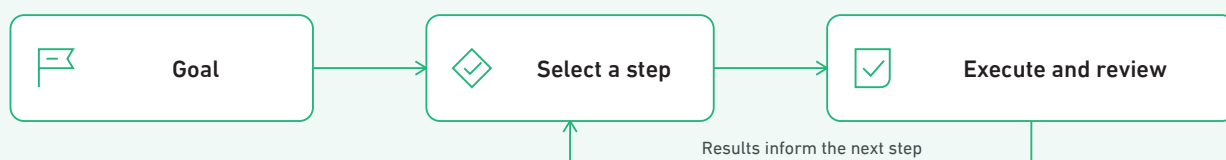
Execute predefined steps and pathways, including programmed branches. For example, a system may be programmed to send a review activity to each learner whose score on a quiz falls below a specified threshold, and an enrichment activity to each learner whose score exceeds that threshold. An automated workflow may incorporate AI tools at certain steps, but this does not necessarily mean that it selects its own course of action.

AI Tool



Responds to a specific request and generates an output, while the human remains responsible for deciding the next step. For example, a teacher may ask an AI tool to develop a specific learning activity and then carry out the steps of the task.

AI Agent



An AI agent is assigned a goal, determines and executes the steps needed to achieve it, reviews the results, and adjusts its course of action accordingly, within the authority delegated to it. For example, an agent may be assigned the task of developing activities to achieve a specific learning objective. It then selects resources and steps, reviews its results, and adjusts its work accordingly.

Understanding Agent Autonomy

Figure (1): Distinguishing Between Automated Workflows, AI Tools, and AI Agents

The AI agent may operate according to predefined steps and is not required to use a language model. What distinguishes it from an automated workflow is its ability to review the output of each step and use it to determine the next step until the goal is achieved. "AI Agent" refers to the system that performs the task, whereas "Agentic AI" describes the approach or capabilities of the system. The two terms do not necessarily represent separate categories. Nor does an AI agent require multiple agents, full autonomy, or self-learning during use. The agent may operate within the conversation, for example, by selecting activities for a learner and adapting the support provided based on the learner's responses, with the aim of achieving a specific learning goal through multiple interactions, without connecting to an external system.



Goal-directed choices



Review of results



Bounded authority

Agent Components and How They Work



A clear goal and two layers of delegation boundaries



System prompts



Technical permissions

Core functions



Select the next step



Execute the step



Review the result

Resources as needed



Information sources



Memory



Agent tools



Other agents

Model or decision engine • Human oversight proportionate to impact

Figure (2): Agent Functions and the Components

The following figure illustrates the core functions of an AI agent and the components it may use to perform its task. The agent begins with a defined goal, selects a step based on the available information, executes it, and reviews the outcome to determine the next step or conclude the task within the boundaries of its delegated authority. The agent may draw on information sources, use memory to retain information needed for the task, use tools and systems to perform tasks, or collaborate with other agents to accomplish the task. The agent's architecture varies according to its intended purpose; it is not necessarily required to have long-term memory or to connect to external tools or other agents.

The agent relies on a model or decision engine to determine its steps. Its operating boundaries are established through two layers. The first consists of prompts that guide the model, together with technical permissions that determine what the agent can actually access and execute. The second layer enforces restrictions when the model fails to follow its instructions.

How a Learning Agent Works



Illustrative example: a middle-school mathematics learning assistant

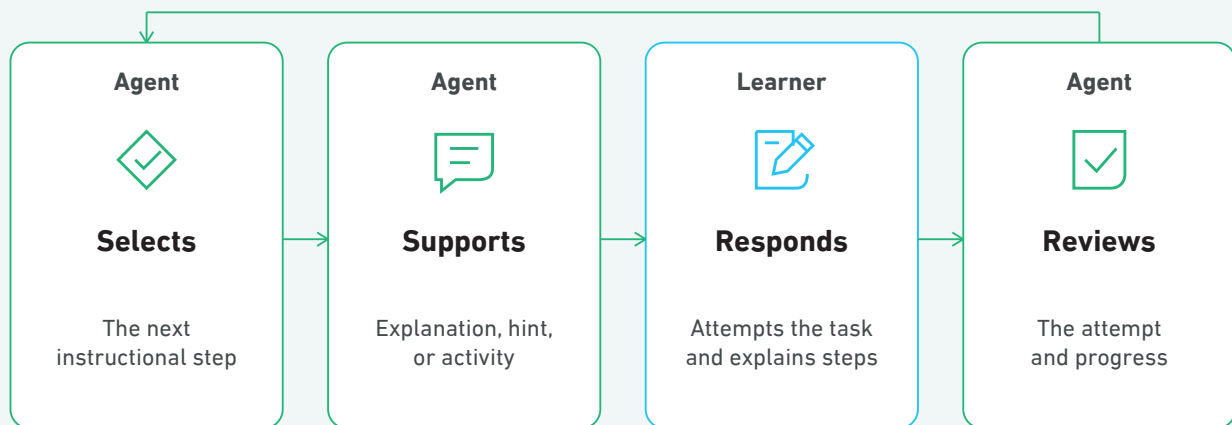


Teacher

Defines learning objectives and assistance boundaries

The agent's operating cycle

The next step adapts to the learner's response



The cycle repeats as needed



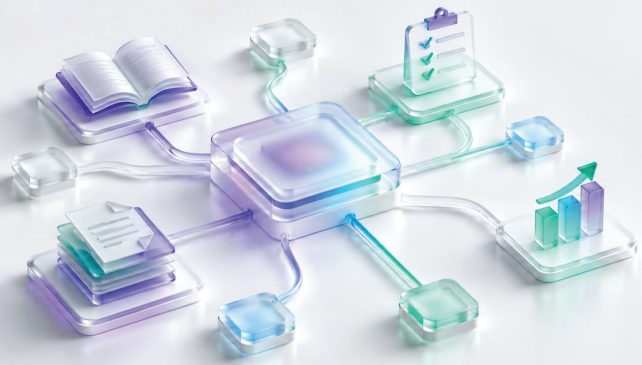
Return to the teacher - Human-in-the-Loop

The agent pauses for guidance or approval, then continues according to the teacher's decision.

Figure (3): The Agent's Operating Cycle and Boundaries for Human Judgment

The following figure illustrates how an agent operates using the example of a mathematics learning assistant for middle school students. The teacher defines the learning objectives and boundaries of the provided assistance. The agent then selects an instructional step and provides support, reviews the student's attempt, and adjusts the next step based on the student's response. This process may incorporate a Human-in-the-Loop approach, whereby the agent pauses to seek the teacher's guidance or approval before proceeding.

Areas of Application



The following table presents examples of eight areas in which AI agents may be used in learning and training, including examples of agent tasks, intended users, and beneficiaries. A single use case may fall within more than one area, in which case the requirements relevant to each applicable area must be applied.

Area	Examples of Agent Tasks	Users and Beneficiaries
Teaching and Educational Practitioner Support	Planning lessons, adapting activities, and identifying learning difficulties.	Used by the practitioner. Learners are affected by decisions relating to teaching.
Learning, Practice, and Personalization	Selecting activities suited to the learner's level and adapting them according to the learner's progress.	Used by the learner. The learner's learning opportunities and level of support are affected.
Evaluation and Measurement	Preparing evaluation tasks, analyzing performance, and suggesting grades.	Used by the practitioner or evaluating entity. The learner is affected by the decision.
Guidance and Support	Providing academic and career guidance and adapting content.	Used by the learner or a specialist, such as an academic or career mentor. The learner is affected by the recommendations provided.
Educational Content and Resources	Collecting, verifying, producing, and updating resources	Used by the content creator. Learners and beneficiaries are affected.
Research and Knowledge Management	Collecting and analyzing evidence and organizing knowledge.	Used by a researcher or employee. Participants and individuals targeted by the findings are affected.
Educational Services and Administration	Processing registration, scheduling, transactions, and communications.	Used by an employee or service applicant. Individuals whose transactions or records are involved are affected.
Governance and Decision Support	Analyzing performance indicators and resource-use efficiency and comparing decision alternatives.	Used by leadership or the relevant institution. Individuals and institutions are affected by the decisions made.

Use Characteristics



The requirements applicable to an agent's use within a given area vary according to the users and beneficiaries, the data, the permissions granted to the agent, and the anticipated impact of its use. The following table outlines the use characteristics that help determine the applicable requirements and the level of protection, supervision, and verification.

Component	What Should Be Specified
 Purpose and Area of Application	Task, intended outcome, relevant areas of application, and usage limits.
 Individuals and Context of Use	Users and beneficiaries, sector, age groups, accessibility needs, and whether the use is institutional or personal.
 Data and Systems	Sensitivity and sources of the data, data retained by the agent, and the tools, systems, and parties associated with it.
 Autonomy and Delegation	The tasks the agent is authorized to select and perform, the resources available to it, who grants its authority, and the tasks requiring human approval.
 Potential Impact	The potential effects of use or failure on learning, rights, opportunities, safety, and services, including the severity and scope of potential harm and the possibility of remedying it.

Autonomy and Impact



The agent's level of autonomy and the impact of its use help determine the level of protection and verification. This document uses three descriptive levels of autonomy:

1



Provide outputs or recommendations.

2



Execute tasks after obtaining human approval.

3



Execute tasks within defined boundaries without requiring approval for each action.

These levels measure what the agent is permitted to execute without prior human verification, rather than how it selects steps while performing a task. The usage may involve more than one level; each task is therefore classified according to the authority granted to the agent.

Impact levels

Impact is assessed independently of autonomy. For example, a recommendation to reject a learner's admission may have a high impact even if the agent is not authorized to implement it.

Limited impact

Impact is limited when the agent's outputs can be corrected without significant consequences.

Moderate impact

Impact is moderate when they affect the learning pathway or service quality in ways that require expert review.

High impact

Impact is high when it affects a consequential decision or the safety of individuals; involves data whose misuse or unauthorized disclosure could cause serious harm; or has widespread effects or consequences that are difficult to remedy.

Protection and Verification Matrix



The following matrix illustrates how the required level of protection and verification varies according to the agent's autonomy and the impact.

Agent autonomy level	Limited impact	Moderate impact	High impact
Provides outputs or recommendations (The individual decides and acts).	Verification proportionate to the task.	Expert review before approval.	Rigorous verification and an accountable human decision.
Executes after human approval (The individual approves each task).	Verification of the approved task.	Review the results and verify whether they can be corrected.	Approval by an authorized decision-maker following review.
Executes within defined permissions (The individual monitors and has the authority to stop execution).	Restricted permissions and the ability to stop execution.	Stronger evidence and monitoring of impacts.	Autonomous execution must not be permitted for actions affecting consequential decisions covered under Standard 5. Autonomous execution is limited to reversible actions, subject to restricted delegation, continuous monitoring, and the ability to stop execution immediately.

The Standards



The standards define the fundamental requirements for using AI agents in learning and training to achieve verifiable educational and institutional value. Thus, maintaining the role and responsibility of human judgment, safeguarding users and beneficiaries, and protecting data and rights. These standards apply to various fields of AI agent utilization mentioned in the previous section; they address aspects of quality, protection, and responsibility that should be considered regardless of the agent's task or context of use.

These standards are structured into seven interconnected dimensions, as demonstrated in the following figure. Each dimension includes specific standards and verifiable requirements. When evaluating any use, all seven dimensions are considered, and the applicable requirements are determined based on the characteristics of the use, the agent's level of autonomy, the potential impact of its tasks and outputs, and the respective responsibilities and authorities of the parties involved. This structure enables applying the standards according to the type of use, while stating the reason for the non-applicability of any requirement when its application conditions are not met.



Figure (4): Dimensions of Key Standards

1

Dimension 1

Value and Contextual Relevance

S1

Clarity of Purpose and Value

Standard Description

The use of the agent is linked to a specific and verifiable need and value.

Requirements

1

The institution specifies the task, the desired outcome, the users, the beneficiaries, the scope of use, and the reason for selecting the agent to perform the task.

2

The institution defines indicators to measure the achieved value, whether it is an improvement in learning, service, or work efficiency, while taking into account effort, cost, and unintended consequences.

3

The benefits of using an agent are based on supporting evidence; the quality or speed of the outcome alone is not proof of improved learning.



Responsible Parties

The institution, the practitioner, or the user, each according to their role.



Scope of Application

All types of use, according to their scale and impact.



Verification Evidence

Clear definition of the purpose and expected value, indicators to measure its achievement, and evidence supporting the announced benefits.

1

Dimension 1

Value and Contextual Relevance

S2

Relevance to Learning and Training

Standard Description

The use of the agent is appropriate to the task, the beneficiaries, and the educational and cultural context in Saudi Arabia.

Requirements

1

The institution verifies the suitability of the agent's use for learning objectives or service requirements, and its consistency with the role of the practitioner and the responsibilities of the competent authority.

2

The use of the agent should be considered based on the users' levels, experiences, languages, digital skills, and needs, while verifying the correctness and clarity of Arabic when used.

3

The institution verifies the appropriateness of the content and interaction method to local values and context. In addition to verifying the applicability of the submitted evidence regarding the agent's suitability to the target group and environment.



Responsible Parties

The institution, the practitioner, and the provider, each within the scope of their role and responsibilities.



Scope of Application

When selecting the agent, designing its use, and offering it to new beneficiaries.



Verification Evidence

Results of expert review and tests using samples representing the target groups and tasks, and evidence of addressing shortcomings.

1

Dimension 1

Value and Contextual Relevance

S3

Requirements and Risks Alignment

Standard Description

The agent's authority, protective measures, and required evidence are aligned with the impact and risks of use.

Requirements

1

The institution defines the users, beneficiaries, type of data, level of autonomy, and the potential impact of error or misuse. It also reviews these factors whenever the context of use changes.

2

The institution assesses the likelihood and severity of potential harm, as well as the residual risks following the implementation of protective measures. In doing so, it distinguishes impact from the level of autonomy and takes into account the extent of the harm, the difficulty of remedying it, and the groups requiring additional protection.

3

The institution identifies the party authorized to accept the risks or escalate them to the management level. It restricts or suspends use when the risks exceed manageable levels and defines the scope of any trial before authorizing broader deployment.



Responsible Parties

The institution, in collaboration with the provider and relevant experts.



Scope of Application

All institutional uses, according to their impact.



Verification Evidence

Risk assessment clarifies the reasons for its classification and is reflected in the limits of use, review level, and protection.

2

Dimension 2

The Human Role and Scope of Delegation

S4

Delegation and Authorities

Standard Description

The agent operates within clearly defined tasks, authorities, and limits established by the party granting its authorization.

Requirements

1

The institution defines the tasks the agent is authorized to perform, the data and resources it may access, the permitted execution time and resource consumption for each task, the authorization period or termination conditions, and the actions requiring human approval.

2

Any tasks the agent delegates to other tools or agents must remain within the original scope of authorization and must not be expanded by instructions contained in external content, memory, or services.

3

The responsible party verifies the effectiveness of the authorization limits and the possibility of modifying and withdrawing them through the components associated with the task.



Responsible Parties

The institution, practitioner, or user granting the delegation, within the scope of the authorized permissions.



Scope of Application

Whenever an agent chooses or performs acts on behalf of another party.



Verification Evidence

Defined boundaries and aligned authorities, and tests showing how to handle the attempts to bypass them.

2

Dimension 2

The Human Role and Scope of Delegation

S5

Human Supervision and Core Decisions

Part 1 of 2

Standard Description

The ultimate decision remains with the authorized individual, in addition to effective human supervision over what the agent provides or executes. The human review does not allow usage that is prohibited by the applicable controls.

Requirements

1

Define the decisions that require human approval, including decisions related to admission, final grades, certificates, penalties, rights, and safety.

2

The reviewer has the information, time, and authority to understand, reject, or amend the result, stating the accuracy and the basis of the recommendation.



Responsible Parties

The institution and the decision-maker according to the allowed supervision capabilities.



Scope of Application

Applications involving consequential decisions, each subject to human approval.

Continued on the following page: remaining requirements and verification evidence.

2

Dimension 2

The Human Role and Scope of Delegation

S5

Human Supervision and Core Decisions

Part 2 of 2

Requirements — continued

3

The agent must stop before proceeding with any task that requires approval. The absence of a reviewer or the inability to obtain approval does not constitute consent to proceed.

4

The effectiveness of supervision during use is assessed, including the supervisor's ability to detect errors or boundary violations and intervene promptly to prevent or mitigate harm.



Verification Evidence

Documented examples of human review and formal decision-making, demonstrating the ability to reject a recommendation or prevent its implementation.

2

Dimension 2

The Human Role and Scope of Delegation

S6

User Role and Competence

Standard Description

Usage maintains the role of the learner, practitioner, and decision-maker, ensuring they perform the thinking, estimation, and decision-making required by the task.

Requirements

1

The institution clearly defines the agent's responsibilities and distinguishes them from the learning, judgment, and accountability that remain with the individual.

2

The institution ensures that users and supervisors understand the agent's capabilities and limitations, how to verify its outputs, and how to seek assistance when needed.

3

The institution shall assess whether the use of the agent undermines essential skills or the ability to exercise judgment and make decisions and adjust the level of assistance accordingly, while taking legitimate support into account.



Responsible Parties

The institution, the practitioner, and the decision-maker.



Scope of Application

Uses that affect learning, professional practice, or decision-making.



Verification Evidence

Clear distribution of roles, and evidence of the user or supervisor's ability to perform tasks that remain within their responsibility.

3

Dimension 3

Responsibility and Transparency

S7

Defining Responsibilities

Standard Description

Clarity of responsibility and use outcomes, including cases where more than one provider or agent is involved.

Requirements

1

The institution appoints an individual responsible for overseeing the use of the agent and clearly allocates responsibilities for design, provision, decision-making, monitoring, and error handling among the parties involved.

2

The provider documents the components and versions relevant to the task in sufficient detail to establish traceability to those who performed and authorized it.

3

The service provision arrangements include the verification evidence, change and incident reporting, support, and service termination provisions, while ensuring that each party retains its applicable obligations.



Responsible Parties

The institution and provider, including internal development and external services.



Scope of Application

Institutional use, its contracts, and its affiliations.



Verification Evidence

Clear assignment of responsibility for each consequential outcome, with established procedures for sharing evidence and handling errors.

3

Dimension 3

Responsibility and Transparency

S8

Disclosure and Interpretation

Part 1 of 2

Standard Description

The responsible institution shall provide users and beneficiaries with clear information about the agent's role, limitations, and designated point of contact, to the extent required for them to understand the service or outcome.

Requirements

1

The disclosure explains the agent's purpose and limitations, as well as how data is processed through its use, in language appropriate to the target audience.

2

Content generated or recommended by the agent is clearly distinguished from human approval, and an understandable explanation is provided of the basis for material outputs, including the sources, data, and rules relied upon and the steps performed by the agent, without requiring the disclosure of protected information or the model's internal reasoning.



Responsible Parties

The institution and the provider and the practitioner when presenting results or providing service.



Scope of Application

The user and the beneficiary, according to the type of the interaction and the outcome.

Continued on the following page: remaining requirements and verification evidence.

3

Dimension 3

Responsibility and Transparency

S8

Disclosure and Interpretation

Part 2 of 2

Requirements — continued

3

When a child uses or interacts with the agent, information is presented in a manner appropriate to the child's age and the nature of the impact, taking into account the roles of the guardian and the practitioner and obtaining any required approvals in accordance with applicable regulations.



Verification Evidence

Clear disclosure tailored to the target audience, with an explanation of the basis and limitations of the output that influences decisions.

3

Dimension 3

Responsibility and Transparency

S9

Reviewing Results and Objecting

Standard Description

Tasks and consequential outcomes are traceable, reviewable, and correctable, without imposing technical burdens on users or beneficiaries beyond their capabilities.

Requirements

1

Records and supporting evidence are retained to document the tasks performed, inputs and sources used, authorizations granted, relevant component versions, and human approvals where required. Where necessary to understand the outcome, the steps are linked in sequence.

2

A designated channel is available for contacting a human representative and requesting a review or correction within a timeframe appropriate to the impact of the decision.

3

The institution addresses errors and their consequences, notifies affected beneficiaries where necessary, and limits retained data to the minimum required for review and compliance with applicable obligations.



Responsible Parties

The institution and the provider, according to the information and authority each possesses.



Scope of Application

Uses that affect rights, decisions, records, or services.



Verification Evidence

Documented examples demonstrating that the consequential outcomes can be traced, reviewed, and corrected, without requiring all conversations to be retained or made accessible to everyone.

4

Dimension 4

Fairness and Protection

S10

Use and Interaction Safety

Part 1 of 2

Standard Description

The agent takes into account the safety, age, and needs of users and beneficiaries, with clearly defined limits on the support it provides and referral pathways to qualified professionals, such as psychologists, healthcare professionals, or social workers.

Requirements

1

The institution verifies the suitability of the agent's content and interactions and addresses the risk of harmful, misleading, or inappropriate guidance, particularly in sensitive health, psychological, social, and professional contexts

2

Specialized and safety-critical tasks remain subject to appropriate supervision and authorization. The agent does not replace a qualified specialist, and its conclusions alone are not used as the basis for a diagnosis.



Responsible Parties

The institution, provider, or supervising specialist.



Scope of Application

Interacting with users or tasks that may affect their safety, according to the source of the risk.

Continued on the following page: remaining requirements and verification evidence.

4

Dimension 4

Fairness and Protection

S10

Use and Interaction Safety

Part 2 of 2

Requirements — continued

3

When a child uses or is impacted by the agent, the effects on the child's safety, rights, and cognitive, psychological, and social development shall be assessed in light of the interaction and its potential impact. Parental or guardian consent does not replace the need to verify that the use is safe.

4

Support is available when signs of harm emerge, with the limits of the available response clearly defined, privacy safeguarded, and monitoring restricted to what is necessary and legally justified.



Verification Evidence

Targeted review of potential harmful scenarios, with clearly defined support and referral boundaries.

4

Dimension 4

Fairness and Protection

S11

Safeguarding User Autonomy

Standard Description

The agent respects users' choices and does not exploit their trust, vulnerabilities, or data to manipulate them or foster harmful dependence on the agent.

Requirements

1

The interaction design does not imply that the agent is conscious or capable of forming a reciprocal personal relationship, nor does it use emotional pressure to prolong engagement or influence user behavior.

2

The virtual nature of educational characters and simulations is clearly disclosed, and educational recommendations are kept separate from advertising and commercial interests. Users can end the interaction at any time without being misled or deceived, and children's data and interactions are not used for commercial targeting.

3

Usage patterns that undermine learning, decision-making, or human relationships are addressed, while recognizing that reliance on assistive technologies by people with disabilities is not inherently harmful.



Responsible Parties

The provider, the institution, and the practitioner in designing, presenting, and following up on the interaction.



Scope of Application

Direct interaction, personalization, and recommendations influencing choices.



Verification Evidence

Reviewing interaction and recommendation features, clarity of business interests, and addressing harmful dependency patterns.

4

Dimension 4

Fairness and Protection

S12

Fairness and Equal Opportunities

Standard Description

The use of the agent preserves access to learning opportunities and services and addresses unjustified disparities among users and beneficiaries.

Requirements

1

The institution examines differences in performance quality and outcomes across relevant groups, languages, and contexts and addresses any unjustified disparities.

2

The agent's conclusions about past capabilities or experiences remain subject to correction and review and do not become a permanent judgment that restricts the beneficiary's opportunities.

3

Alternative processing options are available when unjustified barriers arise, including disparities in digital skills and access to resources. Access to basic educational opportunities does not require users to purchase paid personal services.



Responsible Parties

The institution and provider and the practitioner when relying on classification, customization, or evaluation and measurement.



Scope of Application

Uses that may have different effects across groups or affect individuals' opportunities.



Verification Evidence

Results documenting differences across relevant groups, the measures taken to address them, and the limits of the verification performed.

4

Dimension 4

Fairness and Protection

S13

Accessibility and Participation

Standard Description

Users, including people with disabilities, can access basic features and participate fairly.

Requirements

1

The institution considers accessibility needs in interaction, content, disclosure, consent, and requesting assistance, according to applicable requirements.

2

The institution verifies the ability to complete essential tasks with support tailored to the target groups.

3

An alternative procedure is available when accessibility barriers arise, preserving the intended skill or purpose of the service without placing the burden of accessibility failures on users or limiting their opportunities as a result.



Responsible Parties

The institution, the provider, and the practitioner, according to the service provided.



Scope of Application

Services and experiences directed to users.



Verification Evidence

Completing essential tasks and requesting help and review, while stating and addressing known obstacles.

5

Dimension 5

Data, Privacy, and Rights

S14

Data Processing Legitimacy

Part 1 of 2

Standard Description

Data is processed for a specific and legitimate purpose and to the extent necessary, in accordance with relevant regulations, rules, and controls.

Requirements

1

The institution defines the roles of data controllers and processors, the legal basis and purpose of processing, the data required, data subjects' rights, and the data retention period.

2

The institution verifies the restrictions applicable to the data, environment, and tool classification, and the consent or legitimacy of the purpose does not constitute an open license for participation or model training.



Responsible Parties

The institution and the provider are according to their legal roles, and anyone who processes the data of others within the scope applies to them.



Scope of Application

Collecting, using, storing, sharing, or transferring data.

Continued on the following page: remaining requirements and verification evidence.

5

Dimension 5

Data, Privacy, and Rights

S14

Data Processing Legitimacy

Part 2 of 2

Requirements — continued

3

The agent's data sources, the information it transmits to providers, tools, or other agents, and any transfers outside Saudi, including onward transfers where applicable, are assessed before access or data sharing is enabled.

4

Data subjects can exercise their rights, including requesting data correction or destruction, in accordance with applicable regulations, and the impact of data processing is evaluated when required.



Verification Evidence

Clearly documented justifications and limits for all consequential data processing activities, with evidence of compliance with data restrictions and data subjects' rights.

5

Dimension 5

Data, Privacy, and Rights

S15

Archived Information Management

Standard Description

The accuracy of information stored or retrieved by the agent is appropriate to the task, and its use is limited to the authorized purpose and scope of authority, with each individual's data protected.

Requirements

1

Specify what information is retained in memory, the purpose of retaining it, and how long it needs to be kept, while allowing it to be corrected or deleted in accordance with applicable rights and obligations.

2

Individuals' data must not be combined, used for tasks, or shared with agents or services beyond the authorized scope and permissions.

3

The agent's conclusions are clearly distinguished from documented information, and the validity of outdated information is verified before any impactful action is taken.



Responsible Parties

The institution and the provider.



Scope of Application

Storing, retrieving, or transferring information between system components.



Verification Evidence

The ability to identify information sources, define the limits of data use, correct inaccurate information, and verify that individuals' data remains separate.

5

Dimension 5

Data, Privacy, and Rights

S16

Content and Output Rights

Standard Description

The agent observes intellectual property rights and the terms of use of sources and outputs according to the relevant provisions.

Requirements

1

Access to and use of content, including its modification, publication, and use in model training, are verified according to the applicable license terms or legal exemptions.

2

Clarify the contractual rights and responsibilities related to inputs and outputs, without assuming exclusive ownership of any automatically generated output.

3

The information needed to attribute sources, disclose the agent's contributions, and address rights-related claims is retained, while protecting confidential information.



Responsible Parties

The institution, the provider, the practitioner, the researcher, and the user, according to their respective roles and responsibilities.



Scope of Application

Gathering, producing, publishing, or reusing content.



Verification Evidence

Applicable licenses or legal grounds authorizing use and publication, along with sources that can be reviewed. Technical accessibility alone does not establish the right to use the content.

6

Dimension 6

Security and Reliability

S17

Output Accuracy and Limitations of Use

Standard Description

The agent's outputs are sufficiently reliable for the task, and its capability limitations are clearly communicated before its outputs are used.

Requirements

1

The accuracy, completeness, and relevance of the results are verified in the targeted tasks, including addressing errors, ambiguity, and conflicting information. Verification of the agent covers not only the result itself but also the execution path and compliance with the authorized scope. The same task is repeated as often as warranted by the potential impact of its use to assess the consistency of performance.

2

The agent distinguishes verified information from unconfirmed information and does not present unverified sources or execution results as facts or completed actions.

3

Define the circumstances requiring the agent to stop or refer the matter to a specialist when information is insufficient or safe execution is not possible.



Responsible Parties

The institution, the provider, and the practitioner responsible for the outputs relied upon.



Scope of Application

All uses, according to purpose and impact.



Verification Evidence

Results of repeated testing on real-world tasks, demonstrating success and failure rates, compliance with authorized implementation procedures, and the limitations of relying on the outputs.

6

Dimension 6

Security and Reliability

S18

Secure Access and Task Execution

Part 1 of 2

Standard Description

Technical and regulatory controls protect the ecosystem's data and resources, and the agent's actions are limited to the authorization granted to it.

Requirements

1

Access is restricted to what is required for the task. The identity of the agent, the authorizing party, and any associated parties are verified according to the sensitivity of the transaction, with the agent's authority to act on their behalf and the validity of the action clearly established. The agent is assigned its own dedicated access permissions and credentials, limited to the minimum scope and duration to perform the task. It must not use user credentials or permanent credentials. System prompts must not contain login credentials or confidential information, as these may be exposed through the agent's interactions.

2

The effectiveness of protection against prompt injection through content or tools, data leakage, privilege escalation, and resource misuse is verified. Security must not rely solely on the model's compliance with language instructions.



Responsible Parties

The institution, the provider, and the developer, in accordance with the scope of control each party has over its respective system.



Scope of Application

Operating the agent or linking it to resources, systems, or services.

Continued on the following page: remaining requirements and verification evidence.

6

Dimension 6

Security and Reliability

S18

Secure Access and Task Execution

Part 2 of 2

Requirements — continued

3

The agent's outputs are treated as untrusted inputs and validated before being passed to connected systems or implemented by them. All actions performed by the agent, including code execution, web browsing, and system interface control, are carried out in an isolated environment that limits the impact of errors or exploitation.

4

Applicable national cybersecurity controls are implemented, and their effectiveness is tested in the context of agent use and its integrations.



Verification Evidence

Tests demonstrating the effectiveness of access controls and compliance with authorized permissions during task execution. This includes testing the handling of untrusted sources; documenting known vulnerabilities and mitigating their impact before deployment; verifying that outputs are inspected before execution; confirming the isolation of execution environments; and restricting the scope and duration of data access granted to the agent.

6

Dimension 6

Security and Reliability

S19

Use Suspension and Addressing Errors

Standard Description

The possibility of restricting or suspending the agent, withdrawing its authorizations, and addressing its impacts, while continuing learning and essential services or providing a better alternative.

Requirements

1

The responsible party can suspend or restrict the affected use and revoke the associated authorization. New and pending tasks that fall outside the authorized scope, including those delegated to other components, are blocked. Ongoing tasks are stopped where feasible, with a record of the actions completed and any tasks that could not be halted.

2

Tasks are classified according to their reversibility, with enhanced safeguards applied before executing actions whose effects are difficult to reverse.

3

The consequences of errors or failures are addressed, evidence is preserved, and an alternative is provided if the service is interrupted or terminated.



Responsible Parties

The institution and the provider are based on the authorizations and affected services.



Scope of Application

Uses involving task execution or tasks on which learning or service delivery depends.



Verification Evidence

Demonstrated ability to halt execution and revoke authorization, address the consequences of completed actions, and maintain service continuity or provide a suitable alternative.

6

Dimension 6

Security and Reliability

S20

Integration Safety and Multi-Agent

Standard Description

The integration of agents, tools, and systems preserves data integrity, maintains authorization boundaries, and ensures clear accountability.

Requirements

1

Identify external components and parties that may affect the system, along with the data and permissions exchanged between them, and verify compatibility and limits of use.

2

Test interactions among components and the sequencing of tasks, including conflicting objectives, error propagation, and repeated execution of consequential tasks. The success of each component individually or consensus among agents on a single result is not sufficient.

3

Ensure that the service can be terminated or transitioned to an alternative provider while preserving the rights and data for continuity, without mandating a specific product or protocol.



Responsible Parties

The institution, the provider, and the developer.



Scope of Application

Integration with systems and services, or using more than one agent.



Verification Evidence

Test results demonstrating the reliability of integrations with connected systems, defined responsibilities, and the ability to disconnect integrations or migrate.

7

Dimension 7

Verification and Continuous Improvement

S21

Pre-Deployment Verification

Part 1 of 2

Standard Description

The decision to authorize use is based on evidence proportionate to the intended purpose and potential impact, clearly distinguishing established findings from aspects requiring further verification.

Requirements

1

The institution establishes approval criteria and evidence requirements before deployment, covering performance, safety, rights, and the beneficiaries, and clearly documents the limitations of testing. Approval of the agent is based on more than a single trial or demonstration.

2

Distinguish between benefits claimed by the provider and evidence obtained through independent testing, desk-based reviews, and field trials. Findings are not generalized to other versions or contexts without verifying their applicability.



Responsible Parties

The institution, the provider, and any party announcing the impact of using the agent.



Scope of Application

Institutional deployment or scaling, and claims regarding the effectiveness of agent use.

Continued on the following page: remaining requirements and verification evidence.

7

Dimension 7

Verification and Continuous Improvement

S21

Pre-Deployment Verification

Part 2 of 2

Requirements — continued

3

Claims of improved learning outcomes are supported by evidence of individual learning, such as knowledge retention and the ability to apply knowledge in relevant contexts. Service and administrative applications are measured according to their intended purposes.



Verification Evidence

Evidence supporting approval and intended-purpose criteria, distinguishing task-completion quality from actual learning outcomes, and user satisfaction from demonstrated effectiveness. In addition, when automated verification tools are used, their validity, accuracy, and limitations are documented.

7

Dimension 7

Verification and Continuous Improvement

S22

Performance Monitoring and Remediation

Standard Description

The institution monitors actual use and associated risks and addresses identified deficiencies before adverse impacts persist or use is expanded.

Requirements

1

A defined set of indicators to measure objective achievement, monitor damage and differences across groups, and enable practitioners, users, and beneficiaries to report issues.

2

Address incidents and complaints, identify their root causes and impacts, and modify usage restrictions or suspend use as warranted by the findings of the review.

3

Link continuity and expansion with sufficient evidence. Improvements in average performance or user satisfaction do not offset fundamental failures in safety or the protection of rights.



Responsible Parties

The institution and the provider, according to their respective responsibilities.



Scope of Application

During use and when considering scaling up.



Verification Evidence

Decisions that respond to results and complaints, with documented addressing of issues.

7

Dimension 7

Verification and Continuous Improvement

S23

Review Changes

Part 1 of 2

Standard Description

The suitability of use is reviewed when the agent's capabilities, authorities, components, or context are significantly changed.

Requirements

1

Specify the changes that require review, such as changing the model, provider, or a data source that affects the results, permissions, or beneficiaries. Monitor performance indicators to detect any unannounced change in the agent's behavior, including changes resulting from the agent's previous usage stored in its memory.

2

The provider discloses relevant changes and limitations in advance, and the institution determines whether existing evidence remains sufficient or further verification is required. The provider shall not implement material changes to the model on which the agent relies without providing advance notice and an opportunity to verify the changes before they take effect or offering the option to retain the previously approved version



Responsible Parties

The institution, the provider, and the developer.



Scope of Application

Change or renewal or termination of support, according to its actual effect.

Continued on the following page: remaining requirements and verification evidence.

7

Dimension 7

Verification and Continuous Improvement

S23

Review Changes

Part 2 of 2

Requirements — continued

3

Use shall be restricted where the effects of a material change or continued compliance with applicable requirements following that change cannot be verified.



Verification Evidence

A documented decision clarifying its justifications regarding continued use after change, without requiring a full re-examination for every routine update that does not change the agent's behavior in the targeted tasks.

Use Case Requirements

This section reviews requirements based on the type of tasks performed in the agent use cases identified in Section Three. The requirements of each field are applied according to the tasks performed by the agent; one use may combine more than one field. These requirements address the specific characteristics of each educational or service task, with reference to Section Four for requirements relating to data, security, rights, and individual protection.



Teaching and Supporting the Educational Practitioner



Learning, Practice, and Customization



Evaluation and Measurement



Guidance and Support



Content and Educational Resources



Research and Knowledge Management



Services and Educational Management



Governance and Decision Support

Requirements follow the tasks performed; a single use may span more than one area.

1

Area of Application 1

Teaching and Supporting the Educational Practitioner

Use Case Requirements

Standard Description

The agent supports the practitioner's work, while the practitioner retains responsibility for selecting and aligning learning experiences with learners' needs.

Requirements

1

The practitioner can understand the rationale behind the proposed plan or activity, review it, and adapt it before implementing it with learners.

2

Material changes to learning opportunities are supported by evidence, and the agent's assessment of a learner's needs is not used as the sole basis for such changes.

3

Workload reduction is measured by accounting for the time required for review and correction, as well as the quality of the resulting work. This effect is assessed separately from learning improvements.



Responsible Parties

The institution, the practitioner, and the provider.



Scope of Application

Planning, teaching, training, and monitoring learning needs.



Verification Evidence

Reviewable plans, clear justifications for practitioner choices, and measurement of time and quality to verify the stated benefit.

2

Area of Application 2

Learning, Practice, and Customization

Use Case Requirements

Standard Description

Agent-supported learning activities develop learners' capabilities without replacing the learning or practice they must undertake themselves to achieve the intended learning outcomes.

Requirements

1

The type and timing of assistance are tailored to the learner's needs, and they should be given the opportunity to try, clarify their answer, and receive feedback.

2

The learning pathway remains subject to review, and learners are not restricted to a particular level or denied opportunities based on unverified assumptions.

3

Define the indicators of learning and the level of assistance provided, and verify independent performance when it is a component of the competency. Furthermore, clarify the limitations of generalizing simulation results to real-world performance.



Responsible Parties

The institution, the practitioner, and the provider.



Scope of Application

Learning support, practice, simulation, and customization.



Verification Evidence

Performance evidence demonstrates what the learner has learned, specifies the level of assistance provided, and identifies the limitations of the results. The quality of work completed by the agent on the learner's behalf is not sufficient on its own.

3

Area of Application 3

Evaluation and Measurement

Use Case Requirements

Standard Description

The use of the agent preserves the validity, fairness, and accuracy of assessments of an individual's learning and competence.

Requirements

1

Define what is being assessed and which tasks the agent is permitted to perform. Rules governing the agent's use are communicated before the assessment, including cases where using the agent is part of the competency.

2

Verify the validity of the assessment, the evaluation criteria, the accuracy of scores and grades, and differences across groups. Suspected breaches of academic integrity are not determined solely on the basis of automated detection tools.

3

Provide evidence on the learner's performance, distinguishing the agent's contribution, and is not limited to the output or usage log solely. Practical competence required for the relevant skill or profession is demonstrated under appropriate conditions and in accordance with the requirements of the competent authority. Simulation alone is insufficient where actual performance is required.



Responsible Parties

The institution, the practitioner, the evaluation body, and the provider.



Scope of Application

Designing, implementing, and interpreting the results of the assessment. The substantive results are adopted and reviewed according to the relevant standards.



Verification Evidence

Tasks and assessment conditions appropriate for measuring the intended competency, in addition to evidence of the individual's performance that forms the basis for the assessment.

4

Area of Application 4

Guidance and Support

Use Case Requirements

Standard Description

The agent provides targeted support within clearly defined limits, preserves user choice, and enables users to return to a qualified specialist, such as an academic or professional advisor.

Requirements

1

Verify that information on programs, qualifications, and opportunities is up to date; ensure that alternatives and the limitations of recommendations are clearly communicated; and distinguish guidance from official qualification recognition or guarantees of employment.

2

Recommendations remain subject to review and must not restrict access to opportunities or constitute a medical or psychological diagnosis by an unqualified person.

3

Content adapted for accessibility preserves its intended meaning and the skills it is designed to develop. Any significant loss of information resulting from the adaptation is clearly disclosed.



Responsible Parties

The institution, the specialist, and the provider.



Scope of Application

Academic and career guidance, support, and accessibility.



Verification Evidence

Recommendations and content verified for accuracy and contextual suitability, with access to human assistance when needed.

5

Area of Application 5

Content and Educational Resources

Use Case Requirements

Standard Description

Content generated or modified by the agent is verified before approval and can be corrected wherever it is used.

Requirements

1

Review the content for accuracy and suitability to the curriculum, intended competencies, and target audience before approving it for its intended use.

2

Maintain clear records of the approved version, information sources, and changes, enabling the educational resource to be corrected or affected copies to be withdrawn.

3

Verify the sources cited by the agent and the applicable rights to use them. Merely listing the agent as a reference does not prove the source's existence, validity, or legitimacy of its use.



Responsible Parties

The institution, the practitioner or content producer, and the provider.



Scope of Application

Producing, updating, adopting, and publishing resources.



Verification Evidence

Resources that are identifiable, traceable, and available for review, with mechanisms to correct errors wherever they are published or used.

6

Area of Application 6

Research and Knowledge Management

Use Case Requirements

Standard Description

The use of the agent preserves the integrity, impartiality, and reviewability of the research methodology and findings.

Requirements

1

Verify the existence, validity, and currency of sources. Research data is distinguished from generated conclusions and summaries.

2

The agent's substantive contributions to data collection, analysis, and writing are disclosed, and methodological choices that may affect the results are documented, while accountability remains with the researcher or the institution.

3

Consider research ethics, participant rights, and confidentiality, and complete required approvals before contact, experimentation, or data collection.



Responsible Parties

The institution, the researcher, the knowledge management teams, and the provider.



Scope of Application

Research, production, and regulation of institutional knowledge.



Verification Evidence

Reviewable sources, methodology, and conclusions, with approvals obtained according to the type of research.

7

Area of Application 7

Services and Educational Management

Use Case Requirements

Standard Description

The agent performs services and operations in accordance with applicable rules and authorized permissions, while ensuring the integrity of records and protecting the rights of the parties involved.

Requirements

1

The agent distinguishes between providing information and preparing a request, on the one hand, and submitting or approving a request or modifying a record, on the other. Service delivery is governed by the applicable official rules and procedures.

2

Verify the outcome of each task in the relevant system and prevent duplicate submissions or unintended commitments upon error or retry.

3

Service value is measured based on accuracy, completion time, effort required from the service requester, and continuity of service. The impact of errors on records and access to opportunities is also reviewed.



Responsible Parties

The party, the provider, and the service employee.



Scope of Application

Registration, scheduling, correspondence, and operations related to learning and training.



Verification Evidence

Accurate, traceable, and correct procedures, with measures demonstrating the achieved service benefits.

8

Area of Application 8

Governance and Decision Support

Use Case Requirements

Standard Description

The agent supports institutional decision-making by providing clear evidence, traceable sources, and defined limitations while the decision and accountability remain with the authorized decision-maker.

Requirements

1

Document the sources, coverage, and quality of indicators. Observed facts are distinguished from expectations and assumptions, and correlation between results is distinguished from evidence of a causal relationship.

2

Examine the impact of the recommendation on different categories, institutions, and resources, and present alternatives rather than relying on a single aggregate indicator.

3

Any decision of compliance, imposition of a penalty, or procedure with official consequences must have a legal basis and be made by the authorized decision-maker following the required review. The agent's recommendation does not, in itself, constitute an approved decision.



Responsible Parties

The institution, the leadership, the decision-maker, and the provider.



Scope of Application

Institutional analysis, governance, and decision support.



Verification Evidence

A decision with an interpretable rationale and clearly defined evidentiary limitations, identifying the approving authority and the evidence on which the decision was based.

Conclusion



Value and accountability

These standards provide guidance on the use of AI agents in learning and training, supporting the achievement of educational and institutional value while preserving human accountability, protecting beneficiaries' rights, and promoting safe and reliable use.



Application and verification

Their application should be proportionate to the type and characteristics of each use case and its potential impact, with ongoing review of suitability, verification of outcomes, and remediation of any emerging risks or deficiencies.



Continuous improvement

These standards remain adaptable in light of evidence and practical experience gained through their implementation, keeping pace with the technological advancements and the changing needs of learning and training in Saudi Arabia.



المركز الوطني
للتعليم الإلكتروني
National eLearning Center

Standards for AI Agents in Learning and Training

This document is also available in Arabic.



Standards for AI Agents in Learning and Training © 2026 National eLearning Center.
Licensed under the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International License (CC BY-NC-SA 4.0).

NELC.gov.sa