

Simulazione d'Esame – Risposte

**Simulazione d'Esame – Set A
Versione 2.2**

ISTQB® AI Testing Syllabus

Compatibile con il Syllabus versione 2.0

International Software Testing Qualifications Board



Avviso di copyright

Avviso di copyright © International Software Testing Qualifications Board (di seguito denominato ISTQB®).

ISTQB® è un marchio registrato di International Software Testing Qualifications Board.

Tutti i diritti riservati.

Gli autori dichiarano con la presente di trasferire il copyright a ISTQB®. Gli autori (come attuali titolari del copyright) e ISTQB® (come futuro titolare del copyright) hanno concordato le seguenti condizioni di utilizzo:

Possono essere copiati estratti di questo documento per un uso non commerciale se la fonte è riconosciuta.

Qualsiasi Training Provider accreditato può utilizzare questa Simulazione d'Esame nei propri corsi di formazione se gli autori e ISTQB® sono riconosciuti come fonte e titolari del copyright della Simulazione d'Esame, e a condizione che qualsiasi pubblicità di tale corso di formazione sia effettuata solo dopo che l'Accreditamento ufficiale dei materiali di formazione sia stato ricevuto da un Member Board riconosciuto da ISTQB®.

Qualsiasi individuo o gruppo di individui può utilizzare questa Simulazione d'Esame in articoli e libri, a condizione che gli autori e ISTQB® siano riconosciuti come fonte e titolari del copyright della Simulazione d'Esame.

È proibito qualsiasi altro utilizzo di questa Simulazione d'Esame senza aver prima ottenuto l'approvazione scritta di ISTQB®.

Qualsiasi Member Board riconosciuto da ISTQB® può tradurre questa Simulazione d'Esame a condizione che riproduca il sopra menzionato Avviso di copyright nella versione tradotta della Simulazione d'Esame.

Responsabilità del documento

L'Examination Working Group ISTQB® è responsabile di questo documento.

Questo documento è mantenuto da un core team di ISTQB® composto dal Syllabus Working Group e dall'Exam Working Group.

Riconoscimenti

Questo documento è stato prodotto da un core team di ISTQB®: Klaudia Dussa-Zieger, Stuart Reid, Vipul Koch, Kyle Siemens, Qin Liu, Werner Henschelchen, Jarosław Hryszko

Il core team ringrazia il team di review dell'Exam Working Group, il Syllabus Working Group e i Member Board per i loro suggerimenti e contributi.

Cronologia delle revisioni

Modello di layout usato per Simulazione d'Esame – Versione Risposte:	Data: 27 novembre 2024 2.12
---	--------------------------------

Versione	Data	Osservazioni
1.0	2021/10/01	Rilascio per GA
2.0 Beta	2026/01/05	Revisione Beta
2.0	2026/04/17	Rilascio per GA
2.1	2026/05/20	Aggiornamento della domanda n. 10
2.2	2026/07/20	Sostituzione della domanda n. 19

Indice dei contenuti

Avviso di copyright	2
Responsabilità del documento	2
Riconoscimenti	2
Cronologia delle revisioni	3
Indice dei contenuti	4
Introduzione	6
Scopo del documento	6
Istruzioni	6
Tabella Riassuntiva delle Risposte	7
Risposte	8
1	8
2	8
3	9
4	10
5	10
6	11
7	12
8	13
9	13
10	14
11	15
12	16
13	16
14	17
15	18
16	18
17	18
18	19
19	20
20	20
21	21
22	21
23	22
24	22
25	23
26	23
27	24
28	24
29	25
30	25
31	26
32	26
33	27
34	28
35	29
36	29
37	29
38	30
39	31
40	32
Appendice: Risposte alle domande aggiuntive	33
Tabella Riassuntiva delle Risposte	34
A1	36



A2.....	38
A3.....	38
A4.....	39
A5.....	40
A6.....	41

Introduzione

Scopo del documento

Le domande di esempio, le risposte e le relative giustificazioni contenute in questa Simulazione d'Esame sono state create da un team di esperti in materia e da scrittori esperti di domande, con l'obiettivo di:

- Supportare i Member Board e gli Exam Board ISTQB® nelle loro attività di scrittura delle domande
- Fornire ai training provider e ai candidati all'esame esempi di domande d'esame

Queste domande non possono essere utilizzate così come sono in nessun esame ufficiale.

Nota, gli esami reali possono includere un'ampia varietà di domande e questa Simulazione d'Esame **non** intende includere esempi di tutti i possibili tipi, stili o lunghezze di domande; inoltre questa Simulazione d'Esame può essere più difficile oppure meno difficile di qualsiasi esame ufficiale.

Istruzioni

In questo documento potete trovare:

- La Tabella Riassuntiva delle Risposte, che include per ciascuna risposta corretta:
 - Livello K, Obiettivo di apprendimento e valore del punteggio
- Gli insiemi di risposte, che includono per tutte le domande:
 - Risposta corretta
 - Giustificazione per ciascuna opzione di risposta
 - Livello K, Obiettivo di apprendimento e valore del punteggio
- Insiemi di risposte aggiuntive, che includono per tutte le domande [non si applica a tutte le Simulazioni d'Esame]:
 - Risposta corretta
 - Giustificazione per ciascuna opzione di risposta
 - Livello K, Obiettivo di apprendimento e valore del punteggio
- *Le domande sono contenute in un documento separato*

Tabella Riassuntiva delle Risposte

Numero domanda (#)	Risposta corretta	LO	Livello K	Punti
1	a	AI-1.1.1	K2	1
2	c, e	AI-1.1.2	K2	1
3	a	AI-1.1.3	K2	1
4	b	AI-1.1.4	K2	1
5	a	AI-1.1.5	K2	1
6	b	AI-1.1.6	K2	1
7	c	AI-2.1.1	K2	1
8	b	AI-2.1.2	K2	1
9	c	AI-2.2.1	K2	1
10	c	AI-3.1.1	K2	1
11	d	AI-3.1.2	K2	1
12	d	AI-3.1.4	K2	1
13	c	AI-3.2.1	K2	1
14	b	AI-3.2.3	K2	1
15	c	AI-3.3.1	K3	2
16	a	AI-3.4.1	K2	1
17	a	AI-4.1.1	K2	1
18	b	AI-4.1.2	K2	1
19	b	AI-4.1.3	K2	1
20	b	AI-4.2.1	K2	1

Numero domanda (#)	Risposta corretta	LO	Livello K	Punti
21	d	AI-4.2.2	K3	2
22	b	AI-4.3.1	K2	1
23	a	AI-4.3.2	K2	1
24	a	AI-5.1.1	K2	1
25	c	AI-5.1.2	K2	1
26	d	AI-5.1.3	K2	1
27	b	AI-5.1.4	K2	1
28	c	AI-5.1.5	K3	2
29	b	AI-5.1.6	K2	1
30	b	AI-6.1.1	K2	1
31	a	AI-6.1.2	K2	1
32	a	AI-6.1.3	K2	1
33	c	AI-6.1.4	K2	1
34	b	AI-6.1.5	K3	2
35	b	AI-6.1.7	K2	1
36	c	AI-6.1.8	K2	1
37	c	AI-6.1.9	K2	1
38	d	AI-6.1.10	K2	1
39	b, d	AI-7.1.1	K2	1
40	b	AI-7.1.2	K2	1

Risposte

Numero domanda (#)	Risposta corretta	Spiegazione / Motivazione	Obiettivo di apprendimento (LO)	Livello K	Numero di punti
1	a	a) È corretta. Questa opzione differenzia correttamente i sistemi basati su AI dai sistemi convenzionali. I sistemi basati su AI/ML apprendono i pattern presenti nei dati, e molti possono adattarsi tramite autoapprendimento, mentre i sistemi convenzionali utilizzano una logica predefinita e producono output prevedibili. Nota: questo non si applica ai sistemi esperti rule-based. b) Non è corretta. Sebbene i sistemi basati su AI possano talvolta elaborare i dati più velocemente, ciò non è sempre vero, e il loro vantaggio risiede più nell'adattabilità e nella gestione della complessità che nella velocità. c) Non è corretta. Questa opzione inverte le caratteristiche dei due tipi di sistemi. I sistemi basati su AI sono probabilistici e meno spiegabili, mentre i sistemi convenzionali sono deterministici e più semplici da interpretare. d) Non è corretta. I sistemi convenzionali sono più adatti per attività critiche in cui è importante poter spiegare i risultati; i sistemi basati su AI in genere hanno scarsa trasparenza e spiegabilità, e sono quindi più appropriati per l'uso in ambiti non regolamentati. Inoltre, i sistemi basati su AI sono spesso più adatti per problemi più complessi.	AI-1.1.1	K2	1
2	c, e	a) Non è corretta. La narrow AI non è autoapprendente in tutti i casi; è specifica per un'attività. La general AI non è limitata a problemi specialistici, e la super AI non è limitata ad attività predefinite. b) Non è corretta. La narrow AI non opera in modo indipendente dall'intervento umano, la general AI non è esclusiva della robotica, e	AI-1.1.2	K2	1

		la super AI non si limita a migliorare il processo decisionale umano, ma ipoteticamente supera del tutto l'intelligenza umana. c) È corretta. Questa opzione differenzia accuratamente i tre tipi di AI, evidenziando la natura specifica per attività della narrow AI, la versatilità simile a quella umana della general AI, e la natura ipotetica e le capacità superiori della super AI. d) Non è corretta. Sebbene la narrow AI sia correttamente identificata come specifica per un'attività, la general AI, se raggiunta, avrebbe molte applicazioni pratiche. Inoltre, non è chiaro quale relazione avrebbe la super AI con i modelli di AI generativa. e) È corretta. Tutta l'AI odierna è narrow AI. Non è chiaro quando verrà raggiunta la general AI. Una volta raggiunta la general AI, data la disponibilità di conoscenza su Internet, è quasi certo che si arriverà anche alla super AI.			
3	a	a) È corretta. L'AI è il termine più ampio. Si riferisce al concetto generale di creare macchine in grado di svolgere attività che normalmente richiedono l'intelligenza umana. Questo include un'ampia gamma di approcci e tecniche, non solo il machine learning. Il ML è un sottoinsieme dell'AI. È un approccio specifico per realizzare l'AI, in cui alle macchine viene data la capacità di apprendere dai dati senza essere esplicitamente programmate. Invece di scrivere regole specifiche per ogni situazione, gli algoritmi ML apprendono pattern e formulano previsioni sulla base dei dati. Il DL è un sottoinsieme del machine learning. È un tipo specializzato di ML che utilizza reti neurali artificiali con più livelli (da cui il termine "profondo"). b) Non è corretta. L'AI e il ML esistevano come concetti e campi di studio molto prima che il deep learning diventasse rilevante. Il deep learning è un progresso relativamente recente all'interno dei più ampi campi del machine learning e dell'AI.	AI-1.1.3	K2	1

		c) Non è corretta. DL e ML non sono intercambiabili. Il DL è un tipo specifico di ML. Inoltre, l'AI non rappresenta un “approccio distinto”: ML e DL sono approcci per implementare l'AI. d) Non è corretta. Sebbene sia vero che l'AI comprende sia il ML sia il DL, descriverli come “che operano in parallelo” non è corretto. Il DL è un tipo di ML, non una metodologia parallela allo stesso livello in grado di operare in parallelo.			
4	b	a) Non è corretta. Descrive l'analisi e la comprensione, che non sono capacità della GenAI, e non si concentra sull'aspetto generativo dell'AI generativa. b) È corretta. Questa opzione descrive accuratamente l'AI generativa, evidenziando la sua capacità di creare nuovi contenuti sulla base di pattern appresi dai dati di addestramento. c) Non è corretta. L'AI generativa si concentra sulla creazione di nuovi contenuti, non sul miglioramento di contenuti esistenti come precursore di attività AI più tradizionali, come la classificazione e la previsione. d) Non è corretta. L'AI generativa non è limitata a testo e immagini; ha applicazioni in sanità, scoperta di farmaci, simulazione di dati e altri ambiti.	AI-1.1.4	K2	1
5	a	a) È corretta. Questo confronto evidenzia accuratamente i vantaggi delle GPU rispetto alle CPU per le attività di machine learning, grazie alle loro capacità di elaborazione parallela, mentre le CPU eccellono nel calcolo general-purpose. b) Non è corretta. Sebbene le CPU abbiano velocità di clock più elevate, sono meno efficienti delle GPU per le attività di ML perché mancano delle capacità di elaborazione massicciamente parallela richieste per attività di gestione dei dati come la moltiplicazione di matrici.	AI-1.1.5	K2	1

		c) Non è corretta. L'hardware specifico per l'AI come gli ASIC è più adatto all'edge computing piuttosto che all'addestramento dei modelli ML, che tipicamente avviene nel cloud. d) Non è corretta. I processori neuromorfici sono progettati per imitare la struttura del cervello e non si basano sull'architettura di von Neumann; pertanto questa affermazione non è accurata.			
6	b	Date le seguenti affermazioni sullo sviluppo e l'hosting di modelli AI: i. Riduce i costi di sviluppo utilizzando risorse di public cloud, eliminando la necessità di investimenti in hardware locale ⇒ Questo non è un approccio ibrido poiché riguarda solo lo sviluppo sul cloud. ii. Utilizza lo sviluppo locale dei componenti di preparazione dei dati per i dati sensibili, per una maggiore sicurezza, prima di passare al cloud per l'addestramento dell'intero sistema ⇒ Questo è <u>ibrido</u> perché abbiamo sviluppo locale per la preparazione dei dati e addestramento sul cloud. iii. Comporta costi più bassi perché per lo sviluppo locale vengono utilizzati laptop e i costi hardware iniziali sono contenuti ospitando i modelli AI su public cloud ⇒ Questo è <u>ibrido</u> perché lo sviluppo locale avviene su laptop e l'hosting su public cloud. iv. Semplifica lo sviluppo e l'hosting standardizzando i processi su server locali, eliminando la necessità di configurazioni complesse basate sul cloud ⇒ Questo non è ibrido poiché tutto avviene su server locali.	AI-1.1.6	K2	1

		v. Garantisce la massima sicurezza ospitando i modelli AI su private cloud, evitando così i rischi associati alle vulnerabilità dell'hardware locale ⇒ Questo non è ibrido poiché tutto avviene su private cloud. Quindi, l'opzione corretta è la b) ii e iii.			
7	c	Considerando ciascuno degli esempi in ordine: A. Il tasso di successo di un operatore remoto nel forzare un drone ad attivare il protocollo di atterraggio sicuro quando il suo sistema di navigazione AI mostra un comportamento pericoloso. Questa misura valuta l'efficacia dell'intervento umano sul funzionamento del sistema durante un comportamento AI pericoloso, in linea con la definizione di intervenibilità: la capacità di consentire un intervento esterno nei processi automatizzati. Questo corrisponde all'intervenibilità (4). B. Il tempo medio necessario per annullare con successo la decisione automatizzata di un sistema di gestione delle frodi di bloccare la transazione di un cliente. Questa misura riguarda la rapidità con cui un utente può annullare la decisione di un sistema automatizzato e il grado in cui un utente può controllare o influenzare le azioni del sistema. Questo è un indicatore diretto della controllabilità dell'utente (2). C. L'F1-score di un modello di object detection in un'auto a guida autonoma con pioggia intensa. L'F1-score in condizioni avverse (pioggia intensa) verifica quanto bene il modello si comporti in scenari impegnativi, il che è una misura della robustezza dell'AI (1). D. Il tempo necessario a un motore di raccomandazione di e-commerce per aggiornare i propri suggerimenti in base a una nuova tendenza di moda in rapida diffusione. La velocità con cui un sistema adatta le proprie raccomandazioni alle nuove tendenze	AI-2.1.1	K2	1

		riflette la sua capacità di adattare la propria funzionalità a requisiti mutevoli, il che corrisponde all'adattabilità funzionale (3). In tal caso, la combinazione corretta è: 1C – 2B – 3D – 4A, il che rende corretta l'opzione c).			
8	b	a) Non è corretta. Il problema reale è l'opposto: i requisiti sono generalmente troppo vaghi e spesso sono forniti solo implicitamente dai dati, il che porta a una scarsa tracciabilità. b) È corretta. Questa imprevedibilità è una sfida fondamentale in termini di sicurezza fisica quando si distribuisce l'AI in sistemi legati alla sicurezza fisica, poiché complica la verifica, la validazione e la garanzia continua di un funzionamento sicuro. c) Non è corretta. Il problema è che i sistemi self-learning non smettono di adattarsi dopo il deployment, e quindi si allontanano dal comportamento originariamente testato. d) Non è corretta. Gli standard maturi legati alla sicurezza fisica attualmente non prevedono alcuna disposizione per l'AI, e alcuni ne limitano esplicitamente l'uso.	AI-2.1.2	K2	1
9	c	a) Non è corretta. Questo riguarda l'intervenibilità, come definita nella ISO 25059, che descrive la capacità dell'utente di controllare il robot di pattugliamento autonomo quando nota che potrebbe urtare una scultura. Fornisce inoltre un intervallo di tempo specifico e misurabile per l'intervento. b) Non è corretta. Questo riguarda la correttezza funzionale, definita nella ISO 25059. Descrive come il sistema si adatta ai cambiamenti di umidità per mantenere l'ambiente desiderato. Fornisce un valore di umidità specifico e misurabile e un intervallo di tempo entro cui il sistema di controllo deve reagire. c) È corretta. È l'opzione MENO probabile perché si concentra su alcune delle sotto-caratteristiche dell'usabilità, come l'apprendibilità e l'operabilità, che sono caratteristiche di qualità generiche	AI-2.2.1	K2	1

		applicabili alla maggior parte dei sistemi, e non sono direttamente associate alle caratteristiche specifiche dell'AI nella ISO 25059 (ossia la controllabilità dell'utente e la trasparenza, che sono sotto-caratteristiche dell'usabilità per i sistemi basati su AI). Il criterio di accettazione è inoltre più soggettivo rispetto alle altre opzioni, e quindi più difficile da testare. d) Non è corretta. Questo riguarda la trasparenza, definita nella ISO 25059. Descrive la capacità del sistema di fornire informazioni aggiuntive per spiegare il proprio processo decisionale.			
10	c	Considerando le descrizioni dei sistemi di esempio: A. L'importo speso può essere considerato la funzione di ricompensa per questo sistema, che modifica il proprio comportamento per aumentare l'importo speso. Si tratta quindi di un esempio di Apprendimento per rinforzo (2). B. L'app utilizza testo in quella che può essere considerata una lingua sorgente e una corrispondente traduzione 'corretta' di questa sorgente, impiegando una forma di apprendimento supervisionato senza che venga menzionata alcuna funzione di ricompensa esplicita. La classificazione utilizza dati di addestramento paralleli — coppie di frasi corrispondenti in due lingue (ad es. “Hello” in inglese e “Bonjour” in francese) — in cui ciascuna coppia ha un'etichetta che collega il testo sorgente al testo di destinazione. Il modello apprende le corrispondenze tramite addestramento supervisionato, codificando le caratteristiche del testo di input e prevedendo l'output di traduzione corretto, minimizzando gli errori su molte coppie di questo tipo per generalizzare a nuove frasi. Si tratta quindi di un esempio di Classificazione (3).	AI-3.1.1	K2	1

		C. La regressione viene utilizzata in questo scenario per prevedere un esito continuo (in questo caso, il tempo prima del guasto di un'apparecchiatura) sulla base di un insieme di variabili di input (dati dei sensori e registrazioni storiche di manutenzione). Analizzando la relazione tra queste variabili, il modello ML può identificare pattern e tendenze che indicano quando un'apparecchiatura è a rischio di guasto. Si tratta quindi di un esempio di Regressione (4). D. Analizzando le interazioni degli utenti (ad es. like, commenti, contenuti condivisi) e gli interessi dichiarati, la piattaforma di social network può identificare pattern e somiglianze tra gli utenti. Questi pattern possono poi essere utilizzati per raggruppare gli utenti in community o “cluster” che condividono interessi e comportamenti comuni. Si tratta quindi di un esempio di Clustering (1). Quindi, 1D – 2A – 3B – 4C è corretto, e pertanto l'opzione c) è quella corretta.			
11	d	1. Le prestazioni del modello vengono testate utilizzando i dati di validazione. Questo fa parte dell'attività 'Valutazione del modello'. (D) 2. Viene identificata l'origine dei dati di test utilizzati per testare il modello. Questo fa parte dell'attività 'Preparazione e test dei dati'. (B) 3. I dati di test vengono utilizzati per determinare che i criteri di prestazione concordati siano soddisfatti. Questo fa parte dell'attività 'Test del modello'. (C)	AI-3.1.2	K2	1

		4. Il modello viene testato sulla piattaforma di destinazione. Questo fa parte dell'attività 'Distribuzione del modello'. (A) Quindi, l'opzione corretta è: 1D – 2B – 3C – 4A, e pertanto d) è corretta.			
12	d	a) Non è corretta. L'approccio RAG non aggiunge nuovi livelli alla rete neurale per memorizzare la documentazione; recupera invece dati esterni rilevanti al momento dell'inferenza e integra l'input del modello con queste informazioni. b) Non è corretta. Sebbene il fine-tuning con dati di alta qualità possa ridurre il bias, non garantisce la prevenzione di output ingiusti, poiché il bias può persistere o addirittura essere introdotto durante il processo di fine-tuning. c) Non è corretta. Il fine-tuning può essere eseguito su tutti i livelli oppure solo su un sottoinsieme. Spesso vengono aggiornati solo alcuni livelli, mentre altri restano bloccati (frozen), a seconda dell'attività e dei dati disponibili. d) È corretta. Il RAG si basa sulla cura e sull'indicizzazione di dati esterni rilevanti prima dell'uso, ma non richiede di modificare l'architettura o i parametri del modello pre-addestrato stesso; integra invece gli input del modello con le informazioni recuperate a runtime.	AI-3.1.4	K2	1
13	c	a) Non è corretta. Questa affermazione evidenzia un passaggio fondamentale nella preparazione dei dati: raccogliere e consolidare i dati da fonti diverse prima che inizi qualsiasi elaborazione o analisi. Tuttavia, la preparazione dei dati in genere comprende anche la pulizia, la trasformazione e l'organizzazione dei dati, oltre alla raccolta dei dati grezzi. b) Non è corretta. Questa affermazione non è accurata perché il feature engineering viene generalmente eseguito prima (o durante) l'addestramento del modello, non dopo. Lo scopo del feature	AI-3.2.1	K2	1

		engineering è creare o modificare le feature per migliorare le prestazioni del modello ML durante la fase di addestramento. c) È corretta. Questa opzione identifica correttamente due tecniche comuni di pre-processing dei dati: l'augmentation (aumentare i dati creando versioni modificate) e il sampling (ridurre i dati selezionando sottoinsiemi). d) Non è corretta. Questa affermazione coglie solo in parte lo scopo dell'EDA, che è comprendere le caratteristiche dei dati e individuare problemi tramite visualizzazione e statistiche riassuntive. Tuttavia, l'EDA non è “testing esplorativo”, bensì un processo analitico che orienta le successive fasi di preparazione dei dati.			
14	b	a) Non è corretta. Il dataset di addestramento viene utilizzato principalmente per adattare i parametri del modello, non per ottimizzare gli iperparametri. Il dataset di validazione viene tipicamente utilizzato per calibrare gli iperparametri, e il dataset di test non viene mai utilizzato per generare dati di addestramento. b) È corretta. Il dataset di addestramento viene utilizzato per adattare o creare il modello, il dataset di validazione viene utilizzato per calibrare gli iperparametri del modello e prevenire l'overfitting, e il dataset di test è riservato alla valutazione delle prestazioni del modello su dati mai visti prima. Questa opzione descrive accuratamente i ruoli distinti e sequenziali di ciascun dataset nel flusso di lavoro ML. c) Non è corretta. Il dataset di addestramento non viene utilizzato per la valutazione finale del modello: questo è il ruolo del dataset di test. Inoltre, il dataset di test non viene utilizzato per calibrare gli iperparametri, poiché ciò comporterebbe il rischio di overfitting sul dataset di test. d) Non è corretta. Il ruolo primario del dataset di addestramento non è garantire la generalizzazione, bensì adattare il modello. Il dataset di validazione non viene utilizzato per il deployment, e il dataset di test	AI-3.2.3	K2	1

		non viene utilizzato per una valutazione iniziale, bensì per la valutazione finale dopo lo sviluppo del modello			
15	c	a) Non è corretta. Vedere l'opzione c) per la formula e il calcolo corretti. b) Non è corretta. Vedere l'opzione c) per la formula e il calcolo corretti. c) È corretta. La formula per la precisione = $TP / (TP + FP) * 100 = 78 / (78 + 22) = 78 / 100 * 100$ d) Non è corretta. Vedere l'opzione c) per la formula e il calcolo corretti.	AI-3.3.1	K3	2
16	a	a) È corretta. Questo è il passaggio successivo nel ciclo di addestramento ML. La loss viene ripropagata nella rete per modificare i valori dei pesi e dei bias. b) Non è corretta. L'addestramento procede con il batch di dati successivo o completa l'epoca. Non rielabora gli stessi dati solo per confermare il valore della loss. c) Non è corretta. Le funzioni di attivazione vengono scelte durante la progettazione del modello e di norma non vengono modificate durante un'esecuzione di addestramento. d) Non è corretta. L'azzeramento dei pesi avviene all'inizio dell'addestramento (inizializzazione), non dopo il calcolo del valore della loss per un batch di dati.	AI-3.4.1	K2	1
17	a	Considerando i sistemi basati su AI forniti come esempio: i. Un sistema che apprende dai dati in tempo reale per migliorare le proprie previsioni di guasto e aggiorna automaticamente i piani di manutenzione. Si tratta di una forma di sistema basato su AI adattivo, che modifica i piani sulla base dei dati in tempo reale.	AI-4.1.1	K2	1

		ii. Un filtro antispam in un'app di posta elettronica, che identifica lo spam sulla base di regole predefinite. Si tratta di una forma di sistema basato su AI di tipo locked, che genera risultati di test prevedibili, poiché si basa sul seguire regole predefinite di identificazione dello spam che non cambieranno finché il modello rule-based non verrà ricostruito. iii. Un motore di raccomandazione su un servizio di streaming che aggiorna i propri suggerimenti in base ai cambiamenti nelle abitudini di visione e nelle preferenze dell'utente. Si tratta di una forma di sistema adattivo, che si adatta sulla base delle abitudini di visione e delle preferenze dell'utente. iv. Un assistente personale che apprende dal proprio utente. Si tratta di un sistema basato su AI adattivo, poiché adatta i propri parametri sulla base delle interazioni degli utenti con il sistema. v. Un sistema rule-based per la diagnosi medica. Si tratta di una forma di sistema basato su AI di tipo locked, che non modificherà le regole del dominio finché il modello rule-based non verrà ricostruito. Pertanto, il filtro antispam e il sistema di diagnosi medica sarebbero più semplici da testare, poiché sono entrambi forme di sistemi locked e generano risultati di test prevedibili. E, quindi, l'opzione a) è corretta			
18	b	a) Non è corretta. Sebbene i sistemi basati su AI possano essere grandi e complessi, la ragione per utilizzare un approccio statistico deriva principalmente dalla loro natura non deterministica, non dall'impraticabilità dell'automazione o dalle dimensioni. b) È corretta. I sistemi basati su AI sono non deterministici e richiedono dataset di test ampi e rappresentativi per raggiungere una significatività statistica, in particolare nella valutazione statistica della correttezza funzionale.	AI-4.1.2	K2	1

		c) Non è corretta. Questa opzione travisa il concetto. Un singolo test case non è sufficiente per valutare la qualità del sistema basato su AI. È necessaria una valutazione statistica su molte previsioni. d) Non è corretta. Questo contraddice le best practice. Un dataset di test separato (da addestramento e validazione) è essenziale, non opzionale, ed è parte del rigore statistico richiesto.			
19	b	a) Non è corretta. Lo sviluppo esplorativo dei sistemi AI può spesso comportare che i requisiti di sistema cambino durante lo sviluppo; tuttavia, i sistemi AI self-learning non devono necessariamente essere sviluppati utilizzando un approccio di sviluppo esplorativo. b) È corretta. I sistemi basati su AI self-learning aggiornano continuamente i propri modelli interni dopo il deployment. Questo fa sì che il comportamento 'corretto' del sistema basato su AI cambi man mano che il sistema evolve, e questo comportamento non si riflette in specifiche aggiornate. c) Non è corretta. Ciascun sistema AI self-learning può apprendere in modi diversi, poiché ognuno incontrerà esperienze operative diverse; tuttavia, ciò normalmente significa che avranno comportamenti attesi leggermente diversi, non che determinare tali comportamenti attesi richieda un giudizio soggettivo. d) Non è corretta. Alcuni sistemi AI self-learning possono essere complessi, ma questo non è il loro attributo distintivo: alcuni possono essere relativamente semplici in termini di comportamento appreso. Pertanto, non esiste una ragione sistemica per cui i loro output attesi non possano essere determinati da tester umani.	AI-4.1.3	K2	1
20	b	a) Non è corretta. I modelli di AI generativa non producono output deterministici, quindi la corrispondenza esatta con risultati attesi predefiniti non è un approccio adatto per questo testing. b) È corretta. A causa della variabilità e complessità degli output della GenAI, il testing si concentra sulla coerenza, sulla conformità alle	AI-4.2.1	K2	1

		regole e sulla plausibilità, piuttosto che sulla corrispondenza con output attesi fissi. Input diversi, prompt opzionali e parametri influenzano tutti i risultati dei test. c) Non è corretta. Sebbene i modelli di AI generativa siano probabilistici, il testing (ad es. testing esplorativo, metamorphic testing e adversarial testing) può comunque essere utilizzato. d) Non è corretta. La review manuale è possibile, ma anche metodi automatizzati, inclusi altri strumenti GenAI o sistemi di riconoscimento delle immagini, possono valutare efficacemente i contenuti generati. Pertanto, affermare che il testing automatizzato “non si applica” non è accurato.			
21	d	a) Non è corretta. Il red teaming viene tipicamente eseguito prima del deployment. b) Non è corretta. Nella domanda non viene fornita alcuna indicazione sul fatto che la sicurezza sia più o meno importante del testing per bias. c) Non è corretta. Il red teaming viene eseguito in modo proattivo prima del deployment. d) È corretta. Il red teaming dovrebbe essere condotto prima del deployment, non dopo, e più tipi di vulnerabilità (sicurezza e bias) dovrebbero essere testati insieme.	AI-4.2.2	K3	2
22	b	a) Non è corretta. Sebbene il testing di sistema avrebbe potuto rilevare questo difetto, il testing dei dati di input è il livello di test più efficace per rilevarlo, poiché sarebbe stato individuato in una fase precedente. b) È corretta. Il testing dei dati di input si concentra sulla qualità e sulla rappresentatività dei dati. Sembra che le aree del Regno Unito con un'altitudine superiore a 1.250 metri non fossero adeguatamente rappresentate nei dati di addestramento.	AI-4.3.1	K2	1

		c) Non è corretta. Il testing di integrazione dei componenti verifica la presenza di difetti nelle interfacce e nelle interazioni tra i componenti, e non si concentra su valori di dati specifici. d) Non è corretta. Il testing del modello ML avrebbe potuto identificare questo difetto, ma il testing dei dati di input sarebbe il livello di test più efficace per rilevarlo, poiché sarebbe stato individuato prima.			
23	a	a) È corretta. La sicurezza e l'usabilità sono caratteristiche di qualità generiche e possono applicarsi a qualsiasi sistema. Il bias dei dati e le prestazioni del modello sono rischi potenziali specifici dei sistemi ML. b) Non è corretta. Sia il bias dei dati sia il bias algoritmico sono rischi associati esplicitamente a un sistema ML e, pertanto, non si applicano ai sistemi convenzionali. c) Non è corretta. Il risk management sia per i sistemi non-AI sia per i sistemi AI (compresi i sistemi ML self-learning) dovrebbe essere di natura dinamica. d) Non è corretta. La correttezza funzionale può essere il fattore di rischio primario sia per i sistemi convenzionali sia per i sistemi ML, ma questo dipende dal contesto.	AI-4.3.2	K2	1
24	a	a) È corretta. Lo scenario descrive la necessità di verificare l'origine dei dati e di determinare che non siano stati manomessi. Questo corrisponde al testing di provenienza dei dati. b) Non è corretta. Sebbene il team probabilmente eseguirebbe il testing di rappresentatività dei dati su un nuovo dataset, questa attività verifica se le caratteristiche dei dati corrispondono a quelle del mondo reale, non se la loro origine documentata sia accurata, o se si tratti di dati grezzi validi. c) Non è corretta. Il feature testing valuta il potere predittivo delle feature, il che è una questione diversa dalla verifica della fonte e dell'integrità del dataset stesso.	AI-5.1.1	K2	1

		d) Non è corretta. Il dataset constraint testing verifica la coerenza interna del dataset (ad es. intervalli, tipi), ma non può verificare da dove provengano i dati o se siano stati alterati illegittimamente.			
25	c	a) Non è corretta. Questo approccio si concentra sulla logica del modello, non sulle caratteristiche dei dati di addestramento. Non individua direttamente i pattern di bias presenti nei dati. b) Non è corretta. Questo approccio esamina gli output del modello anziché i dati di addestramento stessi. È utile per identificare il bias nelle previsioni, ma non per la rilevazione iniziale del bias nei dati. c) È corretta. Questo approccio può aiutare a evidenziare dove potrebbe essere introdotto il bias, ma si basa su controlli di processo piuttosto che su prove effettive tratte dai dati. d) Non è corretta. Questo approccio valuta gli esiti o le previsioni in presenza di cambiamenti ipotetici, mirando al bias negli esiti piuttosto che valutare direttamente il dataset di addestramento.	AI-5.1.2	K2	1
26	d	a) Non è corretta. Entrambi i tipi di pipeline trarrebbero beneficio da un approccio completo su più livelli. Una pipeline operativa richiederebbe un ampio testing di sistema end-to-end, non solo la validazione dei singoli script. b) Non è corretta. Entrambi i tipi di pipeline utilizzerebbero una gamma di test. Limitare l'una al solo approccio di test a livello di sistema e l'altra al solo approccio di test a livello di componente e integrazione non segue un approccio su più livelli. c) Non è corretta. La configuration management verifica che vengano utilizzate le versioni corrette “in fase di addestramento, test e produzione”, il che implica che sia importante per entrambi i tipi di pipeline garantire coerenza e riproducibilità, non solo per quelle operative. d) È corretta. Lo scopo della pipeline determina la strategia di test. Le pipeline di addestramento danno priorità alla generazione di dati di	AI-5.1.3	K2	1

		alta qualità, mentre le pipeline operative live si concentrano su aspetti non funzionali come l'efficienza prestazionale, la scalabilità e la robustezza dell'AI.			
27	b	a) Non è corretta. Un dataset di riferimento non deve necessariamente essere universalmente applicabile o derivare esclusivamente da benchmark di settore. Tipicamente viene calibrato per corrispondere alle caratteristiche operative e contestuali della popolazione target del sistema AI. b) È corretta. Lo scopo principale del dataset di riferimento è fornire una baseline statistica, che consenta ai professionisti di confrontare oggettivamente le distribuzioni, le correlazioni tra feature e la copertura dei dati di addestramento rispetto a quanto atteso in ambienti operativi reali. c) Non è corretta. Il campionamento stratificato viene tipicamente applicato al dataset di riferimento per creare una baseline completa, piuttosto che direttamente ai dati di addestramento stessi. Il dataset di riferimento mira a garantire che tutti i sottogruppi rilevanti siano rappresentati, fungendo da standard per valutare la rappresentatività degli altri dataset. d) Non è corretta. Il dataset di riferimento non è pensato come fonte principale per il testing finale del modello ML o per il testing con elevata indipendenza. Esiste per valutare la rappresentatività dei dati prima dell'addestramento del modello.	AI-5.1.4	K2	1
28	c	a) Non è corretta. Un vincolo di intervallo valida solo che i valori di un singolo attributo rientrino in limiti specificati, ma non verifica la relazione matematica tra <code>monthly_payment</code> e <code>loan_amount</code>. b) Non è corretta. Un vincolo di conteggio verifica la quantità di valori non nulli, non la correttezza matematica dei campi calcolati. c) È corretta. Questo scenario richiede di validare una relazione matematica tra due attributi (<code>monthly_payment</code> e <code>loan_amount</code>). Il	AI-5.1.5	K3	2

		vincolo di correlazione è specificamente progettato per verificare che i valori di un attributo siano correlati ai valori di un altro attributo secondo una regola definita: in questo caso, che monthly_payment sia uguale a loan_amount/324. d) Non è corretta. Un vincolo di duplicati identifica valori identici, ma avere rate mensili duplicate potrebbe essere legittimo (richiedenti diversi potrebbero avere lo stesso importo di rata).			
29	b	a) Non è corretta. Le annotazioni multiple normalmente non sono automatizzate. Le annotazioni multiple si basano tipicamente su annotatori umani e non garantiscono la coerenza; al contrario, ne evidenziano l'incoerenza. b) È corretta. Le annotazioni multiple prevedono che i data point vengano etichettati in modo indipendente da più annotatori. Il confronto tra queste etichette porta a individuare disaccordi tra gli annotatori, evidenziando potenziali difetti. c) Non è corretta. Il confronto delle distribuzioni delle etichette fa parte dell'analisi della distribuzione dei dati, non delle annotazioni multiple. d) Non è corretta. Segnalare i data point con un'elevata loss è il compito dell'analisi della loss del modello, non delle annotazioni multiple.	AI-5.1.7	K2	1
30	b	Considerando ciascuno dei rischi in ordine: A. Il modello ML potrebbe comportarsi in modo diverso per diversi gruppi demografici. Questo corrisponde meglio al Testing per bias, che esamina esplicitamente l'equità tra popolazioni diverse. (2) B. Input leggermente modificati al modello ML potrebbero causare risposte piuttosto diverse e inattese. Questo si abbina correttamente all'Adversarial testing, che testa deliberatamente la capacità di un modello ML di gestire piccole perturbazioni dell'input. (3)	AI-6.1.1	K2	1

		C. Le previsioni effettuate dal modello ML potrebbero risultare inaccurate in alcuni casi. Questo viene affrontato al meglio dal Testing delle prestazioni funzionali ML, che valuta direttamente se il modello ML produce output accurati. (1) D. L'accuratezza del modello ML potrebbe essere diminuita significativamente da quando è stato distribuito. Questo corrisponde adeguatamente al Drift testing, che monitora i cambiamenti delle prestazioni nel tempo in produzione. (4) Pertanto, la corrispondenza corretta tra rischi e approcci di test è: 1C – 2A – 3B – 4D, e quindi b) è l'opzione corretta.			
31	a	a) È corretta. Questo fa parte del requisito/documentazione non funzionale. b) Non è corretta. Non rientra nell'ambito della documentazione del modello. c) Non è corretta. La documentazione del modello dovrebbe contenere l'origine dei dati di addestramento. d) Non è corretta. La documentazione delle modifiche apportate dal sistema self-learning potrebbe essere prodotta solo dal sistema stesso al momento dell'aggiornamento.	AI-6.1.2	K2	1
32	a	a) È corretta. Quando vengono eseguiti più test a parità di accuratezza misurata (83%), il margine di errore tipicamente diminuisce se il livello di confidenza rimane invariato. Questo perché una dimensione campionaria maggiore può ridurre il margine di errore e/o aumentare la nostra confidenza statistica nel valore misurato. Se il modello ML aveva precedentemente raggiunto un'accuratezza dell'83% con un margine di errore di $\pm 4\%$ a un livello di confidenza del 94%, e sono stati eseguiti test aggiuntivi mantenendo la stessa misurazione di accuratezza dell'83% e mantenendo il livello di confidenza al 94%, questo comporterebbe un margine di errore inferiore (ad es. $\pm 2\%$). Ciò	AI-6.1.3	K2	1

		indica che possiamo essere più certi che l'accuratezza reale sia effettivamente intorno all'83%. b) Non è corretta. Quando vengono eseguiti più test, il margine di errore tipicamente diminuisce e il livello di confidenza tipicamente aumenta anziché diminuire, supponendo che l'accuratezza misurata rimanga la stessa. Passare da un livello di confidenza del 94% al 92% indicherebbe che sono stati eseguiti meno test o che i risultati dei test erano più variabili, il che contraddice lo scenario in cui sono stati eseguiti test aggiuntivi. c) Non è corretta. Se fossero stati eseguiti test aggiuntivi e l'accuratezza fosse rimasta all'83%, il margine di errore tipicamente diminuirebbe e non aumenterebbe, se il livello di confidenza fosse mantenuto al 94%. Il margine di errore aumentato di $\pm 6\%$ rende questa opzione meno probabile rispetto all'opzione in cui diminuisce (a). d) Non è corretta. Lo scenario afferma che l'accuratezza del modello ML per i test aggiuntivi è stata misurata all'83%, non all'85%. Un aumento della percentuale di accuratezza indicherebbe che il modello ML ha avuto prestazioni migliori nei test aggiuntivi, il che contraddice l'informazione fornita secondo cui il modello ML “ha fornito previsioni corrette meno volte del previsto”, pur mantenendo la stessa accuratezza misurata dell'83%.			
33	c	a) Non è corretta. Implica erroneamente che la conoscenza interna del modello ML sia irrilevante, mentre tale conoscenza può migliorare il processo di testing consentendo di creare adversarial example più mirati. b) Non è corretta. Suggerisce erroneamente che i metodi manuali siano l'unico approccio, ignorando le tecniche automatizzate, che sono anch'esse fondamentali nell'adversarial testing.	AI-6.1.4	K2	1

		c) È corretta. Nell'adversarial testing, gli adversarial example vengono spesso creati per identificare le vulnerabilità perturbando input funzionanti. d) Non è corretta. Non coglie il senso dell'adversarial testing, che utilizza deliberatamente valori di input aggiornati (cioè test) per creare adversarial example.			
34	b	Il test case T1 differisce dal test case sorgente per il cambiamento nei requisiti relativi alla fotocamera 3D; ora è più specifico. Deve essere inclusa una fotocamera 3D. Ciò significa quindi che i risultati attesi di follow-up possono includere al massimo i risultati originali del test (i telefoni precedentemente consigliati, ma con fotocamera 3D). Anche il test case T2 differisce dal test case sorgente per il cambiamento nei requisiti relativi alla fotocamera 3D; anch'esso è più specifico. Non deve essere inclusa alcuna fotocamera 3D. Ciò significa quindi che i risultati attesi di follow-up possono includere al massimo i risultati originali del test (i telefoni precedentemente consigliati, ma senza fotocamera 3D). Poiché T1 elenca i telefoni <u>con</u> fotocamera 3D, i telefoni rimanenti del test case sorgente devono essere quelli senza fotocamera 3D, e quindi devono trovarsi in T2. Pertanto, T1 e T2 combinati dovrebbero contenere tutte le fotocamere del test case sorgente, ma senza alcuna sovrapposizione tra i due. a) Non è corretta. La fotocamera SnapHappy M3 manca negli output combinati di T1 e T2 b) È corretta. Non c'è sovrapposizione tra gli output di T1 e T2, e nessuna fotocamera manca. c) Non è corretta. Le fotocamere SnapHappy sono elencate per entrambi i test case.	AI-6.1.5	K3	2

		d) Non è corretta. L'output dei due test case è identico, nonostante i requisiti differenti relativi alla fotocamera 3D.			
35	b	a) Non è corretta. Sebbene il dynamic testing sia appropriato quando la ground truth è disponibile, non analizza le proprietà statistiche dei dati di input. b) È corretta. Poiché la ground truth non è disponibile, il drift testing statico, che non la richiede, è l'unica opzione praticabile. c) Non è corretta. L'organizzazione “non dispone di un meccanismo per tracciare se un cliente se ne va effettivamente”. Pertanto, non dispone di risultati effettivi con cui confrontarsi, il che rende impossibile eseguire il dynamic testing. d) Non è corretta. Lo static testing esamina le distribuzioni dei dati, mentre confrontare le metriche di prestazione con la ground truth è il meccanismo del dynamic testing.	AI-6.1.7	K2	1
36	c	a) Non è corretta. Il modello ML ha buone prestazioni sui dati di validazione, quindi è improbabile che si tratti di un caso di underfitting. b) Non è corretta. La deriva concettuale si riferisce a cambiamenti che si verificano dopo l'addestramento del modello ML e le fasi di validazione. c) È corretta. Le prestazioni scarse sui dati di test e le buone prestazioni sui dati di validazione suggeriscono overfitting. d) Non è corretta. I criteri di accettazione dovrebbero essere coerenti tra i diversi insiemi di dati, quindi è improbabile che criteri di accettazione bassi portino a una differenza tra i risultati dei test con dati di validazione e dati di test indipendenti.	AI-6.1.8	K2	1
37	c	a) Non è corretta. Secondo il syllabus, “il testing A/B non genera test case e non fornisce alcuna indicazione su come i test debbano essere progettati, sebbene negli test vengano spesso utilizzati input	AI-6.1.9	K2	1

		operativi". Il testing A/B confronta le risposte di due varianti piuttosto che generare test case. b) Non è corretta. Il testing A/B non si concentra sulle interazioni tra componenti all'interno di un sistema. Confronta due versioni diverse di un sistema, non componenti all'interno di un singolo sistema. c) È corretta. Come indicato nel syllabus, "ogni volta che il sistema viene aggiornato, il testing A/B viene utilizzato per verificare che la variante aggiornata funzioni altrettanto bene, o meglio, della versione precedente". Questo descrive accuratamente come viene utilizzato il testing A/B nel contesto dei sistemi ML. d) Non è corretta. Il testing A/B non analizza la struttura interna dell'algoritmo. È un "approccio di test statistico che tipicamente richiede il confronto dei risultati di test provenienti da più esecuzioni per determinare la differenza tra i programmi". Si concentra sul confronto degli output di due varianti, non sull'esame delle loro strutture interne.			
38	d	a) Non è corretta. Variare gli iperparametri è una modifica minore allo stesso modello ML. Non lo cambia in modo sostanziale. Uno pseudo-oracolo dovrebbe idealmente essere significativamente diverso per rilevare i difetti in modo efficace. Questa opzione è troppo simile al SUT. b) Non è corretta. Il fine-tuning è una modifica dello stesso modello ML, che tipicamente comporta un ulteriore addestramento su un dataset specifico. Non è un cambiamento sufficientemente sostanziale da creare uno pseudo-oracolo davvero indipendente. Si basa comunque sullo stesso modello e framework di base, aumentando il rischio di difetti condivisi. c) Non è corretta. La retrieval-augmented generation (RAG) potenzia un modello linguistico integrandolo con un meccanismo di recupero. Sebbene il RAG aggiunga complessità, si basa comunque sul modello ML di base in fase di test. L'architettura fondamentale e il	AI-6.1.10	K2	1

		processo di addestramento del modello di base rimarranno gli stessi. Sebbene il RAG introduca una differenza, si tratta comunque di un'estensione del modello originale piuttosto che di un'alternativa indipendente. d) È corretta. Utilizzare un framework di sviluppo ML diverso significa utilizzare librerie diverse, potenzialmente algoritmi sottostanti diversi o implementazioni diverse degli algoritmi, e un ambiente di sviluppo ML diverso. Questa opzione massimizza l'indipendenza dello pseudo-oracolo dal SUT, riducendo il rischio di difetti condivisi e aumentando la potenziale efficacia del testing back-to-back.			
39	b, d	a) Non è corretta. Questo descrive un rischio di scarsa efficienza prestazionale. La mitigazione corretta è il performance (efficiency) testing, distinto dal testing delle prestazioni funzionali ML (che valuta la correttezza funzionale). b) È corretta. Questo descrive un potenziale difetto di una libreria utilizzata. Il testing delle prestazioni funzionali ML può valutare il comportamento del modello ed evidenziare anomalie derivanti dall'uso di una nuova libreria difettosa che causa un cambiamento nei risultati dei test. c) Non è corretta. Questo descrive la necessità di mitigare il rischio di un'installazione difettosa del framework ML. La mitigazione appropriata per questo caso è lo smoke testing, non il testing completo delle prestazioni funzionali ML. d) È corretta. Questo descrive il rischio di una scarsa interpretazione dei risultati dei test dovuta alla natura stocastica del processo di apprendimento. Il testing delle prestazioni funzionali ML, utilizzando più esecuzioni e l'analisi statistica, è la mitigazione per questo rischio. e) Non è corretta. Questo descrive una selezione non ottimale dell'algoritmo. La mitigazione appropriata è una review di idoneità	AI-7.1.1	K2	1

		dell'algoritmo o il testing A/B, non il testing delle prestazioni funzionali ML di un singolo algoritmo non ancora selezionato.			
40	b	a) Non è corretta: inverte i ruoli dei due tipi di test. Il canary testing serve direttamente un piccolo gruppo di utenti reali con il nuovo modello ML, mentre lo shadow testing elabora il traffico live in background senza mai influenzare le risposte agli utenti. b) È corretta: lo shadow testing offre un modo privo di rischi per osservare come si comporta un nuovo modello ML con i dati di produzione, mentre il canary testing espone deliberatamente un piccolo sottoinsieme di utenti al nuovo modello per osservarne le prestazioni e l'impatto nel mondo reale. c) Non è corretta: entrambi i tipi di test vengono utilizzati per validare le prestazioni e la correttezza funzionale di un modello sul traffico live. La differenza principale riguarda il profilo di rischio, non il fatto che uno sia per il performance testing e l'altro per il testing di integrazione dei componenti. d) Non è corretta: rappresenta in modo fondamentalmente errato il canary testing come un'attività offline. Sia il canary testing sia lo shadow testing sono metodi online che utilizzano dati utente live. La distinzione chiave è se la risposta del nuovo modello ML venga effettivamente inviata all'utente oppure venga solo registrata per un'analisi successiva.	AI-7.1.2	K2	1

Appendice: Risposte alle domande aggiuntive

Tabella Riassuntiva delle Risposte

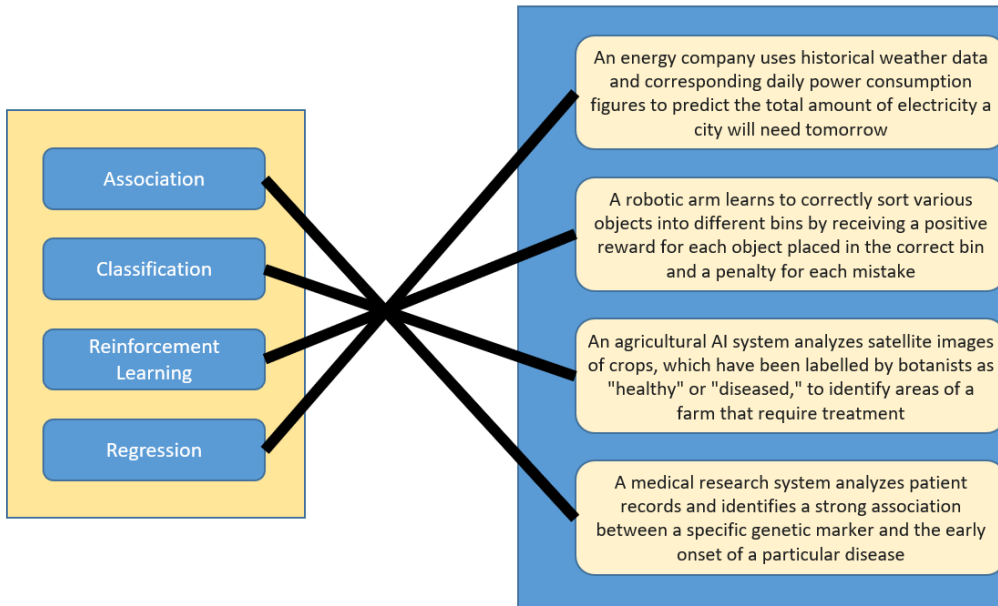
Numero domanda (#)	Risposta corretta	LO	Livello K	Punti
A1	Nuovo tipo di domanda	AI-1.1.2	K2	1
A2	a	AI-1.1.7	K2	1
A3	c	AI-1.1.8	K2	1
A4	Nuovo tipo di domanda	AI-3.1.1	K2	1
A5	Nuovo tipo di domanda	AI-3.1.2	K2	1
A6	c	AI-3.4.3	K2	1

Risposte

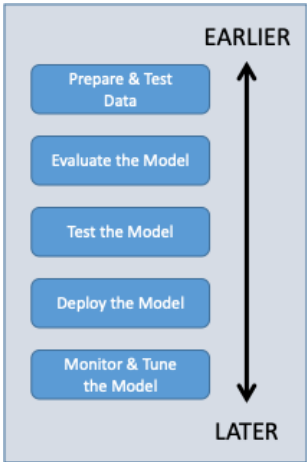
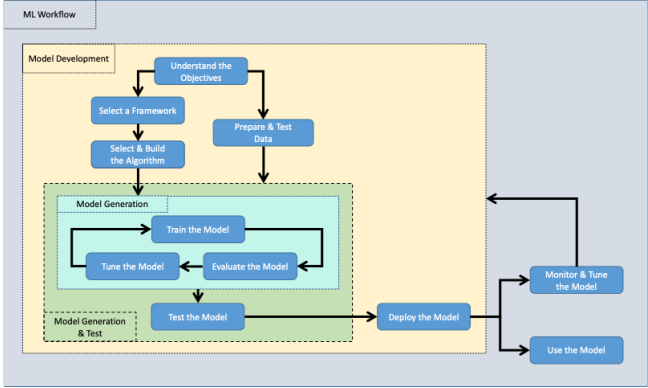
Numero domanda (#)	Risposta corretta	Spiegazione / Motivazione	Obiettivo di apprendimento (LO)	Livello K	Numero di punti
A1		Narrow AI A system that examines radiological images to detect the specific signatures of cancerous tumors. A language translation model that can convert written text from French to Spanish. General AI A system that manages complex daily schedules, learns new recipes from a video, and holds conversations about novels it has just read. A system that can independently learn any field of science and collaborate with human scientists by proposing novel hypotheses and original experiments. Super AI An artificial mind that generates entirely new forms of art, music, and mathematics that are incomprehensible to humans. Spiegazione / Motivazione	AI-1.1.2	K2	1

		Considerando ciascuno dei sistemi di esempio in ordine: - Una mente artificiale che genera forme completamente nuove di arte, musica e matematica incomprensibili per gli esseri umani. Questo sistema dimostra creatività e intelligenza ben oltre le capacità umane, producendo output non comprensibili dalle persone. Tali capacità superano sia l'intelligenza generale a livello umano sia qualsiasi general AI attuale, e si tratta quindi di un esempio di super AI. - Un sistema che gestisce programmi giornalieri complessi, apprende nuove ricette da un video e conversa su romanzi che ha appena letto. Questa AI svolge un'ampia gamma di attività complesse che richiedono apprendimento, percezione, pianificazione e comprensione del linguaggio, paragonabile alla versatilità umana. Mostra adattabilità e competenza in ambiti nuovi. Si tratta quindi di un esempio di AGI (general AI). - Un sistema in grado di apprendere autonomamente qualsiasi campo scientifico e di collaborare con scienziati umani proponendo nuove ipotesi ed esperimenti originali. Questo sistema ha la capacità di padroneggiare in modo autonomo e flessibile molteplici discipline scientifiche e di condurre ricerca creativa. La sua competenza copre l'apprendimento, il ragionamento e il pensiero innovativo al livello di un esperto umano. Si tratta quindi di un esempio di AGI (general AI). - Un sistema che esamina immagini radiologiche per rilevare le specifiche caratteristiche dei tumori cancerosi. Questa AI eccelle in un'attività ben definita e specialistica (l'analisi di immagini mediche), senza una comprensione o capacità più ampie al di fuori del proprio dominio. La sua applicazione è limitata al riconoscimento di pattern nei dati radiologici e non si estende a una cognizione generalizzata. Si tratta quindi di un esempio di narrow AI.			
--	--	--	--	--	--

		Un modello di traduzione linguistica in grado di convertire testo scritto dal francese allo spagnolo. Questo modello è costruito per svolgere una singola funzione: tradurre tra due lingue. Dimostra competenza in un'unica attività delimitata e non possiede una comprensione completa o un'adattabilità che vada oltre la conversione linguistica. Si tratta quindi di un esempio di narrow AI.			
A2	a	a) È corretta. La costruzione del modello è una funzione chiave, e i framework di sviluppo ML forniscono “strumenti per definire l'architettura del modello ML”, inclusa la possibilità di specificare la struttura del modello ML, come un albero decisionale. b) Non è corretta. Il termine “eliminano” è un'esagerazione. I framework di sviluppo ML supportano il pre-processing dei dati, ma non ne eliminano la necessità. La gestione dei dati è una funzione di supporto che non esegue un'automazione completa, eliminando l'attività. c) Non è corretta. I framework di sviluppo ML esistono a diversi livelli di astrazione, incluso “un'API di livello superiore, che semplifica la creazione del modello”. Ciò suggerisce che la programmazione non è un requisito universale, e la scelta del framework dipende dalla “competenza degli utenti”. d) Non è corretta. Non è vero che i framework di sviluppo ML vincolino i modelli ML sviluppati a quel framework. In generale, qualsiasi modello ML sviluppato può essere distribuito ovunque (entro limiti ragionevoli).	AI-1.1.7	K2	1
A3	c	a) Non è corretta. La ISO/IEC/IEEE TR 29119-11 non riguarda il testing dell'AI, ma, come tutti gli standard, fornisce linee guida e non impone sanzioni in caso di mancata conformità. b) Non è corretta. I principi OCSE costituiscono un insieme di linee guida non vincolanti. c) È corretta. Se il sistema rientra nell'ambito dell'EU AI Act (cioè verrà utilizzato in Europa), il mancato rispetto di un approccio basato sul rischio può comportare sanzioni severe.	AI-1.1.8	K2	1

		d) Non è corretta. La ISO/IEC 25059 descrive le caratteristiche di qualità specifiche dei sistemi AI, ma, come tutti gli standard, fornisce linee guida piuttosto che imporre sanzioni in caso di mancata conformità.			
A4		 Spiegazione / Motivazione Considerando ciascuna delle descrizioni dei sistemi di esempio dall'alto verso il basso: • Association (Associazione) - Questo sistema utilizza l'associazione analizzando registrazioni di pazienti non etichettate per identificare	AI-3.1.1	K2	1

		dipendenze tra attributi diversi, come un gene e una malattia. Non prevede una classe, ma rivela invece la forza della relazione tra le variabili all'interno dei dati. • Classification (Classificazione) - Apprende da un dataset di immagini pre-etichettate con categorie discrete ('sane' o 'malate'). Il suo scopo è categorizzare le nuove immagini in una di queste classi specifiche e non numeriche, il che la distingue dalla regressione. • Reinforcement learning (Apprendimento per rinforzo) - Il sistema apprende attraverso l'interazione diretta con il proprio ambiente fisico, non da un dataset preesistente. Il comportamento del robot viene modellato nel tempo da un sistema di ricompense (per il posizionamento corretto) e penalità (per il posizionamento errato). • Regression (Regressione) - Il modello ML viene addestrato su dati etichettati per prevedere un valore numerico specifico e continuo: la quantità di elettricità. L'output del modello non è una categoria ma una quantità su una scala, il che è la caratteristica chiave della regressione.			
A5		Risposta corretta	AI-3.1.2	K2	1

		 Spiegazione/Motivazione 			
A6	c	a) Non è corretta. Il limite principale riguarda il ragionamento interno del modello, un problema generale applicabile a molte attività, non specificamente a quelle che coinvolgono il feedback umano.	AI-3.4.3	K2	1

		b) Non è corretta. La copertura strutturale è un metodo per valutare l'accuratezza dei test di un singolo modello, non per confrontare modelli diversi. Sebbene le dimensioni del modello possano influire sui numeri di copertura grezzi, i valori calcolati vengono tipicamente presentati come percentuali, e pertanto questo non costituisce un limite rispetto al suo scopo previsto. c) È corretta. Affronta direttamente i limiti chiave delle misure di copertura utilizzate in isolamento. Le reti neurali possono apprendere correlazioni spurie, portando ad attivazioni corrette per ragioni sbagliate, il che significa che il modello potrebbe raggiungere un'elevata copertura strutturale pur prendendo decisioni basate su feature irrilevanti. d) Non è corretta. La copertura strutturale è un complemento utile che dovrebbe essere combinato con altri metodi. È pensata per integrare, non sostituire, altre attività di test, come la valutazione della correttezza funzionale e dell'utilità degli output del modello.			
--	--	---	--	--	--