Guía Desarrollo Hackathon2

















Fecha: V.2 6 Mayo 2025





1.	Introducción	3
2.	Solicitud de Acceso al programa Microsoft Founders y creación de la cuenta de Azur	e4
	2.1 Criterios de Elegibilidad	4
	2.2 Consejos para una Aplicación Exitosa	4
	2.3 Revisión y tiempos para su evaluación	5
	2.4 Ayuda y soporte técnico del programa	5
	2.5 Proceso de Aplicación Paso a Paso	5
3.	Canjear los créditos	11
	3.1 Cómo canjear esta ventaja:	11
4.	Guías sobre servicios y recursos comunes	15
	4.1 Crear un Workspace en Azure	15
	4.2 Creación de una instancia de computación	21
	4.3 Crear un notebook	23
	4.4 Creación de una instancia de Azure Al Services	25
	4.5. Creación de una instancia de Azure Al Search	28
	4.6. Implementación de Code-Carbon en la solución	32
	4.7. Petición de ampliación de cuota para GPUs	32
5.	Desarrollo de un caso de uso utilizando Fine-Tuning	35
	5.1 Descripción del caso de uso	35
	5.2 Servicios necesarios en Azure	36
6.	Requisitos para desarrollar una PoC de RAG	36
	6.1 Descripción del caso de uso	36
	6.2 Servicios necesarios en Azure	36
7.	Requisitos para desarrollar una PoC de TAG	37
	7.1 Descripción del caso de uso	37
	7.2 Servicios necesarios en Azure	37





1. Introducción

Este documento tiene como finalidad definir los requisitos técnicos, servicios necesarios y consideraciones para los equipos que participarán en el hackathon sobre IA Generativa utilizando Azure, así como una serie de guías sobre cómo desarrollar los casos de uso planteados.

El hackathon se centrará en cuatro líneas de trabajo principales: Prompt-Engineering, Fine-Tuning, Retrieval Augmented Generation (RAG) y Table Augmented Generation (TAG). Los equipos deberán desarrollar su caso de uso basado en una de ellas. Podrán implementar los casos de uso utilizando las tecnologías y servicios recomendadas en el documento o cualquiera de elección propia, siempre y cuando se desarrolle en Azure. Cada uno de estos casos combinan la necesidad de comprensión profunda de los modelos generativos con la integración de servicios en la nube, datos relevantes y su implementación técnica. Los equipos deberán aprovechar modelos externos disponibles en plataformas como Hugging Face, ampliando así las posibilidades de personalización y experimentación. Además, Los casos de uso que se planteen deberán ser aquellos cuya forma óptima de desarrollo sea mediante IA generativa, en comparación con otras técnicas como la analítica o la IA tradicional.

El hackathon se desarrollará en la nube de Microsoft Azure, un entorno que ofrece una amplia gama de herramientas y servicios diseñados específicamente para respaldar la construcción, implementación y escalabilidad de soluciones basadas en IA. Cada equipo contará con acceso a una cuenta de Microsoft for Startups Founders con 1000 dólares en créditos de Azure, que deberán utilizar durante todo el desarrollo del reto. Los equipos deben realizar una solicitud en el programa, tal y como se indica en el punto 2, en la siguiente página, para que les puedan proporcionar acceso a los servicios necesarios para abordar el caso de uso seleccionado, garantizando flexibilidad y capacidad técnica para cada desafío.

En este documento se detallan los requisitos técnicos necesarios para cada caso de uso, los servicios de Azure que se necesitarán y una guía general para estimar los recursos necesarios. También se proporcionan recomendaciones sobre la configuración inicial de las cuentas de Azure, las mejores prácticas para la integración de modelos externos y consideraciones clave sobre el manejo de costos y eficiencia.

El éxito del hackathon dependerá de la capacidad de los equipos para combinar la tecnología de Azure con sus ideas creativas y habilidades técnicas, todo ello teniendo en mente en todo momento la componente de sostenibilidad, ajustando el desarrollo de sus casos de uso al budget establecido.







2. Solicitud de Acceso al programa Microsoft Founders y creación de la cuenta de Azure

El programa <u>Microsoft for Startups Founders Hub</u> se está utilizando como base para el desarrollo del hackathon, ya que permite proporcionar a los equipos participantes las herramientas necesarias, acceso a recursos técnicos, y un entorno en Azure que facilita la creación y despliegue de sus soluciones. Este enfoque asegura que los equipos cuenten con el soporte adecuado para desarrollar sus ideas y maximizar su impacto durante la competición.

Además, una vez finalizado el proceso de competición los participantes que hayan adquirido una calidad de los proyectos considerable pueden seguir beneficiándose de las herramientas del programa.

2.1 Criterios de Elegibilidad

Para participar en el Hackathon y acceder al programa de Microsoft, los requisitos básicos son los siguientes:

- Participar en el desarrollo de un producto o servicio basado en software. El software debe ser de propiedad propia y no estar licenciado por terceros.
- Haber recibido menos de 10 000 USD en créditos gratuitos de Azure anteriormente.
- Tener la sede central en uno de los países cubiertos por Azure.
- No haber pasado por una ronda de financiación Serie D o posterior.
- Operar con fines de lucro.
- No ser una institución educativa, entidad gubernamental, blog personal, tienda de desarrollo, consultoría, agencia, empresa de minería de criptomonedas o relacionada con Bitcoin.

2.2 Consejos para una Aplicación Exitosa

Para realizar una solicitud exitosa, el participante debe seguir estas recomendaciones:

- **Describir la idea desde una perspectiva empresarial.** Explica tu propuesta como si fuera un proyecto de negocio, resaltando su viabilidad y potencial de impacto.
- Registrar solo un participante por equipo. Cada equipo debe designar un único representante o "líder" encargado del registro.
- **Proporcionar el máximo nivel de detalle.** Explica tu idea de manera clara y estructurada, asegurándote de que se comprenda fácilmente tu visión.
- **Destacar el propósito y la visión del proyecto.** Explica de forma precisa cuál es el propósito de tu startup y cómo generará impacto en el mercado.
- **Incluir el papel de la tecnología de Microsoft.** Describe cómo herramientas como Azure pueden potenciar el desarrollo y crecimiento de tu startup.
- Revisar los datos personales. Asegúrate de que la información de contacto proporcionada sea correcta, ya que será el medio para comunicarse contigo.







 Estandarizar los nombres del proyecto. Utiliza un nombre sencillo que describa la idea de negocio.

Ejemplo de idea:



2.3 Revisión y tiempos para su evaluación

En promedio, se tardan tres días laborables en revisar una aplicación. Los tiempos de respuesta pueden retrasarse si es necesario comprobar la información de la aplicación. Se le notificará tan pronto como se tome una decisión.

2.4 Ayuda y soporte técnico del programa

Para cualquier inconveniente o consulta técnica:

- 1. Selecciona el icono "?" en la esquina superior derecha de la pantalla.
- Selecciona "Enviar una incidencia de soporte técnico". Adjunta documentos relevantes o captura de pantalla de los errores según sea necesario.
 También puedes enviar la consulta al siguiente correo electrónico: startups@microsoft.com

2.5 Proceso de Aplicación Paso a Paso

Acceso a una guía de vídeos con toda la información <u>Founders Hub Tips and Tricks - YouTube</u> entre ellos está un vídeo somo como enviar la solicitud al programa. También se encuentra más detalle en la web de Microsoft para la Aplicación a Microsoft for Startups Founders Hub: <u>Información general sobre el proceso de aplicación de Microsoft for Startups | Microsoft Learn</u>

PASO 1: ACCEDER AL PORTAL DE APLICACIÓN

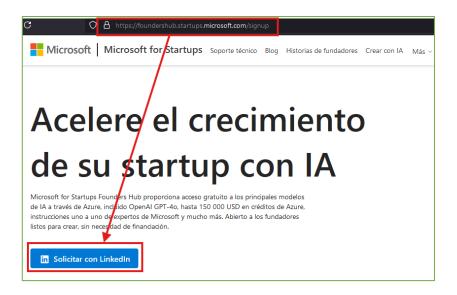
Para acceder al programa, un representante del equipo debe seguir los pasos indicados.

• Ve al sitio oficial de Microsoft for Startups Founders Hub.









Haz clic en el botón de "Apply with LinkedIn". (Si no dispones de perfil de LinkedIn, es necesario crearte uno)

NOTA IMPORTANTE: La solicitud está redactada bajo el supuesto de que el participante es una startup. Sin embargo, esto no implica que quienes aún no tengan constituida su startup no puedan participar. Cada vez que se mencione el término "startup", deberás completar la información correspondiente al proyecto que has ideado para concursar en la competición.

PASO 2: COMPLETAR EL FORMULARIO DE REGISTRO

Debemos completar un formulario, que no nos tomará más de diez minutos, en el que se nos solicitará:



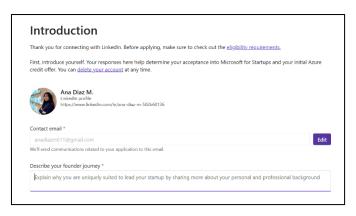
1. Información general:

- o Nombre, correo electrónico, número de contacto.
- o Breve descripción acerca de tu experiencia.









2. Detalles de la startup (proyecto):

Los campos obligatorios están marcados con*

- o Nombre del proyecto* y web de la startup (opcional).
- o Cuál es el modelo de negocio principal de tu startup (B2B/B2C).
- o Descripción del producto o servicio.
- o Sector en el que opera (FinTech, HealthTech, e-commerce, etc.).
- o Nivel de experiencia con Azure*
- o Fase de desarrollo (idea, prototipo, producto lanzado).

3. Dirección

Si no eres una empresa consolidada, deberás poner tu dirección habitual.

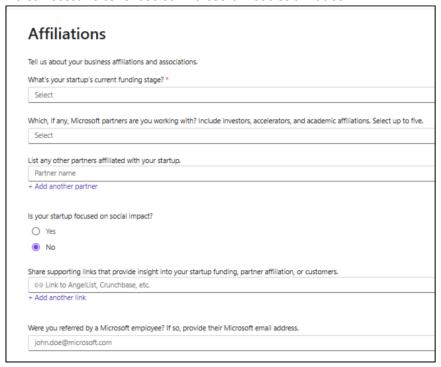








4. No es necesario tener socios Microsoft ni socios afiliados



Info Adicional:

More about you These questions help Microsoft provide relevant resources, such as mentor communities, when available. Your responses won't affect decisions about your participation in Microsoft for Startups. Are you a technical or non-technical founder? * Technical founder What's your self-identified gender? Female

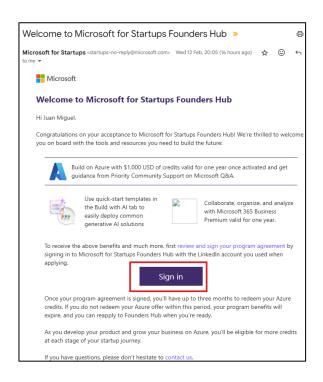
5. Revisión y envío de la solicitud:

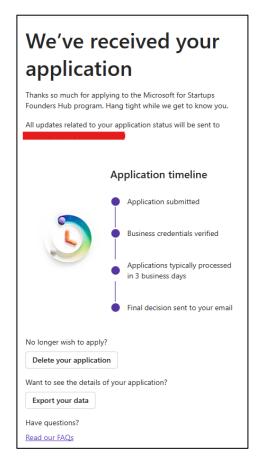
Una vez finalizado el proceso, verás en la web una confirmación de que tu solicitud ha sido recibida correctamente y se encuentra en revisión. También te comunicarán por correo electrónico. Recibirás una respuesta en menos de cuatro días laborables.



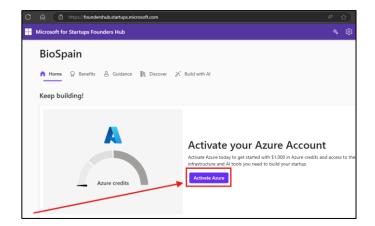








Una vez que ingreses al portal de Microsoft Founders Hub con tu cuenta: https://foundershub.startups.microsoft.com/ podrás activar tu cuenta de Azure desde la sección "Home" haciendo clic en "Activate Azure". Al hacerlo, obtendrás \$1,000 en créditos de Azure y acceso a la infraestructura y herramientas de IA necesarias para desarrollar tu proyecto/idea.

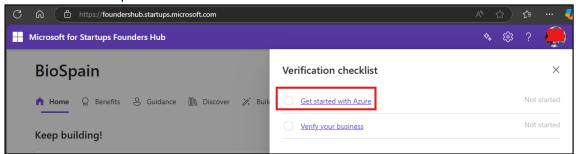




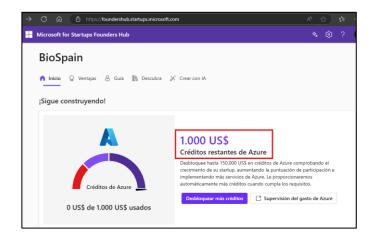




Seleccionamos "Empezar con Azure"



y al finalizar veremos la pantalla de confirmación en la que nos dice que tenemos los 1000\$ US disponibles



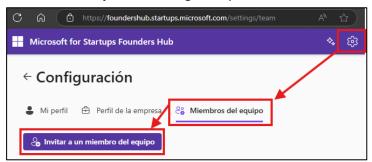
PASO 3: INCLUIR AL RESTO DE MIEMBROS DEL EQUIPO:







Una vez activada la cuenta, deberás incluir al resto de miembros de tu equipo desde el apartado de configuración del portal, seleccionado el icono de engranaje en la esquina superior derecha de la pantalla. Este icono se encuentra junto a la imagen de perfil.



- 1. Selecciona "Invitar a un miembro del equipo" en la pestaña "Miembros del equipo".
- 2. Agrega la dirección de correo electrónico del nuevo miembro del equipo y haga clic en "Enviar invitación".
- 3. El miembro del equipo recibirá una invitación por correo electrónico. Deben aceptar la invitación para que se complete este paso.

3. Canjear los créditos

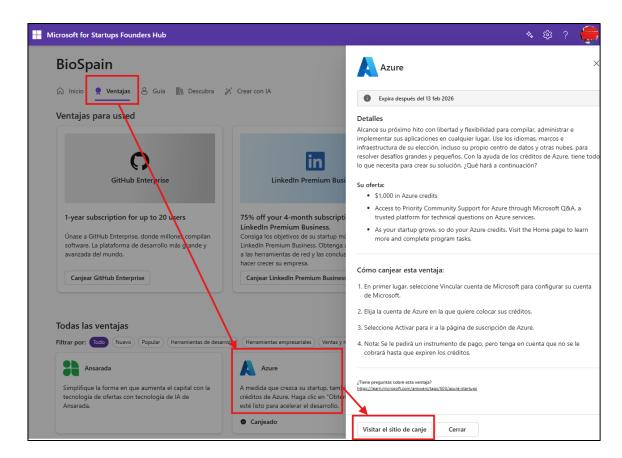
3.1 Cómo canjear esta ventaja:

- 1. En primer lugar, seleccione Vincular cuenta de Microsoft para configurar su cuenta de Microsoft.
- 2. Elija la cuenta de Azure en la que quiere colocar sus créditos.
- 3. Seleccione Activar para ir a la página de suscripción de Azure.
- 4. Nota: Se le pedirá un instrumento de pago, pero tenga en cuenta que no se le cobrará hasta que expiren los créditos.





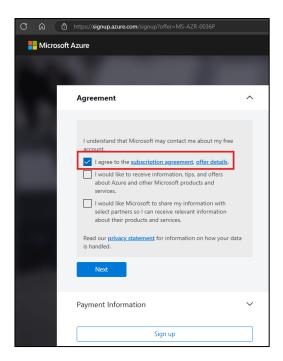




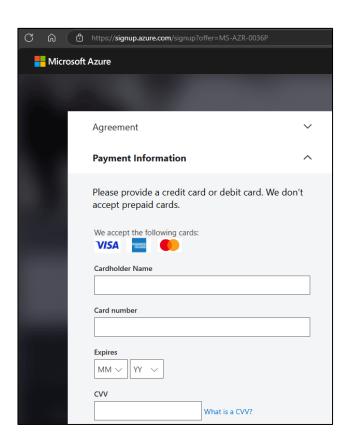
Aparece la multiselección en la sección de acuerdo donde debes seleccionar de forma obligatoria que se acepta el contrato de subscripción y los detalles de la oferta.







Posteriormente se dirigirá a la sección de pago, donde deberás introducir los datos de tu tarjeta bancaria para poder continuar y vincular el patrocinio.









Una vez introducido el medio de pago quedará conectado el patrocinio a tu cuenta Azure.

Desde el portal de Azure en el apartado de "Subscripciones" se podrá comprobar que tiene activado el patrocinio.

Nota:

Incluir un medio de pago es un **trámite necesario** para activar y conceder los créditos dentro del entorno Azure. **No se realizará ningún cargo al participante**, salvo en caso de superar el límite de **1.000** \$ asignado. Para evitar cualquier coste adicional, recomendamos seguir **las buenas prácticas y aclaraciones indicadas.** La organización **no se hace responsable** de posibles cargos que puedan generarse por exceder el crédito otorgado.

Aclaraciones IMPORTANTES sobre el Uso de Recursos Cloud en el Hackathon

Como parte del hackathon, los equipos tendrán acceso a recursos en la nube para el desarrollo de sus soluciones basadas en IA. Sin embargo, es fundamental gestionar adecuadamente el uso de estos recursos para evitar sobrecostes innecesarios.

Límite de gasto

Cada equipo debe asegurarse de no superar un límite máximo de **1.000 \$** en consumo de recursos cloud durante el hackathon. Es responsabilidad de cada equipo monitorizar su gasto y optimizar el uso de las máquinas virtuales y otros servicios.

Monitorización del gasto

Los participantes pueden revisar el consumo de recursos y el gasto acumulado en tiempo real a través del **panel de control de costes** de la plataforma cloud. Se recomienda revisar esta información periódicamente para evitar sorpresas. Las instrucciones detalladas sobre cómo acceder a estos datos estarán disponibles en la guía oficial del hackathon. <u>Costo y uso de Azure Monitor - Azure Monitor | Microsoft Learn</u>

Optimización y apagado automático de máquinas

Para evitar gastos innecesarios, es imprescindible **apagar las máquinas virtuales cuando no se estén utilizando**. Además, se recomienda configurar el **apagado automático** para prevenir consumos accidentales. <u>Apagado automático de una máquina virtual - Azure Virtual Machines</u> | Microsoft Learn







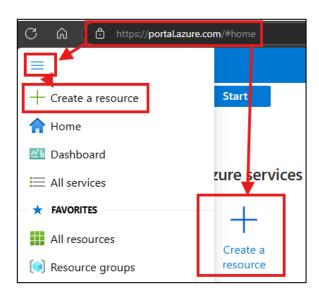
4. Guías sobre servicios y recursos comunes

Una vez que los equipos dispongan de una cuenta de Azure, para poder desarrollar cualquiera de los casos de uso planteados, deberán seguir una serie de pasos comunes como punto de partida.

4.1 Crear un Workspace en Azure

Para poder comenzar con el desarrollo, lo primero que deben hacer los equipos es crear un workspace. Para más información sobre qué es un workspace o espacio de trabajo, consultar este enlace ¿Qué es un área de trabajo de Azure Machine Learning?. Para poder crearlo, es necesario seguir los siguientes pasos:

1) En la consola de <u>Microsoft Azure</u>, dentro de la sección "Azure services" selecciona "Create a resource".



En el panel lateral, en la sección "Analytics", selecciona "Azure Machine Learning" y haz click encima de "Create" del nombre del servicio.

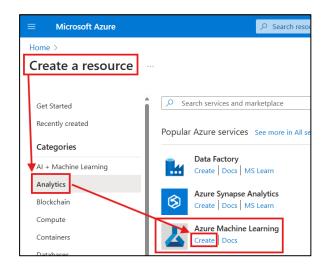
- 3) Rellena la información del workspace.
 - a. Subscription: Selecciona la opción por defecto.
 - b. Resource group: Pon el nombre "PNAV-Hackathon".
 - c. Name: Elige un nombre <u>que identifique a tu equipo</u> en el hackathon2 para el workspace. Por ejemplo: "WS-NombreEquipo",
 - d. Region: West Europe --> es importante indicar la región, ya que si se selecciona otra región es posible que no estén los mismos servicios disponibles.
 - e. Storage account: Mantén el valor por defecto.







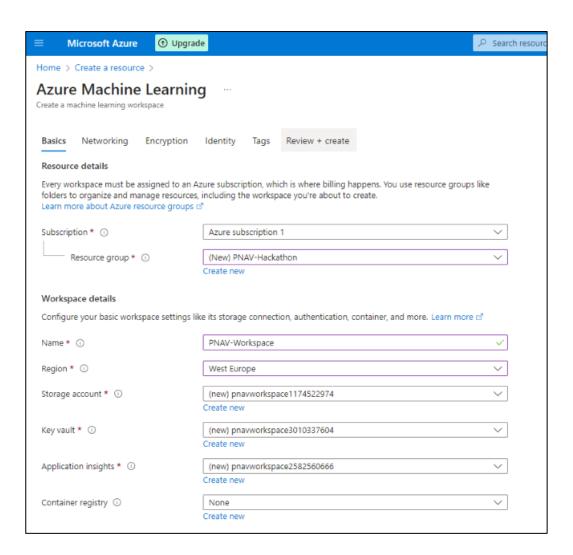
- f. Key vault: Mantén el valor por defecto.
- g. Application insights: Mantén el valor por defecto.
- h. Container registry: Mantén el valor por defecto.



Y selecciona "Next: Networking".



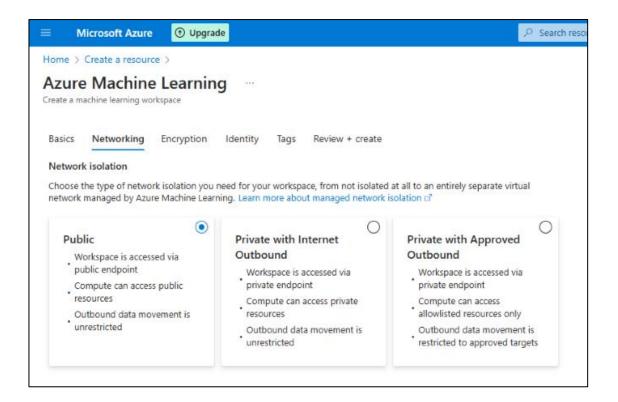




4) En la sección "Networking", selecciona la opción "Public". A continuación, haz clic en "Next: Encryption".

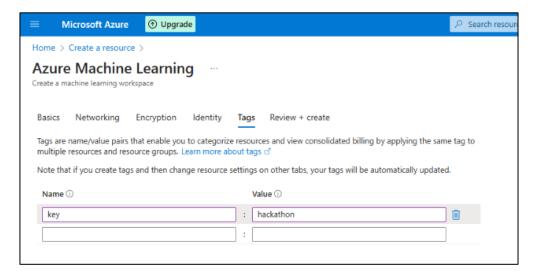






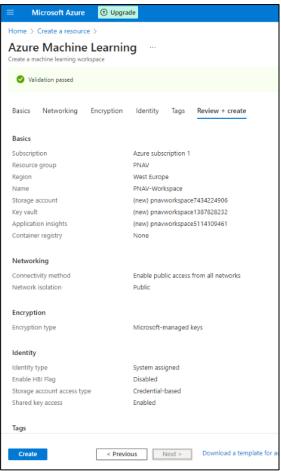
- 5) En la sección "Encryption" mantén la configuración por defecto. Selecciona "Next: Identity".
- 6) En la sección "Identity" mantén la configuración por defecto. Selecciona "Next: Tags".
- 7) En la sección "Tags", añade la siguiente etiqueta:
 - Name: "key"
 - Value: "hackathon"

Haz click en "Next: Review + create".

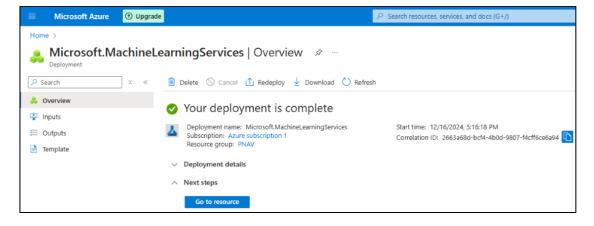








- 8) A continuación, se someterá al workspace a un proceso de validación que durará unos segundos. Si la validación es correcta, aparecerá el siguiente mensaje:
- 9) El proceso de despliegue puede durar varios minutos (en torno a 2 o 3). Cuando haya finalizado se podrá ver lo siguiente.

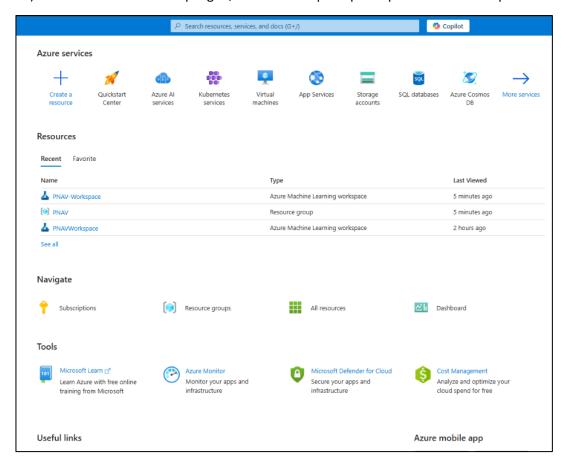




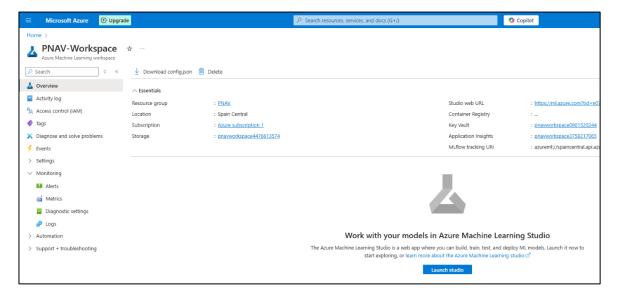




10) Una vez finalizado el despliegue, en la consola principal se podrá ver el workspace creado.



11) Para abrir el workspace, desde la consola principal, haz click en el nombre. Después, haz click en "Launch studio".





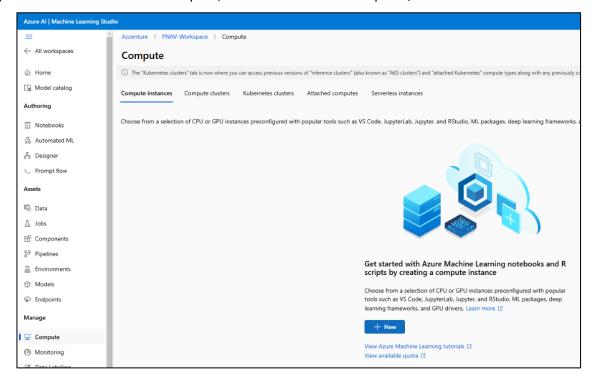




4.2 Creación de una instancia de computación

Para poder ejecutar código, entrenar o inferir modelos es necesario crear una instancia de computación. Para más información sobre qué es una instancia de computación, consultar <u>este enlace</u>. Es recomendable crear varios tipos de instancias, en función de las necesidades de ejecución.

1) En la consola del workspace, en la sección "Compute", haz click en "+ New"



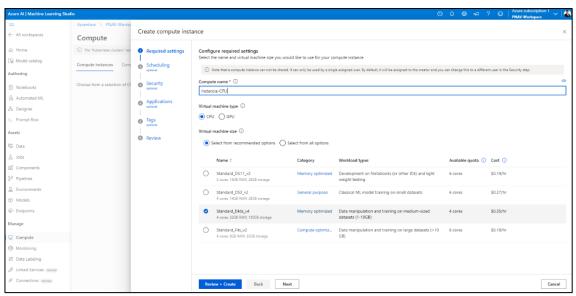
- 2) Configura la instancia de computación. Para ello:
 - a. Escoge un nombre para la instancia de computación.
 - b. Selecciona el tipo de instancia (CPU) dependiendo del tipo de tareas a las que se vaya a destinar el recurso.
 - Selecciona si se desea elegir entre opciones recomendadas o todas las opciones.
 Hay que recordar que existen límites de créditos (1000\$ por cuenta) que no deberéis sobrepasar.
 - d. Escoge el tamaño de la instancia.

Y haz click en "Next".



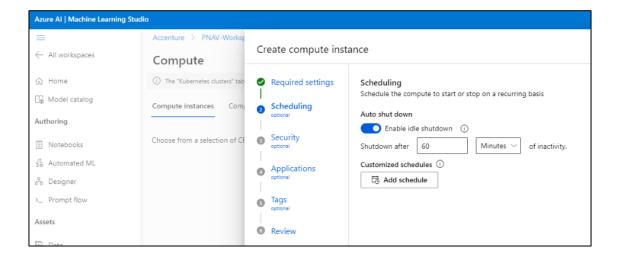






3) En la sección "Scheduling", mantén los ajustes por defecto de un shutdown después de 60 minutos de inactividad. Si no son esos los ajustes por defecto, cámbialos.

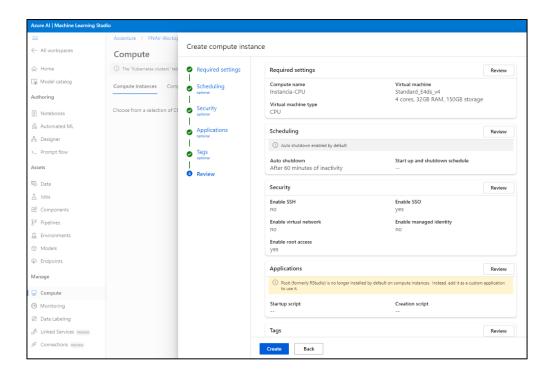
Y haz click en "Next".



4) En las pestañas "Security", "Applications" y "Tags" mantén los ajustes por defecto. En la sección "Review", comprueba que la configuración es correcta y selecciona "Create"

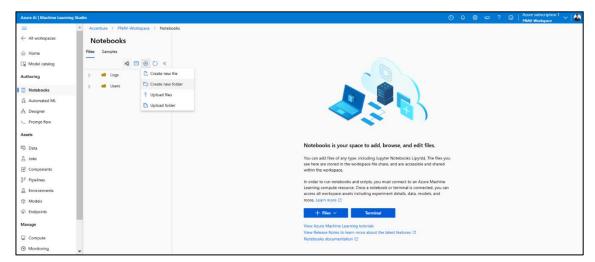






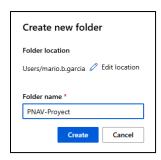
4.3 Crear un notebook

1) En la consola del workspace, en la sección "Notebooks", haz click en "+" y después en "Create new Folder". En "Folder name", pon un nombre para la carpeta donde se guardarán los ficheros del proyecto y haz click en "Create".

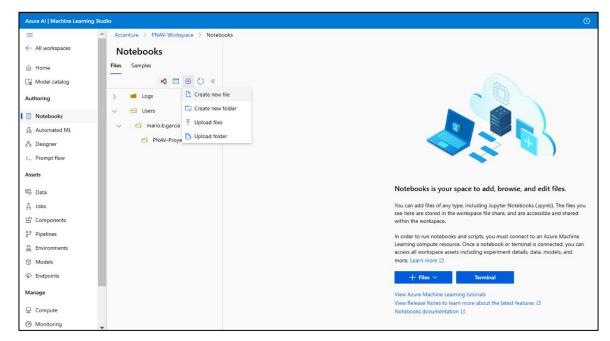




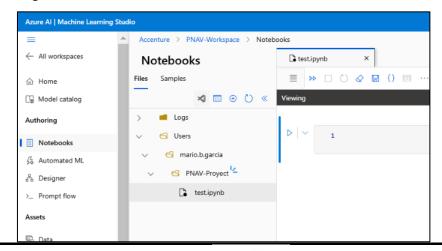




2) Una vez creada la carpeta del proyecto, haz click en "+" y después en "Create new file". En "File name", pon un nombre al documento. Selecciona "Notebook (*.ipynb)" en "File type" y haz click en "Create".



3) Haz doble click en el nombre del notebook para abrirlo. En la barra superior del notebook, escoge la instancia de computación creada y haz click en el icono de "Start compute" para ejecutar el código en ella.









El proceso de arranque puede tardar de varios segundos a un par de minutos en función de la demanda.

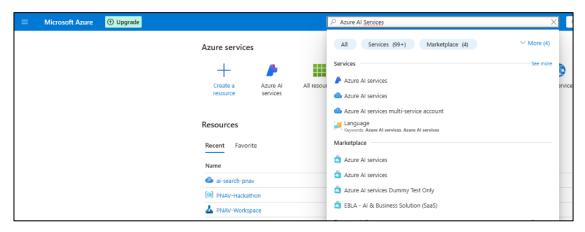
4) Cuando la máquina haya arrancado, se podrá ver el siguiente mensaje:



4.4 Creación de una instancia de Azure Al Services

Para poder desarrollar los casos de uso de RAG y TAG, es necesario crear primero, una instancia de Azure Al Services. Para más información sobre qué es una instancia de computación, consultar <u>este enlace</u>. Para poder crear la instancia, se deben seguir los siguientes pasos:

1. En la consola de Azure, en la barra búsqueda, escribe "Azure Al services" y selecciona la opción marcada.

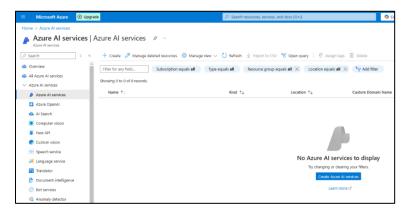






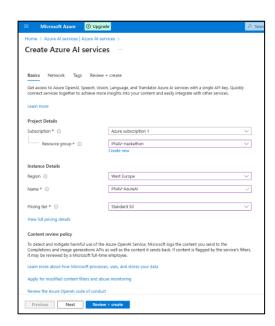


2. Haz click en "Create Azure Al Services".



- 3. Configura la instancia de Azure Al Services. Para ello:
 - Subscription: Selecciona la opción por defecto
 - Resource group: Selecciona el mismo que hayas utilizado para el resto de los recursos, en este caso "PNAV-Hackathon"
 - Region: Selecciona la misma región que hayas utilizado anteriormente:
 West Europe
 - Name: Asigna un nombre a la instancia, puedes poner "PNAV-AzureAI"
 - Pricing tier: selecciona la opción por defecto

Haz click en "Review + create"

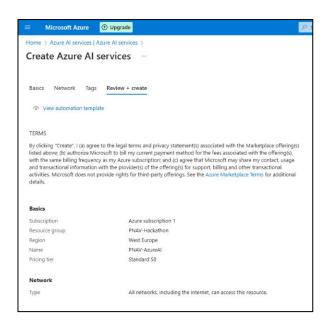


4. A continuación, se someterá a la instancia a un proceso de validación que durará unos segundos. Cuando la validación termine, haz click en "Create".

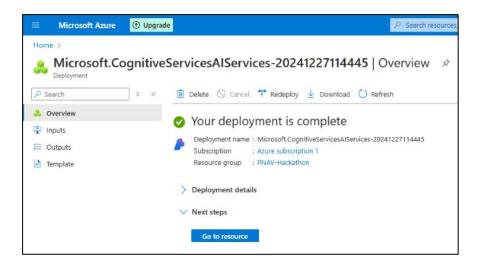








- 5. Cuando el proceso de creación termine, podrás ver el siguiente mensaje: Para poder acceder al recurso, haz click en "Go to Resource".
 - 6. Si se hace click en "Go to Resource" se podrá ver la consola principal del servicio.



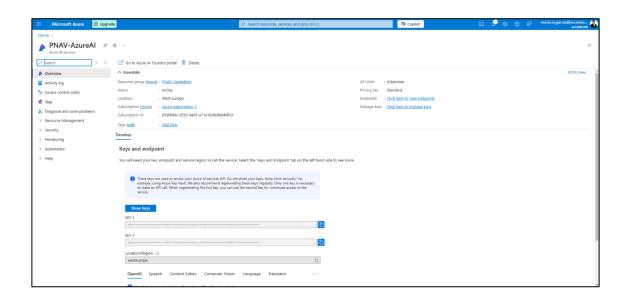




4.5. Creación de una instancia de Azure Al Search

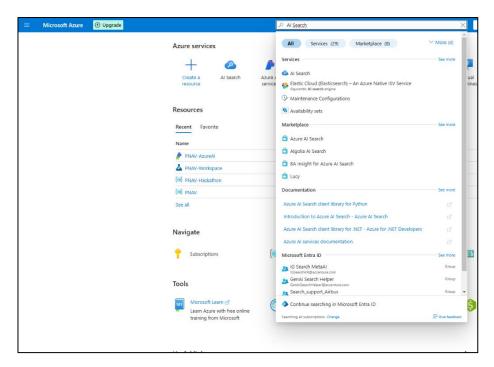
Si se quiere diseñar una solución con una arquitectura basada en RAG, será necesario crear una instancia de Azure Al Search para poder almacenar los documentos y embeddings.

1) En la consola de Azure, en la barra búsqueda, escribe "Al Search" y selecciona la opción marcada.

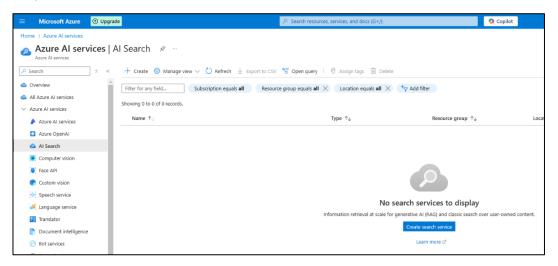








2) Después, haz click en "Create search service"

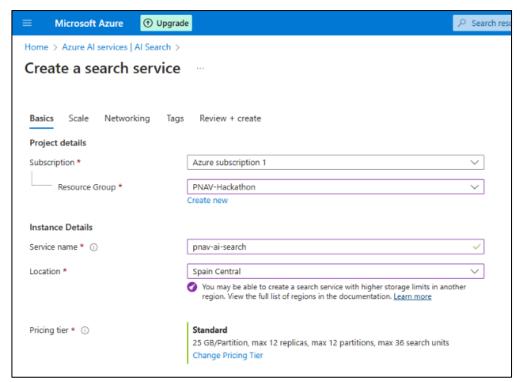


- 3) Configura la instancia de Azure Al Services. Para ello:
 - Subscription: Selecciona la opción por defecto
 - Resource group: Selecciona el mismo que hayas utilizado para el resto de los recursos, en este caso "PNAV-Hackathon"
 - Region: Selecciona la misma región que hayas utilizado anteriormente.
 - Name: Asigna un nombre a la instancia, puedes poner "pnav-ai-search"

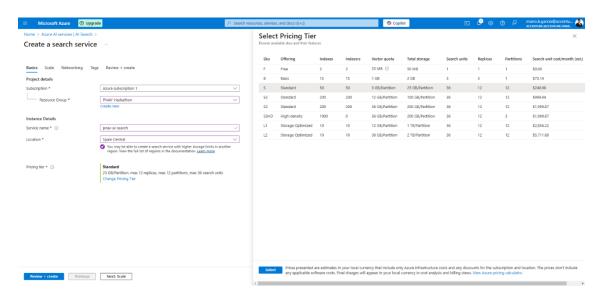








1) Haz click en "Change Pricing Tier" para ver las diferentes tarifas. Se abrirá un desplegable con todas las opciones. Si se quiere utilizar el servicio para hacer pruebas, se puede utilizar la opción gratuita ("Free"), que no tiene coste asociado. En cambio, si se quiere utilizar para el desarrollo del caso de uso, lo recomendable es escoger la opción básica ("Basic") que tiene un coste aproximado de \$75 al mes y los recursos que ofrece son suficientes. Selecciona la opción que desees y haz click en "Select".



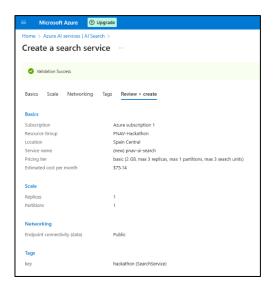
2) En el resto de las pestañas se puede configurar la instancia. Escoge los parámetros que más se ajusten a tu caso de uso.







- a. Scale: puedes modificar tanto el número de réplicas como las particiones. Cuidado, aumentarlas incurrirá en mayores costes.
- b. Networking: puedes escoger si quieres que la contectividad al endpoint sea pública o privada
- c. Tags: puedes incluir la etiqueta "key": "hackathon"
- 3) Una vez configurado, si se hace click en la sección "Review + create" la instancia pasará por un breve proceso de validación. Una vez validado se podrá ver el mensaje "Validation Success".
- 5) Una vez validado, haz click en "Create".
 - 6) Cuando ha finalizado el despliegue podrás ver la siguiente pantalla.









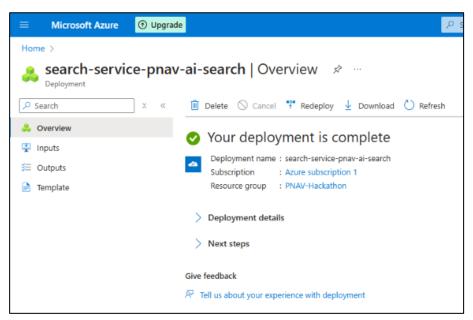
4.6. Implementación de Code-Carbon en la solución

Code-Carbon estima las emisiones de CO₂ generadas por el uso de recursos computacionales, mediante la medición del consumo energético de un programa. Los participantes deberán utilizar Code-Carbon durante todo el desarrollo para cuantificar las emisiones de CO₂ de las soluciones desarrolladas.

Para ello, deberán seguir los pasos descritos en el notebook 203 codecarbon implementation.ipynb en el repositorio GitHub del hackathon.

Todas las ejecuciones de código se deberán cuantificar con code-carbon. Los equipos deberán reportar periódicamente las emisiones generadas.

4.7. Petición de ampliación de cuota para GPUs



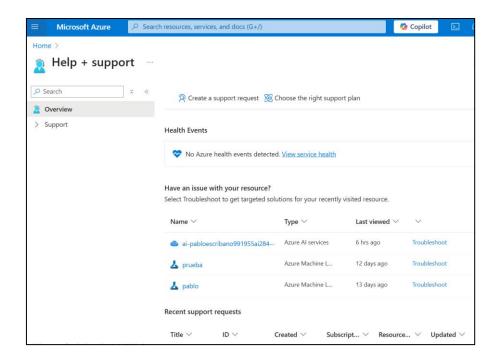
El hackathon está planteado para implementar soluciones con LLMs ligeros (2 y 7 b de parámetros) por lo tanto, cualquier solución debería poder ser implementada utilizando las instancias de CPU habilitadas en el programa. Sin embargo, en caso de que el caso de uso no pueda ser desarrollado técnicamente mediante las instancias de CPUs disponibles en el programa de Founders, se puede solicitar una ampliación de cuota para tener acceso a un clúster de GPUs del programa.

Para ello, hay que abrir un ticket desde la sección Help + support.

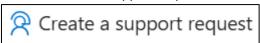








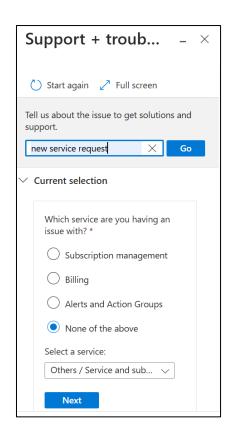
Una vez en este servicio, pulsa en "Create a support request":



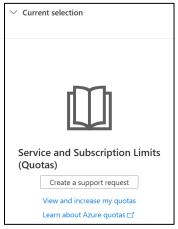
Se te abrirá la siguiente pestaña:







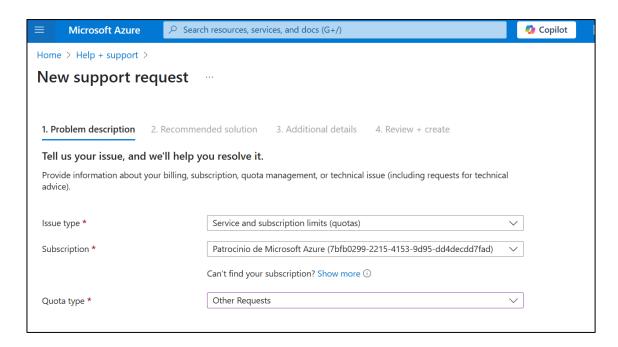
Una vez aquí, busca new service request, y selecciona la opción "None of the above". En Select a service, selecciona "Service and Suscription limits (quotas)". Pulsa a Next, y se te abrirá más abajo la opción de crear la petición en la siguiente pantalla:



Pulsa en "Create a support request", y se te abrirá una nueva pantalla con la creación del ticket:







Rellena los campos de la misma forma que en la imagen. Finalmente, en el siguiente paso, explica por qué técnicamente no se puede realizar con CPUs, orientándolo a la start up descrita en la solicitud de acceso del programa de Founders. Con ello se creará un ticket, y se pondrán en contacto contigo. Te pedirán lo siguiente:

- GPU VM SKU: El código de la instancia de GPU que solicitas. En estos links se encuentra más información sobre estas instancias: <u>Virtual machine sizes overview - Azure Virtual</u> <u>Machines | Microsoft Learn</u> y <u>Understand compute targets - Azure Machine Learning |</u> <u>Microsoft Learn</u>
- New limit: El límite de cuota que necesitéis para vuestra solución.
- Region: (Europe West or Europe North). Seleccionad Europe West para menor latencia.

5. Desarrollo de un caso de uso utilizando Fine-Tuning

5.1 Descripción del caso de uso

La finalidad de este caso de uso será entrenar modelos personalizados de IA generativa a partir de modelos preexistentes mediante *fine-tuning*. Los equipos podrán escoger cualquier modelo *open source* disponible en en Hugging Face, teniendo en cuenta el tamaño de cada uno. Se recomendará no utilizar LLMs de más de 7b de parámetros.

Los equipos dispondrán de un ejemplo en el git proporcionado, en el siguiente enlace.







5.2 Servicios necesarios en Azure

Para poder desarrollar una POC de *fine-tuning* los equipos deberán utilizar, como mínimo **Azure Machine Learning Studio** para crear y utilizar los notebooks e instancias de computación necesarias. Los notebooks deberán tener acceso a internet para poder instalar librerías y cargar modelos de Hugging Face.

Se valorará el uso de servicios o recursos adicionales disponibles.

Se valorará el uso óptimo de los recursos. Para tareas de ejecución de código básicas, se recomienda utilizar instancias de CPU. Para tareas de computación más exigentes, como entrenamiento, reentrenamiento o inferencia de modelos más pesadas, se deberá utilizar una CPU según los requisitos.

6. Requisitos para desarrollar una PoC de RAG

6.1 Descripción del caso de uso

El caso de uso consiste en la implementación de sistemas que combinen capacidades de recuperación de información con generación de texto para ofrecer respuestas mejoradas basadas en un corpus de datos específico. Los participantes deberán utilizar modelos de Hugging Face, no los disponibles en el catálogo de Azure.

Los equipos dispondrán de un ejemplo en el git propocionado, en el <u>siguiente</u> enlace.

6.2 Servicios necesarios en Azure

Para poder desarrollar una POC de RAG los equipos podrán utilizar:

- Azure Machine Learning Studio, para crear y utilizar los notebooks e instancias de computación necesarias. Los notebooks deberán tener acceso a internet para poder instalar liberías y cargar modelos de Hugging Face.
- **Azure ML Services**, para utilizar Azure Al Search, guardar los datos y hacer las búsquedas.

Se valorará el uso de servicios o recursos adicionales disponibles:

- Bases de datos vectoriales externas como ChromaDB o Pinecone
- Contenedores de archivos como Blob Storage
- Etc.

Se valorará el uso óptimo de los recursos. Para tareas de ejecución de código básicas, se recomienda utilizar instancias de CPU.







7. Requisitos para desarrollar una PoC de TAG

7.1 Descripción del caso de uso

Desarrollar un caso de uso capaz de generar texto con LLMs basado en datos tabulares. Los participantes deberán utilizar modelos de Hugging Face, no los disponibles en el catálogo de Azure.

7.2 Servicios necesarios en Azure

Para poder desarrollar una POC de TAG los equipos podrán utilizar:

- Azure Machine Learning Studio, para crear y utilizar los notebooks e instancias de computación necesarias. Los notebooks deberán tener acceso a internet para poder instalar liberías y cargar modelos de Hugging Face.
- Azure ML Services, para utilizar Azure Al Search, guardar los datos y hacer las búsquedas.

Se valorará el uso de servicios o recursos adicionales disponibles:

- Bases de datos relacionales como Azure SQL Database o MySQL
- Contenedores de archivos como Blob Storage
- Etc.

Se valorará el uso óptimo de los recursos. Los equipos deberán utilizar un tipo de instancia virtual acorde a las necesidades de ejecución.

