

RLHF y RLAIIF, Revolución Silenciosa de la Retroalimentación Educativa

Díaz Rodríguez, Elizabeth <sup>1</sup>

<sup>1</sup> Universidad Ana G. Méndez, [ediaz79@uagm.edu](mailto:ediaz79@uagm.edu)

**Resumen**

El aprendizaje reforzado a partir de comentarios se ha consolidado como una técnica innovadora en el campo del aprendizaje automático, permitiendo los modelos de inteligencia artificial (IA). La investigación compara el RLHF (*Reinforcement Learning Human Feedback*) y el RLAIIF (*Reinforcement Learning from AI Feedback*). La mayoría de las investigaciones manifiestan una inclinación preferencial hacia el modelo RLAIIF por su escalabilidad (Khedri & Höglund, 2023; Lee et al., 2022 & Zhichao et al., 2024). Otros investigadores proponen un enfoque híbrido, integrando ambas alternativas (Dakota, M.,2024). Ambos enfoques se complementan para mejorar el proceso de aprendizaje.

*Palabras claves:* retroalimentación, RLHF, RLAIIF, Inteligencia Artificial

**Abstract**

Reinforcement learning from feedback has emerged as an innovative technique in machine learning, enhancing artificial intelligence (AI) model training. Current research compares two key approaches: RLHF (Reinforcement Learning from Human Feedback) and RLAIIF (Reinforcement Learning from AI Feedback). Most studies demonstrate a preference for RLAIIF due to its superior scalability (Khedri & Höglund, 2023; Lee et al., 2022; Zhichao et al., 2024).

However, other researchers advocate for a hybrid approach that strategically combines both methods (Dakota, 2024). These complementary frameworks can synergistically improve machine learning processes.

*Keywords:* feedback, RLHF, RLAIIF, Artificial Intelligence

### **RLHF y RLAIIF, Revolución Silenciosa de la Retroalimentación Educativa**

¿Qué entendemos por retroalimentación? La retroalimentación en el ámbito educativo es la reacción que provee el educador al estudiante sobre su labor en una actividad o tarea. El propósito es guiar, reforzar habilidades, destrezas y competencias. Además, proporciona herramientas para la evolución progresiva del aprendizaje (Escuela de profesores del Perú, 2024; Núñez-Valdés et al., 2024; Umáximo, s.f.). Descubrimos que sucede directamente debido a la intervención del educador. La naturaleza es personalizada, contextualizada y se adapta a las necesidades y características específicas de cada estudiante. Se convierte en una interacción directa y humana donde las emociones y motivaciones de los involucrados da forma a la interacción (Núñez-Valdés et al., 2024). Se busca fomentar la reflexión crítica, ya sea sincrónica o asincrónica, basada en la observación continua (Escuela de profesores del Perú, 2024). Luna Acuna y colaboradore (2022) resumen el concepto como “un proceso oportuno, adecuado al contexto y tipo de tarea o actividad de aprendizaje. Involucra la comunicación fluida entre docente, estudiantes y padres de familia. Además, favorece la reflexión constructiva del discente y la autorreflexión del docente respecto a la calidad de su práctica pedagógica” (p. 1).

La invención de la máquina, la revolución industrial y agrícola sistematizó las tareas. La globalización y las TIC (Tecnologías de la Información y Comunicación) encontraron la razón de ser, para mejorar la eficacia de los modelos de aprendizaje automática.

### **Planteamiento del problema**

El *aprendizaje reforzado* a partir de comentarios se convirtió en una técnica de aprendizaje automático en la que los modelos de inteligencia artificial (IA) aprenden comportamientos a través de comentarios humanos directos en lugar de funciones de recompensa más tradicionales, alineando mejor la IA con los objetivos y expectativas humanos de retroalimentación (Sciencenow, 2025). Consiste en entrenar el modelo basándose en las preferencias o correcciones proporcionadas por seres humanos. En lugar de simplemente predecir las secuencias de palabras mediante la revisión de datos, la IA se alinea estrechamente con las ideas humanas de lo que constituye una respuesta buena o útil de acuerdo con los estándares humanos. El concepto se sugirió por primera vez en OpenAI en 2019 y es una evolución del aprendizaje reforzado (RL). El aprendizaje reforzado a partir de comentarios de humanos y el aprendizaje reforzado que ajusta un modelo de recompensa basado en las preferencias derivadas de un modelo de lenguaje estándar son métodos de aprendizaje automático (ML) para entrenar sistemas de IA, pero difieren significativamente en la forma en que guían el proceso de aprendizaje. El RL tradicional se basa en señales de recompensa procedentes del entorno, lo que significa que la IA recibe comentarios de sus acciones dentro de un conjunto predefinido de automatizaciones y aprende a maximizar estas recompensas a través de pruebas y errores. Los comentarios automatizados definen lo que es preciso o natural, pero no se ajustan necesariamente con las complejas preferencias humanas (Sciencenow, 2025).

Por el contrario, la retroalimentación basada en reacciones personalidad y sensibilidad del individuo incorpora comentarios humanos directos en el ciclo de aprendizaje, lo que proporciona a la IA conocimientos reales y contextualmente relevantes sobre lo que los humanos consideran resultados deseables o de alta calidad. Este método permite que la IA aprenda no solo a realizar tareas, sino también a adaptar sus respuestas de acuerdo con las opiniones humanas, lo que la hace más eficaz para aplicaciones en las que es esencial una comprensión similar a la humana (Sciencenow, 2025). A este tipo de retroalimentación se le conoce con sus siglas en inglés como *reinforcement learning human feedback* (RLHF).

No obstante, el *Reinforcement Learning from AI Feedback* (RLAIIF) es otra técnica de aprendizaje automático que representa un avance significativo utilizando la IA (Data Camp, 2024; Lee et al., 2024). En general, el RLAIIF (Lee et al. 2024) introduce un enfoque escalable para entrenar, sin depender de etiquetas de retroalimentación humana. Aprovechando una *constitución* de principios codificados en lenguaje natural con la capacidad de autocrítica de un modelo inicial útil, los autores demuestran resultados prometedores en la dirección del comportamiento del modelo lingüístico, al tiempo que mejoran la transparencia y la flexibilidad.

La RLAIIF muestra potencial para ampliar la supervisión a medida que los modelos sean más capaces, complementando y automatizando parcialmente la supervisión humana. También mitiga problemas como la especificación errónea de la recompensa y el comportamiento evasivo observados en anteriores enfoques de retroalimentación humana. Sin embargo, mantener una sólida alineación humana, gestionar el cambio distributivo y garantizar una supervisión fiable de los sistemas avanzados de IA siguen siendo retos pendientes (Data Camp, 2024).

El RLHF y el RLAIIF son dos enfoques que buscan optimizar la retroalimentación para mejorar el rendimiento de los modelos de RL. El desarrollo de técnicas como el RLHF y el RLAIIF refleja la tendencia hacia modelos de IA más sofisticados y alineados con las necesidades humanas. Cabría entonces preguntar ¿Cuáles son las características que señala la investigación entre RLHF y RLAIIF en la optimización de modelos de aprendizaje automático? ¿Cuáles son los desafíos?

### **Justificación**

En la actualidad, el aprendizaje por refuerzo, con sus siglas en inglés RL (*Reformente Learning*) se ha convertido en una herramienta fundamental para el desarrollo de modelos de inteligencia artificial en diversas áreas, desde la robótica hasta el procesamiento del lenguaje natural. Sin embargo, la efectividad de estos modelos depende en gran medida de la calidad y el tipo de retroalimentación recibida durante el proceso de entrenamiento. RLHF permite aprovechar la intuición y el juicio humano para afinar las respuestas del modelo (Sciencenow, 2025), mientras que RLAIIF utiliza la rapidez y eficiencia de otros modelos de IA para proporcionar una retroalimentación constante y menos costosa (Data Camp, 2024).

### **Propósito**

Los hallazgos de esta investigación proporcionarán una información valiosa para investigadores y profesionales en el conocimiento de los métodos de retroalimentación basados en el AI. Se espera contribuir al campo del aprendizaje por refuerzo, impulsando nuevas innovaciones y aplicaciones prácticas.

## Revisión de Literatura

La retroalimentación, conocida como refuerzo *del aprendizaje*, por sus siglas en inglés RI (*reinforcement learning*) es reconocida por la literatura especializada como uno de los elementos esenciales para promover el aprendizaje. El enfoque desde un principio fue acompañar el proceso, progreso y producto del aprendizaje para mejorar la comprensión del estudiantado sobre la información y el contenido trabajado en el aula. Además, de fomentar la coconstrucción y el desarrollo de habilidades complejas para fortalecer la trayectoria académica (Herrera-Araya, 2003).

Si bien la investigación sobre la temática no es profusa ni pletórica, es posible detectar debates y tensiones que impactan la producción de conocimiento en su estudio. Por mucho tiempo el dilema se centró en si la práctica fuera de naturaleza monológica, unidireccional y centrada en el profesorado, o dirigida hacia un enfoque socio-constructivista que la define como un fenómeno social complejo (Herrera-Araya, 2023).

La retroalimentación automatizada surge como respuesta a varias necesidades y desafíos en diversos campos, particularmente en la educación y la inteligencia artificial, especialmente como una solución para mejorar la eficiencia, la personalización y el aprendizaje autónomo en la educación, aprovechando los avances tecnológicos para transformar la experiencia educativa tanto para estudiantes como para educadores (Vocerol, 2018).

El aprendizaje por refuerzo, *reinforcement learning* (RL) es un modelo de inteligencia artificial. Funciona con un algoritmo de aprendizaje automático resolviendo problemas mediante prueba y error. ¿Cómo lo hace? Basándose en datos obtenidos. Primero entrena distintos escenarios para tomar una serie de decisiones. En función de las acciones realizadas recibe

recompensas o penalizaciones. Y en base a esas respuestas busca maximizar la recompensa. Pese a que no es un concepto nuevo, los avances tecnológicos han mejorado considerablemente los resultados que se obtienen, en diversas áreas. La efectividad de estos modelos depende en gran medida de la calidad y el tipo de datos recibidos durante el proceso.

RLHF se ha utilizado comúnmente para alinear los modelos de lenguaje con las preferencias individuales mediante la incorporación de anotaciones humanas, mientras que RLAIIF propone usar un enfoque basado en IA para reemplazar la retroalimentación humana (Khedri & Höglund, 2023). El RLHF (*reinforcement learning for human feedback*) es el aprendizaje automático, por refuerzo a partir de la retroalimentación con revisión individualizada por un instructor, experto, profesor o educador. RLHF, por sus siglas en inglés, es una técnica para alinear un agente inteligente con las distinciones muy específicas del individuo o en todo caso de los usuarios. Implica entrenar un modelo de recompensa para representar las preferencias, que luego se puede utilizar para entrenar otros modelos a través del aprendizaje por refuerzo. El ranking se realiza preferentemente de forma comparativa; por ejemplo, se evalúan varias respuestas a una pregunta, clasificándolas en función de las condiciones de ajuste. El proceso de RLHF requiere la participación de seres humanos, quienes actúan como anotadores para clasificar, ordenar y comparar la salida de un modelo de lenguaje, con el objetivo de entrenarlo y enseñarle si lo que produce se considera deseable o no (Khedri & Höglund, 2023; Nanobaly, 2024).

El aprendizaje por refuerzo a partir de la retroalimentación de la IA en RLAIIF, (*reinforcement learning for artificial intelligence feedback* por sus siglas en inglés) es una técnica de aprendizaje automático en la que los modelos de IA brindan retroalimentación a otros modelos de IA durante el proceso de aprendizaje por refuerzo. En lugar de depender únicamente

de la información humana, RLAIIF aprovecha las capacidades de los sistemas de IA existentes, como los grandes modelos de lenguaje, para evaluar acciones y guiar el aprendizaje. Esta retroalimentación de la IA puede adoptar diversas formas, como generar recompensas, clasificar respuestas o sugerir mejoras directamente. Al automatizar el ciclo de retroalimentación, RLAIIF tiene el potencial de optimizar la capacitación, reducir costos y mejorar el rendimiento de varios sistemas de IA, incluidos los grandes modelos de lenguaje. RLAIIF utiliza un enfoque basado en IA para reemplazar la retroalimentación humana. Ofrece beneficios potenciales, para una mayor eficiencia de tiempo y recursos (Khedri & Höglund, 2023).

LLM (siglas en inglés para *Large Language Model*), es un modelo de lenguaje de aprendizaje profundo con una red neuronal, monitoreado por miles de millones de parámetros, entrenados en grandes cantidades de texto sin etiquetar mediante aprendizaje autosupervisado o aprendizaje semisupervisado (*What is a Large Language Model (LLM)*, 2023). RLHF aprovecha la intuición y el juicio humano para afinar las respuestas del modelo, mientras que RLAIIF utiliza en LLM la rapidez y eficiencia de otros modelos de IA para proporcionar una retroalimentación constante, menos costosa aprovechando los modelos de IA, mejorando el rendimiento y la escalabilidad. (Data Camp, 2024). La retroalimentación humana se considera costosa y limitada en términos de tiempo y disponibilidad, mientras que la retroalimentación de IA ofrece una alternativa más rápida y escalable.

La calidad de la retroalimentación determina si el desarrollo del modelo es robusto y preciso. Comprobar cuál de los dos métodos proporciona una mejor retroalimentación mejorará el rendimiento y la generalización de los modelos de aprendizaje automático, beneficiando al sistema educativo público, privado y empresarial. A continuación, presentaremos una tabla

comparativa de las características de RLAIIF (*Reinforcement Learning from AI Feedback*) y RLHF (*Reinforcement Learning from Human Feedback*)(sapien, 2024) en la literatura. La tabla 1 compara las características entre RLHF y RLAIIF, en términos de escalabilidad (adaptabilidad y eficiencia), costo, matiz en la retroalimentación, y uso recomendado.

**Tabla 1**

*Características Comparativas Entre la Retroalimentación RLAIIF Y RLHF* (Lee et al., 2024; sapien, 2024)

<b>Criterios</b>	<b>RLHF</b>	<b>RLAIIF</b>
Fuente	proporcionada por humanos	generada por modelos de IA
Escalabilidad	limitada por la naturaleza humana	alta por la automatización
Costo	alto por necesidad anotadores humanos	bajo por la automatización
Matiz	capacidad de captar matices humanos	criterios algorítmicos
Uso	decisiones éticas y alineadas con valores humanos	sistemas automatizados a gran escala
Velocidad	más lento por la anotación humana	rápido por la generación automática de respuestas
Sesgos	sujeto a la naturaleza humana	depende del modelo de IA utilizado
Rendimiento	enfoque humano, considerada menos eficiente a gran escala	rendimiento comparable o superior a RLHF en ciertas tareas

### **Método**

La metodología de la investigación tiene un enfoque integral y sistemático. Se buscó obtener y presentar información precisa y útil mediante una combinación de metodologías

documentales para proporcionar respuestas claras y efectivas. La investigación combinó revisión de literatura, análisis de contenido, referencia cruzada, uso de herramientas de inteligencia artificial y citación de fuentes confiables. Esta aproximación integral asegura que la información sea exacta y relevante, minimizando errores al verificar datos a través de múltiples fuentes. Además, el uso de tecnologías avanzadas permite procesar grandes volúmenes de datos de manera eficiente, lo que resulta en respuestas rápidas. Al sintetizar y resumir la información de manera clara y concisa, se facilitó la comprensión del usuario, aumentando la confianza en la información presentada gracias a la citación de fuentes respetadas. Esta metodología brindó un enfoque sólido, ajustado al investigador, mejorando la calidad y utilidad de las respuestas (Sobrido Prieto & Rumbo-Prieto, 2018).

Primero se examinaron y analizaron diversas fuentes escritas para reunir información relevante. La información fue interpretada y categorizada en tablas para identificar patrones distintivos, referenciando datos de múltiples fuentes para asegurar precisión y coherencia. Estas técnicas cualitativas permitieron proporcionar respuestas detalladas, coherentes y basadas en evidencia, aprovechando el análisis de la información disponible.

### **Resultados**

La siguiente tabla resume la información derivada de estudios recientes de RLHF y RLAIIF. La tabla muestra la descripción de la metodología utilizada en las investigaciones seleccionadas y una síntesis de los datos y conclusiones obtenidos en cada estudio de los dos modelos de retroalimentación.

**Tabla 2**

*Comparativa de las características de RLHF y RLAI: Análisis basado en los datos de cada investigación*

<b>Título</b>	<b>Metodología</b>	<b>RLHF</b>	<b>RAIF</b>	<b>Conclusión</b>
<i>RLAIF vs. RLHF: Scaling Reinforcement Learning from Human Feedback with AI Feedback</i> (Lee et al., 2022)	experimental, comparativa	excelente desempeño	reducción de costos, generación de diálogos útiles, d-RLAIF logra una tasa de éxito del 66 % en la muestra	RLAIF logra mejoras comparables a RLHF en tres tareas de generación de texto.
<i>Comparison Between RLHF and RLAI in Fine-Tuning a Large Language Model</i> (Khedri & Höglund, 2023)	revisión sistemática, se recopila, analiza y categoriza una amplia gama de técnicas de	comprensión más profunda de los matices y valores humanos	recibió una alta aceptación, amigable, aumenta la eficiencia y escalabilidad, pero carecer de la sutileza humana	RLAIF alternativa prometedora para el ajuste fino de modelos de lenguaje grandes, busca escalabilidad y eficiencia. RLHF es valioso para aplicaciones de comprensión profunda de las preferencias humanas.
<i>A Comprehensive Survey of LLM Alignment Techniques: RLHF, RLAI, PPO, DPO and More</i> (Zhichao et al., 2024)	revisión sistemática, recopilación y análisis con una gama de técnicas de alineación de modelos de lenguaje (LLM) paara una visión integral del campo.	técnica basada en modelos de recompensa para alinear los LLMs con las preferencias humanas;	variante generada por IA, útil en escenarios donde la retroalimentación humana es limitada	RLHF es ideal para aplicaciones que requieren una comprensión profunda de las preferencias humanas, mientras que RLAI es más adecuado para situaciones donde la velocidad y la escalabilidad son cruciales.
<i>Generative Reward Models</i> (Dakota,2024)	evalua su efectividad con métricas cuantitativas, datos empíricos y observaciones controladas	costoso por requerir mucha intervención humana	las etiquetas sintéticas no siempre se alinean bien con la realidad del usuario	propone un enfoque híbrido que unifica las metodologías RLHF y RLAI

## Discusión de resultados

La característica óptima que predomina en las investigaciones presentadas para el modelo RLHF es la aplicación de comprensión de las preferencias humanas. La particularidad óptima para el modelo RLAIIF se enfoca a situaciones donde la escalabilidad y la eficiencia son las mejores alternativas para ambientes donde la velocidad y el procesamiento masivo son fundamentales.

¿Cuáles son los desafíos que presentan ambos modelos? Los principales desafíos identificados giran en torno a las siguientes interrogantes: ¿cómo mantener la calidad de la alineación con valores humanos mientras se minimizan los costos?; ¿cómo mejorar la generalización de RLAIIF para tareas subjetivas o ambiguas?; ¿cómo diseñar sistemas flexibles que integren retroalimentación humana e IA según las necesidades específicas?

Las investigaciones revelan que la selección entre RLHF y RLAIIF constituye un *trade-off* fundamental entre calidad interpretativa y eficiencia operacional. Por un lado, RLHF logra una comprensión de los matices humanos, como se menciona en Khedri & Höglund (2023) y Lee et al. (2022). Sin embargo, la intervención humana, incrementa los costos y limita su escalabilidad, tal como se indica en Dakota (2024). Por otra parte, RLAIIF reduce costos y mejora la eficiencia, como se observa en Lee et al. (2022) y Zhichao et al. (2024), pero podría carecer de la sutileza humana y generar etiquetas no alineadas con la realidad del usuario, según Khedri & Höglund (2023) y Dakota (2024). Un desafío adicional es la alineación con preferencias humanas complejas. RLHF es superior en aplicaciones que requieren interpretación de valores éticos o culturales, pero depende de la disponibilidad de humanos para generar retroalimentación, como se recalca en Zhichao et al. (2024). Khedri & Höglund (2023) advierten que la naturaleza basada

en IA de RLAIIF, si bien permite escalabilidad y velocidad, conlleva riesgos significativos de propagación de sesgos. Para mejorar la generalización de RLAIIF en tareas subjetivas o ambiguas, se enfrenta el reto de superar a RLHF en tareas que requieren empatía o creatividad. Las etiquetas sintéticas en RLAIIF pueden ser inconsistentes con expectativas reales, como se observa en Dakota (2024).

En cuanto a soluciones, Dakota (2024) propone un modelo híbrido RLHF-RLAIIF, combinando lo mejor de ambos métodos. Sin embargo, esto introduce complejidad en el diseño y entrenamiento. La elección entre RLHF y RLAIIF depende del contexto: escalabilidad versus precisión humana, como se discute en Zhichao et al. (2024). Este enfoque híbrido puede ayudar a equilibrar los costos y la calidad, pero requiere una cuidadosa evaluación de las necesidades específicas de cada aplicación.

### **Conclusiones**

La mayoría de las investigaciones manifiestan una inclinación preferencial hacia el modelo RLAIIF por su escalabilidad (Khedri & Höglund, 2023; Lee et al., 2022 & Zhichao et al., 2024). Sin embargo, los investigadores del estudio titulado *Generative Reward Models* proponen un enfoque híbrido, integrando ambas alternativas (Dakota, M.,2024). El RLAIIF o RLHF proporcionan una solución escalable y automatizada, que puede complementar la enseñanza tradicional. Ambos enfoques se complementan para mejorar el proceso de aprendizaje.

### **Opinión personal**

A medida que la IA continúa evolucionando, es probable encontrar diferentes innovaciones, optimizando la eficacia, precisión y la ética de los sistemas de IA. La personalización del aprendizaje y la toma de decisiones éticas en aplicaciones donde interviene

la IA continuarán siendo un desafío. Sin embargo, la realidad es evidente. La comercialización de la educación ha ido en aumento (González, 2013), por lo tanto, el mercado no permite tomar las mejores herramientas ni las más óptimas para la enseñanza y aprendizaje al priorizar la rentabilidad y la competencia sobre la calidad intrínseca del aprendizaje.

### Referencias

Data Camp. (2024, 29 julio). *RLAIIF: ¿Qué es el aprendizaje por refuerzo a partir de la retroalimentación de la IA?* Datacamp.com.

<https://www.datacamp.com/es/blog/rlaifreinforcement-learning-from-ai-feedback>

Dakota, M., Van Phung, D., Rafael, R., Chase, B., Nathan, L., Louis, C., Jan-Philipp, F., Chelsea, F., & Alon, A. (2024). Generative Reward Models. *arXiv [cs.LG]*.

<https://arxiv.org/html/2410.12832v1>

Escuela de profesores del Perú. (2024, mayo 10). *¿Qué es la retroalimentación?, mejores formas de retroalimentación*. Escuela de Profesores Del Perú. [https://epperu.org/que-es-la-](https://epperu.org/que-es-la-retroalimentacion-mejores-formas-de-retroalimentacion/)

[retroalimentacion-mejores-formas-de-retroalimentacion/](https://epperu.org/que-es-la-retroalimentacion-mejores-formas-de-retroalimentacion/)

González, A. (2013, julio 29). El gran negocio de la educación. *Colima medios*.

<https://colimamedios.com/el-gran-negocio-de-la-educacion/>

Herrera-Araya, David. (2023). Perspectivas e investigación reciente sobre retroalimentación en el aula: Consideraciones para un enfoque pedagógico y dialógico. *Revista Electrónica*

*Educare*, 27(1), 589-608. <https://dx.doi.org/10.15359/ree.27-1.14547>

- Khedri, J., & Höglund, S. (2023). *Comparison Between RLHF and RLAIIF in Fine-Tuning a Large Language Model School of Electrical Engineering and Computer Science (EECS)*.  
<https://www.diva-portal.org/smash/get/diva2:1782683/FULLTEXT01.pdf>
- Lee, H., Phatale, S., Mansoor, H., Mesnard, T., Ferret, J., Lu, K., Bishop, C., Hall, E., Carbune, V., Rastogi, A., Prakash, S. (2024, julio 21). *RLAIIF vs. RLHF: Proceedings of the 41st International Conference on Machine Learning*. Guide Proceedings.  
<https://dl.acm.org/doi/10.5555/3692070.3693141>
- Luna Acuña, M. L., Peralta Roncal, L. E., Gaona Portal, M. del P., & Dávila Rojas, O. M. (2022). La retroalimentación reflexiva y logros de aprendizaje en educación básica: una revisión de la literatura. *Ciencia Latina Revista Científica Multidisciplinar*, 6(2), 3242-3261.
- Nanobaly. (2024). *Aprendizaje RLHF para LLM y otros modelos*. Innovatiana.com.  
<https://es.innovatiana.com/post/rlhf-our-detailed-guide>
- Núñez-Valdés, K., Gerson Núñez-Valdés, & Castillo-Paredes, A. (2024). Retroalimentación en el contexto educativo: Una revisión sistemática. *Formación Universitaria*, 17(2), 61–72.  
<https://doi.org/10.4067/s0718-50062024000200061>
- sapien. (2024). *RLAIIF vs. RLHF: A Detailed Comparison of AI Training Methods*. Sapien.io.  
<https://www.sapien.io/blog/rlaif-vs-rlhf-understanding-the-differences>
- Sobrido Prieto, M., & Rumbo-Prieto, J. M. (2018). La revisión sistemática: pluralidad de enfoques y metodologías. *Enfermería Clínica*, 28(6), 387–393.  
<https://doi.org/10.1016/j.enfcli.2018.08.008>

Umáximo. (s.f.). *Retroalimentación educativa: ¿Cómo transformar el aprendizaje*

*escolar?* Umaximo.com. Recuperado 7 marzo, 2025, de

<https://www.umaximo.com/post/retroalimentacion-educativa-como-transformar-el-aprendizaje-escolar>

Vocerol. (2018). *Automatización de la retroalimentación: ¿Cómo la IA puede mejorar la*

*evaluación del desempeño en los sistemas de gestión del aprendizaje?* Psicosmart.pro.

<https://psicosmart.pro/articulos/articulo-automatizacion-de-la-retroalimentacion-como-la-ia-puede-mejorar-la-evaluacion-del-desempeno-en-los-sistemas-de-gestion-del-aprendizaje-202980>

*What is a Large Language Model (LLM)*. (2023, junio 4). GeeksforGeeks.

<https://www.geeksforgeeks.org/large-language-model-llm/>

Zhichao, W., Bin, B., Pentyala, S. K., Kiran, R., Sougata, C., Shubham, M., Zixu, Zhu, Xiang-

Bo, M., Sitaram, A., Na, & Cheng. (2024). A Comprehensive Survey of LLM Alignment

Techniques: RLHF, RLAIIF, PPO, DPO and More. En *arXiv [cs.CL]*.

<http://arxiv.org/abs/2407.16216>