



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

ADE

Facultad de Administración  
y Dirección de Empresas /UPV

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Análisis de viabilidad de un programa simple de apoyo al  
desarrollo de proyectos de ciencia de datos

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Núñez Molina, Iván

Tutor/a: Grau Gadea, Gonzalo Francisco

CURSO ACADÉMICO: 2022/2023

# Resumen

El Trabajo de Fin de Grado aquí presentado realiza un análisis de viabilidad para estudiar el potencial de mercado de un programa informático. Este programa está destinado a actuar como herramienta de soporte para el desarrollo de proyectos de ciencia de datos (*Big Data*), especialmente para estudiantes universitarios o, más generalmente, individuos interesados en iniciarse en la ciencia de datos y que no tienen aún los conocimientos necesarios para utilizar herramientas más complejas, almacenando los datos de forma local.

Para llegar a una conclusión, realizaremos varios análisis que cubrirán las diversas áreas de conocimiento necesarias para tomar una decisión informada, partiendo desde el macroentorno y acercando nuestro punto de vista hasta llegar a las especificaciones del producto. La importancia del estudio está magnificada por el tamaño reducido de este mercado, y por tanto la poca información disponible que hay al respecto.

Para empezar, se realizará un análisis PESTEL para determinar cuáles son los factores más importantes en el macroentorno de este producto. Veremos especialmente que la naturaleza del producto hace que algunos de estos factores sean prácticamente irrelevantes, mientras que otros son absolutamente cruciales.

A continuación, llevaremos a cabo un análisis de las 5 fuerzas de Porter, con el fin de acercarnos más al producto y descubrir qué agentes específicos del mercado pueden suponer un peligro para el éxito de nuestro programa informático.

Finalmente, se hará un análisis de necesidades seguido de una propuesta concreta apoyada en este análisis. Este proceso se llevará a cabo tanto para el contexto actual como para el futuro, a fin de trazar una breve hoja de ruta para el desarrollo continuado de este proyecto.

**Palabras clave:** Análisis de viabilidad, Big Data, Ciencia de datos, Educación, Programa informático

---

# Abstract

The Final Degree Project presented here carries out a feasibility analysis to study the market potential of a computer program. This program is intended to act as a support tool for the development of data science (*Big Data*) projects, especially for university students or, more generally, individuals interested in getting started in data science and who do not yet have the knowledge needed to use more complex tools, storing data locally.

To reach a conclusion, we will carry out several analyses that will cover the various areas of knowledge needed to make an informed decision, starting from the macroenvironment and bringing our point of view closer to reaching the product specifications. First of all, we will do a superficial study of the market to get an idea of its current state. The importance of this study is magnified by the small size of this market, and therefore the little information available about it.

Next, a PESTEL analysis will be carried out to determine which are the most important factors in the macroenvironment of this product. In particular, we will see that the nature of the product makes some of these factors practically irrelevant, while others are absolutely crucial.

We will also carry out a Porter's 5 forces analysis, in order to get closer to the product and discover which specific market players may pose a threat to the success of our software.

Finally, a needs analysis will be carried out followed by a concrete proposal supported by this analysis. This process will be carried out both for the current context and for the future, in order to draw a brief roadmap for the continued development of this project.

**Key words:** Feasibility analysis, Big Data, Data science, Education, Software

---

# Índice general

---

<b>Índice general</b>	<b>3</b>
<b>Índice de figuras</b>	<b>5</b>
<b>Índice de tablas</b>	<b>5</b>
<hr/>	
<b>1 Introducción</b>	<b>1</b>
1.1 Motivación . . . . .	1
1.2 Objetivos . . . . .	2
<b>2 Análisis del macroentorno (PESTEL)</b>	<b>3</b>
2.1 Factores políticos . . . . .	3
2.2 Factores económicos . . . . .	4
2.3 Factores sociales . . . . .	6
2.4 Factores tecnológicos . . . . .	7
2.5 Factores ambientales . . . . .	9
2.6 Factores legales . . . . .	10
<b>3 Análisis del microentorno (5 fuerzas de Porter)</b>	<b>13</b>
3.1 Competencia actual . . . . .	13
3.2 Competencia potencial . . . . .	15
3.3 Productos sustitutivos . . . . .	16
3.4 Poder de negociación de los proveedores . . . . .	16
3.5 Poder de negociación de los clientes . . . . .	17
<b>4 Análisis de necesidades</b>	<b>19</b>
4.1 Resumen DAFO . . . . .	19
4.1.1 Debilidades . . . . .	19
4.1.2 Amenazas . . . . .	20
4.1.3 Fortalezas . . . . .	21
4.1.4 Oportunidades . . . . .	21
4.2 Necesidades actuales . . . . .	22
4.3 Necesidades futuras . . . . .	22
<b>5 Propuesta de estrategia</b>	<b>25</b>
5.1 Estrategia actual . . . . .	25
5.2 Estrategia futura . . . . .	26
5.2.1 Éxito moderado . . . . .	26
5.2.2 Éxito mayor . . . . .	27
<b>6 Conclusiones</b>	<b>29</b>
<b>Bibliografía</b>	<b>31</b>
<hr/>	
Apéndice	
<b>A Objetivos de Desarrollo Sostenible</b>	<b>35</b>
A.1 Relación con ODS . . . . .	35
A.2 Reflexión . . . . .	35



## Índice de figuras

---

2.1	Inversión privada en IA por sector, 2021 Vs. 2022 (NetBase Quid, 2022) . . .	5
2.2	Inversión privada en IA, 2013 a 2022 (NetBase Quid, 2022) . . . . .	5
2.3	Comparativa de la misma foto, sin y con el borrador mágico de Google (elaboración propia) . . . . .	9
3.1	Instancias de IA introducidas en al menos un sector de la empresa, 2018 a 2022 (McKinsey & Company Survey, 2022) . . . . .	16
5.1	Diagrama explicativo de la metodología ágil [Santa Lucía Impulsa, 2021] .	26

## Índice de tablas

---

2.1	Inversión privada en IA, 2013 a 2022 (NetBase Quid, 2022) . . . . .	6
3.1	Comparativa de competidores (características) (elaboración propia) . . . .	14
3.2	Comparativa de competidores (valoración ponderada) (elaboración propia)	14
3.3	Comparativa de competidores (fechas de salida) (elaboración propia) . . .	15
4.1	Tabla de análisis DAFO (elaboración propia) . . . . .	19
A.1	Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS) (elaboración propia) . . . . .	35



---

---

# CAPÍTULO 1

## Introducción

---

Con la II<sup>o</sup> Guerra Mundial en Europa, se desarrollaron lo que después se conocería como los primeros ordenadores electrónicos de la historia: el Colossus británico, utilizado durante la guerra para descifrar códigos alemanes (entre ellos el famoso código Enigma), y el ENIAC estadounidense, presentado en 1946 y financiado por el ejército. Ambos eran aparatos masivos, que ocupaban una habitación entera y requerían de personal entrenado para ser utilizados. La industria de la computación llegó al gran público en los años 70 con el desarrollo del primer ordenador personal, y pasó gradualmente de acercarse tímidamente a nuestras casas a estar presente en la gran mayoría de ámbitos de nuestra vida bajo la forma de dispositivos electrónicos como el teléfono móvil, la lavadora programable, o incluso el microondas.

Casi desde su concepción, la industria informática ha estado mitificada como rompedora y radical, llevando constantemente innovaciones tecnológicas a los usuarios. Este mito ha sido perpetuado por la propia industria con lugares como Silicon Valley, que es casi un lugar sagrado para la tecnología a nivel mundial, o personas como Steve Jobs, prácticamente reverenciados como gurús de la innovación y los negocios.

Sin embargo, lo cierto es que hace tiempo que la industria de la computación pasó de ser el emergente *underdog* para convertirse en la más importante del mundo: a fecha de 22 de abril del 2022, 6 de las 10 empresas con mayor capitalización bursátil en el mundo eran tecnológicas, recogiendo casi dos tercios (8.379,27 de 13.301,71 miles de millones de dólares) del capital bursátil de estas 10 empresas [[Statista Research Department, 2023](#)]. Es innegable que este sector ha alcanzado cierta madurez, lo que conlleva un crecimiento más lento, si bien la pandemia del COVID-19 prolongó brevemente la época de crecimiento masivo de las tecnológicas debido a las necesidades particulares que generó. Esta madurez implica que el mercado está más saturado con nuevos productos y servicios, que incrementan la competitividad: las innovaciones propuestas son mayoritariamente incrementales en lugar de radicales, y es difícil ver a las empresas líderes del sector como Apple o Microsoft perdiendo ese liderazgo en el futuro cercano.

Esta normalización de la industria tecnológica nos permite acercarnos a ella de la misma forma en que lo haríamos con una industria más tradicional, utilizando herramientas como el análisis PESTEL o las 5 fuerzas de Porter para establecer la viabilidad de un producto, por ejemplo, de *software*.

### 1.1 Motivación

---

Como estudiante de Ingeniería Informática además de ADE, tengo un gran interés en el desarrollo *software*: he hecho mis pequeños proyectos por mi cuenta, y he empleado



muchos de los conocimientos de informática que he aprendido en la carrera. Sin embargo, también he aprendido en la carrera de ADE que a la hora de sacar un proyecto serio al mercado, el propio producto es sólo parte del trabajo: no basta con tener una buena idea, hay que saber cómo encauzarla, desarrollarla y exponerla al mercado para que sobreviva y tenga éxito.

He decidido aprovechar la oportunidad que me presenta el Trabajo de Fin de Grado para poner en práctica estos conocimientos, y llevar a cabo el proceso que rodea el lanzamiento de un nuevo producto de software de principio a fin.

## 1.2 Objetivos

---

Este trabajo tiene como objetivo averiguar si un programa de apoyo al desarrollo de proyectos de ciencia de datos sería viable en el mercado informático actual. El propósito de este programa es facilitar el trabajo de los desarrolladores, proporcionándoles herramientas útiles que les ahorren tiempo. Además, este programa estará específicamente enfocado a estudiantes y principiantes: por tanto, reducimos nuestro público objetivo para ofrecer un producto mejor adaptado a sus necesidades.

Cabe destacar que el producto analizado en este trabajo no está enfocado al público general, si no que es una herramienta destinada a ser utilizada para crear otros productos (por ejemplo, crear un modelo de predicción de mercados para carteras de negocio). De esta manera, al reducir nuestros clientes, acotamos también nuestros potenciales competidores.

---

---

## CAPÍTULO 2

# Análisis del macroentorno (PESTEL)

---

El análisis PESTEL busca realizar un análisis sistemático de los factores más importantes que pueden afectar al producto, sin tener en cuenta el propio mercado del producto. A esto se le refiere como el "macroentorno", y puede ser determinante a la hora de prever el éxito o el fracaso de nuestro programa. Estos factores son el político, económico, social, tecnológico, ambiental y legal.

### 2.1 Factores políticos

---

La informática moderna ha realizado avances en potencia de procesamiento impensables hace apenas dos décadas; en una de esas situaciones en las que la herramienta precede al uso, hemos empleado esta potencia para alimentar proyectos de ciencia de datos que serían imposibles de llevar a cabo hace sólo 10 años. Los grandes modelos de Inteligencia Artificial (una de las aplicaciones más populares de la ciencia de datos, aunque no la única) como el reciente ChatGPT-3 requieren de un *hardware* de muy alto nivel, e incluso un principiante puede tener problemas para proyectos pequeños si intenta utilizar su ordenador personal para ello. Por tanto, un *hardware* avanzado y moderno es necesario para estar a la altura de estos requerimientos.

Por estas razones, el avance general de la ciencia de datos depende en gran medida de la disponibilidad de este *hardware*, y a su vez este *hardware* está condicionado por la industria de los semiconductores, que llevan varios años siendo el cuello de botella en su fabricación. La escasez de estos componentes, esenciales en la fabricación de cualquier componente electrónico, ya ha estado dando problemas a industrias tan grandes como la automovilística, que ha tenido que realizar parones en su producción varias veces en los últimos años [Cordero, 2021]. Es entonces cuando entran en juego las tensiones geopolíticas en el sudeste asiático: la empresa taiwanesa TSMC (Taiwan Semiconductor Manufacturing Company) lleva varios años poseyendo más del 50 % del mercado de los semiconductores [Alsop, 2023], lo que crea una dependencia enorme de la industria informática en esta única empresa. Esta dependencia ha convertido a la empresa en un importante activo político para Taiwán, ya que convierte a la isla en un agente esencial para la economía mundial, una influencia que el gobierno taiwanés aprovecha para protegerse de la agresión indirecta de la República Popular China [Woo & Blanchard, 2022]. Estas tensiones políticas suponen un riesgo para la provisión mundial de semiconductores, que puede fluctuar mucho según los eventos futuros, condicionar los avances en ciencia de datos, y por tanto afectar al éxito de nuestro producto.

Por otro lado, la fragilidad de la industria de los semiconductores se ha empezado a considerar como un asunto de interés nacional e internacional, y los otros dos polos del poder mundial, Estados Unidos y Europa, han comenzado a actuar para paliar esta fragilidad.

Por la parte de los Estados Unidos, el gobierno estadounidense ha anunciado una inversión privada de más de 80.000 millones de dólares en manufacturación local de semiconductores por parte de varias empresas de talla mundial como Intel o Samsung [The White House, 2021], aunque el largo plazo de estas inversiones pone en duda su fiabilidad: Estados Unidos salió en 2020 del gobierno republicano de Donald Trump, durante el cual se establecieron medidas proteccionistas que dificultaron las exportaciones del país; en sus propias palabras, *"We should never be reliant on a foreign country for the means of our own survival. [...] This crisis has underscored just how critical it is to have strong borders and a robust manufacturing sector. [...] Our goal for the future must be to have American medicine for American patients, American supplies for American hospitals, and American equipment for our great American heroes."* [Trump White House, 2020]. Estas declaraciones de un presidente de los Estados Unidos plantean dudas a la hora de garantizar que este país vaya a ser una fuente estable de semiconductores en el mercado global.

Sin embargo, la Unión Europea también ha centrado sus esfuerzos en resolver este problema: la Comisión Europea ha puesto en marcha una iniciativa a nivel internacional que busca impulsar el desarrollo de la industria de los semiconductores, a través de la Ley Europea de Chips [Comisión Europea, DG CNECT, 2022]. Esta ley tiene como objetivo potenciar la investigación en tecnologías punteras de microchips e incrementar la inversión en fabricación local, como dijo la presidenta de la Comisión Europea, Ursula von der Leyen, en el Discurso sobre el estado de la Unión de 2021: *"The aim is to jointly create a state-of-the-art European chip ecosystem, including production. That ensures our security of supply and will develop new markets for ground-breaking European tech"* [von der Leyen, 2021]. Dada la mayor estabilidad política de la Unión Europea respecto al gobierno estadounidense, podemos considerar esta propuesta como un compromiso fiable para garantizar el suministro futuro de semiconductores.

## 2.2 Factores económicos

---

La ciencia de datos y su retoño más popular, la IA (Inteligencia Artificial) están mucho más presentes en nuestra vida diaria de lo que parece: los anuncios que nos muestra YouTube, las sugerencias de compra de Amazon, las recomendaciones de canciones de Spotify, o asistentes de voz como Siri son sólo algunas de sus aplicaciones más visibles. Sin embargo, la inteligencia artificial tiene aplicaciones que van mucho más allá de lo esperado, debido a su versatilidad y su creciente complejidad, que le permite abarcar problemas mucho más variados año tras año: el modelo GPT-2, construido en 2019 por OpenAI, fue desarrollado utilizando 1.500 millones de parámetros <sup>1</sup>. Su sucesor, GPT-3, lanzado en 2022, fue desarrollado utilizando 175.000 millones de parámetros, 117 veces más [Narayanan, y otros, 2021].

La industria de la IA es masiva y tiene presencia en multitud de sectores, con inversiones en 2022 de 6.000 millones de dólares en la industria médica, 2.500 millones en marketing, o 5.000 millones en ciberseguridad [Maslej, y otros, 2023]. Cabe destacar que la inversión en IA en la mayoría de sectores se redujo en el año 2022 respecto a 2021, pasando de una inversión total de 276.140 millones de dólares en 2021 a 189.590 millones

---

<sup>1</sup>Los "parámetros" utilizados en el desarrollo de un modelo de IA, más conocidos como "pesos", son una serie de ponderaciones estadísticas que determinan la calidad del modelo final. Cuantos más parámetros, más preciso es el modelo, pero también más cuesta de entrenar.

en 2022. Esta reducción no está distribuida equitativamente, como se observa en la figura 2.1:

**Private Investment in AI by Focus Area, 2021 Vs. 2022**

Source: NetBase Quid, 2022 | Chart: 2023 AI Index Report

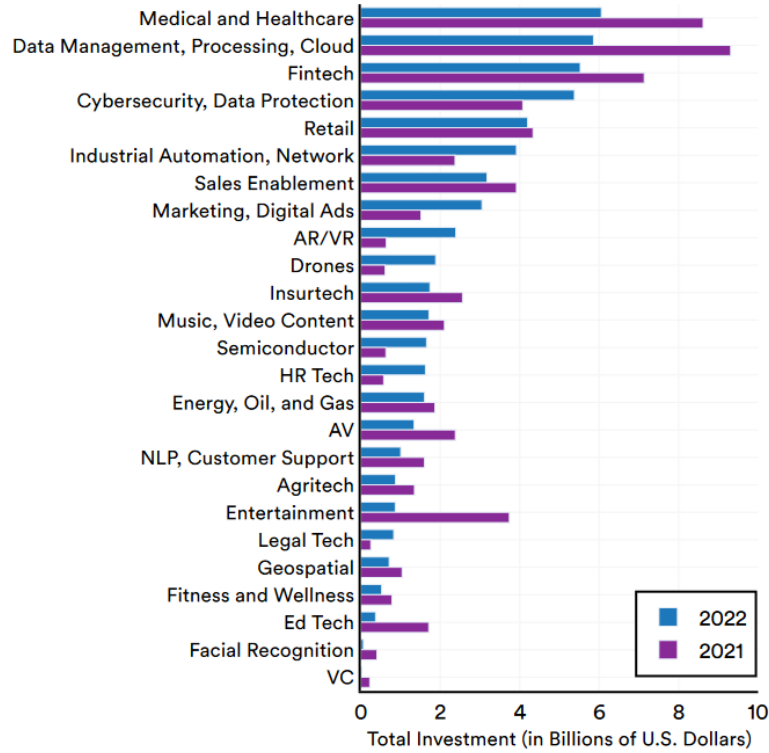


Figura 2.1: Inversión privada en IA por sector, 2021 Vs. 2022 (NetBase Quid, 2022)

Si bien algunas industrias como la médica o el entretenimiento han reducido sus inversiones, en otras como marketing o ciberseguridad se han incrementado sustancialmente. En cualquier caso, la tendencia de la última década es alcista, como se ve en la figura 2.2:

**Private Investment in AI, 2013–22**

Source: NetBase Quid, 2022 | Chart: 2023 AI Index Report

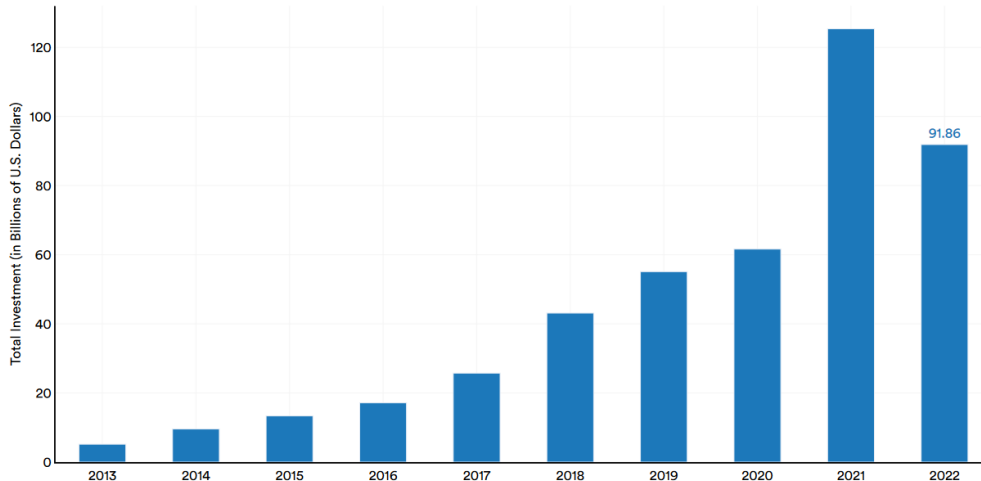


Figura 2.2: Inversión privada en IA, 2013 a 2022 (NetBase Quid, 2022)

Año	Inversión (miles de millones de \$)	Crecimiento (%)
2013	5.168545793	
2014	9.558396415	84.93
2015	13.34679382	39.63
2016	17.13074732	28.35
2017	25.72191773	50.15
2018	43.09941534	67.56
2019	55.09039152	27.82
2020	61.61122052	11.84
2021	125.3568742	103.46
2022	91.86206767	-26.72

**Tabla 2.1:** Inversión privada en IA, 2013 a 2022 (NetBase Quid, 2022)

A la vista de los datos, el crecimiento de la inversión es constante hasta llegar el 2021, cuando se duplica en tan solo un año. Podemos suponer que esto se debe al volcado de la sociedad en el uso de la tecnología durante la pandemia del COVID-19, en cuyo caso la reducción del año 2022 sería una vuelta a la normalidad. Si observamos de nuevo la figura 2.2, vemos que el año 2021 aparenta ser un dato anómalo en una tendencia creciente que se ve continuada por los datos del año 2022. Es fácil por tanto concluir que la industria de la inteligencia artificial y la ciencia de datos está en pleno auge: si bien ha estado presente en las empresas desde hace más de una década, es solo recientemente cuando se ha introducido en nuestros hogares con productos dirigidos directamente al consumidor general, incrementando así masivamente su mercado.

## 2.3 Factores sociales

Aunque las herramientas derivadas de la ciencia de datos han estado desde la concepción de esta disciplina dirigidas a un público principalmente empresarial (como por ejemplo, los anuncios personalizados de Google en base a tus búsquedas previas), recientemente se han introducido a un público más general a través de *software* complejo envuelto en una interfaz más amigable:

**ChatGPT** Un modelo creado para conversar con el usuario simulando a un humano. Ha sido entrenado con información general de internet, y es capaz de responder a casi cualquier pregunta de manera convincente. Desarrollado por OpenAI.

**Midjourney** Un modelo que genera imágenes a partir de un *prompt*, o lo que es lo mismo, una palabra/frase que describe la imagen que el usuario quiere generar. Este modelo ha sido entrenado con imágenes de internet, tanto fotografías como dibujos, y tiene la capacidad de emular estilos artísticos concretos e incluso artistas específicos.

Este tipo de aplicaciones han resultado ser tremendamente populares entre el consumidor medio, con ChatGPT alcanzando los 600 millones de visitantes únicos en enero, el mes de su lanzamiento, y 1.600 millones en marzo [Similarweb, 2023].

Sin embargo, esta explosión de popularidad ha puesto de manifiesto problemas éticos que hasta ahora se habían mantenido en la sombra respecto al tratamiento de la información. Para conseguir los niveles de precisión que convierten estos modelos en algo más que meras curiosidades, las empresas que los desarrollan necesitan cantidades inmensas de datos con los que entrenarlos. Estos datos no siempre han sido adquiridos de forma

moral; pese a que la falta de legislación al respecto no lo hace explícitamente ilegal, se ha generado una conversación sobre la ética de esta práctica que pone en duda si se debería permitir. Por ejemplo, Midjourney genera imágenes a partir de una frase. Sin embargo, para poder hacer esto, ha sido entrenada con millones de imágenes que tienen palabras similares asociadas: si utilizas como *prompt* la frase "Wood cabin in the forest", el modelo habrá aprendido que las imágenes que otras personas han creado y llamado de formas similares tienen una serie de características, y reproducirá una imagen con estas características. El planteamiento ético es el siguiente: si este modelo de IA ha podido crear una nueva imagen gracias a haber aprendido de las imágenes de un fotógrafo, ¿debería este fotógrafo recibir una retribución económica? ¿O al menos poder optar a sustraer sus fotografías del entrenamiento de estos modelos? El consenso de fotógrafos y artistas es que sí, mientras que las empresas están en contra por razones de costes y desarrollo.

Si bien los modelos conversacionales como ChatGPT no tienen estos problemas de autoría, sufre de otros muy diferentes. ChatGPT está diseñado para responder con seguridad a las preguntas de los usuarios en base a la información que ha aprendido, pero esta seguridad no siempre está justificada: en abril de 2023, Jonathan Turley, profesor universitario de Washington D.C., se quejó públicamente de que ChatGPT había creado una falsa acusación por acoso sexual sobre él: *"I learned that ChatGPT falsely reported on a claim of sexual harassment that was never made against me on a trip that never occurred while I was on a faculty where I never taught. ChatGPT relied on a cited Post article that was never written and quotes a statement that was never made by the newspaper."* [Turley, 2023]. Las implicaciones de estas "alucinaciones", como se les llama, están claras: consultas de este tipo pueden dar resultados completamente falsos, que se pueden dar por ciertos con facilidad debido a la supuesta objetividad de estos sistemas. Si bien este ejemplo es claramente falso, ya que es fácil de comprobar con una búsqueda y su severidad plantea dudas inmediatamente, este tipo de alucinaciones pueden ocurrir con hechos que pasan desapercibidos.

Por último, es importante saber que este aire de objetividad que se le da a los sistemas de inteligencia artificial como ChatGPT no es más que eso, un mito: están entrenados con datos creados por seres humanos (en este caso, texto escrito por seres humanos), y por tanto está sujeto a los mismos sesgos que los autores de estos textos: no sólo hay que prestar atención a la veracidad de lo que nos dice el sistema, sino que también hay que ser consciente de que el desarrollador del sistema puede haberlo creado con ciertos sesgos ideológicos, ya sea de forma intencionada o no. Todos estos factores de riesgo han contribuido a crear un ambiente de desconfianza alrededor de estos modelos, que si bien tienen capacidades fascinantes, conllevan muchos riesgos en su uso que deben ser paliados antes de que se considere como una tecnología fiable.

## 2.4 Factores tecnológicos

---

El sector de la ciencia de datos ha surgido como una subsección de la industria informática, que debido a su creciente demanda se ha independizado a su propia industria (aunque el solapamiento entre ambas es considerable). Esto se debe a que el desarrollo de nuevas herramientas con una base en la ciencia de datos se ha convertido en la nueva gallina de los huevos de oro para las grandes empresas tecnológicas, que buscan convertirse en pioneras y coger la delantera en este mercado creciente.

Algunas de las aplicaciones de esta tecnología, junto a varias empresas que las emplean, son:

**Perfil de clientes** El gran negocio de la última década, el tratamiento y negocio con datos de clientes, se ha convertido en tal gracias al desarrollo de sistemas de IA. Siste-

mas como las famosas *cookies* son como las empresas recaban información sobre sus clientes; sin embargo, después estos datos son tratados con herramientas de *software* para convertirse en información utilizable, que después se emplea para personalizar publicidad, recomendaciones de compra, o directamente para ser vendida a otras empresas que la utilizaran para estos objetivos.

EMPRESAS: Amazon, Google.

**Asistentes de voz** Los asistentes de voz llevan varios años siendo un estándar en los teléfonos inteligentes de media-alta gama, con Siri como su ejemplo más famoso, pero a mediados de los 2010 se popularizaron los "*home assistants*", o asistentes de hogar, cuyas instancias más populares han sido con diferencia Amazon Alexa y Google Assistant. Estos productos utilizan inteligencia artificial para reconocer los comandos del usuario y tomar acciones.

EMPRESAS: Amazon, Apple, Google.

**Toma de decisiones** La tendencia de la dirección empresarial en el último siglo ha ido de métodos mayoritariamente intuitivos que confían en el criterio de expertos, a utilizar grandes cantidades de datos para tomar decisiones más informadas. La IA permite tratar estos datos con facilidad y llegar a conclusiones que cumplan los objetivos de la empresa de la forma más eficaz posible. Una industria en la que se ha popularizado esto es en la inversión en bolsa, con sistemas de inteligencia artificial controlando portafolios completos.

EMPRESAS: Q.ai<sup>2</sup>.

**Predicción de eventos** La aplicación insignia de la inteligencia artificial en las aulas universitarias es la predicción de eventos: a partir de un conjunto de datos, un algoritmo extrae patrones de conducta y es capaz de predecir con cierta precisión si, dados nuevos datos, ocurrirá cierto evento concreto o no. Esto es aplicable a muchas áreas, como la agricultura, donde la predicción del tiempo puede ser esencial para proteger o aprovechar las cosechas.

EMPRESAS: Envision Digital<sup>3</sup>, EOS Data Analytics<sup>4</sup>.

**Edición fotográfica** La edición de imágenes también se ha beneficiado del desarrollo de la inteligencia artificial, creando todo tipo de soluciones de selección de elementos, coloreado inteligente y multitud de opciones más. Herramientas como el "Borrador mágico" de Google en sus teléfonos permiten seleccionar manualmente elementos en una foto, y dejar que el sistema los sustituya por una amalgama de los elementos de su alrededor. El resultado sin embargo no es un borrón de colores, sino (generalmente) un relleno inteligente que concuerda con el entorno. Se puede ver un ejemplo en la figura 2.3.

EMPRESAS: Adobe, Google.

---

<sup>2</sup><https://learn.tryq.ai/blog/how-q-ai-uses-ai>

<sup>3</sup><https://www.envision-digital.com/enweather-agriculture/>

<sup>4</sup><https://eos.com/products/crop-monitoring/>



(a) Sin borrador.

(b) Con borrador.

**Figura 2.3:** Comparativa de la misma foto, sin y con el borrador mágico de Google (elaboración propia)

**Edición de vídeo en directo** La pandemia del COVID-19 ha popularizado las plataformas de comunicación por voz y vídeo. En un esfuerzo por tomar la delantera en el mercado y utilizando una tecnología similar a la edición fotográfica, una de las características adicionales de estas plataformas es la posibilidad de modificar tu fondo cuando estás en una videollamada con una imagen personalizada. Esto se consigue también a través de inteligencia artificial, que está entrenada para reconocer el contorno de una persona y recortar toda la imagen a su alrededor, para sustituirla por la imagen seleccionada. Esta tecnología también es muy popular en, por ejemplo, Instagram, donde se pueden utilizar multitud de filtros que se superponen al rostro del usuario.

EMPRESAS: Meta, Microsoft, Zoom.

Todas estas diferentes aplicaciones de la tecnología muestran la flexibilidad de la inteligencia artificial, y nos da una idea del futuro que tiene. Su potencial es tanto grande como variado: podemos suponer que estas herramientas continuarán popularizándose, con lo que más y más personas se interesarán por este sector.

## 2.5 Factores ambientales

---

Uno de los factores más menospreciados en el desarrollo de modelos de inteligencia artificial es su impacto medioambiental: como se ha mencionado anteriormente, los modelos de IA de las grandes empresas se alimentan de cantidades masivas de datos, y el procesamiento de estos datos requiere cantidades igualmente masivas de energía. En 2020, el consumo de energía del conjunto de centros de tratamiento de datos en Estados Unidos fue de más de 70.000 millones de KWh<sup>5</sup> [Shehabi, y otros, 2016].

<sup>5</sup>Esta cifra aglutina el consumo de todos los centros de datos de Estados Unidos, no sólo los que están dedicados al entrenamiento de modelos de inteligencia artificial.



Además, el procesamiento de estos datos consume grandes cantidades de agua: se estima que el entrenamiento de un modelo LaMDA (un modelo de Google) consume entre 500.000 y 2.000.000 de litros de agua, dependiendo tanto de la ubicación de los centros de datos como del periodo del año [Li, Yang, Islam, & Rei, 2023]. Esto se debe a que la mayoría de este consumo proviene de la refrigeración de los sistemas de *hardware*, que se sobrecalientan debido al gran consumo energético mencionado anteriormente. Estas cifras pueden reducirse realizando una gestión inteligente y medio-ambientalmente consciente de los recursos, por ejemplo entrenando estos modelos en invierno, cuando la temperatura natural minimiza el sobrecalentamiento. Sin embargo, no sólo el entrenamiento tiene un coste, sino también el uso regular: se estima que, para procesar entre 20 y 50 preguntas, ChatGPT necesita consumir alrededor de 500 mL de agua [Gendron, 2023].

El impacto medioambiental de estos sistemas se está convirtiendo en una preocupación creciente, debido a la escala que están alcanzando. Los efectos del cambio climático ya son perceptibles en prácticamente todo el mundo, y si bien la industria de los datos y la inteligencia artificial está todavía en crecimiento, podría verse limitada por su consumo de recursos.

## 2.6 Factores legales

---

El auge en los años recientes de los sistemas de inteligencia artificial de uso general ha conllevado un incremento de la legislación al respecto, en un esfuerzo por regular su funcionamiento y prevenir/solucionar los problemas que conlleva la concepción de un nuevo paradigma de desarrollo tecnológico. Como parte de la Agenda España Digital 2026, el gobierno español ha lanzado la Estrategia Nacional de Inteligencia Artificial (ENIA). La ENIA tiene como objetivo potenciar el desarrollo de la industria nacional de ciencia de datos, y simultáneamente *"Preparar a España para las transformaciones socioeconómicas que origina la IA"* a través de legislación [MINECO, s.f.].

La ENIA es una propuesta gubernamental que busca establecer un plan de desarrollo de la inteligencia artificial a nivel nacional en varios ámbitos, tanto económicos y sociales, como legales. Una de sus metas es establecer un marco regulador que permita el desarrollo de esta industria a la vez que garantiza los derechos de los españoles, adaptando las protecciones con este fin que ya existen a este nuevo paradigma tecnológico: *"Todo sistema de Inteligencia Artificial, debe respetar los derechos fundamentales, equidad en el acceso y prevención contra la discriminación, así como también todos los derechos del mundo 'analógico'."* [MINECO, 2020].

Además de legislación específica sobre inteligencia artificial, España también ha puesto en marcha la Oficina del Dato; recordemos que, si la inteligencia artificial es el cerebro, los datos son los cientos y miles de experiencias que le permiten aprender y por tanto, actuar. Esta oficina tiene como objetivo armonizar y facilitar el uso de los datos de los consumidores entre los diferentes sectores industriales del país, a la vez que vela por el cumplimiento de la ley en el trato de estos datos [MINECO, s.f.].

Además del nivel nacional, también se han puesto en marcha iniciativas a nivel internacional: la Comisión Europea ha manifestado su intención de crear un marco regulador propio que permita a los países miembros de la Unión Europea introducir una legislación cohesiva, que proteja los intereses de sus ciudadanos de forma uniforme a lo largo del territorio. A través de estas medidas, la Comisión Europea espera que la UE se convierta en el polo económico líder a nivel mundial, tanto en desarrollo de sistemas de inteligencia artificial como en regulación y legislación [Comisión Europea, 2020].

Es evidente que la dirección general de los diferentes actores gubernamentales es la misma: establecer una regulación que proteja los intereses de los ciudadanos, a la par que reconoce el enorme potencial tanto económico como social de la inteligencia artificial y fomenta un desarrollo responsable de esta.



---

---

## CAPÍTULO 3

# Análisis del microentorno (5 fuerzas de Porter)

---

Mediante un análisis de las 5 fuerzas de Porter, estamos descendiendo a un nivel de observación más cercano al producto, y buscando dentro del mercado qué factores pueden suponer una amenaza a nuestro programa. Para esto, analizaremos horizontalmente qué empresas pueden suponer una amenaza para nosotros (tanto en el presente como en el futuro) y qué productos podrían reemplazar al nuestro, y analizaremos verticalmente la influencia que pueden ejercer sobre nosotros tanto los proveedores como los clientes.

### 3.1 Competencia actual

---

Para analizar la competencia actual de nuestro producto, primero debemos definir claramente nuestro producto. El software que vamos a desarrollar es una plataforma que proporciona al desarrollador de ciencia de datos las herramientas necesarias para sus proyectos. Algunas de estas herramientas y características, en las que basaremos una comparativa, son las siguientes:

- Registro de modelos: permite llevar un control más ajustado de los diferentes modelos (proyectos) en los que está trabajando.
- Visualización de métricas: proporciona una interfaz adaptada para visualizar las métricas que determinan el éxito o fracaso del modelo, frecuentemente incluyendo tablas y gráficos.
- Control de versiones: permite llevar un control exhaustivo de las diferentes versiones de un proyecto, al igual que de los factores que llevan a los resultados de una versión (esencial para garantizar la reproducibilidad de los resultados).
- Open source: acceso libre al código fuente de la plataforma.
- Seguimiento en producción: recaba información del rendimiento del modelo incluso después del entrenamiento, cuando ya ha sido lanzado y está siendo activamente utilizado.
- Reportes integrados: ofrece una herramienta que genera automáticamente reportes de rendimiento adaptados al ojo humano.
- Pipelines: permite reproducir todos los pasos del entrenamiento de un modelo con un solo clic, facilitando el proceso.

Una vez definidas las áreas de comparación, procedemos a buscar cómo los principales programas del mercado se desempeñan en estas áreas a través de los portales web de estas herramientas.

CARACTERÍSTICAS	Iterative Studio <sup>1</sup>	MLflow <sup>2</sup>	Comet ML <sup>3</sup>	Weights & Biases <sup>4</sup>	Metaflow <sup>5</sup>	Kedro <sup>6</sup>	Dataiku <sup>7</sup>
Registro de modelos	✓	✓	✓	✓	×	×	✓
Visualización de métricas	✓	✓	✓	✓	✓	✓	✓
Control de versiones	✓	✓	✓	✓	×	✓	✓
Open source	×	✓	×	×	✓	×	×
Seguimiento en producción	×	×	✓	×	×	×	✓
Reportes integrados	×	×	×	✓	✓	×	✓
Pipelines	✓	×	✓	✓	✓	✓	✓

**Tabla 3.1:** Comparativa de competidores (características) (elaboración propia)

A primera vista, observamos que el *software* más completo es Dataiku, que cumple con todas las características que valoramos (a excepción de ser *Open source*, que no afecta al desempeño del programa). Para establecer un ranking del resto de programas, vamos a realizar una valoración ponderada simple de estos:

- 3 puntos: Registro de modelos, Control de versiones, Pipelines
- 2 puntos: Visualización de métricas
- 1 punto: Open source, Seguimiento en producción, Reportes integrados

CARACTERÍSTICAS	Iterative Studio	MLflow	Comet ML	Weights & Biases	Metaflow	Kedro	Dataiku
Registro de modelos	3	3	3	3	0	0	3
Visualización de métricas	2	2	2	2	2	2	2
Control de versiones	3	3	3	3	0	3	3
Open source	0	1	0	0	1	0	0
Seguimiento en producción	0	0	1	0	0	0	1
Reportes integrados	0	0	0	1	1	0	1
Pipelines	3	0	3	3	3	3	3
TOTAL	11	9	12	12	7	8	13

**Tabla 3.2:** Comparativa de competidores (valoración ponderada) (elaboración propia)

Si bien Dataiku es el *software* más completo, Comet ML y Weights & Biases le siguen de cerca y ofrecen características muy similares, con Iterative Studio inmediatamente después de estos tres. Esta competitividad en un sector tan pequeño puede resultar un problema para nuestro proyecto, ya que el auge actual de la inteligencia artificial implica una carrera de desarrollo en la que puede ser difícil mantenerse a la altura de las demandas del mercado. Sin embargo, esta expansión de la IA implica también el descubrimiento de nuevas aplicaciones, lo que a su vez puede llevar a la aparición de nuevos nichos de mercado dentro del sector: una forma de esquivar esta competitividad podría ser la especialización en uno de estos nichos, de forma que limitáramos las capacidades del *software* para centrarnos en ofrecer servicio para una aplicación concreta de forma eficaz.

<sup>1</sup><https://iterative.ai/>

<sup>2</sup><https://mlflow.org/>

<sup>3</sup><https://www.comet.com/site/>

<sup>4</sup><https://wandb.ai/site>

<sup>5</sup><https://metaflow.org/>

<sup>6</sup><https://kedro.org/>

<sup>7</sup><https://www.dataiku.com/>

## 3.2 Competencia potencial

Al ser un mercado en plena expansión, la entrada de competidores potenciales es inevitable y difícilmente previsible. Podemos suponer que las grandes empresas líderes en inteligencia artificial eventualmente desarrollarán sus propias herramientas, o realizarán una integración vertical adquiriendo una que ya esté en el mercado, como se desarrolla en la sección 3.5.

Sin embargo, tenemos que tener en cuenta que nuestro sector no es el de desarrollo de proyectos de inteligencia artificial, en el que participan grandes empresas como Google o Microsoft, sino el de las herramientas de apoyo a este tipo de proyectos: el desarrollo de estas herramientas, aunque se beneficia del conocimiento de expertos a la hora de determinar las características del producto y de programarlas, requiere de muchísimo menos personal y recursos. Estas barreras de entrada más reducidas implican que la aparición de nuevos competidores es un factor para el que debemos prepararnos. Si volvemos a los programas similares comparados anteriormente, vemos que todos han sido lanzados recientemente:

Programa	Iterative Studio	MLflow	Comet ML	Weights & Biases	Metaflow	Kedro	Dataiku
Fecha de salida	2022 <sup>8</sup>	2019 <sup>9</sup>	2017 <sup>10</sup>	2017 <sup>11</sup>	2019 <sup>12</sup>	2019 <sup>13</sup>	2013 <sup>14</sup>

**Tabla 3.3:** Comparativa de competidores (fechas de salida) (elaboración propia)

Excepto Dataiku, todos los programas han sido lanzados en los últimos 6 años, con el más reciente cumpliendo apenas un año. La inversión de las empresas en recursos de inteligencia artificial no hace más que crecer, como se ha visto en la figura 2.1: esta inversión ha dado como resultado un uso mayor de tecnologías de IA en las compañías, mostrado en la figura 3.1, en la que se observa que el número medio de estas tecnologías introducidas en la empresas se ha duplicado en los últimos 4 años. Por tanto, tenemos la certeza de que nuestro sector de mercado se ampliará en el futuro próximo, así como los competidores a los que nos enfrentaremos.

<sup>8</sup><https://www.businesswire.com/news/home/20220726005244/en/Iterative-Introduces-First-Generative-AI-based-Machine-Learning-Model-Registry>

<sup>9</sup><https://github.com/mlflow/mlflow/releases/tag/1.0.0>

<sup>10</sup><https://www.comet.com/site/about-us/>

<sup>11</sup><https://www.crunchbase.com/organization/weights-biases>

<sup>12</sup><https://github.com/Netflix/metaflow/releases/tag/2.0.0>

<sup>13</sup><https://github.com/kedro-org/kedro/releases/tag/0.14.0>

<sup>14</sup><https://www.crunchbase.com/organization/dataiku>

### Average Number of AI Capabilities That Respondents' Organizations Have Embedded Within at Least One Function or Business Unit, 2018–22

Source: McKinsey & Company Survey, 2022 | Chart: 2023 AI Index Report

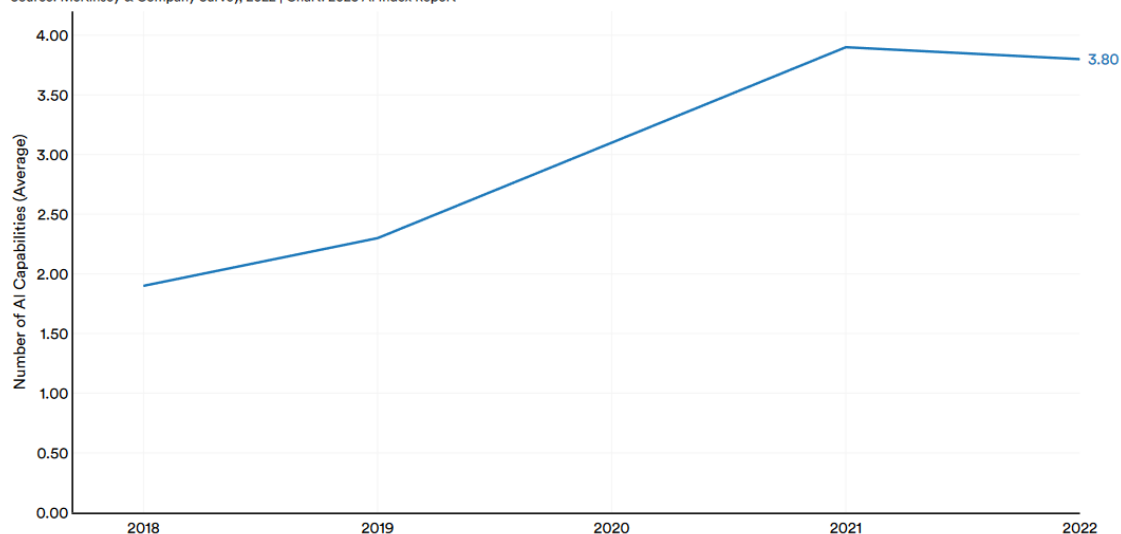


Figura 3.1: Instancias de IA introducidas en al menos un sector de la empresa, 2018 a 2022 (McKinsey & Company Survey, 2022)

## 3.3 Productos sustitutos

Debido a las características del mercado, difícilmente podrían encontrarse productos alternativos que no se consideraran como competidores directos. Todas las etapas de la historia de la informática han comenzado como aplicaciones caóticas y desorganizadas de una idea revolucionaria, que con el tiempo se organizaron y estandarizaron para permitir nuevos avances.

Es cierto que el campo de la ciencia de datos aún es joven, pero la industria informática ha alcanzado un punto de madurez en el que la creación y aplicación de estándares es algo esperado, y que por tanto llevará menos tiempo de lo que podría haber tardado hace 20 años: cualquier gran proyecto de datos que se precie, especialmente en empresas de alcance internacional que tratan con cantidades masivas de información, utilizan estrategias y procesos claramente definidos. Es en base a estos procesos que se diseñan las herramientas de apoyo de las que hablamos en este trabajo, y por tanto es difícil imaginar cómo podría aparecer un producto sustitutivo; al fin y al cabo, el producto está diseñado no conforme a las especificaciones de un cliente, sino a las especificaciones de todo un sector.

## 3.4 Poder de negociación de los proveedores

Si bien la industria tecnológica incluye productos tanto de *hardware* como de *software*, el sector de mercado en el que nos estamos moviendo trata exclusivamente de desarrollo de *software*. Una de sus características distintivas, y que explica el origen de los mitos de Steve Jobs con su garaje, es el hecho de que el desarrollo de programas informáticos es una actividad casi exclusivamente intelectual y que apenas requiere de recursos físicos más allá de un ordenador.

Si bien esto no es aplicable a empresas como Apple o Microsoft, que tanto por su expansión a otros mercados como simplemente por su tamaño ya no siguen este principio, sí lo es para una idea como la nuestra. Incluso a la hora de escalar el proyecto y

ampliar a un equipo de desarrollo, los únicos recursos que necesitaría una empresa dedicada a nuestro programa de ciencia de datos serían ordenadores y una localización donde trabajar, e incluso esto se ha puesto en entredicho en los últimos años con el auge del teletrabajo. Lo genérico de estos recursos nos vuelve prácticamente inmunes a presiones de proveedores, ya que al ser recursos utilizados en la inmensa mayoría de sectores del mundo, los proveedores son también numerosos y competitivos entre sí.

### 3.5 Poder de negociación de los clientes

---

En este área de la informática es incluso más importante de lo normal tener en cuenta a los grandes jugadores, ya que el desarrollo de ciencia de datos es una actividad que generalmente requiere grandes cantidades de recursos y de datos, algo que no está al alcance de muchas pequeñas y medianas empresas: tiene una barrera de entrada muy alta, lo cual reduce el número de clientes potenciales, y consecuentemente incrementa su poder de negociación.

La combinación entre un número reducido de clientes potenciales y el tamaño inherente de estos clientes convierte este factor en el más relevante de las 5 fuerzas de Porter: las grandes empresas utilizan frecuentemente servicios y productos externos durante sus operaciones, pero darles este servicio o producto supone una presión constante para estar a la altura, o de lo contrario arriesgarse a perder al cliente debido simplemente a la facilidad que tiene para cambiar de producto o desarrollar el suyo propio. Una vez más, la rápida expansión del sector supone una amenaza debido a la aparición constante de nuevos requerimientos, con el fin de satisfacer necesidades que aparecen casi de un día para otro.

Por ejemplo, Google tiene una historia consolidada de integración vertical con una serie de adquisiciones que siguen el mismo patrón:

- Adquirió Android Inc. en 2005 con el fin de entrar en el mercado de *software* móvil [Elgin, 2005].
- Adquirió YouTube en 2006 por 1.650 millones de dólares [NBC News, 2006].
- Adquirió Where2Technologies en 2004, que se convertiría poco después en Google Maps [Kiss, 2009].
- Adquirió Keyhole en 2004, cuya tecnología se utilizaría para lanzar en 2005 Google Earth [Google, s.f.].

Google es una empresa que tiende a expandir sus servicios a través de adquisiciones, entrando así rápidamente en nuevos mercados mediante un producto relativamente establecido y desarrollándolo después. No sería descabellado suponer que, dada la creciente importancia de la inteligencia artificial, Google quisiera disponer de una herramienta *in house* que poder personalizar a su antojo. Esta práctica no es exclusiva de Google, sino que otras grandes tecnológicas como Microsoft repiten el patrón, con más de 200 adquisiciones por parte de la empresa fundada por Bill Gates desde 1986.





---

---

## CAPÍTULO 4

# Análisis de necesidades

---

Una vez hecho un análisis general del macroentorno y del microentorno del producto, complementaremos este último enfocándonos en los factores que podrían influenciarlo en el futuro, y estableciendo un plan a grandes rasgos que nos ayudará a lidiar de forma estructurada con estos factores, ya sea aprovechando factores positivos o paliando factores negativos.

### 4.1 Resumen DAFO

---

El primer paso de este proceso será un análisis DAFO: Debilidades, Amenazas, Fortalezas, Oportunidades. Este análisis realiza una comparativa en dos ejes entre factores externos (AO)/internos (DF), y factores positivos (FO)/negativos (DA). En cada sección, añadiremos un subapartado con un resumen de CAME (Corregir debilidades, Afrontar amenazas, Mantener fortalezas, Explotar oportunidades) que desarrollaremos posteriormente.

DEBILIDADES	AMENAZAS
<ul style="list-style-type: none"><li>· Equipo pequeño de desarrollo</li><li>· Falta de experiencia profesional</li></ul>	<ul style="list-style-type: none"><li>· Barreras bajas de entrada a la industria<ul style="list-style-type: none"><li>· Factores políticos (2.1)</li><li>· Factores sociales (2.3)</li></ul></li><li>· Factores ambientales (2.5)</li><li>· Factores legales (2.6)</li></ul>
FORTALEZAS	OPORTUNIDADES
<ul style="list-style-type: none"><li>· Público objetivo de nicho</li><li>· Requerimientos de producto simplificados</li></ul>	<ul style="list-style-type: none"><li>· Mercado en rápida expansión</li><li>· Creación de nuevos nichos de mercado<ul style="list-style-type: none"><li>· Factores económicos (2.2)</li><li>· Factores sociales (2.3)</li><li>· Factores tecnológicos (2.4)</li></ul></li></ul>

**Tabla 4.1:** Tabla de análisis DAFO (elaboración propia)

#### 4.1.1. Debilidades

Las debilidades son factores internos negativos.

- **Equipo pequeño de desarrollo** Al ser este un proyecto a pequeña escala con un único desarrollador, nuestra baja capacidad de trabajo y de reacción al mercado supone un problema grave a la hora de querer realizar un proyecto completo y de

calidad. Este factor es inevitable, ya que pese a que se puede ahorrar mucho trabajo con algo de ingenio y una gestión inteligente de los recursos, este ahorro tiene un límite a partir del cual lo único que se puede hacer para contribuir al proyecto es realizar el propio proyecto.

- **Falta de experiencia profesional** Como estudiante, pese a una formación teórica extensa, carezco de la experiencia práctica que proporciona trabajar en la propia industria y que me permitiría adelantarme a muchos problemas que irán surgiendo durante el desarrollo, especialmente en cuanto a requerimientos por parte de las empresas: está claro que si fuera un usuario de estas herramientas, estaría mucho más al tanto de las necesidades de estos usuarios, sobre todo cuando tienes requerimientos técnicos tan específicos.
- **CORREGIR** Debido a la dificultad de superar estos problemas, lo único que podemos hacer para paliarlos es establecer pautas para minimizarlos: estableceremos un orden de prioridades en la introducción de características al programa para asegurarnos de que cubrimos las bases más importantes, y aprenderemos sobre estas características secuencialmente antes de introducirlas.

#### 4.1.2. Amenazas

Las amenazas son factores externos negativos.

- **Barreras bajas de entrada a la industria** Si recordamos que nuestro sector competitivo no es “la ciencia de datos” sino las herramientas de ayuda a estos proyectos, nos damos cuenta de que la barrera de entrada es considerablemente más baja: no son necesarias grandes cantidades de datos ni un equipo legal que se asegure de que se utilizan de acuerdo a la legislación vigente, sino simplemente un equipo de desarrollo con buenas ideas y capacidades, y *hardware*.
- **Factores políticos (2.1)** Como se menciona en la sección correspondiente, la inestabilidad de la oferta de semiconductores es un condicionante decisivo en el desarrollo de la industria, y uno al que no podemos hacer frente sin cambiar radicalmente nuestro producto.
- **Factores sociales (2.3)** El uso de datos de individuos en los proyectos de Inteligencia Artificial pone en tela de juicio lo ético de estas prácticas, lo que supone un grave riesgo para nuestro producto; igual que el mercado está ahora en crecimiento, un giro de la opinión pública puede suponer que esta tendencia se revierta.
- **Factores ambientales (2.5)** Aunque de manera indirecta, las preocupaciones por el impacto medioambiental del entrenamiento de modelos de IA podría impactar la opinión pública de la industria, especialmente con la creciente preocupación por el cambio climático y sus efectos.
- **Factores legales (2.6)** En relación con la preocupación social por la invasión a la privacidad individual que puede suponer el uso de datos de entrenamiento, las limitaciones legales impuestas por entidades nacionales e internacionales pueden suponer un duro golpe al crecimiento del sector a nivel mundial.
- **AFRONTAR** Muchas de estas amenazas son situaciones de talla internacional. Para minimizar sus efectos, sólo podemos intentar adelantarnos a sus consecuencias a través de un desarrollo conservador y preventivo, o lo que es lo mismo, crear un producto de calidad que cumpla con las expectativas de nuestros clientes para generar fidelización. Veremos en otra sección por qué esto puede no ser posible.

### 4.1.3. Fortalezas

Las fortalezas son factores internos positivos.

- **Público objetivo de nicho** El público objetivo que hemos seleccionado es reducido y específico, además de principiante, lo que significa que son menos exigentes. Esto nos da una gran ventaja a la hora de crear un producto que dé respuesta a sus necesidades, además de tener inherentemente menos competencia por ser un público más pequeño.
- **Requerimientos de producto simplificados** Como hemos mencionado anteriormente, nuestro público objetivo consiste principalmente de individuos que estén interesados en iniciarse en la ciencia de datos. Esto significa que tendrán que aprender el funcionamiento de este tipo de proyectos, y no utilizarán todas las características propias de este tipo de programas, o estarán más interesados en versiones simplificadas que en versiones completas y complejas. Esto nos facilita los requisitos de desarrollo.
- **MANTENER** Debemos invertir un tiempo considerable en el diseño del programa, para convertirlo en un referente en materia de programas educativos aprovechando la falta de competidores actual. Una vez obtenido este puesto, será más fácil mantenerlo con publicidad boca a boca.

### 4.1.4. Oportunidades

Las oportunidades son factores externos positivos.

- **Mercado en rápida expansión** Si bien el sector profesional lleva ya varios años haciendo uso de la ciencia de datos en sus actividades de negocio, el interés reciente por parte del sector público la ha catapultado a los telediarios y periódicos. Podemos considerar que la ciencia de datos ha cruzado el umbral a partir del cual pasa de ser exclusivamente una oscura herramienta del sector profesional, para convertirse en algo en lo que los jóvenes estudiantes de informática se pueden interesar. La creación reciente de este nicho de mercado nos presenta una rara oportunidad para introducirnos cuando aún no hay demasiados competidores, en un mercado en el que de por sí hay espacio para pocos.
- **Creación de nuevos nichos de mercado** Precisamente la aparición de este pequeño público objetivo nos da una idea del crecimiento potencial del sector de la ciencia de datos. Igual que la industria informática general ha creado una cantidad sorprendente de necesidades profesionales (por ejemplo, empresas y productos especializados en diseño, programación en diferentes lenguajes, soporte técnico, *hardware*, DevOps, la propia ciencia de datos...), podemos esperar que los avances en ciencia de datos creen progresivamente nuevos nichos de mercado que podrían resultar interesantes en el futuro.
- **Factores económicos (2.2)** El crecimiento de la industria durante los últimos años ha sido constante y masivo, y es el principal motivo por el que este proyecto tiene futuro. Pese a las barreras impuestas por otros factores, los beneficios potenciales derivados de este tipo de tecnología son tales que es difícil imaginar un mundo en el que no serían explotados por las empresas.
- **Factores sociales (2.3)** De la misma forma que la imagen pública puede suponer un obstáculo para el desarrollo de la IA, su popularidad también puede convertirse en

una ventaja: si bien tiene muchos problemas de carácter ético y legal, su explosión de usuarios ha ocurrido por lo interesante y útil de estas herramientas, así como por su facilidad de uso (una vez la herramienta está desarrollada, que no el desarrollo en sí mismo). Esta combinación de simpleza y utilidad hace casi inevitable que pasen de ser curiosidades a ser de uso común, lo que nos favorece.

- **Factores tecnológicos (2.4)** La Inteligencia Artificial se cuele en cada vez más sectores a pasos agigantados, y nuevas aplicaciones de esta tecnología se desarrollan continuamente. El auge económico de esta industria depende directamente de sus capacidades tecnológicas, que no paran de incrementarse y habilitar nuevos usos con un gran potencial.
- **EXPLOTAR** La actual naturaleza cambiante de la industria de la ciencia de datos supone una oportunidad que podemos explotar si actuamos con velocidad: esto implica tanto llevar un desarrollo lo más acelerado posible, como mantenerse al día de los cambios del panorama informático para modificar dicho desarrollo si fuese necesario. Finalizar nuestro programa con un nivel de calidad razonable en un momento en el que hay pocos competidores podría suponernos una ventaja enorme.

## 4.2 Necesidades actuales

---

El desarrollo de un programa informático tiene ciertas características que lo separan de otras áreas más industriales, principalmente el hecho de que el producto final no sea físico nos permite ignorar todo tipo de gestión de compraventa de materias primas, logística, distribución o mano de obra que tradicionalmente suponen una buena parte de los gastos de una empresa mercantil al uso.

Esto hace que el peso relativo del trabajo intelectual (diseño y programación) sea enorme, con lo que es el área donde más debemos centrarnos para garantizar el éxito del producto. Esta situación excepcional me permite trabajar en el desarrollo de este programa como un equipo de una sola persona, minimizando costes sin imposibilitar el éxito del producto. Esto no significa que el desarrollo vaya a ser muy eficaz, pero sí lo más eficiente posible en cuanto a uso de recursos económicos.

Todo esto significa que, actualmente, la programación de este sistema de gestión de ciencia de datos sólo requiere a un desarrollador durante el tiempo suficiente para aprender las tecnologías necesarias y realizar el propio desarrollo: recordamos que nuestros requerimientos no son los de una gran empresa, sino que queremos realizar un producto simplificado e intuitivo, lo cual simultáneamente simplifica nuestras necesidades, como se ve en la sección 4.1.3.

## 4.3 Necesidades futuras

---

En cuanto a perspectivas de futuro, sería inútil pensar en necesidades en el caso del fracaso del producto, así que asumiremos que será un éxito y tendremos que expandirnos.

Para planificar correctamente esta expansión, tendremos que cubrir una serie de necesidades que de otra forma supondrían obstáculos importantes:

- **Análisis de mercado.** Tendremos que decidir si continuamos el desarrollo de un programa especializado en ser simple e intuitivo para nuestro público actual de

estudiantes y principiantes, o si nos introducimos en otras áreas relacionadas del mercado. Para tomar esta decisión será esencial realizar un estudio de mercado concienzudo, que nos indique qué tipo de producto tendría la mejor recepción o la menor competencia.

- **Conocimiento experto.** Al ser la ciencia de datos un campo en constante evolución y crecimiento, es importante mantenerse al día e incluso por delante de los nuevos descubrimientos y cambios en los paradigmas de desarrollo de la inteligencia artificial y modelos de *Machine Learning*. Esto nos permitirá, una vez decidido el nuevo producto a desarrollar (o si continuamos con el actual), hacerlo de manera que incremente su valor y potencialmente resulte atractivo para otros usuarios.
- **Incremento de la capacidad productiva.** Llegado cierto punto, el hecho de que el desarrollo informático no requiera de materias primas o fábricas no es suficiente para justificar una plantilla de una sola persona; la ampliación/desarrollo de la plataforma necesitaría de personal nuevo para cubrir unos plazos de entrega razonables, además de para desarrollar e implementar a una velocidad razonable los cambios que, como ya hemos mencionado, son frecuentes en este sector.
- **Gestión de empresa.** Esta expansión del proyecto en una empresa con potencialmente varios productos y una plantilla propiamente dicha genera la necesidad de dedicar una parte de esa plantilla a la gestión de la propia empresa, especialmente en los ámbitos legal y económico (finanzas, contabilidad, salarios, clientes...).



---

---

## CAPÍTULO 5

# Propuesta de estrategia

---

A continuación, vamos a plantear una estrategia destinada a lanzar el producto en la actualidad, seguida de una estrategia a largo plazo que buscaría continuar el éxito del producto.

### 5.1 Estrategia actual

---

Para llevar este proyecto a buen puerto nos será necesario, como en cualquier otro, obtener los mejores resultados posibles utilizando los menores recursos posibles. Para cumplir este objetivo, estableceremos un orden de prioridad en el desarrollo de las herramientas del programa que nos permita crear un prototipo utilizable lo antes posible, que después completaremos progresivamente. Este orden de prioridad estará basado en la utilidad de las características en cuestión para nuestros potenciales clientes, y las separará en tres categorías: esencial (necesaria para un prototipo inicial), importante (añadidos que incrementarían considerablemente el valor del producto) y extra (añadidos que aportan una utilidad poco frecuente en el mercado). La descripción de cada característica puede encontrarse en la sección [3.1](#).

1. Control de versiones [esencial]: uno de los pilares centrales del MLOps, ya que facilita mucho el manejo de las diferentes versiones de un modelo de Inteligencia Artificial, algo crucial en el proceso iterativo para encontrar un modelo de calidad.
2. Registro de modelos [esencial]: permite gestionar y realizar comparativas entre diferentes proyectos, con el mismo objetivo de encontrar un modelo de calidad.
3. Pipelines [esencial]: simplifica enormemente el proceso iterativo de mejora, creando "tuberías" hechas de varias etapas que permiten al desarrollador ejecutar todo el proceso fácilmente cuando modifica alguna de las etapas.
4. Visualización de métricas [importante]: proporciona una interfaz gráfica que presenta de forma intuitiva métricas que dan información sobre el rendimiento del modelo. Son métricas que se pueden calcular manualmente, pero que se utilizan mucho en todos los proyectos de ciencia de datos.
5. Reportes integrados [extra]: un paso más en la "traducción" de los resultados numéricos del modelo al lenguaje humano, en el que se añade una interpretación rudimentaria automática a la visualización de métricas.
6. Seguimiento en producción [extra]: proporciona información en tiempo real sobre el rendimiento de un modelo una vez se está utilizando para su propósito final. Esta



función puede ser importante en algunos casos específicos, pero no es algo que esté generalizado en la industria.

Para implementar estas características, utilizaremos una metodología ágil que se basa en "sprints": se determinan objetivos semanales que responden a una problemática pre-establecida, se implementan en esa semana, se realizan pruebas de funcionamiento y calidad, y a la semana siguiente se repite el proceso. Esta metodología se utiliza actualmente en muchas empresas debido al enfoque que tiene en los requerimientos del cliente: en lugar de tener un concepto de alto nivel que la empresa tiene que plasmar en un producto, el cliente participa en cada momento del desarrollo y la empresa se asegura de que cada sprint resulta en un paso más para cumplir los diversos requerimientos del cliente. Esto hace que el documento de requerimientos tenga que ser extenso y subdividirse en porciones factibles para desarrollar en un periodo de una semana.



**Figura 5.1:** Diagrama explicativo de la metodología ágil [Santa Lucía Impulsa, 2021]

A nivel de planificación tradicional de empresa, disponemos de numerosas ventajas mencionadas anteriormente: no tenemos la necesidad inmediata de adquirir instalaciones, maquinaria o personal, sino que podemos comenzar el trabajo limitando nuestros recursos a un desarrollador que puede incluso trabajar de forma remota. Esto hace que este proyecto tenga un riesgo muy bajo debido a la inversión inicial casi nula que necesita.

## 5.2 Estrategia futura

Poniendo la vista en el futuro, la situación cambia. Vamos a considerar dos situaciones según el desempeño del producto en el mercado, y establecer una guía de conducta para tratar de incrementar sus probabilidades de éxito en cada caso.

### 5.2.1. Éxito moderado

A falta de datos exactos de mercado, asumimos que un "Éxito moderado" estaría definido por unos ingresos pequeños, del orden de 3 cifras mensuales o menos. En este caso, nos encontraríamos en una situación difícil, aunque no necesariamente definitiva.

Los ingresos, por pequeños que sean, se convierten mayoritariamente en beneficios gracias a los gastos reducidos de nuestra actividad. Sin embargo, un beneficio reducido nos impide la expansión, lo que a su vez nos dificulta mantener el programa actualizado en una industria que cambia a un ritmo acelerado. Esto será un problema si no somos capaces de compensar esta carencia de alguna otra forma.

Para paliar esto, tenemos dos opciones. En primer lugar, podemos dedicar más tiempo al desarrollo de características para el programa a cambio de reducirlo en otras áreas, como diseño gráfico o *testing* de errores. Es una práctica arriesgada, pero que puede permitirnos ofrecer un producto más completo, aunque menos intuitivo. Esto tiene la ventaja de que puede incrementar las ventas a corto plazo, permitiéndonos incrementar los ingresos lo suficiente como para financiar una pequeña expansión de personal, pero a la vez supone apostar a que este incremento llegará lo suficientemente rápido como para poder solucionar los errores de código que vayan surgiendo y mejorar la interfaz gráfica. Este último punto es especialmente importante, ya que no podemos olvidar que nuestro producto está destinado a usuarios principiantes como alumnos, con lo que una interfaz amigable e intuitiva es esencial para facilitar su uso.

En segundo lugar, podemos tomar una dirección completamente opuesta y optar por la opción más segura: ya que nuestro programa está enfocado a usuarios noveles, centrarnos en mejorar nuestro producto para este público y cimentar este nicho de mercado. Esto implicaría dedicar mayores esfuerzos a la interfaz gráfica, y especialmente a realizar un diseño inteligente que facilite el uso del programa a los ojos de alguien que no está familiarizado con convenciones que pueden darse por sentadas en la industria. Si bien esto no es tarea fácil, ya que requiere simultáneamente de conocimientos de estas convenciones (para poder evitarlas) y de ciertas nociones de educación y aprendizaje, es una inversión de tiempo que será mucho más duradera en el futuro: las innovaciones en el sector suelen estar aplicadas a áreas avanzadas (mejoras de rendimiento de los entrenamientos, nuevas formas de configurar los experimentos o los datos, etc), mientras que las bases de la ciencia de datos están ya bastante asentadas.

### 5.2.2. Éxito mayor

Esta vez asumiremos que los ingresos son muy superiores, del orden de 4 cifras altas mensuales. Es una situación muy favorable ya que, de nuevo, la mayoría de estos ingresos se convierten directamente en beneficio. La estrategia está mucho más clara: tenemos la oportunidad de expandir el proyecto con la inclusión de nuevo personal.

**Equipo de desarrollo** Necesitaríamos a 1-2 personas que se dedicaran al desarrollo del programa, tanto para añadir funcionalidades que no se pudieron incluir inicialmente como para mantenerlo actualizado respecto a los cambios y novedades de la industria.

**Equipo financiero/contable/legal** Al menos 1 persona se dedicará a la gestión de los recursos de la empresa, y a mantenernos al día con la regulación vigente. Esto es necesario por motivos obvios, ya que la documentación necesaria para crear una empresa y registrar las operaciones diarias puede ser compleja. Sin embargo, nuestro tamaño reducido no justifica el uso de nuestros recursos limitados en este empleado. Lo que realmente lo justifica es, de nuevo, el panorama cambiante de la industria: como hemos visto en el apartado 2.6, organizaciones gubernamentales, tanto nacionales como internacionales, están redactando legislación que puede afectarnos. Es importante mantenernos no sólo al día sino un paso por delante, para prepararnos para estas leyes y evitar problemas legales y económicos.

Una vez realizada la expansión, el futuro de la empresa depende realmente del buen uso de estos recursos, tanto humanos como económicos. Para establecer unos ingresos estables, contactaremos directamente con universidades y organizaciones educativas que puedan estar interesadas por nuestro *software*, ya que un solo contrato puede suponer muchos más ingresos que esforzarnos en personas individuales. Además son ingresos garantizados durante un periodo de tiempo acordado, con una opción muy probable a extender dicho contrato si el programa funciona como es debido y no surge un competidor capaz de ofrecer un mejor servicio.

Ofreceremos también una versión para uso individual, con una licencia que impida el uso académico/profesional, a un precio reducido o incluso gratuita. El objetivo de esto no es tanto generar ingresos, aunque son bienvenidos, sino más bien convertir nuestra herramienta en una opción popular a la hora de iniciarse en la ciencia de datos. Esta popularidad traducida en una base de usuarios creciente se puede utilizar como argumento a nuestro favor a la hora de trabajar con nuestros verdaderos clientes, las organizaciones educativas, de dos formas: creando una equivalencia entre el tamaño de nuestra base de usuarios y la calidad del programa, e insistiendo en que la mejor elección para los alumnos es el programa que más probabilidades tienen de usar en su tiempo libre (debido, una vez más, a una gran base de usuarios) ya que les ahorraría tiempo y facilitaría el proceso educativo.

---

---

## CAPÍTULO 6

# Conclusiones

---

Llevar a cabo este proyecto tiene desventajas importantes que hacen tambalearse su propia premisa, pero también una serie de ventajas que lo convierten en muy atractivo. Hemos tratado de aprovechar las ventajas y paliar las desventajas.

El campo de la ciencia de datos, tanto a nivel industrial como a nivel tecnológico, es muy inestable y está en constante desarrollo: el análisis PESTEL (capítulo 2) pone de manifiesto esta naturaleza cambiante en varios ámbitos diferentes, y nos muestra que tiene un riesgo inherente difícil de reducir al nivel de empresa (legislación en constante movimiento, presión social, tensiones políticas...). Las decisiones que hemos tomado al respecto han sido, principalmente:

- Centramos en un nicho de mercado reducido, para crear un producto diseñado específicamente para ellos. Pese a reducir nuestros clientes potenciales totales, incrementamos las probabilidades de que los miembros de este mercado sean clientes.
- Optar actualmente por una estructura de desarrollo que se centra en sacarle el máximo partido al tiempo disponible, estableciendo un orden de introducción de características según su importancia. Si necesitamos presentar un prototipo temprano a un cliente, podremos mostrarle algunas de las funcionalidades más importantes desde un periodo temprano en el proceso de desarrollo.
- Establecer planes de desarrollo de la empresa diferentes según nuestro rendimiento económico, para cubrir un mayor número de posibilidades del futuro del proyecto; si prevemos múltiples escenarios, estaremos preparados para afrontar cualquiera de ellos gracias a un análisis y trabajo previos.

Por otro lado, el desarrollo *software* a pequeña escala siempre ha contado con ciertas ventajas que no tienen otras áreas industriales. Además, algunos de los factores que nos suponen un problema también se convierten en positivos con tan solo cambiar el enfoque:

- El mercado es cambiante porque está en rápida expansión, lo que incrementa los clientes y por tanto las oportunidades para ofrecer nuestro servicio. Es un caso más favorable que un mercado ya establecido, ya que la oferta de la competición tarda un tiempo en satisfacer la demanda creciente de los clientes, una ventana de tiempo que podemos aprovechar para captar a parte de esos clientes y establecernos nosotros mismos como un nombre de referencia en nuestro nicho de mercado.
- El desarrollo del proyecto aquí presentado puede realizarse por una sola persona. Aunque el resultado final será indudablemente peor que el que podría haberse conseguido con un equipo más extenso, también reduce muchísimo el presupuesto

necesario inicial y minimiza el riesgo, permitiéndonos tantear el terreno antes de realizar una inversión grande.

- Al ser un mercado que demanda un producto tan específico y técnico como es “una herramienta de apoyo al desarrollo de proyectos de ciencia de datos”, es extremadamente difícil crear un producto sustitutivo que llegue a desplazar al nuestro o similares. Esto significa que, aunque otros competidores surjan o incluso obtengan una mayor cuota de mercado que nosotros, será difícil expulsarnos completamente del mercado; las mejoras serán incrementales, lo que facilita incluirlas en nuestro producto para intentar defender o recuperar una posición fuerte.

En última instancia, lo más importante es que la situación actual presenta una oportunidad de oro que se puede aprovechar de forma relativamente sencilla, al menos inicialmente. Sin embargo, esta oportunidad es efímera y de la misma forma que lo hemos hecho nosotros, otros llegarán a esta misma conclusión. Como toda oportunidad de mercado, cuanto más rentable es, menos tiempo ofrece a los interesados para explotarla.

# Bibliografía

---

- [Alsop, 2023] Alsop, T. (14 de Marzo de 2023). *Leading semiconductor foundries revenue share worldwide from 2019 to 2022, by quarter*. Obtenido de Página web de Statista: <https://www.statista.com/statistics/867223/worldwide-semiconductor-foundries-by-market-share/>
- [Comisión Europea, 2020] Comisión Europea. (19 de Febrero de 2020). *White Paper on Artificial Intelligence: a European approach to excellence and trust*. Obtenido de Página web de la Comisión Europea: [https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf)
- [Comisión Europea, DG CNECT, 2022] Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías. (8 de Febrero de 2022). *Propuesta de REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO por el que se establece un marco de medidas para reforzar el ecosistema europeo de semiconductores (Ley de Chips)*. Obtenido en Página web de EUR-Lex: <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:52022PC0046>
- [Cordero, 2021] Cordero, D. (24 de Agosto de 2021). *La falta de chips fuerza el parón de plantas españolas de automóviles*. Obtenido de Página web de El País: <https://elpais.com/economia/2021-08-24/la-falta-de-chips-fuerza-el-paron-de-plantas-espanolas-de-automoviles.html>
- [Elgin, 2005] Elgin, B. (15 de Agosto de 2005). *Google Buys Android for Its Mobile Arsenal*. Obtenido de Página web de Bloomberg Businessweek: [https://web.archive.org/web/20110205190729/http://www.businessweek.com/technology/content/au\\_g2005/tc20050817\\_0949\\_tc024.htm](https://web.archive.org/web/20110205190729/http://www.businessweek.com/technology/content/au_g2005/tc20050817_0949_tc024.htm)
- [Gendron, 2023] Gendron, W. (14 de Abril de 2023). *ChatGPT needs to "drink" a water bottle's worth of fresh water for every 20 to 50 questions you ask, researchers say*. Obtenido de Página web de Business Insider: <https://www.businessinsider.com/chat-gpt-generative-ai-water-use-environmental-impact-study-2023-4>
- [Google, s.f.] Google. (s.f.). *Our history in depth*. Obtenido de Página web de Google UK: <https://web.archive.org/web/20160406123606/http://www.google.co.uk/about/company/history/#2004>
- [International Trade Administration of the USA, 2021] International Trade Administration of the USA. (2021). *Global Artificial Intelligence Market Report*. Obtenido de Página web de International Trade Administration: <https://www.trade.gov/global-artificial-intelligence-market-report>
- [Kiss, 2009] Kiss, J. (17 de Junio de 2009). *Secrets of a nimble giant*. Obtenido de Página web de The Guardian: <https://www.theguardian.com/technology/2009/jun/17/google-sergey-brin>

- [Li, Yang, Islam, & Rei, 2023] Li, P., Yang, J., Islam, M., & Ren, S. (2023). *Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models*. Riverside, CA: UC Riverside.
- [Maslej, y otros, 2023] Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). *The AI Index 2023 Annual Report*. Stanford, CA: AI Index Steering Committee, Institute for Human Centered AI, Stanford University.
- [MINECO, s.f.] Ministerio de Asuntos Económicos y Transformación Digital. (s.f.). ENIA. Obtenido de Página web del Ministerio de Asuntos Económicos y Transformación Digital: <https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx>
- [MINECO, 2020] Ministerio de Asuntos Económicos y Transformación Digital. (Noviembre de 2020). *ENIA Estrategia Nacional de Inteligencia Artificial*. Obtenido de Página web de La Moncloa: <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>
- [MINECO, s.f.] Ministerio de Asuntos Económicos y Transformación Digital. (s.f.). *Oficina del Dato*. Obtenido de la Página web del Ministerio de Asuntos Económicos y Transformación Digital: <https://portal.mineco.gob.es/es-es/digitalizacionIA/oficina-del-dato/Paginas/oficina-del-dato-se-digitalizacion-ia.aspx>.
- [Narayanan, y otros, 2021] Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Beranuer, J., Catanzaro, B., Phanishayee, A., & Zaharia, M. (2021). Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM. *The International Conference for High Performance Computing, Networking, Storage, and Analysis* (págs. 1-15). St. Louis, Missouri: Association for Computing Machinery.
- [NBC News, 2006] NBC News. (9 de Octubre de 2006). *Google buys YouTube for \$1.65 billion*. Obtenido de Página web de NBC News: <https://www.nbcnews.com/id/wbna15196982>
- [Santa Lucía Impulsa, 2021] Santa Lucía Impulsa. (3 de Febrero de 2021). *Metodología Agile: ¿Qué es y para qué sirve?*. Obtenido de la Página web de Santa Lucía Impulsa: <https://www.santaluciaimpulsa.es/metodologia-agile-que-es-para-que-sirve/>
- [Shehabi, y otros, 2016] Shehabi, A., Smith, S., Sartor, D., Brown, R., Herrlin, M., Koomey, J., Masanet, E., Horner, N., Chelikowsky, J., & Lintner, W. (2016). *United States Data Center Energy Usage Report*. Berkeley, CA: Lawrence Berkeley National Lab.
- [Similarweb, 2023] Similarweb. (Mayo de 2023). *chat.openai.com Traffic and Engagement Analysis*. Obtenido de Página web de Similarweb: <https://www.similarweb.com/website/chat.openai.com/#traffic>
- [Statista Research Department, 2023] Statista Research Department. (9 de Junio de 2023). *The 100 largest companies in the world by market capitalization in 2023*. Obtenido de Página web de Statista: <https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-capitalization/>
- [The White House, 2021] The White House. (21 de Enero de 2021). *FACT SHEET: Biden-Harris Administration Bringing Semiconductor Manufacturing Back to America*. Obtenido de Página web de The White House: <https://www.whitehouse.gov/briefing-room/statements-releases/2022/01/21/fact-sheet-biden-harris-administration-bringing-semiconductor-manufacturing-back-to-america-2/>

- [Trump White House, 2020] Trump White House. (24 de Marzo de 2020). *Remarks by President Trump, Vice President Pence, and Members of Coronavirus Task Force in Press Briefing*. Obtenido de Página web de Trump White House: <https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-vice-president-pence-members-coronavirus-task-force-press-briefing-10/>
- [Turley, 2023] Turley, J. (6 de Abril de 2023). *Defamed by ChatGPT: My Own Bizarre Experience with Artificiality of 'Artificial Intelligence'*. Obtenido de Página web de Jonathan Turley: <https://jonathanturley.org/2023/04/06/defamed-by-chatgpt-my-own-bizarre-experience-with-artificiality-of-artificial-intelligence/>
- [von der Leyen, 2021] von der Leyen, U. (15 de Septiembre de 2021). *2021 State of the Union Address by President von der Leyen*. Obtenido de Página web de la Comisión Europea: [https://ec.europa.eu/commission/presscorner/detail/ov/SPEECH\\_21\\_4701](https://ec.europa.eu/commission/presscorner/detail/ov/SPEECH_21_4701)
- [Woo & Blanchard, 2022] Woo, R., & Blanchard, B. (15 de Agosto de 2022). *Angry China stages more drills near Taiwan as U.S. lawmakers visit*. Obtenido de Página web de Reuters: <https://www.reuters.com/world/asia-pacific/china-stages-more-drills-near-taiwan-us-lawmakers-visit-2022-08-15/>





---

# APÉNDICE A

## Objetivos de Desarrollo Sostenible

---

### A.1 Relación con ODS

---

Objetivos de Desarrollo Sostenible	Alto	Medio	Bajo	No Procede
ODS 01. Fin de la pobreza				×
ODS 02. Hambre cero				×
ODS 03. Salud y bienestar				×
ODS 04. Educación de calidad	×			
ODS 05. Igualdad de género				×
ODS 06. Agua limpia y saneamiento				×
ODS 07. Energía asequible y no contaminante				×
ODS 08. Trabajo decente y crecimiento económico		×		
ODS 09. Industria, innovación e infraestructuras	×			
ODS 10. Reducción de las desigualdades				×
ODS 11. Ciudades y comunidades sostenibles				×
ODS 12. Producción y consumo responsables		×		
ODS 13. Acción por el clima				×
ODS 14. Vida submarina				×
ODS 15. Vida de ecosistemas terrestres				×
ODS 16. Paz, justicia e instituciones sólidas				×
ODS 17. Alianzas para lograr objetivos				×

**Tabla A.1:** Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS) (elaboración propia)

### A.2 Reflexión

---

El proyecto aquí presentado puede relacionarse, en mayor o menor medida, con varios Objetivos de Desarrollo Sostenible. Cabe destacar que en algunos casos, la reflexión sobre el ODS no está relacionado tanto con la propia herramienta, sino con el uso que se le puede dar: como plataforma para el desarrollo de proyectos de ciencia de datos, su efecto se notará principalmente en el uso que se le dé para llevar a cabo estos proyectos, y no en la plataforma en sí misma.

**Educación de calidad** El usuario al que está orientado esta herramienta es un usuario principiante, frecuentemente un estudiante, que necesita una introducción específicamente diseñada para el campo de la ciencia de datos: al facilitar la entrada, esperamos que más alumnos mantengan su interés por un campo de estudio que

puede resultar muy complicado si no se presenta de forma adecuada. Esto es importante no sólo por el interés profesional que puede tener el usuario, sino porque la relevancia de esta industria está en aumento día tras día; aprendiendo a utilizar estas herramientas, tendrá también una idea aproximada de cómo se están utilizando sus propios datos por las grandes empresas, que utilizan herramientas similares para extraer información.

**Trabajo decente y crecimiento económico** El uso de datos como fuente de información para las empresas y el entrenamiento de modelos de Inteligencia Artificial es una práctica que ha llegado para quedarse: esta herramienta proporciona una introducción adecuada a los principiantes interesados, de forma que se les facilite la entrada a un mundo extremadamente técnico, lleno de tecnicismos e ideas muy alejadas conceptualmente de lo que una persona se suele encontrar en su día a día. Los proyectos derivados de nuestra herramienta pueden ser tremendamente variados, pero su uso en las empresas suele tener uno de dos objetivos: optimizar el uso de recursos y agilizar procesos, reduciendo el coste económico (una mejora del funcionamiento interno de la empresa), o bien mejorar la precisión de sus campañas publicitarias (una mejora del funcionamiento externo de la empresa).

**Industria, innovación e infraestructuras** La revolución tecnológica causada por el uso de datos ha generado indirectamente beneficios masivos para las empresas más grandes del mundo, y su uso se está expandiendo a todo tipo de industrias, hasta el punto de que ha tenido repercusión social en espacios completamente separados de la prensa especializada. Ahora mismo es uno de los campos de la informática con el desarrollo más rápido, y esto se debe a que los casos de uso aplicables son cada vez más. Nuestra plataforma está simplificada con el fin de ser accesible a principiantes, pero aún así debe ser lo suficientemente completa como para permitir el desarrollo de proyectos de ciencia de datos complejos de principio a fin, si bien con ciertas limitaciones inevitables por su naturaleza simplificada.

**Producción y consumo responsables** El potencial de los datos como recurso es inmenso, y debe ser regulado por una legislación responsable y consciente de la vulnerabilidad de la ciudadanía a las grandes compañías. Nuestra herramienta pretende cumplir esta legislación a rajatabla, no sólo por las consecuencias legales de no hacerlo, sino porque está destinada a trabajar con las futuras generaciones de trabajadores: tenemos el deber, junto a las organizaciones educativas, de enseñar prácticas que sean tanto eficaces como responsables.