Herramientas para la Computación en la Nube Dirigida a Inteligencia Artificial

Tema 1. Introducción, contexto y características de proveedores y servicios de IA en la nube

Índice

1.1. Introducción y objetivos

La **inteligencia artificial** (IA) y la **computación en la nube** constituyen dos de las principales tendencias *(megatrends)* del sector tecnológico en la actualidad. Estas dos disciplinas están estrechamente relacionadas. Por un lado, los proveedores de computación en la nube han incorporado servicios específicos de IA en sus porfolios. Por otro, la elevada demanda de recursos de computación, asociada a los procesos de IA, hace inviable para muchas organizaciones y particulares realizar la inversión en infraestructura adecuada. Por ello, el **pago por uso** que ofrecen los proveedores de nube se convierte en una alternativa muy atractiva.

Sin embargo, el mercado de IA en la nube es vasto y complejo, por lo que resulta de vital importancia adquirir los conocimientos necesarios para determinar en qué casos estos servicios son en efecto una opción recomendable y, en ese caso, qué criterios se deben contemplar para elegir al **proveedor idóneo.**

En las próximas secciones, se describe el estado del **mercado de servicios de IA** en la nube, cuáles son los principales proveedores que ofrecen estos servicios, de qué capacidades disponen y en qué situaciones se recomienda contratarlos como alternativa a la construcción de capacidades propias.

Finalmente, se tratan las cuestiones esenciales de **privacidad y seguridad** para proteger los datos y los algoritmos generados, y garantizar que estos no incumplen las regulaciones existentes en la materia, u otras específicas, que están apareciendo en torno al uso responsable de la IA.

Ideas clave

Este planteamiento permite al alumno alcanzar los siguientes objetivos:

- Comprender el mercado de IA en la nube, sus principales *players* y las capacidades de IA que ofrecen.
- Reconocer en qué casos debe utilizarse la nube como alternativa a una solución *on premise*.
- Diseñar estrategias que optimicen el coste de desarrollo de una aplicación con capacidades de IA, mediante la integración de servicios disponibles en la nube.
- Ser capaces de asesorar a un cliente que desee incorporar un programa de IA basada en servicios de computación en la nube.

1.2. El mercado de servicios de inteligencia artificial en la nube

Definición y características

El mercado de servicios de IA en la nube se refiere a todos aquellas capacidades que ofrecen los proveedores de computación en la nube para posibilitar la **adopción de esta tecnología** bajo un enfoque flexible, ágil y escalable.

El enfoque es **flexible** porque estos servicios se prestan en modalidad de **pago por uso,** es decir, su coste depende del grado de utilización, de los niveles de servicio garantizados y de las funcionalidades disponibles. En otros casos, se lleva a cabo una provisión previa de instancias que proporcionan una determinada capacidad de computación. Esta capacidad puede ajustarse de forma dinámica para adaptarla a las necesidades reales de la organización en cada momento. Con este planteamiento, las organizaciones que desean incorporar capacidades de IA evitan costosas inversiones en infraestructura, tiempo y personal especializado.

El enfoque es **ágil** porque los tiempos de provisión de las capacidades son muy reducidos, en ocasiones de minutos. El usuario dispone de todas las funcionalidades para generar y productizar sus modelos, o consumir directamente los modelos preentrenados para diferentes propósitos.

Finalmente, el enfoque es **escalable** porque los proveedores de nube disponen de mucha más capacidad de computación de la que puede necesitar cualquier organización, por lo que, si la adopción de IA en una organización crece exponencialmente, estos proveedores pueden asignar tantos recursos como sean necesarios para satisfacer la demanda.

Beneficios obtenidos

Las características de los servicios de IA en la nube descritas en la sección anterior propician un considerable número de beneficios para las organizaciones que desean incorporar estas capacidades en sus procesos de negocio y aplicaciones. A continuación, se enumeran los principales beneficios:

- **Reducción del plazo** requerido para la adopción versus elevados plazos de provisión y puesta en marcha asociados a cualquier tecnología, pero especialmente en una tan compleja como la IA.
- **Control de costes:** costes vinculados al consumo y la capacidad requerida en cada momento versus la necesidad de inversiones *up front (a priori)* y el coste/esfuerzo necesario para mantener la infraestructura desplegada.
- **Eficiencia en costes:** el coste responde a la necesidad real versus infraestructura sobredimensionada para absorber los picos de actividad, pero infrautilizada la mayor parte del tiempo restante.
- **Reutilización:** consumo de modelos preentrenados por el proveedor para resolver problemáticas conocidas y comunes a múltiples organizaciones versus la dificultad de entrenar modelos desde cero.
- **Costes de personal:** el consumo de servicios en la nube reduce las necesidades de personal especializado, especialmente en el caso de AutoML (IA automatizada o asistida sin necesidad de utilizar lenguajes de programación).
- ***Software* como servicio:** los proveedores en la nube proporcionan plataformas integradas para la creación y gestión de modelos frente a la necesidad de instalar, configurar y mantener *software* que realice estas funciones en infraestructura propia.
- Disponibilidad de los **avances más recientes** en el campo de la IA sin necesidad de realizar inversiones adicionales.

Grado de adopción actual

La popularización de la IA en la nube augura una cifra de negocio de $134 800 M en 2025, lo que supone un crecimiento de más del doble respecto al mercado del *software* en general. Por ello, cada uno de los principales proveedores de computación en la nube ha respondido a esta tendencia con su propia plataforma de IA y una colección de **servicios asociados.** En particular:

- Microsoft con Azure AI.
- Amazon con AWS SageMaker.
- Google con Vertex AI.

Más allá de la evolución futura que anticipan los analistas, la realidad es que la IA en la nube ya ha experimentado un crecimiento explosivo y un ritmo de adopción elevado. En concreto, en 2023, el 70 % de los entornos de computación en la nube contratados por organizaciones incluyeron servicios de IA, una cifra comparable a la de otros servicios mucho más consolidados y maduros, como la arquitectura de aplicaciones Kubernettes.

En el caso de la **IA generativa,** y en particular de los servicios ofrecidos por **Azure** en este ámbito, el consumo de los servicios ha crecido un 228 % entre junio y octubre de 2023 frente al 13 % de crecimiento experimentado por el resto de los servicios de IA de Azure y el 45 % de los servicios de Google Vertex AI. Estos datos confirman que, en gran medida, la citada explosión en el uso de la IA en la nube está estrechamente **ligada a la IA generativa.** Por tanto, hoy en día no puede afirmarse que exista una auténtica adopción generalizada y homogénea de todas las capacidades que la IA ofrece en la nube.

Asimismo, muchas de las organizaciones que han adoptado la IA en la nube se encuentran aún en fase de **experimentación o pruebas,** como revela el hecho de que un 32 % de los entornos tienen menos de 10 instancias de IA desplegadas. En torno a un 28 % han evolucionado a una etapa más consolidada, aunque lejos de incorporar la IA como un sistema productivo más, y un 10 % pueden considerarse gestionados por usuarios experimentados y con usos más ligados a necesidades productivas. En los próximos años comprobaremos si las organizaciones superan esta fase inicial para hacer realidad los objetivos de transformación que propicia la IA.

El proveedor de IA en la nube que **mayor penetración de mercado** ha experimentado es **Microsoft:** el 54 % de los entornos de Azure incluyen servicios de IA generativa (OpenAI). Le sigue Amazon Web Services (AWS), con un 53 % de entornos con servicios de SageMaker, y Google Vertex AI, con un 44 %. De estos datos se deduce que, si se considera el espectro más amplio de servicios de IA en la nube, y no específicamente las variantes generativas, la mayor cuota corresponde a Amazon.

Por otro lado, muchas organizaciones (69 %) utilizan entornos de nube para hospedar ***software* propio** relacionado con IA, en lugar de contratar los servicios de IA de sus proveedores. Dicho de otro modo, emplean infraestructura en la nube para crear capacidades propias de IA. Se trata de una opción intermedia que no explota todas las ventajas de la IA en la nube, pero que, a tenor de los datos, resulta atractiva para un número considerable de empresas.

En estos casos, los principales paquetes de *software* de IA instalados en la nube son:

- **Hugging Face Transformers:** permite descargar modelos preentrenados o entrenarlos para un propósito más específico. Proporciona interoperabilidad entre PyTorch, TensorFlow y JAX. Los modelos cubren escenarios habituales, como procesamiento de lenguaje natural, respuesta a preguntas, visión artificial, reconocimiento y transcripción de audio, extracción de información de documentos, clasificación de vídeos y OCR.

- **LangChain:** permite a los desarrolladores integrar modelos *large language models* (LLM) en sus aplicaciones, disponibilizar APIs en la aplicación para facilitar el consumo de estos modelos y monitorizar de forma continua la calidad de los resultados para mejorarlos a lo largo del tiempo.

- **Tensorflow Hub:** es un repositorio de modelos preentrenados con el popular *framework* de IA TensorFlow, que pueden ser reutilizados para adaptarlos a escenarios concretos y desplegarlos en cualquier entorno. Se consumen instanciando módulos, principalmente a través de Python. También es posible publicar modelos propios para que puedan ser reutilizados por otros usuarios.

Finalmente, otras organizaciones (42 %) emplean sus entornos de nube únicamente para **desplegar modelos** entrenados con infraestructura propia o descargados de algún repositorio de modelos. Dicho de otro modo, prescinden de las plataformas de IA en la nube para generar modelos, incorporando modelos propios, explotando únicamente las funcionalidades relacionadas con *model serving* (operacionalización). En estos casos, el modelo externo más popular que se despliega en entornos de nube es **BERT,** un modelo de clasificación de textos que permite identificar categorías o sentimientos en el contenido.

Por último, es importante destacar que esta rápida adopción no se ha visto acompañada de las correspondientes inversiones en **gestión y gobernanza de la**

IA. Esta carencia propicia escenarios de riesgo, ineficiencias, falta de madurez y ausencia de control sobre el uso de estas capacidades.

- **Escenarios de riesgo:** la falta de control sobre los procesos de toma de decisiones de la IA puede ocasionar graves perjuicios a una organización cuando esta adopte decisiones erróneas.
- **Ineficiencias:** múltiples departamentos pueden estar contratando de forma independiente las mismas capacidades de IA a los mismos proveedores, multiplicando la cifra de inversión realmente necesaria para la adopción.
- **Falta de madurez:** muchos de los modelos generados no superan la fase de experimentación por la ausencia de experiencia real en la operacionalización de estos modelos en ambientes productivos.
- **Ausencia de control:** la inexistencia de un gobierno eficaz de IA abre las puertas al incumplimiento de regulaciones relacionadas con la privacidad, el tratamiento de datos o las prácticas anticompetitivas, así como a un mayor impacto de incidentes de ciberseguridad.

Si bien esta situación es habitual en cualquier tecnología de rápida penetración, las organizaciones deben hacerle frente en los próximos años para no deteriorar su postura de seguridad, preservar el grado de excelencia en el control interno y asegurar el cumplimiento de las normativas y regulaciones aplicables.

1.3. Porfolio de capacidades de IA disponibles en la nube

En este apartado, se describen los servicios de IA más frecuentes que proporcionan los proveedores de computación en la nube.

AutoML

Las capacidades de AutoML hacen referencia a la posibilidad de diseñar modelos personalizados sin necesidad de utilizar un lenguaje de programación o disponer de los conocimientos especializados en IA. Es generalmente usado por desarrolladores para **incorporar capacidades de IA** de una forma ágil a sus aplicaciones.

Estos modelos pueden utilizarse de forma independiente para resolver una problemática de negocio concreta o combinarse con otros servicios del proveedor de nube para completar aspectos no contemplados por estos.

Las **funcionalidades clásicas** de un servicio AutoML son las siguientes:

Preparación automática de datos

Incluye el **manejo de datos** ausentes, la limpieza de datos para descartar los incorrectos o incompletos, el enriquecimiento de datos con información de otras fuentes y la adaptación a formatos adecuados para la posterior construcción de modelos.

En AutoML, estas actividades se realizan de **forma automática** o muy **dirigida.** Generalmente, se complementan con funcionalidades de visualización de datos para comprender la naturaleza de estos y poder evaluar el resultado de los procesos de preparación aplicados.

Ingeniería de variables

Incluye la **clasificación automática** de las variables existentes y la detección de su grado de importancia, así como la generación de nuevas variables —en forma de metadatos— que podrían resultar eficaces durante el proceso posterior de construcción de modelos.

Generación automática de modelos

Incluye la capacidad de generar modelos de forma automática a partir de los datos y unos parámetros de configuración. Se deben especificar las variables predictoras y aquello que se quiere predecir. El propio servicio analiza los datos proporcionados y selecciona el modelo o modelos idóneos para generar la capacidad predictiva. También mide el rendimiento e identifican valores óptimos de los hiperparámetros.

En función del grado de sofisticación, el servicio puede combinar varios modelos en uno con la finalidad de proporcionar un rendimiento conjunto superior al de los modelos individuales.

Productización de modelos

Incluye la **generación de un *workflow*** que integre todas las fases del *pipeline* analítico convencional: preparación de datos, entrenamiento del modelo, provisión de infraestructura para desplegarlo, puesta en producción, creación de APIs para procesar peticiones, medición del rendimiento en producción —que incluye la detección de *concept drift*— y gestión integral del ciclo de vida de los modelos. El servicio también puede **reemplazar modelos obsoletos** por otros que proporcionen mejores predicciones, en el contexto de un proceso de mejora continua que se ejecuta forma automática. Y, por último, este también podría evaluar el impacto de la **calidad del dato** en la precisión de dichas predicciones.

IA responsable

Permite la **detección de sesgos en los datos** *(biases)* que pueden deteriorar la calidad de las predicciones o afectar o discriminar a un colectivo de clientes. También se detecta el uso de datos no autorizados por las regulaciones de IA para la creación de algoritmos. Para ello, se incluye la posibilidad de interpretar las predicciones, es decir, entender cómo ha llegado el modelo a la conclusión a partir de los datos de entrada.

En la Figura 1, se muestra de forma esquemática el proceso de AutoML en la plataforma Azure de Microsoft.

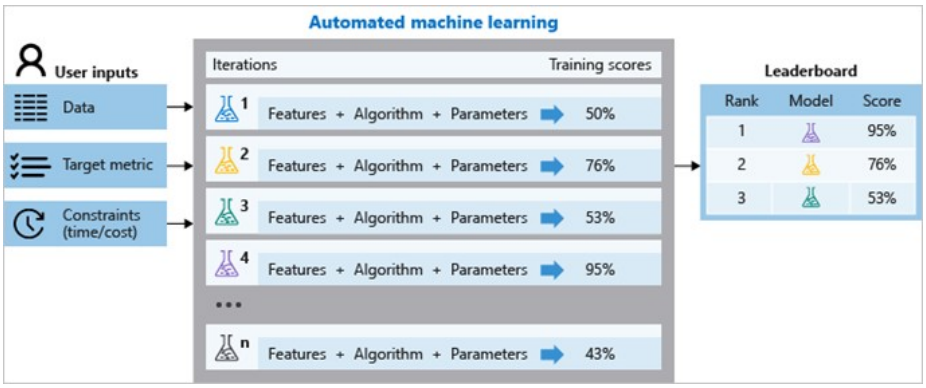

Figura 1. Proceso de AutoML en Azure Machine Learning. Fuente: Manashgoswami, 2023.

Servicios para usuarios especializados

Estos servicios se dirigen a expertos en el ámbito de la ciencia de datos y la inteligencia artificial que desean construir sus **propios modelos,** en lugar de utilizar las facilidades de AutoML o consumir los modelos preentrenados del proveedor.

Estos usuarios demandan funcionalidades que permitan gestionar el **ciclo de vida de los modelos,** que incluyen actividades similares a las descritas para el caso de AutoML: preparación de datos, ingeniería de variables, generación de modelos, selección y optimización de modelos (también denominado «gestión de experimentos»), productización (u operacionalización) de modelos y monitorización continua del rendimiento de los modelos en producción.

Para cubrir este amplio espectro de actividades, los proveedores de nube disponen de plataformas que integran todas las herramientas necesarias. Muchos de los componentes utilizados para implementar estas herramientas son **propietarios** y, por tanto, diferentes en cada proveedor; sin embargo, otros son de naturaleza ***open-source*** y están disponibles en todos los proveedores.

El ejemplo más claro de componente *open-source* son los **Jupyter Notebooks,** que se utilizan para cubrir las etapas correspondientes a la construcción de los modelos. Esta herramienta permite a un usuario experimentado escribir código fuente en varios lenguajes de programación (principalmente Python) para implementar un *pipeline* analítico a través del que se entrenan y prueban modelos con diferentes algoritmos.

La principal ventaja del formato de *notebook* es que permite ejecutar paso a paso dicho *pipeline* y, por tanto, combinar en un único documento el código fuente y los resultados obtenidos. Esta característica fomenta la colaboración entre especialistas y la creación de una importante **base de conocimiento.** A partir de un *notebook* es sencillo extraer y reutilizar código de otros como punto de partida para resolver una problemática propia.

En la Figura 2, se muestra un ejemplo de entorno Jupyter en AWS SageMaker.

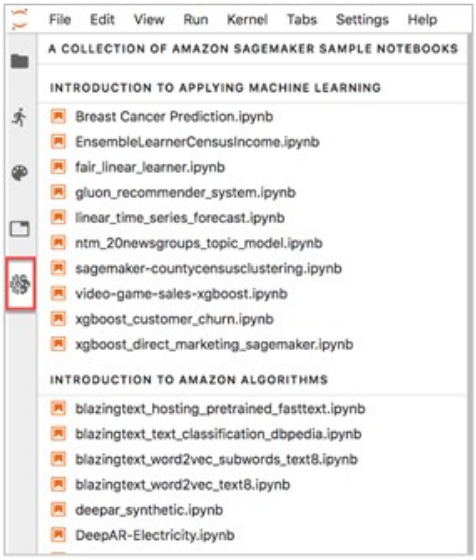

Figura 2. Colección de Jupyter Notebooks de ejemplo en AWS SageMaker. Fuente: Amazon SageMaker, s. f.

Procesamiento del lenguaje

Los servicios de lenguaje se basan en modelos de IA en la nube diseñados para **procesar mensajes en lenguaje natural** (audio o texto) con un fin determinado.

Los **fines** más habituales en los que se especializan estos modelos son los siguientes:

- **Conversión de voz a texto:** permite transformar un audio en texto. De esta forma, se pueden utilizar a continuación otros modelos para comprender el contenido del texto.
- **Extracción de tókenes:** consiste en la detección de palabras, expresiones o símbolos clave que permiten esclarecer la intención del mensaje.
- ***Part of speech:*** permite identificar la categoría gramatical de las palabras que forman parte del mensaje para aumentar la comprensión de este.

- **Traducción:** realiza la conversión de un texto de un idioma a otro. La IA es especialmente eficaz en esta tarea, porque puede aprender los patrones estructurales de cada idioma para generar traducciones de mucha calidad.
- **Análisis de sentimientos:** se refiere a la detección del tono de un mensaje, que puede ser positivo, negativo o neutro.
- **Análisis de texto o documentos:** permite extraer metadatos en forma de conceptos clave o temáticas, y también resumir textos a partir de ellos.

Visión artificial

Los servicios de visión artificial en la nube permiten identificar **objetos, personas, acciones, anomalías, colores** y otros patrones similares en una imagen. El resultado del proceso de detección se transforma en metadatos, que pueden emplearse para **clasificar o agrupar conjuntos** de imágenes o vídeos.

La visión artificial también incluye el **procesamiento de secuencias de vídeo,** con las mismas capacidades que en el caso de las imágenes. En algunos casos, estos modelos permiten seguir la trayectoria de objetos que aparecen en el vídeo o transcribir el audio del vídeo a texto. Asimismo, es posible identificar texto manuscrito o escrito a máquina en imágenes o vídeos y extraerlos para su posterior procesamiento y clasificación.

Por último, en esta categoría se incluyen los servicios de IA generativa especializados en la **creación de imágenes** a partir de una definición en texto *(prompt)* o a partir de imágenes similares. Estas imágenes pueden a su vez emplearse como datos de entrada etiquetados para el entrenamiento de los modelos que extraen información de imágenes.

Otros servicios

En esta categoría se incluyen aquellos servicios de IA de menor peso o relevancia en el porfolio de los proveedores de nube, pero que resuelven **problemáticas de negocio reales** de forma eficaz, por lo que pueden resultar útiles a las organizaciones.

- **Detección de anomalías:** se trata de servicios que analizan un flujo de eventos o datos para identificar anomalías en ellos. Los modelos utilizados sintetizan los patrones habituales para ser capaces de detectar circunstancias excepcionales que no se ajustan a dichos patrones. Se utilizan en el ámbito de la prevención de fraude, la monitorización de sistemas y la ciberseguridad.

- **Sistemas de recomendación:** se trata de servicios que proporcionan recomendaciones personalizadas de contenido o productos. Estos modelos capturan las preferencias y patrones de comportamiento de los usuarios para ofrecerles otros contenidos o productos que puedan interesarles.

1.4. Principales proveedores de servicios de inteligencia artificial en la nube

El mercado de IA en la nube ha atraído a las principales empresas especializadas en computación en la nube. En los últimos años, estas organizaciones han construido y disponibilizado sus respectivas plataformas de IA. También han incorporado progresivamente **nuevas funcionalidades,** a un ritmo vertiginoso, para adaptarse a las preferencias de los usuarios y a los avances registrados en el ámbito de la IA.

En algunos casos, estos proveedores han modernizado e integrado sus plataformas para atraer a más usuarios (por ejemplo, Google con Vertex AI) o han firmado acuerdos estratégicos con compañías punteras de IA para canalizar sus innovaciones a través de servicios en la nube (por ejemplo, Amazon con Anthropic o Azure con OpenAI).

Aunque el número de proveedores de nube que ofrecen servicios de IA es elevado, en este documento se han considerado únicamente los más utilizados, o bien aquellos que presentan alguna característica diferencial. En todos los casos se describen las capacidades ofrecidas en torno a cuatro grandes bloques que agrupan los servicios descritos en la sección previa.

- Procesamiento de lenguaje natural.
- Visión artificial.
- AutoML.
- Servicios para usuarios especializados.

Alibaba Cloud

Alibaba, grupo propietario de la popular plataforma de comercio electrónico **AliExpress,** ofrece a través de su plataforma de computación en la nube servicios de IA en el ámbito de procesamiento del lenguaje, visión artificial y AutoML.

Se comercializan **modelos preentrenados** para resolver problemática de diferentes sectores (banca, salud, administración pública, entre otros). Estos modelos pueden integrarse en aplicaciones para que dispongan de capacidades avanzadas de IA.

El punto fuerte de Alibaba Cloud en IA es el **grado de personalización** de sus servicios: es posible combinar los modelos preentrenados con herramientas de bajo nivel para usuarios especializados, con el fin de adaptarse a cualquier escenario de negocio. Por otro lado, es el proveedor que ofrece más idiomas en sus servicios de IA para traducción de textos.

Amazon Web Services

Amazon ofrece a través de su plataforma de **IA SageMaker** la posibilidad de automatizar todo el ciclo de vida de producción de modelos en IA y *machine learning.*

Esta capacidad permite a las organizaciones trasladar el resultado de los desarrollos de IA al ámbito de las operaciones de una manera ágil. Es decir, permite productizar los modelos en plazos más cortos y con menos recursos involucrados.

Google

Las capacidades de IA de **Google Cloud Platform** incluyen procesamiento de lenguaje, visión artificial, procesamiento de datos estructurados, procesamiento de documentos, servicios basados en IA para atención al cliente y AutoML, junto con una colección de herramientas de bajo nivel, como TensorFlow, para los usuarios que disponen de un perfil especializado en ciencia de datos e IA. Adicionalmente, Google es líder en investigación de IA y en su uso responsable.

Microsoft

La plataforma **Azure AI** de Microsoft incluye servicios de lenguaje, visión y AutoML. Su principal ventaja es un modelo escalonado de precios que permite iniciar la experiencia con una inversión relativamente baja, pudiendo crecer progresivamente a medida que las capacidades de IA brindan los resultados esperados. Esta estrategia le ha permitido liderar el mercado.

H20.ai

A diferencia de los anteriores, H20 es un proveedor de nicho que solo ofrece servicios de IA en la nube. Permite el desarrollo, implementación y gestión de capacidades de IA en la nube o en infraestructura del cliente *(on premise),* y también en entornos híbridos (en los que una parte de la infraestructura está en la nube y otra en el centro de proceso de datos del cliente). Asimismo, está fuertemente especializado en AutoML para procesamiento de datos estructurados, series temporales, imágenes, vídeo, audio, texto y documentos.

IBM

Los servicios de IA en la nube de IBM pertenecen a **tres categorías:** procesamiento de lenguaje para responder a preguntas y generar narrativas (chatbots); clasificación y detección de objetos, acciones y anomalías en secuencias de vídeo; y AutoML. Para los usuarios más avanzados, incluye funcionalidades de calidad del dato, linaje y supervisión de modelos, detección de sesgos y explicabilidad de dichos modelos.

Oracle

La plataforma de nube **Oracle Cloud** ofrece capacidades de IA para procesamiento de voz, visión, documentos y traducción y servicios de AutoML que, como en el caso de H20.ai, pueden desplegarse en la nube, en el centro de datos del cliente y de forma híbrida entre ambos.

Por otro lado, las características diferenciales de los tres principales proveedores (Google, Microsoft, Amazon) son las siguientes:

- **Servicio gestionado:** el proveedor gestiona internamente la provisión, escalado y mantenimiento de la infraestructura necesaria para proveer las funcionalidades de creación y gestión de modelos.
- **Proceso extremo a extremo:** cubre todas las etapas del ciclo de vida de la IA con herramientas especializadas e integradas para cada una de ellas.
- **Escalabilidad y flexibilidad:** permite incrementar o reducir la potencia de los recursos que prestan los servicios de IA para adaptarse dinámicamente a las cargas de trabajo de IA.
- **Extensa biblioteca de algoritmos:** estos algoritmos agilizan la tarea de prototipado y experimentación de los usuarios avanzados.
- **Soporte para lacreación de algoritmos propios** con plataforma *open-source,* como TensorFlow o PyTorch.
- Capacidades de **AutoML** para todas las etapas del proceso.
- **Monitorización y gestión de trazas *(logs)*** de los modelos desplegados en producción para detectar incidentes o degradación del rendimiento.
- **Integración con el ecosistema** de servicios de computación en la nube del proveedor, lo que permite construir soluciones de negocio completas con un único proveedor.

Inteligencia generativa

En el ámbito específico de IA generativa, se incluye en la Tabla 1 un resumen esquemático de las capacidades de los tres principales proveedores.

GenAI Category	GenAI Component	Amazon Web Services	Google Cloud	Microsoft Azure
Foundation Models	Runtime	Amazon Bedrock	Vertex AI	Azure OpenAI
	Text / Chat	TBD	PaLM	GPT
	Code	TBD	Codey	GPT
	Image Generation	TBD	Imagen	DALL-E
	Translation	TBD	Chirp	None
Model Catalog	Commercial	Amazon SageMaker JumpStart Amazon Titan	Vertex AI Model Garden	Azure ML Foundation Models
	Open Source	Amazon SageMaker JumpStart Hugging Face	Vertex AI Model Garden	Azure ML Hugging Face
Vector Database		Amazon RDS (pgvector)	Cloud SQL (pgvector)	Azure Cosmos DB Azure Cache
Model Deployment & Inference		Amazon SageMaker	Vertex AI	Azure ML
Fine-tuning		Amazon Bedrock	Vertex AI	Azure OpenAI
Low-code/No-code Development		TBD	Gen App Builder	Power Apps
Code Completion		Amazon Code Whisperer	Duet AI for Google Cloud	GitHub Copilot

Tabla 1. Capacidades de IA generativa de los principales proveedores de computación en la nube. Fuente: Msv, 2023.

La información de la Tabla 1 se estructura de la siguiente forma:

- **Modelos principales:** modelos utilizados como motor para la prestación de las capacidades de IA generativa. Se subdividen en *runtime* (plataforma común), chatbots, código fuente, generación de imágenes y traducción.
- **Catálogo de modelos:** oferta de modelos preentrenados, diferenciando aquellos de naturaleza *open-source* de los propietarios.
- **Base de datos de vectores o *embeddings:*** la IA generativa transforma palabras en representaciones numéricas que se almacenan en una base de datos. En este punto, se indica qué base de datos utiliza cada proveedor.

- **Despliegue de modelos e inferencia:** se refiere al entorno utilizado para poner en producción los modelos, de manera que puedan atender peticiones de usuarios o aplicaciones.
- **Ajuste de parámetros:** se refiere al proceso de optimización de los parámetros de los modelos para maximizar su rendimiento en las labores de generación de texto, imágenes, vídeo o traducciones.
- **Integración de IA con *low-code/no-code:*** se refiere a los mecanismos que permiten integrar los modelos en funcionalidades de aplicaciones.
- **Asistente de código para desarrolladores:** se refiere a las capacidades para completar código fuente de forma automática a partir de un fragmento o crearlo a partir de una definición del propósito en lenguaje natural.

Como se puede observar, en muchos de los aspectos, Amazon se encontraba un paso por detrás del resto de proveedores en 2023. Sin embargo, el anuncio reciente de su plataforma **BedRock** sugiere que en la actualidad dispone de una oferta comparable a la de sus competidores.

1.5. Ventajas e inconvenientes de la nube frente al desarrollo de capacidades on premise

Una de las principales decisiones que las organizaciones deben tomar en su proceso de adopción de la IA es el **entorno** en el que desean tener disponibles estas capacidades. En la actualidad, existen tres alternativas: en la nube, en su centro de datos o en un entorno híbrido, es decir, con algunos componentes alojados en la nube y otros en su centro de datos.

Para tomar la decisión adecuada en cada caso, es importante conocer las **ventajas y desventajas** de las dos primeras opciones, así como identificar los **escenarios** en los que tiene sentido plantear un entorno híbrido.

En la Tabla 2, se identifican las características deseables y el grado en que uno u otro entorno las satisface.

Característica	Entorno de nube	Entorno *on premise*
Entorno de trabajo integrado	Disponible desde la contratación del servicio.	No disponible, debe construirse *from scratch* con procesos de instalación, configuración y mantenimiento de varios paquetes de *software*.
Personalización del entorno de trabajo	Limitado por el soporte del proveedor.	Posible, aunque requiere esfuerzo adicional para el *setup* del entorno.
Disponibilidad de modelos preentrenados para casuísticas habituales	Disponible y actualizado periódicamente con las últimas innovaciones.	No disponible. Se deben entrenar modelos desde cero o incorporar desde repositorios externos de modelos.
Personalización de modelos para resolver problema especializados	Limitado a determinadas categorías de modelos (por ejemplo, IA generativa).	Viable para los modelos construidos *in-house*.
Necesidades de personal especializado en IA y ciencia de datos	Menores, gracias al AutoML y la integración *out-of-the-box* de herramientas	Elevadas, dado que se requiere un especialista diferente para cada actividad del proceso y la necesidad de implementar *from scratch* cada integración necesaria.
Facilidad de integración de capacidades de IA en aplicaciones	Elevada, puesto que los proveedores integran la IA dentro de sus ecosistemas de desarrollo de aplicaciones.	Dependiente de las características de entorno, pero en general reducida puesto que requiere diseños *ad hoc*.

Ideas clave

Característica	Entorno de nube	Entorno *on premise*
Escalado de capacidades para adaptarse a necesidades crecientes del negocio	Rápida y adaptada al crecimiento deseado.	Lenta y costosa, porque implica adquirir, provisionar y configurar nueva infraestructura.
Coste de la computación masiva y continua para procesos de IA	Muy elevado.	Menor que en el caso de la nube, aunque requiere realizan inversiones *up front* para disponer de la capacidad antes de poder usarla.
Cumplimiento de regulaciones y estándares de calidad	Los proveedores ofrecen un elevado nivel de cumplimiento y robustos sistemas de controles para satisfacer estándares de calidad habituales de la industria.	Requiere un esfuerzo adicional de inversión y mantenimiento.
Cumplimiento de regulaciones locales	Puede ser limitado e insuficiente. Varios proveedores tienen restricciones regulatorias en geografías como Europa.	Viable, siempre que se adopten las medidas correspondientes.
Protección de datos y privacidad	Reducida. En función de los datos procesados, las normativas pueden prohibir el tratamiento de estos datos por parte de proveedores de nube.	Viable desde el punto de vista legal, aunque requiere inversión para adoptar las medidas necesarias.

Característica	Entorno de nube	Entorno *on premise*
Coste de las transferencias de datos (relacionado con *data gravity*)	Puede ser muy elevado por los volúmenes que se demandan en IA y por la necesidad de enviar dichos datos a otra geografía en la que el proveedor aloja sus servicios.	Despreciable, puesto que no es necesario transferir los datos a otra ubicación para su tratamiento.
Nivel de protección frente a incidentes de seguridad	Los proveedores de nube se caracterizan por tener sistemas robustos de ciberseguridad y un amplio catálogo de controles para reducir el impacto de un incidente.	Requiere de un esfuerzo considerable, si bien se reduce cuando la organización ya dispone de estándares de seguridad elevados.
Fiabilidad del entorno	Muy elevada por los agresivos niveles de servicio (SLAs) a los que se comprometen los proveedores, y por sus sistemas de redundancia.	Dependiente del grado de madurez de la organización en IT y del apetito de inversión para disponer de sistemas de redundancia.
Predictibilidad del rendimiento	Limitada. La infraestructura se comparte con otros clientes, lo que puede causar lentitud en momentos de pico de actividad.	Completa, dado que la infraestructura no es compartida con otras organizaciones.

Tabla 2. Comparativa de características entre entorno de nube y *on premise.* Fuente: elaboración propia.

De la Tabla 2, se derivan las siguientes conclusiones:

- Un proveedor de nube es siempre la mejor opción en la fase de **experimentación o pruebas** cuando el objetivo es comprobar si la adopción de la IA en la organización proporciona los beneficios esperados.

- Una vez que se supera la fase de experimentación, **la opción de nube puede resultar muy costosa** en relación con la inversión que requiere un despliegue propio, debido al incremento notable de los costes cuando se alcanzan determinados umbrales de requisitos de procesamiento. Si la infraestructura de IA va a utilizarse de forma intensiva, la opción *on premise* conduce a una importante optimización de costes.

- Si el **grado de personalización** que se requiere es elevado, la opción *on premise* es el camino por seguir. Sin embargo, si la organización puede materializar los beneficios de la IA, explotando directamente los servicios ofrecidos por el proveedor de nube, consumiendo los modelos preentrenados por este o usando sus plataformas y herramientas, el entorno de nube es la mejor alternativa.

- Si los datos que van a procesarse con algoritmos de IA son sensibles, su **protección legal** puede ser limitada en entornos de nube. Asimismo, cada vez aparecen más restricciones al tratamiento de datos fuera de la geografía en la que opera la organización, lo que puede hacer completamente inviable un entorno de nube para disponer de capacidades de IA. En general, es importante considerar el concepto de ***data gravity*** en la decisión: los datos «atraen» hacia ellos mismos a las aplicaciones y servicios que los tratan.

- Si el **presupuesto disponible** para la adopción de IA en fases tempranas es limitado, la opción de nube es la mejor alternativa. Si el presupuesto es considerable por la **criticidad de la IA para el negocio** de la organización, debe valorarse la opción *on premise* para asegurarse de que los costes están controlados y cualquier aspecto de privacidad o protección de datos puede gestionarse internamente, sin depender de terceros.

Por último, los **entornos híbridos** son útiles en los siguientes casos:

- Cuando la regulación impida el tratamiento de determinados datos fuera de la organización o el territorio en el que opera. En ese caso, se puede dividir el proceso de creación de modelos en aquellas tareas que pueden realizarse a través de los servicios de IA del proveedor de nube y aquellas que deben realizarse en el centro de datos para cumplir con la regulación.

- Cuando el coste de una actividad determinada sea inferior si se realiza en el centro de datos de la organización, en relación con el coste de los servicios correspondientes en el proveedor de nube.

1.6. Aspectos de seguridad y privacidad de los entornos de nube

Aspectos de seguridad

El uso de servicios de IA en la nube trae consigo el **almacenamiento y procesamiento de información** potencialmente sensible en sistemas de terceros (el proveedor de nube), fuera del control de la organización propietaria de dicha información. Por otro lado, los activos generados en el proceso de IA (modelos) también se alojan en estos sistemas de terceros.

Estas dos circunstancias desencadenan la necesidad de evaluar las **medidas de seguridad** que implanta el proveedor de nube para incrementar las garantías de que datos y modelos no acaben expuestos como resultado de una brecha de seguridad. Sin embargo, el consumo de servicios en la nube implica un modelo de **responsabilidad compartida,** en el que la seguridad también involucra al cliente de los servicios.

En la Tabla 3, se indican los **requisitos exigibles al proveedor** de nube en materia de seguridad.

Categoría	Descripción
Estándares	Cumplimiento de estándares de ciberseguridad y certificaciones como ISO 27001, SOC 2, PCI DSS, GDPR. Implica que el proveedor dispone de políticas, procedimientos y controles para proteger la confidencialidad, integridad y disponibilidad de la información.
Auditorías e informes de seguridad	El proveedor realiza auditorías de seguridad periódicas y genera informes de resultados que tienen asociado un plan de acción para garantizar la mejora continua.
Plan de respuesta a incidentes	El proveedor dispone de un plan de respuesta y recuperación ante incidentes de seguridad.
Cifrado	El proveedor de cifrar los datos almacenados y en tránsito con protocolos de cifrado robustos como AES 256 y TLS 1.3. El proveedor debe proporcionar la opción de que el cliente traiga sus propias claves de cifrado.
Política de *backup*, retención y destrucción de datos	El proveedor debe disponer de políticas de *backup* y retención de los datos, así como procedimientos de destrucción segura de los datos.
Soberanía del dato	El proveedor debe cumplir con las regulaciones de protección de datos de las regiones donde los datos son almacenados y procesados.
Acuerdos de nivel de servicio	El proveedor debe indicar el nivel de servicio proporcionado y su garantía de disponibilidad, así como las penalizaciones aplicables en caso de incumplimiento. El proveedor debe disponer de una estrategia de redundancia acorde con los niveles de servicio garantizados.

Tabla 3. Requisitos de seguridad aplicables al proveedor. Fuente: elaboración propia.

Por otro lado, en la Tabla 4, se indican las **medidas de seguridad** que puede adoptar el **cliente** de servicios de IA en la nube para complementar las medidas de seguridad del proveedor.

Categoría	Descripción
Monitorización de seguridad	La actividad de los servicios y sistemas utilizados para la aplicación de IA deben estar monitorizados para identificar actividad sospechosa e incidentes de seguridad.
Cifrado	Los datos utilizados en los servicios de IA deben estar cifrados tanto en los sistemas de almacenamiento como cuando son transmitidos a través de redes.
Análisis de seguridad	Se deben utilizar los sistemas de análisis de seguridad proporcionados por el proveedor de nube para identificar y corregir debilidades en los servicios de IA.
Antivirus y *antimalware*	Se deben activar las funcionalidades de detección de virus y *malware* que proporciona el proveedor de nube.
Acceso remoto	El acceso a los servicios de IA debe realizarse a través de un canal seguro, lo que puede implicar el uso de VPNs.
Endpoint detection and response	Se deben activar las funcionalidades de protección de *endpoints* que proporciona el proveedor de nube.
Separación de entornos	Se deben crear, mantener y proteger diferentes entornos (desarrollo, preproducción, producción) involucrados en el ciclo de vida de gestión de modelos.
***Firewall* de aplicación**	Si el modelo está productizado, el servicio que sirve las predicciones debe estar protegido por un *firewall* de aplicación.
Autenticación y autorización	Se deben establecer mecanismos de autenticación segura para el acceso a los servicios de IA y una correcta separación de privilegios para que cada usuario solo tenga acceso a la información y funcionalidades que necesita.
Seguridad de red	Las redes en las que se operan los servicios de IA deben estar correctamente segmentadas para aislar el impacto de un incidente de seguridad en una de ellas respecto al resto.

Tabla 4. Requisitos de seguridad aplicables al cliente. Fuente: elaboración propia.

Aspectos de privacidad

La **regulación de la privacidad** tanto en el ámbito de la inteligencia artificial como en los entornos de computación en la nube es una temática de muy reciente aparición. Si bien se están realizando progresos a marchas forzadas, hoy en día no existe un punto de consenso entre países ni jurisdicciones. Por ello, la información vertida en este documento debe ser **sometida a revisión futura** para incorporar los nuevos avances y acuerdos que permitan disponer de una postura consolidada y homogénea para la toma de decisiones.

La evolución de esta postura afecta especialmente a los proveedores de nube, puesto que ostentan una posición dominante en el mercado de la IA. Estas empresas toman el **rol de un tercero *(third-party)*** que provee herramientas, modelos y algoritmos cuyo uso inapropiado puede introducir riesgos en la privacidad de los datos de ciudadanos, gobiernos y empresas.

Para gestionar correctamente la cuestión de la privacidad de IA en la nube, es preciso contemplar tanto los requisitos asociados a nuestro caso de uso como aquellos que son exigibles al proveedor de nube. El objetivo es tener tantas **garantías** como sea posible de que no se incurra en ilegalidades o vulneración de derechos.

En la Tabla 5, se incluyen aquellos **requisitos** que deben evaluarse en relación con el caso de uso o aplicación de la IA que empleará los servicios de IA en la nube.

Principio	Riesgo identificado	Medida recomendada
Uso legítimo, igualitario, legal y claramente definido	Uso de IA para propósitos no autorizados. Resultados con sesgo discriminatorio por uso inadecuado de variables.	Seleccionar *inputs* con criterios de legalidad e idoneidad para el propósito. Trazas relaciones de causalidad justificadas entre los datos de entrada y las predicciones obtenidas.
Transparencia y explicabilidad	Recoger datos personales no fiables, imprecisos, incompletos o no autorizados. Incapacidad de explicar la razón de una decisión o evitar una decisión con resultados desastrosos.	Incluir el derecho a ser informado durante la recopilación de datos. Utilizar algoritmos o mecanismos que permitan explicar las causas de una decisión tomada por la IA.
Gobernanza y trazabilidad	Puesta en producción de modelos cuyos errores pueden tener un impacto negativo muy superior al previsto. Ausencia de responsabilidades en la organización frente a dichos errores.	Definir procesos de validación integrales de los modelos, especialmente de aquellos cuyos errores puedan ocasionar impacto negativo en las personas. Definir claramente roles y responsabilidades de los procesos de IA.

Principio	Riesgo identificado	Medida recomendada
Minimización de datos	Uso de variables irrelevantes que influyen negativamente en la precisión de las predicciones, con perjuicio para las personas. Introducción de sesgos que afectan a colectivos o minorías (no suficientemente representadas en los datos de entrenamiento).	Utilizar únicamente las variables necesarias para el propósito. Excluir datos de mala calidad o inapropiados.
Limitación del propósito	Introducir nuevos usos de los modelos que conduzcan a escenarios no deseables, como desinformación, información incorrecta o difamación de personas.	Restringir el propósito del modelo al caso de uso originalmente identificado para el que se han llevado a cabo las comprobaciones oportunas en materia de privacidad.
Precisión de los datos personales	El uso de datos incorrectos puede conducir a decisiones tomadas por la IA que causan perjuicio a las personas.	Las personas deben tener el derecho a corregir los datos que han autorizado a tratar a un tercero. Los datos deben tener asociados estrictos controles de calidad antes de ser usados.
Limitación de la retención de los datos	La conservación de datos no necesarios incrementa la exposición a incidentes de ciberseguridad relacionados con fuga de datos.	Implantar procedimientos de eliminación segura de datos y asegurarse de que se ejecutan siempre que un conjunto de datos usados en procesos de IA deja de ser necesario.

Principio	Riesgo identificado	Medida recomendada
Confidencialidad, integridad y disponibilidad	La ausencia de una postura de seguridad madura abre la puerta a ciberataques que desencadenan incidentes de fuga de información con datos personales.	Ampliar la postura de seguridad para contemplar los procesos de IA y todos sus componentes.
Respeto de la privacidad del usuario final	Una compañía que no aplique suficientes controles puede incurrir en incumplimientos regulatorios que provocan sanciones y pérdida de reputación.	Definir e implementar controles específicos para garantizar la privacidad en los procesos de IA, extremo a extremo.

Tabla 5. Principios de privacidad para aplicaciones de IA. Fuente: elaboración propia.

Por otro lado, todos los proveedores de computación en la nube, al margen de si los servicios son de IA o no, deben cumplir con las exigencias de la **ISO 27018,** como estándar de referencia en la protección de datos personales en la nube.

Entre las **directrices** de este estándar, destacan las indicadas a continuación:

- Los proveedores de nube no pueden utilizar los datos personales con **propósitos de *marketing,*** a menos que el propietario de los datos (la persona) lo autorice de manera expresa.
- Los proveedores de nube deben **proteger el tránsito** de datos personales a través de redes públicas al almacenarlos en dispositivos o al restaurarlos.
- Los proveedores de nube deben firmar un **acuerdo de confidencialidad** para el tratamiento de datos personales. Además, deben proporcionar formación específica a los empleados involucrados en su tratamiento.
- Si un proveedor de nube sufre una **brecha de seguridad** que afecta a los datos personales, los propietarios deben ser informados inmediatamente.

- El proveedor de nube debe informar de todas las **empresas que realizan tratamiento** de los datos personales recopilados.

Finalmente, la recientemente publicada **Ley de la IA de la Unión Europea** establece las líneas maestras de actuación para una IA responsable en Europa, y aplica de manera directa a los principales proveedores de nube que prestan servicios de IA.

El EU AI Act establece un **sistema de clasificación del nivel de riesgo** asociado a una determinada aplicación de la IA. Los requerimientos fijados para las aplicaciones de **mayor riesgo** incluyen áreas de control en los siguientes ámbitos:

- Implantación de un sistema de gestión de riesgos.
- Implantación de controles de gobernanza de datos.
- Generación de documentación técnica precisa.
- Preservación de registros.
- Transparencia e información precisa al usuario.
- Supervisión humana a las decisiones tomadas por la IA.
- Precisión de los datos, robustez de los sistemas y medidas de ciberseguridad.
- Implantación de un sistema de gestión de calidad.
- Análisis de impacto de derechos fundamentales.

Proveedores como **Microsoft** han anunciado que los principios de su estrategia de privacidad en el ámbito de IA están alineados con las directrices del EU AI Act. Otros, como **Amazon,** han publicado guías para analizar el riesgo incurrido por el uso de sistemas de IA en la nube. Por último, **Google** mantiene en su sitio web una colección de recursos que explican su compromiso con el cumplimiento de regulaciones de IA, incluyendo el EU AI Act.

1.7. Cuaderno de ejercicios

1. En el artículo cuyo enlace se facilita abajo, se proporciona un *state-of-the-art* de amplio espectro en el ámbito de IA en la nube. ¿Cuáles son las líneas de actividad que recomiendan los autores para el futuro a partir de los datos recabados en su estudio?

Zangana, H. M. y Zeebaree, S. R. (2024). Distributed Systems for Artificial Intelligence in Cloud Computing: A Review of AI-Powered Applications and Services. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM), 5*(1), 11-30.
https://ojs.unikom.ac.id/index.php/injiiscom/article/download/11883/4150

Son las siguientes:

- Integración de técnicas de explicabilidad en la IA para comprender con qué criterios toma la IA las decisiones.
- Estandarización de las métricas de rendimiento.
- Mejora de los protocolos de seguridad.
- Enfoques que garanticen la escalabilidad.
- Consideraciones éticas y eliminación de sesgos.
- Colaboración para mejorar la seguridad IoT, es decir, en el *edge computing.*
- Detección de las tendencias emergentes.
- Servicios diseñados con foco en el usuario.
- Transferencia de conocimientos entre diferentes ámbitos o temáticas.

- Iniciativas de formación y concienciación.
- Plataformas robustas para analizar trazas generadas.

2. En el siguiente enlace se proporcionan criterios para comparar los *offerings* de los proveedores de nube. En particular, incluye ejemplos de aspectos que deben considerarse durante el proceso de evaluación. Indica los tres aspectos más relevantes para un proyecto que pretende resolver una problemática común, para la cual puede haber modelos creados, y que demanda una considerable capacidad de computación, pero solo en momentos puntuales.

How can you compare cloud platforms with AI services? (s. f.). *LinkedIn.* https://www.linkedin.com/advice/0/how-can-you-compare-cloud-platforms-ai

---

Los tres aspectos clave, tomados de la sección «Feature comparison», serían:

- *«Does it provide pre-trained models, APIs, or tools for specific AI tasks?»:* porque al ser una problemática común, pueden existir modelos preentrenados que nos eviten tener que construirlo desde cero.
- *«Does it offer GPU, TPU, or other specialized hardware for AI?»:* porque si la tarea demanda mucha capacidad de computación, que el proveedor disponga de *hardware* especializado permite satisfacer dicha demanda más fácilmente.
- *«Does it offer pay-as-you-go or subscription pricing models for AI?»:* porque si la demanda de capacidad de computación es puntual, nos interesa un servicio de *pay-as-you-go,* de manera que no tengamos que incurrir en costes fijos y solo generemos gasto cuando necesitamos la infraestructura.

3. De los beneficios obtenidos por el uso de IA en la nube, ¿cuáles son los que están directamente relacionados con optimización de costes?

Existen dos beneficios relacionados con ahorro de costes:

- El hecho de que los costes se vinculan al consumo real y la capacidad requerida en cada momento, evitando de esta manera realizar inversiones considerables en infraestructura propia e incurrir en costes fijos.
- El menor coste de personal porque las tecnologías de IA se entregan como servicio (es el proveedor quien contrata a los especialistas) y, especialmente, por la disponibilidad de capacidades AutoML, en las que un usuario sin conocimientos avanzados en ciencia de datos puede construir modelos.

4. ¿Cuáles son los siete puntos que deben considerarse en materia de seguridad a la hora de elegir un proveedor de IA en la nube?

Son los siguientes:

- El proveedor cumple los principales estándares de ciberseguridad.
- El proveedor realiza auditorías de seguridad periódicas.
- El proveedor dispone de un plan de respuesta ante incidentes.
- El proveedor cifra datos almacenados y en tránsito con protocolos robustos.
- El proveedor dispone de políticas de *backup,* retención de datos y destrucción segura de datos.
- El proveedor cumple con las regulaciones de protección de datos de las regiones donde los datos son almacenados y procesados.
- El proveedor se compromete a un nivel de servicio y disponer de una estrategia de redundancia acorde con dicho nivel.

5. En el siguiente enlace se incluye un artículo que evalúa cómo la IA está cambiando el mercado de computación en la nube. ¿Cuál es el consejo de Maynard Williams, *managing director* en Accenture, para las empresas que quieran adoptar la IA en sus procesos de negocio?

Law, M. (2023, diciembre 1). How AI is changing the cloud landscape. *Technology Magazine.* https://technologymagazine.com/articles/how-ai-is-changing-the-cloud-landscape

---

Maynard Williams aconseja contemplar primero el problema de negocio que se desea resolver, y solo después evaluar las tecnologías y procesos necesarios para abordarlo. En particular, se debe evitar formular el problema de la siguiente manera: «¿Cómo podemos empezar a utilizar la IA lo antes posible?».

1.8. Referencias bibliográficas

Amazon SageMaker (s. f.). *Ejemplo de cuadernos.* https://docs.aws.amazon.com/es_es/sagemaker/latest/dg/howitworks-nbexamples.html

International Organization for Standardization (2019). *Information technology — Security techniques— Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors* (estándar ISO 27018:2019). https://www.iso.org/standard/76559.html

Manashgoswami (2023, junio 7). What is automated ML? AutoML - Azure Machine Learning. *Microsoft Learn.* https://learn.microsoft.com/en-us/azure/machine-learning/concept-automated-ml?view=azureml-api-2

Msv, J. (2023, junio 30). Generative AI cloud platforms: AWS, Azure, or Google? *The New Stack.* https://thenewstack.io/generative-ai-cloud-services-aws-azure-or-google-cloud/

Parlamento Europeo (2024, junio 18). Ley de la IA de la UE: primer reglamento sobre inteligencia artificial. Unión Europea. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Sate of AI in the cloud 2024

Cohen, A. y Moore, M. (2024, enero 17). Wiz Research presents its latest report: "State of AI in the cloud 2024." *Wiz.* https://www.wiz.io/blog/key-findings-from-the-state-of-ai-in-the-cloud-report-2024

El informe de Wiz Research proporciona las claves para entender, a partir de datos empíricos, el proceso de adopción de la IA en la nube por parte de las organizaciones. Revela también qué proveedores están experimentando mayor crecimiento y en qué áreas, así como los patrones de consumo de estos servicios, lo que permite anticipar la evolución futura de la IA en la nube en los próximos años.

2023 Gartner report Magic quadrant for cloud AI developer services

Scheibmeir, J., Sicular, S., Batchu, A., Fang, M., Baker, V. y O'Connor, F. (2023). *2023 Gartner report Magic quadrant for cloud AI developer services.*

El informe de Gartner constituye un estudio pormenorizado de las capacidades de cada proveedor de nube en el ámbito de los servicios de IA. Se establecen criterios de evaluación objetivos y se revelan las fortalezas y debilidades identificadas en cada proveedor. Con esta información, resulta más sencillo asesorar en materia de IA en la nube a un potencial cliente que desee emprender este camino.

1. ¿Por qué se considera que el enfoque de IA en la nube es ágil?

A. Porque los entornos de trabajo que facilita el proveedor de nube se ejecutan más rápidamente que su equivalente en un centro de datos.

B. Porque las capacidades de IA pueden provisionarse en minutos.

C. Porque es sencillo decidir qué proveedor de nube utilizar, independientemente del caso de uso o las necesidades específicas.

D. Porque el acceso a los servicios de IA puede realizarse directamente desde un navegador web.

2. ¿En qué consiste la ingeniería de variables en AutoML?

A. En normalizar los valores de cada variable existente en el *dataset.*

B. Es la única actividad para la que el AutoML no proporciona ningún tipo de automatización.

C. En la clasificación automática de las variables existentes y la detección de su grado de importancia, así como la generación de nuevas variables.

D. Consiste en identificar el significado de cada variable en el contexto del caso de uso que se pretender resolver con el modelo de IA.

3. ¿Qué porcentaje de entornos de nube dispone en la actualidad de servicios de IA contratados según Wiz Research?

A. En torno al 70 %.

B. En torno al 99 %.

C. En torno al 25 %.

D. Residual, prácticamente el 0 %.

4. ¿Cuál es uno de los beneficios de contratar servicios de IA en la nube?

A. Que el coste es siempre menor respecto a la provisión de capacidades de IA en un centro de datos propio.

B. Que parte del trabajo de creación de modelos lo realizan los ingenieros del proveedor.

C. Que en general se obtienen mejores rendimientos en los modelos proporcionados por el proveedor respecto a los que podemos crear nosotros.

D. El consumo de modelos preentrenados por el proveedor para resolver problemáticas conocidas.

5. ¿Cuál es un escenario en el que se recomienda desplegar capacidades de IA *on premise* en lugar de hacerlo a través de un proveedor de nube?

A. Cuando se dispone de un presupuesto bajo.

B. Si el grado de personalización que se requiere es elevado.

C. Cuando solo se desea experimentar con la IA, pero aún no hay grado de madurez suficiente para llevarla a producción.

D. Cuando no dispongamos de personal suficientemente especializado.

6. ¿Cuál de los siguientes es un requisito exigible al proveedor en materia de seguridad?

A. La separación de entornos: desarrollo, preproducción y producción.

B. El uso de VPNs para acceder a los servicios de IA de forma segura.

C. La auditoría de los datos que subimos a sus servicios para garantizar que no incumplen regulaciones de privacidad.

D. Cumplimiento de estándares de seguridad y posesión de certificaciones relacionadas con seguridad.

7. ¿Cuál de los siguientes es un requisito de seguridad que debe adoptar la propia organización que contrata servicios de IA en la nube?

A. El parcheado de los sistemas a través de los cuales se prestan los servicios de IA.

B. La actualización de los sistemas operativos en los servidores.

C. Establecer una correcta separación de privilegios para que cada usuario solo tenga acceso a la información y funcionalidades que necesita.

D. La disponibilidad de mecanismos de autenticación robustos para acceder a los servicios de IA.

8. ¿Cuál es una de las características diferenciales de los principales proveedores de nube en relación con proveedores de nicho en el ámbito de IA?

A. La disponibilidad de modelos preentrenados.

B. La integración con el resto de los servicios de computación en la nube, lo que permite construir una solución de negocio con un único proveedor.

C. Las capacidades de AutoML.

D. Las capacidades de visión artificial.

9. ¿Para qué sirve la etapa de extracción de tókenes en los sistemas de procesamiento del lenguaje natural?

A. Para detectar de palabras, expresiones o símbolos clave que permitan esclarecer la intención del mensaje.

B. Para revisar ortográficamente el contenido del mensaje.

C. Para transformar el mensaje en código fuente interpretable por una máquina.

D. Para descartar el contenido irrelevante del mensaje.

10. ¿En qué consiste la minimización de datos en relación con la privacidad?

A. Consiste en utilizar el menor volumen de datos posible para reducir el impacto de una brecha que afecte a la privacidad.

B. Consiste en reducir el rango de las variables numéricas para evitar cifras elevadas que pueden interferir con los algoritmos de IA utilizados.

C. Consiste en comprimir los datos antes de enviarlos al servicio de IA.

D. Consiste en utilizar únicamente variables relevantes en el proceso de entrenamiento, evitando aquellas que puedan generar sesgos con perjuicio para las personas.