

Simulacro de Examen: AWS Certified AI Practitioner (AIF-C01)

Dominio 3: Aplicación de Modelos Fundacionales (150 Preguntas)

Material de Estudio Intensivo

Instrucciones

Este cuestionario contiene 150 preguntas de opción múltiple. Selecciona la respuesta correcta para cada una. La clave de respuestas se encuentra en la última página del documento.

Preguntas

1. ¿Qué característica de los modelos fundacionales (FM) los diferencia principalmente de los modelos de ML tradicionales?
 - a) Su capacidad para realizar una sola tarea específica como análisis de sentimiento.
 - b) Su naturaleza de uso general y adaptabilidad a múltiples tareas mediante peticiones.
 - c) Que requieren menos datos de entrenamiento que los modelos tradicionales.
 - d) Que son sistemas deterministas.
2. Al elegir un modelo previo al entrenamiento, ¿cuál es el equilibrio principal que debe considerar un arquitecto?
 - a) Solo la exactitud del modelo.
 - b) El equilibrio entre tiempo de entrenamiento, coste y rendimiento.
 - c) El color de la interfaz de usuario.
 - d) El número de usuarios concurrentes únicamente.
3. En un sistema de vehículos autónomos donde las decisiones deben ser instantáneas, ¿cuál es la restricción de diseño más crítica?
 - a) El coste de almacenamiento.
 - b) La latencia de inferencia.
 - c) El número de idiomas soportados.
 - d) La antigüedad del dataset.
4. ¿Por qué un modelo K-vecinos más cercanos (KNN) podría no ser ideal para inferencias en tiempo real en problemas de grandes dimensiones?

- a) Porque no usa GPUs.
 - b) Porque realiza la mayor parte del trabajo computacional durante la etapa de inferencia, aumentando la latencia.
 - c) Porque es un modelo fundacional.
 - d) Porque solo funciona con texto.
5. ¿Qué mide la métrica de "precisión media promedio" (mAP) en el contexto de visión artificial?
- a) La velocidad de carga del modelo.
 - b) La capacidad del modelo para localizar y clasificar correctamente varios objetos en una imagen.
 - c) El sentimiento de los usuarios hacia el modelo.
 - d) El número de parámetros del Transformer.
6. ¿Por qué se dice que los modelos fundacionales son cajas negras?
- a) Porque su código fuente no es público en AWS.
 - b) Debido a su extrema complejidad que impide la interpretabilidad matemática directa de cada decisión.
 - c) Porque solo se pueden ejecutar de noche.
 - d) Porque no aceptan entradas de imagen.
7. ¿Cuál es la diferencia técnica entre interpretabilidad y explicabilidad?
- a) Son sinónimos en AWS.
 - b) La interpretabilidad es entender la mecánica interna; la explicabilidad intenta aproximar esa mecánica con modelos más simples.
 - c) La explicabilidad es solo para modelos de regresión.
 - d) La interpretabilidad no requiere matemáticas.
8. ¿Qué parámetro de inferencia controla la aleatoriedad y diversidad de la respuesta en Amazon Bedrock?
- a) Longitud de respuesta.
 - b) Temperatura.
 - c) Secuencia de parada.
 - d) Cifrado KMS.
9. Si desea una respuesta altamente determinista y conservadora de un LLM, ¿qué valor de temperatura debería elegir?
- a) Un valor alto (ej. 1.0).
 - b) Un valor bajo (ej. 0.1).
 - c) El valor por defecto de la GPU.
 - d) Cero absoluto.

10. ¿Qué hace el parámetro "Top P" (Nucleus Sampling)?

- a) Elige siempre la palabra más larga.
- b) Selecciona del subconjunto de tokens cuya probabilidad acumulada suma el valor P.
- c) Limita la salida a P caracteres.
- d) Borra los tokens menos frecuentes del dataset.

11. ¿Cuál es la función del parámetro "Top K"?

- a) Limitar el pool de candidatos a los K tokens más probables.
- b) Multiplicar el número de parámetros por K.
- c) Detener la inferencia después de K segundos.
- d) Repetir la respuesta K veces.

12. ¿Qué técnica permite mitigar alucinaciones proporcionando contexto de una base de conocimientos externa en tiempo de inferencia?

- a) Fine-tuning masivo.
- b) RAG (Generación Aumentada por Recuperación).
- c) RLHF.
- d) Poda de parámetros.

13. ¿Cuál es un requisito previo indispensable para crear una base de datos vectorial?

- a) Un servidor de correo.
- b) Un modelo de machine learning (generalmente de incrustación/embedding).
- c) Una suscripción a Netflix.
- d) Desactivar el firewall.

14. ¿Qué servicio de AWS permite implementar RAG de forma gestionada conectando modelos a datos empresariales?

- a) Amazon S3 solo.
- b) Bases de conocimiento de Amazon Bedrock (Knowledge Bases).
- c) AWS Shield.
- d) Amazon WorkSpaces.

15. ¿Qué extensión de Amazon RDS para PostgreSQL permite almacenar y buscar incrustaciones vectoriales?

- a) postgis.
- b) pgvector.
- c) uuid-ossp.
- d) pgcrypto.

16. ¿Qué servicio de AWS ofrece GraphRAG para analizar relaciones topológicas complejas en los datos?
- a) Amazon OpenSearch.
 - b) Amazon Neptune Analytics.
 - c) Amazon Aurora.
 - d) Amazon DocumentDB.
17. ¿Qué herramienta de Bedrock orquesta tareas de varios pasos, llama a APIs y consulta bases de conocimiento automáticamente?
- a) Bedrock Playgrounds.
 - b) Agents para Amazon Bedrock.
 - c) Amazon Titan Lite.
 - d) AWS Glue.
18. Un modelo que no puede reservar un vuelo por sí solo necesita un Agente porque:
- a) El modelo no sabe hablar.
 - b) El modelo carece de la lógica de orquestación y acceso a flujos de trabajo específicos de la organización.
 - c) Los modelos son demasiado pequeños.
 - d) Las APIs no son compatibles con texto.
19. ¿Cuál es la principal ventaja de usar Amazon OpenSearch Serverless para vectores?
- a) Que es más lento pero seguro.
 - b) Que permite realizar búsquedas vectoriales sin gestionar la infraestructura de la base de datos.
 - c) Que solo acepta datos estructurados.
 - d) Que no tiene coste alguno.
20. ¿Qué mide la métrica BERTscore utilizada en Amazon Bedrock?
- a) La velocidad de la CPU.
 - b) La similitud semántica entre la respuesta generada y una referencia humana.
 - c) El número de faltas de ortografía.
 - d) La temperatura de la red.
21. ¿Cómo define AWS la Ingeniería de Peticiones?
- a) El proceso de programar el microchip de la GPU.
 - b) La práctica de crear y optimizar las entradas (instrucciones, palabras, signos) para guiar eficazmente a los LLMs.
 - c) Un método para borrar la memoria del modelo.
 - d) La construcción física de servidores.

22. ¿Qué es una petición "Few-shot"?

- a) Una petición muy corta.
- b) Una petición que incluye algunos ejemplos para ayudar al modelo a calibrar su resultado.
- c) Una petición enviada por un usuario sin cuenta de AWS.
- d) Una petición que no recibe respuesta.

23. ¿Qué técnica de prompting pide al modelo que desglose su proceso de razonamiento en pasos intermedios?

- a) Zero-shot.
- b) Cadena de pensamiento (Chain of Thought).
- c) Prompt Tuning.
- d) Inyección de prompts.

24. ¿Qué describe el ".Espacio Latente" de un modelo de lenguaje?

- a) El tiempo de espera del usuario.
- b) El conocimiento codificado y los patrones estadísticos que capturan relaciones entre datos.
- c) El disco duro de Amazon S3.
- d) La papelera de reciclaje de los modelos.

25. ¿Por qué un modelo puede alucinar a pesar de que su ".espacio latente." esté funcionando correctamente?

- a) Porque el modelo está roto.
- b) Porque el tema de la petición no está presente o es insuficiente en su conocimiento preentrenado, forzándolo a elegir la coincidencia estadística más próxima.
- c) Porque el usuario escribió mal una palabra.
- d) Debido a un exceso de memoria RAM.

26. ¿Cuál es un riesgo de seguridad donde un usuario intenta eludir las barreras de protección del modelo mediante entradas ingeniosas?

- a) Poda.
- b) Jailbreak.
- c) RAG.
- d) Inferencia por lotes.

27. ¿Qué técnica avanzada reemplaza el texto de la petición por un soporte de incrustación continua optimizado durante el entrenamiento?

- a) Fine-tuning masivo.
- b) Ajuste de peticiones (Prompt Tuning).

- c) Few-shot prompting.
 - d) Agentes de Bedrock.
28. ¿Qué diferencia al "secuestro" (hijacking) de una inyección de petición normal?
- a) El secuestro intenta cambiar la instrucción original de la petición por una nueva.
 - b) El secuestro solo ocurre en dispositivos móviles.
 - c) No hay diferencia.
 - d) El secuestro es un término legal, no técnico.
29. ¿Qué técnica de prompting es ideal para problemas matemáticos o de pensamiento lógico complejos?
- a) Clasificación simple.
 - b) Cadena de pensamiento (Chain of Thought).
 - c) Zero-shot.
 - d) Resumen automático.
30. ¿Qué recurso de Amazon Bedrock permite establecer filtros de seguridad para categorías dañinas y palabras bloqueadas?
- a) Amazon S3.
 - b) Barreras de protección (Guardrails).
 - c) Amazon Titan Image Generator.
 - d) AWS Glue.
31. ¿Qué es el ".envenenamiento" (poisoning) en el contexto de peticiones?
- a) Incluir instrucciones dañinas en fuentes externas que el modelo podría ingerir (webs, correos).
 - b) Apagar el sistema de seguridad de AWS.
 - c) Borrar los datasets de entrenamiento.
 - d) Un virus en el hardware.
32. ¿Cuál es una práctica recomendada fundamental para un ingeniero de peticiones?
- a) Ser lo más vago posible para que el modelo sea creativo.
 - b) Ser específico e incluir instrucciones claras sobre formato, estilo y contexto.
 - c) No usar ejemplos.
 - d) Usar solo una palabra.
33. ¿Qué es la técnica "Least-to-Most.en ingeniería de peticiones?
- a) Pedir al modelo que trabaje menos.
 - b) Desglosar una pregunta compleja en subproblemas sencillos y resolverlos secuencialmente.

- c) Usar el modelo más pequeño disponible.
 - d) Reducir el número de tokens al mínimo.
34. ¿Qué técnica solicita al modelo que genere primero los datos pertinentes antes de completar la tarea?
- a) Chain of Thought.
 - b) Indicaciones de conocimientos generados (Generated Knowledge).
 - c) Zero-shot.
 - d) Poda.
35. ¿En qué consiste la técnica de Refinamiento automático”(Self-refinement)?
- a) El modelo se apaga y se reinicia.
 - b) El modelo soluciona el problema, critica su propia solución y la vuelve a resolver.
 - c) El desarrollador corrige el código.
 - d) Amazon Bedrock actualiza la versión del modelo.
36. ¿Qué es un .^{Estímulo} Direccional.^{en} un prompt?
- a) Un mapa de la ciudad.
 - b) Una sugerencia o pista (palabras clave) para guiar al modelo hacia el resultado deseado.
 - c) La dirección IP del servidor.
 - d) Un comando de voz.
37. ¿Por qué es importante equilibrar la simplicidad y la complejidad en una petición?
- a) Para gastar menos electricidad.
 - b) Para evitar respuestas vagas (por ser demasiado simple) o confusiones (por ser demasiado compleja).
 - c) Para que el modelo cargue más rápido.
 - d) No es importante.
38. ¿Cuál es el beneficio de usar plantillas de peticiones?
- a) Reducen el coste de AWS.
 - b) Permiten reutilizar estructuras exitosas para diferentes casos prácticos de forma consistente.
 - c) El modelo aprende más rápido.
 - d) Eliminan la necesidad de bases de datos.
39. ¿Qué técnica utiliza una búsqueda en árbol para explorar múltiples pasos siguientes posibles del modelo?
- a) Cadena de pensamiento.
 - b) Árbol de pensamientos (Tree of Thoughts).

- c) Mayéutica.
 - d) Zero-shot.
40. ¿Qué técnica de prompting busca mejorar el razonamiento de sentido común mediante árboles de explicaciones?
- a) Chain of Thought.
 - b) Peticiones Mayéuticas.
 - c) Least-to-Most.
 - d) Prompt Tuning.
41. ¿Qué proceso de entrenamiento utiliza datos no estructurados y aprendizaje autosupervisado a escala de petabytes?
- a) Refinamiento (Fine-tuning).
 - b) Entrenamiento Previo (Pre-training).
 - c) RLHF.
 - d) Prompt Engineering.
42. ¿Cuál es el objetivo principal del Refinamiento” (Fine-tuning)?
- a) Crear un modelo desde cero.
 - b) Adaptar un modelo preentrenado a una tarea o conjunto de datos personalizado mediante aprendizaje supervisado.
 - c) Borrar los pesos originales del modelo.
 - d) Reducir el número de GPUs.
43. ¿Qué término describe cuando un modelo refinado para una tarea olvida cómo realizar otras tareas generales?
- a) Error de sistema.
 - b) Olvido catastrófico.
 - c) Poda radical.
 - d) Alucinación profunda.
44. ¿Qué conjunto de técnicas permite adaptar modelos fundacionales congelando la mayoría de los parámetros originales?
- a) Full Fine-tuning.
 - b) PEFT (Parameter-Efficient Fine-Tuning).
 - c) RAG.
 - d) Zero-shot.
45. ¿Qué técnica de PEFT crea matrices de ”bajo rango.”^{entrenables} en cada capa de la arquitectura del Transformer?
- a) ReFT.

- b) LoRA (Low-Rank Adaptation).
 - c) Poda.
 - d) Softmax.
46. ¿Qué significa congelar un modelo durante el refinamiento?
- a) Guardarlo en un servidor frío.
 - b) Preservar los parámetros y ponderaciones originales sin permitir que cambien durante el nuevo entrenamiento.
 - c) Detener la inferencia.
 - d) Borrar el dataset.
47. ¿Cuál es la ventaja del Refinamiento Multitarea?
- a) Es más barato.
 - b) Ayuda a mitigar el olvido catastrófico al entrenar el modelo en varias tareas simultáneamente.
 - c) No requiere datos etiquetados.
 - d) Solo funciona en Amazon Titan.
48. ¿Qué proceso de refinamiento se usa para alinear las respuestas del modelo con los valores humanos de honestidad y amabilidad?
- a) Poda.
 - b) RLHF (Aprendizaje por refuerzo a partir de retroalimentación humana).
 - c) Inferencia por lotes.
 - d) Tokenización.
49. ¿Qué servicio de AWS ofrece la capacidad de refinar modelos de generación de texto con datasets específicos del dominio?
- a) Amazon SageMaker JumpStart.
 - b) Amazon Polly.
 - c) AWS Shield.
 - d) Amazon Macie.
50. ¿Qué herramienta de AWS permite preparar datos mediante flujos "no-code" (sin código)?
- a) AWS Glue (Spark).
 - b) Amazon SageMaker Canvas.
 - c) Amazon EMR.
 - d) Amazon RDS.
51. ¿Cuál es la función de SageMaker Clarify en la preparación de datos para el refinamiento?

- a) Traducir los datos.
 - b) Detectar sesgos potenciales (género, raza, edad) y desequilibrios en el dataset.
 - c) Aumentar la velocidad de la red.
 - d) Borrar archivos duplicados.
52. ¿Qué es el ^{en}trenamiento previo continuo?
- a) No dejar nunca de entrenar.
 - b) Seguir entrenando un FM con datos no etiquetados para acumular mayor conocimiento y adaptabilidad.
 - c) Un error en la canalización de ML.
 - d) Entrenar solo los fines de semana.
53. ¿Qué modelos de Amazon Titan admiten el preentrenamiento continuo en Bedrock?
- a) Solo Titan Image.
 - b) Titan Text Express y Titan Text Lite.
 - c) Ninguno.
 - d) Solo modelos de terceros.
54. ¿Para qué se usa el "dataset de retención" (holdout set) al final del refinamiento?
- a) Para borrar el modelo.
 - b) Para realizar una evaluación final del rendimiento con datos que el modelo nunca ha visto.
 - c) Para entrenar los hiperparámetros.
 - d) Para guardar copias de seguridad.
55. Durante el refinamiento, ¿qué se calcula para actualizar las ponderaciones del modelo?
- a) La temperatura.
 - b) La pérdida (loss) entre la finalización del modelo y la etiqueta de entrenamiento.
 - c) El número de tokens.
 - d) La latencia de red.
56. ¿Qué biblioteca de código abierto se usa a menudo con SageMaker para entrenamiento distribuido paralelo?
- a) DeepSpeed.
 - b) React.
 - c) WordPress.
 - d) Excel.
57. ¿Qué técnica de refinamiento se centra en intervenciones en las representaciones ocultas. ^{en} lugar de pesos?

- a) LoRA.
 - b) ReFT (Representation Finetuning).
 - c) RAG.
 - d) Poda.
58. ¿Qué describe la "hipótesis de la representación lineal"?
- a) Que todos los modelos son líneas rectas.
 - b) Que los conceptos están codificados en subespacios lineales de representación en la red neuronal.
 - c) Que el aprendizaje es infinito.
 - d) Que el modelo no tiene capas ocultas.
59. ¿Cómo ayuda el refinamiento para la adaptación de dominios?
- a) Cambia el nombre del modelo.
 - b) Permite que el modelo aprenda jerga técnica, términos específicos de industria y datos especializados.
 - c) Reduce el precio de AWS.
 - d) Elimina la necesidad de internet.
60. ¿Cuál es un paso crítico en la preparación de datos antes del refinamiento?
- a) Formatear los datos como incrustaciones vectoriales.
 - b) Dividir los datos en entrenamiento, validación y prueba.
 - c) Comprar más discos duros.
 - d) Traducir todo al inglés.
61. ¿Por qué es difícil evaluar modelos fundacionales con métricas deterministas como la exactitud simple?
- a) Porque no usan números.
 - b) Porque su resultado no es determinista y no se puede comparar fácilmente con una única etiqueta fija.
 - c) Porque AWS no lo permite.
 - d) Porque son demasiado pequeños.
62. ¿Qué técnica de optimización de inferencia consiste en reducir el tamaño del LLM para que cargue más rápido?
- a) RAG.
 - b) Reducción del tamaño del modelo (ej. destilación o cuantización).
 - c) Aumentar la temperatura.
 - d) Usar prompts más largos.

63. ¿Qué métrica se usa específicamente para evaluar la calidad de los resúmenes automáticos comparándolos con referencias humanas?

- a) BLEU.
- b) ROUGE.
- c) RMSE.
- d) Accuracy.

64. ¿Para qué tarea se diseñó originalmente el algoritmo BLEU?

- a) Generación de imágenes.
- b) Traducción automática de idiomas.
- c) Detección de fraude.
- d) Clasificación de spam.

65. ¿Qué mide el punto de referencia GLUE (General Language Understanding Evaluation)?

- a) La temperatura del servidor.
- b) La capacidad del modelo para generalizar en diversas tareas de lenguaje natural.
- c) El precio de los tokens.
- d) El número de GPUs necesarias.

66. ¿Qué añade "SuperGlue" respecto a GLUE?

- a) Más memoria.
- b) Tareas adicionales complejas como razonamiento en varias frases y comprensión lectora.
- c) Soporte para audio.
- d) Mayor velocidad.

67. ¿Qué mide el benchmark MMLU (Massive Multitask Language Understanding)?

- a) Solo la gramática.
- b) El conocimiento del mundo y las capacidades de resolución de problemas en temas como historia, matemáticas y legislación.
- c) La latencia de red.
- d) El consumo de energía.

68. ¿Cuál es el enfoque del benchmark BIG-bench?

- a) Tareas muy sencillas.
- b) Tareas extremadamente difíciles que van más allá de las capacidades actuales de los modelos (biología, razonamiento lógico, etc.).
- c) Solo traducción al español.
- d) Limpieza de bases de datos.

69. ¿Qué ventaja ofrece el benchmark HELM (Holistic Evaluation of Language Models)?
- a) Es el más barato.
 - b) Ofrece transparencia y orientación sobre qué modelo funciona mejor para tareas específicas como detección de sesgos.
 - c) Elimina la necesidad de evaluación humana.
 - d) Solo funciona con modelos de Amazon.
70. ¿Cómo ayuda Amazon SageMaker Clarify en la evaluación de LLMs?
- a) Permite crear trabajos de evaluación para comparar métricas de calidad de modelos basados en texto de JumpStart.
 - b) Traduce los resultados.
 - c) Borra los modelos con bajo rendimiento.
 - d) Genera facturas automáticamente.
71. ¿Qué es la "Perplejidad" (PPL) en la evaluación de modelos de lenguaje?
- a) Un estado de confusión del usuario.
 - b) Una medida de qué tan bien un modelo es capaz de predecir la siguiente palabra en una secuencia.
 - c) El tamaño del prompt.
 - d) La latencia de inferencia.
72. En términos de Perplejidad (PPL), ¿qué indica una puntuación más baja?
- a) Que el modelo es peor.
 - b) Que el modelo predice mejor y comprende más el lenguaje humano.
 - c) Que el modelo es más caro.
 - d) Que no tiene datos.
73. ¿Qué capa de la pila de aplicaciones de GenAI proporciona computación, red y almacenamiento?
- a) Capa de Aplicación.
 - b) Capa de Infraestructura.
 - c) Capa de Orquestación.
 - d) Capa de Marketing.
74. ¿Por qué una empresa necesitaría almacenamiento adicional en su pila de GenAI?
- a) Para guardar películas.
 - b) Para recopilar finalizaciones y comentarios de usuarios para futuros ajustes y evaluaciones.
 - c) Para ralentizar el sistema.
 - d) No es necesario almacenamiento adicional.

75. ¿Qué capa de la pila incluye la interfaz web o API REST que consumen los usuarios finales?

- a) Capa de Infraestructura.
- b) Capa de Aplicación.
- c) Capa de Modelo.
- d) Capa de Datos.

76. ¿Cuál es una solución para el problema del conocimiento desactualizado.^{en} un LLM sin reentrenarlo?

- a) Apagar el modelo.
- b) Usar RAG (Generación Aumentada por Recuperación) para acceder a datos externos actuales en tiempo de inferencia.
- c) Cambiar el nombre del modelo.
- d) Borrar la base de datos.

77. ¿Qué mide la métrica de "fidelidad."^{en} una evaluación de Bedrock?

- a) La velocidad de respuesta.
- b) Si la respuesta se fundamenta correctamente en el contexto proporcionado sin inventar datos (alucinaciones).
- c) El número de usuarios leales.
- d) El color de las imágenes generadas.

78. ¿Qué componente se encarga de transferir las entradas del usuario al LLM y devolver las finalizaciones en una aplicación compleja?

- a) El disco duro.
- b) La biblioteca de orquestación.
- c) El monitor de energía.
- d) La tarjeta de red únicamente.

79. ¿Cuál es la mayor desventaja de reducir el tamaño de un modelo (destilación)?

- a) El modelo se vuelve más grande.
- b) Puede haber una reducción en el rendimiento o exactitud del modelo.
- c) El modelo deja de funcionar en AWS.
- d) Los tokens se vuelven más caros.

80. En la evaluación de modelos, ¿qué papel juegan los trabajadores humanos?

- a) Programar el modelo.
- b) Realizar evaluaciones manuales comparativas de respuestas para determinar preferencias de calidad.
- c) Ninguno.

- d) Reemplazar a la IA totalmente.
81. ¿Qué modelo fundacional de Anthropic es el más inteligente y avanzado según los benchmarks?
- a) Claude 3 Haiku.
 - b) Claude 3.5 Sonnet.
 - c) Claude 2.0.
 - d) Claude Titan.
82. ¿Cuál es la principal característica del modelo Claude 3 Haiku?
- a) Es el más lento pero preciso.
 - b) Es el modelo más rápido y compacto, con respuesta casi instantánea.
 - c) Solo genera audio.
 - d) Requiere 100 GPUs para funcionar.
83. ¿Qué familia de modelos de Amazon Nova se centra en la generación de contenido visual (imagen/video)?
- a) Amazon Nova Micro.
 - b) Amazon Nova Canvas y Reel.
 - c) Amazon Nova Pro.
 - d) Amazon Nova Text.
84. ¿Qué distingue a los modelos de Amazon Nova de comprensión?
- a) Solo aceptan texto.
 - b) Aceptan entradas multimodales (texto, imagen, video) y generan salidas de texto.
 - c) Son sistemas de reglas.
 - d) No usan redes neuronales.
85. ¿Qué modelo fundacional es ideal para la conversión de texto a imagen y puede ejecutarse incluso en un smartphone?
- a) GPT-4.
 - b) Stable Diffusion.
 - c) BERT.
 - d) Claude Opus.
86. ¿Qué es el "Brevity Penalty" en el cálculo de la puntuación BLEU?
- a) Una multa por escribir poco.
 - b) Un factor que penaliza traducciones generadas que son demasiado cortas comparadas con la referencia humana.
 - c) Un error de memoria.

- d) Una técnica de compresión.
87. ¿Cuál es la principal limitación de la métrica BLEU?
- a) Solo funciona en inglés.
 - b) Compara solapamiento de tokens en lugar de significado semántico profundo.
 - c) Es muy cara de calcular.
 - d) No existe limitación.
88. ¿Qué característica del modelo BLOOM destaca sobre otros modelos?
- a) Es un modelo pequeño.
 - b) Es un modelo multilingüe masivo (46 idiomas) fruto de una colaboración científica global.
 - c) Solo sabe programar en Python.
 - d) Es propiedad exclusiva de Amazon.
89. ¿Qué plataforma actúa como un "centro comunitario" para compartir casi 200,000 modelos de ML?
- a) Facebook.
 - b) Hugging Face.
 - c) Amazon Prime.
 - d) Wikipedia.
90. ¿Cuál es un desafío ético de los modelos fundacionales entrenados con datos de internet?
- a) Que son demasiado lentos.
 - b) El sesgo (Bias), ya que pueden captar discursos de odio o matices inapropiados del dataset de entrenamiento.
 - c) Que no saben sumar.
 - d) El consumo de agua de los servidores.
91. ¿Qué servicio de AWS ayuda a monitorizar la latencia de invocación de Bedrock en tiempo real?
- a) AWS Glue.
 - b) Amazon CloudWatch (métrica InvocationLatency).
 - c) AWS Shield.
 - d) Amazon RDS.
92. ¿Qué permite el "Model Invocation Logging." en Amazon Bedrock?
- a) Borrar los prompts de los usuarios.
 - b) Recolectar metadatos, peticiones y respuestas para auditoría y análisis.
 - c) Aumentar la temperatura del modelo.

d) Traducir logs automáticamente.

93. ¿Qué ventaja ofrece Amazon MemoryDB en el almacenamiento de vectores?

- a) Es la opción más barata.
- b) Ofrece la latencia más baja (milisegundos de un dígito) con el mayor recall”.
- c) Solo funciona con imágenes.
- d) No requiere RAM.

94. ¿Qué algoritmos de búsqueda vectorial utiliza principalmente Amazon OpenSearch?

- a) Solo búsqueda lineal.
- b) HNSW (Pequeño mundo navegable jerarquizado) e IVF (Archivo invertido).
- c) Burbuja y QuickSort.
- d) Ninguno.

95. ¿Para qué sirve la extensión ”pgvector.” en Amazon Aurora?

- a) Para enviar correos.
- b) Para realizar búsquedas de similitud (Euclídea, Coseno) en datos vectoriales usando SQL.
- c) Para comprimir la base de datos.
- d) Para borrar tablas automáticamente.

96. ¿Qué describe el término Chunking. en la preparación de datos para RAG?

- a) Borrar los datos.
- b) Dividir documentos extensos en fragmentos más pequeños y significativos para su indexación.
- c) Cambiar el nombre de los archivos.
- d) Encriptar el disco duro.

97. ¿Cuál es un beneficio de usar citas. en las respuestas de RAG de Bedrock?

- a) El texto se ve más bonito.
- b) Permite al usuario verificar la fuente original y reduce la probabilidad de alucinaciones.
- c) Aumenta el coste.
- d) No tiene beneficios.

98. ¿Qué es el .envenenamiento de datos”(data poisoning) en el entrenamiento previo?

- a) Manipular los datos de entrenamiento para que el modelo aprenda patrones incorrectos o maliciosos.
- b) Un error de hardware.
- c) Borrar la base de datos.

- d) Usar datos antiguos.
99. ¿Qué es la inversión de modelo?
- a) Un ataque para intentar extraer datos confidenciales del dataset de entrenamiento a partir de las salidas del modelo.
 - b) Dar la vuelta a la GPU.
 - c) Cambiar el orden de las capas.
 - d) Desinstalar el modelo.
100. ¿Cuál es la prioridad número uno de AWS en infraestructura de IA?
- a) La velocidad.
 - b) La seguridad y confidencialidad de las cargas de trabajo del cliente.
 - c) El diseño de los logos.
 - d) El bajo coste únicamente.
101. ¿Qué componente de AWS garantiza que el tráfico de red de IA no pase por el internet público?
- a) Amazon VPC y AWS PrivateLink.
 - b) Google Chrome.
 - c) Un cable USB.
 - d) Facebook.
102. ¿Qué describe mejor a los .^agentes.^{en} Bedrock?
- a) Empleados de Amazon que responden chats.
 - b) Software autónomo que descompone tareas, planea pasos y ejecuta acciones vía APIs.
 - c) Robots físicos.
 - d) Virus informáticos.
103. ¿Qué es la .^autoatención global.^{en} Vision Transformers (ViT)?
- a) Mirar solo una esquina de la foto.
 - b) Capturar relaciones entre cualquier par de parches de la imagen, sin importar su posición.
 - c) Ignorar los colores.
 - d) Apagar la cámara.
104. ¿Qué técnica de Fine-tuning es la más eficiente en términos de computación y almacenamiento según el material avanzado?
- a) Refinamiento completo.
 - b) PEFT / LoRA.
 - c) ReFT (Representation Finetuning).

- d) Zero-shot.
105. ¿Qué describe el benchmark "SuperGlue" como mejora sobre GLUE?
- a) Es más pegajoso.
 - b) Tareas que requieren un razonamiento más profundo y lógica en contextos complejos.
 - c) Solo para imágenes.
 - d) Soporta más idiomas.
106. ¿Qué servicio de AWS se usa para "Generar código con ayuda de IA basándose en comentarios?"
- a) Amazon Q Developer.
 - b) Amazon Bedrock Playgrounds.
 - c) Amazon S3.
 - d) AWS Glue.
107. ¿Qué es un "Logit." en la salida de un Transformer?
- a) Un error de red.
 - b) Puntuaciones numéricas para cada token antes de ser normalizadas por Softmax.
 - c) El nombre de la base de datos.
 - d) Una unidad de medida de calor.
108. ¿Qué describe la técnica "Self-Refinement." en prompts?
- a) Pedir al usuario que corrija el prompt.
 - b) El modelo genera una respuesta, la critica internamente y la mejora en un bucle iterativo.
 - c) Cambiar el modelo por uno más grande.
 - d) Reiniciar AWS.
109. ¿Qué permite el ".aprendizaje en contexto" (In-context learning)?
- a) Entrenar el modelo durante años.
 - b) Que el modelo aprenda una tarea nueva a partir de los ejemplos proporcionados en la ventana de contexto del prompt.
 - c) Borrar los datos del cliente.
 - d) Usar solo modelos pequeños.
110. ¿Qué es la "Curse of Dimensionality" mencionada en los recursos de bases de datos vectoriales?
- a) Que los datos se vuelven infinitos.
 - b) La dificultad de encontrar similitudes cuando los vectores tienen demasiadas dimensiones, haciendo que los datos parezcan dispersos.

- c) Un virus en el servidor.
d) El coste de la electricidad.
111. ¿Cuál es la ventaja de los procesadores "AWS Graviton3" para la inferencia?
- a) Son más coloridos.
b) Hasta un 50 % de ahorro en coste y un 60 % menos de energía con rendimiento similar a x86.
c) No requieren ventiladores.
d) Solo funcionan en Bedrock.
112. ¿Qué describe el nivel 1 de MLOps?
- a) Todo es manual.
b) Automatización de la canalización de entrenamiento continuo.
c) CI/CD completo con múltiples modelos.
d) Solo lectura de datos.
113. ¿Qué describe el nivel 2 de MLOps?
- a) Nivel inicial de un estudiante.
b) CI/CD completo y orquestación automatizada para empresas que escalan con múltiples canalizaciones.
c) No usar automatización.
d) Solo para bases de datos SQL.
114. ¿Qué es un "Modelo autorregresivo como GPT"?
- a) Un modelo que se conduce solo.
b) Un modelo que predice el siguiente valor de una secuencia basándose en todos los anteriores.
c) Un modelo que no usa datos.
d) Un sistema basado en reglas.
115. ¿Qué técnica de prompting divide una pregunta compleja en partes lógicas que imitan el pensamiento humano?
- a) Cadena de pensamiento (Chain of Thought).
b) Zero-shot.
c) Poda.
d) Inferencia por lotes.
116. ¿Qué técnica de prompting explora varios caminos de razonamiento y elige el mejor?
- a) Árbol de pensamientos (Tree of Thoughts).
b) Mayéutica.

- c) Prompt Tuning.
 - d) RAG.
117. ¿Qué métrica penaliza más los errores grandes en un modelo de regresión de FM?
- a) MAE.
 - b) RMSE.
 - c) Exactitud.
 - d) F1.
118. ¿Qué es un "stop sequence." en Bedrock?
- a) Una multa de tráfico.
 - b) Una cadena de caracteres que indica al modelo que deje de generar más tokens.
 - c) El fin de la suscripción de AWS.
 - d) Una pausa de 5 minutos.
119. ¿Cuál es el beneficio de .^Amazon Titan Text Lite?
- a) Es el modelo más inteligente del mundo.
 - b) Es un modelo rentable y eficiente para tareas de texto más ligeras como resúmenes simples.
 - c) Solo genera imágenes.
 - d) Es de código abierto.
120. ¿Qué servicio se usa para monitorizar desviaciones de datos y de concepto en producción?
- a) SageMaker Model Monitor.
 - b) Amazon S3.
 - c) AWS Glue.
 - d) Amazon Polly.
121. ¿Cuál es la mayor limitación de un modelo con muchas .^alucinaciones?
- a) Es muy lento.
 - b) Proporciona respuestas que parecen correctas pero son objetivamente falsas, dañando la confianza del usuario.
 - c) Gasta mucho dinero.
 - d) No tiene parámetros.
122. ¿Qué arquitectura neuronal permite al modelo "mirar hacia atrás y hacia adelante."^{en} una oración al mismo tiempo?
- a) Unidireccional.
 - b) Bidireccional (como BERT).

- c) Regresión Lineal.
d) Árbol binario.
123. ¿Qué técnica de PEFT es considerada un reemplazo directo” que aprende intervenciones 10x-50x más eficientes?
- a) LoReFT.
 - b) RAG avanzado.
 - c) Fine-tuning completo.
 - d) Inferencia por lotes.
124. ¿Qué describe el ”Holistic Evaluation of Language Models” (HELM)?
- a) Una técnica de respiración.
 - b) Un punto de referencia que combina múltiples métricas para mejorar la transparencia y evaluación de modelos.
 - c) Un servicio de seguridad de red.
 - d) Una marca de auriculares.
125. ¿Qué significa que un modelo fundacional sea ”Multipropósito”?
- a) Que tiene muchos dueños.
 - b) Que puede realizar tareas de contenido (texto, imagen, audio) y razonamiento con el mismo modelo base.
 - c) Que es muy caro.
 - d) Que no necesita electricidad.
126. ¿Cómo mitiga el Refinamiento Multitarea.^{el} olvido catastrófico?
- a) Entrenando al modelo para que no piense.
 - b) Entrenando con ejemplos de diversas tareas a la vez, manteniendo la capacidad de generalización.
 - c) Borrando las capas antiguas.
 - d) Usando solo datos de Wikipedia.
127. ¿Cuál es el propósito de .^Amazon Sumerian” según el material de IA?
- a) Minería de datos históricos.
 - b) Producción virtual y creación de contenidos 3D.
 - c) Traducción de lenguas muertas.
 - d) Almacenamiento de logs.
128. ¿Qué es la ”Normalización de capas.” en un Transformer?
- a) Hacer que el modelo se vea igual.
 - b) Mantener las salidas de las capas dentro de un rango determinado para facilitar el entrenamiento y evitar sobreajuste.

- c) Borrar los datos corruptos.
 - d) Cambiar el idioma del sistema.
129. ¿Qué es una Conexión Residual en redes neuronales?
- a) Un cable suelto.
 - b) Atajos que permiten el flujo de información de una parte de la red a otra omitiendo operaciones intermedias.
 - c) Una conexión a internet lenta.
 - d) El fin de una transacción.
130. ¿Cuál es el objetivo final de usar MLOps en una organización?
- a) Gastar más dinero.
 - b) Automatizar y estandarizar procesos para entregar modelos de alta calidad de forma repetible y fiable.
 - c) Eliminar al departamento de TI.
 - d) Usar la tecnología más antigua.

Clave de Respuestas (Dominio 3)

1. b	31. a	61. b	91. b	121. b
2. b	32. b	62. b	92. b	122. a
3. b	33. b	63. b	93. b	123. b
4. b	34. b	64. a	94. b	124. b
5. b	35. b	65. a	95. b	125. b
6. b	36. b	66. a	96. a	126. b
7. b	37. b	67. b	97. b	127. b
8. b	38. b	68. a	98. b	128. b
9. b	39. b	69. b	99. b	129. b
10. b	40. a	70. a	100. b	130. b
11. a	41. b	71. b	101. b	131. b
12. b	42. b	72. b	102. b	132. b
13. b	43. b	73. b	103. b	133. b
14. b	44. b	74. b	104. b	134. a
15. b	45. b	75. b	105. b	135. b
16. b	46. b	76. b	106. b	136. b
17. b	47. b	77. b	107. b	137. b
18. b	48. b	78. b	108. b	138. b
19. b	49. a	79. b	109. a	139. b
20. b	50. b	80. b	110. b	140. a
21. b	51. b	81. b	111. b	141. b
22. b	52. b	82. a	112. b	142. b
23. b	53. b	83. b	113. b	143. b
24. b	54. b	84. b	114. b	144. a
25. b	55. b	85. b	115. a	145. b
26. b	56. a	86. b	116. b	146. b
27. b	57. b	87. b	117. b	147. b
28. a	58. b	88. a	118. b	148. b
29. b	59. b	89. b	119. a	149. b
30. b	60. b	90. a	120. b	150. b