Redefiniendo la experiencia de búsqueda y el descubrimiento de productos en el sector del e-commerce mediante un sistema de recomendación basado en RAG combinado con BM25 y Embeddings

Isabel Ayala Bueno
Miguel Ángel Jiménez
Álvaro Regaño

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Entendiendo el sector del e-commerce **3**
Estado actual del e-commerce en España 3
Canales de compra online 4
Motivación de compra del consumidor 5
Experiencia de compra online 6
Comportamiento del consumidor antes de realizar compras online 7
Análisis de la Inteligencia Artificial en el contexto actual del e-commerce **9**
¿Qué percepción tienen los consumidores sobre el uso de la IA en un e-commerce? 10
¿Cómo se está introduciendo la IA en el comercio electrónico? 13
El descubrimiento de productos **15**
Descubrimiento por Navegación: 15
Descubrimiento por Búsqueda: 20
El rol de la búsqueda en el embudo de conversión 21
Tipos de búsqueda 22
El motor de búsqueda **35**
Funcionamiento de los algoritmos de búsqueda léxica: 36
Intent 41
Entity 43
Implementación del Intent y Entity a nivel de posicionamiento 44
Implementación del Intent y Entity en el motor de búsqueda interno 47
Diseñar la solución **50**
Design Thinking 50
Mapa de proceso 50
User Persona 51
Inferir el algoritmo 52
Definir la solución **53**
RAG 53
Arquitectura del RAG 55
Fine-tuning 56
Impacto del RAG en el e-commerce 57
Sistema de recomendación de productos basado en RAG combinado con BM25 y Embeddings 58
Diseño de la UI 58
Criterios de evaluación del modelo 59
Modelo de lenguaje local 60
Arquitectura de un LLM local + RAG 61
Implementación del modelo 62
Conclusiones **65**

Certificado en Inteligencia Artificial

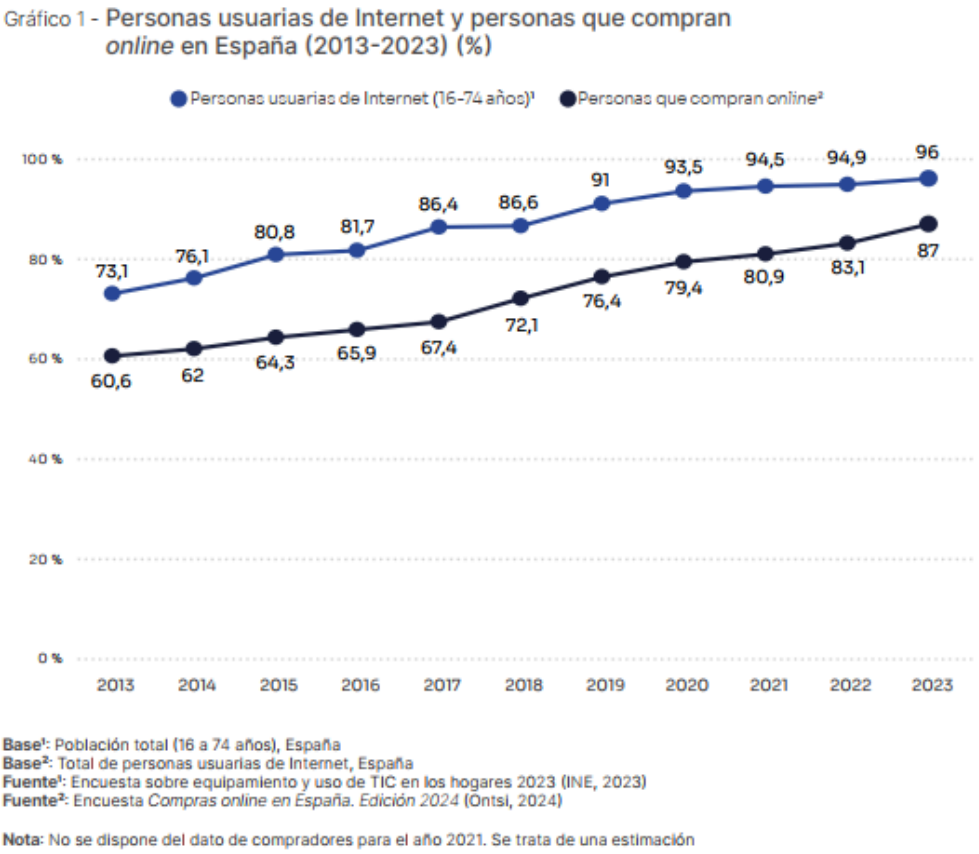
EOI & Samsung Innovation Campus

Entendiendo el sector del e-commerce

El comercio electrónico o e-commerce es la práctica de **comprar y vender bienes y servicios totalmente en línea**. Actualmente abarca numerosos tipos de mercado, desde ropa, artículos deportivos y cotidianos, libros, muebles, hasta la inversión en acciones y diversos productos financieros. Como podemos ver, se trata de un sector muy flexible. Además, podemos acceder a él a través de una gran variedad de dispositivos (móvil, tablet, portátil y muchos más). Todos estos factores hacen que el mercado del comercio electrónico sea **muy competitivo**.

Estado actual del e-commerce en España

Hoy en día, las compras por Internet son un destacado método de venta. Según el estudio **"Compras online en España"**, de la edición 2024, realizado por el **Observatorio Nacional de Tecnología y Sociedad**, el **87% de los usuarios de Internet compra online**, lo que se traduce en 30 millones de españoles.



Base¹: Población total (16 a 74 años), España
Base²: Total de personas usuarias de Internet, España
Fuente¹: Encuesta sobre equipamiento y uso de TIC en los hogares 2023 (INE, 2023)
Fuente²: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

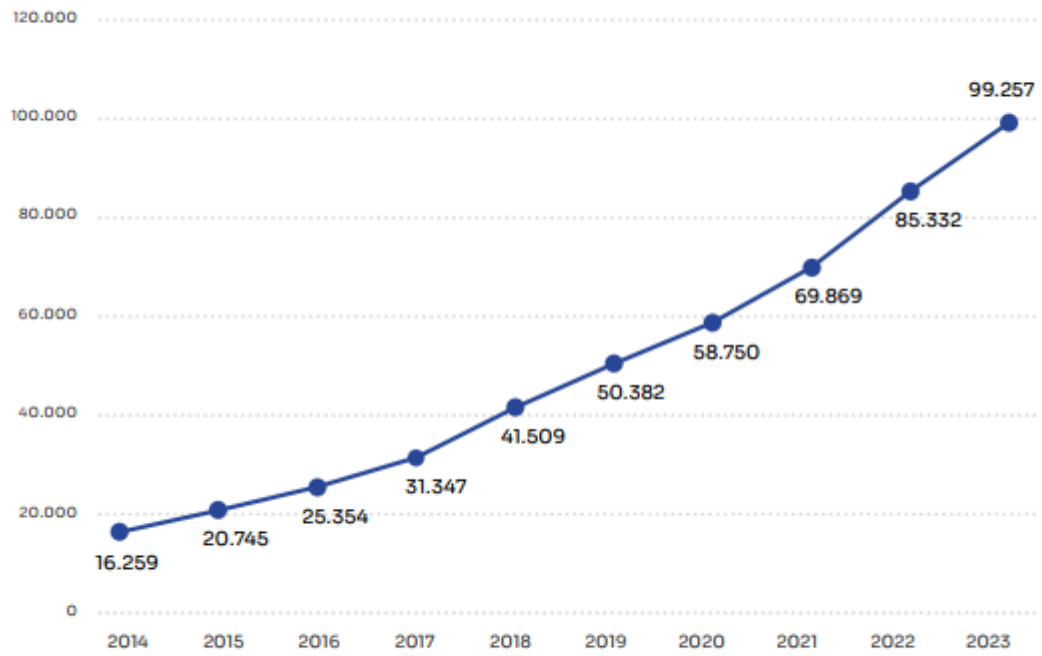
Nota: No se dispone del dato de compradores para el año 2021. Se trata de una estimación

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

El comercio electrónico business to customer (B2C) no ha hecho más que crecer, experimentando un ritmo de avance continuo en los últimos 10 años. Este sector es tan potente que incluso un 46,8% de las personas que no suelen realizar compras en línea se está planteando hacerlo en los próximos meses. Por ello, no es de extrañar que el **volumen de negocio B2C en España** haya aumentado en algo más de 13.900 millones de euros, llegando así a la cifra actual de **2023** de **99.257 millones de euros**.

Gráfico 10 - **Volumen de negocio del comercio electrónico B2C en España (2014-2023) (Millones €)**

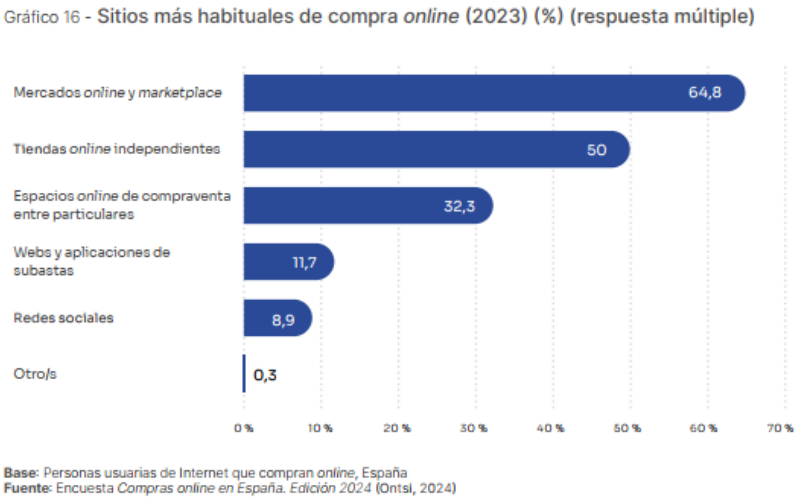


Fuente: Panel de hogares Ontsi (Ontsi, 2011-2017);
Encuesta *sobre equipamiento y uso de TIC en los hogares 2023* (INE, 2023);
Datos estadísticos de comercio electrónico 2023 (CNMC, 2024);
Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

Canales de compra online

Los compradores saben muy bien dónde adquirir su productos o servicios. Siguiendo con la tendencia anual, el **marketplace** se posiciona como plataforma

número uno en las preferencias de los consumidores con un nivel de preferencia del **64,8%** seguido de las tiendas online independientes con un 50%.
Sin embargo, algo a destacar de este último año es la **irrupción de las redes sociales como canal de compra** con un nivel de preferencia del **8,8%**. Esto evidencia el alto grado de penetración de las redes sociales y los teléfonos móviles en las actividades cotidianas de las personas, sugiriendo un posible **descubrimiento de carácter más conversacional del producto**.



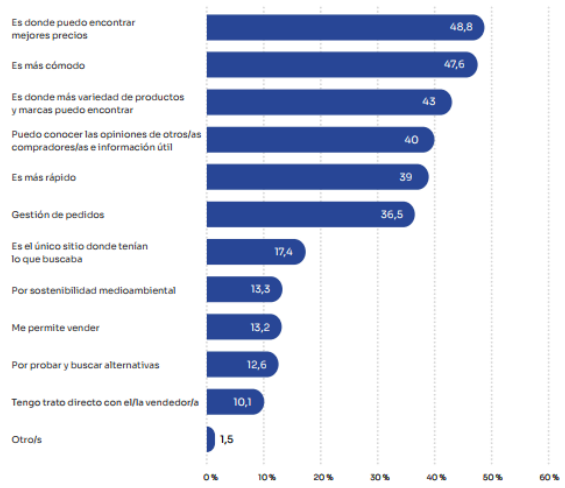
Motivación de compra del consumidor

¿Qué es lo que hace tan atractivo al marketplace para los consumidores? Los precios (48,8%), la comodidad (47,6%), la variedad de productos y marcas (43%) y la **posibilidad de acceder a las opiniones de otros usuarios (40%)** son los principales alicientes de los usuarios seguidos por la **sencillez en el proceso de compra (30,9%)** y la facilidad para **comparar información (28,6%)** como factores intrínsecos de la experiencia de compra online.

Certificado en Inteligencia Artificial

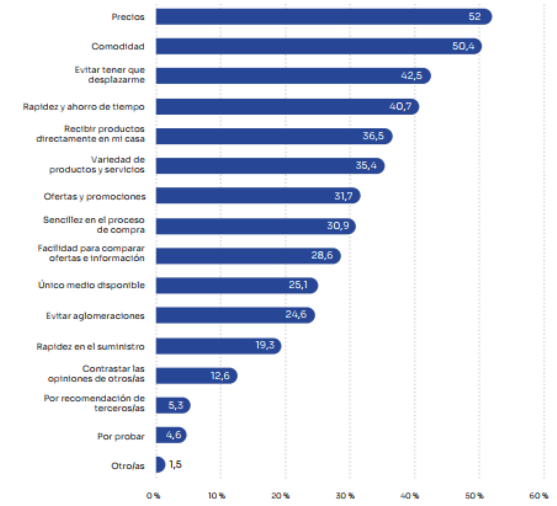
EOI & Samsung Innovation Campus

Gráfico 17 - Principales motivos para comprar en *marketplace* (2023) (%) (respuesta múltiple)



Base: Personas usuarias de Internet que compran *online*
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

Gráfico 60 - Principales razones para comprar *online* (2023) (%) (Respuesta múltiple)



Base: Personas usuarias de Internet que compran *online*, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

Experiencia de compra online

Para tener una visión holística de la experiencia de compra online, es necesario conocer la percepción de los clientes durante todo el proceso, tanto en sus puntos fuertes para afianzarlos como en los débiles para mejorarlos.

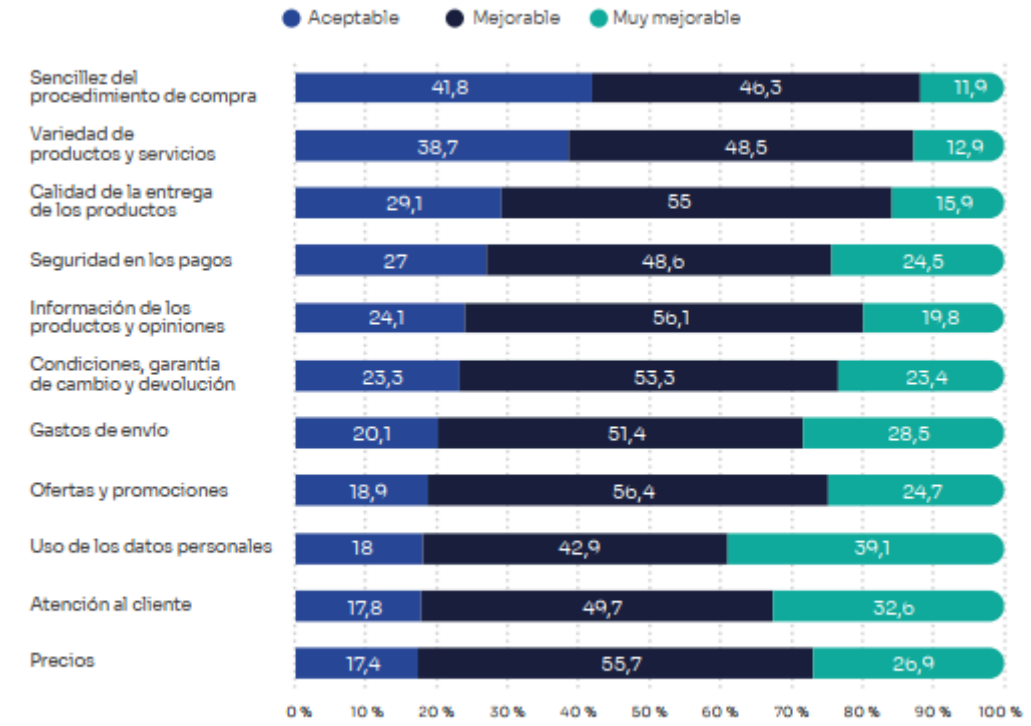
Con un **41,8%**, la **sencillez del procedimiento de compra** constituye el elemento mejor valorado de toda la transacción en línea, aunque hay que prestar atención al hecho de que esta cifra ha experimentado un ligero descenso en la valoración de hasta 12 puntos desde 2020 (53,1%) y 2022 con un 48,9%.

En cuanto a las facetas **menos valoradas** con necesidad de mejora, podemos destacar las ofertas y promociones, el **uso de los datos personales**, la **atención al cliente** y los **precios**.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Gráfico 59 - **Aspectos satisfactorios y mejorables de las compras *online* (2023)**



Base: Personas usuarias de Internet que compran *online*, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

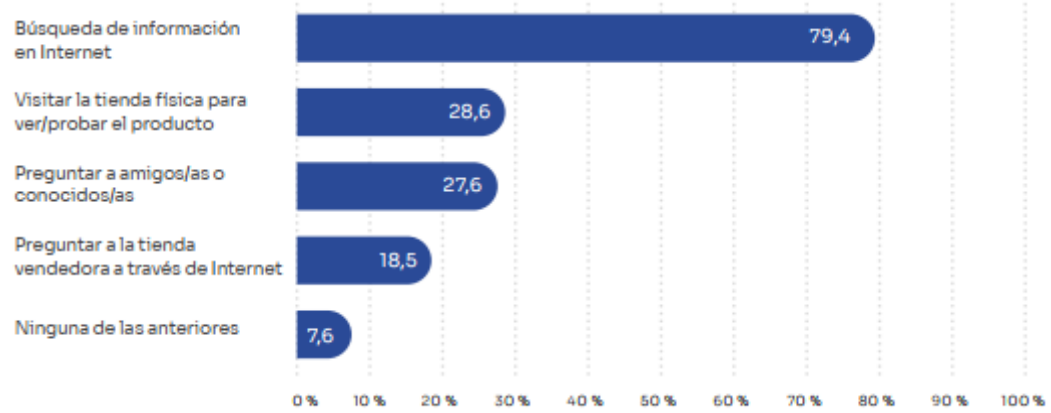
Comportamiento del consumidor antes de realizar compras online

El comercio electrónico ha transformado la forma en que los clientes compran y consumen productos y servicios. La comodidad que nos ofrece el móvil frente al desplazamiento físico ha supuesto que este dispositivo se convierta en el más utilizado para **buscar información previa a la compra**, siendo esta práctica la más común (**79,4%**) antes de realizar cualquier compra online.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Gráfico 12 - Acciones habituales antes de realizar compras *online*, 2023 (%) (Respuesta múltiple)



Base: Personas usuarias de Internet que no compran *online*, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

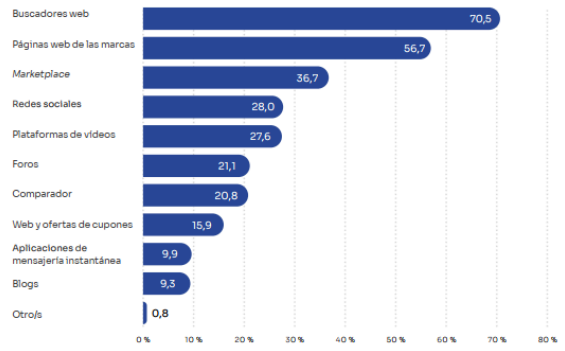
Como podemos ver en el siguiente gráfico, los **buscadores web** se posicionan como el canal preferido por los consumidores (**70,5%**) para informarse sobre un producto o servicio antes de realizar la compra.
Otras fuentes de información comunes son las propias **páginas web de las marcas** (**56,7%**) y los marketplace (36,7%). También tienen un peso importante en el proceso de búsqueda las redes sociales (28,0%) y las plataformas de vídeos (27,6%).

Por otro lado, es importante conocer qué tipo de información previa a la compra buscan los consumidores y cómo influye en el proceso de adquisición.
El gráfico a continuación nos muestra que, antes de realizar una compra en línea las personas se informan principalmente sobre el **precio** (**83,1%**) y las **características del producto** (**73,2%**) y las **opiniones** de los compradores (**63%**).

Certificado en Inteligencia Artificial

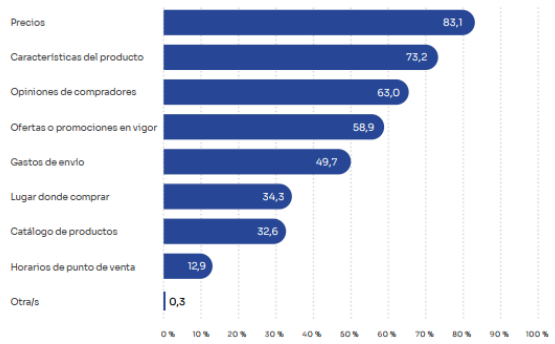
EOI & Samsung Innovation Campus

Gráfico 14 - **Canales de búsqueda de información sobre bienes y servicios (2023) (%)**



Base: Personas compradoras en línea que buscan información *online* sobre bienes y servicios, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

Gráfico 15 - **Tipo de información que se busca *online* sobre bienes y servicios (2023) (%)**



Base: Personas compradoras en línea que buscan información *online* sobre bienes y servicios, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

Análisis de la Inteligencia Artificial en el contexto actual del e-commerce

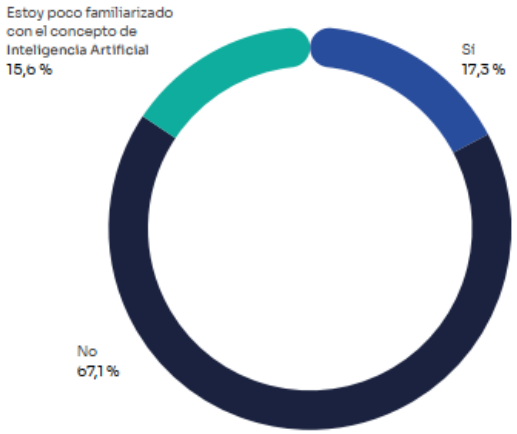
La Inteligencia Artificial (IA) se encuentra cada vez más presente en todos los sectores de la sociedad y, según el **“Informe del Estado del Comercio en España 2024”** de **Shopify**, el campo del e-commerce no se queda atrás. Sin embargo, la IA no solo está transformando el modelo de compra digital sino que las **expectativas** de los consumidores también están cambiando. Estos nuevos consumidores **buscan valor y calidad**, una **experiencia de compra coherente** omnicanal, y la **integración de la IA**, manteniendo al mismo tiempo un toque personal y humano durante toda la experiencia de compra.

Como podemos observar en el siguiente gráfico, la IA ya comienza a estar integrada en el comercio electrónico. El **17,3%** de los consumidores (pertenecientes al segmento de edad de 16 a 24 años) **utilizan la IA** para realizar sus compras onlines, aunque aún se necesita más tiempo para que gane presencia (el 67,1% afirma aún no haberla utilizado).

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Gráfico 43 - Distribución de personas que compran *online* que utilizan la Inteligencia Artificial (2023) (%) (Respuesta múltiple)

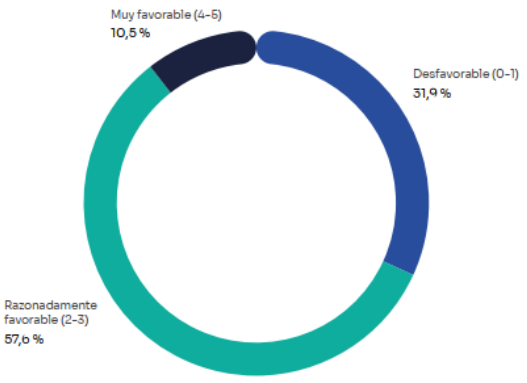


Base: Personas usuarias de Internet que compran *online*, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

¿Qué percepción tienen los consumidores sobre el uso de la IA en un e-commerce?

Aunque la IA todavía no se haya introducido completamente como herramienta para realizar compras en línea, podemos afirmar que la población tiende a valorar positivamente esta tecnología. El **57,6%** de los usuarios de e-commerce considera **razonadamente favorable la influencia de la IA** y el 10,5% muy favorable.

Gráfico 44 - Valoración de la influencia de la Inteligencia Artificial en las compras *online* (2023) (%)



Base: Personas usuarias de Internet que compran *online* y están familiarizadas con la IA, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

Aunque aún podamos observar algunas resistencias por parte de la población (el 31,9% restante considera desfavorable el uso de la IA) podemos destacar toda una serie de nuevas tendencias y expectativas. Según el Informe de Shopify, los compradores online:

1. **Muestran interés y expectativas positivas:**

   Los consumidores muestran un interés considerable en cómo la IA puede mejorar su experiencia de compra.

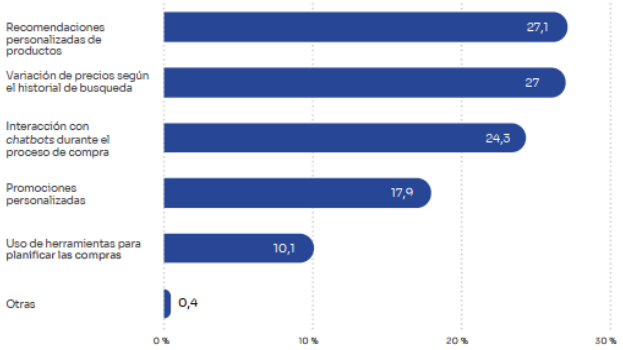
   - Un **72%** de los consumidores de e-commerce esperan que la **IA** haga **más eficientes sus compras**, facilitando el **descubrimiento de productos** y **agilizando los procesos de compra**. En este sentido, los motores de búsqueda de productos, potenciados por IA, pueden ayudar a los clientes a encontrar rápidamente lo que buscan mediante funciones predictivas como la de autocompletar o las vistas previas de productos.
   - Un **55%** esperan que las marcas ofrezcan **experiencias de compra tecnológicamente avanzadas**, superando el promedio europeo del 47%.
   - Además, los consumidores creen que la IA puede ayudarles a encontrar ofertas, obtener **información sobre productos**, y descubrir nuevas marcas.

   El siguiente gráfico nos muestra que los consumidores de comercio electrónico destacan positivamente el uso de la IA para recibir **recomendaciones personalizadas** de productos (**27,1%**) y **precios dinámicos** basados en el historial de búsqueda (**27%**).

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Gráfico 45 - Usos de la Inteligencia Artificial detectados en el proceso de compra en línea por personas familiarizadas con el concepto de IA que compran *online* (2023) (%)



Base: Personas que compran *online* y están familiarizadas con el concepto de IA, España
Fuente: Encuesta *Compras online en España. Edición 2024* (Ontsi, 2024)

2. **Personalización:**

   Los consumidores valoran la personalización, por ello es importante que la plataforma de e-commerce sea flexible y se **adapte a sus necesidades** durante su experiencia de compra. Esto constituye una gran oportunidad, puesto que con la IA podemos analizar el comportamiento y las preferencias del usuario para ofrecer recomendaciones personalizadas en tiempo real.

3. **Confianza en la IA:**

   Los consumidores consideran que las empresas integren la IA en su proceso de compra es una inversión valiosa. Perciben que son empresas que **priorizan al cliente** y, por ello, están dispuestos a aprender más sobre ella.

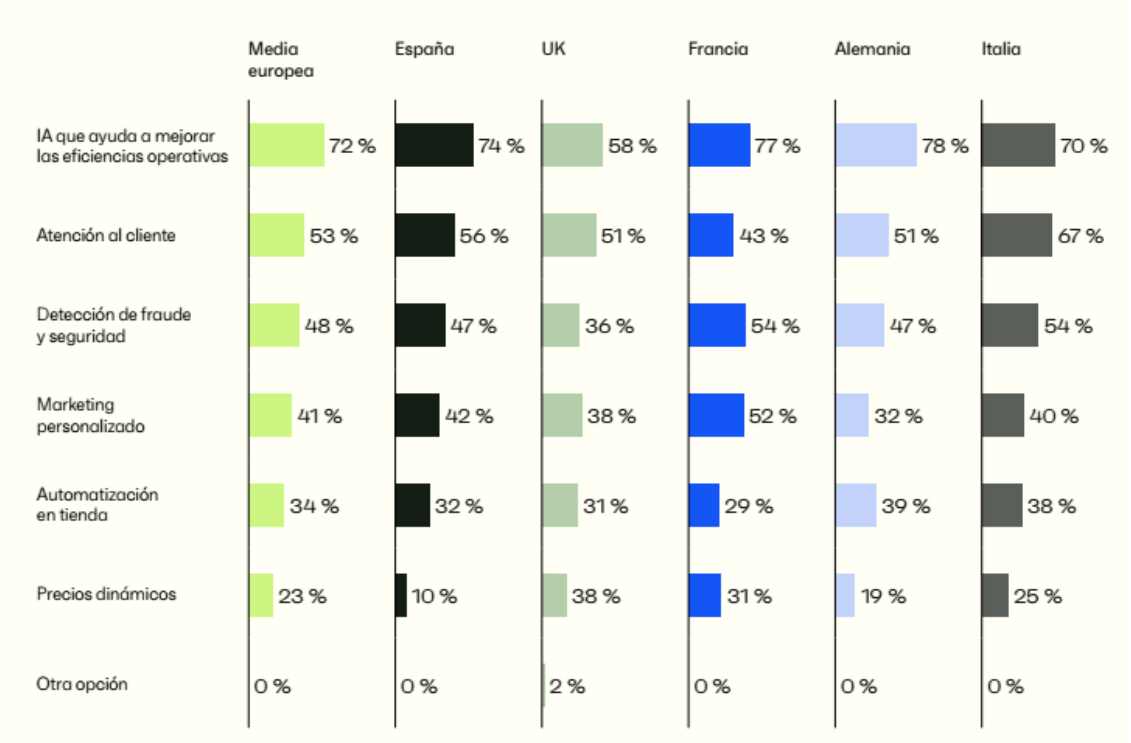
   Sin embargo, también es importante que las empresas sean transparentes en el uso de la IA y compartan ejemplos de cómo utilizan los datos de sus clientes.

A modo de resumen, los nuevos consumidores están cada vez más interesados en las aplicaciones de IA que mejoran la experiencia de compra, como la

personalización o la **eficiencia en la búsqueda**. No obstante, también valoran el trato humano y la transparencia en el uso de la tecnología, a la vez que son conscientes de la importancia de la privacidad de sus datos. Por ello, para satisfacer las expectativas de este perfil de consumidor, los e-commerce tienen que encontrar un equilibrio entre la innovación tecnológica y la confianza y calidad del servicio.

¿Cómo se está introduciendo la IA en el comercio electrónico?

Según el Informe de Shopify, el **36% de las empresas españolas están invirtiendo o planean invertir en IA en el próximo año** y hasta un 90% tienen programado mantener o aumentar su presupuesto tecnológico. En el siguiente gráfico, podemos ver en qué funciones quieren las empresas implementar la IA:



Como podemos observar, la **principal ventaja** que los gestores de e-commerce perciben de la IA es que puede **aumentar la eficiencia operativa**. En este sentido, la IA puede ayudar a los comerciantes a optimizar procesos automatizando tareas, personalizando la experiencia del cliente o gestionando el inventario y mejorando la toma de decisiones. Además, la IA también permite a las empresas reducir costes a través de precios dinámicos, optimizando así, sus estrategias de precios para maximizar ingresos y mantener la competitividad.

Sin embargo, las empresas también tienen presente los desafíos que la IA presenta en relación a la privacidad de los datos:

1. **Preocupaciones sobre la autenticidad:**

   Algunas empresas muestran preocupación por la posible **falta de autenticidad o interacción humana** que puede resultar de la implementación de la IA. Sin embargo, esta preocupación no es generalizada, ya que la mayoría de las empresas no la ven como su mayor desafío.

2. **Brecha de habilidades:**

   A pesar del interés en la IA, un **65%** de empresas notan que existe una brecha de habilidades a la hora de utilizar la IA de forma eficaz y consideran que la **falta de personal capacitado en IA** es un desafío importante.

A modo de conclusión, podemos decir que la IA se está integrando de forma muy eficiente en el sector del comercio electrónico, transformando a su paso las expectativas de los consumidores. Si bien la IA ofrece muchas **oportunidades** para **mejorar la eficiencia y la personalización**, también plantea **desafíos** como la necesidad de abordar las **preocupaciones éticas y de privacidad** en relación a los datos de los clientes, la **complejidad de su integración** y la necesidad de contar con **personal especializado** para implementar estas soluciones de manera efectiva. A pesar de todos estos retos, los consumidores del comercio electrónico

esperan con interés la incorporación de la IA en toda la experiencia de compra. Confían en que la IA haga **más eficientes sus compras**, facilite el **descubrimiento de productos** y **agilice el proceso de compra**.

Por todo esto, en este proyecto vamos a explorar una funcionalidad clave en cualquier plataforma de comercio electrónico que puede resolver de manera eficiente estas necesidades de los consumidores: **el buscador**. Al fin y al cabo, para comprar un producto, los clientes, primero, deben poder encontrarlo. Así veremos, a continuación, **cómo la IA puede mejorar la eficiencia de compra a través de la búsqueda y el descubrimiento de productos**.

El descubrimiento de productos

El **Product Discovery** o *Descubrimiento de Productos* es el **proceso mediante el cual los usuarios de un e-commerce encuentran un servicio o artículo que desean comprar**. Este proceso puede tomar diversas formas, desde investigar productos hasta explorar una categoría de productos específica dentro del sitio web.
Para ello, los usuarios pueden tomar dos caminos diferentes, pero compatibles entre sí:

- La **navegación** a través del menú, categorías y filtros
- La **búsqueda** mediante el buscador o motor de búsqueda interno

Descubrimiento por Navegación:

En el sector de e-commerce, diseñar la *home* o página de inicio y una navegación por categorías que resulte fácil y sencilla es un verdadero reto. Requiere una **arquitectura de la información** y un **sistema de etiquetado** sólidos. Además, es

importante contar con páginas de categorías intermedias bien optimizadas y un diseño estético y equilibrado. Y esto es sólo el principio.

Sin embargo, por mucho que nos enfoquemos en que el diseño de la web cumpla los criterios estéticos establecidos o que las páginas intermedias sean relevantes, **si los elementos básicos de la navegación no están bien definidos, la tasa de conversión o la reputación de la marca pueden verse afectadas**. A modo de ejemplo, si los usuarios no encuentran un determinado tipo de producto, pueden llegar a la conclusión de que el sitio simplemente no posee ese artículo. Este error en el diseño de la arquitectura del sitio web no solo provoca la pérdida inmediata de una venta, sino también numerosas pérdidas futuras, pues el sitio web ya se habrá posicionado y creado una mala experiencia en la mente del consumidor. Así, es poco probable que un usuario vuelva a un sitio que cree que no tiene el producto que necesita.

Statistics indica que el **75% de los usuarios abandonan los sitios web de comercio electrónico si no encuentran lo que están buscando en menos de 15 segundos**.
Incluso afirman que hay una relación directa entre las ventas y la calidad de la jerarquización de los productos en la web. **Los sitios que están presentan una mala organización de su catálogo de productos venden 50% menos que los que sí lo están**. Esto demuestra que un índice de venta bajo no se debe sólo a malas estrategias de marketing, sino también a cómo se presenta la información de cara a potenciales consumidores.

La **taxonomía** forma parte de la arquitectura de la información de un sitio web. Podríamos definirla como: **“un sistema que se enfoca en organizar y etiquetar el contenido para que sea fácil de entender y facilitar la navegación”**. Es decir, en un comercio electrónico del sector de la moda, cuando el usuario llega a la página de inicio, se puede encontrar que la ropa está organizada por categoría de

hombres y mujeres. Después de hacer esta primera selección, puede ver que la ropa se vuelve a dividir en camisas, pantalones, zapatos, etc.

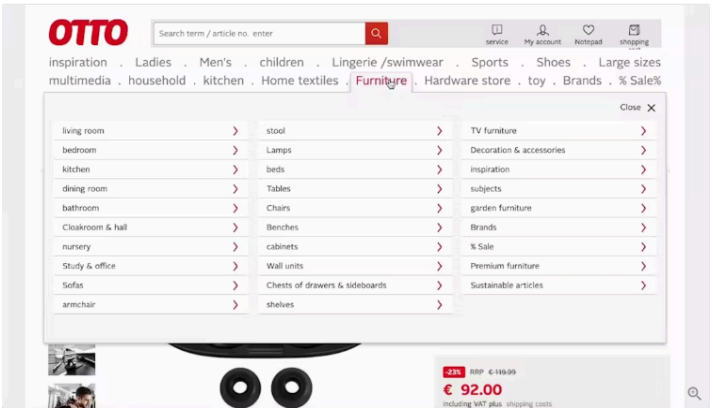
Llegados a este punto, podemos entender que una taxonomía correcta permitirá a los usuarios encontrar fácilmente lo que están buscando a través de la navegación por categorías y filtros de los productos del e-commerce.

Según el **Instituto Baymard de UX Research**, en cuanto a la taxonomía de categorías, el **76% de los sitios web de comercio electrónico presentan un rendimiento "mediocre" o "pobre"**, dificultando, así, a los usuarios encontrar los productos que buscan. Esto se debe a que los sitios web a menudo categorizan los productos de manera incorrecta. Entre otras cuestiones, esto da lugar a que los usuarios no puedan ni mezclar ni combinar diferentes criterios para obtener la lista de productos que desean. A continuación, vamos a profundizar en los principales problemas de taxonomía que comprometen la navegación y el descubrimiento de productos a los usuarios:

- **Categorías y subcategorías mal organizadas:** El 85% de los sitios web no dividen las categorías y subcategorías en partes manejables y presentan la información a los usuarios mediante largas listas que resultan abrumadoras. Esto dificulta que los usuarios obtengan una visión general del catálogo de productos y seleccionen las opciones más relevantes. **La mayoría de los usuarios se sentían abrumados cuando se les presentaban más de 10 subcategorías entre las que elegir**.

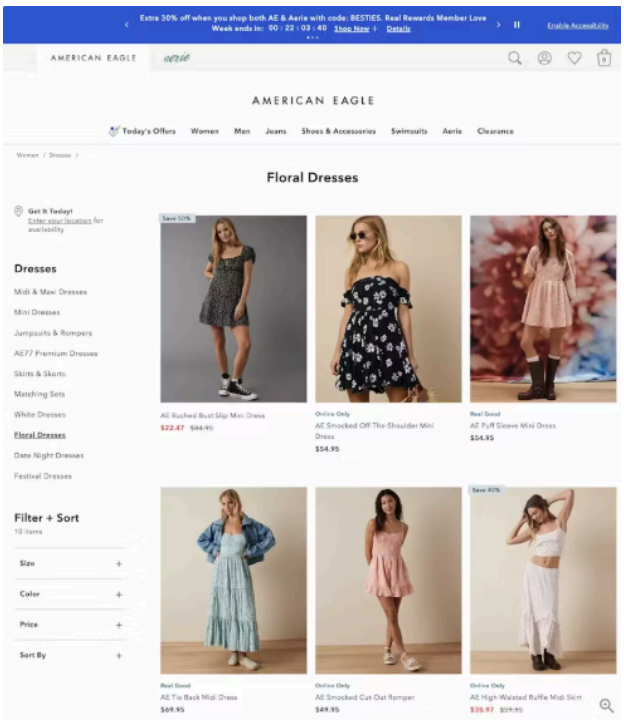
Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus



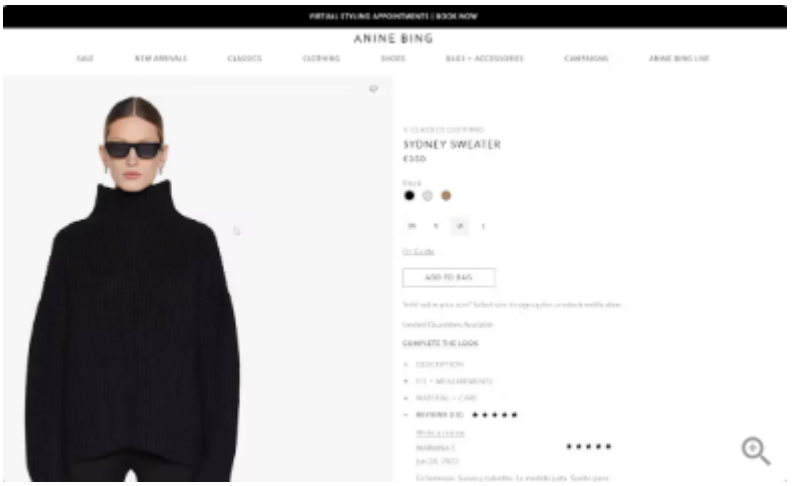
"Podría estar menos saturado... más fácil de navegar. Si estás buscando algo, tienes que hacer mucho esfuerzo para encontrarlo". Presentar una lista de 29 subcategorías como lo hace el sitio alemán Otto en la categoría “Muebles”, hace que los usuarios tengan dificultades para elegir el mueble más relevante.

- **Sobrecategorización de productos:** El 91% de los e-commerce implementan incorrectamente tipos de un mismo producto con atributos compartidos como categorías separadas. Para entenderlo mejor, visualizaremos el siguiente ejemplo:



En American Eagle, los “Vestidos florales” se encuentran como subcategoría de “Vestidos”. Esto conlleva que un usuario no pueda obtener una visión general de “Todos los vestidos”. Tampoco le permite personalizar la búsqueda a través de una lista con “Vestidos florales” y “Minivestidos”.

- **Falta de claridad en la navegación principal:** El 91% de los sitios no muestran dónde se encuentra el usuario dentro de la jerarquía del sitio en la navegación principal.

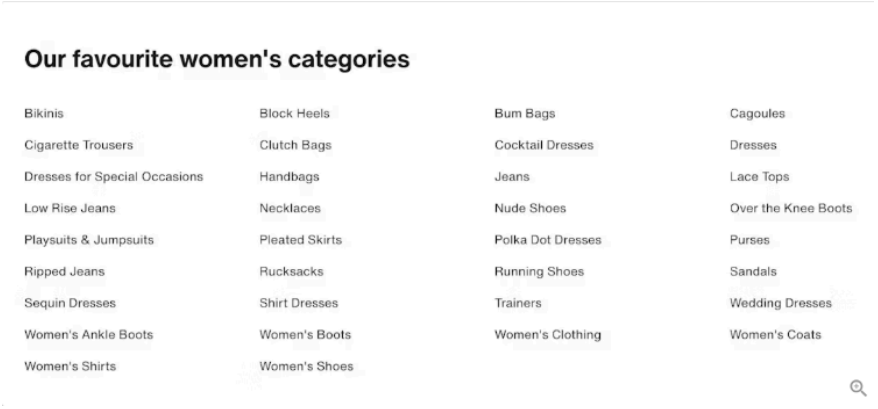


Cuando un usuario llega a esta página en la web de Anine Bing, le puede resultar difícil saber la ubicación en la jerarquía global de la web. Los breadcrumbs, “Ropa clásica”, no están ubicados en su posición habitual, justo encima del título del producto). Por ello, si el usuario llega directamente a esta página y quiere ver productos similares, tendrá que adivinar en qué categoría principal de navegación deben hacer clic para encontrarlos.

- **Miniaturas de subcategorías difíciles de interpretar:** En cuanto a las páginas intermedias de categorías y subcategorías, si estas carecen de miniaturas o poseen miniaturas difíciles de interpretar, los usuarios suelen tener más problemas para determinar qué subcategoría quieren explorar. Y si estas subcategorías sólo presentan texto, aún les resulta más difícil decidir por cuál navegar, por lo que una guía visual siempre es importante.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus



Un ejemplo de mala práctica la podemos encontrar en esta página de Zalando, ya que no presenta ninguna imagen para presentar las subcategorías.
Como hemos podido observar, una gran parte de los comercios web presentan problemas de taxonomía y navegación, por lo que si el descubrimiento de productos se realiza por navegación y resulta complicado, los usuarios recurrirán a la barra de búsqueda. Y si, en este caso, el buscador tampoco proporciona resultados relevantes, la tasa de rebote del comercio aumentará.

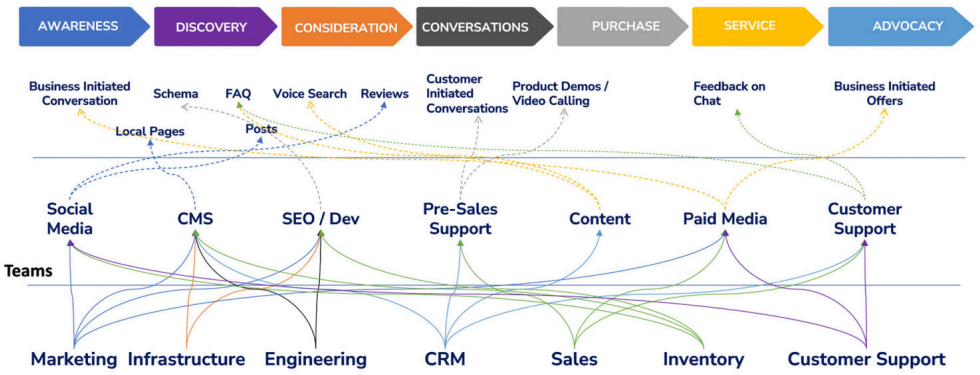
Descubrimiento por Búsqueda:

La búsqueda de productos mediante el motor de búsqueda del comercio electrónico es un tipo de consulta a la que **los usuarios suelen recurrir cuando han identificado una necesidad de un producto pero no están seguros de qué marca o modelo específico necesitan o cuando saben exactamente qué producto quieren comprar.** De esta manera, un buscador representa para los usuarios un atajo en su proceso de búsqueda. Sepan o no exactamente lo que quieren comprar, recurren a esta herramienta con la esperanza de llegar antes a un potencial producto. Supone una alternativa mucho **más rápida que la navegación web**, siendo este último un proceso un poco más costoso para el usuario ya que necesita entender la arquitectura del sitio web para encontrar el producto que está buscando.

Por todo esto, es fundamental que un algoritmo de búsqueda funcione correctamente. Para ello, debe ser capaz de **entender la intención del usuario** y **proporcionar resultados relevantes y precisos** que satisfagan sus necesidades. Si por alguna razón el algoritmo falla, los usuarios pueden llegar a frustrarse y abandonar el sitio, lo que se traduce para el comercio en la **pérdida de oportunidades de negocio**. Por ello, podemos decir que el buen funcionamiento del algoritmo de búsqueda **afecta directamente la experiencia de usuario, la visibilidad de un sitio web y las conversiones de ventas**.

El rol de la búsqueda en el embudo de conversión

Profundizando un poco más en este tema, podemos decir que la búsqueda juega un papel crucial en el ***funnel* o embudo de conversión**. Este se define como el **proceso que sigue un usuario desde que descubre un producto o servicio hasta que realiza una compra**.



Así, vamos a explorar qué papel juega la búsqueda en cada una de las etapas de este embudo:

- **Descubrimiento:** Al inicio del embudo, la búsqueda ayuda a los usuarios a **encontrar información sobre productos o servicios que pueden resolver sus necesidades o problemas**. De esta manera, la búsqueda puede ayudar

a descubrir nuevos productos o marcas. En este punto, las **consultas pueden ser informativas**, donde los usuarios buscan aprender sobre un tema, o pueden ser más orientadas a un problema, donde el usuario ya tiene una necesidad pero no sabe la solución exacta.

- **Consideración:** A medida que los usuarios avanzan en el embudo, la búsqueda les permite **comparar opciones, leer reseñas y buscar información adicional sobre los productos o servicios que están considerando**. En esta etapa, las **búsquedas pueden ser de tipo comercial**, con el usuario comparando diferentes productos o servicios antes de tomar una decisión.

- **Decisión o Conversión:** En la última parte del embudo, **la búsqueda facilita que los usuarios encuentren el producto o servicio específico que desean comprar y realicen la transacción**. Aquí, las **búsquedas son transaccionales**. Los usuarios buscan términos como "comprar ahora," "barato," o "envío gratis". Por ello, el motor de búsqueda debe llevar a los usuarios directamente a las páginas de producto o al proceso de compra.

- **Fidelización:** Una buena experiencia de búsqueda puede **aumentar la probabilidad de que los usuarios regresen al sitio web en el futuro y realicen más compras**. La personalización de los resultados de búsqueda puede ayudar a que los clientes se sientan comprendidos, y mejorar las conversiones desde el inicio del proceso de compra.

Tipos de búsqueda

Una vez entendido la importancia de un buen algoritmo de búsqueda en todo el proceso de compra, desde el descubrimiento de la necesidad hasta la fidelización del comercio electrónico, vamos ahora a **explorar los diferentes tipos de**

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

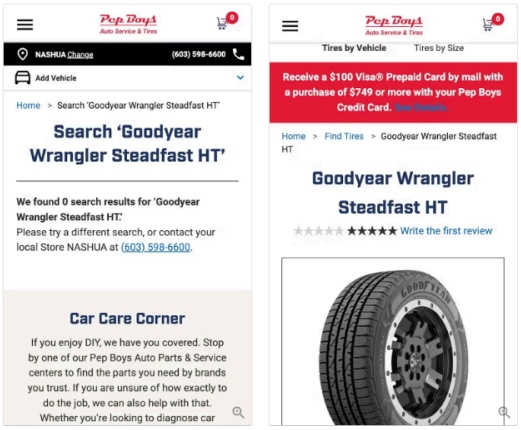
búsqueda que un usuario puede realizar cuando decide optar por utilizar el motor de búsqueda interno de un comercio en línea.

Para ello, nos vamos a fijar en el **"Estudio de usabilidad para la búsqueda de productos" de Baymard de 2024**. En este estudio de más de **5.550 horas de investigación**, se analizaron **12 sitios web** pertenecientes tanto a minoristas masivos como a sitios más pequeños. Los participantes fueron: Staples, HP, Office Depot, Best Buy, B&H Photo, Newegg, GAP, H&M, AEO, Nordstrom, L.L. Bean, Urban Outfitters.
A lo largo de este proceso de investigación, se llevaron a cabo **más de 219 sesiones de pruebas de usabilidad** en las que se encontraron **más de 1.000 problemas relacionados con la búsqueda de productos**.

El proceso de descubrimiento de productos puede incluir diferentes tipos de búsquedas, como:

- **Búsquedas exactas:**

  El **33%** de los sitios **no admiten búsquedas exactas** de nombres de productos o números de modelo.



En la web de Pep Boys, si un usuario realiza una búsqueda exacta de un neumático concreto, no obtiene ningún resultado (primera imagen), aunque los productos se venden en el sitio y se pueden encontrar navegando (segunda imagen).

Cuando los **usuarios saben exactamente qué producto quieren comprar, suelen introducir el título o el número de modelo** del producto tal cual en el campo de búsqueda. Incluso, durante el estudio, algunos usuarios primero buscaban el producto en cuestión en un sitio externo, y después, copiaban y pegaban el nombre en el campo de búsqueda del e-commerce para asegurarse de que lo escribían bien.

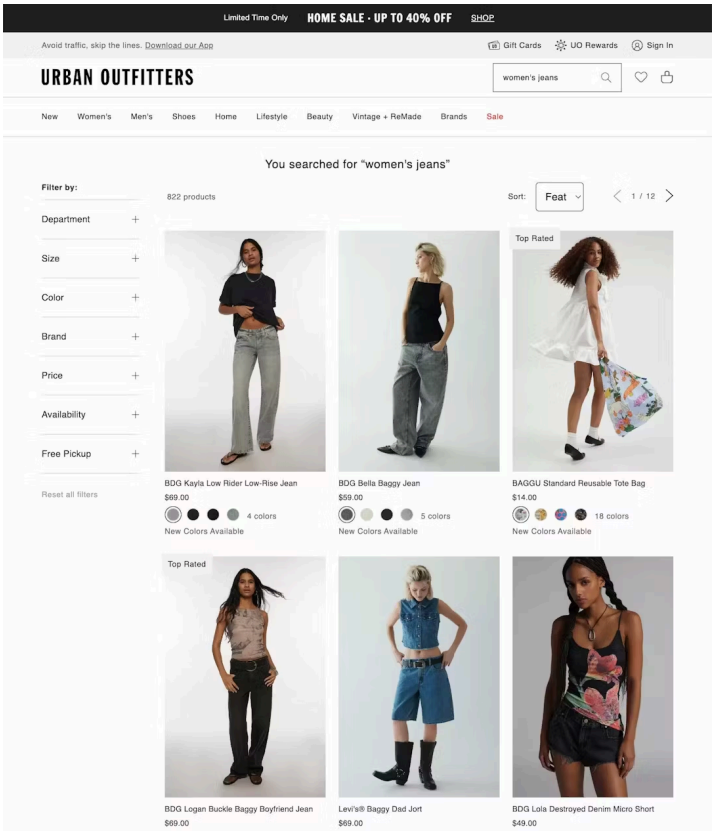
En este caso, como el usuario sabe exactamente el nombre del producto, podemos pensar que se trata de un problema sencillo a resolver a nivel técnico. El **algoritmo de búsqueda** simplemente tendría que encontrar la **concordancia de palabras clave y atributos del producto** para mostrárselo finalmente al usuario. Sin embargo, este algoritmo tiene que tener en cuenta otros casos de uso que se pueden dar en la búsqueda, como las faltas de ortografía o los nombres alternativos de productos.
Sin embargo, a pesar de ser la solución más fácil de soportar técnicamente, un tercio de los sitios web obtuvieron resultados mediocres o peores.

- **Búsquedas por tipo de producto:**

El 29% de los sitios tienen problemas con las búsquedas por tipo de producto (por ejemplo, "vaqueros de mujer" o "portátiles").

Cuando los usuarios no están buscando un producto concreto, sino un tipo de producto para explorar, recurrirán a la **búsqueda por categoría o subcategoría de productos** (por ejemplo, «sandalias» u «ordenadores portátiles»).

Normalmente, cuando un usuario quiere realizar este tipo de búsquedas, optará primero por utilizar el **menú de navegación**. No obstante, si este menú no está bien categorizado o jerarquizado y el usuario tiene dificultades para encontrar la categoría de producto que necesita, se escogerá el **buscador**.



En Urban Outfitter's, si un usuario busca “vaqueros de mujer”, el sitio web muestra 822 productos con artículos irrelevantes y sin sugerencias de categorías a simple vista.

Por situaciones como esta, durante el estudio, los participantes suponían que los resultados pobres o limitados con los que se encontraban, eran la selección completa del sitio para esos productos y no intentaban seguir navegando.

- **Búsquedas por características:**

> El 34% de los sitios no admiten búsquedas por atributos de productos (por ejemplo, "sofás de tela" o "chaqueta negra").

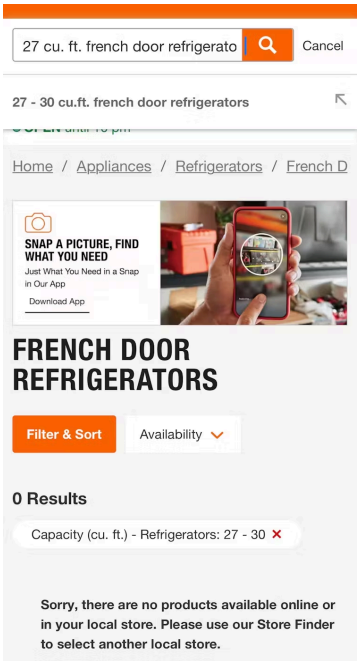
La búsqueda por características es un tipo de consulta en la que **el usuario incluye uno o más atributos del producto en la barra de búsqueda**. Por ejemplo, si el usuario está buscando un tipo de chaqueta específico, escribirá en el buscador algo como “chaqueta de cuero”.

Cuando hablamos de características nos referimos a una amplia gama de atributos de producto como el **color** (“vestidos rojos”), **material** (“escaleras de aluminio”), **talla** (“pantalones de la talla 34”), **especificaciones de rendimiento** (“móvil de 12 GB de RAM”), **formato** (“DVD de La historia Interminable”), **precio** (“mochila de 20€”) o **marca** (“pintalabios de Revlon”), entre otros.

Normalmente, **estos atributos se combinan con diferentes tipos de búsqueda**. Se puede dar el caso en el que un usuario realice una búsqueda por “Tipo de producto”, aunque también se han observado combinaciones con consultas temáticas, sintomáticas y de compatibilidad, y luego combinar esta búsqueda con una característica del producto. De esta manera, el usuario obtendría, en los resultados de búsqueda, el subconjunto de productos deseado. Un ejemplo de este tipo de consultas sería “champú voluminizador sin parabenos” o “chaqueta azul transpirable Decathlon”.

En cuanto a implementación, la solución ideal sería aplicar cualquier característica buscada como filtro en la página de resultados (en el caso de que exista un filtro específico para dicha característica).
Siguiendo esta práctica, la percepción y experiencia del usuario mejoraría ya que aumenta la transparencia (el usuario puede ver lo que se incluye) y

el control (el usuario puede activar o desactivar los filtros) del mismo. Esto contribuye a una experiencia de compra más personalizada, lo que logrará una mayor satisfacción en el usuario y asegura un posible retorno, lo que se traduce, en consecuencia, a una mayor tasa de retención. Sin embargo, podemos ver que casi el 40% de los e-commerce no prestan atención a este tipo de búsqueda.



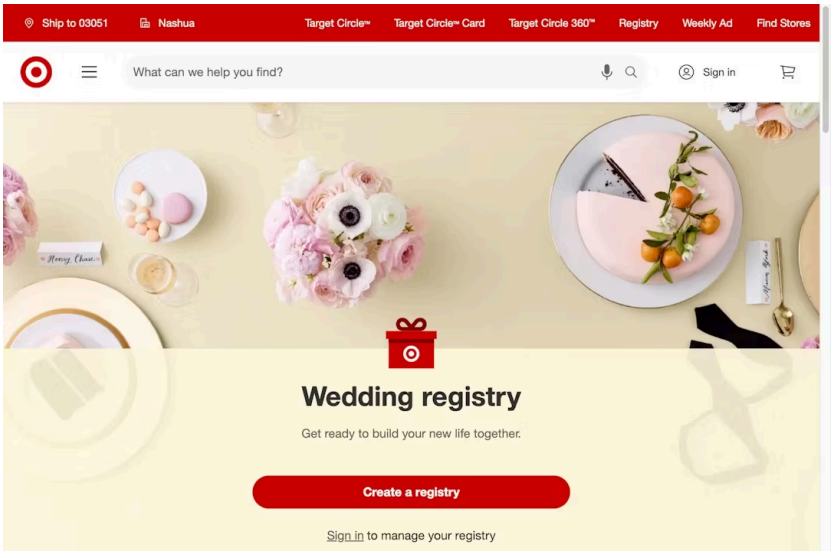
En Home Depot, si el usuario introduce en el buscador una consulta en la que se incluya la capacidad y el tipo de frigorífico, el sitio devuelve 0 productos, a pesar de que existen filtros con esas características.

- **Búsquedas por caso de uso:**

El 36% de los sitios tienen problemas con las búsquedas por cómo, dónde o cuándo se utilizará un producto (por ejemplo, "regalo de boda" o "ordenador portátil para juegos").

Este tipo de búsqueda describe **cómo, dónde y cuándo se utilizará el producto**. Algunos ejemplos donde se estaría poniendo en práctica este tipo de búsqueda serían consultas por lugares de uso del producto (“muebles de salón”), condiciones ambientales (“chaqueta de primavera” o “saco de dormir para tiempo frío”), acontecimientos (“regalo de boda”) o actividades (“pantalón de fútbol”).

Por norma general, **los usuarios que realizan este tipo de búsqueda no están familiarizados con el catálogo de productos y quieren encontrar rápidamente un producto que satisfaga una necesidad específica**. Por ello, al realizar la búsqueda utilizando sus modelos mentales sobre en qué contexto utilizarían el producto, los resultados que les puede devolver el algoritmo de búsqueda pueden llegar a ser ausentes, escasos o irrelevantes. Esto genera, en los usuarios, una impresión de que los productos buscados simplemente no están disponibles en el e-commerce y obliga a los usuarios a tener que reescribir la consulta y volver a intentarlo.



En Target, cuando un usuario busca "regalo de boda", la web le redirige a la página de registro de bodas, sin que pueda buscar en ella productos apropiados para la ocasión.

Para que este tipo de búsquedas que no incluyen palabras clave o filtros proporcionen resultados relevantes, el algoritmo tiene que interpretar tanto el significado de la propia consulta como el etiquetado interno de los productos. Esto se puede conseguir haciendo que los atributos actúen como valores de filtrado reales.

Sin embargo, el **36%** de los sitios web de compra, tienen problemas para gestionar este tipo de búsquedas más contextuales si aquellos atributos descriptivos que incluye el usuario en su consulta no forman parte del título del producto.

- **Búsquedas por abreviaturas y símbolos:**

El 50% de los sitios no admiten búsquedas de abreviaturas o símbolos comunes (por ejemplo, "13 centímetros" vs "13 cm").

Cuando los usuarios quieren encontrar un producto con unas **características medibles específicas** o que se encuentre en un rango de precio, a menudo utilizan atajos lingüísticos o, más comúnmente conocidos como abreviaturas o símbolos.

En el caso de las abreviaturas, resultan un recurso más fácil de soportar a nivel técnico. Solo requiere relacionar los diferentes términos. Por ejemplo: “mm” se emparejaría con “milímetros”.

Los símbolos, en cambio, pueden tener significados diferentes en función de su contexto y orden en la consulta de búsqueda, por lo que pueden resultar un poco más complicados. Con esto nos referimos a que, por ejemplo, el símbolo “-” puede indicar un signo menos (“saco de dormir para -10 grados”) o un intervalo (“jerséis de 20 a 50 euros”).

En este último caso, el signo “-” no sólo cambia su significado como resta, sino que, actúa como una instrucción de filtrado para acotar la lista de resultados.







El algoritmo de búsqueda de REI no entiende la similitud entre “11 ft.” y “11 foot” ya que no devuelve los mismos resultados que “11’”.

Por todo esto, es importante que el algoritmo de búsqueda pueda entender correctamente este tipo de alternativas a la hora de generar resultados. Sin embargo, el 50% de los sitios web no admiten ni siquiera las búsquedas más básicas de abreviaturas o símbolos en el sitio web.

- **Búsquedas por compatibilidad:**

El 31% de los sitios no facilitan la búsqueda de productos compatibles o accesorios (por ejemplo, "cargador para portátil Dell").

A menudo, los usuarios no conocen el nombre del **accesorio** que necesitan, sino sólo los datos del producto que ya poseen. Por ello, los usuarios suelen realizar búsquedas de artículos compatibles. Esto es, introduciendo el nombre o la marca de un producto que poseen junto con el tipo de accesorio o repuesto que buscan.

No obstante, hay que tener en cuenta que el algoritmo de búsqueda tiene que combinar dos o más productos compatibles para dar una respuesta adecuada.

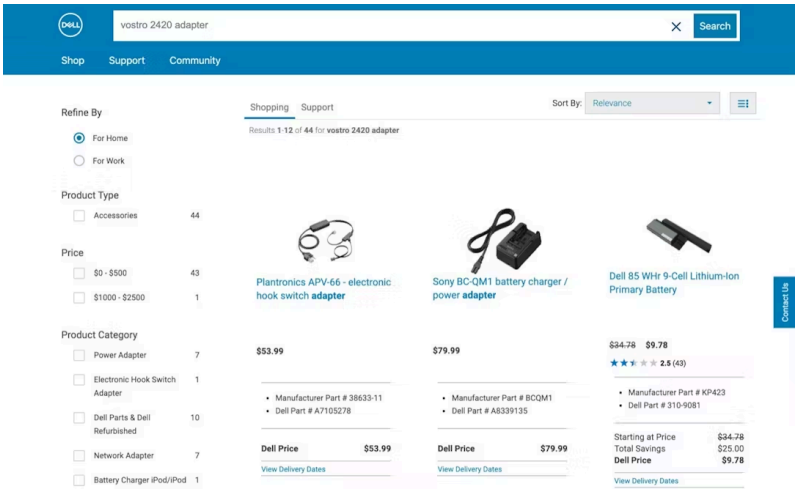
El usuario suele generar este tipo de búsqueda de la siguiente manera:

- Combina la marca y el modelo o serie del producto principal (“Portátil Dell XPS 13 Touch”) junto con un tipo de accesorio (“adaptadores”).
- Combina la marca y el tipo del producto principal (“portátil Dell”) junto con un tipo de accesorio (“protector de pantalla”).
- Utiliza sólo el nombre del producto principal (“macbook pro”) y espera encontrar accesorios compatibles entre los resultados.
- Utiliza sólo el nombre del accesorio o de la pieza (“adaptador para portátil” o “correa para aspiradora”).

Normalmente, los usuarios que saben el modelo exacto de su producto suelen incluirlo para buscar accesorios compatibles. Pero, también puede darse el caso de que los usuarios no conozcan o recuerden el modelo concreto de su producto, por lo que también es importante la compatibilidad con tipo y marca más genérico.
Dado que las relaciones de compatibilidad pueden ser complejas, puede ser difícil identificarlas y analizarlas a partir de las consultas enviadas en el campo de búsqueda. Por ello, el 31% de los e-commerce no saben introducir bien técnicamente este tipo de búsquedas, por lo que los usuarios pueden tener dificultades para encontrar accesorios o piezas relevantes.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus



Si un usuario realiza una búsqueda en Dell de “adaptador Vostro 2420”, la web mostrará productos irrelevantes (un adaptador de teléfono, un adaptador de cámara y una batería incompatible).

- **Búsquedas por síntomas:**

El 38% de los sitios no ayudan a los usuarios que buscan productos basándose en un problema o síntoma ("dolor de rodilla").

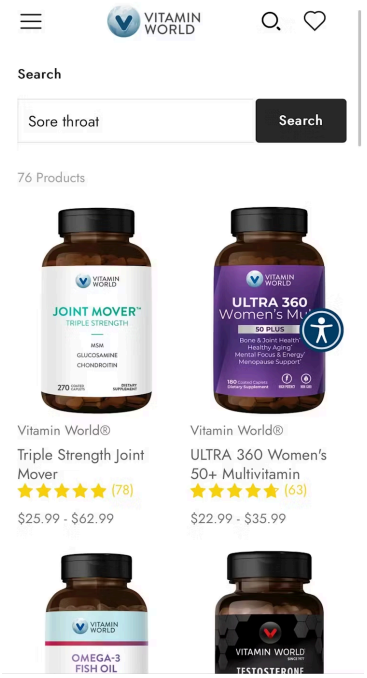
Con la información que tenemos hasta ahora, sabemos que, cuando los usuarios conocen el producto concreto que buscan, suelen recurrir a las búsquedas exactas. En caso contrario, si no conocen el producto exacto o no están seguros de cuál quieren, suelen buscar por tipo de producto.

Sin embargo, hay veces que los usuarios ni siquiera saben qué tipo de producto necesitan. Por lo que, solo saben **describir el problema** que tienen (“alfombra manchada”) y lo que quieren es explorar posibles soluciones. Este tipo de consultas se dan en la búsqueda por síntomas.

Este tipo de búsqueda es importante porque, aunque no se suelen dar a menudo, constituyen el **último recurso del usuario cuando no sabe qué**

solución buscar y, si falla, resultaría casi imposible encontrar productos relevantes en el e-commerce. Sobre todo si la nomenclatura de las categorías del sitio web no se basa en problemas o síntomas (como suele ocurrir).

De hecho, en muchos sitios, la búsqueda de un síntoma (“dolor de rodilla”), devuelve, en primer lugar, cualquier producto relacionado con la palabra clave “rodilla” y muestra una gran cantidad de productos irrelevantes relacionados con la búsqueda, lo que hace difícil para el usuario hacerse una idea general de las opciones pertinentes.



En Vitamin World, una consulta con “Dolor de garganta” devuelve resultados irrelevantes. Cuando esto ocurre, algunos usuarios optan por abandonar el sitio en lugar de refinar la búsqueda o navegar por la lista.

- **Búsquedas no relacionadas con productos:**

Certificado en Inteligencia Artificial

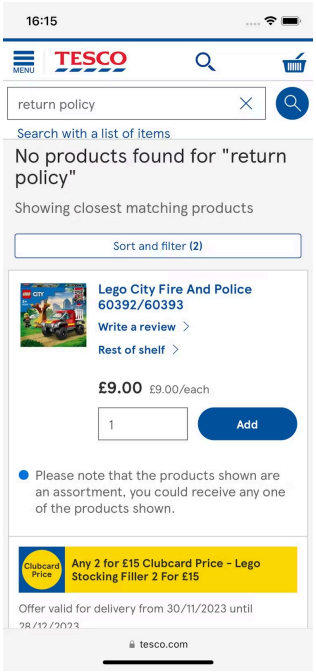
EOI & Samsung Innovation Campus

> El 50% de los sitios tienen problemas con búsquedas de información que no es un producto, como la política de devoluciones.

Tal y como su nombre indica, este tipo de consultas se producen cuando **los usuarios buscan algo que no es un producto**, como la política de devoluciones o la información de envío, y al no encontrarlo algunos usuarios abandonan el sitio ya que consideran que es información importante antes de hacer una compra.

Aunque la función principal de la búsqueda en un comercio electrónico es encontrar productos relevantes para el usuario, el motor de búsqueda no debería limitarse solo eso ya que, los usuarios esperan que la barra de búsqueda sirva para encontrar todo tipo de información sobre el sitio web.

Además, durante los test, los usuarios a menudo buscaban este tipo de contenido cuando no lo podían encontrar en donde esperaban (menús de navegación, footers, páginas de productos, etc.)



Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

> En Tesco, cuando un usuario busca por la política de devoluciones, el sitio web le devuelve una lista de resultados irrelevantes que, en su lugar..

Como hemos podido notar a lo largo de todo el análisis, a pesar de la importancia de poseer un buen algoritmo de búsqueda en un e-commerce , los **resultados** que ofrecen son, en un gran número de casos, **irrelevantes o escasos**, lo que provoca en el usuario **frustración** y **abandono**.
Esto se debe a una falta de soporte para los diferentes tipos de búsqueda.

El motor de búsqueda

El motor de búsqueda es una herramienta que ayuda a los usuarios de un e-commerce a encontrar productos a través de **palabras clave o *keywords***. Para ello, el buscador utiliza un algoritmo que combina técnicas de **emparejamiento exacto y ranking de palabras** entre la consulta que un usuario ingresa en la barra de búsqueda y las palabras clave almacenadas en las descripciones y metadatos de los productos de la base de datos del sitio web. El objetivo final es proporcionar resultados relevantes al usuario basados en lo que ha buscado.

Este tipo de búsqueda más tradicional, también conocida como **búsqueda léxica**, utiliza algoritmos tradicionales de Natural Language Processing o NLP que buscan, precisamente, lo que acabamos de comentar, **coincidencias exactas de palabras o frases** entre el catálogo de productos y la consulta del usuario.

A continuación, vamos a describir brevemente el funcionamiento interno de estos algoritmos para poder encontrar, más adelante, una solución más eficiente:

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Funcionamiento de los algoritmos de búsqueda léxica:

- **Indexación y catalogación de productos:** El primer paso es catalogar e indexar todos los productos disponibles en el sitio web, incluyendo sus nombres, descripciones, categorías y atributos de producto. Esta información se almacena en una base de datos para que el algoritmo pueda acceder a ella cuando un usuario realiza una búsqueda.
- **Coincidencia de palabras clave:** Cuando un usuario realiza una consulta, el algoritmo compara las palabras clave de la consulta con las palabras clave de la base de datos. Si encuentra una coincidencia, el producto se incluye en los resultados de búsqueda. Como ya hemos mencionado, los algoritmos como **BM25** y **TF-IDF** utilizan la frecuencia de los términos (TF) y la frecuencia inversa del documento (IDF) para calcular las puntuaciones de relevancia correspondientes a cada par documento-consulta. En concreto:
    - El TF-IDF calcula la importancia de cada palabra basada en dos factores:
        - TF: mide cuántas veces aparece una palabra en un documento.
        - IDF: mide cómo de única o rara es esa palabra en todos los documentos.
    - El BM25 es un algoritmo de puntuación que ajusta la relevancia de los resultados considerando factores como la longitud de los documentos. De esta manera, consigue evitar que textos muy largos dominen los resultados y adaptarse mejor a las expectativas de los usuarios.
- **Puntuación de relevancia:** A continuación, el algoritmo calcula una puntuación de relevancia para cada producto que coincide con la consulta. Esta puntuación indica el nivel de relevancia de un

producto en relación con la consulta inicial, y se utiliza para ordenar los resultados. Así, se muestran los productos más relevantes primero.

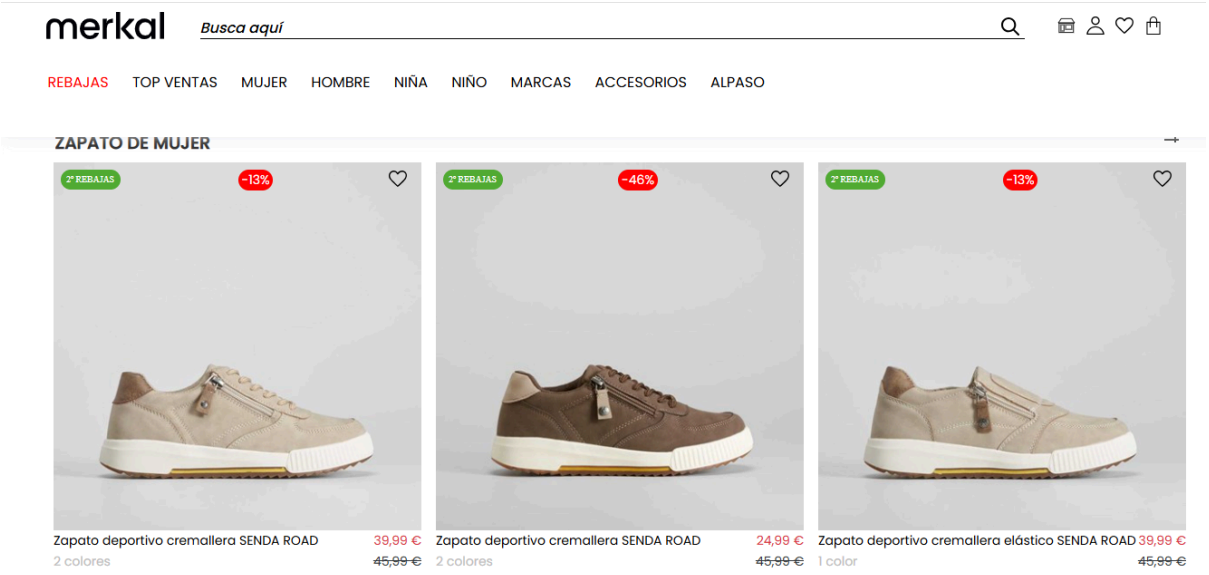
- **Presentación de resultados:** Como acabamos de comentar, una vez hecho el ranking de productos según su relevancia, se presentan al usuario en primer lugar los productos más relevantes para su consulta.

Este tipo de técnicas suelen proporcionar buenos resultados y ofrecen diferentes **ventajas** como:

- **Baja latencia:** Este tipo de algoritmos de búsqueda son **muy veloces**, lo que significa que manejan consultas y proporcionan resultados con tiempos de respuesta extremadamente breves. En la era de la inmediatez en la que nos encontramos, esto es vital ya que, como hemos observado previamente, los usuarios buscan resultados precisos e inmediatos a la hora de buscar productos.
- **Interpretabilidad de resultados:** Estos algoritmos basados en TF-IDF o BM25 son bastante transparentes. En otras palabras, resulta **sencillo comprender** por qué un producto lo ha considerado relevante para una consulta dado que los resultados se pueden explicar de manera matemática. A modo de resumen, podríamos simplificar el funcionamiento de estos algoritmos de la siguiente manera: “este producto contiene las palabras clave de la consulta en su descripción, por lo considero que que es relevante y te lo muestro en los primeros resultados”.
- **Buena capacidad de generalización:** Con esto nos referimos a que los algoritmos de búsqueda léxica **son bastante eficaces en una amplia variedad de categorías**, ya que solo requieren hallar palabras exactas. Dicho de una manera más sencilla, funcionan igual de bien tanto al buscar libros como ropa o dispositivos electrónicos.

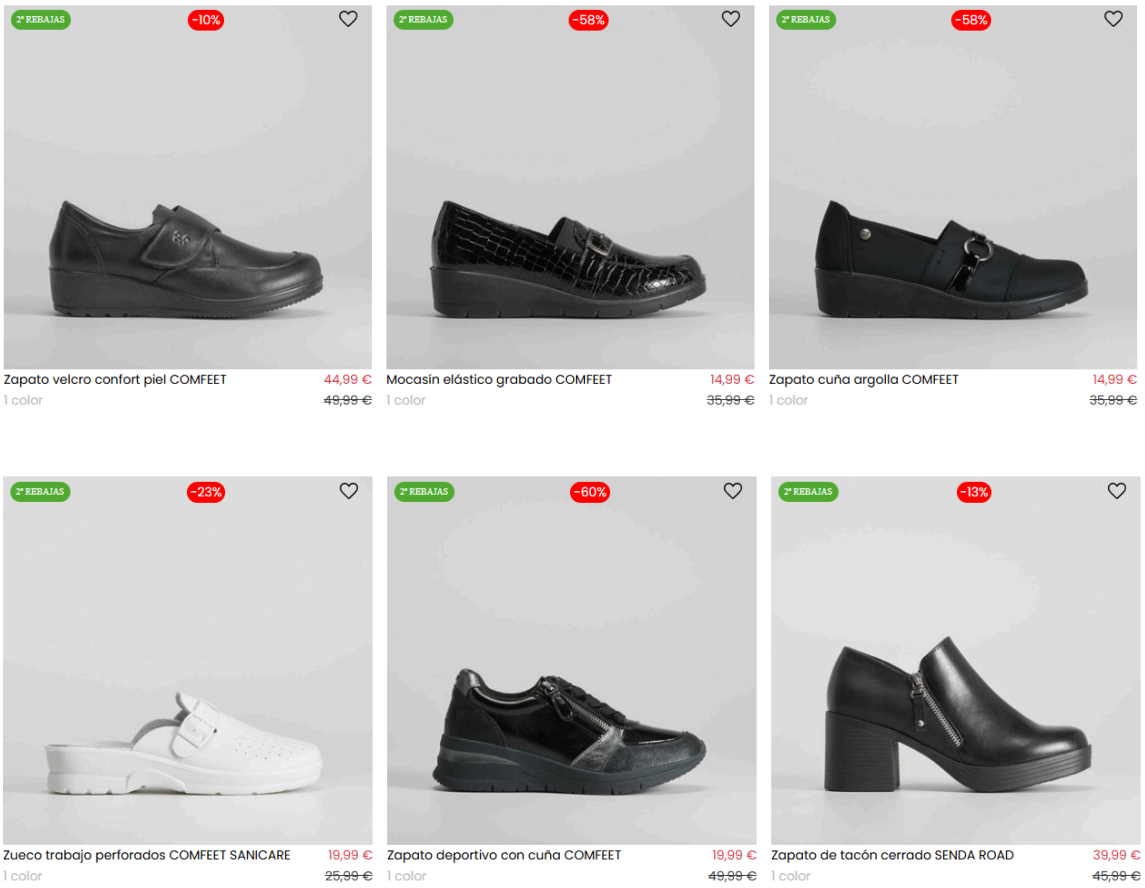
Sin embargo, esta búsqueda basada en *keywords* también tiene **limitaciones**. Como ya hemos visto previamente, este tipo de búsqueda se enfoca exclusivamente en la coincidencia de palabras clave, **sin tener en cuenta el contexto o el significado intrínseco** de las mismas. Esto podría dar lugar a **resultados irrelevantes**, especialmente si los usuarios utilizan sinónimos, frases, o consultas más complejas.

A modo de ejemplo, podríamos pensar en un usuario que introduce en el buscador del e-commerce Merkal el término **"zapato de mujer"**. Si nos basamos en el funcionamiento del algoritmo TF-IDF que hemos explicado anteriormente, el algoritmo buscaría en la base de datos las descripciones de productos, título o cualquier otro campo indexado que contengan las palabras exactas "zapato" y "mujer".



Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus



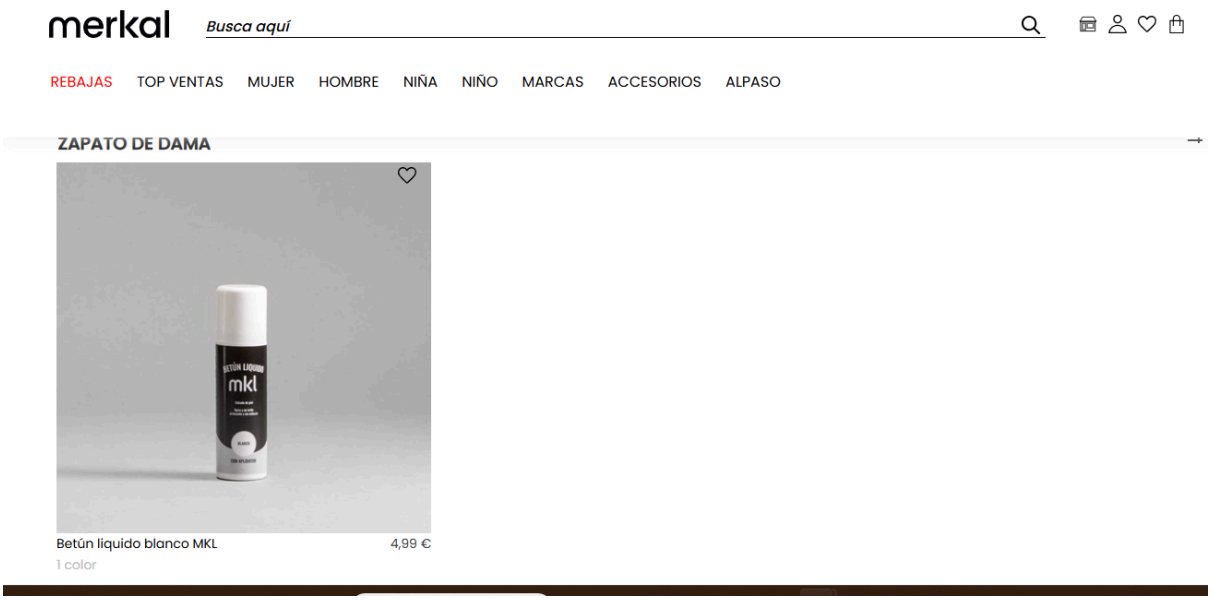
Como podemos ver en los resultados, el buscador nos devuelve toda una serie de zapatos de diferentes características:

- Zapato deportivo cremallera SENDA ROAD
- Zapato velcro confort piel COMFEET
- Mocasín elástico grabado COMFEET
- Zueco trabajo perforados COMFEET SANICARE
- Zapato de tacón cerrado SENDA ROAD

El algoritmo nos muestra resultados muy variados, aquellos que contienen en su descripción o título la palabra “zapato” o “mujer” (no nos enseña botas o botines, por ejemplo). Sin embargo, si cambiamos el término “mujer” de la búsqueda por un sinónimo, “dama”, el resultado que obtenemos es el siguiente:

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

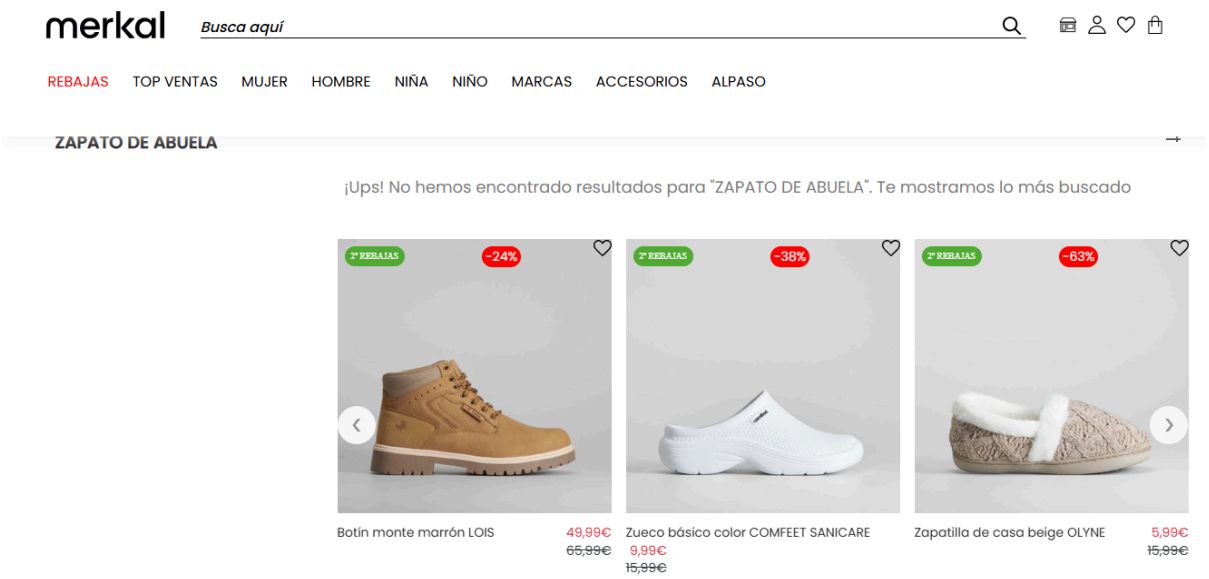


Aquí queda claro que el algoritmo de búsqueda no entiende sinónimos ni variaciones en las palabras. Es más, nos muestra un producto irrelevante, un complemento para el calzado.

Si realizamos una búsqueda más con “zapato de abuela”, vemos que el algoritmo nos lanza un mensaje de error. No ha encontrado una coincidencia exacta entre las palabras de la consulta y la base de datos, por lo que no ha entendido nuestra consulta. A cambio, nos ofrece los productos más buscados.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus



Llegados a este punto, nos podríamos plantear la siguiente pregunta: **¿Cómo podríamos lograr que el algoritmo entendiera cualquier tipo de consulta sin necesidad de incluir cada palabra clave con sus respectivas variaciones y sinónimos en la base de datos y así lograr una experiencia de búsqueda más eficiente y satisfactoria?**

Antes de resolver esta cuestión, vamos a entender, primero, cómo tienen que estar optimizados los títulos, descripciones y metadata de cada producto en una web de comercio electrónico para que el buscador interno pueda detectarlos correctamente y devolver resultados relevantes al usuario y, al mismo tiempo, el motor de búsqueda de Google pueda posicionar bien dicha web.

Para ello, vamos a analizar dos conceptos clave: Intent y Entity

Intent

El ***intent*** o **intención de búsqueda** se refiere al **propósito, intención, motivación u objetivo que tiene un usuario al realizar una consulta en un motor de búsqueda.** Esta intención representa las necesidades, deseos o problemas que el usuario busca satisfacer o resolver al ingresar la consulta.

Por ejemplo, en la consulta "Necesito unos botines de fútbol Adidas" la intención del usuario es ver una lista de botines de fútbol.

Para poder reconocer un *intent*, a nivel gramatical normalmente está formado por un verbo y un sustantivo, como en este caso: Necesito botines.

Podemos encontrar diferentes categorías de *intent* dependiendo de la etapa del funnel en la que se encuentre el usuario. Estas categorías son: **informativa**, **transaccional**, **comercial**, y **navegacional**.

- **Ejemplos:**
    - Un usuario con **intención informativa** busca aprender sobre un tema, como "cómo hacer una lasaña".
    - Un usuario con **intención transaccional** busca realizar una compra, como "comprar iphone 16".
    - Un usuario con **intención comercial** busca comparar opciones antes de comprar un producto, como "mejor smartphone 2024".
    - Un usuario con **intención navegacional** busca una página web específica, como "iniciar sesión Netflix".

A nivel de **SEO**, la intención de búsqueda resulta ser una pieza clave, puesto que afecta directamente la calidad de los resultados. Los motores de búsqueda actuales como Google han mejorado sus algoritmos para entender mejor esta intención detrás de las consultas. Para ello, analizan varios factores, como las **palabras clave** utilizadas y el **contexto de la búsqueda**. Por ello, **si el contenido de un sitio web se alinea con la intención de búsqueda del usuario, los motores de búsqueda lo recompensarán con una mejor posición en los resultados de búsqueda y una menor tasa de rebote**.

El análisis de la intención nos puede revelar **qué tipo de contenido es el más adecuado** para satisfacer la necesidad del usuario (por ejemplo, un blog, una página de producto, o una página de comparación). En el caso del e-commerce, la

página de producto será el elemento clave donde poner en práctica las mejores **técnicas de posicionamiento y optimización** para que no solo sea la web del comercio electrónico fácil de encontrar y posicionar en Google, sino que también, el propio **motor de búsqueda interno** pueda **mostrar con mayor precisión y relevancia los resultados de búsqueda del usuario mejorando así, su experiencia de compra**.

Pero antes de indagar en estas técnicas, tenemos que entender otro concepto importante en cualquier página de producto de un e-commerce: Entity

Entity

Una entidad o *entity* es una **pieza de información que proporciona contexto o detalles relevantes para la intención del usuario. Puede ser una persona, un lugar, un objeto, un evento, un concepto, o una expresión numérica**. Por ejemplo, si un usuario escribe: "Necesito unos botines de fútbol Adidas.", las entidades son “botines de fútbol” y “Adidas”. Otros ejemplos de entidades podrían ser:

- "Nueva York" como entidad que representa una ubicación.
- "zapatos de tacón alto" como entidad que representa un producto.
- "iPhone 16" como entidad que representa un producto específico.

La relación que podemos encontrar entre *intent* y *entity* radica en que la **intención** responde al *por qué*, mientras que la **entidad** es el *qué* de la búsqueda. Si nos fijamos en el último ejemplo, un usuario puede tener una intención específica, como por ejemplo comprar, relacionada con una entidad específica, un "iPhone 16".

Para satisfacer completamente la consulta de un usuario, los motores de búsqueda externos deben entender tanto la intención como las entidades involucradas. Como hemos mencionado anteriormente, **el *intent* nos permite**

saber qué tipo de contenido espera el usuario, mientras que las *entities* nos ayudan a crear el contenido correcto.

Una vez hemos entendido la diferencia entre *intent* y *entity* y cómo trabajan juntos en la consulta del usuario, vamos a explorar cómo debemos utilizarlas en las páginas de producto para, en primer lugar, lograr un buen posicionamiento web y, en segundo lugar, lograr que el motor de búsqueda interno del e-commerce entienda bien el catálogo de productos para devolver resultados relevantes y precisos al usuario.

Implementación del Intent y Entity a nivel de posicionamiento

Los componentes básicos de cualquier página de producto de un e-commerce son las **descripciones**, **títulos**, **etiquetas o *tags*** y **metadatos**. Por ello, es crucial que incluyamos tanto el ***intent*** como la ***entity*** en su contenido. Integrando estos dos conceptos, los motores de búsqueda pueden comprender el producto y mostrarlo a los usuarios correctos, **mejorando la visibilidad y el posicionamiento** del comercio, así como para **enriquecer la experiencia de usuario** al facilitar su proceso de compra.

Entities:

Las entidades son la base de la **descripción del producto** ya que permiten una identificación clara del mismo. Por ello, es importante que introduzcamos información específica y sin ambigüedades sobre lo que es el producto. Esto incluye su **nombre, tipo, características, marca, modelo, materiales, tamaño, color, especificaciones técnicas y cualquier otro atributo relevante**. Por ejemplo, en vez de poner solo "chaqueta", podemos escribir "chaqueta de cuero negra para hombre talla L" en la descripción.

Intent:

El uso del *intent* es más sutil pero igualmente importante. Por ello, debemos asegurarnos de enriquecer las descripciones mediante los diferentes tipos de intención de búsqueda que hemos mencionado anteriormente. Esto es:

- **Si la intención es informativa significa que el usuario está buscando información sobre el producto:** debemos incluir detalles que resuelvan sus dudas y respondan a preguntas comunes. Por ejemplo, si un usuario busca "cómo limpiar una chaqueta de cuero", incluye instrucciones de cuidado en la descripción.
- **Si la intención es transaccional significa que el usuario está listo para comprar:** introduciremos palabras clave que reflejen esa intención, como "comprar ahora", "oferta", "envío gratis", etc.
- **Si la intención es comercial significa que el usuario está comparando productos:** incluiremos información que resalte las ventajas y beneficios del producto, como reseñas, comparaciones y precios.
- **Si la intención es navegacional significa que el usuario ya conoce la marca del producto:** debemos usar el nombre de la marca en el título y la descripción para facilitar su búsqueda.

Para asegurarnos de que estamos escribiendo tanto los títulos como las descripciones, etiquetas o metadatos de los productos en un lenguaje que tanto los motores de búsqueda externos como los usuarios puedan entender, debemos **utilizar lenguaje natural.** En este sentido, es importante utilizar **palabras clave** en las descripciones pero también hay que tener en cuenta cómo buscaría el usuario el producto e introducir **frases que se parezcan a las preguntas de los usuarios**. Por ejemplo, en lugar de solo usar palabras clave como "zapatos deportivos", podemos usar frases como "zapatillas para correr cómodas" o "zapatos deportivos para gimnasio". De esta forma, **enriquecemos las keywords** para que se adapten a diferentes contextos, pudiendo llegar a utilizar las ***long tail keywords*** o palabras clave de cola larga. Estas son palabras clave más específicas que nos pueden ayudar a atraer a usuarios con una intención de compra más

clara. Un ejemplo podría ser, "vestido de verano rojo talla mediana con estampado floral" en lugar de solo "vestido rojo". No obstante, tampoco tenemos por qué limitarnos solo al uso de palabras clave exactas, podemos utilizar también sinónimos y términos relacionados.

Implementación:

- **Título del Producto:**
  - Tiene que ser conciso e incluir el ***entity* principal del producto** (nombre, tipo) y **palabras clave relevantes**.
  - Ejemplo: "Chaqueta de cuero negra para hombre - Talla L".
- **Descripción del Producto:**
  - Debe ser detallada e incluir todos los **atributos relevantes del producto** (*entities*) y **palabras clave de intención** que destaquen los beneficios y resuelvan las posibles dudas del usuario. Para ello, las frases de la descripción deben responder a ¿qué es?, ¿cómo se usa?, ¿para qué sirve? o ¿por qué debo comprarlo?
- **Etiquetas:**
  - Las etiquetas **son el equivalente a palabras clave**, pero más acotadas. Utilizar etiquetas relevantes nos ayuda a clasificar correctamente el producto pero en su nomenclatura, se tienen que utilizar palabras que los usuarios podrían usar en sus búsquedas.
  - Ejemplo: "chaqueta", "cuero", "hombre", "negra", "talla L", "otoño".
- **Metadatos:**
  - El título y descripción tienen que contener palabras clave que resalten el valor del producto y estén optimizadas para SEO.
  - Los metadatos, como la descripción meta, pueden influir en el **CTR** (Click Through Rate), por lo que deben ser descriptivos y atractivos.

Ejemplo real:

Producto: *Zapatillas de running para mujer marca X modelo Y.*

- **Título:** Zapatillas Running Mujer X Modelo Y - Máximo Confort
- **Descripción:** Las zapatillas de running para mujer marca X modelo Y están diseñadas para ofrecer el máximo confort y rendimiento en tus carreras diarias. Fabricadas con materiales transpirables y una suela con amortiguación superior, estas zapatillas proporcionan un ajuste perfecto y una excelente respuesta en cada paso. Ideales para corredoras de todos los niveles. ¡Compra ahora y experimenta la diferencia!
    - (Aquí se incluyen palabras clave de intención transaccional, como "compra ahora", y palabras clave de intención informativa como "máximo confort", "transpirables" y "amortiguación superior" así como la entidad del producto).
- **Etiquetas:** "zapatillas running", "zapatillas mujer", "zapatillas deportivas", "running", "confort", "marca X", "modelo Y".
- **Metadatos:**
    - Título: Zapatillas Running Mujer X Modelo Y - ¡Máximo Confort y Rendimiento!
    - Descripción: Descubre las zapatillas de running para mujer marca X modelo Y. Confort, amortiguación y ajuste perfecto para tus carreras. ¡Compra ahora y mejora tu rendimiento!

Implementación del Intent y Entity en el motor de búsqueda interno

Como ya hemos comentado, el motor de búsqueda interno puede utilizar algoritmos como **TF-IDF** o **BM25** para encontrar **palabras clave coincidentes entre la consulta del usuario y los productos de catálogo del e-commerce**. A continuación, veremos cómo optimizar el título, descripción, etiquetas y metadatos con *entities* e *intents* relevantes asegura recibir resultados más relevantes.

Si incluimos *entities* relevantes, facilitamos que el motor de búsqueda comprenda qué es el producto y lo relacione con las consultas de los usuarios. Tanto el TF-IDF como el BM25 utilizan la frecuencia de los términos y la frecuencia inversa del documento para calcular la relevancia. Por ello, **cuanto más se repitan las entidades relevantes en un documento** (en este caso, en la descripción del producto) **y menos comunes sean en el resto del catálogo, mayor será su puntuación de relevancia**.

Sin embargo, como ya hemos visto, el contenido de los elementos de la página de producto no solo deben contener *entities* específicas. También son importantes los *intents*. Estos utilizan el **lenguaje natural** en el que los usuarios se expresan normalmente para buscar productos en lugar de usar solo palabras clave aisladas. Por ello, **mejoran la capacidad del motor de búsqueda de ofrecer resultados relevantes**. Por ejemplo, si un usuario busca "comprar zapatillas de running baratas", la descripción del producto debería incluir palabras clave que reflejen esa intención transaccional, como "comprar", "baratas", "ofertas" o "envío gratis" . Al mismo tiempo, debería también incluir palabras clave de intención informativa como "comodidad" y "transpirables", incrementando así la relevancia para usuarios que están buscando reseñas o información antes de comprar.

Como ya hemos comentado brevemente, tanto TF-IDF como BM25 consideran la frecuencia con la que aparecen los términos (*entities* e *intents*) en el texto. **Al incluir de manera natural y estratégica las entidades e intenciones en el título, descripción y metadatos, se aumenta la frecuencia de los términos relevantes.**

Además, si tenemos en cuenta que el TF-IDF penaliza los términos que son muy comunes en todo el corpus de documentos (el BM25 lo hace de forma más sofisticada), **al utilizar palabras clave específicas y menos comunes, pero que describen con precisión las entidades del producto y reflejan la intención de compra del usuario, se mejora la puntuación de relevancia dando lugar a mejores resultados.**

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

Ejemplo real:
Si un usuario busca "organizar cocina", el motor de búsqueda interno debería mostrar organizadores de cocina, estantes, divisores y otros productos relacionados. Esto se puede lograr incluyendo estas entidades y una intención informativa/transaccional en las descripciones y títulos de dichos productos.

Para un producto como un "vestido de verano rojo talla mediana con estampado floral", las *entities* serían: vestido, verano, rojo, talla mediana, estampado floral. La intención puede ser transaccional (compra) o informativa (ver detalles y reseñas), por lo que se deben incluir palabras clave como "comprar vestido de verano rojo" y "vestido rojo con estampado floral opiniones".

Certificado en Inteligencia Artificial

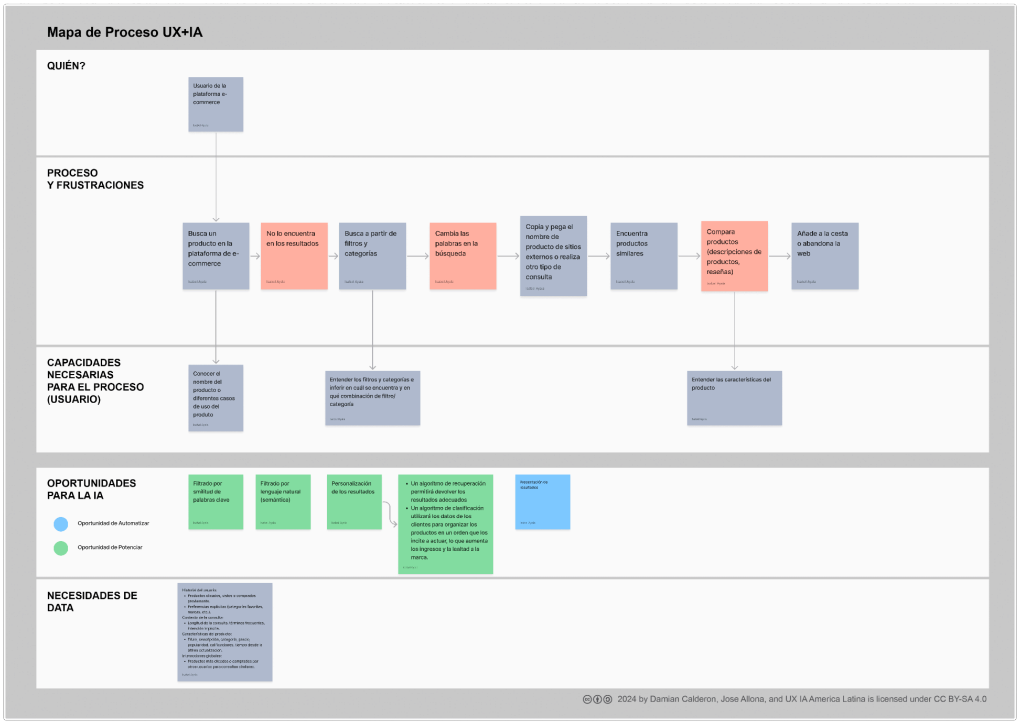
EOI & Samsung Innovation Campus

Diseñar la solución

Design Thinking

Mapa de proceso

En primer lugar, para entender mejor las posibles **frustraciones** o ***pain points*** del usuario, realizamos un **Mapa de Proceso** donde construimos el *journey* que tiene que realizar el usuario para cumplir su objetivo final: comprar el producto.
Como hemos visto anteriormente, los principales *pain points* tienen que ver con la **dificultad del algoritmo de proporcionar resultados relevantes ante diferentes tipos de consulta**, por lo que el usuario tiene que recurrir al sistema de navegación o a un refinamiento de su consulta para encontrar posibles productos candidatos hasta que compra el producto, o abandona el comercio.



Para poder diseñar una solución que se ajuste bien a esas particularidades, es necesario primero entender cuáles son las **capacidades actuales del usuario** para llevar a cabo su tarea, cuáles son las **oportunidades** que nos puede brindar la **IA** y cuáles son las **necesidades de data** que tenemos que tener en cuenta para crear una **solución verdaderamente efectiva y alineada con el usuario y la tecnología**.

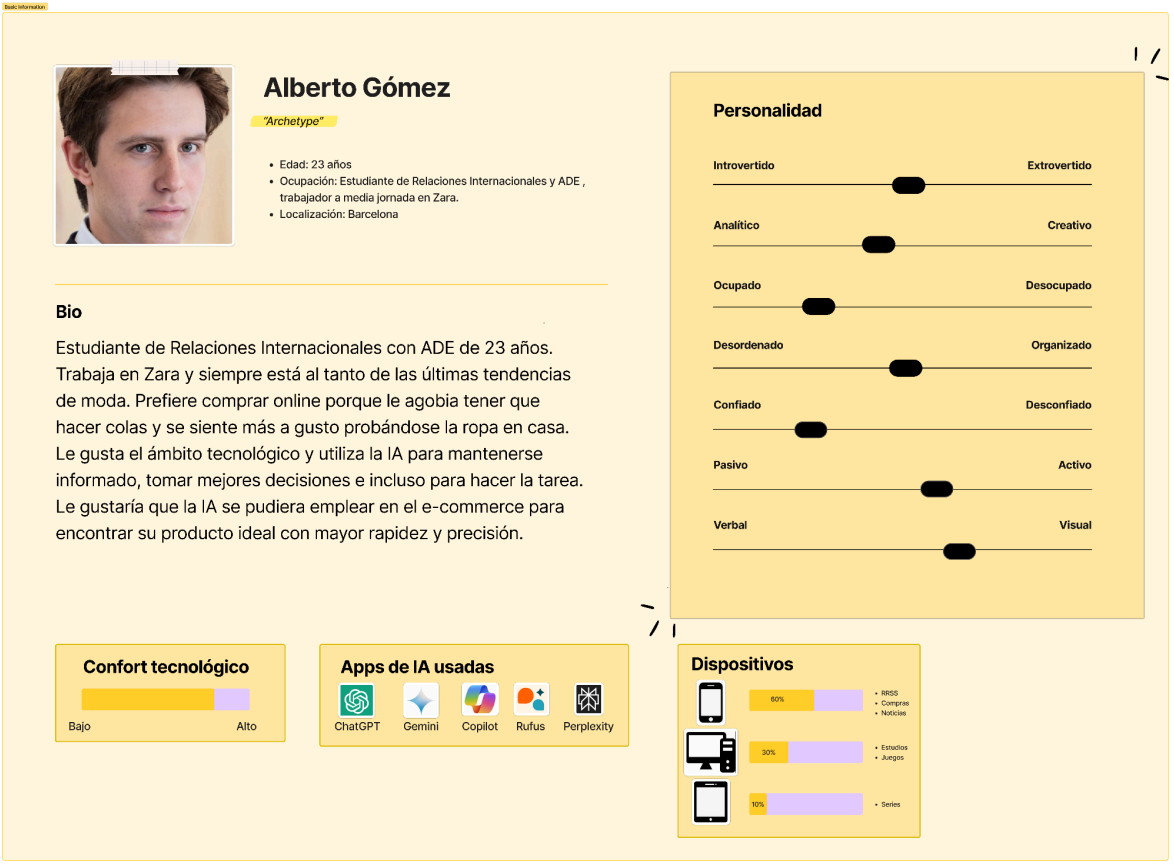
En este caso, las **capacidades del usuario** tienen que ver con el **entendimiento de la arquitectura de la web** y su habilidad para aplicar **filtros** cuando es necesario. En cuanto a la **IA**, podemos utilizarla para realizar un **filtrado léxico y semántico** y una **clasificación personalizada** de los productos más relevantes. Para ello, como **necesidades específicas de data** encontramos el **catálogo de productos** y diferentes métricas de **popularidad**, **interacción**, **longitud de la consulta** o **historial del usuario**.

User Persona

Para entender en más profundidad al perfil de usuario que utilizará nuestra solución, creamos un **User Persona**. Este está compuesto por varios apartados en los que definimos su **biografía** y **rasgos de personalidad** así como su capacidad de desempeño a nivel tecnológico, es decir, su nivel de **confort con la tecnología**, las **aplicaciones de IA** que más utiliza y los **dispositivos electrónicos** que maneja en su día a día. Conocer todos estos factores nos ayudará a **diseñar una solución y una interfaz alineada con sus necesidades y expectativas**.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus

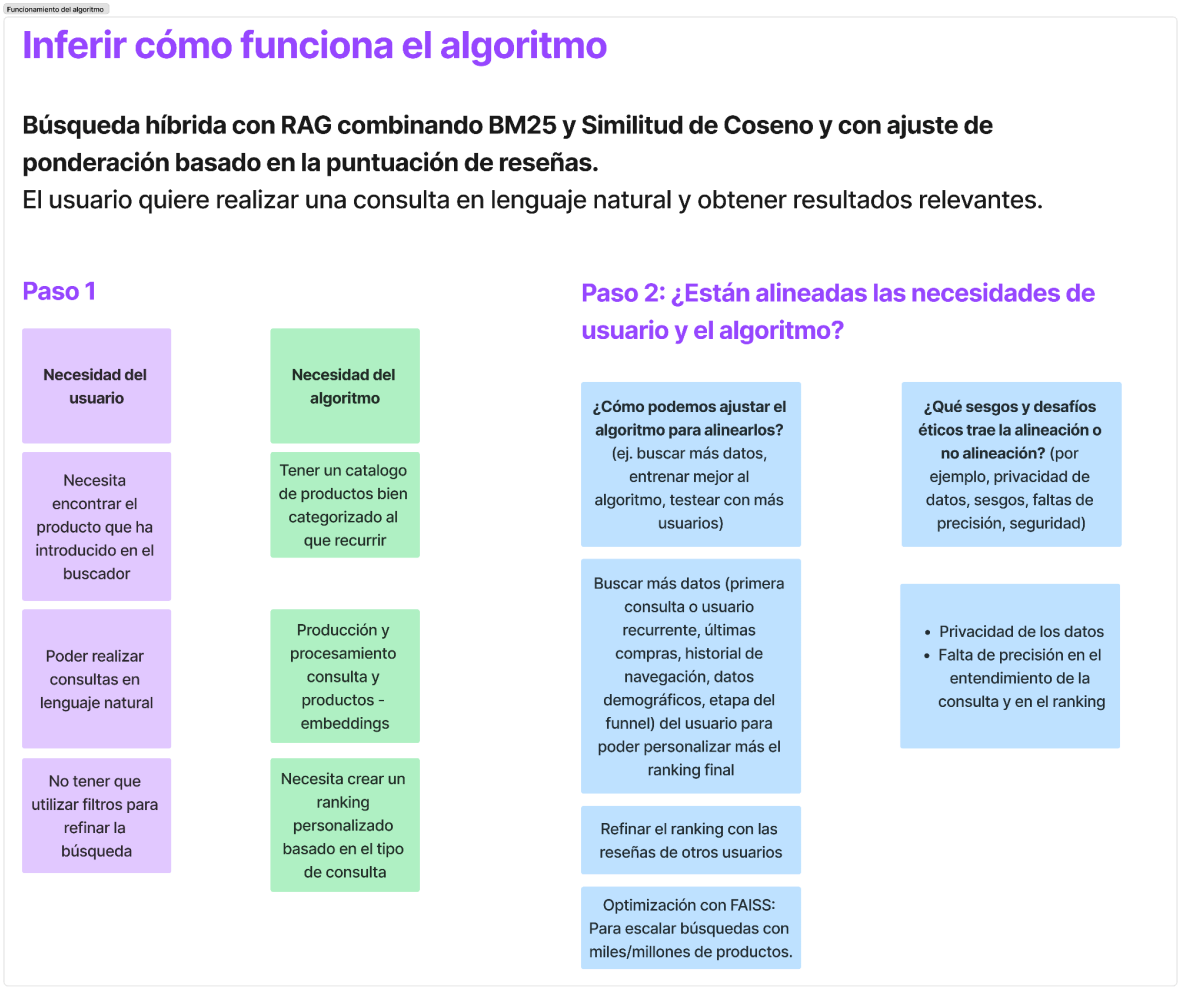


Inferir el algoritmo

Antes de diseñar el algoritmo, realizamos un ejercicio donde destacamos las **necesidades** tanto del **usuario** como del **algoritmo**. De esta forma, pudimos definir qué criterios debemos tener en cuenta para alinear a ambos y con qué **retos** nos podemos encontrar.

Certificado en Inteligencia Artificial

EOI & Samsung Innovation Campus



Definir la solución

RAG

Como hemos comentado en capítulos anteriores, los algoritmos de búsqueda y recomendación actuales, como **BM25**, son **modelos más estáticos**. Miden la coincidencia de diferentes términos en una serie de documentos para determinar la relevancia del mismo. No obstante, el e-commerce es un sector muy dinámico. Para realizar una buena recomendación, **el algoritmo se tiene que adaptar rápidamente a datos cambiantes en tiempo real**, como las interacciones del

usuario y el número de compras de un determinado producto, además de tener en cuenta **opiniones sesgadas** de productos que pueden haber afectado negativamente a la **puntuación de relevancia** del mismo.
En la era de la **personalización**, que un e-commerce sea capaz de manejar esta gran cantidad de datos dinámicos y ofrecer al usuario recomendaciones alineadas a su perfil, impactará directamente en la **experiencia de compra, el aumento de ventas y el posicionamiento en la mente del consumidor**.

De esta manera, nos planteamos la siguiente pregunta:
¿Cómo podríamos mejorar la capacidad de recuperación de información en tiempo real de tal forma que sea más eficiente, relevante y acorde al perfil del usuario?

El RAG se presenta como una solución potente a todos estos desafíos. Como describiremos a continuación, esta técnica nos permite realizar recomendaciones actualizadas y adaptadas a las preferencias de cada usuario.

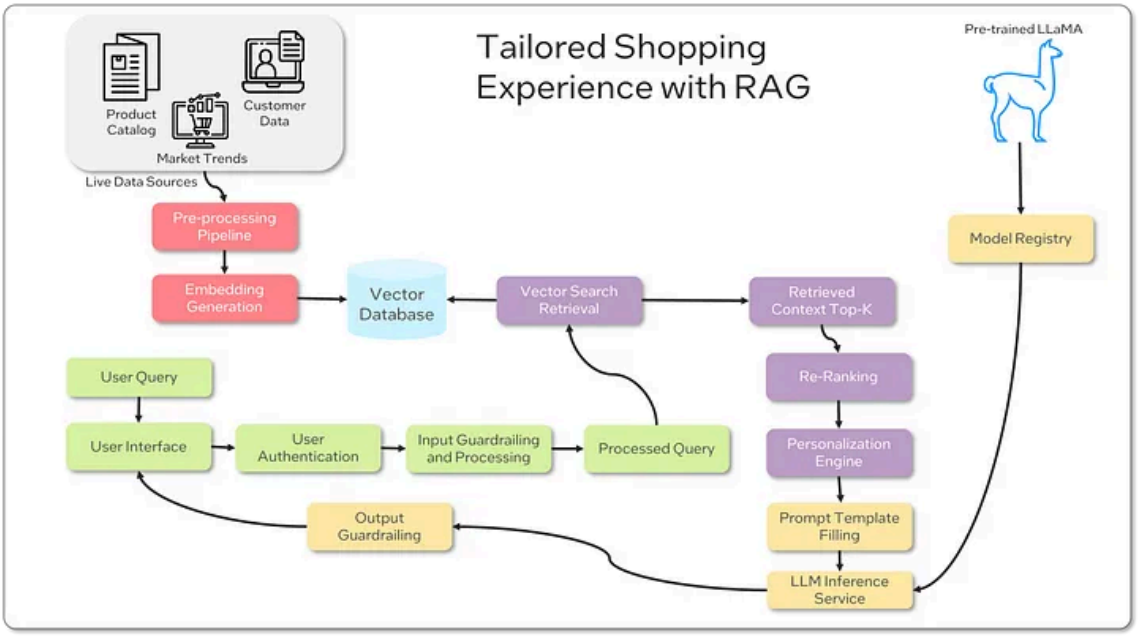
Un **RAG** (Retrieval-Augmented Generation) es un marco de optimización que combina tanto la **recuperación de información** como la **inteligencia artificial generativa**. Como su propio nombre indica, consta de dos partes fundamentales:

- **Recuperador:** Se encarga de buscar y extraer la información más relevante y similar a la consulta del usuario dentro de la base de datos.
- **Generador:** Recibe ese contenido filtrado como contexto para crear una respuesta coherente y conversacional, utilizando lenguaje natural.

Tal y como sus siglas indican, se añade un mecanismo de **recuperación** de datos para **aumentar** el LLM al que va unido, ayudándole, de esta manera, a **generar** respuestas precisas y relevantes para el usuario.

Una vez tenemos claro esto, podemos profundizar un poco más en su funcionamiento interno.

Arquitectura del RAG



Autor: Eduardo Alvarez

1. En primer lugar, se **recopilan datos** sobre las tendencias de mercado y las preferencias de los usuarios y comportamiento de compra en la web de diversas fuentes como el catálogo de productos, redes sociales y sistemas de monitorización de tráfico como los CRM.
2. A continuación, estos datos se transforman en **embeddings**, vectores densos multidimensionales que capturan la relación entre los datos, y se almacenan en una **base de datos vectorial**.
3. La capa RAG se encarga de recuperar la información pertinente de la base de datos vectorial a partir de una serie de algoritmos y técnicas que evalúan la **similitud de los embeddings** de los documentos y de la consulta del usuario. Es en este momento cuando el sistema puede realizar un **ranking** que garantice que las recomendaciones sean pertinentes para el mercado actual y personalizadas para cada usuario.

4. Con la información obtenida, el sistema genera dinámicamente una **respuesta con sugerencias de productos personalizados mediante un LLM**. Para modular el tipo de respuesta que queremos y ajustar a los manuales de voz de la marca, podemos instruir al LLM a que siga una serie de pautas. Esto lo conseguimos diseñando un **System Prompt** y unos **guardarraíles** que permiten definir los límites de la interacción entre el usuario y la compañía.

Como podemos inferir, esta arquitectura resulta muy potente para **mejorar la satisfacción y experiencia de compra del cliente** ya que podemos crear un sistema inteligente completamente personalizado al sector donde opere el e-commerce.
Podríamos pensar que otra solución válida podría ser utilizar un **LLM fundacional** y realizar un ***fine-tuning*** para que domine el ámbito de actuación específico del e-commerce.

Fine-tuning

Mientras el RAG conecta un LLM con datos dinámicos de la empresa a modo de contexto para aumentar el prompt engineering, el *fine-tuning* reentrena un modelo específico con un conjunto estático de datos etiquetados con la finalidad de mejorar su rendimiento ante tareas específicas del sector.

Usualmente, cuando proveemos de una serie de materiales e información específica a un LLM, este, al comprender mejor el dominio específico, sus características y su terminología, puede generar respuestas más precisas. Sin embargo, no podemos olvidar que estos **LLMs**, como GPT-3 y GPT-4, están **preentrenados con un conjunto fijo y masivo de datos** que periódicamente se actualiza. Mientras tanto, solo pueden recurrir a la última actualización de conocimiento que tienen. Por ello, al no tener un acceso continuo a nuevos datos no pueden mejorar su capacidad de aprendizaje y a menudo pueden generar

respuestas inexactas o alucinaciones, especialmente si le preguntamos por información muy específica, reciente o compleja. Este tipo de alucinaciones se podrían definir como un **fenómeno en el que los modelos de IA generativa no pueden encontrar fácilmente la información referente a la consulta del usuario en su conjunto de datos de entrenamiento**, por lo que, al utilizar algoritmos estadísticos de predicción, intentan "adivinar" estadísticamente la respuesta a la consulta, dando lugar a errores de precisión.

Todas estas limitaciones las podemos solventar con el **RAG**, ya que nos permite **acceder a una base de datos dinámica sin necesidad de preentrenar el LLM sobre un conocimiento estático**. De esta manera, puede generar respuestas más actualizadas y precisas. A continuación, vamos a explorar en más profundidad cómo la implementación del RAG puede mejorar la calidad del descubrimiento de productos de un e-commerce.

Impacto del RAG en el e-commerce

Una arquitectura RAG aplicada al sector del e-commerce permite a los comercios mejorar, de forma notable, la experiencia de búsqueda. Como ya sabemos, a modo de resumen, **el RAG recupera productos relevantes de la base de datos** (en este caso, el catálogo de productos) **y genera respuestas descriptivas incluso si las palabras clave de la consulta del usuario no coinciden exactamente con los títulos o descripciones de los productos**. Por ejemplo, si el usuario escribe en la barra de búsqueda "chaqueta ligera de senderismo para clima frío", el RAG puede recuperar productos relacionados con la consulta, resumir sus características y proporcionar una recomendación aunque las palabras "ligera" o "clima frío" no aparezcan en la descripción original.

Y no solo esto, al poder conectarse a una base de datos dinámica y actualizarse constantemente, puede **refinar las recomendaciones** de productos basadas en **tendencias del mercado**, **objetivos de negocio** y hasta puede aprender de

interacciones individuales de los usuarios como su **historial de compra**, su **comportamiento de navegación**, preferencias de **color** e incluso puede añadir comentarios sobre el **tamaño**, **materiales**, **compatibilidad** o **casos de uso reales** de los productos si estos se adecúan a la consulta. En este sentido, un RAG podría saber si un televisor funciona con Alexa, si una impresora es compatible con un tipo de cartucho de tinta, o si una prenda se puede lavar en la lavadora.

Si todo esto ya parece increíble, el RAG podría, incluso **comparar productos** y recuperar **reseñas de usuarios** para mejorar aún más sus recomendaciones y ayudar a los usuario a tomar mejores decisiones. Por ejemplo, podría resumir que un portátil está bien valorado por la duración de la batería y la calidad de la pantalla, aunque algunos usuarios pudieran notar que se recalienta demasiado al utilizar diferentes programas durante mucho tiempo.

Todas estas capacidades, sin duda proporcionan un mayor contexto al RAG al que los sistemas de recomendación tradicionales no pueden optar. Y esto, inevitablemente tiene **consecuencias positivas tanto en el negocio como en los usuarios**, lo que se traducirá en un **aumento del tráfico en la web** y **mayor tasa de conversión y satisfacción del usuario**.

Sistema de recomendación de productos basado en RAG combinado con BM25 y Embeddings

En el siguiente enlace explicamos, paso a paso, la construcción del modelo.

Código del modelo

Diseño de la UI

Para finalizar, realizamos un prototipo donde podemos ver cómo se implantaría nuestra solución en la interfaz de un e-commerce.
UI Design

- En primer lugar, en el buscador principal de la página instalamos un toggle que nos permite activar la búsqueda por IA. Esto lo realizamos para dar un control a los usuarios sobre qué tipo de búsqueda quieren realizar y, en el caso de que experimenten frustración con nuestro modelo, puedan cambiar a la búsqueda léxica habitual sin que interceda en las métricas de conversión del negocio.
- En segundo lugar, creamos la interfaz de chat donde se desarrollaría la búsqueda y conversación con el modelo.
- Por último integramos un sistema de valoración binaria para poder recoger métricas sobre el éxito de su interacción.

Criterios de evaluación del modelo

Para evaluar la experiencia global tanto a nivel de programación del modelo como de usuario, establecimos 3 tipos de criterios de evaluación de la respuesta:

- **Criterios generales**
    - User prompt: Consulta exacta, mixta o semántica
    - Respuesta:
        - ¿Proporciona productos relevantes?
        - ¿La recomendación está alineada a las características del producto?
        - ¿Responde de manera fluida y utiliza un tono natural y conversacional?
- **Consulta exacta**
    - User prompt: “Glidden paint”
    - Respuesta:
        - ¿Entiende adecuadamente mayúsculas/minúsculas?
        - ¿Ignora signos de puntuación o espacios extra?
- **Consulta semántica**
    - User prompt: “I'm remodeling my bathroom and need a stylish white sink with a faucet included. My maximum budget is less than $1,000”

    - Respuesta:
        - ¿La respuesta abarca todos los aspectos relevantes de la consulta?
        - ¿Mantiene coherencia y continuidad con el contexto de la consulta? (mismas entidades sin introducir información irrelevante)

Modelo de lenguaje local

Un modelo de lenguaje grande o **Large Language Model** por sus siglas en inglés (LLM) es un modelo de lenguaje entrenado sobre grandes cantidades de texto que permiten hacer una amplia variedad de tareas, desde análisis de sentimiento hasta razonamiento matemático. Anteriormente hemos hablado de los LLM como modelos disponibles en la nube, como ChatGPT, Gemini o DeepSeek. En esta parte desarrollamos lo que se conoce como **LLM local**. En contraste con los anteriores, los LLM locales son de gran utilidad dado que ofrecen **mayor seguridad que los modelos online**. Los datos de entrenamiento están almacenados en servidores locales, evitando filtraciones o ataques, por lo que estos modelos son de suma importancia para empresas donde los datos son confidenciales.

Debido a estas propiedades, los LLMs locales presentan las siguientes ventajas dentro del e-commerce:

- **Ofrecen un mayor control sobre los datos y privacidad**, dado que al ejecutarse localmente, no requieren enviar datos de clientes o productos a servidores externos, garantizando mayor **seguridad y confidencialidad**. Como hemos indicado anteriormente, se pueden entrenar o afinar con datos internos sin riesgo de filtraciones.

- Permiten una **búsqueda optimizada basada en recomendaciones** pues mediante el uso del entendimiento semántico, pueden mejorar la búsqueda léxica interpretando sinónimos, errores tipográficos o búsquedas contextuales. A la vez, Facilitan la personalización de resultados según el comportamiento del usuario.
- La propia tienda puede **automatizar restricciones** en búsquedas según políticas de la misma, como devolver productos dentro de ciertos plazos o según condiciones específicas. En otras palabras, pueden **adaptarse a reglas de negocio** específicas y responder preguntas sobre términos de servicio.
- Dado que el modelo esta **entrenado sobre la base de datos local,** presenta una **baja latencia** dado que no depende de servidores externos, mejorando la experiencia de usuario en grandes catálogos de productos.
- **Amplia personalización** ajustándose a necesidades específicas sin depender de modelos de OpenAI o Google. Por ejemplo, se pueden integrar **embeddings de productos** para mejorar recomendaciones y agrupaciones de búsqueda.
- Finalmente, los modelos locales **evitan costos de API en la nube** y pueden ejecutarse en infraestructura propia, optimizando gastos en procesamiento de lenguaje natural. Igualmente, pueden escalar en función del hardware disponible, ajustando su rendimiento a las necesidades del negocio.

Arquitectura de un LLM local + RAG

Hay diferentes métodos para trabajar con el modelo de lenguaje local. Existe modelos preentrenados en grandes cantidades de texto que permiten realizar las búsquedas sin necesidad de entrenarlos sobre la base de datos. En este caso, se debe realizar un **fine-tuning** del modelo, pasando por una técnica de **Low-Ranking Adaptation** (LoRA), que permite adaptar partes del modelo sin

alterar la estructura interna del modelo. De esta forma el modelo preentrenado se adapta a la base de datos del e-commerce.

El flujo de trabajo de un **LLM local** en el **buscador de una página web de e-commerce** puede dividirse en varias etapas, combinando procesamiento de texto, recuperación de información y generación de respuestas. A continuación aparece el flujo típico de trabajo de un RAG+LLM:

1. En primer lugar, se lanza una pregunta al modelo (entrada de usuario), que puede incluir especificaciones del producto, así como otras cuestiones: propiedades del envío, devolución, opiniones, etc.
2. Esta consulta es procesada y transformada en una representación entendible por el modelo. Este proceso comienza con la tokenización de la consulta, dividiéndola en palabras clave.
3. A continuación, se procede a la búsqueda de productos en la base de datos. Esta búsqueda combina dos enfoques diferentes:
    a. Por un lado, se realiza una búsqueda consultando la base de datos y filtrando a través de filtros clásicos.
    b. Por otro, el RAG realiza una recuperación semántica de los resultados más relevantes utilizando embeddings (vectores numéricos).
4. Finalmente, el LLM local elabora la respuesta basada en los productos encontrados y según los filtros que el usuario haya aplicado (respuesta contextualizada).

Implementación del modelo

Los modelos de lenguaje locales son modelos preentrenados y exportados a un archivo que contiene los parámetros del modelo optimizados. De esta forma, al cargar el archivo de parámetros, se pueden lanzar consultas al modelo si necesidad de conexión a internet. Estos archivos son bastante grandes, y su uso está limitado a la potencia de la máquina. En nuestro caso no hemos podido

combinar el RAG con el LLM local debido a la falta de memoria. A continuación describimos el proceso que hemos seguido.

Existen numerosos modelos locales ya entrenados que podemos encontrar en internet. Hemos escogido la página Hugging Face, donde encontramos modelos como Llama, Mistral o Gemma. Nos enfocamos en la biblioteca `transformers` de Hugging Face, que permite una implementación flexible y optimizada para diferentes hardware.

Antes de instalar el modelo, es necesario configurar el entorno y asegurarse de que Python y las librerías necesarias están instaladas: `torch`, `transformers`, `sentencepiece`. Una vez instaladas las dependencias, cargamos entonces un modelo de lenguaje y generamos el tokenizador. En este caso, usaremos **Mistral 7B**, un modelo eficiente para generación de texto:

```
# Nombre del modelo en Hugging Face
model_name = "mistralai/Mistral-7B-v0.1"
# Cargar el tokenizador
tokenizer = AutoTokenizer.from_pretrained(model_name)
```

```
# Cargar el modelo en la CPU
device='cpu'
model = AutoModelForCausalLM.from_pretrained(model_name,
torch_dtype=torch.float16, device_map="auto").to(device)
prompt = <ejemplo de prompt>
```

Una vez compuesto el prompt del sistema a través del RAG, se tokeniza con el modelo de lenguaje:

```
# Tokenizar la entrada
inputs = tokenizer(prompt , return_tensors="pt").to(device)
# Generar una respuesta
outputs = model.generate(**inputs, max_new_tokens=100)
# Decodificar y mostrar la salida
print(tokenizer.decode(outputs[0], skip_special_tokens=True))
```

Conclusiones

En este trabajo hemos querido ofrecer una visión del estado del e-commerce actual y cómo implementar una posible solución, identificando ciertos problemas que surgen a nivel de consumidor.

El **e-commerce** es uno de los muchos mercados que ya se benefician de la inclusión de la **inteligencia artificial** (IA) en sus engranajes. Esto facilita la optimización de procesos clave como la **búsqueda de productos**, la personalización de recomendaciones y la gestión de inventarios. La IA se ha convertido en un elemento fundamental para la transformación de la experiencia de compra, haciendo que las plataformas sean más intuitivas, eficientes y personalizables. Desde la mejora de **motores de búsqueda** hasta la generación de **recomendaciones** precisas, la IA desempeña un papel central en la mejora de la interacción entre los usuarios y los negocios.

Sin embargo, a pesar de estos avances, todavía existen desafíos que afectan tanto la experiencia del usuario como a la **eficiencia** de la compraventa. Uno de los principales problemas lo encontramos en la **precisión y relevancia** de los resultados de búsqueda. Los motores de búsqueda tradicionales suelen basarse en **coincidencias de palabras clave**, lo que puede llevar a resultados irrelevantes o poco útiles. La falta de comprensión del **contexto e intención** del usuario dificultan la conversión de visitas en compras efectivas. Además, un sistema de búsqueda ineficiente puede hacer que los usuarios **abandonen** la plataforma antes de encontrar el producto que necesitan. Por ello, la integración de algoritmos avanzados en los motores de búsqueda permite optimizar la relevancia de los resultados y, en consecuencia, **mejorar la satisfacción del usuario**.

Para enfrentar estos desafíos, hemos identificado los **pain points** más comunes en la experiencia del usuario y hemos diseñado un algoritmo que optimiza la

búsqueda de productos, actuando como un asistente para mejorar la experiencia de compra. Nuestra solución combina un **enfoque híbrido** de **recuperación aumentada de generación** con **embeddings semánticos** para realizar búsquedas más precisas, encontrando artículos que realmente se alineen con las necesidades del usuario.

Además, hemos integrado un **modelo de lenguaje natural** (LLM) que interpreta las consultas del usuario y elabora respuestas detalladas, proporcionando recomendaciones personalizadas y explicaciones contextualizadas. Al principio, consideramos la posibilidad de implementar un LLM local para mejorar la privacidad y la eficiencia del sistema. Sin embargo, los modelos LLM requieren de un rendimiento computacional elevado, lo que nos llevó a adoptar una solución basada en GPT-4. Este cambio nos permitió asegurar respuestas más precisas y adecuadas al contexto, ofreciendo una experiencia de usuario significativamente mejorada.

Con esta implementación, logramos un sistema de búsqueda más avanzado, intuitivo y efectivo, reduciendo la fricción en el proceso de compra y facilitando una experiencia más satisfactoria para los usuarios del e-commerce.