

Google Cloud Certified: Generative AI Leader

Nota: Los cursos oficiales y Certificados de Google Generative AI Leader para aprobar antes del examen.

Nivel: Beginner Duración: 90 min Fuente: guía de estudio oficial de Microsoft Learn

1. Fundamentals of gen AI

30–30%

Describe core generative AI (gen AI) concepts and use cases

- La IA es el campo amplio; el machine learning aprende patrones a partir de datos y el NLP permite entender y generar lenguaje. La IA generativa usa modelos fundacionales, preentrenados con enormes volúmenes de datos y adaptables: los LLM se centran en texto, los multimodales combinan texto, imagen, audio y vídeo, y los de difusión crean imágenes eliminando ruido. La ingeniería de prompts guía la respuesta; el prompt tuning aprende un prefijo. Impago de un préstamo. El no supervisado descubre patrones en datos sin etiquetar, por ejemplo al segmentar clientes. El aprendizaje por refuerzo aprende por ensayo y error con recompensas y penalizaciones, como un robot.
- El ciclo de vida del ML va de la ingesta de datos (Pub/Sub para flujos de eventos) a su preparación (BigQuery, que aprende a agarrar objetos). El semisupervisado combina pocos datos etiquetados con muchos sin etiquetar. Dataflow), el entrenamiento (AutoML sin código o entrenamiento personalizado en Gemini Enterprise Agent Platform), el despliegue en endpoints para inferencia online o por lotes y la gestión del modelo: Model Registry.
- Para elegir un modelo fundacional se parte del caso de uso: la modalidad (texto, imagen, vídeo o audio), la ventana para las versiones, Model Monitoring para la deriva y Pipelines para automatizar el flujo.
- de contexto necesaria, la calidad, la latencia y el coste, la seguridad y el cumplimiento normativo, y la posibilidad de ajustarlo con datos propios. Un modelo más grande no siempre es mejor: Model Garden reúne modelos de Google.
- La IA generativa destaca en crear contenido (texto, imágenes, vídeo y código), resumir documentos y opiniones, de partners y abiertos para compararlos, descubrir información mediante búsqueda conversacional y automatizar tareas repetitivas. Para previsiones o clasificaciones con datos estructurados suele bastar la IA tradicional, y ambas se combinan: la tradicional segmenta

Describe how various data types are used in gen AI and the business implications

- La IA depende de la calidad de sus datos: deben estar completos, ser consistentes entre sistemas, válidos, actualizados y relevantes para las condiciones reales de uso; si no, aparecen sesgos y alucinaciones. También deben ser accesibles: los silos y la variedad de formatos frenan los proyectos, y Google recomienda una plataforma.
- Los datos estructurados siguen un esquema fijo de filas y columnas y se consultan con SQL, como inventarios o registros financieros. Los no estructurados no tienen modelo predeterminado: correos, PDF, imágenes, audio, vídeo o imágenes de satélite, y son la mayoría de los datos de una empresa. JSON y XML son semiestructurados. Google
- Los datos etiquetados asocian cada ejemplo a su respuesta correcta (una categoría, un sentimiento o un valor) y recomienda BigQuery para los estructurados y Cloud Storage para los no estructurados.
- son la base del aprendizaje y el ajuste supervisados; los no etiquetados carecen de ella y alimentan el aprendizaje no supervisado y el autosupervisado de los modelos fundacionales. Etiquetar es caro, por lo que se combinan

Identify the core layers of the gen AI landscape and the business implications

- La capa de infraestructura aporta el cómputo, el almacenamiento y la red con los que se entrenan y sirven los modelos. En Google Cloud se agrupa en AI Hypercomputer, que integra aceleradores (GPU de NVIDIA y TPU diseñadas por Google), software abierto y opciones de consumo flexibles, como Spot VMs o reservas, para
- La capa de modelos la forman los modelos fundacionales: modelos grandes preentrenados con datos masivos que maximizar el goodput y controlar el coste.
- Los LLM son un tipo de modelo fundacional; otros trabajan con imágenes, audio o vídeo. Difieren en tamaño, modalidad y coste, así que elegir modelo es un equilibrio por caso de uso, y adaptar uno
- La capa de plataformas une modelos, datos y agentes y oculta la complejidad de la infraestructura. En Google Cloud preentrenado suele ser más rápido y barato que entrenar desde cero.
- es Gemini Enterprise Agent Platform, la evolución de Vertex AI, organizada en cuatro pilares: crear (Agent Studio, Agent Development Kit, Model Garden), escalar (Agent Runtime), gobernar (Agent Registry, Agent Identity, Agent Gateway) y optimizar (evaluación, simulación y observabilidad).
- Un agente de IA persigue un objetivo con autonomía: usa un modelo como motor de razonamiento, planifica, recuerda y actúa sobre sistemas externos mediante herramientas. A diferencia de un asistente, no espera a que el usuario decida cada paso, y a diferencia de un bot, no sigue reglas fijas. Google advierte que tareas deterministas como resumir o clasificar no necesitan un flujo agéntico.
- uso diario, como Gemini en Gmail o Docs, y la aplicación Gemini Enterprise, puerta de entrada única para buscar en los datos de la empresa con permisos, chatear y usar agentes. Aporta valor rápido sin desarrollo propio y Google

Identify the use cases and strengths of Google's foundation models

- Gemini es la familia de modelos multimodales más capaces de Google, creada por Google DeepMind: entiende texto, imágenes, audio, vídeo y código en una misma petición y razona sobre ellos. Su ventana de contexto larga permite analizar grandes volúmenes de documentos o vídeos sin trocearlos, y es el motor de razonamiento de
- Gemma es una familia de modelos abiertos y ligeros creada con la misma investigación que Gemini. Al ser de asistentes y agentes como Gemini Enterprise. Se usa como servicio gestionado en Agent Platform.
- pesos abiertos, se puede ajustar con el framework que se prefiera y ejecutar en hardware propio, móviles o servicios alojados. Tiene variantes especializadas, como MedGemma (medicina), ShieldGemma 2 (seguridad) o
- Imagen es el modelo especializado de Google para generar imágenes a partir de texto. Sus puntos fuertes son la EmbeddingGemma (búsqueda), y Google recomienda Agent Platform para servirla sin equipo de MLOps.
- calidad de imagen, el fotorealismo, los estilos artísticos, la tipografía y los diseños de marca, logotipos y productos, con baja latencia; también permite editar con máscaras y ampliar la resolución. Para la edición conversacional en
- Veo es el modelo de Google DeepMind para generar vídeo a partir de texto o imágenes, con estilos visuales y varios turnos. Google recomienda Gemini, y lo generado con Imagen lleva la marca de agua SynthID, cinematográficos muy variados. Puede animar fotos existentes, crear el vídeo entre un fotograma inicial y otro final, ampliar y editar clips y generar audio nativo con diálogos y efectos. Las empresas lo usan para acelerar anuncios y contenido de marketing, y sus vídeos llevan la marca de agua SynthID.

2. Google Cloud's gen AI offerings

35–35%

Describe Google Cloud's strengths in the field of gen AI

- Google se define como una empresa «AI-first» con un enfoque integral (full-stack): infraestructura propia con TPU y GPU, investigación de Google Research y Google DeepMind, la familia de modelos Gemini, nativamente multimodal, y productos que la llevan a millones de usuarios. Esa integración, guiada por sus Principios de IA, permite trasladar
- Google Cloud preserva las plataformas de la empresa: los prompts, las respuestas y los datos de ajuste del cliente no se usan para entrenar modelos sin su permiso, ofrece indemnización de propiedad intelectual, filtros de seguridad, Model Armor y Sensitive Data Protection, el grounding reduce las alucinaciones y todo funciona
- Gemini está integrado en toda la familia de productos de Google: Google Workspace (Gmail, Docs, Meet, Vids), Gemini en BigQuery y Looker para datos, Gemini Code Assist y Gemini Cloud Assist para equipos técnicos, Gemini en Security Command Center para seguridad, y Gemini Enterprise como puerta de entrada única que conecta con datos de
- Workspace, Meritos sin Google Sales Cloud de la Red de Integración, más de 200 modelos de Google, abiertos como Gemma y de terceros como Claude, intercambiables sin rehacer la lógica de los agentes; estándares abiertos como MCP para conectar agentes con sistemas y A2A para que agentes de distintos proveedores colaboren; y tecnologías
- Abilvas computeren el sistema de supercomputación integrada de Google Cloud, combina hardware optimizado (TPU diseñadas por Google, GPU de NVIDIA, red y almacenamiento), software abierto como PyTorch y JAX y modelos de consumo flexibles, como Spot VMs o reservas, para maximizar el goodput. Las TPU son ASIC propios para cálculos matriciales a gran escala y se apoyan en centros de datos globales y eficientes.

Describe how Google Cloud's prebuilt gen AI offerings enable AI powered work

- La app de Gemini ayuda a redactar, resumir y analizar, e incluye Deep Research para informes a partir de cientos de fuentes y Canvas para crear prototipos y textos. Los Gems son versiones guardadas de Gemini con instrucciones y archivos propios que ahorran tiempo en tareas repetitivas; con archivos de Drive usan siempre la versión más
- Gemini Enterprise (antes Agentspace) une en sí todas las limitadas que respeta permisos sobre Workspace, Share-Point, Jira y otros sistemas, un asistente de IA y agentes: prediseñados, creados sin código o propios con ADK, todos bajo un gobierno central. Gemini Notebook Enterprise sirve para trabajar sobre fuentes elegidas a mano, y
- Gemini en Workspace integra la IA en cada app que ya se usaba: resume un audio y redacta en Gmail, crea borradores en Docs, analiza datos en Sheets, toma notas con decisiones y siguientes pasos en Meet y ayuda a crear vídeos con Vids. Los datos de la empresa no se usan para entrenar los modelos y Gemini solo recupera contenido al que el usuario ya tiene acceso.

Describe how Google Cloud's gen AI offerings improve the customer experience

- Agent Search (antes Vertex AI Search) permite crear búsquedas y recomendaciones con calidad Google sobre webs y datos propios: entiende la intención, da respuestas de IA con citas y sirve como RAG listo para usar, también en medios. Grounding with Google Search conecta Gemini con la web en tiempo real para información posterior al
- Customer Engagement Search (o ya Gemini Enterprise Customer Experiences) a propósito del servicio al cliente: Conversational Agents ofrece autoservicio por voz y chat con IA generativa y flujos deterministas; Agent Assist ayuda en directo a los agentes humanos con respuestas, coaching y resúmenes; Conversational Insights analiza temas, sentimiento y calidad; y Contact Center as a Service enruta las interacciones por todos los canales.

Describe how Google Cloud empowers developers to build with AI

- Gemini Enterprise Agent Platform, evolución de Vertex AI, reúne Model Garden, una biblioteca para descubrir, probar, ajustar y desplegar modelos de Google, de socios y abiertos; Agent Search (antes Vertex AI Search), búsqueda semántica con respuestas generativas citadas sobre datos propios sin gran experiencia en ML; y AutoML,
- Google Cloud ofrece RAG que define un motor de búsqueda semántica y por palabras clave listo para usar, y RAG Engine, un servicio gestionado que conecta datos privados con los LLM. Para máximo control están las API de bajo nivel: Layout Parser, embeddings, Vector Search, la API de ranking, la de generación
- Agent Platform y la de todo el grounding, que más se usa se construye con ASK (código abierto) u otros frameworks, con Agent Studio (low-code) o partiendo de Agent Garden; se escala en Agent Runtime con Sessions y Memory Bank; se gobierna con Agent Registry, Agent Identity y Agent Gateway; y se optimiza con evaluación, simulación y observabilidad.

Define the purpose and types of tooling for gen AI agents

- Las herramientas permiten al agente obtener información actualizada y actuar sobre sistemas externos. Con las funciones, el modelo indica qué llamar y con qué parámetros y la aplicación cliente ejecuta la llamada; las extensiones las ejecuta la propia plataforma; los almacenes de datos dan respuestas basadas en webs y
- Un agente puede usar APIs propias, de socios o públicas, Speech-to-Text y Text-to-Speech para voz, Translation y Document Translation (conserva el formato), Document AI para extraer campos de documentos, Vision e Video Intelligence para imágenes y vídeo, y Natural Language para sentimiento y entidades. Cloud Run functions (antes
- Google AI Studio) ejecuta prompts rápidos los modelos y Gemini by experimentales con prompts usando una cuenta de Google y una clave de API. Agent Studio, en Agent Platform, es la opción empresarial: compara prompts y modelos, diseña agentes low-code y ofrece IAM, SLA, soporte 24/7, cumplimiento como HIPAA, residencia de datos, CMEK y modelos de terceros en Model Garden.

3. Techniques to improve gen AI model output

20-20%

Describe how to proactively overcome foundation model limitations

- Los modelos fundacionales dependen de la calidad de sus datos y del prompt, solo conocen lo que había en su entrenamiento (fecha de corte) y pueden alucinar, es decir, generar contenido plausible pero falso. También
- El ground truth es la calidad del modelo con fuentes verificables (Google Search, Agent Search, RAG Engines) y la actualización de datos a los privados sin reentrenar. Google recomienda empezar por el diseño de prompts y
- Pasarlo a un fin único y fijar datos verificables y de las páginas establecidas. Es decir, cada que aumenta la precisión se debe actualizar el Gemini Evaluation Service. La verificación se vigila con Model Monitoring y Feature Store mantiene características coherentes y versionadas. El éxito se mide con KPI de calidad del modelo, del sistema y

Describe prompt engineering techniques and how they drive better results

- La ingeniería de prompts se prepara con iteraciones verificadas y para evaluar las respuestas para guiar al modelo hacia el resultado deseado sin modificarlo. Todo prompt necesita una tarea; las instrucciones del sistema, los
- Ejemplos del solo texto o instrucciones de alto nivel y precisión de control y precisión de formato, estilo o tono, ejemplos de los ejemplos prácticos se mejoran en la precisión de rol define un personaje, idealmente en las
- La tradición de pensamiento crítico al modelo que divide en partes o en temas, en que mejora la precisión de los problemas complejos o muestra cómo llegar al resultado; el razonamiento debe ir antes de la respuesta y puede separarse con un formato de salida. ReAct alterna razonar, actuar con herramientas externas y observar el resultado,

Identify grounding techniques and their use cases

- El grounding conecta a las fuentes de datos de fuentes verificables: reduce las alucinaciones, ancla las respuestas a tus datos y aporta enlaces a las fuentes. Puede usar datos propios de la empresa (documentos, bases de datos,
- RAG (Retrieval-Augmented Generation) permite el uso de tus fuentes de datos como contexto de consulta antes de la que el público, genera. Así reduce las alucinaciones y permite responder con datos privados o actualizados sin reentrenar.
- Agent Search (antes de Text-to-SQL) les que el RAG sirve para las interacciones con salidas de Google sobre datos empresariales y las APIs de RAG (como el uso de embeddings, Vector Search, ranking y check grounding) permiten
- La temperatura controla la aleatoriedad de la salida. RAG Engines que se ajustan a la temperatura de salida y la longitud de la salida y las secuencias de parada controlan la longitud, y los filtros de seguridad bloquean la salida dañina según umbrales por categoría sin cambiar el comportamiento del modelo.

4. Business strategies for a successful gen AI solution

15-15%

Describe the Google Cloud-recommended steps to successfully implement a transformational gen AI solution

- La IA generativa crea contenido nuevo: texto (marketing, resúmenes, traducción), imágenes (Imagen), vídeo (Veo), música (Lyria), código, búsqueda conversacional y contenido o recomendaciones personalizadas según el perfil
- Google Cloud recomienda una estrategia de implementación de IA generativa que se centra en la privacidad, la seguridad, la fiabilidad, la tolerancia a fallos y la gobernanza de los datos. La IA generativa puede proporcionar una experiencia de usuario única y personalizada, pero también puede generar riesgos de seguridad y privacidad. Google Cloud ofrece herramientas y servicios para ayudar a las empresas a implementar IA generativa de forma segura y responsable.
- Google Cloud ofrece una amplia gama de servicios de IA generativa, desde modelos de lenguaje preentrenados hasta plataformas de desarrollo de aplicaciones. Gemini Enterprise Agent Platform (antes Vertex AI) sirve para
- Google Cloud ofrece una amplia gama de servicios de IA generativa, desde modelos de lenguaje preentrenados hasta plataformas de desarrollo de aplicaciones. Gemini Enterprise Agent Platform (antes Vertex AI) sirve para
- Google Cloud ofrece una amplia gama de servicios de IA generativa, desde modelos de lenguaje preentrenados hasta plataformas de desarrollo de aplicaciones. Gemini Enterprise Agent Platform (antes Vertex AI) sirve para
- Google Cloud ofrece una amplia gama de servicios de IA generativa, desde modelos de lenguaje preentrenados hasta plataformas de desarrollo de aplicaciones. Gemini Enterprise Agent Platform (antes Vertex AI) sirve para

Define secure AI and its importance in protecting AI systems from malicious attacks and misuse

- La seguridad de la IA que integra desde el diseño y acompañar al desarrollo, el despliegue y la operación. Cada fase tiene sus riesgos: el envenenamiento de datos se mitiga pronto con saneamiento y controles de integridad; la
- Se debe ir en prompts de Google para proteger los sistemas de IA. La IA generativa puede proporcionar una experiencia de usuario única y personalizada, pero también puede generar riesgos de seguridad y privacidad. Google Cloud ofrece herramientas y servicios para ayudar a las empresas a implementar IA generativa de forma segura y responsable.
- La IA generativa puede proporcionar una experiencia de usuario única y personalizada, pero también puede generar riesgos de seguridad y privacidad. Google Cloud ofrece herramientas y servicios para ayudar a las empresas a implementar IA generativa de forma segura y responsable.

Describe the importance of responsible AI in business

- La IA responsable es clave para que el negocio y la confianza frene la adopción y muchas empresas abandonan proyectos por motivos éticos. Google recomienda revisar cada caso en su contexto desde el diseño, con equipos
- Que los sistemas sean transparentes y se identifiquen los datos (como el postulado) como un medio para identificar las personas de la desanonimización sustituye identificadores por tokens que mantienen los cruces y, con la clave,
- La IA responsable es clave para que el negocio y la confianza frene la adopción y muchas empresas abandonan proyectos por motivos éticos. Google recomienda revisar cada caso en su contexto desde el diseño, con equipos
- Que los sistemas sean transparentes y se identifiquen los datos (como el postulado) como un medio para identificar las personas de la desanonimización sustituye identificadores por tokens que mantienen los cruces y, con la clave,
- La IA responsable es clave para que el negocio y la confianza frene la adopción y muchas empresas abandonan proyectos por motivos éticos. Google recomienda revisar cada caso en su contexto desde el diseño, con equipos