

Validación de la capacidad de ChatGPT en la clasificación de competencias lingüísticas en español según el MCER

Validation of ChatGPT's Ability to Classify Spanish Language Competencies According to the CEFR

Tatyana Martynova
Universidad de Granada
martynovatk@correo.ugr.es
ORCID: [0009-0008-3884-3966](https://orcid.org/0009-0008-3884-3966)

Resumen: Este estudio analiza la capacidad de ChatGPT para clasificar y generar contenido en español según los niveles del Marco Común Europeo de Referencia para las Lenguas (MCER), desde A1 hasta C2. La investigación se centra en cinco tareas: 1) evaluación del contenido gramatical en programas educativos; 2) clasificación de materiales audiovisuales; 3) distribución del vocabulario; 4) adaptación de textos a distintos niveles; y 5) elaboración de preguntas para exámenes de Diplomas de Español como Lengua Extranjera (DELE). Los resultados muestran que ChatGPT tiene una comprensión básica de los descriptores del MCER y destaca en la adaptación de textos y en la formulación de preguntas orales. Además, demuestra capacidad para simplificar contenidos y proponer temas relevantes. Sin embargo, presenta limitaciones significativas, como la falta de precisión en la clasificación de vocabulario y materiales audiovisuales, la superficialidad en los programas educativos para niveles avanzados y cambios innecesarios en la adaptación de textos. Se concluye que ChatGPT puede ser una herramienta valiosa para complementar el aprendizaje de idiomas, pero su uso requiere supervisión docente para garantizar la fiabilidad de los contenidos. Se destaca la necesidad de profundizar en investigaciones futuras que permitan optimizar sus capacidades con el objetivo de mejorar su eficacia en entornos educativos formales e informales, y garantizar su alineación con los estándares del MCER.

Palabras clave: ChatGPT; MCER; español como lengua extranjera; clasificación de competencias lingüísticas.

Abstract: This study analyzes ChatGPT's ability to classify and generate content in Spanish according to the levels of the Common European Framework of Reference for Languages (CEFR), from A1 to C2. The research focuses on five key tasks: evaluating grammatical content in educational programs, classifying audiovisual materials, distributing vocabulary, adapting texts to different proficiency levels, and developing questions for DELE exams. The results show that ChatGPT has a basic understanding of CEFR descriptors, standing out in text adaptation and the formulation of oral questions. Additionally, it demonstrates the ability to simplify content and propose relevant topics. However, it also presents significant limitations, such as a lack of precision in classifying vocabulary and audiovisual materials, superficiality in educational programs for advanced levels, and unnecessary changes when adapting texts. It is concluded that ChatGPT can be a valuable tool to complement language learning, but its use requires teacher supervision to ensure content reliability. The need for further research to optimize its capabilities is highlighted, with the aim of improving its effectiveness in both formal and informal educational contexts and ensuring its alignment with CEFR standards.

Keywords: ChatGPT; CEFR; Spanish as a foreign language (SFL); classification of language competencies.

INTRODUCCIÓN

En los últimos años la inteligencia artificial (IA), en particular la inteligencia artificial generativa, se ha convertido en una parte esencial de diversas áreas de la vida, incluida la educación. Uno de los ejemplos más destacados del uso de la IA en el aprendizaje es ChatGPT, un modelo de lenguaje tipo chatbot capaz de generar texto, analizar preguntas y comunicarse en tiempo real con la persona que lo usa. En el contexto de la enseñanza de idiomas extranjeros, ChatGPT ha abierto nuevas posibilidades, ofreciendo apoyo personalizado, acceso constante y herramientas innovadoras que facilitan el aprendizaje autónomo o asistido.

Han aparecido varios estudios dedicados a la eficacia del uso de ChatGPT en el entorno educativo y académico (Alemán Ariza et al., 2024; Bozkurt, 2024; Larverde Gonzalez y Rodríguez Hernández, 2025; Montoya, 2024; Moreno Padilla, 2019; Orozco Castaño, 2022). Además, se han desarrollado investigaciones centradas específicamente en el aprendizaje de idiomas, sobre todo en inglés (Chicaíza Chicaíza et al., 2023; Kohnke et al., 2023; Shaikh et al., 2023), así como intentos de estimar y valorar las capacidades de esta herramienta mediante experimentos prácticos: la calidad del *feedback* proporcionado por ChatGPT (Hernández, 2024),

la habilidad de crear preguntas para exámenes (Rivera-Rosas et al., 2024), la evaluación de escritura (vanderLinde y Mera Cury, 2024), entre otras.

Este estudio busca evaluar la capacidad de ChatGPT para clasificar y generar contenido en español de acuerdo con los niveles del Marco Común Europeo de Referencia (MCER), desde A1 hasta C2. Pretende ofrecer tanto un diagnóstico preliminar sobre el grado de alineación entre las respuestas generadas y los descriptores del MCER como una propuesta metodológica para futuras investigaciones más desarrolladas.

Los niveles del MCER (Consejo de Europa, 2002) proporcionan una descripción estándar y global de las competencias lingüísticas, lo que permite medir habilidades en comprensión, expresión y escritura en diferentes idiomas. Así, al aprender español, el estudiantado debe comprender bien las competencias específicas de su nivel: desde el uso correcto de las estructuras gramaticales hasta el dominio de un vocabulario adaptado y la capacidad de interactuar en situaciones reales. Este enfoque asegura una progresión lógica y eficiente en el aprendizaje del idioma. Además, ayuda a que la comunidad estudiantil, el cuerpo docente, las personas que diseñan los cursos y las que crean materiales didácticos, así como quienes elaboran los exámenes, tengan inevitablemente implicación en este proceso, que supone centrarse en una dimensión concreta (Consejo de Europa, 2002, p. 10). No obstante, investigaciones recientes han subrayado que su aplicación puede presentar márgenes de subjetividad (De la Torre García, 2016; Pineda Sánchez et al., 2024), especialmente en la asignación de contenidos o vocabulario a determinados niveles. A pesar de ello, sigue siendo un marco de referencia fundamental para la enseñanza de idiomas en Europa y otras regiones.

Para evaluar ChatGPT, este estudio propone cinco tareas clave:

- Evaluar el contenido gramatical de un programa educativo generado por ChatGPT para un nivel específico y verificar si cumple con las necesidades gramaticales correspondientes.
- Analizar la capacidad de comprensión de los niveles: determinar si el modelo puede identificar correctamente el nivel de un video o material audiovisual.
- Examinar la distribución del vocabulario: comprobar si el vocabulario presentado está correctamente clasificado según los niveles del MCER.
- Medir la capacidad de adaptar textos: comprobar si ChatGPT puede ajustar un mismo texto a diferentes niveles manteniendo coherencia y adecuación.

- Crear preguntas para exámenes: analizar si el modelo es capaz de generar preguntas útiles para la preparación de exámenes DELE en distintos niveles.

Derivado de estas tareas, y con el objetivo de orientar el análisis empírico, se plantean algunas preguntas de investigación alineadas con el objetivo central del estudio. La pregunta general integradora es la siguiente: ¿en qué medida puede ChatGPT generar y clasificar contenidos didácticos en español como lengua extranjera conforme a los descriptores del MCER, desde el nivel A1 hasta el C2? Por su parte, las preguntas específicas son:

1. ¿En qué grado los programas educativos elaborados por ChatGPT integran contenidos gramaticales pertinentes y adecuados a cada nivel del MCER? ¿Pueden ser utilizados en procesos de autoaprendizaje para guiar al estudiantado a través de las necesidades gramaticales propias de cada nivel?
2. ¿Es capaz ChatGPT de determinar con precisión el nivel de dificultad de materiales audiovisuales?
3. ¿Clasifica ChatGPT el vocabulario de manera coherente con los criterios establecidos en el *Plan curricular del Instituto Cervantes* (2006)? ¿Puede considerarse fiable su uso en tareas en las que el nivel léxico desempeña un papel crucial?
4. ¿Es capaz ChatGPT de adaptar textos escritos de forma coherente y progresiva según los diferentes niveles de competencia lingüística definidos por el MCER?
5. ¿En qué medida las preguntas orales generadas por ChatGPT para la preparación de los exámenes DELE reflejan los objetivos comunicativos y las exigencias específicas de cada nivel?

Las tareas planteadas incluyen tanto capacidades productivas como receptivas y se centran en las actividades lingüísticas establecidas por el MCER, que son la comprensión, la expresión, la interacción y la mediación (Consejo de Europa, 2002, p. 14). Además, ofrecen una visión integral en la creación de programas y en la comprensión de las necesidades de los exámenes y evaluaciones. Este enfoque

no solo mide las capacidades individuales de ChatGPT, sino que también permite proponer mejoras metodológicas para su integración en la enseñanza de idiomas.

Al mismo tiempo, hay que recordar que los niveles solo reflejan una dimensión vertical (Consejo de Europa, 2002, p. 18). Por ello, este trabajo no pretende determinar si es posible aprender un idioma utilizando únicamente las capacidades de ChatGPT o si el alumnado no será capaz de lograr el dominio del español si el programa muestra algunos errores, ya que el aprendizaje de un idioma es un proceso más complejo. Sin embargo, entender el grado de conocimiento de las características principales de los niveles lingüísticos por parte de ChatGPT puede esclarecer si tanto docentes como estudiantes que practican el autoaprendizaje pueden confiar en la información proporcionada por el programa, si este es capaz de ofrecer materiales adicionales adecuados y si puede valorar el nivel de quien aprende español para proporcionar un *feedback* razonable.

TAREAS DE VALIDACIÓN DE LA CAPACIDAD DE CHATGPT

A continuación se revisan las cinco tareas de validación de la capacidad de ChatGPT en la clasificación de competencias lingüísticas en español según el MCER, mencionadas anteriormente.

Tarea 1. Evaluación del contenido gramatical en un programa educativo generado por ChatGPT

En la primera tarea se evalúa la capacidad de ChatGPT para crear un programa de español para diferentes niveles de aprendizaje. Además, verificaremos si la gramática proporcionada para cada nivel coincide con los contenidos del *Plan curricular del Instituto Cervantes* (Instituto Cervantes, 2006).

Es relevante señalar que en esta parte, así como en las tareas posteriores, el *prompt* (instrucciones) fue introducido inicialmente en inglés. Esta decisión se fundamenta en varios estudios que evidencian diferencias significativas en las respuestas generadas por ChatGPT según el idioma empleado (Coldewey, 2023; Joshi et al., 2024; Seghier, 2023). Considerando estos hallazgos, se ha optado por el inglés con el objetivo de evaluar la calidad de las respuestas tanto para docentes que hacen uso de esta herramienta como para contextos de autoaprendizaje, especialmente en situaciones en las que el estudiantado que utiliza ChatGPT de forma autónoma posee un nivel inicial de competencia en español. En este escenario hipotético, resulta más probable que las solicitudes se formulen en un idioma

distinto al español, y, dado el carácter global y universal del inglés, su uso en cada una de las tareas permite simular de manera más realista las condiciones en las que esta herramienta podría emplearse en entornos educativos internacionales.

El *prompt* utilizado fue “Create a Spanish A1-level study program with grammar, vocabulary and weekly topics”. Este mensaje se ajustó de forma progresiva para cada nivel, cambiando únicamente la indicación del nivel correspondiente (por ejemplo: “Create a Spanish B1-level study program...” y “Create a Spanish C1-level study program...”), con el objetivo de mantener la coherencia metodológica y facilitar la comparación entre las respuestas generadas.

Analizaremos cómo ChatGPT genera programas para tres niveles diferentes: A1, B1 y C1. A partir de esta información, elaboraremos un cuadro que recopile los temas gramaticales incluidos por ChatGPT en los programas de español para cada nivel y los compararemos con los contenidos establecidos en el *Plan curricular del Instituto Cervantes* (Instituto Cervantes, 2006) (Cuadro 1).

Cuadro 1. Comparación de los temas gramaticales incluidos en los programas de aprendizaje de español generados por ChatGPT para los niveles A1, B1 y C1

A1	B1	C1
<ul style="list-style-type: none"> • Present tense of regular <i>-ar, -er, -ir</i> verbs. • Prepositions of place (<i>en, sobre, bajo, cerca de</i>). • Verb <i>hay</i> (<i>there is/there are</i>). • Introduction to modal verbs (<i>querer, poder</i>). • Singular and plural nouns in context. • Verb <i>ir</i> (<i>to go</i>) and <i>venir</i> (<i>to come</i>). • Imperatives for directions (<i>sigue, gira, toma</i>). • Verb <i>hacer</i> for weather expressions. 	<ul style="list-style-type: none"> • Review of pretérito perfecto (<i>he viajado</i>) and pretérito indefinido (<i>viajé</i>). • Time markers: <i>hace, ya, todavía no, nunca</i>. • Imperfecto for descriptions (<i>era, tenía</i>). • Contrast between “indefinido” and “imperfect”. • Verbs like <i>gustar</i> and <i>parecer</i>. • Subjunctive with opinions (<i>es importante que...</i>). 	<ul style="list-style-type: none"> • Advanced connectors (<i>aunque, sin embargo, no obstante</i>). • Sentence restructuring for variety. • Subjunctive in relative clauses (<i>lo que sea necesario</i>). • <i>Si</i> clauses with the subjunctive. • Passive voice (<i>se dice que, fue construido por</i>). • Advanced use of impersonal constructions. • Concessive clauses (<i>a pesar de que, por más que</i>). • Use of <i>aunque</i> with indicative and subjunctive. • Use of “pluscuamperfecto” and “pretérito perfecto compuesto”.

A1	B1	C1
<ul style="list-style-type: none"> • Adjectives and nouns for weather and clothing. • Verb <i>gustar</i> (to like). • Reflexive verbs (<i>levantarse, bañarse</i>). • Verb <i>doler</i> (to hurt). • Indirect object pronouns (<i>me, te, le</i>). • Near future with <i>ir a</i>. • Recap of present tense verbs. • Review key verb conjugations. • Practice sentence structures. 	<ul style="list-style-type: none"> • Reflexive verbs in present and past tenses (<i>me levanto, me levanté</i>). • Frequency adverbs (<i>normalmente, a menudo</i>). • Imperative (formal and informal commands). • Use of pronouns with commands (<i>hazlo, cómpreselo</i>). • Future tense (<i>hablaré, viviré</i>). • Conditional for hypothetical situations (<i>me gustaría</i>). • Reported speech (<i>dijo que, comentó que</i>). • Use of <i>ser</i> and <i>estar</i> in contexts. • Subjunctive for advice (<i>te recomiendo que...</i>). • Conditional with advice (<i>yo en tu lugar</i>). • Conditional tense (<i>sería, tendría</i>). • Review of past and future tenses. • Practice complex sentences with conjunctions. 	<ul style="list-style-type: none"> • Mastering narrative techniques with adverbs and transitions. • Advanced pronouns (<i>cuyo, el cual</i>). • Vocabulary differences between formal and informal settings. • Use of <i>vos</i> and regional colloquialisms. • Common regional idiomatic expressions. • Advanced relative clauses and syntactic variety. • Use of rhetorical questions and parallel structures. • Review indirect speech (<i>dijo que, comentó que</i>). • Recognizing reduced forms in speech (<i>pa', tó</i>). • Nominalization (<i>la construcción, el análisis</i>). • Use of prepositional phrases for precision. • Advanced conditional sentences for hypothetical cultural comparisons. • Subjunctive for expressing opinions (<i>me sorprende que</i>).

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025). ChatGPT (versión del 20 de enero).

El programa de ChatGPT (OpenAI, 2025) para el nivel A1 ofrece un curso de 12 semanas estructurado en cuatro apartados semanales: *topics* (temas), *grammar* (gramática), *vocabulary* (vocabulario) y *activities* (actividades). Aunque adecuado en general, omite elementos esenciales como demostrativos (*este, esa, aquello*), posesivos (*mi, tus, suyos*) y numerales cardinales, pues incluye solo ordinales. Además, considera objetos indirectos, pero no directos, fundamentales para la construcción básica de oraciones.

ChatGPT (OpenAI, 2025) sugiere recursos adicionales, como libros (*Español en Marcha A1, Aula Internacional 1*), aplicaciones (Duolingo, Memrise, Busuu) y canales de video (SGEL ELE Español para Extranjeros, Easy Spanish). El diseño, enfocado a aprendizaje grupal con actividades interactivas, carece de prácticas individuales, lo que lo hace poco adecuado para el autoaprendizaje. Por ello, en B1 se solicitó un programa específico para autoaprendizaje: “Create a Spanish B1-level self-study program”.

Para B1, el programa mantiene una duración de 12 semanas, sin considerar la mayor carga horaria requerida para este nivel. Aunque muestra avances respecto a A1, presenta lagunas frente al *Plan curricular del Instituto Cervantes* (Instituto Cervantes, 2006), como la falta de variación entre español de España y de (algunas regiones de) América Latina, los grados del adjetivo, los cuantificadores y el pretérito pluscuamperfecto. Sobre el subjuntivo, solo incluye un uso limitado (*es importante que...*), dejando fuera expresiones desiderativas (*ojalá*), de duda (*quizá, tal vez*) y subordinadas con valor desiderativo (*quiero que vengas*).

En cuanto a recursos adicionales, se destacan herramientas útiles como *Gramática de uso del español B1-B2* (Edelsa), aplicaciones (Memrise, Quizlet, Anki) y recursos auditivos (Españolistas, Duolingo Spanish Podcast, Easy Spanish), aunque algunos, como los artículos de *El País*, podrían ser demasiado avanzados para B1.

El programa para C1 mantiene la misma estructura de 12 semanas, lo que no refleja la carga horaria necesaria recomendada por el Instituto Cervantes: unas 180 horas (15 semanales), frente a las 120 horas de B1 (10 semanales) y 60 de A1 (5 semanales) (Instituto Cervantes, s. f.). No obstante, cabe destacar algunos aspectos positivos, ya que ChatGPT recomienda más recursos adicionales para los niveles avanzados y tareas que el alumnado puede realizar utilizándolos. Entre estas recomendaciones se incluye ver programas de comedia y analizar las referencias culturales, leer artículos relacionados con su área en BBC Mundo, escuchar pódcast de *El País* y noticias de RTVE, participar en debates o ver charlas de TED en español, y leer ensayos de la sección de opinión de *El País*, entre otros.

En resumen, aunque ChatGPT es capaz de generar propuestas estructuradas y ofrece recursos complementarios útiles, presenta importantes limitaciones en

cuanto a la exhaustividad y adecuación gramatical, especialmente en niveles superiores. Esto refuerza la necesidad de supervisión docente en su uso didáctico.

Tarea 2. Determinar si el modelo puede identificar correctamente el nivel de un video o material audiovisual

En esta segunda tarea se evalúa la capacidad de ChatGPT para identificar el nivel de dificultad de materiales audiovisuales destinados a la enseñanza de español como lengua extranjera. Los videos seleccionados provienen del canal de YouTube SGEL ELE, que publica contenidos vinculados a manuales didácticos de ELE y clasificados por nivel. Con el fin de evitar sesgos derivados de los títulos, imágenes o enlaces, se utilizó la función «Mostrar texto del vídeo» disponible en la pestaña «Transcripción» de YouTube para extraer el contenido verbal íntegro de cada video. Estos textos se introdujeron en ChatGPT como entrada en formato plano, sin ninguna indicación explícita del nivel. Este procedimiento permitió simular un escenario en el que el modelo debía inferir el nivel únicamente a partir del contenido lingüístico, sin información contextual adicional.

Se solicitó al modelo que clasificara cada fragmento dentro del rango A1-B2, ya que a partir del nivel C1 el uso de materiales no adaptados es común y no necesariamente sigue criterios pedagógicos tan estrictos.

El *prompt* utilizado fue: “Determine the level of this video according to the CEFR”. Este mensaje fue el mismo para todos los casos con el fin de mantener la coherencia metodológica y asegurar una comparación uniforme entre las respuestas generadas.

Como se observa, ChatGPT presenta dificultades significativas para estimar correctamente el nivel en la mayoría de los casos analizados, ya que se equivocó en 15 de las 24 ocasiones (Cuadros 2 y 3). Estas dificultades son menos frecuentes en niveles avanzados, como B2, donde la mayor complejidad de los videos facilita que el modelo identifique el nivel con mayor precisión a partir del tema y el contexto. Sin embargo, en niveles como A2, que abarcan una amplia variedad de temas con vocabulario y gramática básica, ChatGPT tiende a asignar niveles más altos, como B1 o incluso B2. Por ejemplo, en el caso de todos los videos clasificados como A2, el modelo los sobreestimó.

Las *playlists* “Vitamina” y “Agencia ELE”, colecciones de videos basadas en los libros de texto homónimos, presentan mayores desafíos para ChatGPT debido a su estructura más elaborada y a la inclusión de temas no solo cotidianos, sino también polémicos o abstractos. Un ejemplo es el video *Vitamina A2 - Unidad 5*

/ *Vitamina Básico - Unidad 13*, que ChatGPT clasificó como nivel B2, cuando, en realidad, corresponde al nivel A2.

Cuadro 2. Evaluación de los niveles de videos A1 y A2 según la clasificación de ChatGPT

Nombre de video	Nivel del video	Nivel indicado por ChatGPT	Nombre de video	Nivel del video	Nivel indicado por ChatGPT
<i>Compañeros Nueva edición - Unidad "La familia" - La familia de María</i>	A1	A2	<i>Vitamina A2 - Unidad 5 / Vitamina Básico - Unidad 13</i>	A2	B2
<i>Compañeros Nueva edición - Unidad "¿Dónde están las llaves?" - ¿Dónde está mi cámara?</i>	A1	A1	<i>Vitamina A2 - Unidad 2 / Vitamina Básico - Unidad 10 (apartado C)</i>	A2	B2
<i>Vitamina A1/ Vitamina Básico - Unidad 8</i>	A1	A2	<i>Agencia ELE - Vídeo 8: Un viaje de cine</i>	A2	B1
<i>Agencia ELE - Vídeo 5: Una gran cantante</i>	A1	B1	<i>Compañeros Nueva edición - Unidad "El futuro del planeta" - Voluntarios</i>	A2	B1
<i>Agencia ELE - Vídeo 4: Vacaciones en España</i>	A1	B1	<i>Compañeros Nueva edición - Unidad "Sucesos" - Robo en el edificio</i>	A2	B1
<i>NEEM 1/NEEM Básico - Unidad 4 Una casa especial</i>	A1	A2	<i>NEEM 2 - Unidad 2/ NEEM Básico - Unidad 10 Mi amigo Aníbal</i>	A2	B1

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025). ChatGPT (versión del 25 de enero).

Cuadro 3. Evaluación de los niveles de videos B1 y B2 según la clasificación de ChatGPT

Nombre de video	Nivel del video	Nivel indicado por ChatGPT	Nombre de video	Nivel del video	Nivel indicado por ChatGPT
<i>Compañeros Nueva edición - Unidad "¿Qué quieres ser?" - ¿Qué quieres ser?</i>	B1	B1	<i>Vitamina B2 - Unidad 2</i>	B2	B2
<i>Compañeros Nueva edición - Unidad "¡Gastar y gastar!" - El sueño de Pablo</i>	B1	B1	<i>Vitamina B2 - Unidad 9</i>	B2	B2
<i>Agencia ELE - Video 11 - El empleo juvenil</i>	B1	B2	<i>Diverso 3/Diverso Español B Unidad 9 - Igualdad</i>	B2	B1
<i>Vitamina B1 - Unidad 8</i>	B1	B2	<i>NEEM 4 - Unidad 12 - El Greco y la pintura moderna</i>	B2	B2
<i>Vitamina B1 - Unidad 4</i>	B1	B2	<i>NEEM 4 - Unidad 10 - Absuelto por matar en defensa propia</i>	B2	B2
<i>Diverso 2 U6 / Diverso Básico U15 / Voces 3 U1 - ¿Qué haces tú por el medio ambiente?</i>	B1	B1	<i>EEM 4 Nueva edición - Unidades 11 y 12</i>	B2	B2

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025). ChatGPT (versión del 25 de enero).

La explicación proporcionada por ChatGPT para justificar esta clasificación es:

Este texto corresponde al nivel B2, ya que incluye temas más complejos, como la belleza en la edad avanzada, los estereotipos sociales y la importancia de la autorrealización. Además, el texto aborda ideas más abstractas y emplea un vocabulario variado. (OpenAI, 2025, ChatGPT, versión del 25 de enero)

Sin embargo, al analizar los ejemplos destacados por el modelo, se observa que el vocabulario, como *la belleza siempre es belleza* o *la autoestima*, no alcanza el nivel B2. Asimismo, las razones adicionales incluyen:

- Oraciones complejas con subordinadas: *que demuestra que la edad es sólo un número.*
- Uso del presente para describir hechos y acciones: *trabajan para buscar, se ven guapos.*
- Discurso indirecto e ideas abstractas: *demuestran a los más jóvenes lo maravillosa que puede ser esta etapa de la vida.*

Aunque estas características sugieren cierto análisis por parte del modelo, según el MCER (Consejo de Europa, 2002), no son suficientes para justificar la clasificación en el nivel B2. El contenido se mantiene consistente con el nivel A2, ya que la complejidad gramatical y léxica no excede las competencias previstas para este nivel.

En general, se detecta una tendencia de ChatGPT a sobreestimar los niveles, ya que asigna niveles más altos de lo que corresponde según los manuales. Esto podría deberse a que el modelo otorga un peso excesivo al tema y a su percepción subjetiva de complejidad, en lugar de basarse en los descriptores objetivos del MCER.

Tarea 3. Evaluar la distribución del vocabulario

En esta tercera tarea se analiza la capacidad de ChatGPT para clasificar vocabulario en función de los niveles del MCER (Consejo de Europa, 2002), utilizando como referencia el *Plan curricular del Instituto Cervantes* (PCIC) (Instituto Cervantes, 2006). Para ello se seleccionaron 60 palabras previamente categorizadas por nivel en tres grupos (A1-A2, B1-B2 y C1-C2), 20 por cada rango, como se recoge en la Cuadro 4.

Cuadro 4. Clasificación de 60 palabras seleccionadas de tres niveles (A1-A2, B1-B2 y C1-C2) según el *Plan curricular del Instituto Cervantes*

A1-A2	B1-B2	C1-C2
<i>Increíble, precio, entrada, antiguo, redondo, durante, mediodía, alrededor, delgado, talla, igual, creer, vale, aburrido, limpio, oscuro, dibujar, nunca, cambiar, cumpleaños</i>	<i>Frecuente, adivinar, imprescindible, sencillo, rollo, fracasar, equivocado, desastre, acceso, maduro, gritar, agridulce, luminosidad, piedra, ladrillo, proteger, curva, frágil, sobrevivir, origen</i>	<i>Surgir, frialdad, resistir, enderezar, fijo, contiguo, tambalearse, travesía, hilera, sucesivamente, lustro, conmemoración, arcaísmo, actualización, coetáneo, al instante, zanjar, emprender, sinuoso, caducado</i>

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025). ChatGPT (versión del 25 de enero).

Creamos una lista de 60 palabras distribuidas al azar y pedimos a ChatGPT clasificarlas en tres grupos según el nivel (Cuadro 5). Se procedió a ingresarlas mezcladas con el siguiente *prompt*: “Divide these words into three groups based on their levels: A1-A2, B1-B2 and C1-C2”.

equivocado, talla, vale, piedra, cambiar, frialdad, coetáneo, entrada, hilera, durante, emprender, sinuoso, frágil, caducado, luminosidad, delgado, adivinar, resistir, Increíble, zanjar, proteger, oscuro, antiguo, travesía, enderezar, nunca, origen, surgir, gritar, desastre, tambalearse, conmemoración, alrededor, lustro, sencillo, fracasar, ladrillo, aburrido, cumpleaños, frecuente, igual, arcaísmo, creer, rollo, al instante, limpio, curva, agridulce, redondo, dibujar, fijo, actualización, sobrevivir, precio, contiguo, imprescindible, maduro, acceso, sucesivamente, mediodía.

Cuadro 5. Clasificación de 60 palabras seleccionadas de tres niveles (A1-A2, B1-B2 y C1-C2) por ChatGPT

A1-A2	B1-B2	C1-C2
<i>Equivocado</i> , <i>talla, vale, piedra, cambiar, entrada, durante, delgado, adivinar, oscuro, antiguo, nunca, origen, gritar, desastre, alrededor, sencillo, cumpleaños, frecuente, igual, creer, rollo, limpio, curva, redondo, dibujar, precio, mediodía, aburrido</i>	<i>Frialdad, hiler, emprend</i> , <i>frágil, caducado, luminosidad, resistir, increíble (A1-A2), zanjar, proteger, travesía, enderezar, surgir, tambalearse, conmemoración, lustro, fracasar, agridulce, fijo, sobrevivir, contiguo, maduro, acceso</i>	<i>coetáneo</i> <i>sinuoso</i> <i>arcaísmo</i> <i>al instante</i> <i>imprescindible</i> <i>sucesivamente</i> <i>actualización</i> <i>ladrillo</i>

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025). ChatGPT (versión del 25 de enero).

La falta de información explícita sobre el equilibrio en la cantidad de palabras por nivel generó una distribución desigual. ChatGPT asignó 29 palabras al grupo A1-A2, aunque 10 de ellas —marcadas en negritas en el cuadro 5— corresponden realmente al nivel B1-B2 según el PCIC (Instituto Cervantes, 2006). Para el grupo B1-B2, el modelo asignó 23 palabras, superando las 20 originales; de estas, 14 pertenecen en realidad al nivel C1-C2, mientras que una (*increíble*) corresponde al nivel A1-A2. Finalmente, en el grupo C1-C2 ChatGPT clasificó solo ocho palabras, de las cuales dos en realidad pertenecen al nivel B1-B2.

Esta distribución desigual revela que el modelo no identifica con precisión el nivel de dificultad del léxico según los criterios establecidos por el PCIC (Instituto Cervantes, 2006). Además, al no recibir indicaciones sobre la cantidad esperada por grupo, ChatGPT toma decisiones basadas en su propio criterio probabilístico, lo cual introduce un sesgo metodológico difícil de controlar.

En resumen, aunque ChatGPT puede ofrecer agrupaciones generales del vocabulario basándose en su entrenamiento y frecuencia de uso, sus clasificaciones no se ajustan de manera rigurosa a los estándares educativos del MCER (Consejo de Europa, 2002). Este resultado sugiere que para tareas relacionadas con la planificación léxica, la elaboración de materiales didácticos o la evaluación del nivel del estudiantado, la intervención de especialistas sigue siendo indispensable.

Tarea 4. Adaptación del texto a niveles diferentes

En la cuarta tarea se examina la capacidad de ChatGPT para adaptar un mismo texto a distintos niveles de competencia lingüística según el MCER (Consejo de Europa, 2002). La actividad permite observar cómo el modelo modifica el vocabulario, las estructuras gramaticales, los conectores y el estilo discursivo para adecuarse a los niveles A1, A2 y B1 (Cuadro 6).

Para esta prueba se utilizó un texto original de 220 palabras. Se formularon tres *prompts* distintos en inglés; se mantuvo constante la instrucción base y se cambió únicamente el nivel: “Adapt this text to level B1”, “Adapt this text to level A2” y “Adapt this text to level A1”.

Este enfoque coherente permite evaluar las diferencias generadas exclusivamente por la indicación del nivel.

Cuadro 6. Resultados de la adaptación del texto original a diferentes niveles de competencia lingüística (A1, A2, B1)

Fragmento del artículo “Esta es la probabilidad de ganar la Lotería del Niño” (Borreguero, 2025). ChatGPT estima este texto como de nivel B2-C1

*Esta es la probabilidad de ganar la Lotería de El Niño **frente a** la Lotería de Navidad*

*La Lotería de El Niño lo **tiene ya todo preparado** para uno de los sorteos más esperados del año, **con permiso de** la Lotería de Navidad. Mañana **vuelven a girar** los bombos de Lotería y Apuestas del Estado en uno de los **días más observados** de este 2025. **Muchos son los que juegan esperando** recuperar **lo invertido en busca del** famoso Gordo, pero hay que saber que **en este caso hay menos dinero** en juego en el primer premio. Eso sí, **hay más premios**. Esta es **la probabilidad de ganar** la Lotería de El Niño frente a la Lotería de Navidad.*

*Si hablamos **de hacerse millonario** las probabilidades de que **toque el** premio Gordo son de 1 entre 100 000. **Sin embargo, en lo que destaca** este sorteo es en las opciones de ganar algo. **Cerca de un 15% de los décimos** comprados **reciben** algún premio. **Incluidos** los reintegros, las posibilidades de que **te toque** algo es de 1 entre 7. Con la Lotería de El Niño las posibilidades de ganar el primer premio **son las mismas**, pero las de ganar **algo aumentan** a un 38% por número. En **esta ocasión** reparte menos dinero, pero hay más posibilidad de que **te toque algún pellizco**.*

B1

Probabilidad de ganar en la Lotería de El Niño **comparada con** la Lotería de Navidad

La Lotería de El Niño **ya está lista** para uno de los sorteos más esperados del año, **después de la famosa** Lotería de Navidad. Mañana los bombos de Lotería y Apuestas del Estado **girarán de nuevo** en uno de **los días más importantes** de este 2025. **Muchas personas juegan con la esperanza de recuperar lo que gastaron o de ganar el famoso “Gordo”**. Sin embargo, en este sorteo el primer premio **tiene menos dinero**. Pero, eso sí, **hay más premios en total**. Aquí **está la comparación** entre la Lotería de El Niño y la Lotería de Navidad.

Si hablamos **de ganar mucho dinero**, la probabilidad de **conseguir** el premio “Gordo” es de 1 entre 100 000. **Pero lo especial** de este sorteo es que hay más opciones de ganar algo. **Alrededor del 15% de los boletos ganan** algún premio. **Si contamos** también los reintegros, la probabilidad de que **ganes** algo es de 1 entre 7. Con la Lotería de El Niño, las probabilidades de ganar el primer premio **son iguales** a las de la Lotería de Navidad, pero las de ganar **cualquier premio suben** al 38% por número. **Este sorteo reparte menos dinero en total**, pero hay más posibilidades de **ganar un pequeño premio**.

A2

La probabilidad de ganar la Lotería de El Niño **y** la Lotería de Navidad

La Lotería de El Niño **está lista para** uno de los sorteos **más importantes** del año. **Es muy famosa**, como la Lotería de Navidad. Mañana los bombos de la lotería **van a girar otra vez**. **Muchas personas juegan para intentar ganar** el famoso “Gordo”. Pero en esta lotería, el **primer premio tiene menos dinero**. Aunque hay más premios **en total**.

Si hablamos de ganar mucho dinero, la probabilidad de ganar el “Gordo” es de 1 entre 100 000. **Pero esta lotería tiene** más opciones para ganar algo. Cerca del 15% de los boletos **tienen** algún premio. **Con los reintegros**, la probabilidad de **ganar** algo es de 1 entre 7. En la Lotería de El Niño, las posibilidades de ganar el primer premio **son iguales** a las de la Lotería de Navidad. Pero **es más fácil ganar** cualquier premio: 38% por número. **Hay menos dinero en total**, pero más posibilidades de **ganar un premio pequeño**.

A1

La probabilidad de ganar en la Lotería de El Niño

La Lotería de El Niño es un sorteo muy famoso, como la Lotería de Navidad. Mañana es el sorteo. Muchas personas juegan para ganar dinero o el premio grande, el “Gordo”. En esta lotería, el primer premio tiene menos dinero, pero hay más premios pequeños.

Es difícil ganar el “Gordo”. La probabilidad es 1 entre 100 000. Pero con esta lotería es más fácil ganar un premio. El 15% de los boletos tienen premio. Con los reintegros, la probabilidad de ganar algo es 1 entre 7. En la Lotería de El Niño hay más posibilidades de ganar un premio pequeño: 38% por número.

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025).
ChatGPT (versión del 27 de enero).

El texto adaptado a B1 presenta oraciones más simples, como el uso de *pero* en lugar de *sin embargo*, y frases como *muchas personas juegan* en vez de *muchos son los que juegan*. El vocabulario se simplifica: *tocar* se sustituye por *ganar*, *lo invertido* por *lo que gastaron* y *aumentar* por *subir*. Sin embargo, se identifican cambios innecesarios, como añadir *en total a reparte menos dinero*, o sustituir *cerca de* por *alrededor de*, que no aportan mayor significado ni alteran el nivel del texto.

La versión A2, con 167 palabras frente a las 220 originales, simplifica vocabulario, conectores y estructuras. Por ejemplo, *hacerse millonario* cambia a *ganar mucho dinero*, y *el sorteo más esperado* se convierte en *el sorteo más importante*. Expresiones como *juegan esperando* se adaptan a *juegan para intentar ganar*. Además, frases como *vuelven a girar* se sustituyen por *van a girar otra vez*, manteniendo un enfoque sencillo.

La versión A1 se reduce a 112 palabras. El contenido incluye verbos básicos como *ser*, *haber*, *tener*, *jugar* y *ganar*, con repeticiones frecuentes (*ganar* y *premio* aparecen seis veces). Se añaden explicaciones simples, como aclarar “Gordo” como *el premio grande*. Las construcciones son sencillas (*es fácil*, *es difícil*), aunque algunos términos como *posibilidad* o *reintegros* podrían ser complicados para este nivel.

En términos generales, ChatGPT demuestra cierta capacidad para adaptar el texto a distintos niveles, realizando ajustes en extensión, léxico y complejidad gramatical. Asimismo, añade explicaciones culturales útiles en los niveles iniciales. No obstante, la adaptación no siempre responde a criterios consistentes: se detecta el uso de sinónimos innecesarios que no simplifican el texto, así como

la permanencia de palabras o expresiones que podrían considerarse demasiado avanzadas para el nivel objetivo.

En conclusión, aunque el modelo puede generar versiones diferenciadas por nivel, sus adaptaciones no siempre se alinean plenamente con los descriptores del MCER (Consejo de Europa, 2002). Estas inconsistencias refuerzan la necesidad de una revisión experta antes de utilizar estos textos con fines pedagógicos, especialmente en niveles básicos.

Tarea 5. Preguntas para preparación de DELE

El objetivo de esta tarea es diseñar y evaluar preguntas para la práctica oral en la preparación del examen DELE en tres diferentes niveles (A1, B2, C1), ya que entre estos existen diferencias fundamentales en las habilidades necesarias para superar el examen.

En cuanto al nivel A1, la persona usuaria definida por el MCER como básica en las habilidades orales debe contestar preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales (Consejo de Europa, 2002, p. 30). Por otro lado, la persona usuaria independiente de nivel B2 ya debe participar activamente en debates desarrollados en situaciones cotidianas, así como explicar y defender sus puntos de vista. Es decir, ambos niveles se centran en temas personales. Aunque en B2 la variedad de estos temas es más compleja, pueden estar relacionados con la especialidad de la persona usuaria, y es imprescindible saber explicar ventajas e inconvenientes de varias opciones (Consejo de Europa, 2002, p. 31). En lo que se refiere al nivel C1, se espera que la persona usuaria competente utilice el lenguaje para fines sociales y profesionales, que incluya descripciones claras y detalladas sobre temas complejos que incorporen otros temas, que desarrolle ideas concretas y que elabore una conclusión apropiada (Consejo de Europa, 2002, p. 31).

Así, se solicitó a ChatGPT crear cinco preguntas para la preparación de cada uno de estos niveles (Cuadro 7). El *prompt* utilizado fue: “Create 5 oral practice questions for the DELE exam at level A1 / B2 / C1”. Este comando fue adaptado únicamente en lo relativo al nivel con el fin de garantizar coherencia metodológica y permitir la comparación entre resultados.

Cuadro 7. Preguntas de práctica oral creadas por ChatGPT para la preparación del examen DELE: niveles A1, B2 y C1

A1	B2	C1
<p>1. Sobre ti</p> <p><i>¿Cómo te llamas y de dónde eres?</i></p> <p><i>¿Dónde vives ahora y con quién?</i></p> <p>2. Rutina diaria</p> <p><i>¿Qué haces un día normal por la mañana?</i></p> <p><i>¿A qué hora desayunas y qué sueles comer?</i></p> <p>3. Gustos y preferencias</p> <p><i>¿Qué te gusta hacer en tu tiempo libre?</i></p> <p><i>¿Preferes salir con amigos o quedarte en casa? ¿Por qué?</i></p> <p>4. Compras y lugares</p> <p><i>¿Qué tiendas hay cerca de tu casa?</i></p> <p><i>¿Dónde compras normalmente la comida?</i></p> <p>5. Planes futuros</p> <p><i>¿Qué vas a hacer este fin de semana?</i></p> <p><i>¿Te gustaría viajar a otro país? ¿Cuál y por qué?</i></p>	<p>1. Experiencias y desarrollo personal</p> <p><i>¿Cómo te han ayudado tus experiencias personales o profesionales a crecer como persona?</i></p> <p><i>¿Qué desafíos has enfrentado recientemente y cómo los superaste?</i></p> <p>2. Opiniones y temas sociales</p> <p><i>¿Qué opinas sobre el impacto de las redes sociales en la sociedad actual?</i></p> <p><i>¿Crees que la igualdad de género se ha logrado en tu país? ¿Por qué sí o no?</i></p> <p>3. Educación y trabajo</p> <p><i>¿Qué importancia tiene la formación continua en el ámbito profesional?</i></p> <p><i>¿Cómo debería cambiar el sistema educativo para adaptarse a las necesidades del mundo moderno?</i></p> <p>4. Viajes y globalización</p> <p><i>¿Cómo crees que los viajes internacionales influyen en la mentalidad de las personas?</i></p> <p><i>¿Piensas que la globalización tiene más ventajas o desventajas? Justifica tu respuesta.</i></p>	<p>1. Desarrollo personal y sociedad</p> <p><i>¿Crees que el éxito personal está más relacionado con el esfuerzo individual o con las oportunidades externas? ¿Por qué?</i></p> <p><i>¿Qué papel desempeña la educación en la construcción de una sociedad más justa?</i></p> <p>2. Problemas globales y medio ambiente</p> <p><i>¿Hasta qué punto crees que los individuos pueden contribuir a solucionar problemas medioambientales?</i></p> <p><i>¿Cómo afecta el cambio climático a las relaciones entre países?</i></p> <p>3. Cultura y globalización</p> <p><i>¿Consideras que la globalización beneficia o amenaza la diversidad cultural? Justifica tu respuesta.</i></p> <p><i>¿Qué papel juegan las tradiciones en un mundo cada vez más conectado?</i></p> <p>4. Tecnología y futuro</p> <p><i>¿Crees que los avances tecnológicos están cambiando nuestra manera de relacionarnos con los demás?</i></p> <p><i>¿Es positivo o negativo?</i></p> <p><i>¿Cómo imaginas que será el mercado laboral en 20 años con la automatización?</i></p>

A1	B2	C1
	5. Futuro y tecnología <i>¿Cómo imaginas que será la vida dentro de 20 años gracias a los avances tecnológicos?</i> <i>¿Crees que la inteligencia artificial representa más oportunidades o riesgos para la humanidad?</i>	5. Opiniones sobre temas actuales <i>¿Cómo impactan las fake news en la sociedad y qué medidas podrían tomarse para reducir su influencia?</i> <i>¿Qué importancia tiene el pensamiento crítico en un mundo lleno de información?</i>

Fuente: elaboración propia a partir de la información obtenida en OpenAI. (2025). ChatGPT (versión del 27 de enero).

Para cada nivel, ChatGPT destaca por sí mismo cinco temas y formula dos preguntas para cada uno. Además, el programa proporciona consejos útiles para la parte oral de cada nivel, los cuales también analizaremos de manera sucesiva.

En el nivel A1 se observa que todos los temas formulados son básicos, centrados en la rutina diaria y los gustos personales. Para responder a la mayoría de las preguntas se puede utilizar el presente, salvo en las últimas, donde ChatGPT sugiere trabajar con estructuras como *me gustaría* e *ir a + infinitivo*, que también están incluidas en este nivel. En cuanto a los consejos, ChatGPT recomienda el uso de conectores como *pero* y *porque*, y enfatiza la importancia de responder con frases completas.

En el nivel B2 se aprecia que los temas abarcan tanto ámbitos personales como contextos sociales, con la necesidad de formular opiniones personales. Los temas propuestos incluyen tecnología, educación, vida laboral, entre otros, que son muy comunes en la parte oral del DELE. Además, hay preguntas diseñadas para reflexionar sobre ventajas y desventajas, algo que el MCER (Consejo de Europa, 2002) recomienda trabajar en este nivel.

En el nivel C1 algunos temas coinciden con los de B2, como el futuro y la tecnología, la globalización o el desarrollo personal. Sin embargo, hay una clara diferencia en la profundidad de las preguntas asignadas a cada tema. Por ejemplo, en desarrollo personal, mientras que en B2 se pide hablar sobre experiencias personales, en C1 se requiere reflexionar de manera global. Aunque la mayoría de las preguntas en C1 solicitan expresar opiniones, es imprescindible contar con conocimientos sobre temas de actualidad, como la vida social, la economía, los problemas ecológicos y otras tendencias mundiales. En cuanto a las recomendaciones, para B2 ChatGPT sugiere usar conectores como *por lo tanto*, *sin embargo*, *de hecho*, *aunque*, *además*, *en resumen*, *por otro lado*, *desde mi punto de vista*. Por otra

parte, en C1 se recomienda emplear expresiones idiomáticas, una mayor variedad de tiempos verbales, el subjuntivo, el condicional, entre otros recursos avanzados.

En resumen, ChatGPT no solo formula preguntas adecuadas según el nivel, sino que también selecciona temas frecuentes en la parte oral del DELE y ofrece consejos precisos para superar el examen en cada nivel.

CONCLUSIONES

El presente estudio analizó la capacidad de ChatGPT para clasificar y generar contenido en español de acuerdo con los niveles del Marco Común Europeo de Referencia (MCER) (Consejo de Europa, 2002). A través de cinco tareas clave, se exploraron aspectos como la generación de programas educativos, la clasificación de materiales audiovisuales, la distribución del vocabulario, la adaptación de textos y la creación de preguntas para exámenes DELE. Los resultados revelan tanto las fortalezas como las limitaciones de esta herramienta, además de ofrecer perspectivas para su uso e integración en el aprendizaje y la enseñanza de idiomas.

Fortalezas

- **Comprensión general de los niveles:** ChatGPT demuestra una capacidad básica para reconocer y aplicar los descriptores de competencias lingüísticas del MCER, especialmente en la creación de preguntas y adaptación de textos.
- **Propuesta de temas relevantes:** los temas generados suelen alinearse con las necesidades comunicativas y los objetivos de cada nivel, y resultan útiles para actividades de aprendizaje y preparación de exámenes como el DELE.
- **Habilidad para simplificar contenido:** el modelo adapta estructuras gramaticales y vocabulario según los niveles, lo cual facilita contar con materiales más accesibles para niveles básicos o intermedios.
- **Recomendaciones prácticas:** ofrece consejos útiles para la práctica oral, como el uso de conectores, expresiones idiomáticas y variedad de tiempos verbales, lo que puede enriquecer las interacciones comunicativas del estudiantado.

Limitaciones

- **Inexactitud en la clasificación:** ChatGPT tiende a sobreestimar los niveles de dificultad, especialmente en materiales audiovisuales y vocabulario, y a mostrar inconsistencias frente a los estándares educativos del MCER.
- **Falta de profundidad en los programas educativos:** la estructura uniforme de los programas propuestos para diferentes niveles no refleja la carga horaria ni las necesidades específicas de niveles más avanzados.
- **Cambios innecesarios en la adaptación de textos:** aunque el modelo puede simplificar textos, a veces realiza modificaciones redundantes o conserva palabras complejas que no corresponden al nivel objetivo.
- **Dependencia de la supervisión humana:** los resultados evidencian que, sin intervención de personas expertas, la información proporcionada por ChatGPT puede no ser completamente fiable para el diseño de materiales didácticos.

Perspectivas futuras

- ChatGPT puede ser una herramienta valiosa para complementar el aprendizaje, pero requiere mejoras metodológicas para una clasificación más precisa de vocabulario y materiales según estándares como el MCER.
- Es necesario un diseño más detallado de programas educativos, adaptado a las necesidades reales de estudiantes y a la carga horaria correspondiente para cada nivel.
- Para maximizar su utilidad, sería ideal combinar ChatGPT con supervisión docente y recursos didácticos diseñados por expertos en enseñanza de idiomas.

En definitiva, ChatGPT representa una innovación prometedora en el ámbito de la enseñanza de lenguas, pero su uso debe entenderse como complementario y no sustitutivo del trabajo docente. A medida que se desarrollen mejoras en sus algoritmos y capacidades de comprensión pedagógica, el modelo podría consolidarse como un recurso valioso para apoyar el aprendizaje personalizado, la producción

de materiales y la reflexión sobre niveles de competencia lingüística en el marco del MCER.

Los hallazgos de este estudio permiten delimitar implicaciones específicas para el ámbito educativo en el contexto de la enseñanza de lenguas asistida por inteligencia artificial. Para el profesorado, ChatGPT puede funcionar como herramienta auxiliar en etapas iniciales de planificación curricular y en la generación de materiales de práctica, particularmente en las competencias oral y escrita. No obstante, su implementación requiere una validación experta que asegure la pertinencia de los contenidos generados en relación con los descriptores del MCER y los objetivos pedagógicos establecidos. En el caso del autoaprendizaje, la accesibilidad y diversidad temática de los insumos producidos por ChatGPT pueden facilitar la exposición continua a la lengua meta; sin embargo, su uso autónomo exige competencias metacognitivas por parte del estudiantado, así como una capacidad crítica para detectar posibles inconsistencias, omisiones o desviaciones metodológicas. Estas observaciones refuerzan la necesidad de integrar su uso dentro de marcos pedagógicos estructurados, donde la intervención docente actúe como garante de calidad y coherencia formativa.

BIBLIOGRAFÍA

- Alemán Ariza, E., Gómez Miranda, O. y Prieto Montero, A. (2024). ChatGPT y su impacto en el entorno educativo. *Revista Saberes APUDEP*, 7(2), 222-245.
- Borreguero, M. (5 de enero de 2025). Esta es la probabilidad de ganar la Lotería de El Niño frente a la Lotería de Navidad. *20 minutos*. <https://www.20minutos.es/lainformacion/economia-y-finanzas/esta-probabilidad-ganar-loteria-ni-no-5667326/>
- Bozkurt, A. (2024). Genai et al.: Cocreación, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis*, 16(1), 1-10.
- Chicaíza Chicaíza, R. M., Camacho Castillo, L. A., Ghose, G., Castro Magayanes, I. E. y Gallo Fonseca, V. T. (2023). Aplicaciones de ChatGPT como inteligencia artificial para el aprendizaje de idioma inglés: Avances, desafíos y perspectivas futuras. *Latam: Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4(2), 1-13.
- Coldewey, D. (26 de abril de 2023). ChatGPT Lies in Some Languages More than Others. *TechCrunch*. <https://techcrunch.com/2023/04/26/why-chatgpt-lies-in-some-languages-more-than-others/>
- Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Instituto Cervantes; Ministerio de Educa-

ción, Cultura y Deporte; Anaya. https://cvc.cervantes.es/ENSEÑANZA/biblioteca_ele/marco/cvc_mer.pdf

- De la Torre García, N. (2016). La complejidad lingüística en los niveles de competencia del MCER: el caso de la variedad verbal en la expresión escrita en ELE. *Revista Nebrija de Lingüística Aplicada*, (20).
- Hernández, R. J. (2024). Optimizando la efectividad de la retroalimentación de ChatGPT de las producciones escritas de estudiantes ESL: un enfoque desde el docente programador. *YUYAY: Estrategias, Metodologías & Didácticas Educativas*, 3(2), 50–61. <https://doi.org/10.59343/yuyay.v3i2.69>
- Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes*. Instituto Cervantes. https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm
- Instituto Cervantes. (s. f.). *Información sobre el programa académico y niveles del Instituto Cervantes de Leeds*. Recuperado el 20 de enero de 2025 de https://leeds.cervantes.es/es/cursos_espanol/estudiantes_espanol/cursos_generales.htm
- Joshi, S., Ha, E., Rivera, Y. y Singh, V. K. (2024). ChatGPT and Vaccine Hesitancy: A Comparison of English, Spanish, and French Responses Using a Validated Scale. *AMIA Joint Summits on Translational Science Proceedings, 2024*, 266-275.
- Kohnke, L., Moorhouse, B. L. y Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Laverde Gonzalez, J. S. y Rodríguez Hernández, L. A. (2025). AI and Education: Combination to Enhance Knowledge. *Seminars in Medical Writing and Education*, 4(0).
- Montoya, C. (2024). Evaluación transmedia en educación en tiempos de la IA y el ChatGPT. *Cuadernos del Centro de Estudios en Diseño y Comunicación. Ensayos*, (224), 51-63.
- Moreno Padilla, R. D. (2019). La llegada de la inteligencia artificial a la educación. *Revista de Investigación en Tecnologías de la Información*, 7(14), 260-270. <https://doi.org/10.36825/riti.07.14.022>
- OpenAI. (2025). ChatGPT (versión del 20 de enero) [Modelo de lenguaje de gran tamaño]. <https://chat.openai.com/chat>
- OpenAI. (2025). ChatGPT (versión del 25 de enero) [Modelo de lenguaje de gran tamaño]. <https://chat.openai.com/chat>
- OpenAI. (2025). ChatGPT (versión del 27 de enero) [Modelo de lenguaje de gran tamaño]. <https://chat.openai.com/chat>
- Orozco Castaño, C. A. (2022). Sobre los nuevos retos y oportunidades derivados de la implementación de herramientas de inteligencia artificial en la docencia universitaria. *Investigaciones Andina*, 24(45).

- Pineda Sánchez, C., Smith Batson, M. de la C. y Rondón Sánchez, K. (2024). Aplicación del Marco Común Europeo de Referencia en contextos educativos: percepciones docentes. *Didascalía: Didáctica y Educación*, 15(3), 1-15.
- Rivera-Rosas, C. N., Tadeo Calleja-López, J. R., Ruibal-Tavares, E., Villanueva-Neri, A., Flores-Felix, C. M. y Trujillo-López, S. (2024). Exploring the Potential of ChatGPT to Create Multiple-Choice Question Exams. *Educación Médica*, 25(4).
- Seghier, M. L. (7 de marzo de 2023). ChatGPT: Not All Languages Are Equal. *Nature*. <https://www.nature.com/articles/d41586-023-00680-3>
- Shaikh, S., Yayilgan, S. Y., Klimova, B. y Pikhart, M. (2023). Assessing the Usability of ChatGPT for Formal English Language Learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9), 1937-1960. <https://doi.org/10.3390/ejihpe13090140>
- VanderLinde, G. y Mera Cury, T. (2024). El uso de inteligencia artificial y sus desafíos para la evaluación académica: Una revisión de la literatura. *Cuaderno de Pedagogía Universitaria*, 21(41), 126-137.

