

RECOMENDACIONES TURÍSTICAS DE LA IA GENERATIVA: VARIACIONES LINGÜÍSTICAS Y CULTURALES EN LAS REPRESENTACIONES DE DESTINOS TURÍSTICOS GENERATIVE AI TOURISM RECOMMENDATIONS: LINGUISTIC AND CULTURAL VARIATIONS IN THE REPRESENTATIONS OF TOURIST DESTINATIONS

Volha Herasimovich*, Aitziber Pousa-Unanue, Aurkene Alzua-Sorzabal***, Basagaitz Guereño-Omil**** y Ruipu Liang*******

*Universidad Antonio de Nebrija | vherasimovich@nebrija.es | <https://orcid.org/0000-0003-0858-1009>

**Universidad Antonio de Nebrija | apousa@nebrija.es | <https://orcid.org/0000-0001-6596-6293>

***Universidad Antonio de Nebrija, Universidad de Deusto | malzua@nebrija.es, aurkenealzua@deusto.es | <https://orcid.org/0000-0002-6058-9537>

****Universidad de Deusto | b.guereno@deusto.es | <https://orcid.org/0000-0002-6264-6435>

*****Universidad Antonio de Nebrija | lruipu@alumnos.nebrija.es | <https://orcid.org/0009-0009-6813-9977>

ENTREGADO: 30/06/2025 ACEPTADO: 03/11/2025

CC BY-NC-SA 4.0: <https://creativecommons.org/licenses/by-nc-sa/4.0/> 

Resumen: Este estudio analiza las variaciones lingüísticas y culturales en las recomendaciones turísticas generadas por modelos de lenguaje de gran tamaño (LLM), con foco en el comportamiento de ChatGPT ante consultas formuladas en español, euskera, inglés y ruso. Mediante un diseño experimental de métodos mixtos, se evaluaron 210 respuestas generadas a partir de *prompts* estructurados sobre tres destinos del País Vasco: Zumaia, Zarautz y Donostia / San Sebastián. Los resultados evidencian diferencias significativas en la precisión, la exhaustividad y la estructura de las recomendaciones, dependiendo del idioma de consulta y de la forma toponímica. Se observaron mayores tasas de errores y de respuestas genéricas en lenguas extranjeras, especialmente en ruso, mientras que las lenguas locales ofrecieron recomendaciones más detalladas, aunque con cierto riesgo de inexactitud factual. El reconocimiento de destinos se mostró sensible a la escritura, siendo menor en sistemas no latinos. Los resultados reflejan una disparidad lingüística y cultural en la generación automatizada de contenido turístico, probablemente vinculada a sesgos en los datos de entrenamiento. Se concluye que los sistemas de IA generativa deben adaptarse al contexto multilingüe del turismo mediante estrategias colaborativas que integren a desarrolladores, gestores turísticos y expertos culturales, para asegurar representaciones digitales precisas, inclusivas y culturalmente contextualizadas.

Palabras clave: inteligencia artificial generativa, modelos de lenguaje de gran tamaño, destinos turísticos, recomendaciones turísticas, multilingüismo, toponimia

Abstract: This study analyses linguistic and cultural variations in tourism recommendations generated by large language models (LLM), focusing on the behaviour of ChatGPT when responding to prompts in Spanish, Basque, English, and Russian. Using a mixed-methods experimental design, 210 AI-generated responses were evaluated based on structured prompts related to three destinations in the Basque Country: Zumaia, Zarautz, and Donostia / San Sebastián. The results reveal significant differences in the accuracy, comprehensiveness, and structure of recommendations depending on the language of the query and the toponymic form used. Higher rates of errors and generic content were observed in foreign languages, particularly Russian, while local languages yielded more detailed recommendations, though with occasional factual inaccuracies. Destination recognition proved sensitive to the writing system, with reduced performance for non-Latin scripts. These findings reflect a linguistic and cultural disparity in the automated generation of tourism content, likely linked to biases in training data. The study concludes that generative AI systems must be adapted to the multilingual context of tourism through collaborative strategies involving developers, tourism managers, and cultural experts to ensure digital representations that are accurate, inclusive, and culturally contextualised.

Keywords: generative artificial intelligence, large language models, tourism destinations, tourism recommendations, multilingualism, toponymy

1. INTRODUCCIÓN

La representación de los destinos turísticos ya no está determinada exclusivamente por residentes, visitantes o narrativas culturales e históricas, sino que se encuentra crecientemente mediada por tecnologías de inteligencia artificial (Lehto et al., 2025). Estas tecnologías desempeñan un papel central en modelar la información y las experiencias de los visitantes. Dado el impacto creciente de la IA en el comportamiento del consumidor y en la toma de decisiones, resulta fundamental comprender en qué medida la inteligencia artificial genera valor y mejora potencialmente la experiencia del usuario (Grewal et al., 2017; Grundner & Neuhofer, 2021).

En este contexto, los modelos generativos de lenguaje de gran tamaño (del inglés, Large Language Models o LLM), como ChatGPT, Gemini y DeepSeek, han emergido como herramientas eficaces para personalizar experiencias turísticas mediante procesamiento avanzado del lenguaje natural (Doğan & Niyet, 2024; Ren et al., 2025). Entre sus capacidades destacan la generación de respuestas adaptadas, la comprensión contextual y la flexibilidad en entornos no estructurados. Estas características permiten a los turistas acceder a recomendaciones en lenguaje natural sobre destinos, atracciones y experiencias locales, tanto en la fase de planificación como durante y después del viaje (Gursoy et al., 2023). Por ejemplo, durante las fases de inspiración y planificación, el 68 % consultan hasta diez fuentes para informarse, mientras que un 32 % explora más de diez sitios web (Robinson et al., 2024). Las herramientas de IA generativa optimizan estos procesos al reducir la sobrecarga informativa y facilitar recomendaciones personalizadas. La calidad de las recomendaciones es crítica, no solo para mejorar la experiencia del usuario, sino también para preservar la reputación del destino y fomentar la adopción sostenida de estas herramientas.

Pese a su avance, los modelos de lenguaje de gran tamaño (LLM) siguen presentando sesgos lingüísticos y culturales que afectan su neutralidad, especialmente en contextos culturalmente sensibles como el turismo (Fleisig et al., 2024; Z. Liu, 2023; Tao et al., 2024). El rendimiento de los sistemas de recomendación depende directamente de la calidad y representatividad de los datos relativos al ocio y la cultura (Stoica, 2020). La escasez de información digital sobre destinos pequeños o menos conocidos, o en lenguas regionales o minoritarias, puede resultar en recomendaciones incompletas o inexactas, con omisiones relevantes de puntos de interés (PDI).

Los desequilibrios en la existencia de datos relativos al ocio, la cultura y el turismo también pueden inducir sesgos culturales, reflejando visiones del mundo de culturas sobrerrepresentadas. Esto se traduce en recomendaciones que replican estereotipos asociados a raza, género, edad o discapacidad (UN Tourism & Saxion University of Applied Sciences, 2025). Además, el contenido generado por IA puede reflejar estereotipos vinculados a la identidad lingüística, asociando inconscientemente a los hablantes de determinadas lenguas con rasgos, comportamientos o valores específicos. Estos sesgos pueden hacer que usuarios con características distintas (como idioma, género o situación familiar) reciban recomendaciones divergentes de PDI.

Varios gobiernos impulsan iniciativas para reducir la dependencia de modelos internacionales y fomentar el uso de datos locales. En este contexto, el Gobierno de España ha desarrollado ALIA, un modelo de lenguaje entrenado en español, catalán, gallego, valenciano y euskera (ALIA, 2025). No obstante, su orientación hacia usuarios institucionales y profesionales limita su alcance frente a modelos de libre acceso como ChatGPT o DeepSeek, ampliamente utilizados también por turistas internacionales en España. Comprender cómo se construyen y modifican las representaciones digitales de los destinos constituye una prioridad tanto para la investigación como para las políticas públicas (Lehto et al., 2025). Evaluar el impacto de las recomendaciones generadas por IA sobre distintos grupos —viajeros y residentes— y analizar si refuerzan sesgos o desigualdades en accesibilidad, visibilidad o representación cultural es una línea clave en la investigación turística (Lehto et al., 2025). Las directrices éticas en IA subrayan la necesidad de evaluar los sesgos a lo largo del ciclo de vida del modelo, incluyendo los contenidos generados, para prevenir resultados discriminatorios (UN Tourism & Saxion University of Applied Sciences, 2025).

La IA generativa puede facilitar el descubrimiento de PDI relevantes, entendidos como localizaciones de valor cultural, natural, histórico o recreativo (Gursoy et al., 2023; Rico et al., 2019), y mejorar la experiencia del visitante. Sin embargo, los modelos de lenguaje presentan limitaciones importantes, como falta de exhaustividad, exceso de generalidad o desactualización de datos sobre atracciones (Volchek & Ivanov, 2024).

Frente a estos desafíos, iniciativas como GuideGeek proponen soluciones mediante la indexación directa del contenido de las organizaciones de gestión de destinos para integrarlo en las respuestas generadas por IA (GuideGeek, 2025). No obstante, la priorización algorítmica de «experiencias óptimas» puede invisibilizar alternativas significativas, pero menos destacadas, afectando la satisfacción del usuario (Grundner & Neuhofer, 2021).

Este estudio examina las variaciones lingüísticas y culturales en las recomendaciones de visitas turísticas generadas por modelos de lenguaje de propósito general. Se analizan salidas de ChatGPT en cuatro lenguas: español (lengua principal del país), euskera (lengua cooficial local), inglés (lengua franca global) y ruso (lengua con un sistema de escritura distinto), centradas en destinos de Gipuzkoa, caracterizados por la coexistencia de dos lenguas oficiales. El estudio evalúa si las características cuantitativas y cualitativas de las recomendaciones turísticas varían de manera significativa en función de la lengua utilizada.

Los resultados contribuyen al desarrollo de un marco conceptual sobre la interacción entre lengua, cultura y tecnología en el contexto de la IA turística. Se proponen métricas para cuantificar sesgos lingüísticos y culturales, orientando adaptaciones multilingües futuras. Asimismo, se ofrecen recomendaciones prácticas para desarrolladores, responsables de políticas públicas, gestores turísticos y usuarios finales, con el fin de mejorar la fiabilidad, equidad y aplicabilidad de estas tecnologías en la representación y promoción de destinos turísticos.

2. REVISIÓN DE LA LITERATURA

2.1. Disparidades lingüísticas y culturales en la tecnología del lenguaje

Los modelos de lenguaje de gran tamaño (LLM) se utilizan cada vez más en contextos culturalmente sensibles, como la planificación turística y la recomendación de destinos. Estas aplicaciones exigen no solo competencia lingüística, sino también comprensión contextual de la diversidad cultural. Sin embargo, los enfoques de entrenamiento predominantes en procesamiento de lenguaje natural (NLP) ponderan el rendimiento medio sobre grandes corpus, usualmente en inglés y centrados en marcos culturales occidentales. Esto genera lo que se ha denominado disparidad de representación: un rendimiento sistemáticamente inferior en lenguas, grupos étnicos y categorías demográficas subrepresentadas (Blodgett & O'Connor, 2017; Hershcovich et al., 2022; Jørgensen & Søgaard, 2021; Zhang et al., 2021).

Incluso en modelos multilingües, la falta de conciencia cultural en el tratamiento de la lengua perpetúa sesgos. Hershcovich (2022) advierte que etiquetas como «español» o «árabe» suelen reflejar narrativas nacionales dominantes, marginando subgrupos culturales. Sin estrategias de muestreo dirigidas ni metadatos precisos, los LLM tienden a reforzar normas hegemónicas y a excluir variaciones culturales locales.

Fleisig et al. (2024) muestran que las respuestas de ChatGPT en inglés estadounidense estándar son más extensas, precisas y culturalmente informadas que aquellas en dialectos no estándar, como el inglés nigeriano o jamaicano. Esta disparidad se debe a la escasa representación en los datos de entrenamiento y al tratamiento desigual de las variantes lingüísticas. Patrón similar se observa en otras lenguas, donde las respuestas se tornan genéricas o inapropiadas culturalmente. Recientes estudios han pasado de evaluaciones centradas en lenguas a métricas más detalladas de *benchmarking* cultural. Fung et al. (2024) desarrollan un atlas cultural, CultureAtlas, un conjunto de datos con más de 10 000 normas culturales y desviaciones verificadas, abarcando niveles subnacionales y más de 2500 grupos etnolingüísticos. A diferencia de recursos centrados en culturas nacionales, CultureAtlas permite evaluar capacidades culturales en escalas relevantes para el turismo, donde el conocimiento local es clave.

CultureAtlas demuestra que los modelos de lenguaje de última generación, como GPT-4, LLaMA-2 y Vicuna, presentan dificultades considerables para realizar razonamientos culturales detallados, especialmente en contextos con recursos limitados (Fung et al., 2024). Además, los modelos muestran menor cobertura y precisión en culturas con baja disponibilidad de datos, como Bután, Laos o Congo, y peor rendimiento en dominios culturalmente específicos (por ejemplo, gastronomía o vestimenta) respecto a temas más generales como educación o celebraciones (Fung et al., 2024).

Asimismo, el rendimiento disminuye significativamente (más del 50 % en F-score) cuando las normas culturales incluyen factores como edad, religión, estado civil o identidad regional, revelando la limitada capacidad de los LLM para generalizar con variables socioculturales complejas (Fung et al., 2024; Hershcovich et al., 2022). Liu et al. (2024),

corroboran estos hallazgos mostrando que los LLM tienden a memorizar proverbios sin comprender su significado cultural, evidenciando fallos sistemáticos en su razonamiento sobre expresiones figuradas que varían entre las diferentes lenguas.

De manera similar, Rigin et al. (2025) alertan sobre la adopción acrítica de estándares anglocéntricos en dominios como el turismo, lo cual reproduce jerarquías epistémicas y silencia perspectivas locales. Describen este fenómeno como una forma de «esclavitud intelectual» en contextos académicos poscoloniales. Por otro lado, Li et al. (2024) sostienen que el uso de *prompts* (consultas) no inducen comportamiento culturalmente adecuado si los datos de entrenamiento carecen de diversidad. Su modelo CultureLLM muestra que un *fine-tuning* leve con ejemplos específicos mejora el rendimiento en culturas infrarrepresentadas. No obstante, persisten sesgos estructurales, especialmente en lenguas distintas del inglés.

En conjunto, la evidencia indica que la capacidad multilingüe no implica comprensión multicultural (Fung et al., 2024; Z. Liu, 2023), y que ni los *prompts* ni el *fine-tuning* aislados son suficientes sin evaluación rigurosa ni estrategias de recolección y alineación cultural (Li et al., 2024; Masoud et al., 2024). Estas limitaciones son especialmente relevantes para aplicaciones turísticas, donde una descripción generada por IA que ignora normas locales o tergiversa rituales puede alienar usuarios y propagar desinformación.

Fung et al. (2024) enfatizan que el desarrollo de sistemas culturalmente sensibles requiere esfuerzos sistemáticos en recolección de datos, evaluación y normalización, especialmente fuera del ámbito anglófono dominante. Esto exige avanzar hacia un desarrollo colaborativo de tecnologías lingüísticas que involucre no solo a especialistas en IA, sino también a antropólogos, lingüistas y expertos locales. Iniciativas como CultureAtlas representan un avance significativo, pero el progreso sostenido depende de integrar la diversidad cultural y la equidad epistémica en el núcleo del diseño de modelos.

2.2. Interacción orientada a objetivos y personalización según perfiles turísticos

Los *prompts* de los turistas —definidos como instrucciones o solicitudes que se introducen en sistemas de IA generativa para obtener respuestas específicas— pueden entenderse como acciones intencionales dirigidas a objetivos concretos dentro del ciclo del viaje, tales como planificar itinerarios, realizar reservas u obtener información y detalles específicos sobre PDI en los destinos turísticos.

Desde esta perspectiva, Wörndl & Herzog (2022) agrupan las funciones clave de tecnologías aplicadas al turismo, incluidas aplicaciones móviles, en cuatro categorías principales: (1) funciones de búsqueda y acceso a información, (2) reserva de productos y servicios, (3) asistencia en destino y (4) funciones de interacción social. Aunque estas funciones fueron inicialmente asociadas a dispositivos móviles, también explican las motivaciones que impulsan a los turistas a interactuar con agentes conversacionales como ChatGPT. Los *prompts* pueden ser informativos (búsqueda de datos, actualizaciones, confirmaciones), comparativos (evaluación de opciones) o transaccionales (reservas, modificaciones), así como solicitudes de recomendaciones o asistencia contextual.

La personalización de la interacción con sistemas de IA generativa se ve influida por factores como el perfil sociodemográfico y, especialmente, por las motivaciones turísticas, entendidas como los impulsos psicológicos o biológicos que orientan las decisiones de viaje (Cohen et al., 2014; Yoon & Uysal, 2005). Estas motivaciones determinan las preferencias informativas de los usuarios y, por tanto, el tipo de *prompts* que formulan durante la planificación o estancia en destino.

La propuesta de Plog (1974), distingue entre perfiles alocéntricos —turistas orientados a la exploración de entornos poco convencionales— y psicocéntricos o más inclinados a destinos familiares y previsibles. Ambos perfiles recurren a fuentes digitales adaptadas a sus preferencias, lo cual incluye, crecientemente, el uso de herramientas como ChatGPT para la planificación del viaje.

De manera complementaria, el modelo *push-pull* (Crompton, 1979; Dann, 1977, 1981) distinguen entre motivaciones internas (*push*), como el deseo de escapar o experimentar algo nuevo, y factores externos (*pull*), como atributos tangibles o simbólicos del destino. En el alcance de este marco, los PDI, según la clasificación de UN Tourism (2019), pueden estructurarse en seis categorías constituyendo componentes clave en las decisiones del visitante: (1) atracciones culturales o patrimoniales, (2) sitios naturales, (3) instalaciones recreativas, (4) eventos y festivales, (5) espacios gastronómicos y comerciales, y (6) servicios complementarios.

Sin embargo, la percepción de atractivo varía en función de factores internos. Por ejemplo, para algunos viajeros el valor cultural del lugar es determinante, mientras que otros priorizan la accesibilidad o la idoneidad del destino para familias. Estas diferencias afectan directamente la naturaleza de las interacciones con asistentes de IA generativa.

Así, las motivaciones *push* y *pull* permiten comprender las distintas demandas de información formuladas por los turistas. Al planificar un viaje, los visitantes evalúan múltiples elementos del destino (PDI, infraestructuras, servicios, etc.) para tomar decisiones estratégicas, desde la elección del destino hasta aspectos logísticos como el alojamiento (Crompton, 1979; Dann, 1977, 1981).

Este estudio se enmarca en la creciente necesidad de comprender cómo responden los modelos de lenguaje ante las consultas turísticas formuladas por usuarios con distintos perfiles lingüísticos y motivacionales. La investigación se articula en torno a tres preguntas principales:

- PI1 ¿Cómo varían la veracidad y la corrección de las respuestas generadas por IA ante consultas sobre destinos turísticos formuladas en diferentes lenguas (principal, minorizadas, regionales, internacionales y con recursos limitados)?
- PI2 ¿En qué medida la lengua de la consulta afecta la extensión, el detalle descriptivo y la especificidad de las recomendaciones turísticas generadas por IA?
- PI3 ¿Existen diferencias en la priorización de las recomendaciones turísticas según la lengua utilizada en la consulta?

Estas preguntas permiten evaluar el grado de precisión, exhaustividad y relevancia contextual de las respuestas generadas por IA, así como identificar posibles patrones de sesgo lingüístico o cultural en la producción automatizada de contenido turístico.

En paralelo, el avance hacia ecosistemas de turismo inteligente donde la mediación algorítmica constituye una parte creciente de la experiencia del visitante plantea desafíos de gobernanza en contextos multilingües. La armonización de topónimos en distintas lenguas trasciende lo técnico y se configura como un asunto estratégico para la gestión de destinos. Surgen así interrogantes fundamentales para la formulación de políticas: ¿Deben las aplicaciones, los mapas digitales y la señalización incluir todas las variantes oficiales de los nombres de lugar? ¿Qué forma nominal debe emplearse cuando la lengua de consulta del usuario no coincide con las lenguas oficiales del destino? ¿Cómo pueden los organismos turísticos asegurar que los sistemas de IA reflejen sus políticas lingüísticas, sus objetivos de inclusión y sus estrategias de marca?

3. METODOLOGÍA

3.1. Diseño de la investigación

Esta investigación adopta un diseño experimental de métodos mixtos para analizar cómo responde un modelo de IA generativa a *prompts* turísticos formulados en múltiples lenguas y contextos lingüísticos diversos. El objetivo principal es identificar variaciones lingüísticas y culturales en el contenido, la precisión y la estructura de las recomendaciones turísticas generadas por IA. Mediante la integración de técnicas analíticas cuantitativas y cualitativas, el estudio permite un examen sistemático de las salidas del modelo según las distintas lenguas y variantes toponímicas de los destinos turísticos.

El marco experimental se estructuró en torno a un protocolo de *prompting* controlado, mediante el cual un conjunto fijo de *prompts* fue traducido sistemáticamente a cuatro lenguas: inglés, español, euskera y ruso. La selección de lenguas responde a una lógica de representación de diferentes contextos: la lengua principal del país (español), la lengua cooficial local (euskera), una lengua franca global (inglés) y una lengua que emplea un sistema de escritura diferente (ruso). Cada *prompt* emuló consultas turísticas reales, abarcando necesidades informativas generales y específicas (por ejemplo, lugares imprescindibles, experiencias culturales, accesibilidad), lo que permite una evaluación exhaustiva de los tipos de respuesta generados. Para aumentar la validez ecológica de los hallazgos —referida como el grado en que estos contextos se asemejan a las condiciones reales en las que ocurren las acciones simuladas (Schmuckler, 2010)— los *prompts* fueron redactados de manera breve, idiomática y con estructuras comparables entre lenguas, simulando interacciones auténticas de usuarios en búsqueda de información turística mediante agentes conversacionales (Tabla 1).

Al variar tanto la lengua utilizada como la forma toponímica (por ejemplo, oficial vs. cooficial), el estudio introduce un conjunto estructurado de condiciones que permiten evaluar el rendimiento del modelo en el reconocimiento de destinos y su capacidad

para generar recomendaciones relevantes. Este enfoque es especialmente relevante en casos en los que el nombre del destino se presenta en una forma no estándar, ya sea por variaciones ortográficas, transliteración o el uso de un sistema de escritura distinto, lo cual puede no coincidir con los patrones prevalentes en los datos de entrenamiento del modelo.

Tabla 1. *Prompts* utilizados para solicitar recomendaciones turísticas a modelos de IA en cuatro lenguas

| VARIACIÓN DEL PROMPT | INGLÉS | ESPAÑOL | RUSO | EUSKERA |
|-------------------------------------|----------------------------------|--|--|--|
| Visita general | Places to see in N? | ¿Lugares que ver en N? | Что посмотреть в N? | N-en ikusteko lekuak? |
| Atracciones imprescindibles | Must-see places in N? | ¿Lugares imprescindibles en N? | Что обязательно посмотреть в N? | Ikusi beharreko lekuak N-en? |
| Lugares no turísticos | Non-touristy places in N? | ¿Lugares poco turísticos en N? | Нетуристические места в N? | Turistentzat eza-gunak ez diren lekuak N-en? |
| Experiencias culturales locales | Local cultural experiences in N? | ¿Experiencias culturales locales en N? | Как познакомиться с местной культурой в N? | Tokiko kultura bizitzeko esperientziak N-en? |
| Actividades para familias | Things to do with kids in N? | ¿Qué hacer con niños en N? | Чем заняться с детьми в N? | Zer egin hurrekin N-n? |
| Accesibilidad para sillas de ruedas | Wheelchair access in N? | ¿Acceso para sillas de ruedas en N? | Есть ли доступ для инвалидов кресел в N? | Gurpil-aulkientzako sarbidea N-n? |

FUENTE: ELABORACIÓN PROPIA.

En este estudio se han explorado diversas variantes toponímicas a partir de los siguientes criterios:

- Topónimo oficial vs. no oficial: se ha comparado el uso de topónimos oficiales frente a aquellos no oficiales que siguen apareciendo en ciertos documentos y contextos. Por ejemplo, Zarautz es el topónimo oficial en lengua cooficial local, mientras que Zarauz representa la forma histórica en lengua principal del país.
- Flexión gramatical del topónimo: se refiere a la adaptación morfológica de un topónimo a las reglas de una lengua, como ocurre en lenguas con declinación, por ejemplo, el ruso y el euskera. Se ha analizado cómo el uso de la forma base frente a la forma flexionada afecta al reconocimiento por sistemas de IA, que pueden tener dificultades con topónimos flexionados. También se ha evaluado si añadir el país entre paréntesis (por ejemplo, España) mejora la identificación del destino turístico.
- Cambio de sistema de escritura: esta estrategia consiste en emplear un topónimo en una lengua distinta a la del texto principal, utilizando además un

sistema gráfico ajeno. Se ha analizado el uso de un topónimo en sus formas originales (en euskera y en español) dentro de un texto en ruso, empleando el alfabeto latino en lugar del cirílico. El objetivo es evaluar si esta técnica mejora el reconocimiento del destino turístico por parte de la IA.

El diseño contempla comparaciones tanto interlingüísticas como intralingüísticas. Las comparaciones entre lenguas permiten evaluar cómo las diferencias lingüísticas influyen en la cantidad y calidad del contenido generado por IA, mientras que las comparaciones dentro de una misma lengua —particularmente en variantes toponímicas— permiten examinar cómo la sensibilidad a las convenciones de nombrado afecta al reconocimiento del destino y la relevancia de las recomendaciones.

Se seleccionaron tres destinos turísticos del norte de España: Zumaia, Zarautz y Donostia/San Sebastián. Todos están ubicados en Gipuzkoa, una provincia con cooficialidad lingüística (español y euskera) y elevada actividad turística. En primer lugar, estos destinos fueron elegidos por contar con variantes toponímicas multilingües que incluyen tanto denominaciones oficiales como históricas. Zumaia y Zarautz adoptan las formas oficiales en euskera, mientras que las variantes en español (Zumaya y Zarauz) se emplean de manera informal o por parte de visitantes extranjeros. Donostia / San Sebastián representa un caso más complejo, combinando los dos nombres, lo cual genera retos en gestión del destino, como inconsistencias en señalización, plataformas digitales y materiales promocionales. Además, esta dualidad puede dificultar la identificación del destino por parte de los turistas y generar ambigüedad en términos de geolocalización y percepción de marca. En segundo lugar, se tuvo en cuenta la diferencia en la escala urbana: Zumaia, con 10 251 habitantes, representa el núcleo más pequeño; Zarautz, con 23 370 habitantes, se sitúa en una escala intermedia; y Donostia / San Sebastián, con 189 093 habitantes, constituye el entorno urbano de mayor dimensión (INE, 2024).

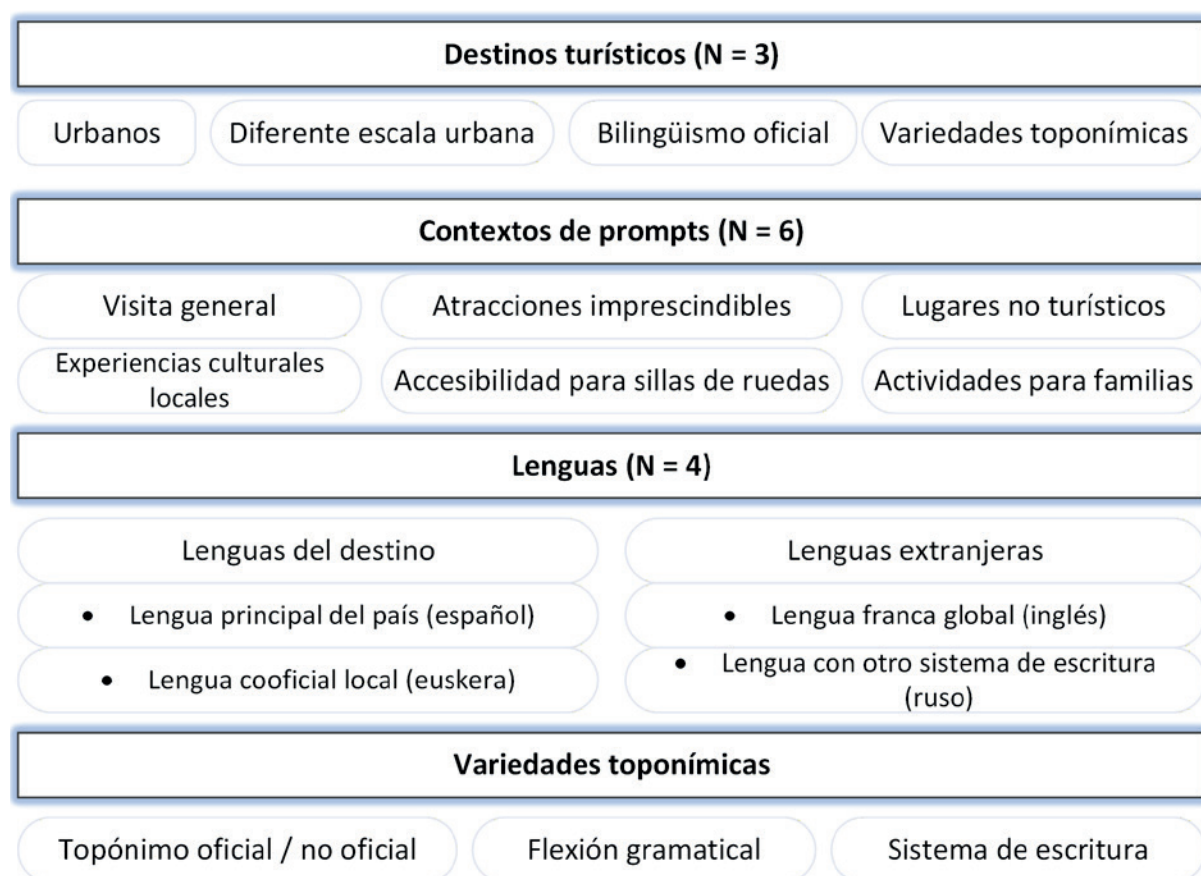
3.2. Fuentes de datos y procedimientos de muestreo

El corpus analizado se compone de 210 textos generados por el modelo GPT-4o de OpenAI. Esta tecnología fue elegida por su relevancia práctica, ya que en ese momento constituía el sistema predeterminado en ChatGPT, ampliamente utilizado por viajeros que empleaban IA generativa en la planificación de sus viajes (Kim et al., 2023; Shin et al., 2025).

Las respuestas fueron generadas mediante la API de OpenAI, lo que permitió mantener condiciones experimentales totalmente controladas. Esta metodología evita influencias derivadas de personalización basada en sesiones previas o historial del usuario. El parámetro probabilístico denominado «temperatura», que regula el grado de aleatoriedad en las respuestas del modelo, se estableció en 0 para garantizar una salida determinista. Esto asegura que, ante una misma entrada y las mismas condiciones, el modelo genere la misma respuesta. Este enfoque permite replicar resultados, identificar sesgos sistemáticos y realizar comparaciones estadísticas entre lenguas sin interferencias aleatorias. Todos los *prompts* fueron enviados el mismo día, garantizando consistencia temporal. Las salidas generadas se almacenaron en formato .docx con nombres de archivo correspondientes a cada *prompt*.

La muestra incluye 210 textos distribuidos por las diferentes lenguas: inglés ($n = 42$), español ($n = 42$), euskera ($n = 18$) y ruso ($n = 108$). La distribución no uniforme de los *prompts* fue intencionada para captar variaciones específicas asociadas a las formas toponímicas. Se empleó una muestra mayor en ruso para analizar las potenciales diferencias en el reconocimiento de distintas formas de topónimos en base a los criterios presentados anteriormente. Para algunas comparaciones, se seleccionó un subconjunto equilibrado; en otros casos, se mantuvieron subconjuntos desiguales, lo cual constituye una limitación mitigada mediante análisis dentro de cada grupo lingüístico. Las decisiones clave del diseño de la investigación para la recolección de datos se presentan en la Figura 1.

Figura 1. El diseño de la investigación para la recolección de datos



FUENTE: ELABORACIÓN PROPIA.

3.3. Análisis de datos

El contenido generado por IA fue codificado manualmente por dos investigadores utilizando el software Atlas.ti 25. Las variables analizadas incluyeron reconocimiento del destino, tipos de recomendaciones turísticas, clasificación de puntos de interés (PDI) y errores de contenido (Tabla 2). Adicionalmente, se emplearon scripts en Python para contabilizar el número de palabras y de elementos sugeridos, generalmente identificados por numeración o viñetas, formatos comunes en las respuestas de GPT-4o.

Tabla 2. Variables utilizadas en el análisis de los resultados generados por la IA

| DIMENSIÓN | VARIABLE | DEFINICIÓN |
|--|-------------------------------|--|
| Reconocimiento del destino | No reconocido | La IA no identifica correctamente el destino. La respuesta sólo contiene consejos generales o menciona puntos de interés (PDI) que no se encuentran en el destino real. No se proporciona ningún PDI real e identificable. |
| | Reconocido | La IA identifica correctamente el destino e incluye al menos un PDI verificable del mundo real situado en él. |
| Recomendaciones turísticas | Puntos de interés (PDI) | Una recomendación, entendida como una frase o conjunto de frases, que hace referencia a un lugar específico con una ubicación identificable (como un edificio, un distrito o una ruta), o a un acontecimiento cuya hora y ubicación pueden identificarse de forma fiable a partir de otras fuentes (por ejemplo, un festival con nombre propio). |
| | Recomendaciones generales | Recomendación, expresada en una frase o grupo de frases, que hace referencia a categorías generales de actividades o atracciones (por ejemplo, «visite museos», «disfrute de los bares locales») sin identificar ningún lugar o acontecimiento concreto. |
| | PDI inexistentes | Recomendación que hace referencia a puntos de interés que en realidad no existen en el destino. |
| | PDI fuera del destino | Recomendación de un punto de interés real situado fuera del destino previsto. Puede reconocerse explícitamente (por ejemplo, indicando la distancia) o presentarse sin aclaraciones. |
| Tipos de PDI y recomendaciones generales | Cultura y patrimonio | Incluye lugares de importancia cultural, histórica o arquitectónica como museos, monumentos, edificios patrimoniales, centros culturales y lugares de culto. |
| | Naturaleza y recreo | Abarca paisajes naturales y espacios para el ocio o las actividades al aire libre, incluidos parques, reservas, miradores, playas, puertos deportivos y parques de aventura. |
| | Eventos | Se refiere a actividades de duración determinada, como festivales y eventos, que pueden tener lugar en sitios específicos o en zonas más amplias de un destino. |
| | Gastronomía y compras | Abarca los lugares que ofrecen comida, bebida y experiencias de venta al por menor, incluidos restaurantes, bares, mercados y tiendas. |
| | Turismo creativo y vida local | Implica actividades participativas relacionadas con la cultura local, incluidos talleres, cursos y experiencias de arte, artesanía, gastronomía, deportes, lengua y vida cotidiana. |
| | Otros | Incluye infraestructuras y servicios turísticos de apoyo, como oficinas de información turística, medios de transporte y alojamientos. |

| DIMENSIÓN | VARIABLE | DEFINICIÓN |
|----------------------|-----------------------------------|---|
| Errores de contenido | Información obsoleta | Una recomendación incluye hechos que antes eran correctos, pero ya no son válidos (por ejemplo, se describe un sitio como visitable pero ahora está cerrado). |
| | Repetición | El mismo PDI aparece más de una vez en la salida con nombres ligeramente diferentes. |
| | Clasificación errónea | El PDI está mal clasificado en cuanto a su categoría o función (por ejemplo, un palacio descrito como un castillo). |
| | Atribución errónea | Se asignan al PDI características o funciones que no tiene, aunque sean típicas de PDI similares en otros lugares (por ejemplo, afirmar que un puerto deportivo local alberga barcos de pesca industrial cuando no es así). |
| | Nombre incorrecto | El PDI existe, pero se hace referencia a él con un nombre incorrecto. |
| | Descripción errónea de los hechos | Los datos concretos sobre el PDI son incorrectos (por ejemplo, se describe un periodo histórico equivocado). |

FUENTE: ELABORACIÓN PROPIA.

Las recomendaciones se categorizaron según una tipología desarrollada para este estudio, basada en el marco teórico de Lew (1987), que distingue tres enfoques en la conceptualización de atracciones turísticas: ideográfico (recursos naturales, culturales o recreativos), organizacional (escala espacial, capacidad, patrones temporales) y cognitivo (percepciones sobre autenticidad, seguridad y riesgo). Este marco permite integrar tanto atributos específicos como cualidades experienciales asociadas a los PDI.

Además, se adoptaron seis categorías amplias de recomendaciones según directrices internacionales (United Nations, Department of Economic and Social Affairs, 2010; World Tourism Organization, 2019): atracciones culturales y patrimoniales, sitios naturales, instalaciones recreativas, eventos y festivales, espacios gastronómicos y comerciales, y otros servicios. Estas categorías se enriquecieron con el concepto de turismo creativo (Richards, 2019), incluyendo actividades participativas vinculadas con la vida local.

3.4 Organización del análisis empírico

El análisis se estructuró en torno a cuatro escenarios analíticos, diseñados para evaluar el comportamiento del modelo de IA generativa en relación con distintos factores lingüísticos, toponímicos y estructurales:

Escenario 1. Reconocimiento de destinos según la lengua y la variante toponímica.

Este escenario evaluó la capacidad del modelo para identificar destinos turísticos en función de la lengua de entrada y la forma del topónimo. Se analizaron 210 textos en español, inglés, euskera y ruso. Para comprobar el reconocimiento en sistemas de escritura distintos, se usaron varias variantes del topónimo: con flexión gramatical; con

flexión gramatical y el país entre paréntesis; en su forma base; y en alfabeto latino dentro de *prompts* en ruso, incorporando así el topónimo utilizado en las lenguas del destino. El análisis permitió identificar las formas con mejor tasa de reconocimiento, especialmente útiles para los Escenarios 2 y 3.

Escenario 2. Sensibilidad toponímica en las respuestas generadas por IA. Se evaluó la influencia de las variantes toponímicas (oficiales e históricas) en las respuestas del modelo. Se compararon textos en español, inglés y ruso. El análisis incluyó dos subconjuntos: (1) los topónimos con dos variantes (Zumaia / Zumaya y Zarautz / Zarauz), con 36 textos por tipo de topónimo, oficial o histórico (24 por lengua); y (2) el topónimo con tres variantes: Donostia / San Sebastián, Donostia, San Sebastián, con 18 textos por cada tipo de topónimo, uno oficial y dos históricos (18 por lengua). En el caso del ruso, para facilitar el reconocimiento por parte del modelo, los topónimos se introdujeron en el alfabeto latino y las preguntas en cirílico.

Escenario 3. Análisis de tipos de recomendaciones turísticas. Este escenario abordó el contenido y la precisión de las recomendaciones generadas por IA. Las recomendaciones turísticas se clasificaron en cuatro categorías: puntos de interés (PDI), recomendaciones generales, PDI inexistentes y PDI fuera del destino. Se definieron los tipos de PDI y de recomendaciones generales según los criterios establecidos en la Tabla 2, donde también se detallan los errores de contenido identificados en las recomendaciones. La muestra incluyó 42 textos en cada una de las siguientes lenguas: español, inglés y ruso, que contenían tanto topónimos oficiales como no oficiales. En el caso del euskera, se analizaron 18 textos, ya que únicamente se emplearon topónimos oficiales cuya forma proviene directamente del propio idioma. En el caso del ruso se utilizaron formas toponímicas en alfabeto latino.

Escenario 4. Análisis cuantitativo de características estructurales. A partir del mismo conjunto de 210 textos analizados en el Escenario 1, se realizó un análisis cuantitativo que consideró la longitud de las respuestas y el número de recomendaciones por texto, diferenciando los resultados según la lengua utilizada. Estos datos se compararon con los resultados cuantitativos obtenidos en los Escenarios 1, 2 y 3.

4. RESULTADOS

Esta sección presenta los resultados derivados del análisis sistemático de las respuestas generadas por el modelo de lenguaje ChatGPT ante *prompts* sobre recomendaciones turísticas formulados en distintas lenguas y variantes toponímicas, recogidas en los escenarios presentados en la metodología.

4.1. Escenario 1: Reconocimiento de los destinos

La lengua del mensaje influyó significativamente en la identificación de destinos, especialmente cuando se empleó un sistema de escritura distinto. En los textos generados en ruso (alfabeto cirílico), los destinos no fueron reconocidos en el 16 % de los

casos, a diferencia de las respuestas en inglés, español y euskera, en las que todos los destinos fueron correctamente identificados.

Las fallas de reconocimiento en ruso se asociaron con el uso del alfabeto cirílico, la presencia de formas flexionadas y el tamaño del destino. Todos los errores se concentraron en los destinos más pequeños: Zumaia (72 % de respuestas irrelevantes) y Zarautz (22 %). Donostia / San Sebastián fue correctamente identificado en todos los casos.

El principal reto fue el reconocimiento de nombres en formas flexionadas, que representaron la mitad de las respuestas irrelevantes (9 de 17). Añadir el nombre del país junto al destino en cirílico mejoró la identificación, aunque no eliminó por completo los errores. La identificación fue plenamente exitosa cuando los nombres se escribieron en su forma local utilizando el alfabeto latino. Estos resultados evidencian las limitaciones de los modelos al procesar topónimos de destinos turísticos pequeños, cuando están escritos en sistemas de escritura distintos al sistema latino utilizado en el propio destino.

4.2. Escenario 2. Sensibilidad toponímica: nombres oficiales e históricos

En primer lugar, se analizaron las respuestas generadas a los *prompts* que incluían topónimos que presentan dos variantes. No se observaron fallos en el reconocimiento de destinos al utilizar topónimos oficiales en lengua cooficial local (Zumaia, Zarautz) ni en lengua principal del país (Zumaya, Zarauz). El número medio de recomendaciones fue similar entre variantes (≈ 8), sin diferencias relevantes en su tipología. Predominaron las recomendaciones de puntos de interés (PDI) específicos (67-69 %), seguidas por recomendaciones generales (25-26 %). Aproximadamente el 5 % de las recomendaciones en cada caso se referían a lugares fuera del destino y cerca de un 1 % mencionaban PDI inexistentes.

En segundo lugar, se analizaron las respuestas generadas a los *prompts* que incluían el topónimo que presenta tres variantes. En este caso, no se hallaron diferencias significativas entre las respuestas correspondientes a las tres formas. La versión bilingüe (Donostia / San Sebastián) generó un número discretamente mayor de recomendaciones por texto (11,4) en comparación con «Donostia» (10,9) y «San Sebastián» (11,2), pero sin diferencias estadísticamente relevantes. El bilingüismo produjo más recomendaciones en español y ruso, mientras que «San Sebastián» obtuvo más en inglés.

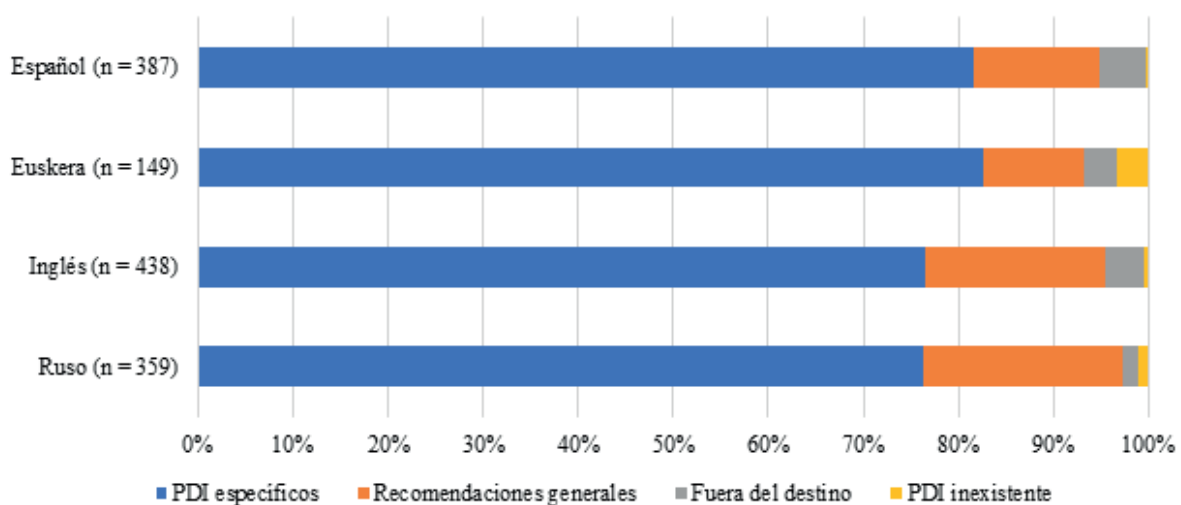
A nivel general, no se detectaron variaciones significativas entre topónimos oficiales e históricos en cuanto a estructura o volumen de contenido generado por IA. La aparente ventaja del topónimo bilingüe requiere validación con una muestra ampliada.

4.3. Escenario 3. Análisis de tipos de recomendaciones turísticas

En todas las lenguas analizadas, las recomendaciones generadas por IA priorizaron PDI específicos sobre recomendaciones generales, lo que evidencia una tendencia a contenidos concretos independientemente del idioma. No obstante, las lenguas principal y cooficial del destino (español y euskera) mostraron una mayor proporción de PDI específicos (81,7 % y 82,6 %, respectivamente) frente al inglés (76,5 %) y al ruso (76,3 %), así como una menor proporción de recomendaciones generales (Figura 2).

El euskera presentó la mayor tasa de PDI inexistentes (3,4 %), mientras que el ruso mostró la menor proporción de PDI ubicados fuera del destino (1,7 %). Estos resultados sugieren que el uso de las lenguas del destino incrementa la especificidad de las recomendaciones, aunque, en el caso de lenguas minoritarias, puede aumentar la aparición de errores, como referencias a lugares ficticios (alucinaciones). En contraste, los textos en lenguas extranjeras tienden a ofrecer información menos localizada.

Figura 2. Distribución de recomendaciones turísticas generadas por IA según la lengua (N = 1333)



FUENTE: ELABORACIÓN PROPIA.

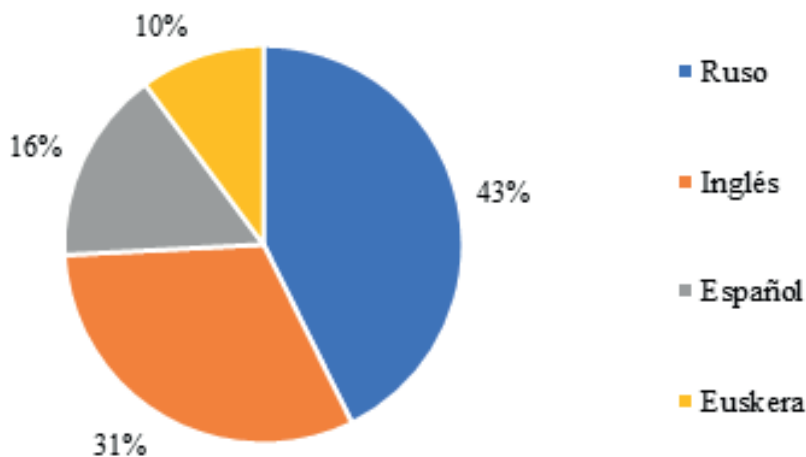
Los textos generados en lenguas extranjeras presentaron una mayor proporción de errores de contenido: el ruso concentró el 43 % del total de errores detectados, seguido del inglés con el 31 %. En comparación, el español y el euskera registraron tasas considerablemente menores (16 % y 10 %, respectivamente), aunque los resultados en euskera deben interpretarse con cautela debido al tamaño reducido de la muestra (Figura 3).

En ruso, el 63 % de los errores correspondieron a nombres incorrectos de PDI, lo que sugiere dificultades en la transcripción o reconocimiento de PDI en un sistema de escritura distinto. En inglés, el 57 % de los errores se debieron a una categorización incorrecta de los PDI, indicando problemas en la clasificación o traducción de tipos de lugares turísticos.

En las cuatro lenguas, las recomendaciones turísticas pertenecían con más frecuencia a las categorías «cultura y patrimonio» y «naturaleza y recreo». Además, los contenidos generados por la IA favorecieron sistemáticamente los PDI específicos frente a las recomendaciones generales en las categorías de «cultura y patrimonio», «naturaleza y recreo» y «eventos». En cambio, las recomendaciones generales predominaron en las categorías «gastronomía y compras» y «turismo creativo y vida local». En estas últimas categorías se mencionaron muy pocos comercios concretos, lo que refleja una tendencia

de los contenidos generados por IA a evitar nombrar establecimientos individuales de comida y bebida y a limitar las referencias a comercios concretos principalmente a los mercados con una trayectoria histórica destacada.

Figura 3. Distribución de errores de contenido según la lengua utilizada



Nota: El porcentaje se refiere al total de errores detectados en la muestra, en todas las lenguas analizadas, correspondientes a las respuestas sobre tres destinos turísticos: Zumaia, Zarautz y Donostia / San Sebastián.

FUENTE: ELABORACIÓN PROPIA.

En particular, la proporción de recomendaciones en «gastronomía y compras» y «turismo creativo y vida local» fue mayor en las lenguas extranjeras (inglés y ruso) que en las lenguas cooficiales del destino (Tabla 3). Por un lado, revela que los contenidos turísticos generados por IA ajustaron su enfoque temático según la lengua. Por otro lado, representa una importante laguna en la representación de los destinos.

Tabla 3. Distribución de recomendaciones turísticas según tipo (N = 1273)

| LENGUA | CULTURA Y PATRIMONIO | NATURALEZA Y RECREO | EVENTOS | GASTRONOMÍA Y COMPRAS | TURISMO CREATIVO Y VIDA LOCAL | OTROS |
|---------|----------------------|---------------------|---------|-----------------------|-------------------------------|-------|
| Euskera | 50 % | 37 % | 5 % | 6 % | 1 % | 1 % |
| Inglés | 42 % | 34 % | 5 % | 14 % | 4 % | 1 % |
| Ruso | 40 % | 37 % | 5 % | 11 % | 7 % | 1 % |
| Español | 44 % | 40 % | 6 % | 7 % | 0 % | 2 % |

Nota: El porcentaje corresponde al total de la fila.

FUENTE: ELABORACIÓN PROPIA.

4.4. Escenario 4. Análisis cuantitativo de características estructurales

Las salidas generadas por IA (N = 210) adoptaron mayoritariamente el formato de lista numerada (salvo dos textos sin reconocimiento de destino). El análisis cuantitativo reveló diferencias relevantes entre las lenguas: el inglés presentó los textos más extensos, con un promedio de 10 ítems y 327 palabras, seguido del español (9 ítems, 302 palabras). Los textos en ruso y euskera fueron más breves, con 8 ítems en promedio; no obstante, las respuestas en ruso fueron un 8 % más largas que en euskera (207 vs. 191 palabras) (Tabla 4).

Tabla 4. Variaciones de escenarios generados por IA según la lengua

| Lengua | Escenario 1. Reconocimiento de destino ¹ | | Escenario 2. Denominación oficial-histórica | | | | Escenario 3. Recomendaciones turísticas ¹ | | Escenario 4. Elementos estructurales y recuento de palabras en los resultados generados por la IA ¹ | | |
|---------|---|----|---|-------|--------------------------|-------|--|-------|--|-----------------|-----------------|
| | | | Zarautz, Zumaia | | Donostia / San Sebastián | | | | | | |
| | TG | NR | TG ² | RT | TG ² | RT | TG ² | RT | TG | Ítems numerados | Conteo palabras |
| | N | N | N | Media | N | Media | N | Media | N | Media | Media |
| Inglés | 42 | 0 | 24 | 9 | 18 | 12 | 42 | 10,4 | 42 | 10 | 327 |
| Español | 42 | 0 | 24 | 8 | 18 | 11 | 42 | 9,2 | 42 | 9 | 302 |
| Ruso | 108 | 17 | 24 | 7 | 18 | 10 | 42 | 8,5 | 108 | 8 | 207 |
| Euskera | 18 | 0 | - | - | - | - | 18 | 8,3 | 18 | 8 | 191 |

Nota. TG: texto generado; RT: recomendaciones turísticas; NR: no reconocido. ¹Los tres destinos estudiados. ²Para el ruso, se seleccionaron aquellas indicaciones en las que el topónimo del destino aparecía transcrito en alfabeto latino, con el fin de facilitar la identificación de los destinos y analizar la respuesta más completa posible.

FUENTE: ELABORACIÓN PROPIA.

El recuento de recomendaciones turísticas confirmó que las salidas más extensas fueron generadas en inglés, seguidas por las generadas en español. Los datos del Escenario 3 recogidos en la Tabla 4 muestran el número promedio de recomendaciones turísticas por texto generado por IA en las cuatro lenguas analizadas. Los *prompts* en inglés produjeron el promedio más alto (10,4 recomendaciones por texto), seguidos por el español (9,2), el ruso (8,5) y el euskera (8,3).

El Escenario 2, que no presenta la limitación de un número desigual de textos y cuenta con la misma cantidad de textos por categoría lingüística, arroja resultados similares. En promedio, se generaron más recomendaciones en inglés que en otros idiomas: 9 frente a 8 en español y 7 en ruso para los casos de Zumaia y Zarautz; y 12 frente a 11 en español y 10 en ruso para el caso de Donostia / San Sebastián (Tabla 4).

5. DISCUSIÓN Y CONCLUSIONES

Este estudio ha puesto en evidencia que las recomendaciones turísticas generadas por modelos de lenguaje de gran tamaño (LLM), como ChatGPT, presentan variaciones

sustanciales según la lengua de la consulta. Las diferencias observadas afectan tanto la veracidad y especificidad del contenido como la forma en que se estructuran y priorizan los puntos de interés. Estas variaciones no solo responden a cuestiones lingüísticas, sino también a factores culturales, tipográficos y geográficos que inciden directamente en la experiencia informativa de los usuarios. Los resultados ponen de manifiesto las limitaciones del contenido de recomendaciones turísticas generadas por inteligencia artificial en distintas lenguas. Revelan casos fallidos de reconocimiento de destinos relacionados con el uso del lenguaje y la ortografía del nombre del destino.

Desde la perspectiva de la planificación y gestión turística, los resultados obtenidos subrayan la necesidad de desarrollar sistemas de IA multilingües capaces de reconocer y armonizar adecuadamente las variantes toponímicas. Esto incluye diferencias ortográficas entre lenguas y sistemas de escritura, así como variantes toponímicas históricas y oficiales, como Zarautz / Zarauz. Si bien el reconocimiento de entidades nombradas funciona adecuadamente para las variantes oficiales e históricas, persisten desafíos importantes en el tratamiento de nombres escritos en sistemas de escritura distintos.

A medida que los destinos turísticos avanzan hacia ecosistemas inteligentes, en los que la mediación algorítmica define buena parte del contacto del visitante con el entorno, surgen nuevos desafíos de gobernanza en contextos multilingües. Actualmente, los destinos son particularmente vulnerables al riesgo de que los modelos de lenguaje de gran tamaño ofrezcan información incompleta, desactualizada y superficial sobre sus atractivos turísticos, puntos de interés y actividades que pueden realizarse en su territorio (Volchek & Ivanov, 2024). Las decisiones sobre qué variantes de nombres de lugar incluir en la señalética digital, en aplicaciones y en mapas, así como qué forma utilizar por defecto cuando la lengua del usuario no coincide con la lengua del destino, resultan cruciales. Estas cuestiones conectan la planificación turística con ámbitos como la política lingüística, la equidad digital y el diseño inclusivo de experiencias turísticas.

En términos de experiencia del usuario, la lengua elegida para interactuar con la IA tiene un impacto directo en la calidad y la utilidad del contenido recibido. Los resultados indican que los usuarios que formulan consultas en lenguas extranjeras, como el inglés o el ruso, tienden a recibir respuestas más genéricas y con una mayor incidencia de errores. En cambio, las consultas en lenguas locales —como el español y, en menor medida, el euskera— tienden a generar respuestas más específicas y culturalmente situadas, aunque también presentan riesgos, como la aparición de los puntos de interés inexistentes en contextos de baja representación digital. Estos hallazgos contradicen la suposición generalizada de que el inglés, debido a su predominancia en los datos de entrenamiento, proporciona recomendaciones superiores. Aunque en inglés las respuestas suelen ser más extensas, la información disponible en las lenguas del destino tiende a ser más detallada y relevante para el visitante.

La necesidad de revisar, adaptar y localizar los contenidos generados por IA en lenguas extranjeras ampliamente utilizadas se vuelve prioritaria para evitar la propagación de

desinformación y asegurar una representación justa y coherente de los destinos. Asimismo, resulta fundamental mejorar los sistemas de traducción, en particular en lo relativo al tratamiento de los nombres geográficos y de las categorías de puntos de interés. Las diferencias temáticas observadas entre lenguas sugieren que la IA adapta, de forma implícita, el enfoque de contenido en función del público objetivo percibido, lo que refuerza la necesidad de implementar estrategias multilingües más complejas, que no se limiten a la traducción literal, sino que consideren aspectos culturales y funcionales del turismo.

El escaso número de negocios específicos mencionados en las recomendaciones analizadas sugiere una tendencia de los LLM a evitar nombrar establecimientos concretos. Esto contrasta claramente con las recomendaciones turísticas generadas por los sistemas de recomendación tradicionales, que suelen ofrecer sugerencias personalizadas que incluyen nombres de restaurantes, cafeterías o tiendas específicas. Por un lado, la omisión de nombres comerciales limita la utilidad práctica del contenido turístico generado por LLM, ya que los turistas suelen buscar orientaciones concretas y aplicables, como dónde comer o hacer compras. Por otro lado, evitar la mención de nombres específicos puede interpretarse como una forma de reducir el sesgo comercial. Al no promocionar negocios particulares, el contenido generado por LLM mantiene un tono neutral que se ajusta a principios de equidad e imparcialidad.

Finalmente, se destaca la importancia de promover enfoques colaborativos en el desarrollo de tecnologías lingüísticas aplicadas al turismo. Este esfuerzo debe involucrar no solo a desarrolladores de IA, sino también a lingüistas, antropólogos, gestores de destinos y responsables de políticas públicas. Solo a través de esta cooperación intersectorial será posible construir marcos inclusivos y culturalmente sensibles que garanticen una representación justa, accesible y contextualizada de los destinos turísticos en la era de la inteligencia artificial generativa.

6. IMPLICACIONES PRÁCTICAS

Los turistas deben tener en cuenta que obtener recomendaciones relevantes para destinos pequeños es más difícil cuando se utiliza un sistema de escritura diferente al local. Esta limitación se ha observado especialmente en lenguas que utilizan alfabetos no latinos, como el cirílico (ruso), pero también es extensible a otros sistemas gráficos como el chino (hanzi), o el árabe. En estos casos, los modelos de inteligencia artificial pueden tener dificultades para reconocer correctamente los nombres de los destinos, especialmente cuando estos son menos conocidos o presentan formas toponímicas no estandarizadas. Por tanto, para los turistas que usan sistemas de escritura no latinos (por ejemplo, el cirílico), se recomienda incluir el nombre del país al escribir un topónimo en su propio sistema de escritura o, alternativamente, escribirlo en el sistema de escritura local.

Los gestores de destinos deben definir estrategias claras sobre el uso de variantes toponímicas en medios digitales, considerando que ese contenido alimenta los

sistemas de IA. Estas decisiones deben tener en cuenta la diversidad lingüística de los visitantes y garantizar una experiencia informativa coherente, accesible y equitativa. Los desarrolladores de IA deben mejorar los sistemas de reconocimiento de entidades toponímicas multilingües.

Cada vez es más esencial que quienes diseñan políticas públicas incorporen criterios de equidad lingüística y de representación cultural en la gobernanza de destinos, para garantizar una gestión inteligente, inclusiva y respetuosa con la diversidad cultural del territorio. Esto incluye promover marcos normativos que regulen el uso de la IA en el turismo, velando por la inclusión de lenguas minorizadas y por la adecuada representación de territorios con baja visibilidad digital.

7. LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

Una limitación importante es el tamaño desigual de los subconjuntos en algunos casos de comparación, debido a la dificultad de seleccionar ortografías de destinos que sean comparables, especialmente en destinos multilingües. Además, se necesitan muestras más grandes y equilibradas para determinar si el uso de topónimos bilingües produce recomendaciones más matizadas.

Este estudio analizó el uso de topónimos de destinos en alfabeto latino dentro de entradas en cirílico. Investigaciones futuras deberían examinar el caso inverso: si el uso de topónimos en cirílico dentro de entradas en inglés (por ejemplo, en el caso de Bulgaria) mejora el reconocimiento en comparación con las ortografías en inglés. También es necesario profundizar en cómo varían las recomendaciones en función del tamaño del destino, ya que los efectos observados pueden diferir significativamente entre destinos grandes y pequeños.

Por último, el análisis no contempla las estructuras lingüísticas que pueden afectar la longitud de las respuestas. Por ejemplo, las respuestas más breves observadas en euskera podrían estar relacionadas con su gramática aglutinante, en la que los marcadores gramaticales se expresan como sufijos y no como palabras separadas.

Financiación: Este trabajo contó con el apoyo del proyecto PID2021-1278930B-100 financiado por MICIU/AEI/ 10.13039/501100011033 y por FEDER, UE.

8. REFERENCIAS

ALIA. (2025). *La infraestructura pública de IA en castellano y lenguas cooficiales*. <https://alia.gob.es/>

Blodgett, S. L., & O'Connor, B. (2017). *Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English* (No. arXiv:1707.00061v1). arXiv. <https://doi.org/10.48550/ARXIV.1707.00061>

Cohen, S. A., Prayag, G., & Moital, M. (2014). Consumer behaviour in tourism: Concepts, influences and opportunities. *Current Issues in Tourism*, 17(10), 872–909. <https://doi.org/10.1080/13683500.2013.850064>

Crompton, J. L. (1979). Motivations for pleasure vacation. *Annals of Tourism Research*, 6(4), 408–424. [https://doi.org/10.1016/0160-7383\(79\)90004-5](https://doi.org/10.1016/0160-7383(79)90004-5)

- Dann, G. M. S. (1977). Anomie, ego-enhancement and tourism. *Annals of Tourism Research*, 4(4), 184–194. [https://doi.org/10.1016/0160-7383\(77\)90037-8](https://doi.org/10.1016/0160-7383(77)90037-8)
- Dann, G. M. S. (1981). Tourist motivation and appraisal. *Annals of Tourism Research*, 8(2), 187–219. [https://doi.org/10.1016/0160-7383\(81\)90082-7](https://doi.org/10.1016/0160-7383(81)90082-7)
- Doğan, S., & Niyet, İ. Z. (2024). Artificial Intelligence (AI) in Tourism. In C. Tanrisever, H. Pamukçu, & A. Sharma (Eds.), *Future Tourism Trends* (Vol. 2, pp. 3–21). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-83753-970-320241001>
- Fleisig E., Smith G., Bossi M., Rustagi I., Yin X., & Klein D. (2024). *Linguistic Bias in ChatGPT: Language Models Reinforce Dialect Discrimination* (No. arXiv:2406.08818v3). arXiv. <https://doi.org/10.48550/arXiv.2406.08818>
- Fung Y., Zhao R., Doo J., Sun C., & Ji H. (2024). *Massively Multi-Cultural Knowledge Acquisition & LM Benchmarking* (No. arXiv:2402.09369v1). arXiv. <https://doi.org/10.48550/ARXIV.2402.09369>
- Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The Future of Retailing. *Journal of Retailing*, 93(1), 1–6. <https://doi.org/10.1016/j.jretai.2016.12.008>
- Grundner L., & Neuhofer B. (2021). The bright and dark sides of artificial intelligence: A futures perspective on tourist destination experiences. *Journal of Destination Marketing & Management*, 19, 100511. <https://doi.org/10.1016/j.jdmm.2020.100511>
- GuideGeek. (2025). *For Destinations – GuideGeek*. <https://guidegeek.com>
- Gursoy, D., Li, Y., & Song, H. (2023). ChatGPT and the hospitality and tourism industry: An overview of current trends and future research directions. *Journal of Hospitality Marketing & Management*, 32(5), 579–592. <https://doi.org/10.1080/19368623.2023.2211993>
- Hershcovich, D., Frank, S., Lent, H., Lhoneux, M., de, Abdou, M., Brandl, S., Bugliarello, E., Piqueras, L. C., Chalkidis, I., Cui, R., Fierro, C., Margatina, K., Rust, P., & Sjøgaard, A. (2022). *Challenges and Strategies in Cross-Cultural NLP* (No. arXiv:2203.10020v1). arXiv. <https://doi.org/10.48550/arXiv.2203.10020>
- INE. (2024). *Cifras oficiales de población de los municipios españoles en aplicación de la Ley de Bases del Régimen Local*. Instituto Nacional de Estadística. <https://www.ine.es/jaxiT3/Tabla.htm?t=2873&L=0>
- Jørgensen A., & Sjøgaard A. (2021). *Evaluation of Summarization Systems across Gender, Age, and Race* (No. arXiv:2110.04384v1). arXiv. <https://doi.org/10.48550/ARXIV.2110.04384>
- Kim, J. H., Kim, J., Kim, C., & Kim, S. (Sam). (2023). Do you trust ChatGPTs? Effects of the ethical and quality issues of generative AI on travel decisions. *Journal of Travel & Tourism Marketing*, 40(9), 779–801. <https://doi.org/10.1080/10548408.2023.2293006>
- Lehto, X. Y., Timothy, D. J., & Xiao, H. (2025). The future of destinations: Rethinking smartness, resisting algorithmic flattening, and reclaiming tourism place. *Journal of Destination Marketing & Management*, 37, 101021. <https://doi.org/10.1016/j.jdmm.2025.101021>
- Lew, A. A. (1987). A framework of tourist attraction research. *Annals of Tourism Research*, 14(4), 553–575. [https://doi.org/10.1016/0160-7383\(87\)90071-5](https://doi.org/10.1016/0160-7383(87)90071-5)
- Li, C., Chen, M., Wang, J., Sitaram, S., & Xie, X. (2024). *CultureLLM: Incorporating Cultural Differences into Large Language Models* (No. arXiv:2402.10946v3). arXiv. <https://doi.org/10.48550/ARXIV.2402.10946>
- Liu, C. C., Koto, F., Baldwin, T., & Gurevych, I. (2024). *Are Multilingual LLMs Culturally-Diverse Reasoners? An Investigation into Multicultural Proverbs and Sayings* (No. arXiv:2309.08591v2). arXiv. <https://doi.org/10.48550/arXiv.2309.08591>
- Liu, Z. (2023). Cultural Bias in Large Language Models: A Comprehensive Analysis and Mitigation Strategies. *Journal of Transcultural Communication*, 3(2), 224–244. <https://doi.org/10.1515/jtc-2023-0019>
- Masoud, R. I., Liu, Z., Ferianc, M., Treleaven, P., & Rodrigues, M. (2024). *Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede's Cultural Dimensions* (No. arXiv:2309.12342v2). arXiv. <https://doi.org/10.48550/arXiv.2309.12342>
- Plog, S. C. (1974). Why Destination Areas Rise and Fall in Popularity. *Cornell Hotel and Restaurant Administration Quarterly*, 14(4), 55–58. <https://doi.org/10.1177/001088047401400409>
- Ren, R., Yingwei, W. X., Yao, X., Cole, S., & Wang, H. (2025). *Whose Journey Matters? Investigating Identity Biases in Large Language Models (LLMs) for Travel Planning Assistan-*

- ce (No. arXiv:2410.17333v3). arXiv. <https://doi.org/10.48550/arXiv.2410.17333>
- Richards, G. (2019). Creative tourism: Opportunities for smaller places? *Tourism & Management Studies*, 15(1SI), 7–10. <https://doi.org/10.18089/tms.2019.15si01>
- Rico, A., Martínez-Blanco, J., Montlleó, M., Rodríguez, G., Tavares, N., Arias, A., & Oliver-Solà, J. (2019). Carbon footprint of tourism in Barcelona. *Tourism Management*, 70, 491–504. <https://doi.org/10.1016/j.tourman.2018.09.012>
- Rigin, S., Kottekkadan, N. N., Thomas, T. K., & Kk, M. N. (2025). Generative AI tools (ChatGPT*) in tourism research: An experimental conversation. *Tourism and Hospitality Management*, 31(2). <https://doi.org/10.20867/thm.31.2.13>
- Robinson, M., Viswanathan, S., Baffoun, S., Qi, A. Y., & Subramaniam, S. (2024). *The travel industry's new trip—How generative AI can redefine customer experiences and unlock new value for organizations*. Accenture. <https://www.accenture.com/content/dam/accenture/final/accenture-com/document-3/The-Travel-Industrys-New-Trip-Final.pdf>
- Schmuckler, M. A. (2010). What Is Ecological Validity? A Dimensional Analysis. *Infancy*, 2(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02
- Shin, S., Kim, J., Lee, E., Yhee, Y., & Koo, C. (2025). ChatGPT for Trip Planning: The Effect of Narrowing Down Options. *Journal of Travel Research*, 64(2), 247–266. <https://doi.org/10.1177/00472875231214196>
- Stoica, A.-A. (2020, May 9). Algorithmic fairness for networked algorithms. *Adaptive Agents and Multi-Agent Systems*. 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand. <https://www.ifaamas.org/Proceedings/aamas2024/>
- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural Bias and Cultural Alignment of Large Language Models. *PNAS Nexus*, 3(9). <https://doi.org/10.1093/pnasnexus/pgae346>
- UN Tourism & Saxion University of Applied Sciences. (2025). *Artificial Intelligence Adoption in Tourism – Key Considerations for Sector Stakeholders*. UN Tourism. <https://doi.org/10.18111/9789284426065>
- United Nations, Department of Economic and Social Affairs. (2010). *International Recommendations for Tourism Statistics 2008*. United Nations. <https://www.e-unwto.org/doi/abs/10.18111/9789211615210>
- Volchek, K., & Ivanov, S. (2024). ChatGPT as a Travel Itinerary Planner. In K. Berezina, L. Nixon, & A. Tuomi (Eds.), *Information and Communication Technologies in Tourism 2024* (pp. 365–370). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-58839-6_38
- World Tourism Organization. (2019). *UNWTO Tourism Definitions*. UNWTO. <https://doi.org/10.18111/9789284420858>
- Wörndl, W., & Herzog, D. (2022). Mobile Applications for e-Tourism. In Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of e-Tourism* (pp. 273–293). Springer International Publishing. https://doi.org/10.1007/978-3-030-48652-5_17
- Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management*, 26(1), 45–56. <https://doi.org/10.1016/j.tourman.2003.08.016>
- Zhang, S., Zhang, X., Zhang, W., & Søgaaard, A. (2021). Sociolectal Analysis of Pretrained Language Models. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 4581–4588. <https://doi.org/10.18653/v1/2021.emnlp-main.375>