

Uso de herramientas de inteligencia artificial en la enseñanza del español como segunda lengua

Xi Wang,
Jinan University, China

Ismael Cardeñosa García,
Jinan University, China

Abstract: *Amidst the growing integration of AI in education, this study examines the use of specific tools like Mizou.com for conversational practice and Grok for assessment assistance in teaching Spanish as a second language. It explores how these tools enhance personalization, accessibility, and efficiency in language education while addressing limitations such as cultural nuances and emotional engagement. Initial results from a 22-student implementation show significant time savings (54.09%–61.82%) in assessment and improved data-driven instruction. The findings suggest that AI fosters inclusive learning environments aligned with Universal Design for Learning principles, though human supervision remains crucial. Future research should explore long-term impacts and cultural competence in AI systems.*

Keywords: Artificial Intelligence, Language Education, Spanish Teaching, Chatbots, Automated Assessment, Inclusive Learning.

Resumen: *En medio de la creciente integración de la inteligencia artificial en la educación, este estudio examina el uso de herramientas específicas como Mizou.com para la práctica conversacional y Grok para el apoyo en la evaluación en la enseñanza del español como segunda lengua. Se analiza cómo estas herramientas mejoran la personalización, la accesibilidad y la eficiencia en la educación lingüística, al tiempo que se abordan limitaciones como los matices culturales y el compromiso emocional. Los resultados iniciales de una implementación con 22 estudiantes muestran un ahorro significativo de tiempo (54,09 %–61,82 %) en la evaluación y una mejora en la instrucción basada en datos. Los hallazgos sugieren que la IA fomenta entornos de aprendizaje inclusivos alineados con los principios del Diseño Universal para el Aprendizaje, aunque la supervisión humana sigue siendo crucial. Investigaciones futuras deberían explorar los impactos a largo plazo y la competencia cultural en los sistemas de IA.*

Palabras clave: Inteligencia artificial, Enseñanza de idiomas, Chatbots, Evaluación automatizada, Aprendizaje inclusivo.

1. Introducción

En la actual era digital, la inteligencia artificial (IA) se está imponiendo con mano de hierro transformando diversos ámbitos, incluyendo la educación. La integración de la IA en el ámbito educativo está marcando una profunda transformación, redefiniendo los enfoques pedagógicos tradicionales y permitiendo experiencias de aprendizaje más personalizadas y eficientes (Ayeni et al., 2024). A nivel global, la IA se ha convertido en un componente estratégico en el desarrollo educativo, ofreciendo un abanico sin precedentes de oportunidades para la personalización y la accesibilidad en el aprendizaje, fomentando la autonomía del estudiante (Limna et al., 2022 y Reinders & White (2016). Como señalan Knoth et al. (2024), ‘Artificial intelligence technologies are rapidly advancing. As part of this development, large language models (LLMs) are increasingly being used when humans interact with systems based on artificial intelligence (AI), posing both new opportunities and challenges’. Según Chapelle (2003), la tecnología siempre ha revolucionado la forma en que se aprenden los idiomas, proporcionando nuevas herramientas que permiten un aprendizaje más interactivo, adaptándose a las necesidades individuales.

Sin embargo, junto con estos beneficios, la IA también presenta desafíos significativos. Uno de los principales escollos de las IA es su capacidad limitada para captar la complejidad cultural y contextual del idioma. Godwin-Jones (2017) argumenta que los sistemas basados en IA son eficaces para tareas como la traducción literal y la corrección gramatical, pero a menudo fallan en comprender los matices culturales y los contextos, que son cruciales a la hora de alcanzar una comunicación completa. Otro aspecto por considerar, como menciona Petersen & Markiewicz (2008), es la falta de conexión emocional y de motivación en las interacciones con IA, lo que puede afectar negativamente al compromiso a largo plazo de los estudiantes.

Además, Gray et al. (2024) señalan que ‘Assessment is a crucial part of education. Traditional marking is a source of inconsistencies and unconscious bias, placing a high cognitive load on the assessors’, lo que subraya la necesidad de herramientas de evaluación automatizadas como las que ofrece la IA.

Este artículo explora el uso de herramientas basadas en IA en la enseñanza del español como segunda lengua (L2), centrándose en la plataforma mizou.com (Mizou) para interacciones conversacionales con los estudiantes, y empleando posteriormente Grok como asistente en la evaluación. Mizou permite a los profesores crear chats de IA personalizados que simulan conversaciones reales, proporcionando a los estudiantes una oportunidad para practicar el idioma de manera interactiva y contextualizada, dentro de unos parámetros definidos previamente por el profesor. Por otro lado, Grok asiste en la evaluación de las conversaciones que, gracias a una base de datos previamente cargada, analiza parámetros como semántica, coordinación de género y número, y coherencia, lo que reduce significativamente el tiempo de corrección manual. A través de un enfoque práctico, se examinan tanto los beneficios como los desafíos de integrar estas tecnologías en el aula, con el objetivo de maximizar su efectividad, valorando su implementación en el aprendizaje de idiomas.

2. Base teórica

La personalización es uno de los aspectos más destacados de la IA en la educación lingüística. Plataformas como Duolingo y Babbel pueden adaptar el contenido a las fortalezas y debilidades de cada estudiante. Según un estudio de Vesselinov &

Grego (2012), los usuarios de Duolingo mostraron mejoras significativas en sus habilidades lingüísticas tras un uso regular de esta aplicación. De manera similar, Mizou permite a los profesores configurar chats de IA de forma personalizada, para que se ajusten a los objetivos específicos de cada tarea, proporcionando a los estudiantes práctica dirigida en áreas donde necesitan mejorar, teniendo posteriormente acceso para analizar la conversación entre el alumno y el chat IA. Por ejemplo, un profesor puede programar un chat para enfocarse en el uso correcto del subjuntivo, programando el chat para que guíe al estudiante a través de escenarios conversacionales indicados para tal fin. Alier et al. (2024) destacan que la IA permite un aprendizaje personalizado que facilita la atención a la diversidad, adaptándose a diferentes estilos y ritmos de aprendizaje. Esta capacidad de adaptación es especialmente valiosa en aulas con estudiantes de diversos niveles de idioma, como en el caso de la enseñanza del español como L2 (Sáenz Valiente, 2024). Además, Song et al. (2024) señalan que ‘Universal Design for Learning (UDL) provides guidelines for making learning more inclusive across disciplines’, lo que respalda la idea de que la IA puede ser utilizada para crear entornos de aprendizaje más inclusivos y accesibles.

La IA mejora la accesibilidad y flexibilidad del aprendizaje de idiomas, permitiendo a los estudiantes practicar en cualquier momento y lugar. Reinders & White (2016) enfatizan que la tecnología fomenta un aprendizaje autónomo y continuo, esencial en la sociedad actual. En el contexto de Mizou, los alumnos acceden a chats de IA a través de enlaces proporcionados por el profesor, lo que les permite interactuar con la tecnología de manera independiente. La IA está abriendo nuevas oportunidades para la enseñanza de idiomas, incluyendo por ejemplo las lenguas de signos, al proporcionar herramientas que hacen del aprendizaje un proceso más accesible. Este enfoque inclusivo resalta el potencial de la IA para democratizar la educación lingüística. Además, Kajiwarra & Kawabata (2024) definen la alfabetización en IA como ‘a set of competencies that enables individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool online, at home, and in the workplace’, lo que proporciona un marco teórico para comprender las competencias necesarias para aprovechar estas herramientas de manera efectiva.

Otra ventaja clave que ha logrado la IA actual es su capacidad de proporcionar retroalimentación instantánea. Los sistemas de IA pueden evaluar respuestas de los estudiantes en tiempo real, corrigiendo errores gramaticales, de pronunciación o de vocabulario al instante. Esto permite a los estudiantes aprender de sus errores en el momento, acelerando el proceso de aprendizaje. Herramientas como Grammarly para escritura, y Rosetta Stone para pronunciación, utilizan IA para ofrecer correcciones precisas y sugerencias de mejora en tiempo real. En el caso de Mizou, los chats de IA proporcionan respuestas que guían al estudiante hacia un uso correcto del idioma, simulando una interacción con un hablante nativo. Crompton et al. (2024) destaca que la retroalimentación inmediata de la IA es una de sus principales fortalezas en la enseñanza de los idiomas, tomando como ejemplo el del inglés, y pudiéndose trasladar a cualquier otro. Esta capacidad mejora la confianza del estudiante y refuerza el aprendizaje, incrementando la eficacia.

La evaluación automatizada es otro ámbito donde la IA está transformando la educación. Herramientas como Grok, utilizadas en este estudio, analizan textos (o diálogos como el caso de este trabajo) basados en parámetros predefinidos, proporcionando una primera capa de corrección que los profesores pueden refinar. Estudios como el de Mizumoto & Eguchi (2023) han demostrado que los sistemas de puntuación automatizada pueden evaluar ensayos con un alto grado de fiabilidad, especialmente en aspectos como la gramática y la coherencia. Estas técnicas de

evaluación basadas en IA pueden identificar patrones en las respuestas de los estudiantes, ayudando a los profesores a detectar áreas de mejora (Martínez-Comesaña et al., 2023). En este estudio, Grok redujo el tiempo de corrección en un 30-40%, permitiendo al profesor centrarse en proporcionar feedback cualitativo. Además, Lee et al. (2024) destacan que 'LLMs have shown significant advancements in educational assessment and automatic scoring... demonstrating efficiency in adaptive learning scenarios, accurately score complex cognitive tasks, and hold the potential to transform educational practices', reforzando la importancia de la IA en la mejora de los procesos de evaluación.

No todo es gramática y coherencia; a pesar de sus avances, la IA sigue teniendo dificultades para comprender y transmitir la riqueza cultural y contextual de un idioma, especialmente matices culturales, el humor, las expresiones idiomáticas y los registros de formalidad, que requieren una comprensión profunda que las máquinas no poseen completamente. Godwin- Jones (2017) señala que esta limitación puede llevar a traducciones incorrectas o interpretaciones erróneas, lo que podría confundir a los estudiantes o proporcionarles información inexacta. En la enseñanza del idioma español, donde las diferencias regionales (por ejemplo, entre el español de México y el de España) son significativas, esta limitación es particularmente relevante. Aunque la IA puede facilitar la comunicación en tiempo real, no sustituye la necesidad de entender las sutilezas culturales que enriquecen el aprendizaje de idiomas (Stollhans, 2024).

Además de captar los matices culturales, la intervención humana es fundamental para mantener la motivación y el compromiso en el aprendizaje de idiomas. Los profesores pueden inspirar, alentar y adaptar su enseñanza según las necesidades de los estudiantes, algo que la IA no puede replicar. Argumentan que la falta de empatía y conexión personal en las plataformas de IA puede resultar en una experiencia de aprendizaje menos enriquecedora y potencialmente desmotivadora para algunos estudiantes (Petersen & Markiewicz, 2008). Un estudio reciente sobre chatbots en la educación lingüística encontró que, aunque son útiles para la práctica, carecen de un entorno emocional que fomente la motivación a largo plazo (Xiao et al., 2024). En el método descrito, la supervisión del profesor mitiga esta limitación al proporcionar feedback personalizado y apoyo emocional tras las interacciones con la IA.

El enfoque híbrido presentado en este trabajo combina la eficiencia de la IA con la supervisión humana posterior, abordando parcialmente las limitaciones de comprensión cultural y conexión emocional. Lyu (2024) encontró que los chatbots mejoran significativamente las habilidades lingüísticas, especialmente en vocabulario y pronunciación, lo que respalda el uso de Mizou. Además, la supervisión posterior del profesor asegura que los estudiantes reciban feedback relevante, facilitando la localización de los errores para después comentarlos en clase.

La integración de la IA en la enseñanza de idiomas debe ser cuidadosamente planificada para maximizar sus beneficios y minimizar sus limitaciones. La IA generativa ofrece innumerables oportunidades para profundizar en el aprendizaje de idiomas, pero debe usarse de forma crítica (Stollhans, 2024). La implementación exitosa de la IA requiere que los profesores estén capacitados para utilizar estas herramientas y entender sus limitaciones. Thoppilan et al. (2022) destaca la importancia de la formación docente para integrar la IA de manera efectiva en el aula. En este estudio, el profesor diseñó y supervisó los chats de Mizou, asegurando que se alinearan con los objetivos pedagógicos. Además, es crucial considerar las implicaciones éticas de la IA, como la privacidad de los datos de los estudiantes y el

riesgo de sesgos en los algoritmos. La equidad en el acceso a la tecnología también debe abordarse para evitar ampliar las brechas educativas.

La IA ofrece un gran potencial en la enseñanza de idiomas, pero su implementación debe ser complementaria a los métodos tradicionales. Los profesores deben crear entornos de aprendizaje más efectivos y enriquecedores, preparando a los estudiantes para comunicarse competentemente en un mundo globalizado.

3. Motivos

La integración de la IA en la enseñanza del español como L2 responde a una serie de motivos fundamentales, que abordan tanto desafíos educativos actuales como aprovechar avances tecnológicos, mejorar el compromiso estudiantil, apoyar a los docentes y contribuir al avance de la investigación académica. Este apartado detalla estas razones, respaldadas por evidencia empírica y contextualizadas en el marco de la enseñanza del español.

El español es uno de los idiomas más hablados y estudiados a nivel global, con casi 500 millones de hablantes nativos y más de 600 millones de hablantes totales, incluyendo aprendices (Instituto Cervantes, 2024). En 2024, solo en Estados Unidos, había alrededor de 8,6 millones de estudiantes de español, seguido por Brasil con 4,05 millones (Instituto Cervantes, 2024). Esta creciente demanda, impulsada por la globalización, el comercio internacional y la riqueza cultural de los países hispanohablantes, requiere métodos de enseñanza escalables y efectivos. Los métodos tradicionales de enseñanza de idiomas enfrentan limitaciones significativas, como numerosas clases, lo que dificulta la atención individualizada y la falta de oportunidades para practicar la conversación fuera del aula. Las herramientas de IA, como Mizou, permiten a los estudiantes practicar de manera autónoma a través de chats personalizados que simulan conversaciones reales, ofreciendo una solución práctica a estas limitaciones. Este enfoque es flexible en múltiples ámbitos, pues se adapta a las necesidades específicas de cada estudiante, garantizando un aprendizaje más inclusivo y efectivo.

Los rápidos avances en la IA han transformado el panorama educativo, ofreciendo herramientas innovadoras que mejoran la enseñanza de idiomas. Estas plataformas utilizan algoritmos de aprendizaje automático para crear experiencias inmersivas en conversaciones personalizadas, mientras que otras herramientas como los chats generativos tipo Grok pueden automatizar la evaluación de textos, analizando aspectos como la semántica y la coherencia. Estas tecnologías permiten a los estudiantes recibir retroalimentación inmediata sobre gramática, vocabulario y pronunciación, elementos esenciales para el dominio de un idioma, pero siempre mejor bajo la supervisión de un profesor, porque puede haber aspectos que necesiten de una reedición. Un meta-análisis reciente encontró que la IA tiene un efecto alto ($g = 0,897$) en el logro del aprendizaje de idiomas, lo que indica que los estudiantes que utilizan herramientas de IA superan significativamente a aquellos que siguen pedagogías tradicionales (Kim & Park, 2023). Aunque este estudio se centró en el inglés, los principios de personalización y retroalimentación inmediata son igualmente aplicables al español, especialmente considerando la complejidad gramatical del idioma, como las conjugaciones verbales y las reglas de concordancia de género y número.

La IA no solo beneficia a los estudiantes, sino que también alivia la carga administrativa de los docentes. Herramientas como Grok pueden ayudar a automatizar parte de la evaluación, realizando una primera revisión, analizando

parámetros predefinidos y reduciendo el tiempo dedicado a la corrección manual. Esto permite a los docentes centrarse en proporcionar retroalimentación cualitativa y diseñar estrategias pedagógicas más efectivas. En un mundo donde la IA está cada vez más presente, desarrollar la alfabetización en IA es esencial. Kajiwara & Kawabata (2024) defienden que al integrar herramientas de IA en la enseñanza, los estudiantes no solo mejoran sus habilidades lingüísticas, sino que también adquieren competencias tecnológicas que serán valiosas en su vida profesional y personal. Este enfoque es particularmente relevante en el contexto educativo, donde los estudiantes deben estar preparados para interactuar con sistemas de IA en diversos ámbitos. La exposición a herramientas como Mizou y Grok fomenta un entendimiento crítico de la IA, promoviendo su uso ético y efectivo. Además, fomentar la alfabetización en inteligencia artificial entre los profesores es crucial para entender el funcionamiento de las herramientas de IA y sus impactos sociales, usándolas finalmente de manera responsable en entornos educativos (Ng et al., 2023).

A pesar del creciente interés en la IA en la educación, como indican Ayeni et al. (2024) aún existen vacíos en la investigación, especialmente en el contexto de la enseñanza de idiomas específicos como el español. La investigación en este campo es crucial para desarrollar estrategias que equilibren los beneficios de la IA con sus limitaciones, como la comprensión limitada de los matices culturales y la falta de conexión emocional. Al combinar la IA con la supervisión docente, este enfoque híbrido maximiza la efectividad del aprendizaje, ofreciendo un modelo que puede ser replicado en otros contextos lingüísticos.

Otro de los grandes potenciales de la IA es que puede democratizar la educación lingüística, al proporcionar acceso a recursos de aprendizaje en áreas remotas, donde los docentes cualificados pueden ser escasos. Plataformas como Mizou permiten a los estudiantes practicar el español a través de enlaces accesibles desde cualquier dispositivo, eliminando barreras geográficas y horarias. La IA ofrece retroalimentación consistente y estandarizada, lo que es particularmente útil en la evaluación de habilidades lingüísticas. Estudios han demostrado que los sistemas de puntuación automatizada son altamente fiables en aspectos como la gramática y la coherencia (Mizumoto & Eguchi, 2023). En la enseñanza del español, donde la precisión gramatical es crucial, esta consistencia mejora la calidad del aprendizaje.

4. Ejemplo de diseño de tarea

La integración de la IA en la enseñanza de idiomas ha transformado las metodologías educativas, ofreciendo herramientas que personalizan el aprendizaje y optimizan la evaluación. Este apartado describe el diseño de una tarea que utiliza la plataforma Mizou para crear chatbots personalizados que simulan conversaciones, en el contexto de la enseñanza del español como L2. Se presenta un enfoque híbrido que combina la interactividad y escalabilidad de la IA con la supervisión docente, promoviendo un aprendizaje autónomo, contextualizado y eficiente. La metodología se alinea con investigaciones recientes que destacan la efectividad de los chatbots en el aprendizaje de idiomas (Koç & Savaş, 2025) y la fiabilidad de la evaluación automatizada (Mizumoto & Eguchi, 2023; Settles et al., 2024). Además, este enfoque se enmarca dentro de las tendencias actuales en la educación de IA en entornos educativos, donde se enfatiza la importancia de la personalización del aprendizaje y la integración de actividades prácticas y basadas en proyectos para fomentar la alfabetización en IA y habilidades de resolución de problemas (Lee & Kwon, 2024). La utilización de chatbots, como los

proporcionados por Mizou, permite simular conversaciones reales, ofreciendo a los estudiantes una práctica interactiva y contextualizada, lo cual es fundamental para el desarrollo de habilidades comunicativas en una L2 (Harry, 2023). Herramientas como Grok facilitan la evaluación automatizada, proporcionando una primera capa de corrección que los profesores pueden refinar, optimizando así el tiempo dedicado a la corrección manual y permitiendo a los educadores enfocarse en actividades más significativas (Harry, 2023).

El diseño de la tarea comienza con la selección de un tema alineado con el calendario académico y los objetivos pedagógicos del curso. Este paso asegura que el contenido sea relevante y se integre con las lecciones previas. Por ejemplo, si en el curso se está abordando la unidad de ‘La comida y restaurantes’, el tema se centra en vocabulario (platos, bebidas, adjetivos) y estructuras gramaticales (como ‘Me gustaría + infinitivo’ o el uso correcto de ‘gustar’) relacionadas con pedir comida en restauración. La selección del tema también considera las necesidades de los estudiantes, adaptándose a su nivel de competencia lingüística y objetivos de aprendizaje, lo que fomenta la instrucción diferenciada (Song et al., 2024). La IA puede analizar el rendimiento pasado de los estudiantes para identificar áreas de dificultad y proporcionar apoyo específico en esas áreas, lo que permite una personalización aún mayor del contenido (Harry, 2023).

5. Configuración del Chatbot

La plataforma Mizou permite una amplia personalización cuya configuración para el chatbot incluye varios componentes clave:

1. Mensaje de Bienvenida

El mensaje de bienvenida establece el tono de la interacción y proporciona una introducción clara al ejercicio. Por ejemplo: ‘¡Hola! Soy tu camarero virtual. Hoy vamos a practicar cómo pedir comida en un restaurante en español. Usa el vocabulario y las frases que hemos aprendido en clase, como ‘Me gustaría’ o ‘La cuenta, por favor.’ Este mensaje es crucial para reducir la ansiedad del estudiante y motivarlo a participar activamente, ajustándose con los principios de la psicología positiva en el aprendizaje de idiomas (Koç & Savaş, 2025). Además, la integración de elementos de gamificación, como recompensas por completar tareas o alcanzar objetivos de aprendizaje, puede aumentar la motivación y el compromiso de los estudiantes, haciendo que la experiencia de aprendizaje sea más interactiva y atractiva (Harry, 2023).

2. Instrucciones para la IA

Las instrucciones definen el rol del chatbot y las expectativas para el estudiante. Por ejemplo: ‘Actúa como un camarero en un restaurante. Saluda al estudiante, ofrécele el menú, toma su pedido, haz preguntas sobre preferencias (por ejemplo, ‘¿Cómo desea la carne?’), corrige errores gramaticales o de vocabulario de manera constructiva, y despídete amablemente. Enfócate en el uso de vocabulario relacionado con comida y estructuras como ‘Me gustaría + infinitivo.’ Estas instrucciones aseguran que la conversación permanezca dentro del tema y promueva las habilidades lingüísticas deseadas, apoyando el enfoque comunicativo (Pedro et al., 2019).

3. Reglas

Las reglas incluyen el contenido que el estudiante debe aprender y las normas para aplicarlo. Por ejemplo:

- Vocabulario: Platos (ensalada, sopa, carne, pescado), bebidas (agua, vino, cerveza), adjetivos (caliente, frío, picante).
- Frases comunes: ‘¿Qué desea pedir?’, ‘Me gustaría...’, ‘La cuenta, por favor’.
- Reglas gramaticales: Uso de ‘gustar’ (‘Me gustaría la pasta’ vs. ‘Me gusta la pasta’), concordancia de género y número, pronombres indirectos.

Este apartado proporciona al chatbot el conocimiento necesario para evaluar respuestas y ofrecer retroalimentación inmediata, mejorando la confianza del estudiante (Crompton et al., 2024).

4. Foto de Perfil y Descripción

Se selecciona una imagen que represente al chatbot, como un camarero sonriente, y se incluye una breve descripción: ‘Este ejercicio te permite practicar cómo pedir comida en un restaurante en español, simulando una interacción real.’ Estos elementos visuales y descriptivos hacen la tarea más atractiva y contextualizada, aumentando el compromiso estudiantil (Reinders & White, 2016).

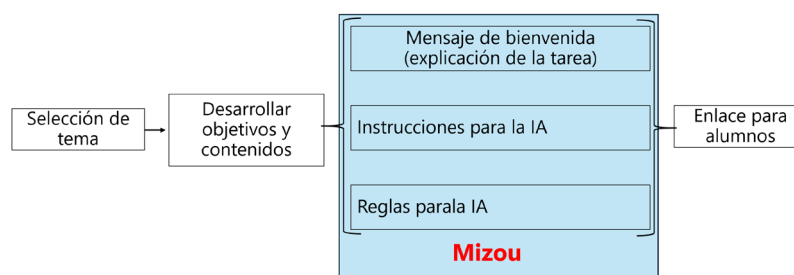


Figura 1: esquema de cómo se plantea el diseño de la tarea a través de la herramienta Mizou.

Tras configurar el chatbot (como se representa en la Figura 1), se guarda y se genera un enlace único para el ejercicio. Mizou permite crear sesiones separadas para diferentes grupos o clases, facilitando la organización en contextos con múltiples cohortes. El docente comparte el enlace, permitiendo a los estudiantes acceder desde cualquier dispositivo con conexión a internet. Esta accesibilidad elimina barreras geográficas y horarias, facilitando el aprendizaje lingüístico (Consejo de Europa, 2024). Mizou ofrece una interfaz donde el docente puede monitorear en tiempo real el progreso de los estudiantes, visualizando los nombres de quienes han iniciado, están en proceso o han completado la tarea. Esta funcionalidad permite identificar rápidamente a los estudiantes que necesitan apoyo adicional, apoyando la gestión del aula y la intervención oportuna (Misiejuk et al., 2024). Los estudiantes interactúan con el chatbot a través de una interfaz de chat, escribiendo mensajes y recibiendo respuestas generadas por la IA en tiempo real. El chatbot guía la conversación según las instrucciones, promoviendo el uso del vocabulario y estructuras objetivo. Por ejemplo, en el tema ‘La comida y restaurantes’, esta interacción simula una situación real, fomentando la competencia comunicativa y reduciendo la ansiedad al practicar en un entorno seguro (Koç & Savaş, 2025).

6. Proceso de evaluación

El proceso de evaluación combina la automatización con la supervisión humana, asegurando una retroalimentación precisa y significativa. Tras completar la tarea, el docente accede a las transcripciones de las conversaciones en Mizou y selecciona una muestra representativa para corregir manualmente. En el caso de este trabajo, los errores se clasifican en tres categorías:

- Semántica: Uso incorrecto de palabras, como ‘vino rojo’ en lugar de ‘vino tinto’.
- Coherencia: Frases que no se ajustan al contexto, como ‘Me gusta la paella’ al ordenar en lugar de ‘Me gustaría la paella’.
- Coordinación de género y número: Errores de concordancia, como ‘una vino’ en lugar de ‘un vino’.

Las conversaciones de ejemplo, junto con sus correcciones, se almacenan en un archivo de texto que actúa como base de corrección, donde se detallan los errores y sus correcciones, y que se utiliza como referencia para la evaluación automática (Mizumoto & Eguchi, 2023). Esta base permite a Grok identificar y corregir errores similares en otras conversaciones, mejorando la consistencia de la evaluación (Settles et al., 2024). La combinación de evaluación automatizada y supervisión humana se alinea con las prácticas comunes en la educación de IA, donde se utilizan métodos mixtos para evaluar tanto los resultados de aprendizaje como las experiencias de los estudiantes (Lee & Kwon, 2024). Este proceso de evaluación automatizada no solo proporciona retroalimentación instantánea a los estudiantes, sino que también ahorra tiempo a los educadores, permitiéndoles concentrarse en proporcionar feedback cualitativo (Harry, 2023).

El docente revisa las correcciones de Grok, ajustando cualquier imprecisión, especialmente en casos donde la IA no capta matices contextuales o culturales. Por ejemplo, Grok podría no distinguir entre el uso formal e informal de ‘tú’ en diferentes contextos hispanohablantes, requiriendo la intervención del docente (Godwin-Jones, 2017). Esta supervisión asegura que el feedback sea preciso y relevante, alineándose con los principios de la evaluación formativa (Black & Wiliam, 1998).

Beneficios de la técnica híbrida en el desarrollo y corrección de ejercicios:

- Personalización: Mizou facilita la personalización de las actividades según los distintos niveles de habilidad y metas, favoreciendo la enseñanza personalizada (Song et al., 2024).
- Interactividad y autonomía: Los estudiantes practican en un entorno conversacional realista, fomentando la autonomía y reduciendo la ansiedad (Koç & Savaş, 2025).
- Eficiencia en evaluación: Grok reduce, en el caso del presente trabajo, el tiempo de corrección en un 30-40%, permitiendo al docente centrarse en el feedback cualitativo (Lee et al., 2024). Los resultados iniciales muestran un ahorro significativo de tiempo en la corrección de tareas, lo cual está en línea con la capacidad de la IA para automatizar tareas repetitivas y permitir a los educadores enfocarse en actividades más significativas (Harry, 2023).
- Escalabilidad: Un solo chatbot puede servir a múltiples estudiantes, ideal para clases numerosas.

- Alfabetización digital: La interacción con IA desarrolla competencias tecnológicas esenciales (Kajiwarara & Kawabata, 2024).

Además, estudios sistemáticos han demostrado que la educación en IA, incluyendo enfoques personalizados y el uso de herramientas interactivas, mejora la alfabetización en IA, las habilidades de resolución de problemas y la reflexión ética sobre el impacto social de la IA (Lee & Kwon, 2024).

7. Desafíos

- Limitaciones culturales: La IA puede fallar en captar matices culturales, como diferencias regionales en el español (Godwin-Jones, 2017).
- Dependencia tecnológica: Requiere acceso a dispositivos e internet, lo que plantea problemas de equidad (Zhu et al., 2024).
- Formación docente: Los docentes necesitan capacitación para diseñar tareas efectivas y supervisar la IA (Pedro et al., 2019).
- Conexión emocional: La falta de empatía en las interacciones con IA puede afectar la motivación a largo plazo (Zhang et al., 2021).
- Calidad de los datos: Es crucial asegurar que los datos utilizados para entrenar los algoritmos de IA sean confiables y precisos para evitar sesgos y garantizar la efectividad de las herramientas personalizadas (Harry, 2023).

8. Principios pedagógicos

- Enfoque comunicativo: Los estudiantes practican el idioma en contextos significativos, simulando situaciones reales (Pedro et al., 2019).
- Evaluación formativa: El feedback inmediato del chatbot ayuda en el aprendizaje continuo de los alumnos (Black & Wiliam, 1998).
- Diseño universal para el aprendizaje (UDL): La personalización y accesibilidad de la tarea atienden a diversos estilos y ritmos de aprendizaje (Song et al., 2024).
- Autenticidad: Aunque la interacción es con IA, el diseño asegura conversaciones lo más naturales posible, imitando interacciones reales (Reinders & White, 2016).

Las transcripciones de múltiples estudiantes permiten al docente identificar patrones de errores comunes, como la confusión entre ‘ser’ y ‘estar’. Estos datos contribuyen para ajustar las lecciones futuras, apoyando la instrucción basada en datos (Misiejuk et al., 2024). La base de corrección, a su vez, se actualiza continuamente, mejorando la precisión de Grok con el tiempo. El diseño de tareas con Mizou y Grok representa una metodología innovadora que maximiza el potencial de la IA en la enseñanza del español. Al combinar práctica conversacional personalizada, evaluación automatizada y supervisión docente, este enfoque mejora la autonomía, el compromiso y la eficiencia. Sin embargo, la supervisión humana es esencial para abordar las limitaciones de la IA, como la comprensión cultural y la conexión emocional, asegurando un aprendizaje rico y significativo.

9. Resultados

La integración de herramientas de inteligencia artificial (IA) como Mizou y Grok en la enseñanza del español como L2 ha generado resultados iniciales prometedores, particularmente en términos de eficiencia en la corrección de tareas y mejora de la instrucción basada en datos. Este apartado presenta los hallazgos de la implementación de este enfoque híbrido, tomando como ejemplo su implementación en varios grupos de algo más de veinte estudiantes, destacando el ahorro de tiempo, la precisión de las correcciones automatizadas, el impacto en la enseñanza y las percepciones estudiantiles. Los resultados se ajustan con investigaciones que subrayan la eficacia de los chatbots y los sistemas de evaluación automatizada en el aprendizaje de idiomas (Koç & Savaş, 2025; Mizumoto & Eguchi, 2023; Settles et al., 2024).

El estudio se llevó a cabo con estudiantes de nivel básico-intermedio en un curso de español como L2. Las tareas se diseñaron utilizando la plataforma Mizou para crear chatbots conversacionales personalizados, enfocados en temas del currículo académico. Las conversaciones generadas fueron evaluadas mediante un enfoque híbrido que combinó la corrección manual y la asistencia automatizada de Grok. Este proceso se aplicó en cinco tareas consecutivas, permitiendo acumular una base de corrección que creció con cada tarea, mejorando la eficiencia y precisión del sistema.

10. Proceso de corrección

Proceso de corrección, por ejemplo para un grupo de 22 estudiantes. La corrección de la tarea estaría compuesta por los siguientes pasos:

1. Corrección Manual: Se seleccionan las conversaciones de cinco estudiantes para una corrección manual detallada, tomando diez minutos por estudiante (50 minutos en total). Estas correcciones se clasificaron en tres categorías: semántica (por ejemplo, ‘vino rojo’ en lugar de ‘vino tinto’), coherencia (por ejemplo, ‘Me gusta la paella’ en lugar de ‘Me gustaría la paella’ al ordenar) y coordinación de género y número (por ejemplo, ‘una vino’ en lugar de ‘un vino’). Estas conversaciones forman la base de corrección, documentada en un archivo de texto.
2. Corrección Automatizada con Grok: Las conversaciones de los diecisiete estudiantes restantes serán corregidas automáticamente por Grok, utilizando la base de corrección. Este proceso se demora alrededor de tres minutos por estudiante para generar las correcciones, que luego serán revisadas por el docente.
3. Revisión Docente: Inicialmente, la revisión de las correcciones de Grok toma tres minutos por estudiante. A medida que la base de corrección crece, la precisión de Grok mejora, reduciendo el tiempo de revisión a dos minutos por estudiante en tareas posteriores.
4. Análisis de Errores Comunes: Al finalizar todas las revisiones, se vuelve a recurrir a Grok para analizar todas las correcciones de forma conjunta, e identificar patrones de errores frecuentes, proporcionando datos para preparar la clase de corrección de esa tarea.



La base de corrección se expandió con cada tarea, incorporando las cinco conversaciones corregidas manualmente por tarea. Por ejemplo, en la tarea uno, la base contenía cinco conversaciones; en la tarea tres, quince; y en la tarea cinco, 25. Esta acumulación permitió a Grok aprender patrones de corrección más precisos, alineándose con los principios del aprendizaje supervisado, donde un mayor conjunto de datos mejora el rendimiento del modelo (Goodfellow et al., 2016).

La eficiencia en la corrección fue un resultado clave de esta implementación. Sin la asistencia de Grok, corregir manualmente las 22 conversaciones habría tomado 220 minutos (22 estudiantes \times 10 minutos). Con el enfoque híbrido, el tiempo total del docente varió según el tamaño de la base de corrección, como se detalla en la Tabla 1.

Tarea	Tamaño base corrección	Tiempo de revisión por estudiante (min)	Tiempo de corrección manual (min)	Tiempo de revisión (min)	Tiempo total del docente (min)	Tiempo ahorrado (min)	Porcentaje de tiempo ahorrado (min)
1	5	3	50	51	101	119	54.09%
3	15	2.5	50	42.5	92.5	127.5	57.95%
5	25	2	50	34	84	136	61.82%

Tabla1: El tiempo de corrección manual corresponde a cinco estudiantes a diez minutos cada uno. El tiempo de revisión corresponde a diecisiete estudiantes al tiempo especificado por estudiante. El tiempo total del docente es la suma de los tiempos de corrección manual y revisión. El tiempo ahorrado se calcula respecto a la corrección manual completa de 22 estudiantes a diez minutos cada uno (220 minutos).

La Tabla 1 muestra que el tiempo total del docente disminuyó a medida que la base de corrección creció. En la tarea uno, con una base de cinco conversaciones, el tiempo total fue de 101 minutos, representando un ahorro del 54.09%. En la tarea tres, con una base de quince conversaciones, el tiempo se redujo a 92.5 minutos (57.95% de ahorro). En la tarea cinco, con una base de 25 conversaciones, el tiempo total fue de 84 minutos, logrando un ahorro del 61.82%. Este aumento en la eficiencia se atribuye a la mayor precisión de Grok, que redujo el tiempo de revisión de tres a dos minutos por estudiante, coincidiendo con estudios que destacan la capacidad de los sistemas de IA para optimizar procesos de evaluación (Lee et al., 2024; Mizumoto & Eguchi, 2023). Los resultados iniciales muestran un ahorro significativo de tiempo en la corrección de tareas, lo cual está en línea con la capacidad de la IA para automatizar tareas repetitivas y permitir a los educadores enfocarse en actividades más significativas (Harry, 2023).

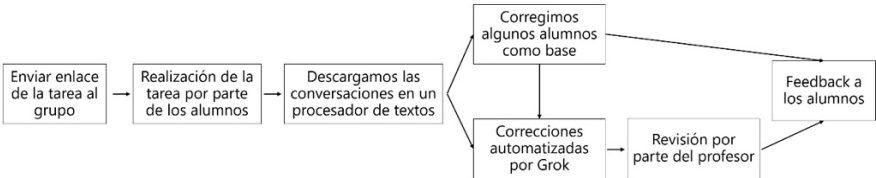


Figura 2: se representa de forma esquemática la realización junto con la posterior revisión de la tarea siguiendo el patrón explicado en el trabajo.

Aunque no se recopilaban datos cuantitativos sobre las percepciones estudiantiles en esta fase inicial, las observaciones cualitativas sugieren que los estudiantes valoraron la oportunidad de practicar en un entorno conversacional seguro proporcionado por Mizou. Gracias a la implementación de la actividad siguiendo el esquema de la Figura 2, los estudiantes informaron sentirse más cómodos cometiendo errores con el chatbot que en interacciones cara a cara, lo que les permitió experimentar con el idioma sin temor a juicios. Sin embargo, algunos estudiantes notaron que el chatbot ocasionalmente no captaba matices culturales, como diferencias regionales en el vocabulario (por ejemplo, ‘tinto’ versus ‘rojo’ para el vino), lo que requería aclaraciones posteriores del docente. Estas observaciones son consistentes con la literatura que señala las limitaciones de la IA en la comprensión de contextos culturales complejos (Godwin-Jones, 2017).

A pesar de los resultados positivos, se identificaron varios desafíos. Primero, la dependencia de la tecnología plantea preocupaciones sobre la equidad en el acceso, ya que no todos los estudiantes pueden tener dispositivos o conexiones a internet confiables (Zhu et al., 2024). Segundo, la formación docente es esencial para maximizar el potencial de estas herramientas, ya que configurar chatbots efectivos y supervisar las correcciones de Grok requiere habilidades técnicas y pedagógicas (Pedro et al., 2019).

Finalmente, la falta de conexión emocional en las interacciones con la IA puede limitar la motivación a largo plazo de algunos estudiantes, lo que subraya la importancia de la supervisión docente para proporcionar apoyo emocional (Zhang et al., 2021).

10. Conclusión

Este estudio ha explorado la integración de herramientas de IA, específicamente Mizou para la práctica conversacional y Grok para la asistencia en la evaluación, en la enseñanza del español como L2. El objetivo fue examinar cómo estas herramientas pueden mejorar la personalización, accesibilidad y eficiencia en la educación lingüística, al tiempo que se abordan sus limitaciones inherentes.

Los resultados iniciales, obtenidos de una implementación con varios grupos de estudiantes, indican que el uso de Mizou permite a los estudiantes participar en conversaciones contextualizadas y personalizadas, fomentando la autonomía y reduciendo la ansiedad asociada con el aprendizaje de un nuevo idioma. La retroalimentación inmediata proporcionada por los chatbots mejora la confianza de los estudiantes y refuerza el aprendizaje de manera efectiva, como respaldan estudios sobre la eficacia de la IA en la educación (Lee et al., 2024). Por otro lado, Grok ha demostrado ser una herramienta valiosa para la evaluación automatizada, reduciendo significativamente el tiempo de corrección manual en un 54.09% a 61.82% a medida que la base de corrección creció. Este ahorro permitió a los docentes centrarse en proporcionar retroalimentación cualitativa, un aspecto crucial para el aprendizaje personalizado (Misiejuk et al., 2024). Además, la identificación de patrones de errores comunes, por ejemplo, como la confusión entre ‘ser’ y ‘estar’, facilitó una preparación de clases más dirigida, alineándose con los principios de la evaluación formativa.

Estos hallazgos tienen implicaciones significativas para la enseñanza de idiomas. La personalización ofrecida por Mizou apoya el UDL, adaptándose a la heterogeneidad de los alumnos (Song et al., 2024). La accesibilidad de estas herramientas elimina barreras geográficas y horarias, democratizando la educación lingüística, especialmente en contextos donde los recursos docentes son limitados.

Además, el uso de IA fomenta la alfabetización digital, una competencia esencial en la sociedad actual, como definen Kajiwaru & Kawabata (2024).

A pesar de los beneficios señalados de la IA, esta presenta limitaciones. Una de las principales es su capacidad limitada para captar matices culturales, como las diferencias regionales en el español, lo que puede confundir a los estudiantes, según Godwin-Jones (2017). Esto requiere intervención docente para proporcionar contexto cultural adecuado. Otro desafío es la dependencia tecnológica, que puede exacerbar las desigualdades educativas si no todos los estudiantes tienen acceso a dispositivos o internet, un tema crítico según Zhu et al. (2024). La formación docente también es esencial, ya que diseñar tareas efectivas y supervisar la IA requiere habilidades específicas, como señalan Pedro et al. (2019). Finalmente, la falta de conexión emocional en las interacciones con la IA puede afectar la motivación a largo plazo, un punto destacado por Zhang et al. (2021). Es crucial asegurar que los datos utilizados para entrenar los algoritmos de IA sean confiables y precisos para evitar sesgos y garantizar la efectividad de las herramientas personalizadas (Harry, 2023).

Se necesitan estudios longitudinales para evaluar el impacto a largo plazo de la IA en la competencia lingüística y la motivación estudiantil. Explorar la efectividad de estas herramientas en otros idiomas, como el inglés o el francés, y en diferentes niveles de competencia podría ampliar su aplicabilidad. Además, mejorar la competencia cultural de la IA mediante datos de entrenamiento más diversos o modelos híbridos es una prioridad. Investigar cómo la IA apoya a aprendices con necesidades diversas, alineado con UDL, podría fomentar entornos más inclusivos. Finalmente, la educación sobre ética de la IA, como proponen Kajiwaru & Kawabata (2024), es esencial para promover un uso responsable, respetando la privacidad y la equidad. Como señalan Lee & Kwon (2024), es fundamental abordar las consideraciones éticas y los desafíos asociados con la implementación de la IA en la educación para garantizar un uso responsable y equitativo de estas tecnologías, asegurando accesibilidad, transparencia y equidad en los sistemas educativos basados en IA.

Referencias

- Alier, M., García-Peñalvo, F.J. and Camba, J.D. 2024. "Generative artificial intelligence in education: from deceptive to disruptive." *International Journal of Interactive Multimedia and Artificial Intelligence* 8 (5): 5–14.
- Ayeni, O.O., Al Hamad, N.M., Chisom, O.N., Osawaru, B. and Adewusi, O.E. 2024, "AI in education: a review of personalized learning and educational technology." *GSC Advanced Research and Reviews* 18 (2): 261–271.
- Black, P. and Wiliam, D. 1998. "Assessment and classroom learning". *Assessment in Education: Principles, Policy and Practice* 5 (1): 7–74, <<https://doi.org/10.1080/0969595980050102>>.
- Chapelle, C. 2003. *English language learning and technology: lectures on applied linguistics in the age of information and communication technology* (Vol. 7). Amsterdam: John Benjamins Publishing.
- Crompton, H., Edmett, A., Ichaporia, N. and Burke, D. 2024. "AI and English language teaching: affordances and challenges." *British Journal of Educational Technology* 55 (6): 2503–2529.
- Godwin-Jones, R. 2017. "Emerging technologies: autonomous language learning." *Language Learning & Technology* 21 (3): 4–11.
- Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y. 2016. *Deep learning* (Vol. 1, No. 2). Cambridge: MIT Press.
- Gray, A., Rahat, A., Crick, T. and Lindsay, S. 2024. "A Bayesian active learning approach to comparative judgement within education assessment." *Computers and Education: Artificial Intelligence* 6, 100226.

- Harry, A. 2023. "Role of AI in education." *Interdisciplinary Journal and Humanity (INJURITY)* 2 (3).
- Instituto Cervantes. 2024. *El español en el mundo: anuario del Instituto Cervantes 2024*. Accessed: 8 November 2025, <https://cvc.cervantes.es/lengua/anuario/anuario_24/el_espanol_en_el_mundo_anuario_instituto_cervantes_2024.pdf>.
- Kajiwara, Y. and Kawabata, K. 2024. "AI literacy for ethical use of chatbot: will students accept AI ethics?" *Computers and Education: Artificial Intelligence* 6, 100251.
- Kim, H. and Park, M. 2023. "Virtual influencers' attractiveness effect on purchase intention: a moderated mediation model of the product-endorser fit with the brand." *Computers in Human Behavior* 143, 107703.
- Knott, N., Tolzin, A., Janson, A. and Leimeister, J.M. 2024. "AI literacy and its implications for prompt engineering strategies." *Computers and Education: Artificial Intelligence* 6, 100225.
- Koç, F.Ş. and Savaş, P. 2025. "The use of artificially intelligent chatbots in English language learning: a systematic meta-synthesis study of articles published between 2010 and 2024." *ReCALL* 37 (1): 4–21.
- Lee, J., Hicke, Y., Yu, R., Brooks, C. and Kizilcec, R.F. 2024. "The life cycle of large language models in education: a framework for understanding sources of bias." *British Journal of Educational Technology* 55 (5): 1982–2002.
- Lee, S.J. and Kwon, K. 2024. "A systematic review of AI education in K-12 classrooms from 2018 to 2023: topics, strategies, and learning outcomes." *Computers and Education: Artificial Intelligence* 6, 100211.
- Limna, P., Jakwatanatham, S., Siripipattanakul, S., Kaewpuang, P. and Sriboonruang, P. 2022. "A review of artificial intelligence (AI) in education during the digital era." *Advance Knowledge for Executives* 1 (1): 1–9.
- Lyu, Y. 2024. "Effectiveness of chatbots in improving language learning: a meta-analysis of comparative studies." *International Journal of Applied Linguistics*.
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocaranza-Prado, I. and Kreibel, D. 2023. "Impact of artificial intelligence on assessment methods in primary and secondary education: systematic literature review." *Revista de Psicodidáctica (English ed.)* 28 (2): 93–103.
- Mizumoto, A. and Eguchi, M. 2023. "Exploring the potential of using an AI language model for automated essay scoring." *Research Methods in Applied Linguistics* 2 (2), 100050.
- Misicjuk, K., Kaliisa, R. and Scianna, J. 2024. "Augmenting assessment with AI coding of online student discourse: a question of reliability." *Computers and Education: Artificial Intelligence* 100216.
- Ng, D.T.K., Su, J., Leung, J.K.L. and Chu, S.K.W. 2023. "Artificial intelligence (AI) literacy education in secondary schools: a review." *Interactive Learning Environments* 32 (10): 6204–6224.
- Pedro, F., Subosa, M., Rivas, A. and Valverde, P. 2019. *Artificial intelligence in education: challenges and opportunities for sustainable development*. Accessed: 8 November 2025, <<http://repositorio.minedu.gob.pe/handle/20.500.12799/6533>>
- Petersen, S.A. and Markiewicz, J.-K. 2008. "Personalised language learning on mobile devices." *Fifth IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education, WMUTE 2008*, Beijing, China, 23–26 March 2008 95–102. Accessed: 8 November 2025, <https://www.researchgate.net/publication/220356402_Personalized_and_Contextualized_Language_Learning_Choose_when_Where_and_What>.
- Reinders, H. and White, C. 2016. "20 years of autonomy and technology: how far have we come and where to next?" *Language Learning & Technology* 20 (2): 143–154.
- Sáenz Valiente, P. 2024. *Uso de la IA para personalizar el aprendizaje y mejorar la educación primaria*. UNIR. Accessed: 8 November 2025, <<https://www.unir.net/revista/educacion/ia-para-personalizar-aprendizaje-mejorar-educacion-primaria/>>.
- Settles, I.H., Jones, M.K., Buchanan, N.T., Dotson, K., Grower, P., O'Rourke, M. and Latimer, K. 2024. "Epistemic exclusion: a theory for understanding racism in faculty research evaluations." *American Psychologist* 79 (4): 539.
- Song, Y., Weisberg, L.R., Zhang, S., Tian, X., Boyer, K.E. and Israel, M. 2024. "A framework for inclusive AI learning design for diverse learners." *Computers and Education: Artificial Intelligence* 100212.

- Stollhans, S. 2024. “With AI translation tools so powerful, what is the point of learning a language.” *The Conversation*. Accessed: 8 November 2025, <<https://theconversation.com/with-ai-translation-tools-so-powerful-what-is-the-point-of-learning-a-language-238068>>.
- Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.T. and Le, Q. 2022. “LaMDA: language models for dialog applications.” *ArXiv preprint*.
- Vesselinov, R. and Grego, J. 2012. *Duolingo effectiveness study*. Accessed: 8 November 2025, <http://static.duolingo.com/s3/DuolingoReport_Final.pdf>.
- Xiao, Y., Zhang, T. and He, J. 2024. “The promises and challenges of AI-based chatbots in language education through the lens of learner emotions.” *Heliyon* 10 (18) e37238.
- Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B. and Perrault, R. 2021. “The AI index 2021 annual report.” *ArXiv preprint arXiv:2103.06312*.
- Zhu, X., Li, J., Liu, Y., Ma, C. and Wang, W. 2024. “A survey on model compression for large language models.” *Transactions of the Association for Computational Linguistics* 12: 1556–1577.